การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสาม บนฟังก์ชันพลังงานเสียง และสเปกตรัมแอลพีซี บนสเกลบาร์กโดยใช้เทคนิค KNN

24 VOWELS RECOGNITION IN THAI SPOKEN LANGUAGE USING 3rd ORDER POLYNOMIAL COEFFICIENT CURVE FITTING ON THE VOICE ENERGY AND SPECTRUM LPC ON BARK SCALE BY KNN TECHNIQUE

พีรพัฒน์ กุลทัพ

PEERAPAT KULTAP

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2548

ISBN 974-15-1614-2

การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียง และสเปกตรัมแอลพีซี บนสเกลบาร์กโดยใช้เทคนิค KNN

24 VOWELS RECOGNITION IN THAI SPOKEN LANGUAGE USING 3rd ORDER POLYNOMIAL COEFFICIENT CURVE FITTING ON THE VOICE ENERGY AND SPECTRUM LPC ON BARK SCALE BY KNN TECHNIQUE

พีรพัฒน์ กุลทัพ

PEERAPAT KULTAP

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2548

ISBN 974-15-1614-2

24 VOWELS RECOGNITION IN THAI SPOKEN LANGUAGE USING 3rd ORDER POLYNOMIAL COEFFICIENT CURVE FITTING ON THE VOICE ENERGY AND SPECTRUM LPC ON BARK SCALE BY KNN TECHNIQUE

PEERAPAT KULTAP

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING
SCHOOL OF GRADUATE STUDIES
KINGMONGKUT INSTITUTE OF TECHNOLOGY LADKRABANG
2005

ISBN 974-15-1614-2

หัวข้อวิทยานิพนธ์	การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสามบนฟังก์ชั่นพลังงานเสียงและสเปกตรัมแอลพีซีบนสเกลบาร์กโดยใช้เทคนิค KNN
นักศึกษา	นายพีรพัฒน์ กุลทัพ
รหัสประจำตัว	43061057
ปริญญา	วิศวกรรมศาสตร์มหาบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้า
พ.ศ.	2548
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร. ไกรสิน ส่งวัฒนา

บทคัดย่อ

วิทยานิพนธ์นี้จะนำเสนอ วิธีการรู้จำหน่วยเสียงสระ 24 เสียงในภาษาไทย ซึ่งประกอบด้วยสระเดี่ยว เสียงสั้นจำนวน 9 เสียง (อะ อิ อึ อุ โอะ เอะ แอะ เออะ เอาะ) สระเดี่ยวเสียงยาวจำนวน 9 เสียง (อา อี อือ อู โอ เอ แอ เออ ออ) สระผสมเสียงสั้นจำนวน 3 เสียง (เอียะ เอือะ อัวะ) และ สระผสมเสียงยาวจำนวน 3 เสียง (เอีย เอือ อัว)

สัญญาณเสียงจะถูกนำมาคำนวณหาค่าสัมประสิทธิ์การฟิตเคิร์ฟแบบพหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียง แล้วใช้วิธีการ KNN เพื่อแยกสระเสียงสั้น สระเสียงยาว จากนั้นจะถูกนำมาคำนวณหาค่าสเปกตรัมแอลพีซีแล้วแปลงให้อยู่บนสเกลบาร์ก นำค่าสเกลบาร์กที่ได้มาคำนวณหาค่าความเข้มในสัญญาณแต่ละความถี่วิกฤตเพื่อนำไปใช้สร้างแบบอ้างอิงการรู้จำสระเดี่ยวเสียงสั้น 9 โมเดล และสระเดี่ยวเสียงยาว 9 โมเดล แล้วนำเสียงทดสอบที่เหลือมาทดสอบกับแบบจำลองใช้กฎการตัดสินใจอย่างง่ายในการบ่งชี้เสียงสระแต่ละเสียง

Thesis Title	24 Vowels Recognition in Thai Spoken Language Using 3rd Order Polynomial Coefficient Curve Fitting on the Voice Energy and Spectrum LPC on Bark Scale by KNN technique
Student	Mr. Peerapat Kultap
Student ID.	43061057
Degree	Master of Electrical Engineering
Programme	Electrical Engineering
Year	2005
Thesis Advisor	Assoc.Prof.Dr. Kraisin Songwatana

ABSTRACT

This thesis presents 24 vowels recognition for Thai spoken language. The Thai language consist of 9 short unmixed vowels (a, i, ω, u, o, e, ε, γ, ɔ̇) , 9 long unmixed vowels (aa, ii, ωω, uu, oo, ee, εε, γγ, ɔ̇ɔ̇) , and 3 short mixed vowels (ia, ωa, ua) and 3 long mixed vowels (i:a:, ω:a:, u:a:)

The coefficient of 3rd Order polynomial curve fitting is calculated for the voice energy. It is then used to classifying short and long vowels by KNN technique. The LPC spectrum for the each frame of the audio signal are then calculated and transformed on to the Bark Scale to determine the Critical Band Intensity. 9 reference models for short unmixed vowels and 9 reference models for long unmixed vowels recognition are created. A decision rule is applied to identify mixed and unmixed vowel in Thai language.

กิตติกรรมประกาศ

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

ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆให้คำแนะนำ คำปรึกษาที่ดีมาโดยตลอดระยะเวลาที่ศึกษาอยู่ให้แก่ศิษย์

ขอขอบคุณ พี่ๆ เพื่อนๆ และน้องๆ ทุกท่านทั้งที่ห้องวิจัย ที่ทำงาน และที่เรียนด้วยกัน สำหรับความเป็นมิตร และความช่วยเหลือต่างๆ ที่มีให้กันระหว่างศึกษาอยู่

ขอขอบคุณเจ้าของเสียงทุกๆ ท่าน ที่ได้เสียสละเวลามาบันทึกเสียงให้

สุดท้ายขอกราบขอบพระคุณ บิดา มารดา และครอบครัว ที่ให้ทุกสิ่งทุกอย่าง กับลูก

ประโยชน์อันถึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้เขียนขอมอบแด่ผู้มีพระคุณทุกท่าน

ขอขอบคุณด้วยใจจริง

พีรพัฒน์ กุลทัพ

สารบัญ

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

สารบัญตาราง (ต่อ)

สารบัญรูป

รูปที่ หน้า

สารบัญรูป (ต่อ)

รูปที่ หน้า

สารบัญรูป (ต่อ)

บทที่ 1

บทนำ

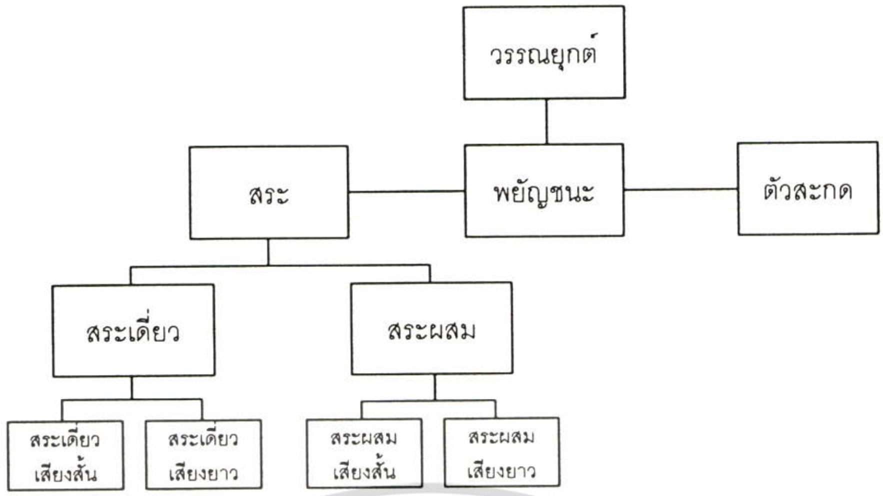
1.1 ความเป็นมาและความสำคัญของปัญหา

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

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

1.1 พิจารณาทั้งหน่วยภาษาที่เปล่งเสียงออกมาทั้งชุด มีทั้งระบบการรู้จำคำพูดเดี่ยว (Isolated word Recognition) และระบบรู้จำคำพูดต่อเนื่อง(Continuous word Recognition) ซึ่งข้อดีของระบบเหล่านี้คือง่าย เนื่องจากมีการหลีกเลี่ยงผลกระทบอันเนื่องจากฐานรากของเสียงภายในคำหรือกลุ่มคำนั้น แต่ข้อเสียคือสามารถรู้จำคำได้จำนวนคำที่จำกัด เนื่องจากต้องใช้เนื้อที่จำนวนมากในการจัดเก็บแบบจำลองอ้างอิง และต้องใช้เวลาในการคำนวณเพื่อเปรียบเทียบมากตามจำนวนของแบบจำลองอ้างอิงที่มีอยู่

1.2 พิจารณาโดยการแยกแยะรายละเอียดของหน่วยเสียง(Phonetic Recognition) วิธีนี้จะพิจารณาลักษณะของหน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ ดังแสดงในรูปที่ 1.1 โดยจะใช้หน่วยเสียงย่อยเหล่านี้เป็นหลักในการรู้จำเสียงพูด ซึ่งวิธีนี้เหมาะสำหรับการพัฒนาไปสู่ระบบการรู้จำคำจำนวนมาก



รูปที่ 1.1 ส่วนประกอบของคำในภาษาไทยโดยแยกตามลักษณะของหน่วยเสียง

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1.2.1 เพื่อศึกษาวิธีการแยกแยะสระเสียงสั้น และ สระเสียงยาว

1.2.2 เพื่อศึกษาวิธีการหาพลังงานของเสียง ที่เหมาะสมที่จะนำมาใช้ในการแยกสระเสียงสั้นและสระเสียงยาว

1.2.3 เพื่อหาตัวแปร หรือค่า พารามิเตอร์ ใดๆ ที่จะใช้เป็นตัวอ้างอิง ของความต่างทางพลังงานเสียงสระเสียงสั้น และ สระเสียงยาว

1.2.4 เพื่อศึกษาวิธีการแยกแยะเสียงสระเดี่ยวและเสียงสระผสมในภาษาไทย

1.2.5 เพื่อพัฒนาระบบการรู้จำเสียงพูดในภาษาไทย โดยใช้หน่วยเสียงสระ

1.2.6 เพื่อลดเนื้อที่หน่วยความจำในการเก็บแบบอ้างอิงและลดเวลาในการเปรียบเทียบกับแบบอ้างอิง

1.2.7 เป็นพื้นฐานในการรู้จำเสียงเพื่อพัฒนาไปสู่การรู้จำเสียงในภาษาไทยทั้งระบบคำพูดเดี่ยวและคำพูดต่อเนื่องต่อไป

1.3 สมมติฐานของการศึกษา

คำในภาษาไทยประกอบด้วยหน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ และอาจมีตัวสะกด ซึ่งก็คือหน่วยเสียงพยัญชนะ คำในภาษาไทยนั้นจะมีหน่วยเสียงสระเป็นองค์ประกอบที่สำคัญที่สุดเนื่องจากหน่วยเสียงสระจะเป็นส่วนที่มีมากที่สุดของคำหากมองในเชิงเวลา และเนื่องจากในภาษาไทยมีทั้งสระเดี่ยว สระผสม และยังแบ่งเป็นสระเสียงสั้นและสระเสียงยาว แต่องค์ประกอบของเสียงสระทั้งหมดนั้นก็เกิดจากส่วนย่อยที่สุดของเสียงสระก็คือเสียงสระเดี่ยว ดังนั้น

การรู้จำเสียงสระสำหรับภาษาไทยทั้งหมดนั้นก็จะต้องเริ่มต้นจากการรู้จำเสียงสระเดี่ยว ซึ่งงานวิจัยนี้ได้นำการรู้จำสระเสียงเดี่ยวมาสร้างให้เกิดเป็นการรู้จำสระเสียงผสมขึ้น

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในการวิจัย

เสียงสระในภาษาไทยมีทั้งหมด 12 คู่เสียง(สั้น,ยาว) ประกอบด้วยเสียงสระเดี่ยว 9 คู่เสียง (สั้น ยาว) คือ “(อะ อา) (อิ อี) (อึ อือ) (อุ อู) (โอะ โอ) (เอาะ ออ) (เอะ เอ) (แอะ แอ) (เออะ เออ)” การแยกแยะสระเสียงสั้นและสระเสียงยาว โดยจะใช้พลังงานเสียงเป็นตัวบ่งชี้ โดยที่พลังงานเสียงจะเป็นตัวบอกถึงลักษณะการเปล่งเสียงตลอดคำว่ามีลักษณะเป็นอย่างไร ซึ่งจะทำให้เห็นถึงช่วงของการหยุด หรือ ช่วงการจบคำ ว่ามีลักษณะเป็นอย่างไร และ หาค่าพารามิเตอร์ ที่ช่วยบ่งชี้ความต่างดังกล่าว เพื่อให้สามารถแยกแยะสระเสียงสั้น และ สระเสียงยาว ได้ และสระเสียงเดี่ยวเป็นส่วนย่อยที่สุดของเสียงสระทั้งหมดในภาษาไทย ดังนั้นตามลักษณะของการออกเสียงแล้ว เสียงสระผสมเกิดจากการนำเสียงสระเดี่ยว 2 เสียงมารวมกัน ซึ่งเสียงสระผสมในภาษาไทยมีทั้งหมด 3 คู่เสียง(สั้น ยาว) คือ เสียงสระ(อัวะ อัว) เกิดจากเสียงสระ “(อุ อู)+(อะ อา)” เสียงสระ(เอียะ เอีย) เกิดจากเสียงสระ “(อิ อี)+(อะ อา)” และเสียงสระ (เอือะ เอือ) เกิดจากเสียงสระ “(อึ อือ)+(อะ อา)” เมื่อสามารถแยกสระเสียงสั้น สระเสียงยาว และรู้จำเสียงสระเสียงเดี่ยวและสระเสียงผสมได้ เราก็สามารถที่จะรู้จำเสียงสระทั้ง 24 เสียงในภาษาไทยได้

1.5 ขอบเขตการวิจัย

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

1.6 ขั้นตอนการศึกษา

1.6.1 ศึกษาระบบการรู้จำภาษาในลักษณะต่างๆ ที่ได้มีการศึกษามาแล้ว

1.6.2 ศึกษาทฤษฎีทางด้านภาษาศาสตร์ และอัลกอลิธึมต่างๆที่ใช้ในการวิเคราะห์เสียงพูด

1.6.3 กำหนดขอบเขตการวิจัย

1.6.4 บันทึกข้อมูลเสียง โดยใช้ผู้พูดทั้งหมด 8 คน เพศชาย 4 คน เพศหญิง 4 คน

1.6.5 เก็บข้อมูล วิเคราะห์เสียง และทดสอบความถูกต้อง

1.6.6 สรุปผลการวิจัย และเสนอแนะแนวทางในการวิจัยต่อไป

บทที่ 2
ทฤษฎีการสร้างเสียงพูดในภาษาไทย

2.1 กล่าวนำ

ในบทนี้จะกล่าวถึงทฤษฎีต่างๆที่เกี่ยวข้องกับงานวิจัยนี้ โดยจะเริ่มจากทฤษฎีทางภาษาศาสตร์ ซึ่งจะกล่าวถึงเรื่องทฤษฎีการสร้างเสียงพูดประกอบด้วย หน่วยเสียงสระ หน่วยเสียงพยัญชนะ หน่วยเสียงวรรณยุกต์ และลักษณะคำ และพยางค์ ในภาษาไทย จากนั้นจะกล่าวถึงทฤษฎีการวิเคราะห์เสียงประกอบด้วย การคำนวณสเปกตรัมแอลพีซี แถบความถี่วิกฤต และความเข้มสัญญาณในแถบความถี่วิกฤต

2.2 ทฤษฎีภาษาศาสตร์

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีเส้นเอ็นในลำคอที่เรียกว่าเส้นเสียง (Vocal Cords) ซึ่งอยู่ในช่องหลอดลม และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่างๆ ผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ปากหรือช่องจมูกออกไปภายนอกซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี้จะเป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (Pass Through) หรือคลื่นเสียงความถี่ไหนที่จะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (Absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้ เนื่องจากรูปร่างของช่องคอและช่องปาก ของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้อย่างอิสระ ทำให้ผู้พูดมีความสามารถในการกำหนดคลื่นเสียงระดับต่างๆ เหล่านี้นั้นเองที่ทำให้เกิดการแปลงเป็นเสียงพูดออกมา

2.2.1 อวัยวะออกเสียง (Organs of Speech)

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

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

2.2.1.1 อวัยวะที่เป็นส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่างๆ อวัยวะที่เป็นส่วนกระทำการที่สำคัญ คือลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

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

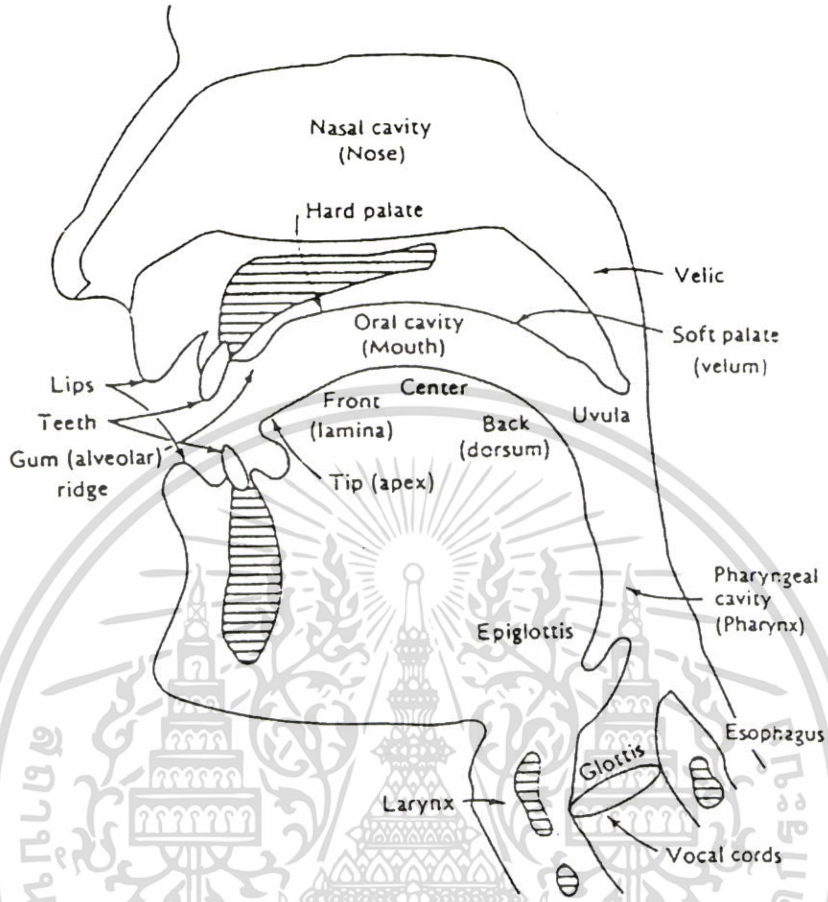
อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูดมีดังแสดงในตารางที่2.1

ตารางที่ 2.1 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูด

อวัยวะ	ลักษณะ / หน้าที่
ริมฝีปาก (Lips)	เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกมา หรือห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่างๆ นี้ ล้วนมีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น
ฟัน (Teeth)	เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่างหรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอผ่านได้ออกมาทำให้เกิดเป็นเสียงชนิดที่เรียกว่าเสียงเสียดแทรกที่เกิดที่ฟันเป็นต้น
ปุ่มเหงือก (Gum ridge)	เป็นส่วนที่อยู่ต่อจากฟันบน
เพดานแข็ง หรือเพดานปาก (Hard palate)	เฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น
เพดานอ่อน (Soft palate)	เป็นส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอ นอกจากเวลาออกเสียงนาสิก เท่านั้นที่เพดานอ่อนจะลดระดับลงมา
ลิ้น (Tongue)	เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งเป็น 3 ส่วนด้วยกันตามหน้าที่ในการออกเสียง
ปลายลิ้น (Blade of the tongue)	เป็นส่วนปลายซึ่งสามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้ง่าย

ตารางที่ 2.1 (ต่อ)

อวัยวะ	ลักษณะ / หน้าที่
หน้าลิ้น (Front of the tongue)	หมายถึงลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง ขณะที่วางลิ้นราบกับปากตอนที่ไม่ได้พูด
หลังลิ้น (Back of the tongue)	หมายถึง ส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ขณะที่วางลิ้นราบกับปาก
ช่องปาก (Oral cavity)	ช่องปากนี้เป็นอวัยวะอีกส่วนหนึ่งที่มีความสำคัญในการพูด กล่าวคือ ช่องปากจะทำหน้าที่เป็นช่องกำทอน (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

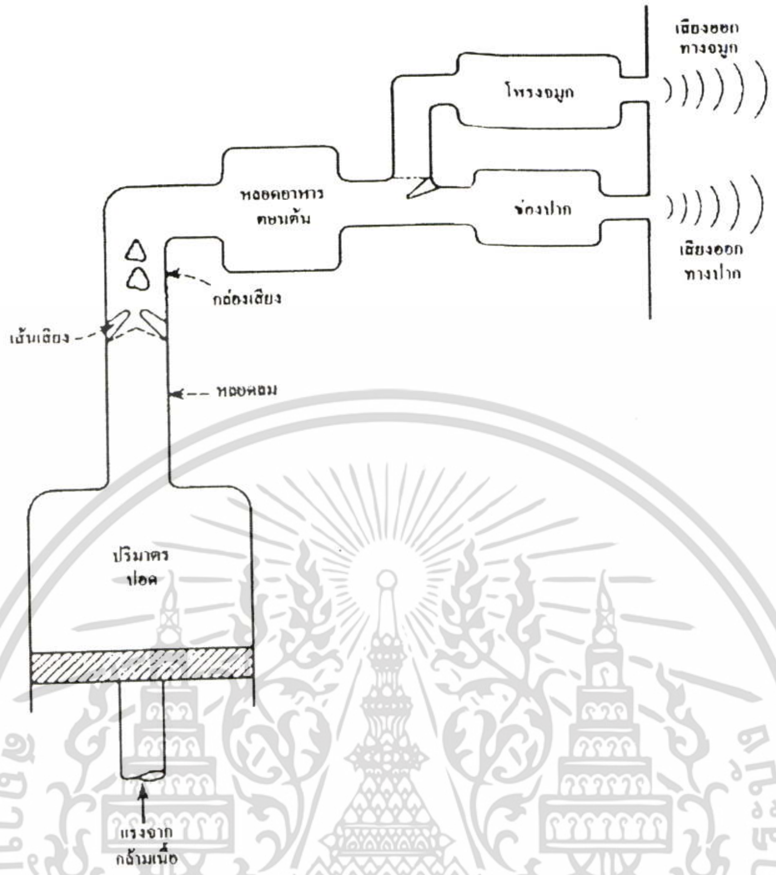
2.2.2 กระบวนการผลิตเสียงพูด

จากระบบเสียงพูดของมนุษย์ สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียงดังรูปที่ 2.2 ซึ่งเราสามารถจำแนกกลไกสร้างเสียงพูด ของมนุษย์ได้ 3 แบบ ดังนี้

2.2.2.1 อากาศไหลจากปอดจะถูกมอดูเลท (Modulate) โดยการสั่นของเส้นเสียงทำให้เกิดเคลื่อนเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (Quasi - periodic pulse - like excitation)

2.2.2.2 อากาศไหลจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (Noise – like excitation)

2.2.2.3 อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศที่มีแรงดันพุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (Transient excitation)



รูปที่ 2.2 แผนภาพระบบเสียงพูดของมนุษย์

2.2.3 เสียงพูดของมนุษย์

เสียงพูดเป็นคลื่นความยาว (Longitudinal wave) เกิดจากการสั่นของอนุภาคตัวกลาง นั้นคือ อากาศ และทิศทางการสั่นของอนุภาคจะอยู่ในทิศเดียวกันกับทิศทางของการเคลื่อนที่ คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงไปตามเวลา

เสียงพูดแบ่งได้เป็น 2 ชนิด ตามการกำเนิดเสียงหรือโหมด (Mode) การกระตุ้น คือ

2.2.3.1 เสียงวอยซ์ หรือเสียงโฆษะ (Voice sound) เกิดจากการบังคับอากาศให้ผ่านสายเสียง ทำให้มีการเปลี่ยนแปลงความตึงหย่อนของเส้นเสียง โดยเส้นเสียงจะสั่นและเกิดเป็นพัลส์ (Pulse) ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงวอยซ์ ตัวอย่างเสียงวอยซ์ ได้แก่ เสียงสระ เสียงพยัญชนะที่ออกเสียงจากลำคอ (Voice consonants)

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

กว้าง (Broad – spectrum noise) ตัวอย่างเสียงอันวอยซ์ ได้แก่ เสียงพยัญชนะที่ไม่ได้เกิดจากลำคอ (Voiceless consonants)

2.2.4 ความยาวของเสียง (Length)

การเกิดเสียงใดเสียงหนึ่งเปล่งออกมาได้นานเท่าใด เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระ เสียงพยัญชนะ หรือเสียงพยัญชนะเสียดแทรก

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

2.2.5 ระดับเสียง สูง – ต่ำ (Pitch)

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

ในการพูดเสียงที่มีระดับ สูง – ต่ำ ได้คือเสียงก้องเท่านั้น เพราะมีการสั่นสะเทือนของเส้นเสียงที่ทำให้เกิดมีความถี่ระดับต่างๆ ได้ ในภาษาไทยระดับเสียง สูง – ต่ำ ของคำเราเรียกว่า "วรรณยุกต์"

2.2.6 ความดัง (Loudness)

ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ

2.2.7 การลงน้ำหนัก (Stress)

การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่ข้างเคียง (เพื่อต้องการเรียกร้องความสนใจเป็นพิเศษ หรือแสดงอารมณ์ อย่างใดอย่างหนึ่ง)

2.2.8 ช่วงต่อของเสียง (Juncture)

ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งเสียงอีกเสียงหนึ่ง 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 ถ้าเราพิจารณาอย่างถ้วนๆ จะพบว่าสามารถแบ่งออกเป็น 2 ประเภทใหญ่ ดังตารางที่ 2.2

ตารางที่2.2 ลักษณะเสียงพูดของมนุษย์

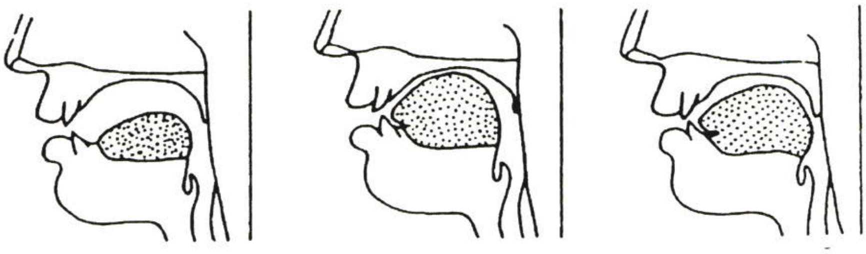
เสียง	ลักษณะ
เสียงเรียง (Segmental sound)	เป็นหน่วยเสียงที่แยกออกจากเสียงอื่นได้โดยเด็ดขาด เพราะมีลักษณะเด่นเฉพาะตัว ในภาษาไทยได้แก่เสียงสระ และเสียงพยัญชนะ
เสียงซ้อน (Supra – segmental feature)	เป็นเสียงที่ทำหน้าที่เป็นส่วนประกอบของเสียงอื่น เพราะไม่สามารถแยกเปล่งเสียงได้ตามลำพัง ในภาษาไทยได้แก่เสียงวรรณยุกต์ และทำนองเสียงเป็นต้น

2.3.1 หน่วยเสียงสระ

2.3.1.1 ลักษณะของเสียงสระ

ลักษณะสำคัญของเสียงสระก็คือ “ เป็นเสียงก้องที่เปล่งเสียงออกมาโดยให้ลมออกทางช่องปากโดยไม่ถูกลิ้นกักหรือขัดขวาง” ดังนั้นเวลาเราออกเสียงสระจะออกเสียงได้สะดวกและออกเสียงได้นาน ทั้งนี้เพราะคุณสมบัติของเสียงสระมีความดังเด่นกว่าเสียงอื่นๆ ที่เรียงอยู่ข้างเสมอ อวัยวะที่เกี่ยวข้องกับการออกเสียงสระได้แก่ลิ้น กับริมฝีปาก ถ้าลิ้นส่วนใดทำหน้าที่เพียงเพียงส่วนเดียวเสียงที่เกิดขึ้นก็จะมีเพียงเส้นเสียงเดียว 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) ส่วนหลังอยู่ใต้เพดานอ่อน (Velum) และส่วนกลางอยู่ระหว่างกลางของสองส่วนนี้ รูปต่อไปนี้แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำส่วนหน้าสูง ส่วนหลังสูง



รูปที่ 2.3 แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง ส่วนหลังสูง

สำหรับภาษาไทยมีหน่วยเสียงสระทั้ง 24 หน่วยเสียง แยกออกเป็นสระเดี่ยว 18 หน่วยเสียง และสระผสม 6 หน่วยเสียง

2.3.1.1.1สระเดี่ยว

เสียงสระเดี่ยว 18 หน่วยเสียง พิจารณาการเกิดเสียงได้เป็น 2 กรณีใหญ่ๆ คือ

2.3.1.1.1.1 การเกิดจากส่วนต่างๆ ของลิ้นหมายถึง ลมผ่านส่วนหน้า ส่วนกลาง หรือส่วนหลังของลิ้น

2.3.1.1.1.2 การเกิดจากลมผ่านลิ้นในขณะที่ลิ้นอยู่ในระดับ สูง กลาง หรือ ต่ำ

ตารางที่ 2.3 เสียงสระเดี่ยวในภาษาไทย

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

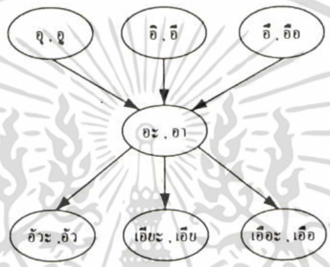
นอกจากนี้ หน่วยเสียงสระเดี่ยว 18 หน่วย สามารถแบ่งตามความสั้น – ยาว ของการออกเสียงได้เป็น

ตารางที่ 2.4 เสียงสระเดี่ยวในภาษาไทยแบ่งตามความสั้น-ยาว

สระ	หน่วยเสียง
สระเดี่ยวเสียงสั้น	อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ
สระเดี่ยวเสียงยาว	อา อี อือ อู เอ แอ โอ ออ เออ

2.3.1.1.2 สระผสม

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



รูปที่ 2.4 แสดงการเกิดเสียงสระผสมในภาษาไทย

2.3.1.2 หน้าที่ของหน่วยเสียงสระในภาษาไทย

หน่วยเสียงสระในภาษาไทยทั้ง 24 หน่วยเสียงนี้ ทำหน้าที่เป็นแกนกลางของพยางค์ หรือคำ กล่าวคือ คำทุกคำในภาษาไทยจะต้องมีเสียงสระอยู่ด้วย และเสียงสระในภาษาไทยจะสามารถเกิดกับเสียงพยัญชนะต้นได้ทุกเสียง และสามารถเกิดกับหน่วยเสียงวรรณยุกต์ได้ทุกหน่วย แต่ไม่สามารถเกิดกับหน่วยเสียงพยัญชนะสะกดได้ทุกหน่วย หน่วยเสียงสระที่ทำให้เกิดคำหรือพยางค์ใช้ได้มากที่สุดในภาษามักเป็นหน่วยเสียงยาว

2.3.2 หน่วยเสียงพยัญชนะ

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

2.3.2.1 ลักษณะของเสียงพยัญชนะ

หน่วยเสียงพยัญชนะ 21 หน่วยเสียงนี้จำแนกเป็นเสียงก้อง เสียงไม่ก้อง เสียงหนัก เสียงเบา และลักษณะการออกเสียง ดังนี้

ตารางที่ 2.5 เสียงพยัญชนะจำแนกตามการออกเสียง

ลักษณะเสียง	พยัญชนะ
เสียงก้อง (โฆษะ)	/ง//ย//บ//ด//ม//น//ร//ล//ว/
เสียงไม่ก้อง (อโฆษะ)	/ก//ค//จ//ช//ซ//ท//ต//ป//พ//ฟ//อ//ฮ/
เสียงหนัก (ธนิต)	/ค//ช//ท//พ/
เสียงเบา (ถวิล)	/ก//จ//ต//ป/

ตารางที่ 2.6 เสียงพยัญชนะ ในภาษาไทย

ที่เกิดเสียง / ลักษณะของเสียง		ริมฝีปาก	ปุ่มเหงือก	เพดานแข็ง	เพดานอ่อน	เส้นเสียง
พยัญชนะกัก	ไม่พ่นลม	ป*	ฏ, ต *	จ	ก*	อ*
	พ่นลม	ผ พ ภ	ฐ ฑ ฒ ถ ท ธ	ฉ ช ฌ	ข ค ฆ	
	ก้อง	บ	ฎ ด			
พยัญชนะไม่กัก	นาสิก	ม *	น ณ *		ง *	
	เสียดแทรก	ฝ ฟ	ซ ศ ษ ส			ฮ ห
	กระทบ		ร			
	ข้างลิ้น		ล ฬ			
	กึ่งสระ	ว *		ญ, ย*		

<u>หมายเหตุ</u> หน่วยเสียง พยัญชนะที่มีเครื่องหมายดอกจันทร์ (*) คือหน่วยเสียงที่สามารถปรากฏ ที่ตำแหน่งท้ายพยางค์ได้

2.3.2.2 หน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย

เสียงพยัญชนะในภาษาไทย 21 หน่วยเสียงนี้สามารถทำหน้าที่ได้ตามตารางที่ 2.7

ตารางที่ 2.7 แสดงหน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย

หน้าที่	พยัญชนะ
พยัญชนะต้นเดี่ยว	/ก//ค//ง//ย//บ//ด//ม//น//ร//ล//ว//จ//ช//ซ//ท//ต//ป//พ//ฟ//อ//ฮ/
พยัญชนะต้นควบ	หน่วยเสียงแรก /ก//ค//ต//ป/ และ /พ/ หน่วยเสียงที่สอง /ร//ล/ หรือ /ว/

ตารางที่ 2.7 (ต่อ)

หน้าที่	พยัญชนะ
พยัญชนะสะกดของพยางค์	/ ป / (แม่กบ) / ต / (แม่ กด) / ก / (แม่กค) / ม / (แม่กม) / ง / (แม่กง) / น / (แม่กน) / ว / (แม่เกอว) / ย / (แม่เกย) 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 นั่นเอง

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

ซึ่งลักษณะของเสียงวรรณยุกต์ สามารถแบ่งออกตามลักษณะระดับเสียงได้ตามตารางที่ 2.8

ตารางที่ 2.8 แสดงลักษณะของเสียงวรรณยุกต์ในภาษาไทย

หน่วยเสียง		สัญลักษณ์	เสียง	ลักษณะ
กลุ่มวรรณยุกต์ระดับ (Level tone)	หน่วยเสียงวรรณยุกต์ระดับ ต่ำ (Low tone)	/ \ /	เอก	หน่วยเสียงนี้จะปรากฏในพยางค์ของภาษาไทย ได้ทุกแบบ
	หน่วยเสียงวรรณยุกต์ระดับกลาง (Mid tone)	/ - /	สามัญ	หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มีตัวสะกดเป็น พยัญชนะกัก (พยางค์คำตาย)

ตารางที่ 2.8 (ต่อ)

หน่วยเสียง		สัญลักษณ์	เสียง	ลักษณะ
กลุ่มวรรณยุกต์ระดับ (Level tone) (ต่อ)	หน่วยเสียงวรรณยุกต์ระดับสูง (High tone)	/ / /	ตรี	หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่ผสมด้วยสระเสียงยาวซึ่งมีตัวสะกดเป็นเสียงกัก
กลุ่มวรรณยุกต์เปลี่ยนระดับ (Contour tone)	หน่วยเสียงวรรณยุกต์เปลี่ยนตก (Falling tone)	/ ^ /	โท	หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มีสระเสียงสั้น และมีเสียงพยัญชนะสะกดเป็นพยัญชนะกัก
	หน่วยเสียงวรรณยุกต์เปลี่ยนขึ้น (Rising tone)	/ v /	จัตวา	หน่วยเสียงนี้จะไม่ปรากฏในพยางค์ที่มีเสียงพยัญชนะสะกด เป็นเสียงกัก

2.4 ลักษณะ พยางค์ และคำในภาษาไทย

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

พยางค์ที่เปล่งออกมาครั้งหนึ่งๆ อาจมีความหมายหรือไม่ก็ได้ แต่เมื่อใดพยางค์ที่ประกอบขึ้นจากเสียงสระพยัญชนะ และวรรณยุกต์ เป็นอย่างน้อยที่สุด และกลุ่มเสียงเหล่านี้มีความหมายและสามารถปรากฏได้โดยลำพัง พยางค์นั้นๆ ก็กลายเป็นคำในภาษาไทย

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

		วรรณยุกต์	
พยัญชนะต้น	(ควบ)	สระ	(พยัญชนะสะกด)

รูปที่ 2.5 แสดงองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์

บทที่ 3

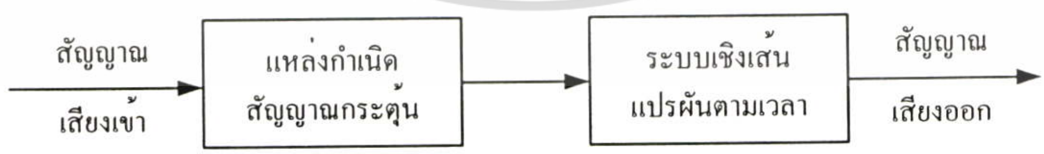
เทคนิคการวิเคราะห์เสียงเชิงเลข

3.1 หลักการประมวลผลในเชิงเวลา

หลักการประมวลผลในเชิงเวลา คือการวิเคราะห์สัญญาณของเสียงพูดโดยตรงจากรูปคลื่นสัญญาณ (waveform) ซึ่งแตกต่างจากการประมวลผลเชิงความถี่ (Frequency – Domain) ตัวอย่างของการประมวลผลของสัญญาณเสียงพูดเชิงเวลา เช่น การหาค่าอัตราการตัดศูนย์(Zero Crossing Rate) การหาค่าพลังงาน (Energy) และค่าออโตคอรีเลชั่น (Autocorrelation) โดยการประมวลผลหาค่าเหล่านี้ เป็นไปโดยง่าย มีขั้นตอนที่ไม่ซับซ้อน แต่ให้องค์ประกอบสำคัญของสัญญาณเสียง

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

ในบทที่ผ่านมาเราได้ทราบการทำงานของระบบเสียงพูดของมนุษย์มาแล้ว ซึ่งเราสามารถแสดงบล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น ดังรูปที่ 3.1 จากรูปมีการแยกภาคแหล่งกำเนิดสัญญาณกระตุ้นออกจากส่วนกำทอนเสียง ซึ่งแทนด้วยระบบเชิงเส้นแปรผันตามเวลา



รูปที่ 3.1 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น

จากรูปที่ 3.1 แหล่งกำเนิดสัญญาณกระตุ้นทำหน้าที่แทน การทำงานของปอดและกล่องเสียง ส่วนนี้จะผลิตขบวนพัลส์ที่มีคาบเวลาพิทช์ ขณะเปล่งเสียงโฆษะ(voice) และให้กำเนิดเสียงซึ่งคล้ายเสียงรบกวนขณะเปล่งเสียงอโฆษะ(unvoice)

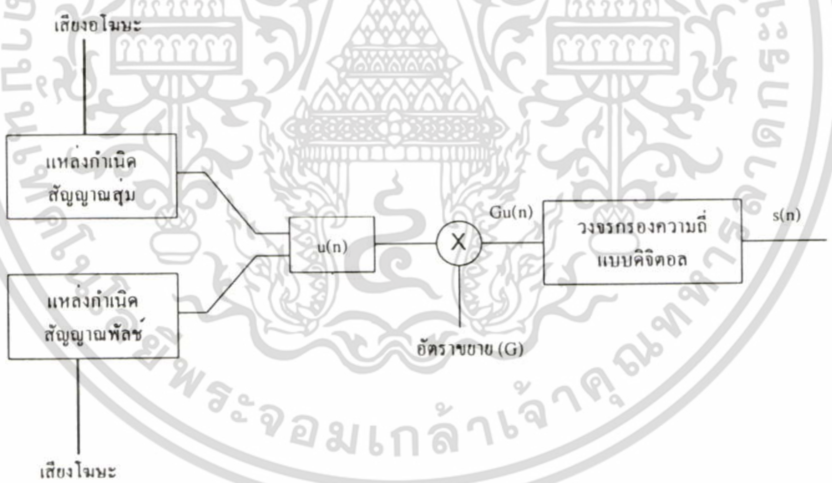
ส่วนที่สองเป็นท่อกำทอนเสียงจะแทนการทำงานของช่องปาก และโพรงจมูก ทำหน้าที่เสมือนตัวกรองสัญญาณ (filter) ที่ยอมให้ความถี่ฟอร์แมนท์ผ่านได้ ซึ่งสามารถแทนด้วยระบบเชิงเส้นแปรผันตามเวลา(time – varying linear system)

3.1.1 แบบจำลองระบบกำเนิดเสียงพูดเชิงเลข

จากแบบจำลองรูปที่ 3.1 ประกอบกับการใช้เทคนิคการประมวลผลสัญญาณเชิงเลข(digital signal processing) สามารถแสดงเป็นแบบจำลองระบบกำเนิดเสียงพูดเชิงเลขดังรูปที่ 3.2

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

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



รูปที่ 3.2 แสดงแบบจำลองของระบบกำเนิดเสียงพูดแบบดิจิตอล

3.1.2 พารามิเตอร์ ของระบบกำเนิดเสียงพูด

การสร้างระบบเสียงพูดจำเป็นต้องวิเคราะห์หาค่าตัวแปรหรือพารามิเตอร์ที่ต้องใช้ตามแบบจำลองและระบบประมวลสัญญาณเสียง ซึ่งประกอบด้วย

3.1.2.1 ดัชนีหรือพารามิเตอร์บอกชนิดเสียงพูดว่าเป็นเสียงโฆษะ(voice)หรืออโฆษะ (unvoice)

3.1.2.2 คาบเวลาพิทซ์ของขบวนพัลส์ในกรณีเสียงโฆษะ(voice)

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

3.1.2.4 พารามิเตอร์ของแบบจำลองท่อกำทอนหรือสัมประสิทธิ์ของวงจรกรอง

3.1.2.5 อัตราการสุ่มตัวอย่างสัญญาณเสียงหรืออัตราเร็วข้อมูล

3.2 การวิเคราะห์สัญญาณเสียงพูดเบื้องต้น

โดยธรรมชาติของสัญญาณเสียงพูดจะไม่เสถียรและเปลี่ยนแปลงตามเวลา(Non-Stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิตอล (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]

3.2.1 กรรมวิธีการปรับบรรทัดฐานแอมปลิจูด (Normalized Amplitude)

กรรมวิธีการปรับบรรทัดฐานแอมปลิจูดของสัญญาณเสียงพูดเป็นการเพิ่มหรือลดขนาดของสัญญาณเสียงพูด เพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน จึงจำเป็นต้องปรับให้มีขนาดของสัญญาณเสียงพูดอยู่ในบรรทัดฐานเดียวกัน เพื่อง่ายต่อการวัดคุณลักษณะและเปรียบเทียบสัญญาณเสียง การปรับบรรทัดฐานแอมปลิจูด (Normalized Amplitude) แสดงดังสมการที่ 3.1

$$\bar{S}[i] = \frac{S[i]}{2^{N-1}} \tag{3.1}$$

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

$S[i]$ คือสัญญาณเสียงพูดดิจิตอล

N คือ จำนวนบิต(Bit) ที่ใช้แทนค่าของสัญญาณเสียงพูด

3.2.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis)

เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ เมื่อเทียบกับแถบความถี่ปฏิบัติงาน (Bandwidth) ไม่เกิน 5 กิโลเฮิรตซ์ ดังนั้น เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal-to-Noise Ratio : SNR) มีค่าค่อนข้างคงที่ตลอดช่วง

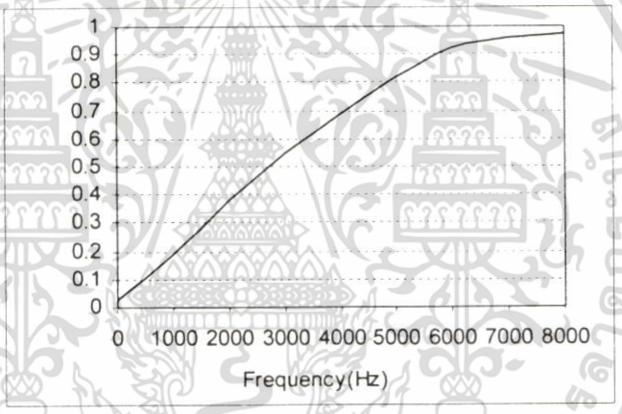
ความถี่ปฏิบัติงานนี้ เราจึงต้องมีการพรีเอมฟาซิส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้นนั่นคือ การพรีเอมฟาซิสก็คือการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (high-pass filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนเป็น

$$H(z) = 1 - a \cdot z^{-1} \qquad \text{โดยที่ } 0.9 < a < 1.0 \tag{3.2}$$

เมื่อเทียบกับรูปที่ 3.3 เราจะได้ว่า

$$s'[i] = s[i] - a \cdot s[i-1] \tag{3.3}$$

ยิ่งค่า a ใกล้ 1 เท่าใด ความถี่สูงก็จะยิ่งถูกขยายมากขึ้นเท่านั้น ค่า a ที่นิยมสำหรับใช้ในการหาพารามิเตอร์ของ LPC คือ 15/16= 0.9375 [3] เมื่อนำฟังก์ชันถ่ายโอนมาพล็อตกราฟของขนาดเทียบกับความถี่จะได้ดังรูปที่ 3.3

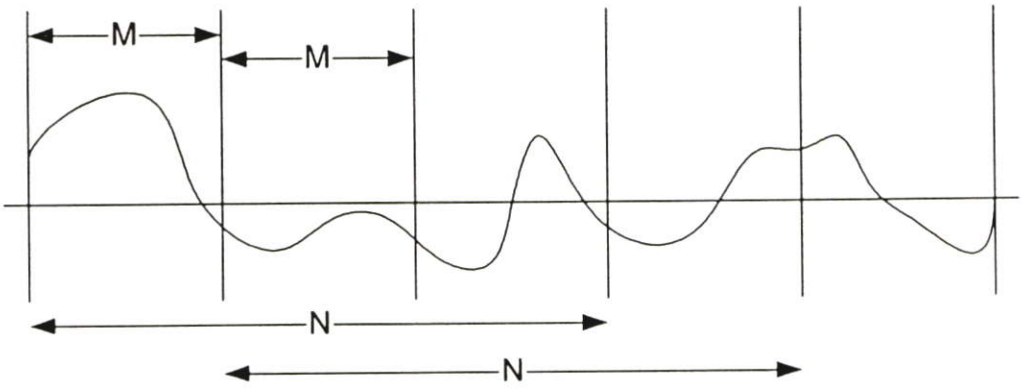


รูปที่ 3.3 แสดงขนาดสเปกตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส

3.2.3 การแบ่งช่วงสัญญาณ (Frame Blocking)

สัญญาณที่ผ่านการพรีเอมฟาซิสแล้ว $S[i]$ จะถูกตัดแบ่งออกเป็นช่วงๆ หรือ เฟรม ช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างสัญญาณ ดังรูปที่ 3.4

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า M โตกว่าค่า N ในการเลื่อนของช่วงในการวิเคราะห์ จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N ก็จะทำให้ตัวอย่างสัญญาณ ทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลง



รูปที่ 3.4 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

วิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N ก็จะทำให้ตัวอย่างสัญญาณ ทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลง

3.2.4 การวินโดว์ (Windowing)

หลักการกำหนดขนาดของวินโดว์

3.2.4.1 วินโดว์จะต้องสั้นพอที่คุณสมบัติของเสียงที่เราสนใจ จะวิเคราะห์นั้นยังไม่เปลี่ยนแปลงในวินโดว์

3.2.4.2 วินโดว์จะต้องยาวพอที่ จำนวนตัวอย่างสัญญาณในวินโดว์ สามารถนำมาคำนวณหาคุณสมบัติที่ต้องการได้

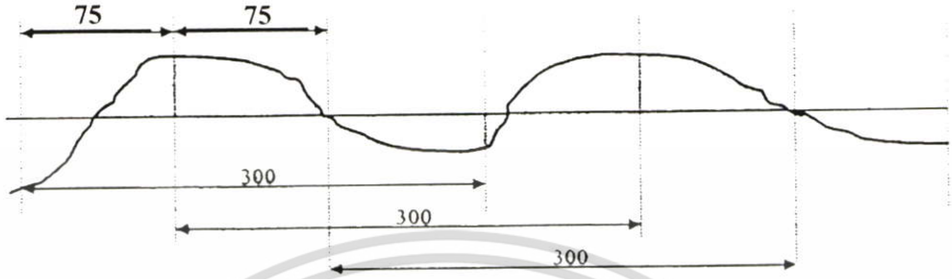
3.2.4.3 วินโดว์ที่ติดกันไม่ควรจะสั้นจนกระโดดข้ามข้อมูลบางส่วนไป แต่ควรเลื่อนวินโดว์ให้น้อยกว่าขนาดของเฟรม

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

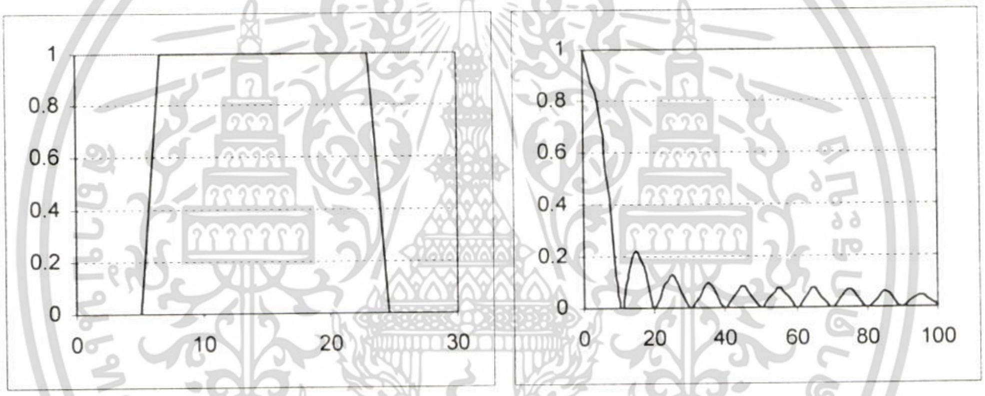
พิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วงใดๆ ที่ตัดมาวิเคราะห์ (รูปที่ 3.5) จะเห็นว่าที่ขอบของเฟรมที่ตัดมานี้มีความไม่ต่อเนื่องของสัญญาณ ถ้ามองในโดเมนความถี่ ก็จะมีความถี่สูงเกิดขึ้นดังนั้นเพื่อที่จะลดองค์ประกอบทางความถี่ที่สูงเหล่านี้ จะคูณด้วยฟังก์ชันวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ในที่นี้จะให้ฟังก์ชันวินโดว์แฮมมิ่ง(Hamming window function) ซึ่งนิยามโดยสมการดังนี้คือ

$$w(n) = 0.54 - 0.46(2\pi n/(N-1)) \text{ เมื่อ } n = 1,2,\ldots,N-1 \qquad (3.4)$$

รูปที่ 3.6 และ รูปที่ 3.7 แสดงองค์ประกอบทางเวลาและทางความถี่ของฟังก์ชันวินโดว์แบบสี่เหลี่ยม และแฮมมิ่ง ตามลำดับ จะเห็นว่าสเปกตรัมของวินโดว์แฮมมิ่งมีริพเพิล (ripple) น้อยกว่าของวินโดว์สี่เหลี่ยม

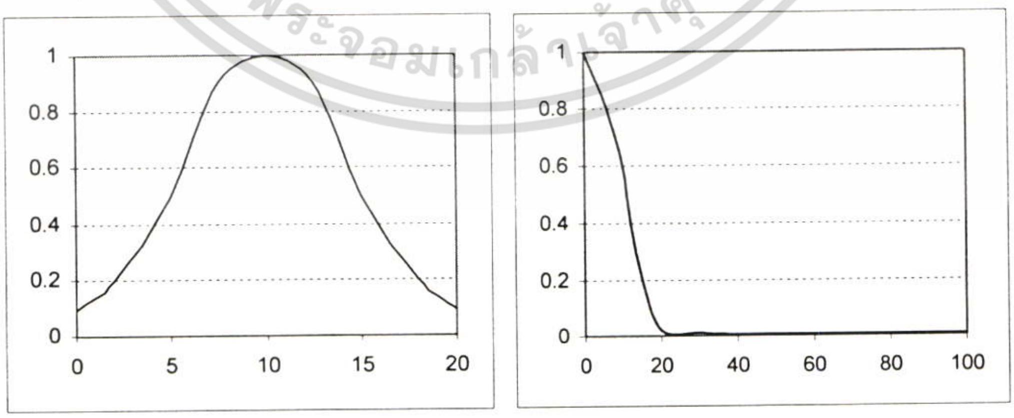


รูปที่ 3.5 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์



(a) (b)

รูปที่ 3.6 แสดงวินโดว์แบบสี่เหลี่ยม(ก) ในโดเมนเวลา (ข) ในโดเมนความถี่



(a) (b)

รูปที่ 3.7 แสดงวินโดว์แบบแฮมมิ่ง (a) ในโดเมนเวลา (b) ในโดเมนความถี่

ดังนั้นเราจะได้ว่า

$$x^{'}(n) = x(n) \cdot w(n) \tag{3.5}$$

3.3 พลังงานของสัญญาณเสียงพูด

พลังงานของสัญญาณเสียงพูด ถูกนำมาใช้ในการวิเคราะห์เสียงพูด และเป็นคุณลักษณะที่นิยมนำมาใช้กันอย่างแพร่หลาย เนื่องจากเป็นวิธีที่คำนวณง่ายและรวดเร็ว พลังงานของสัญญาณเสียงพูดเป็นคุณสมบัติที่แสดงให้เห็นว่ามีสัญญาณเสียง (รวมทั้งสัญญาณรบกวน) เกิดขึ้น ณ เวลานั้นหรือไม่ การคำนวณหาค่าพลังงาน จะทำทีละกรอบเสียงพูด โดย E_n คือ สัญญาณเสียงพูดที่ i ในกรอบเสียงพูด n และในแต่ละกรอบเสียงพูดจะมีสัญญาณเสียงพูดจำนวน K ซึ่งวิธีการคำนวณหาค่าพลังงานของสัญญาณเสียงพูดในวิทยานิพนธ์นี้ เลือกใช้วิธีพลังงานกำลังสอง (Square Energy) ซึ่งเป็นการวัดค่าพลังงานจากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ การหาค่าพลังงานกำลังสองแสดงดังสมการที่ 3.6

$$En = \left[\sum_{i=1}^{w} S_n^2[i] \right] \tag{3.6}$$

3.4 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง (Normalized Energy function)

เพื่อให้ฟังก์ชั่นพลังงานเสียงที่จะนำไปทำการทดสอบบรรทัดฐานค่าพลังงานเดียวกัน ทำได้โดยการกำหนดให้พื้นที่ใต้ฟังก์ชั่นพลังงานเสียงมีค่าเท่ากับหนึ่ง ทุกเสียงทดสอบ ซึ่งแสดงได้ดังสมการต่อไปนี้

กำหนดให้พื้นที่ใต้ฟังก์ชันพลังงานเสียง $E(n)$ มีค่าเท่ากับ 1

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

โดยที่ A คือ normalized factor

N คือ ค่าพลังงานเสียงเริ่มต้น , M ค่าพลังงานเสียงสุดท้าย

สามารถหา normalized factor ได้คือ

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

ดังนั้นจะได้ค่าพลังงานเสียงที่ผ่านการปรับบรรทัดฐานพลังงานเสียง ดังนี้

$$Qn = AEn \tag{3.7}$$

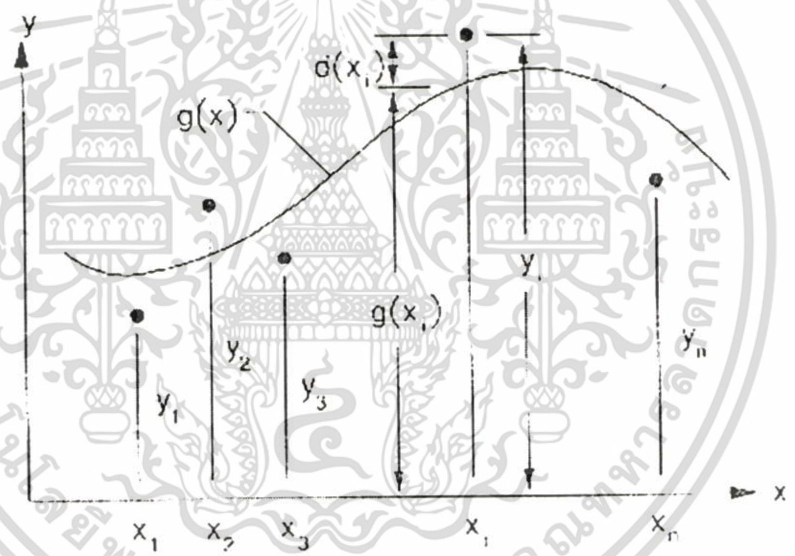
เมื่อ Qn คือ ค่าพลังงานเสียงที่ผ่านการปรับบรรทัดฐานพลังงาน

3.5 การถดถอยแบบพหุนาม

การถดถอยแบบพหุนาม(polynomial regression) เป็นระเบียบวิธี ที่ใช้ประดิษฐ์ฟังก์ชันพหุนามสำหรับข้อมูลที่มีการกระจายโดยทั่วไปที่ไม่อยู่ในรูปแบบของเชิงเส้นหรือสมการกำลัง รูปที่ 3.8 แสดงการใช้ฟังก์ชันพหุนามในรูปแบบทั่วไปกับชุดของข้อมูลที่ไม่ได้อยู่ในรูปแบบเชิงเส้นชุดหนึ่งที่กำหนดมาให้

ชุดของข้อมูลในรูปที่ 3.8 นี้ ประกอบด้วย $x_i, y_i, i = 1,2,\ldots,n$ กล่าวคือมีจำนวนข้อมูลทั้งสิ้น n ข้อมูล ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ m สำหรับข้อมูลชุดนี้

$$g(x) = a_0 + a_1x + a_2x^2 + \ldots + a_mx^m \tag{3.8}$$



รูปที่ 3.8 การถดถอยแบบพหุนามโดยการประดิษฐ์ฟังก์ชันพหุนามจากชุดของข้อมูล

โดย $a_0, a_1, a_2, \ldots, a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่จะประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุดจากข้อมูลทั้งหมดที่กำหนดมาให้ ขั้นตอนในการประดิษฐ์สมการพหุนามนี้ เริ่มจากการหาค่าความผิดพลาด E ทั้งหมดที่เกิดขึ้นจาก n ข้อมูลในรูปแบบดังนี้

$$E = \sum_{i=1}^{n} \left[d(x_i)\right]^2 \tag{3.9}$$

ซึ่งสามารถเขียนให้ประกอบด้วยฟังก์ชันพหุนามดังแสดงในสมการ 3.8 ได้คือ

$$E = \sum_{i=1}^{n} \left[y_i - g(x_i) \right]^2 \tag{3.10}$$

$$E = \sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m) \right]^2 \tag{3.11}$$

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

$$\left.\begin{aligned} \frac{\partial E}{\partial a_0} &= 0 \\ \frac{\partial E}{\partial a_1} &= 0 \\ \frac{\partial E}{\partial a_2} &= 0 \\ \vdots &= \vdots \\ \frac{\partial E}{\partial a_m} &= 0 \end{aligned}\right\} m+1 \text{ สมการ} \tag{3.12}$$

ดังตัวอย่าง เช่น สมการแรกในระบบสมการนี้สามารถประดิษฐ์ขึ้นได้ดังนี้

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

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

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

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

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^m \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \cdots & \sum_{i=1}^{n} x_i^{m+1} \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \cdots & \sum_{i=1}^{n} x_i^{m+2} \\ \vdots & \vdots & \vdots & & \\ \vdots & \vdots & \vdots & & \\ \sum_{i=1}^{n} x_i^m & \sum_{i=1}^{n} x_i^{m+1} & \sum_{i=1}^{n} x_i^{m+2} & \cdots & \sum_{i=1}^{n} x_i^{2m} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \vdots \\ \vdots \\ \sum_{i=1}^{n} x_i^m y_i \end{Bmatrix} \tag{3.13}$$

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

3.6 การคำนวณสเปกตรัมแอลพีซี

เนื่องจากสัญญาณเสียง มีพารามิเตอร์ที่สำคัญต่างๆดังที่กล่าวไว้ข้างต้น การวิเคราะห์การเข้ารหัสโดยการประมาณเชิงเส้น(Linear Predictive Coding : LPC) จึงถูกใช้ในการหาค่าพารามิเตอร์ พื้นฐานต่างๆ เพื่อใช้ในการเข้ารหัสสำหรับส่งสัญญาณเสียงด้วยความถี่บิทต่ำ การย่อข้อมูลและการจดจำเสียงด้วยเหตุที่ว่าวิธีนี้เป็นการวิเคราะห์พารามิเตอร์ต่างๆ ที่ให้ความแม่นยำสูงและสามารถย่อข้อมูลได้อย่างมีประสิทธิภาพ

หลักการพื้นฐานของการประมาณเชิงเส้นคือ การประมาณค่าสัญญาณจากผลรวมเชิงเส้นของสัญญาณก่อนหน้านี้ สมมติว่าสัญญาณเดิมเป็น $s(n)$ การประมาณค่าสัญญาณเป็น $s'(n)$ ดังนั้นเราสามารถอธิบายการ ประมาณเชิงเส้นได้ด้วยสมการต่อไป

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{3.14}$$

เมื่อ α_k เป็นค่าคงที่เรียกวิธีการนี้ว่าการประเมินเชิงเส้นอันดับ p โดยมีเงื่อนไขว่า ค่า α_k ที่ใช้ในการประมาณจะต้องทำให้ผลรวมของกำลังสองของความคลาดเคลื่อน $\{s(n)-s'(n)\}^2$ มีค่าน้อยที่สุด นั่นคือ $\sum e^2(n) = \sum \{s(n)-s'(n)\}^2$ มีค่าต่ำที่สุด ซึ่งการประมาณมีอยู่หลายวิธี ได้แก่

- วิธีโควาเรียนซ์ (Covariance Method)
- วิธีออโตคอร์รีเลชัน (Autocorrelation Method)
- วิธีแลตทิส (Lattice Method)
- วิธีวงจรกรองผกผัน (Inverse Filter Method)
- วิธีการประมาณสเปคตรัม (Spectrum Estimation Method)

และอื่นๆ อีกหลายวิธี แต่ในที่นี้จะเลือกใช้วิธีออโตคอร์รีเลชัน หรือ วิธีอัตตสัมพันธ์ จากหลักการพื้นฐานของการประมาณเชิงเส้น และรูปที่ 3.2 สามารถเขียนสมการได้เป็น

$$s(n) = Gu(n) + \sum_{k=1}^{p} a_k s(n-k) \tag{3.15}$$

โดยที่ a_k เป็นค่าสัมประสิทธิ์ของวงจรกรองเชิงเลข

การประมาณเชิงเส้นโดยใช้สัมประสิทธิ์ α_k คือ

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{3.16}$$

ดังนั้นความคลาดเคลื่อนคือ

$$e(n) = s(n) - s'(n)$$

$$e(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{3.17}$$

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

$$A(z) = \frac{E(z)}{S(z)}$$

$$= 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{3.18}$$

จากสมการ(3.15)-(3.17) จะเห็นได้ว่า ถ้า $\alpha_k = a_k$ แล้ว

$$s(n) = G \cdot u(n) \tag{3.19}$$

ในการประมวลสัญญาณจะมีการแบ่งสัญญาณออกเป็นช่วงสั้นๆ (windowing) ช่วงละประมาณ 20 – 40 มิลลิวินาที ซึ่งจะได้มีการอธิบายต่อไป ดังนั้น ค่าผลรวมของกำลังสองของความคลาดเคลื่อน

$$\begin{aligned} E_n &= \sum_m e_n^2(m) \\ &= \sum_m \left[s_n(m) - s'_n(m)\right]^2 \\ &= \sum_m \left[s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right]^2 \end{aligned} \tag{3.20}$$

โดยที่ n คือ ช่วงที่ n ของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า E_n ต่ำที่สุด จะต้องมีเงื่อนไขว่า

$$\frac{\partial E_n}{\partial \alpha_i} = 0 \quad \text{เมื่อ } i = 1,2,3,\ldots,p$$

จากสมการ(3.20)

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

$\partial E_n / \partial \alpha_i = 0$ ก็ต่อเมื่อ

$$\sum_{k=1}^{p} \alpha_k \sum_m s_n(m-k)s_n(m-i) = \sum_m s_n(m)s_n(m-i) \quad \text{เมื่อ } i = 1,2,3,\ldots,p \tag{3.21}$$

ถ้าเรากำหนดให้ $\varphi_n(i,j) = \sum s_n(m\text{-}k)\, s_n(m\text{-}i)$ เพราะฉะนั้น

$$\sum_{k\text{-}1}^{p} \alpha_k \varphi_n(i,j) = \varphi_n(i,0) \tag{3.22}$$

โดยสมการ(3.20) – (3.21) จะได้ว่า

$$E_n = \sum_{m} s^2_{n}(m) - \sum_{k=1}^{p} \alpha_{k} \sum_{m} s_n(m) s_n(m\text{-}i)$$

และจาก $\varphi_n(i,j) = \sum s_n(m\text{-}k)\, s_n(m\text{-}i)$

$$E_n = \varphi_n(0,0) - \sum_{k=1}^{p} \alpha_n\, \varphi_n(0,k) \qquad (3.23)$$

สมมติว่าใน1เฟรมของสัญญาณที่ตัดมาคำนวณ มี N ตัวอย่าง คือ $s_n(0), s_n(1), s_n(2), \ldots, s_n(n\text{-}1)$ ในที่นี้เราให้ $s_n(m) = 0$ เมื่อ $m < 0$ หรือ $m > N-1$ เพราะฉะนั้น

$$\varphi_n(i,j) = \sum s_n(m\text{-}k)\, s_n(m\text{-}i)$$

$$= \sum_{m=0}^{N\text{-}1\text{-}(i\text{-}k)} s_n(m) s_n(m + i\text{-}k) \qquad 0 \le \mathrm{k} \le p, k \le i \le p$$

ให้

$$R_n(k) = \sum_{m=0}^{N\text{-}1\text{-}k} s_n(m) s_n(m+k) \qquad (3.24)$$

$$\therefore \varphi_n(i,j) = R_n(|i\text{-}k|) \qquad \text{เมื่อ } i = 1,2,3,\ldots,p\,;\ k = 0,1,2,\ldots,p$$

จากสมการ(3.22) จะได้ว่า

$$\sum_{k=1}^{p} \alpha_k R_n(|i\text{-}k|) = R_n(i) \qquad \text{เมื่อ } i = 1,2,3,\ldots,p \qquad (3.25)$$

และจากสมการ(3.23)

$$E_n = R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \qquad (3.26)$$

จากสมาการ(3.25) เขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

$$\begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p\text{-}1) \\ R_n(1) & R_n(0) & \cdots & R_n(p\text{-}2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p\text{-}1) & R_n(p\text{-}2) & \cdots & R_n(0) \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix} \qquad (3.27)$$

หรือ

$$\mathrm{R_n} \cdot \alpha = \mathrm{r}_n \qquad (3.28)$$

$$\mathrm{R}_n = \begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p\text{-}1) \\ R_n(1) & R_n(0) & \cdots & R_n(p\text{-}2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p\text{-}1) & R_n(p\text{-}2) & \cdots & R_n(0) \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \text{และ } \mathrm{r}_n = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix}$$

จากสมการ (3.15) และ (3.17)

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

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

เมื่อ $a_k = \alpha_k$ จะได้ว่า

$$e(n) = Gu(n)$$

$$\therefore E_n = \sum_{m=0}^{n\text{-}1} e^2 m = G \sum_{m=0}^{n\text{-}1} u^2(m) \tag{3.29}$$

$$= \mathrm{R}_n(0) - \sum_{k=1}^{p} \alpha_k \mathrm{R}_n(k)$$

จากสมการ (3.18) เราสามารถหา α โดยตรงดังนี้

$$\alpha = \mathrm{R}_n^{-1} \mathrm{r}_n \tag{3.30}$$

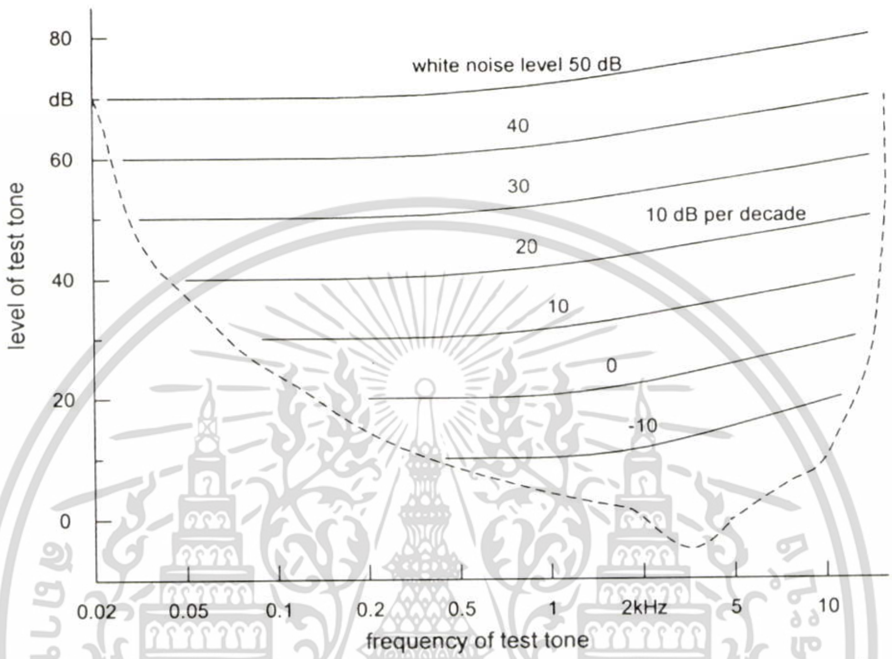
และจากสมการ (3.29) เราสามารถหาค่า G โดยตรงจาก

$$G = \frac{\mathrm{R}_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)}{\sum_{m=0}^{n\text{-}1} u^2(m)} \tag{3.31}$$

3.7 แถบความถี่วิฤต และความเข้มสัญญาณในแถบความถี่วิกฤต

คำจำกัดความของแถบความถี่วิกฤตนั้นได้ถูกกล่าวขึ้นโดย [Fletcher] โดยที่เขาได้ตั้งสมมติฐานไว้ว่าในกรณีที่มีสัญญาณรบกวน(Noise) ทับซ้อนจนสามารถบดบัง(Mask) สัญญาณโทนทดสอบ(Test Tone) ได้นั้น ส่วนของสัญญาณรบกวนที่มีผลในการบดบังสัญญาณโทนทดสอบจริงๆจะเป็นเพียงส่วนที่มีความถี่ใกล้เคียงกับสัญญาณโทนทดสอบเท่านั้น ในการที่จะทำการหาค่า

สัมประสิทธิ์ต่างๆของแถบความถี่วิกฤตได้นั้นยังต้องอาศัยสมมติฐานที่ว่า การบดบังสัญญาณจะเป็นผลก็ต่อเมื่อกำลังงานของสัญญาณรบกวนในแถบสัญญาณเดียวกับสัญญาณโทนทดสอบนั้นเท่ากับกำลังงานของสัญญาณโทนทดสอบ และถือว่ากำลังงานของสัญญาณรบกวนที่อยู่นอกเหนือจากแถบสัญญาณที่ใกล้กับสัญญาณโทนทดสอบไม่มีผลต่อการบดบังสัญญาณ



รูปที่ 3.9 ระดับของสัญญาณโทนทดสอบที่ถูกบดบังโดยสัญญาณ White Noise ที่ระดับสัญญาณต่างๆ และเส้นประเป็นระดับ Threshold in quiet

รูปที่ 3.9 แสดงการบดบังสัญญาณโทนทดสอบโดยใช้สัญญาณรบกวนแถบความถี่กว้าง (Broadband Noise) ในที่นี้ใช้ White Noise ที่ระดับความเข้มสัญญาณต่างๆ โดยเส้นประเป็นระดับ Threshold in Quiet หมายถึงระดับของสัญญาณโทนที่หูของมนุษย์เริ่มที่จะได้ยินเมื่อไม่มีสัญญาณรบกวนใดๆ เช่นหากมีสัญญาณโทนที่ความถี่ 2 kHz โดยไม่มีสัญญาณรบกวนใดๆ หูเราสามารถได้ยินสัญญาณโทนนั้นได้ก็ต่อเมื่อ สัญญาณมีระดับความเข้มประมาณ 0 dB($10^{-12}\,w/m^2$) ขึ้นไป ส่วนเส้นทึบเป็นเส้นที่แสดงระดับ Masking Threshold สำหรับความเข้มสัญญาณ White Noise ที่ระดับต่างๆซึ่ง Threshold นี้บ่งบอกว่าหากสัญญาณโทนมีระดับความเข้มน้อยกว่า Threshold นี้ เราจะไม่สามารถได้ยินสัญญาณโทนนี้ได้ เช่น สัญญาณโทนที่ 2 kHz ขณะที่มีสัญญาณ White Noise ความเข้มสัญญาณ 20 dB เราจะสามารถได้ยินโทนนี้ก็ต่อเมื่อโทนนี้มีระดับความเข้มสัญญาณมากกว่าประมาณ 40 dB ซึ่งจะเห็นได้ว่าในการบดบังสัญญาณโดย White Noise นั้นจะมีระดับ Threshold ที่ไม่เป็นอิสระกับความถี่ทั้งๆที่ตัว White Noise เองนั้นความเข้มสัญญาณของมันไม่ขึ้นกับความถี่ แต่จะเห็นได้ว่าระดับการบดบังสัญญาณจะมีลักษณะที่ไม่ขึ้นกับความถี่ในช่วงที่

สัญญาณมีความถี่ต่ำกว่า 500 Hz ส่วนที่ความถี่ประมาณ 1 kHz ขึ้นไปนั้นจะมีอัตราการเปลี่ยนแปลงอยู่ที่ประมาณ 10 dB/decade

ระบบการรับฟังเสียงของคนเรามีระดับของการถูกบดบังสัญญาณที่ไม่ขึ้นกับความถี่ในช่วงที่ต่ำกว่า 500 Hz ในช่วงความถี่นี้แถบความถี่วิกฤตจะมีความกว้างเท่าๆกันด้วย สำหรับความถี่ที่สูงกว่านี้นั้นจะมีระดับการบดบังสัญญาณเพิ่มขึ้นประมาณ 10 dB/decade นั่นก็หมายความว่าระดับความเข้มของสัญญาณนั้นเพิ่มขึ้นตามความถี่ ดังนั้นความกว้างแถบความถี่ของช่วงสัญญาณนี้เพิ่มขึ้น 10 dB/decade ด้วยเช่นกัน

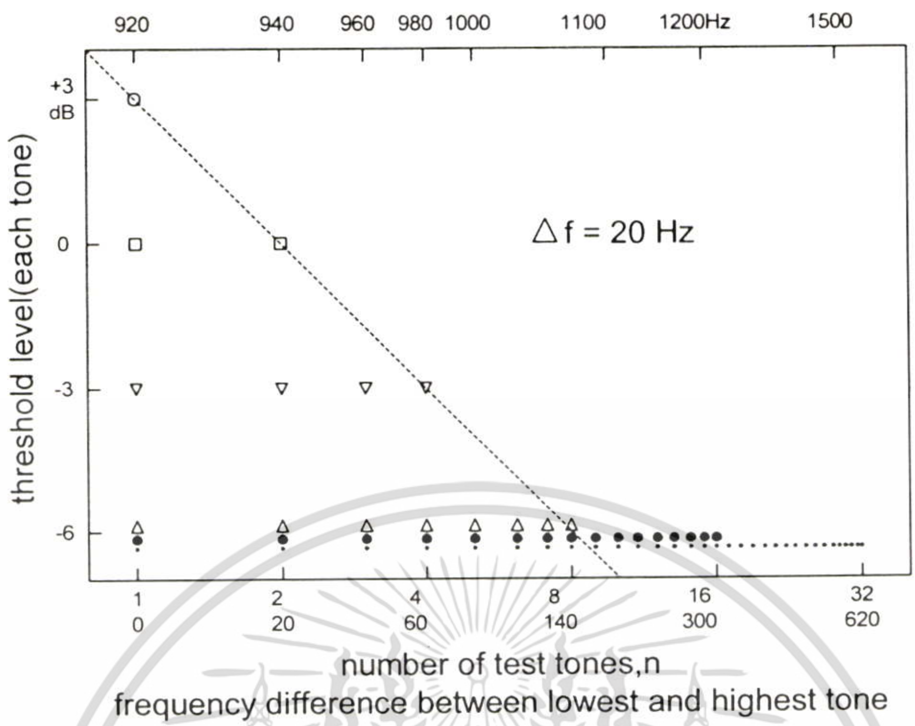
หากเราเชื่อว่าสมมติฐานของ [Fletcher] นั้นถูกต้องแสดงว่าเราต้องสามารถได้ยินเสียงสัญญาณโทนในขณะที่สัญญาณรบกวนที่มีแถบสัญญาณใกล้เคียงกับแถบความถี่วิกฤตที่มีความถี่ของสัญญาณโทนเป็นศูนย์กลางนั้นมีกำลังเท่ากับสัญญาณโทน ซึ่งจะทำให้เราสามารถที่จะคาดคะเนความกว้างของแถบความถี่วิกฤตได้โดย ที่ความถี่ต่ำกว่า 500 Hz นั้นระดับความเข้มของโทนที่ถูกบดบังสัญญาณจะสูงกว่าระดับความหนาแน่นของ White Noise ที่บดบังสัญญาณโทนนั้นอยู่ประมาณ 17 dB ซึ่งจากสมมติฐานนี้ เราสามารถหาความกว้างแถบความถี่วิกฤตได้เป็น $10^{17/10}$ หรือประมาณ 50 Hz

แต่ในความเป็นจริงแล้ว เราสามารถได้ยินเสียงโทนในขณะที่กำลังของเสียงโทนเป็นเพียงครึ่งหนึ่งหรือเพียงหนึ่งในสี่ของกำลังของสัญญาณรบกวนในแถบความถี่ที่ใกล้กับความถี่ของเสียงโทนนั้นเท่านั้น[16]-[17] และการศึกษาโดยใช้ข้อมูลอื่นๆเพิ่มเติม เราสามารถคาดคะเนความกว้างของแถบความถี่วิกฤตได้ใกล้เคียงยิ่งขึ้น ซึ่งในย่านความถี่ต่ำกว่า 500 Hz นั้นความกว้างของแถบความถี่วิกฤตจะมีค่าประมาณ 100 Hz ส่วนที่ความถี่สูงกว่า 500 Hz นั้นจะมีความกว้างของแถบความถี่วิกฤตประมาณ 20 เปอร์เซ็นต์ ของความถี่ศูนย์กลางนั้นๆ

3.7.1 วิธีการในการหาความกว้างแถบความถี่วิกฤต

วิธีการแรกที่ใช้ในการหาความกว้างของแถบความถี่วิกฤตคือ การวัดจากระดับ Threshold โดยตรงซึ่งมีลักษณะเช่นเดียวกับวิธีการอื่นๆที่ใช้การวัดระดับสัญญาณโดยตรงคือ ความกว้างแถบความถี่ หรือระดับสัญญาณที่มีความเกี่ยวข้องกับแถบกว้างความถี่ จะต้องเป็นตัวแปรซึ่งในกรณีนี้ Threshold in Quiet ของสัญญาณโทนความถี่ต่างๆซึ่งมี แอมปลิจูดเท่ากันถูกใช้ในการประมาณความกว้างแถบความถี่วิกฤตที่อยู่ในช่วงประมาณ 1KHz

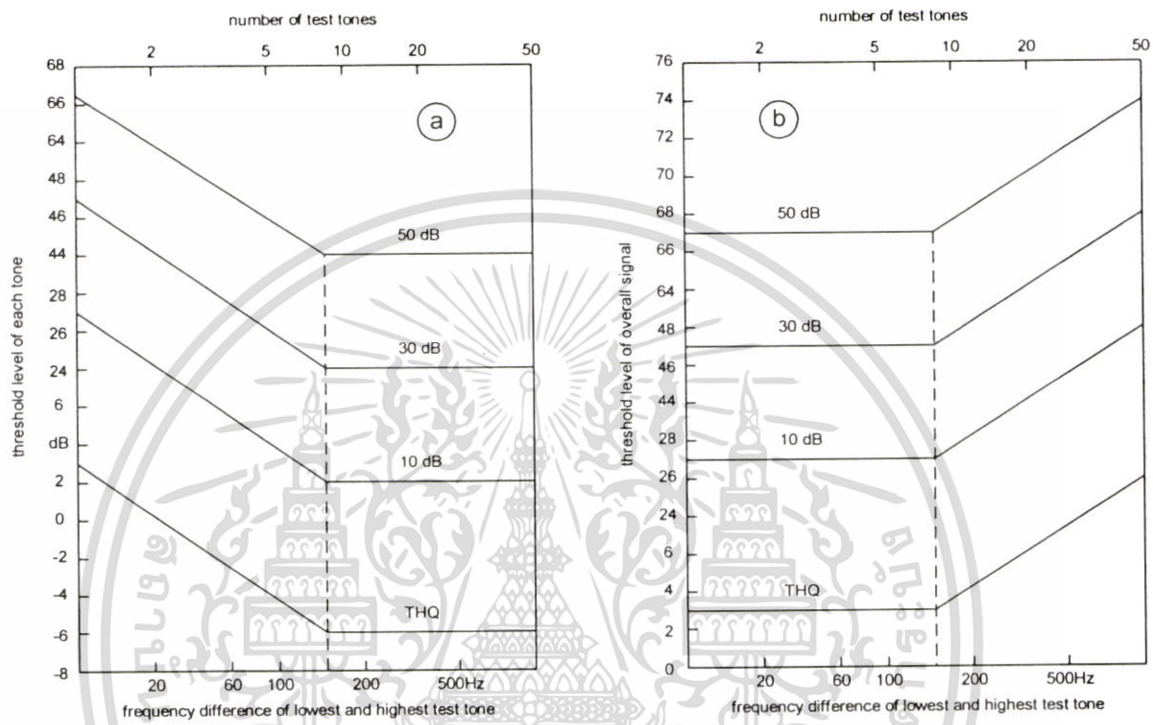
รูปที่ 3.10 เป็นระดับ Threshold in Quiet ของแต่ละโทนเมื่อเปลี่ยนจำนวนของสัญญาณโทนความถี่ต่างๆและความแตกต่างระหว่างความถี่สูงสุดและความถี่ต่ำสุดของโทน จากรูปเป็นการวัดระดับของ Threshold in Quiet โดยอาศัยวิธีการ Tracking สัญญาณโดยเริ่มต้นที่สัญญาณโทนที่ความถี่ 920 Hz เพียงสัญญาณเดียวจะได้ Threshold in Quiet ที่ +3dB เมื่อเพิ่มสัญญาณโทนอีกค่าก็คือ 940 แล้วทำการ Tracking เพื่อวัดระดับ Threshold in Quiet ของสัญญาณที่มีโทนทั้งสองรวมอยู่



รูปที่ 3.10 ความสัมพันธ์ระหว่างระดับ Threshold in Quiet และจำนวนสัญญาณโทนทดสอบ โดยจุดที่มีลูกศรอยู่เป็นจุดที่ใช้ในการคาดคะเนความกว้างแถบความถี่วิกฤต

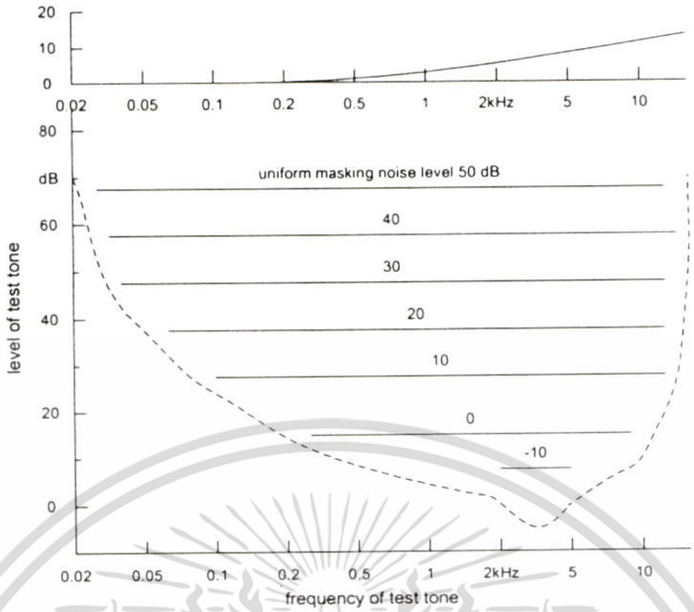
จะได้ระดับ Threshold in Quiet ของแต่ละสัญญาณที่ 0 dB และเมื่อทำตามลักษณะนี้ไปเรื่อยๆ โดยการเพิ่มจำนวนสัญญาณโทนที่ละสองเท่าแล้วทำการ Tracking เพื่อวัดระดับ Threshold in Quiet จะได้ –3dB และ –6dB สำหรับจำนวนโทน 4 และ 8 โทนตามลำดับซึ่งแสดงให้เห็นว่าระดับ Threshold in Quiet ของแต่ละโทนจะลดลงเมื่อมีการรวมกันของสัญญาณโทนมากขึ้น แต่เมื่อเพิ่มจำนวนโทนจนไปถึงจุดหนึ่ง ระดับ Threshold in Quiet ของแต่ละโทนก็จะไม่ลดลงอีกดังที่สังเกตได้ชัดในรูปที่ 3.10 ที่จำนวนโทน 16 และ 32 โทนก็จะไม่มีความแตกต่างกันของระดับ Threshold in Quiet ของแต่ละโทน กับที่ระดับ Threshold in Quiet ของการใช้จำนวนโทนเท่ากับ 8 มากนัก ซึ่งค่านี้เองเป็นค่าที่เราใช้ในการวัดความกว้างแถบความถี่วิกฤต เป็นที่น่าสังเกตว่าช่วงของจำนวนโทนตั้งแต่ 1 ถึง 8 นั้นจะมีระดับ Threshold in Quiet ลดลง 3 dB ทุกๆการเพิ่มขึ้นสองเท่าของจำนวนโทน หมายความว่าที่ระดับ Threshold in Quiet ค่าใดค่าหนึ่งนั้น ระดับแรงดันเสียงรวมทุกโทนค่อนข้างคงที่ไม่ขึ้นกับจำนวนโทน แต่กฎนี้ใช้ได้กับเฉพาะจำนวนโทนที่ไม่เกิน 8 เท่านั้นถ้ามากกว่านี้ระดับ Threshold in Quiet ของแต่ละโทนไม่ได้ลดลงตามการเพิ่มขึ้นของจำนวนโทนจึงทำให้ ระดับแรงดันเสียงโดยรวมเพิ่มขึ้น ซึ่งหมายความว่าระดับ Threshold In Quiet ของระดับการได้ยินเสียงของเราสามารถประมาณได้โดยความเข้มของเสียงโทนทั้งหมด ตราบใดที่โทนต่างๆอยู่ในแถบความถี่วิกฤต ส่วนที่อยู่นอกเหนือแถบกว้างความถี่วิกฤตนี้จะไม่นำมามีผลกับค่า Threshold In Quiet หรืออีกนัยหนึ่งคือ สัญญาณโทนความถี่ 920 Hz ได้ก็ต่อเมื่อมันมีความเข้ม +3 dB ขึ้นไป

หรือสัญญาณ ใดๆที่มีความถี่อยู่ในแถบความถี่วิกฤตที่มีความถี่ศูนย์กลางที่ 920 Hz แล้วมีความเข้มสัญญาณรวมกัน 3 dB ขึ้นไป การคำนวณความกว้างแถบความถี่วิกฤตโดยวิธีนี้สามารถคำนวณได้จากจำนวนของโทน และระยะห่างระหว่างโทนซึ่งอยู่ในรูปจะได้ค่า 140 Hz ซึ่งมาจาก (8-1)*20=140 Hz เป็นความกว้างแถบความถี่วิกฤตสำหรับความถี่ศูนย์กลางที่ 920 Hz



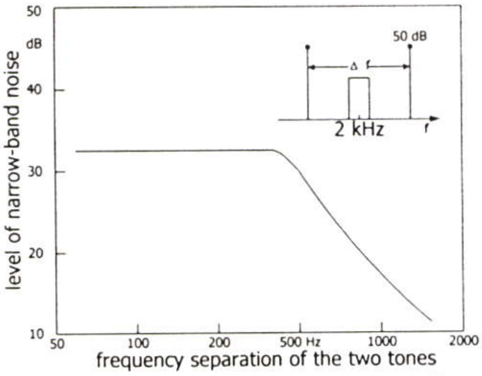
รูปที่ 3.11 ทั้งสองรูปเป็นผลต่อเนื่องจากรูปที่ 3.10 โดยการเพิ่มผลที่ได้จากการใช้ Uniform Masking Noise ด้านซ้ายเป็นระดับ Threshold ของแต่ละโทนส่วนด้านขวาเป็นของสัญญาณรวมทุกโทน

การหาค่าความกว้างแถบความถี่วิฤตโดยใช้ Threshold In Quiet นี้จะให้คำตอบที่ถูกต้องก็ต่อเมื่ออยู่ในย่านความถี่ซึ่งระดับ Threshold In Quiet ไม่ขึ้นกับความถี่ ซึ่งช่วงความถี่นั้นอยู่ในช่วงที่ไม่เกิน 500 Hz ถึง 2KHz เท่านั้นอย่างไรก็ตามหากดูในรูปที่ 3.10 จะเห็นได้ว่า Uniform Masking Noise จะสามารถบดบังสัญญาณซึ่งจะไม่ขึ้นกับความถี่ของโทน ถ้าหากว่าปรากฏการณ์ที่เกิดขึ้นซึ่งอธิบายในรูปที่ 3.10 นั้นเกิดขึ้นไม่ใช่เพียงแค่ที่ Threshold In Quiet เท่านั้น แต่ยังเกิดขึ้นกับระดับของการบดบังสัญญาณโดย Uniform Masking Noise แล้วหมายความว่ามันเป็นไปได้ที่จะเกิดปรากฏการณ์นี้ตลอดช่วงความถี่ของการได้ยินของมนุษย์ได้ด้วย ผลที่ได้จากการวัดจะใกล้เคียงกับรูปที่ 3.10 สำหรับความถี่ของสัญญาณโทนที่ใกล้เคียงกับ 920 Hz ส่วนการใช้ Uniform Masking Noise ที่ระดับต่างๆแสดงในรูปที่ 3.11 จำนวนสัญญาณโทนซึ่งมีระดับสัญญาณเท่ากันจะเป็นค่าในแกนด้านบน

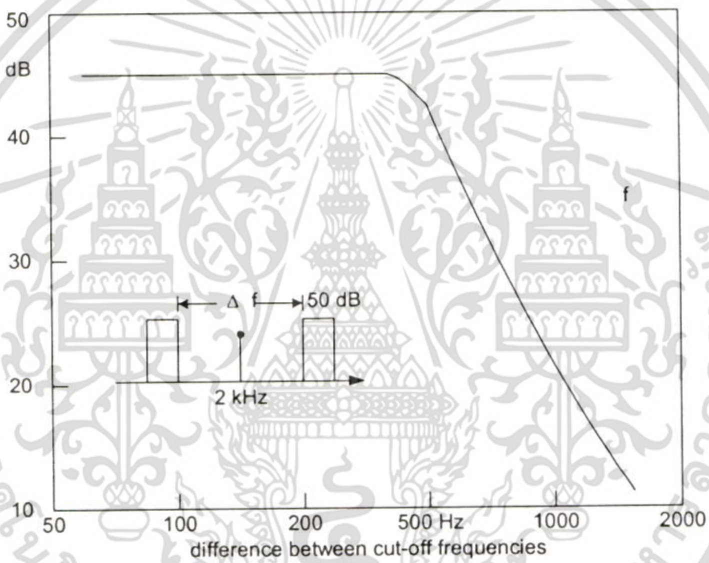


รูปที่ 3.12 ระดับของสัญญาณโทนทดสอบที่ถูกบดบังโดย Uniform Masking Noise ที่ระดับสัญญาณต่างๆ โดยกราฟด้านบนเป็นค่าการลดทอนที่ให้กับ White Noise เพื่อสร้าง Uniform Masking Noise

ส่วนความกว้างแถบสัญญาณที่ได้จากการเปลี่ยนแปลงจำนวนโทนแสดงอยู่ในแกนด้านล่าง ผลแสดงให้เห็นชัดว่าวิธีที่กล่าวผ่านมาทั้งสองวิธีนั้น ระดับ Threshold ของแต่ละโทนจะลดลงในช่วงที่ความแตกต่างของความถี่มีค่าน้อยกว่าค่าๆหนึ่ง ส่วนที่ความแตกต่างของความถี่มีค่าสูงกว่าค่านั้นระดับสัญญาณของแต่ละโทนจะคงที่ ในรูปที่ 3.11a นั้นจะแสดงระดับ Threshold ของแต่ละโทนในขณะที่รูปที่ 3.11b จะแสดงเปรียบเทียบกับระดับแรงดันเสียงรวมทุกโทนแทนที่จะแยกออกสำหรับแต่ละโทน และจะเห็นได้ว่าระดับแรงดันเสียงรวมนั้นจะคงที่ที่ระดับ Threshold in Quiet หรือ Masking Threshold ในบริเวณที่ความแตกต่างของความถี่น้อยกว่าความกว้างแถบความถี่วิฤต ระดับแรงดันของเสียงรวมจะเพิ่มขึ้นเมื่อความแตกต่างของความถี่มากกว่าความกว้างแถบความถี่วิกฤต ซึ่งแสดงให้เห็นว่าองค์ประกอบของเสียงส่วนที่เกินจากแถบความถี่วิกฤตนั้นจะไม่ถูกนำมามีผลในการวัด Threshold in Quiet หรือระดับการบดบังสัญญาณโดย Uniform Masking Noise จึงสามารถกล่าวได้ว่าสิ่งที่มีผลต่อ Threshold in Quiet และระดับของการบดบังสัญญาณคือความเข้มของสัญญาณในแถบความถี่วิกฤตนั้นๆเท่านั้น



รูปที่ 3.13 ระดับ Threshold ของสัญญาณรบกวนแถบความถี่แคบที่มีความถี่ศูนย์กลางอยู่ระหว่างสองโทนที่ใช้ในการบดบังสัญญาณ โดยแสดงเปรียบเทียบกับความแตกต่างของความถี่ของสองโทน



รูปที่ 3.14 ระดับ Threshold ที่สัญญาณโทนทดสอบถูกบดบังโดยสัญญาณรบกวนสองแถบ (Bandpass Noise) โดยเปรียบเทียบกับความแตกต่างความถี่ของ Cutoff ของทั้งสองแถบ

อีกวิธีที่ใช้ในการประมาณความกว้างแถบความถี่วิกฤตคือการใช้สัญญาณโทนสองความถี่ที่มีระดับสัญญาณเท่ากันทำหน้าที่เป็น Masker และให้สัญญาณรบกวนแถบสัญญาณแคบทำหน้าที่เป็นสัญญาณทดสอบ โดยจะทำการวัด Threshold ของการที่สัญญาณทดสอบถูกบดบังเทียบกับความแตกต่างของความถี่ของสัญญาณโทนทั้งสองซึ่งมีสัญญาณทดสอบอยู่ที่ความถี่กึ่งกลาง ดังนั้นช่วงความถี่ของสัญญาณทดสอบจะต้องแคบกว่าความกว้างแถบสัญญาณความถี่วิกฤต ในรูปที่ 3.12 ภาพเล็กแสดงสัญญาณโทนทั้งสองที่ใช้ในการบดบังสัญญาณ และสัญญาณรบกวนแถบสัญญาณแคบที่ใช้เป็นเป็นสัญญาณทดสอบในแกนความถี่ และในรูปใหญ่เป็นข้อมูลระดับ Threshold เมื่อถูกบดบังด้วยสัญญาณโทนที่มีระดับ 50 dB สองค่าและสัญญาณทดสอบมีความถี่ศูนย์กลางที่ 2 kHz

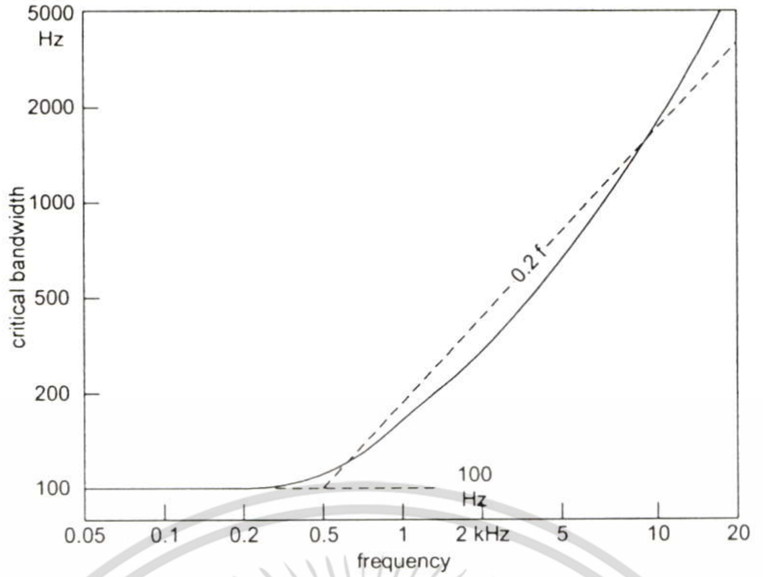
โดยระดับ Threshold ของสัญญาณรบกวนแถบความถี่แคบที่ถูกบดบังสัญญาณโดยสัญญาณโทนทั้งสองนั้นจะแสดงเปรียบเทียบกับความแตกต่างความถี่ของสัญญาณโทนทั้งสอง สำหรับที่ความแตกต่างของความถี่น้อยๆนั้นระดับ Threshold จะไม่ขึ้นกับความแตกต่างของความถี่ แต่เมื่อความแตกต่างของความถี่เกินกว่าระดับหนึ่งไปแล้วระดับ Threshold จะลดลงซึ่งจุดที่ความแตกต่างความถี่ค่านั้นก็คือ ความกว้างแถบความถี่วิกฤต เมื่อทำการวัดโดยเปลี่ยนค่าความถี่ศูนย์กลางของสัญญาณทดสอบไปเรื่อยๆก็จะทำให้เราได้ความสัมพันธ์ระหว่างความกว้างแถบความถี่วิกฤตกับความถี่

อีกวิธีการหนึ่งที่คล้ายกันจะแสดงในรูปที่ 3.14 เราจะใช้สัญญาณรบกวนแถบสัญญาณแคบสองช่วงความถี่มาเป็น Masker แล้วใช้สัญญาณโทนเป็นสัญญาณทดสอบโดยให้ความแตกต่างระหว่าง Higher และ Lower cutoff ของสัญญาณรบกวนแถบความถี่แคบทั้งสองเป็นความแตกต่างความถี่และให้มันเปลี่ยนค่าไปเรื่อยๆขณะที่ทำการวัดระดับ Threshold โดยสัญญาณทดสอบเป็นสัญญาณโทนความถี่ 2 kHz และระดับของสัญญาณรบกวนทั้งสองแถบที่ใช้เป็น Masker มีระดับ 50 dB และสัญญาณรบกวนแถบความถี่แคบที่มีแถบความถี่กว้าง 200 Hz ซึ่งจะเห็นได้ว่าได้ผลเช่นเดียวกันกับที่ได้จากวิธีการก่อนหน้านี้

นอกจากนี้ยังมีวิธีการอื่นๆอีกที่ใช้ในการหาความกว้างแถบความถี่วิกฤต เช่นการใช้การตรวจสอบการเปลี่ยนเฟส, การวัดระดับความดังของเสียง, และการวัดโดยใช้ Short Impulse เป็นต้น ซึ่งวิธีการทั้งหมดนี้จะให้ผลที่ใกล้เคียงกัน

3.7.2 Critical Band Rate Scale

มีวิธีการหลายอย่างที่ใช้ในการประมาณค่าความกว้างแถบความถี่วิกฤตซึ่งได้กล่าวผ่านมาในหัวข้อที่แล้ว ซึ่งแถบความถี่วิกฤตแรกที่อยู่ในช่วงความถี่ต่ำที่สุดของระบบการได้ยินของมนุษย์จะอยู่ที่ประมาณ 20-100 Hz แต่จะเป็นการง่ายกว่าหากเรารวมเอาช่วงความถี่ 0-20 Hz ที่เราไม่สามารถได้ยินได้รวมเข้าไว้ด้วย และกำหนดให้แถบความถี่วิกฤตแรกของการได้ยินของมนุษย์เป็น 0-100 Hz ในรูปที่ 3.15 เป็นข้อมูลที่ได้จากการหาค่าเฉลี่ยจากการใช้วิธีการต่างๆกับกลุ่มตัวอย่างจำนวนมาก แสดงให้เห็นว่าที่ความถี่ต่ำกว่า 500 Hz นั้นแถบความถี่วิกฤตจะมีความกว้างประมาณ 100 Hz ส่วนที่ความถี่สูงกว่านั้นแถบความถี่วิกฤตจะกว้างประมาณ 20 เปอร์เซ็นต์ของความถี่ศูนย์กลาง ส่วนค่าที่ได้จากผลการทดลองจริงๆนั้นอยู่ในตารางที่ 3.1 ที่แสดงทั้งของเขตบนและล่างของแถบความถี่วิกฤต

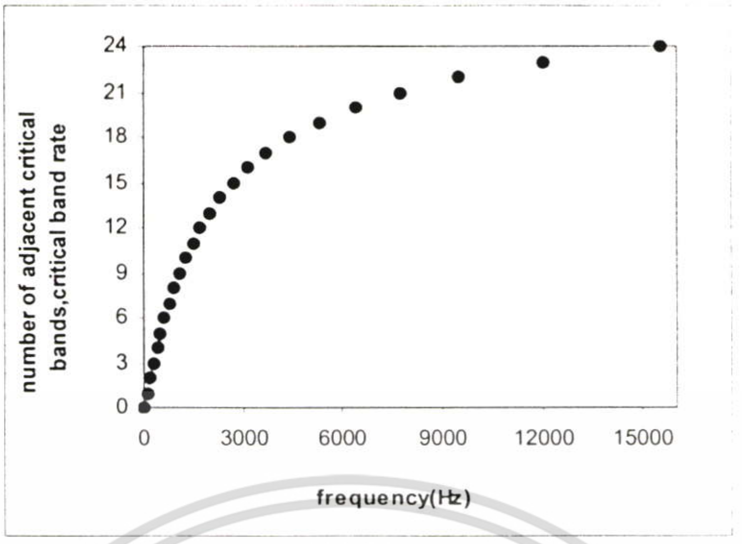


รูปที่ 3.15 ความกว้างแถบความถี่วิกฤตเทียบกับความถี่ และเส้นประแสดงการประมาณ

ความหมายของแถบความถี่วิกฤตนั้น มีความสำคัญต่อการอธิบายลักษณะในการได้ยินของมนุษย์อย่างมาก โดยจะถูกใช้ในแบบจำลองต่างๆในรูปแบบของสเกลที่เรียกว่า Critical Band Rate Scale ซึ่งสเกลนี้อ้างอิงจากหลักความจริงที่ว่า ระบบการรับฟังเสียงของคนเรานั้นจะทำการแบ่งการวิเคราะห์เสียงออกเป็นส่วนต่างๆตามแถบความถี่วิกฤต และถ้าเรายึดเอาแถบความถี่วิกฤตแรกเป็นจุดอ้างอิงแล้วทำการคำนวณแถบความถี่วิกฤตต่อมาเรื่อยๆโดยให้ขอบเขตบนของแถบความถี่วิกฤตปัจจุบันเป็นขอบเขตล่างของแถบความถี่วิกฤตถัดไปก็จะทำให้ได้สเกลที่เรียกว่า Critical Band Rate Scale ซึ่งจากตารางที่ 3.1 จะเห็นได้ว่าช่วงความถี่ถึง 16 kHz นั้นจะถูกแบ่งออกเป็นแถบความถี่วิกฤต 24 แถบ ส่วนในรูปที่ 3.16 เป็นแผนภูมิที่นำข้อมูลในตารางที่ 3.1 มาแสดงให้เห็น ซึ่งใน Critical Band Rate Scale นั้นอันดับของ Critical Band ถูกเรียกเป็น “Bark”

ตารางที่ 3.1 ขอบเขตและความถี่ศูนย์กลาง Critical Band Rate scale

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



รูปที่ 3.16 ความสัมพันธ์ระหว่างลำดับของ Critical Band กับความถี่

ซึ่งความสัมพันธ์ระหว่าง Critical Band Rate scale และ สเกลความถี่มีการประมาณการแปลงระหว่างความถี่และ Critical Band Rate ตลอดช่วงความถี่เสียงที่เราสามารถได้ยินดังต่อไปนี้[16]

$$z / Bark = 13 \text{ arc } \tan(0.76 f / kHz) + 3.5 \text{ arc } \tan(f / 7.5kHz)^2 \tag{3.32}$$

และ

$$\Delta f_G / Hz = 25 + 75[1 + 1.4(f / kHz)^2]^{0.69} \tag{3.33}$$

3.7.3 Critical Band Intensity

ความสามารถในการรับรู้เสียงในย่านความถี่ต่างๆของระบบการได้ยินเสียงของมนุษย์นั้นสามารถอธิบายอย่างคร่าวๆได้โดยความเข้มของสัญญาณเสียงที่ถูกคำนวณโดยแบ่งเป็นย่านความถี่ตามแถบความถี่วิกฤตซึ่งเรียกค่านี้ว่าเป็น Critical Band Intensity โดยค่า Critical Band Intensity สามารถคำนวณได้โดยใช้สมการดังนี้

$$I_G(f) = \int_{f-0.5\Delta f_G(f)}^{f+0.5\Delta f_G(f)} \frac{dI}{df}\, df \tag{3.34}$$

และสามารถเขียนได้เป็น

$$I_G(z) = \int_{z-0.5Bark}^{z+0.5Bark} \frac{dI}{dz}\, dz \tag{3.35}$$

สำหรับการแสดงค่าในสเกล logarithm นั้นเราใช้ I_0 เป็นค่าอ้างอิงดังนั้น Critical Band Level(L_G) สามารถหาได้จาก

$$L_G = 10 \cdot \log \frac{I_G}{I_0} dB \tag{3.36}$$

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

3.8 การหา K-Nearest Neighbor

3.8.1 ความหมายของ Nearest Neighbor

กำหนดให้มีสเปสแบบหลายมิติ (Multi-dimensional Space) ให้เวคเตอร์ข้อมูลสมาชิก (Exemplar) เป็นสมาชิกในเซตของเวคเตอร์ข้อมูลสมาชิก E และให้เซตของเวคเตอร์ข้อมูลสมาชิกประกอบไปด้วยเวคเตอร์ข้อมูลสมาชิกจำนวนจำกัดจำนวนหนึ่ง

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

$$None-nearer(E,\underline{d},\underline{d}') \Leftrightarrow \forall \underline{d}'' \in E \; |\underline{d}-\underline{d}'| \le |\underline{d}-\underline{d}''| \tag{3.37}$$

ในสมการที่5.1 ระยะห่างที่ใช้คือระยะทางแบบ Euclidean ซึ่งสามารถหาได้โดย

$$|\underline{d}-\underline{d}'| = \sqrt{\sum_{i=1}^{i=k_d} (d_i - d_i')^2} \tag{3.38}$$

โดยที่ d_i เป็นองค์ประกอบในมิติหนึ่งของเวคเตอร์ $\underline{d}$

3.9 กฎการตัดสินใจ (Decision Rule)

3.9.1 การแจกแจงความถี่(Frequency Distribution)

หมายถึง การจัดระเบียบข้อมูลที่มีอยู่อย่างกระจัดกระจายให้เป็นระเบียบมากยิ่งขึ้น ด้วยการจัดเรียงลำดับข้อมูลจากค่าน้อยไปมาก หรือจากค่ามากไปน้อย และรวบรวมจำนวนครั้งที่ข้อมูลที่รวบรวมมานั้นมีค่าซ้ำกัน โดยมีวัตถุประสงค์เพื่อนำข้อมูลที่จัดระเบียบแล้วนั้นมาใช้ประโยชน์ให้มากที่สุด

3.9.2 ความถี่สะสม (Cumulative Frequency)

หมายถึง ค่าของการสะสมค่าความถี่ของแต่ละชั้นข้อมูล

3.9.3 ตารางแจกแจงความถี่

หมายถึงวิธีการทางสถิติสำหรับการจัดระเบียบข้อมูลด้วยการเรียงลำดับค่า และแสดงความถี่ของข้อมูลให้อยู่ในรูปของตารางแจกแจงความถี่ โดยมีองค์ประกอบดังนี้

3.9.3.1 พิสัย หมายถึง ผลต่างระหว่างข้อมูลที่มีค่าสูงสุด กับต่ำสุด

$$\text{พิสัย} = \text{ค่าสูงสุด} - \text{ค่าต่ำสุด} \tag{3.39}$$

3.9.3.2 จำนวนชั้น หมายถึง จำนวนชั้นของข้อมูลที่ต้องการ โดยหากต้องการให้มีการกระจายข้อมูลมากให้กำหนดชั้นข้อมูลจำนวนน้อย แต่หากต้องการให้มีการกระจายข้อมูลน้อยให้กำหนดชั้นข้อมูลจำนวนมาก

3.9.3.3 อันตรภาคชั้น(Class Interval) หมายถึง ช่วงค่าในแต่ละชั้นของข้อมูล สามารถคำนวณโดยใช้สูตรดังนี้

$$\text{อันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}} \tag{3.40}$$

หลังจากที่ได้นำข้อมูลที่เก็บรวบรวมได้มาสร้างตารางแจกแจงความถี่เรียบร้อยแล้ว เราสามารถกำหนดค่ากลาง หรือตัวแทนของข้อมูล วิธีการกำหนดค่ากลางของข้อมูลดังกล่าว เรียกว่า “การวัดแนวโน้มเข้าสู่ส่วนกลาง” ซึ่งวิธีในการกำหนดค่ากลางข้อมูลแบ่งออกเป็น 3 วิธี วิธีแรกคือ การคำนวณค่ามัชฌิม(Mean) แบ่งออกเป็น 3 ชนิด ได้แก่ มัชฌิมเลขคณิตหรือค่าเฉลี่ย(Arithmetic Mean) แทนด้วยสัญลักษณ์ $\overline{X}$ (เอ็กซ์บาร์) มัชฌิมเรขาคณิต(Geometric Mean) และมัชฌิมฮาร์โมนิค (Harmonic Mean) วิธีที่สองคือการคำนวณค่ามัธยฐาน(Median) และวิธีที่สามคือการคำนวณค่า **ฐาน**นิยม(Mode)

3.9.4 การคำนวณค่าเฉลี่ยเลขคณิต(Arithmetic Mean)

ในช่วงข้อมูลที่กำหนด ซึ่งตามปกติสูตรทางสถิติที่ใช้ในการคำนวณค่าเฉลี่ย คือ

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{3.41}$$

เมื่อกำหนดให้ $i = 1,2,3,\ldots,N$

$\sum_{i=1}^{N} x_i$ หมายถึง ผลบวกของข้อมูลทั้งหมด

N หมายถึง จำนวนข้อมูลทั้งหมด

3.9.5 การคำนวณค่ามัธยฐาน(Median)

ในช่วงข้อมูลที่กำหนด การหาค่ามัธยฐานสามารถทำได้โดยการเรียงลำดับข้อมูลจากน้อยไปมาก หรือมากไปน้อย จากนั้นหาค่าของข้อมูลที่อยู่ตรงกึ่งกลาง โดยค่าที่ได้ หมายถึง ค่ามัธยฐานนั่นเอง

3.9.6 การหาค่าฐานนิยม(Mode)

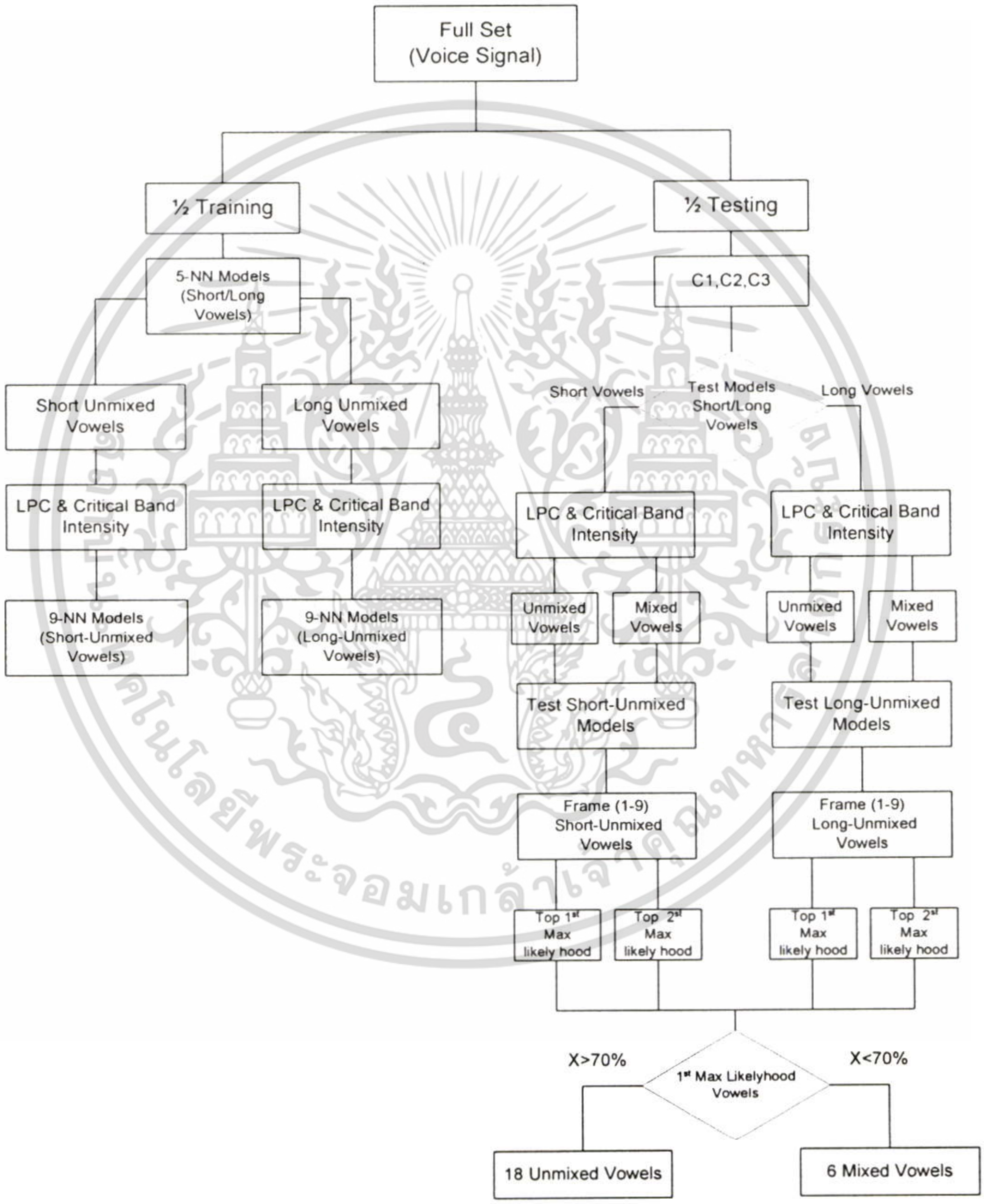
ในช่วงข้อมูลที่กำหนด ซึ่งค่าฐานนิยมสามารถหาได้ด้วยการสร้าง ตารางแจกแจงความถี่ โดยให้สังเกตค่าของข้อมูลที่มีความถี่มากที่สุด ค่านั้นคือค่าฐานนิยมนั่นเอง

บทที่ 4

การทดลองและผลการทดลอง

4.1 กล่าวนำ

ขั้นตอนการรู้จำเสียงสระ 24 เสียง แสดงได้ดังรูปที่ 4.1



รูปที่ 4.1 ขั้นตอนการรู้จำเสียงสระ 24 เสียง

จากรูปที่ 4.1 แสดงขั้นตอนการรู้จำเสียงสระ 24 เสียง โดยจะแบ่งออกเป็นขั้นตอนต่างๆ ดังนี้

4.1.1 ขั้นตอนการเก็บข้อมูลเสียง

4.1.2 ขั้นตอนในการกำหนดขอบเขตพยางค์

4.1.3 ขั้นตอนในการแยกสระเสียงสั้น และ สระเสียงยาว

4.1.4 ขั้นตอนในการวิเคราะห์สระเสียงเดี่ยว และ สระเสียงผสม

4.1.5 ขั้นตอนในการสร้างแบบอ้างอิง และแบบทดสอบ

4.1.6 ขั้นตอนการในทดสอบการรู้จำสระเสียงเดี่ยว แบบเฟรมโดยเทคนิค KNN

4.1.7 ขั้นตอนการทดสอบการรู้จำสระเสียงเดี่ยว และ สระเสียงผสม แบบคำโดยใช้กฎการตัดสินใจอย่างง่าย

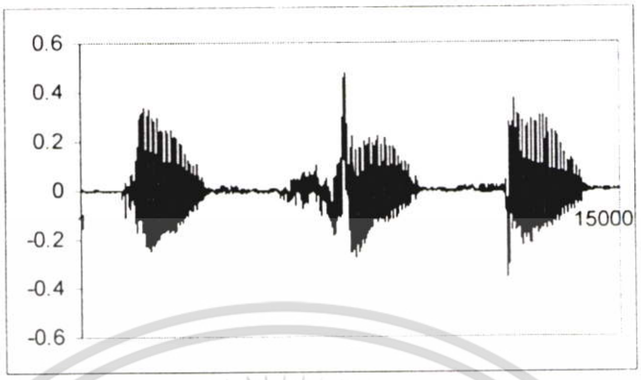
4.2 การเก็บข้อมูลเสียง

การเก็บข้อมูลในวิทยานิพนธ์นี้ จะใช้เสียงจากผู้พูดทั้งหมด 8 คน ประกอบด้วย ผู้พูดเพศชาย 4 คน ผู้พูดเพศหญิง 4 คน ผู้พูดแต่ละคนมีอายุเฉลี่ยอยู่ในช่วง 25-30 ปี ใช้สำเนียงภาษาไทยภาคกลาง ในการออกเสียง โดยจะให้ผู้ทดสอบแต่ละคนอ่านออกเสียงคนละ 504 คำ 2 รอบ โดยให้ผู้ทดสอบแต่ละคนอ่านออกเสียงสระ 24 รูปเสียง ที่ผันตาม พยัญชนะ 21 รูปเสียง ดังนั้นวิทยานิพนธ์นี้จะมีจำนวนคำที่ใช้ทั้งหมด 8,064 คำ ซึ่งจะนำไปสร้างแบบอ้างอิงจำนวน 4,032 คำ และสร้างแบบทดสอบจำนวน 4,032 คำ โดยการอ่านจะให้อ่านออกเสียงแบบไม่ต่อเนื่องคือจะให้อ่านแล้วหยุดสลับกันเพื่อให้คำที่พูดมีลักษณะไม่ต่อเนื่องกัน คำที่จะอ่านจะใช้เสียงพยัญชนะ 1 ตัว แล้วเปลี่ยนเสียงสระไปจนครบ 24 เสียง จากนั้นก็เปลี่ยนพยัญชนะใหม่จนครบ 21 เสียง ซึ่งเสียงที่อ่านนั้นจะมีเสียงวรรณยุกต์รวมอยู่ด้วย ในการบันทึกเสียงซึ่งข้อมูลที่ถูกบันทึกจะบันทึกอยู่ในรูปแบบของไฟล์ข้อมูล ".wav" ซึ่งข้อมูล 1 ตัวอย่างของเสียงจะถูกแทนด้วยข้อมูลขนาด 8 บิท โดยใช้ความถี่ในการแซมปลิ้งเสียงเท่ากับ 11.025 KHz ไฟล์ข้อมูลเสียง ".wav" จะถูกนำไปใช้เป็นอินพุทสำหรับโปรแกรมการคำนวณวิเคราะห์เสียงสระเดี่ยวในภาษาไทยที่เขียนขึ้น

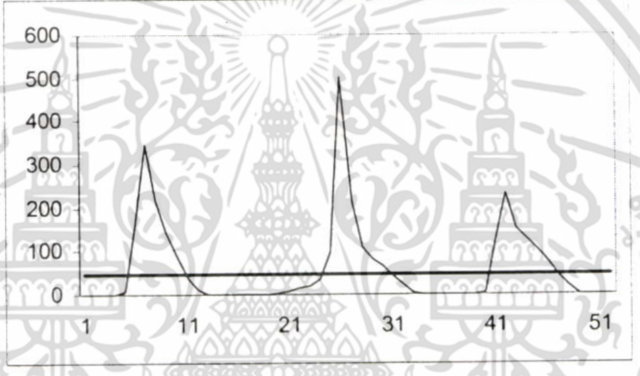
4.3 การกำหนดขอบเขตของพยางค์

หน่วยเสียงสระที่ลักษณะเป็นเสียงก้อง(voice) และมีลักษณะสัญญาณเป็นคาบ ซึ่งเป็นลักษณะที่เห็นได้ชัดในโดเมนของเวลา ดังนั้นการกำหนดขอบเขตของคำเพื่อที่จะวิเคราะห์หน่วยเสียงสระ จึงกำหนดขอบเขตภายในช่วงที่มีความเป็นคาบทั้งหมดของคำหรือพยางค์นั้นๆ รวมถึงการใช้ค่าระดับพลังงานร่วมด้วย โดยเริ่มจากสัญญาณเสียงที่ถูกเก็บมาในลักษณะที่ให้ผู้พูดพูดคำโดดโดยแต่ละคำมีช่วงห่างกันเล็กน้อย ดังรูปที่ 4.1(a) จากนั้นทำการคำนวณหาค่าพลังงานของเสียง

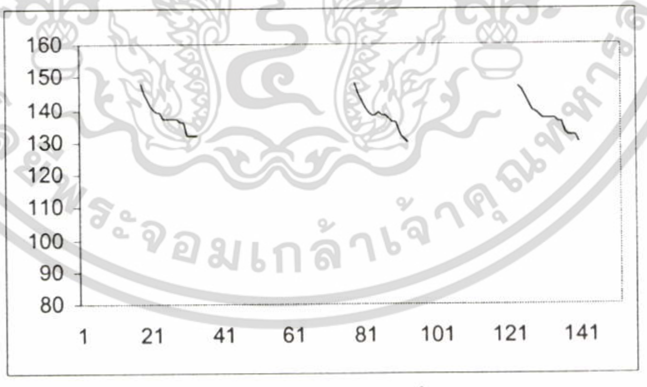
ดังรูปที่ 4.1(b) และค่าขอบเขตความถี่มูลฐานดังรูปที่ 4.1(c) นำค่าขอบเขตของพลังงานและขอบเขตของความถี่มาทับซ้อนกันดังรูปที่ 4.1(d) เพื่อใช้ในการตัดเสียงจะได้เสียงที่ตัดออกมาดังรูปที่ 4.1(e)



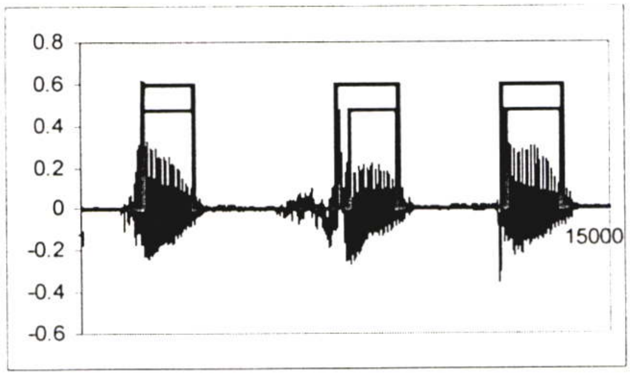
(a) แสดงตัวอย่างเสียงสระเสียง "กะ ขะ งะ"



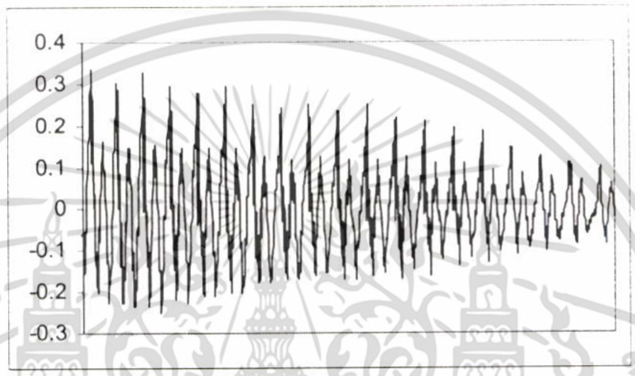
(b) ขอบเขตของค่าพลังงาน



(c) ขอบเขตของค่าความถี่มูลฐาน



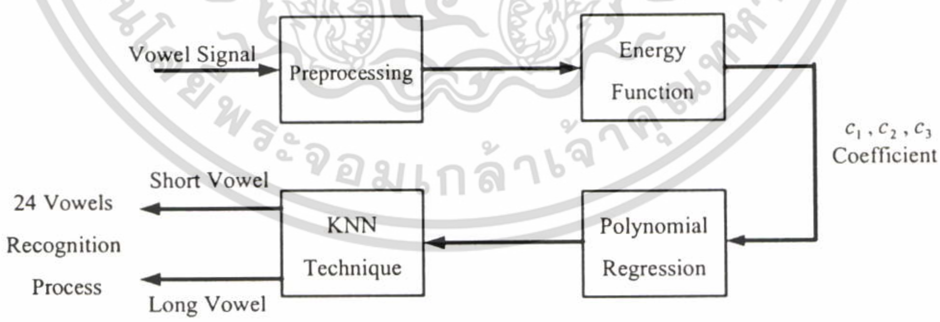
(d) ขอบเขตของทั้งสองวิธี โดยบริเวณที่ใช้คือบริเวณที่มีการทับซ้อนกันของทั้งสองวิธี



(e) แสดงตัวอย่างเสียงสัญญาณที่ผ่านการตัดเสียง

รูปที่ 4.2 การตัดเสียงโดยใช้ค่าพลังงานและความถี่มูลฐาน

4.4 การแยกแยะสระเสียงสั้น และ สระเสียงยาว

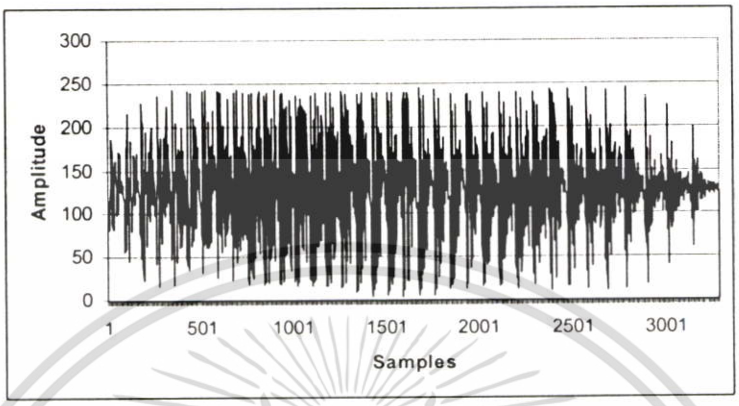


รูปที่ 4.3 แสดงขั้นตอนในการแยกแยะสระเสียงสั้นและสระเสียงยาว

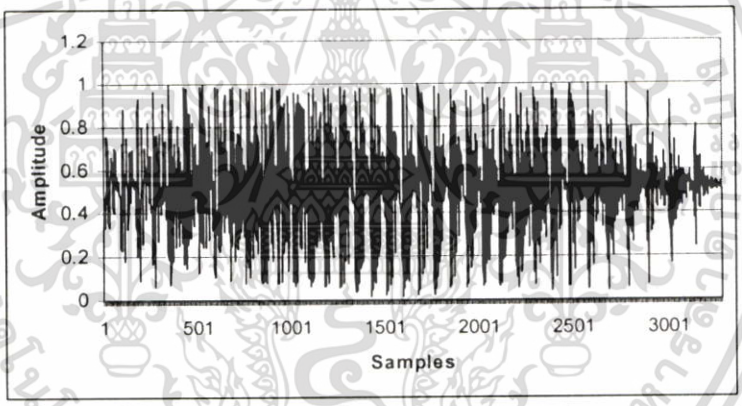
จากรูปที่ 4.2 แสดงขั้นตอนในการแยกแยะสระเสียงสั้นและสระเสียงยาวโดยวิธีการหาค่าพลังงานกำลังสอง และใช้การถดถอยแบบพหุนามอันดับสามมาสร้างแบบอ้างอิงและแบบทดสอบสระเสียงสั้นและสระเสียงยาวโดยใช้เทคนิค KNN ในการแยกแยะเสียงสระ

4.4.1 การเตรียมข้อมูลเสียง

การปรับบรรทัดฐานแอมปลิจูดใช้สมการที่ 2.1 โดยรูปที่ 4.4 และ รูปที่ 4.5 เป็นตัวอย่างสัญญาณเสียงที่เข้ามา และสัญญาณเสียงที่ผ่านการปรับบรรทัดฐาน

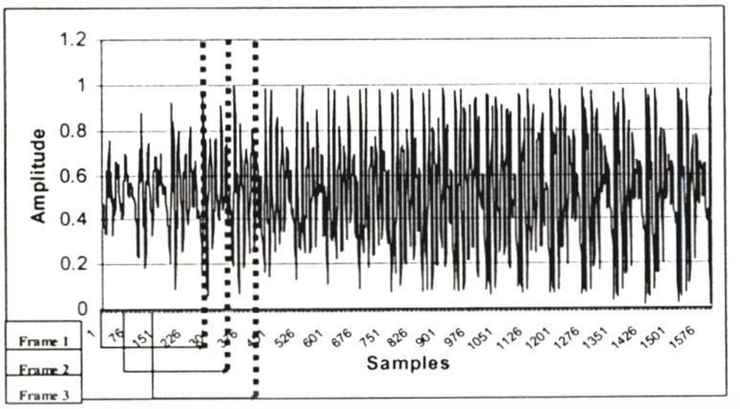


รูปที่ 4.4 สัญญาณเสียง “กะ”

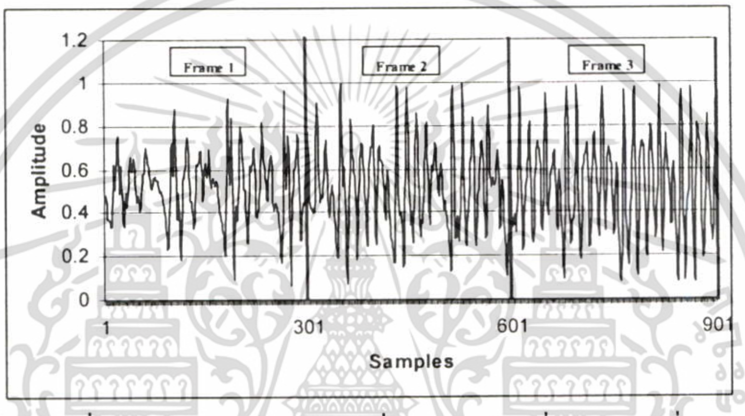


รูปที่ 4.5 สัญญาณเสียง “กะ” ที่ผ่านการปรับบรรทัดฐาน Amplitude

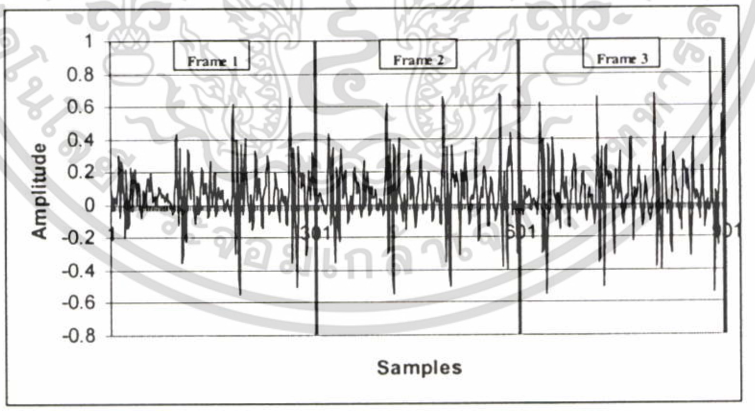
การคำนวณหาค่าพลังงานเสียง เริ่มต้นจากการแบ่งช่วงสัญญาณออกเป็นเฟรม เฟรมละ 300 samples และการวิเคราะห์แต่ละเฟรมจะเลื่อนไปครั้งละ 75 samples(overlap ¾) ตามรูปที่ 4.6 และรูปที่ 4.7 หลังจากนั้นเป็นการปรีเอ็มฟาซิสให้ความถี่สูงมีขนาดสูงขึ้นซึ่งจะได้เป็นดังรูปที่ 4.8 แล้วจึงทำการวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณบริเวณขอบ โดยการใช้แฮมมิ่งวินโดว์ซึ่งจะได้ผลดังรูปที่ 4.9 จากนั้นจึงนำสัญญาณที่ผ่านแฮมมิ่งวินโดว์ ไปคำนวณหาค่าพลังงานเสียงในรูปแบบต่างๆ ตามที่ได้กล่าวมาใน หัวข้อที่ 2.6



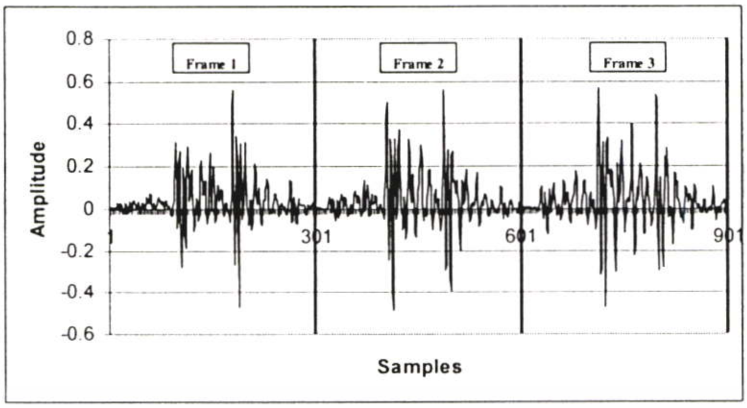
รูปที่ 4.6 สัญญาณเสียง "กะ" ที่แบ่งช่วงสัญญาณ 300 samples ต่อ 1 เฟรม



รูปที่ 4.7 สัญญาณเสียง "กะ" ที่ถูกแบ่งเฟรมที่ 1ถึงเฟรมที่ 3



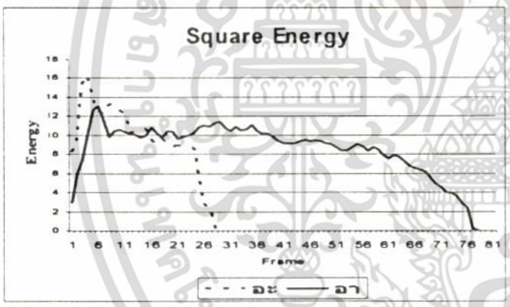
รูปที่ 4.8 สัญญาณเสียง "กะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านการ pre-emphasis



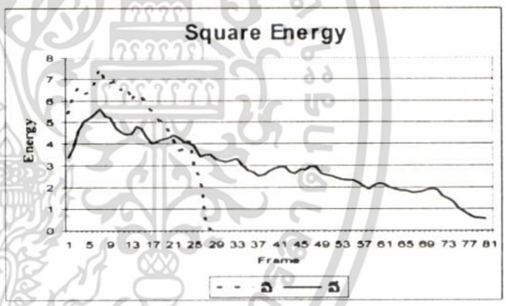
รูปที่ 4.9 สัญญาณเสียง "กะ" ในเฟรมที่ 1 ถึง 3 ที่ผ่านฟังก์ชั่นวินโดว์แบบแฮมมิ่ง

4.4.2 การหาค่าพลังงานเสียงกำลังสอง

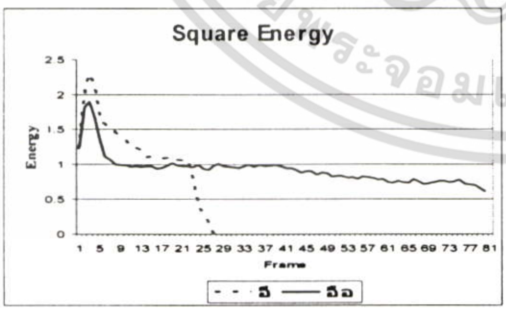
การคำนวณหาค่าพลังงานเสียง จะทำทีละกรอบเสียงพูด โดยมีตัวอย่างกรอบเสียงพูดละ 300 Samples และค่าพลังงานเสียงในกรอบต่อไปจะถูกเลื่อนไปอีก 75 Samples ซึ่งการคำนวณหาค่าพลังงานของสัญญาณเสียงกำลังสอง จะได้ลักษณะพลังงานเสียง ดังนี้



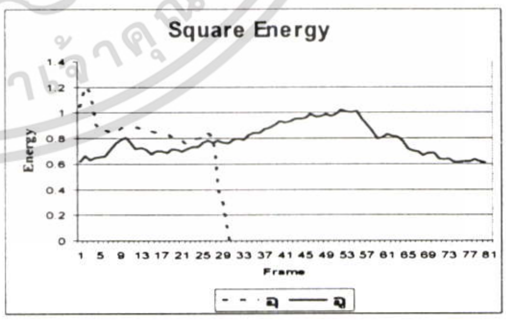
(a) เสียง "อะ-อา"



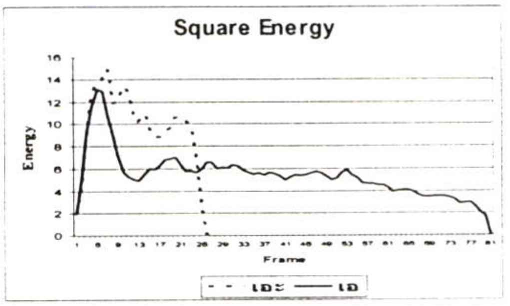
(b) เสียง "อิ-อี"



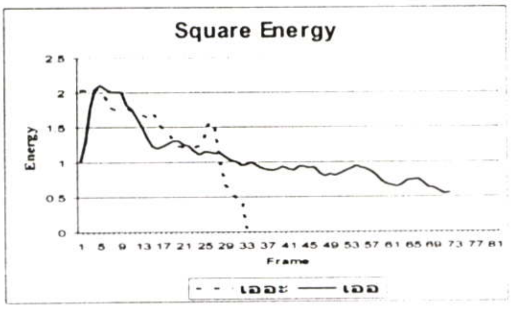
(c) เสียง "อึ-อือ"



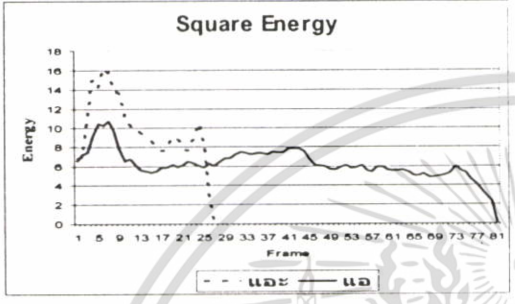
(d) เสียง "อุ-อู"



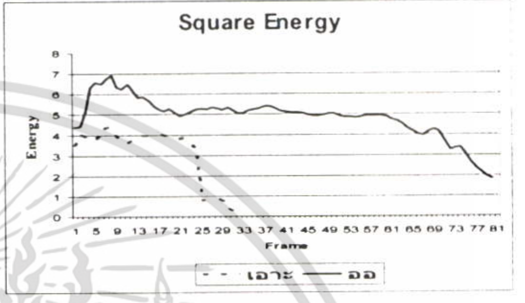
(e) เสียง "เอะ-เอ"



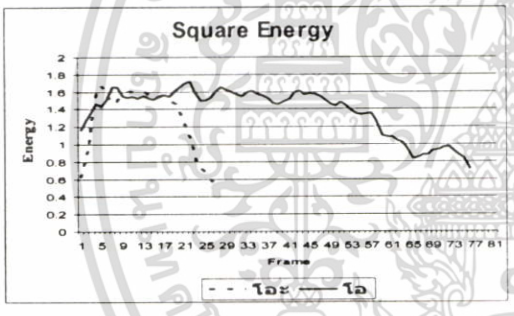
(f) เสียง "เออะ-เออ"



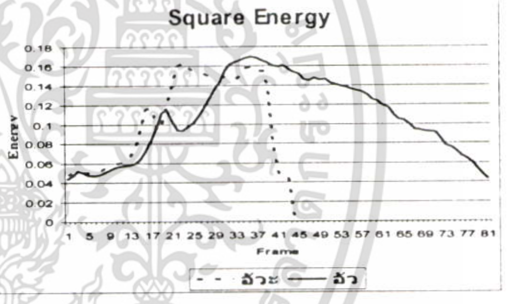
(g) เสียง "แอะ-แอ"



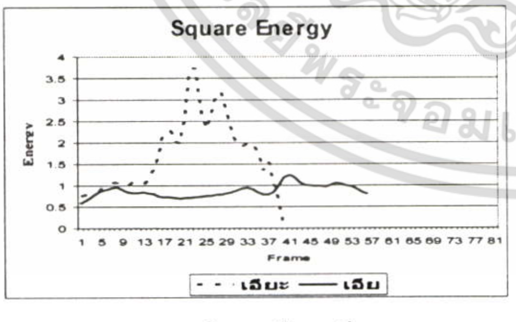
(h) เสียง "เอาะ-ออ"



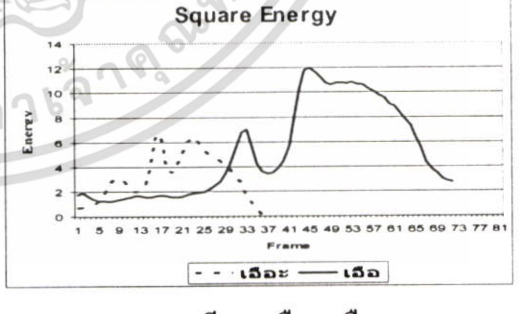
(i) เสียง "โอะ-โอ"



(j) เสียง "อัวะ-อัว"



(k) เสียง "เอียะ-เอีย"



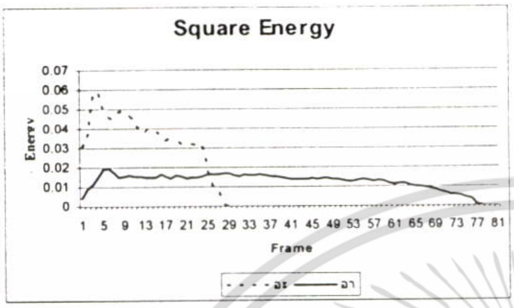
(l) เสียง "เอือะ-เอือ"

รูปที่ 4.10 แสดงลักษณะพลังงานของเสียงสระ

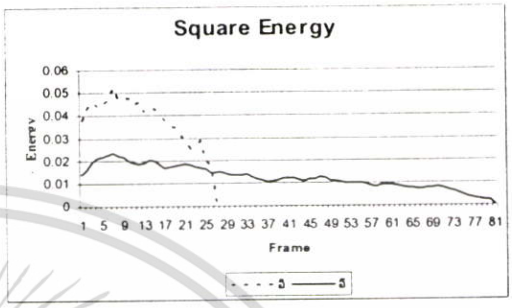
จากรูปที่ 4.10 ลักษณะพลังงานของเสียงสระ จะเห็นได้ว่าลักษณะการหยุดเสียงของสระเสียงสั้นนั้นจะเป็นการหยุดแบบทันทีทันใด ส่วนสระเสียงยาวจะมีการหยุดเสียงแบบค่อยๆ ลดระดับลงช้าๆ

4.4.3 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชั่นพลังงานเสียง (Normalized Energy function)

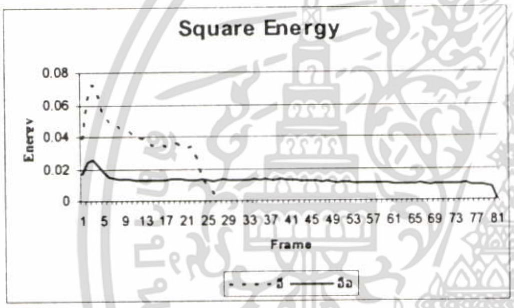
เมื่อได้ลักษณะพลังงานเสียง นำพลังงานเสียงที่ได้มาปรับบรรทัดฐานพลังงาน โดยให้พื้นที่ใต้ฟังก์ชั่นพลังงานเสียง มีค่า เท่ากับ 1 ทุกๆเสียง ตามหัวข้อ 2.10 ซึ่งจะทำให้เห็นความแตกต่างพลังงานเสียงสระเสียงสั้น และ สระเสียงยาว มากขึ้น ดังรูปที่ 4.11



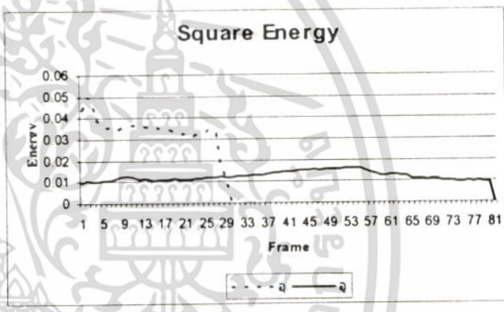
(a) เสียง “อะ-อา”



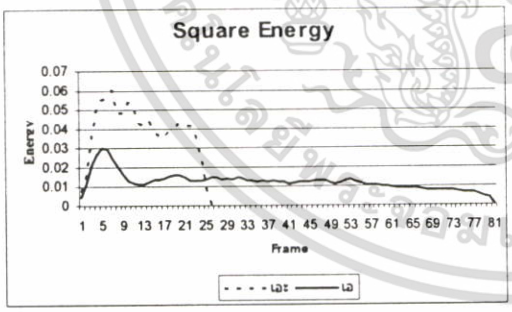
(b) เสียง “อิ-อี”



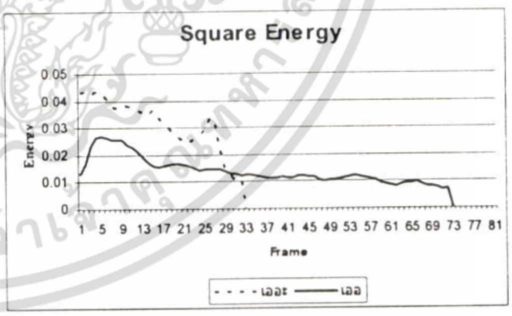
(c) เสียง “อึ-อือ”



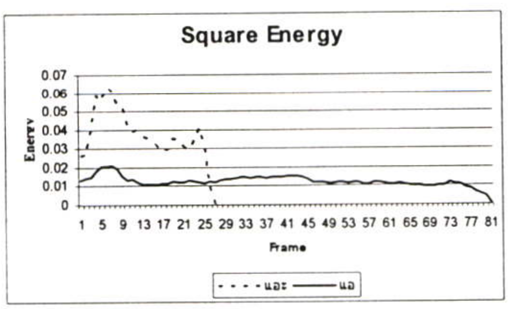
(d) เสียง “อุ-อู”



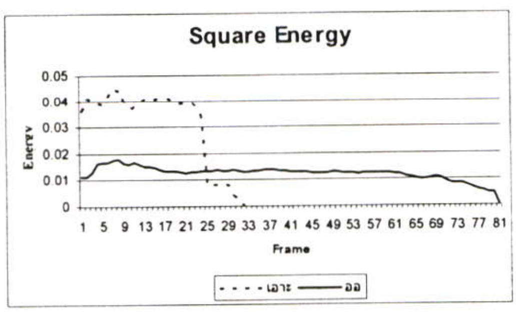
(e) เสียง “เอะ-เอ”



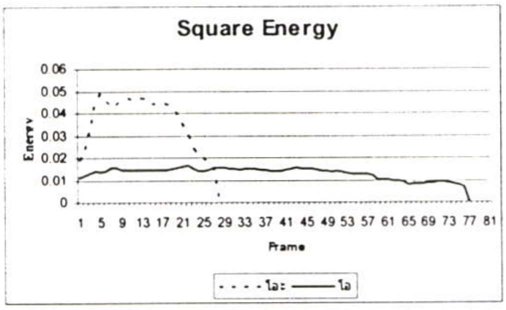
(f) เสียง “เออะ-เออ”



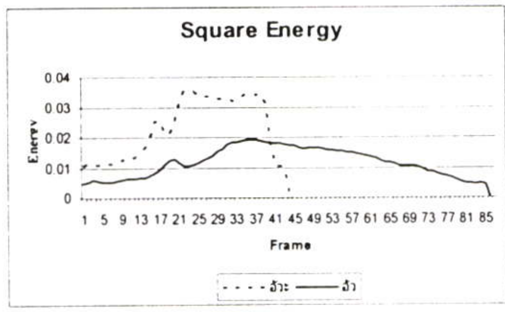
(g) เสียง “แอะ-แอ”



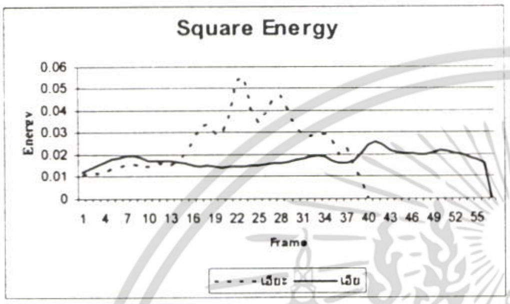
(h) เสียง “เอาะ-ออ”



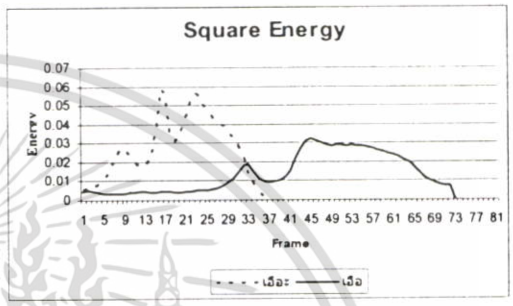
(i) เสียง “โอะ-โอ”



(j) เสียง “อัวะ-อัว



(k) เสียง “เอียะ-เอีย”

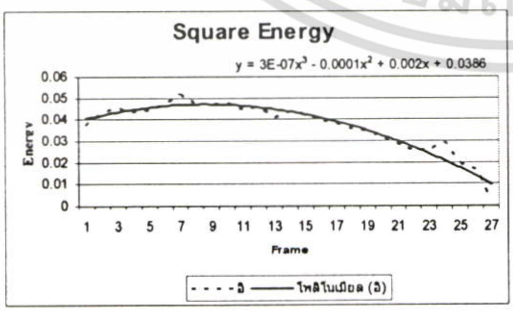


(l) เสียง “เอือะ-เอือ”

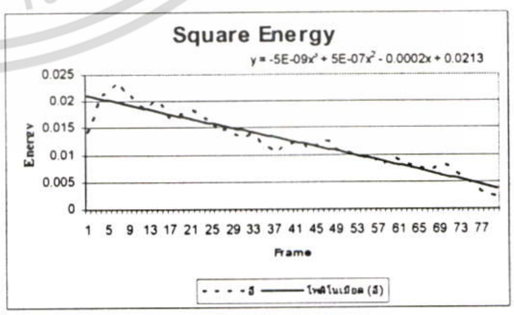
รูปที่ 4.11 แสดงลักษณะพลังงานที่ผ่านการ Normalize

4.4.4 การถดถอยแบบพหุนามของพลังงานเสียง

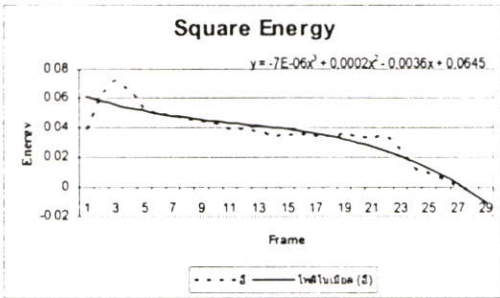
เมื่อได้ลักษณะของพลังงานเสียง นำพลังงานเสียงที่ได้มาปรับบรรทัดฐานทางแอมปลิจูดด้วยค่าสูงสุดของพลังงาน เพื่อให้ในขั้นตอนของการประดิษฐ์ฟังก์ชั่นการถดถอยพหุนามเป็นบรรทัดฐานเดียวกัน ซึ่งฟังก์ชั่นถดถอยพหุนามในงานวิจัยนี้ใช้ฟังชั่นการถดถอยพหุนามอันดับสาม โดยจะได้สัมประสิทธิ์ฟังก์ชั่นการถดถอยพหุนาม เป็นค่า $c3, c2, c1, c0$ ตามที่ได้กล่าวมาในหัวข้อ 3.5 โดยลักษณะฟังก์ชั่นการถดถอยพหุนามของเสียงสระต่างๆ จะเป็นลักษณะดังรูปที่ 4.12



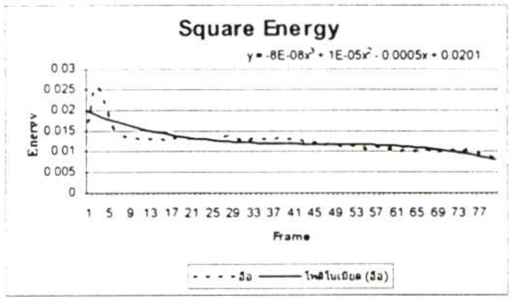
(c) เสียง “อิ”



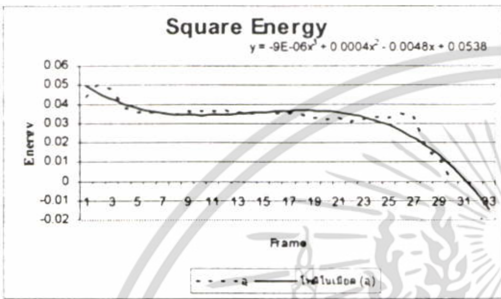
(d) เสียง “อี”



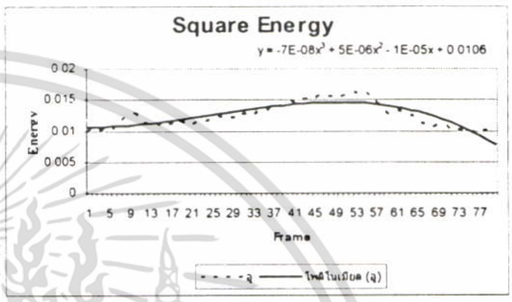
(c) เสียง “อึ”



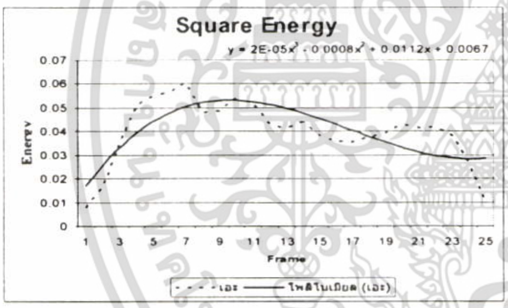
(d) เสียง “อือ”



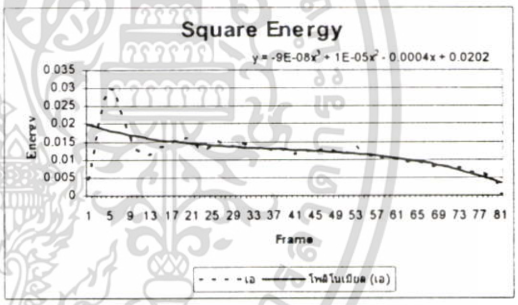
(e) เสียง “อุ”



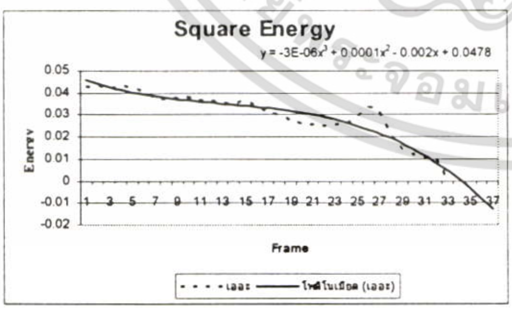
(f) เสียง “อู”



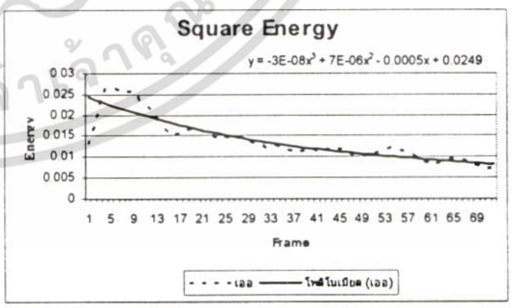
(g) เสียง “เอะ”



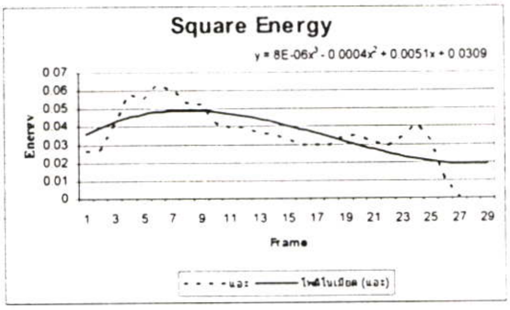
(h) เสียง “เอ”



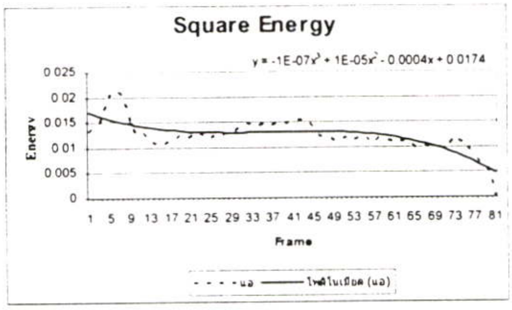
(i) เสียง “เออะ”



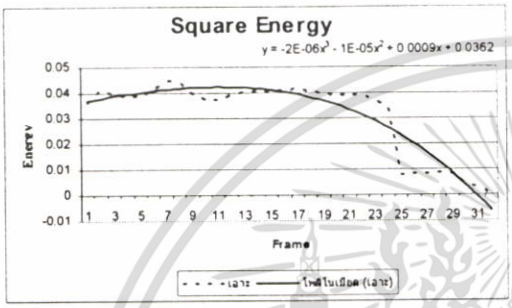
(j) เสียง “เออ”



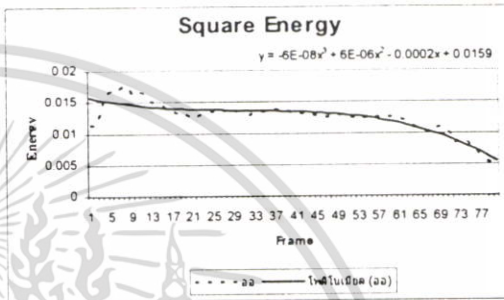
(k) เสียง “แอะ”



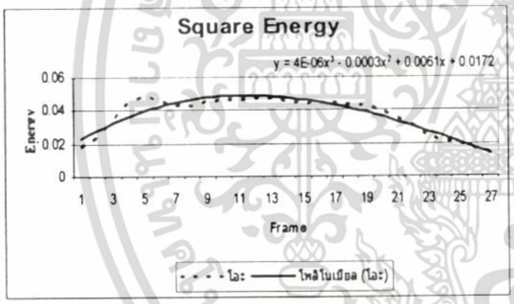
(l) เสียง “แอ”



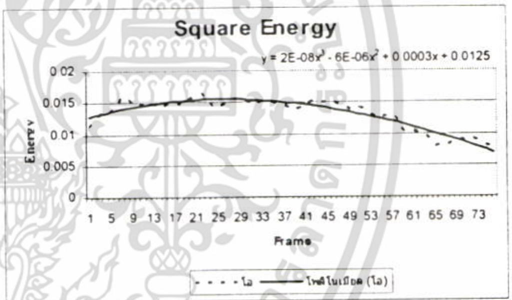
(m) เสียง “เอาะ”



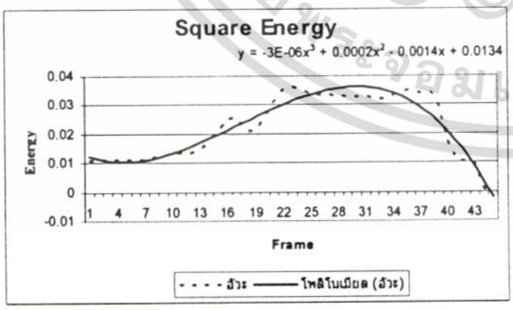
(n) เสียง “ออ”



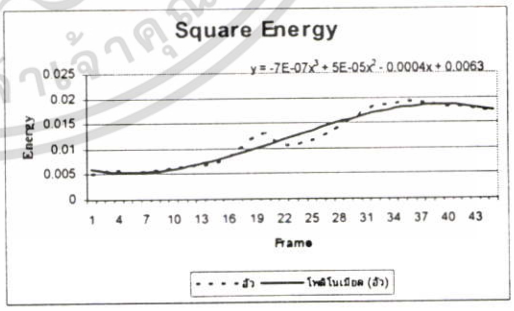
(o) เสียง “โอะ”



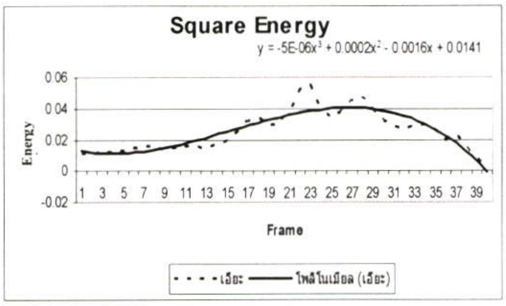
(p) เสียง “โอ”



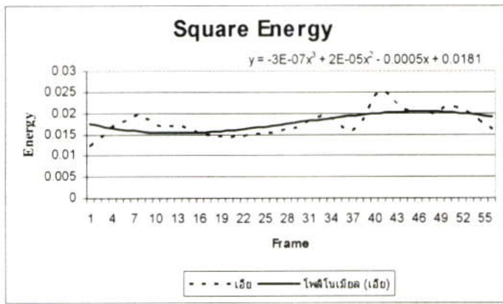
(q) เสียง “อัวะ”



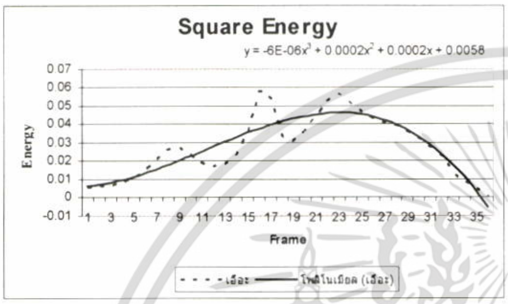
(r) เสียง “อัว”



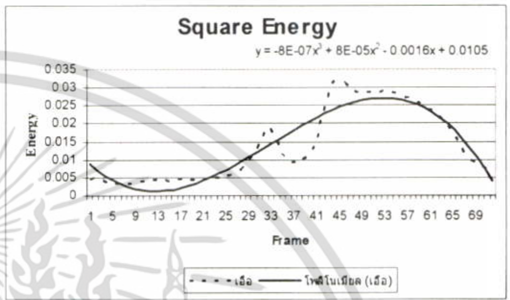
(s) เสียง “เอียะ”



(t) เสียง “เอีย”



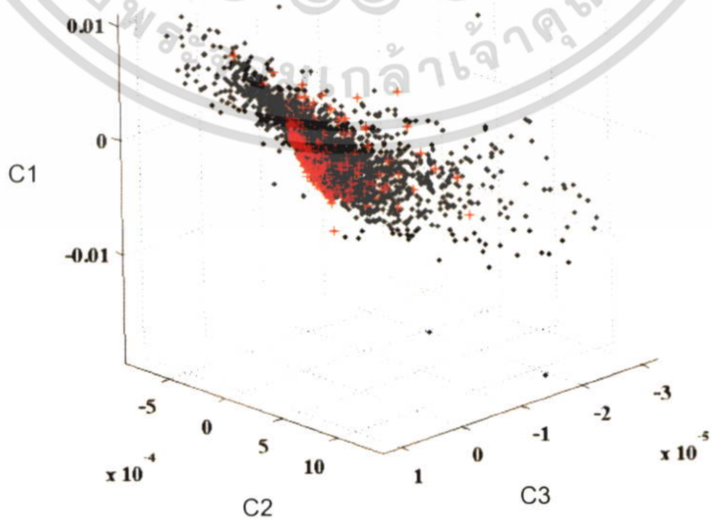
(u) เสียง “เอือะ”



(v) เสียง “เอือ”

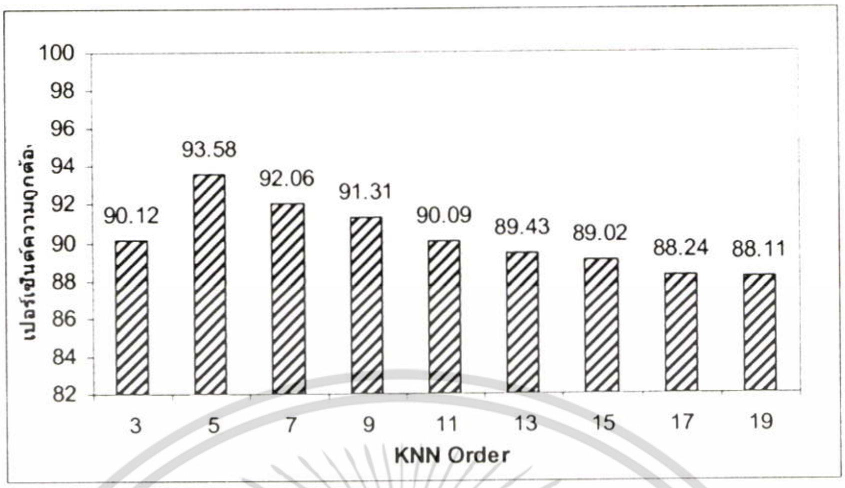
รูปที่ 4.12 ฟังก์ชั่นพหุนามของเสียงสระ

จากรูปการกระจายข้อมูล c_1, c_2 และ c_3 นำค่าสัมประสิทธิ์ c_1, c_2 และ c_3 ไปจัดกลุ่มสระเสียงสั้น และ สระเสียงยาว เป็น Vector 3 มิติ จากสัมประสิทธิ์การถดถอยพหุนาม c_1, c_2 และ c_3 ของตัวอย่างเสียง 1,512 เสียง (สระเสียงสั้น 756 เสียง, สระเสียงยาว 756 เสียง)



รูปที่ 4.13 การกระจายข้อมูลสัมประสิทธิ์ c_1, c_2, c_3

4.4.5 การแยกแยะสระเสียงสั้นและสระเสียงยาวโดยใช้เทคนิค KNN (K-Nearest Neighbor)



รูปที่ 4.14 อัตราการแยกสระเสียงสั้นและสระเสียงยาวของเสียงทดสอบที่ค่า k ต่างๆ

จากรูปที่ 4.14 จะพบว่า ค่า k ที่ให้อัตราการแยกแยะสระเสียงสั้นและสระเสียงยาวมากที่สุดคือ k= 5 ซึ่งให้อัตราการแยกแยะดังนี้

ตารางที่ 4.1 ผลการทดสอบการแยกแยะสระเสียงสั้นและสระเสียงยาว ที่ k= 5

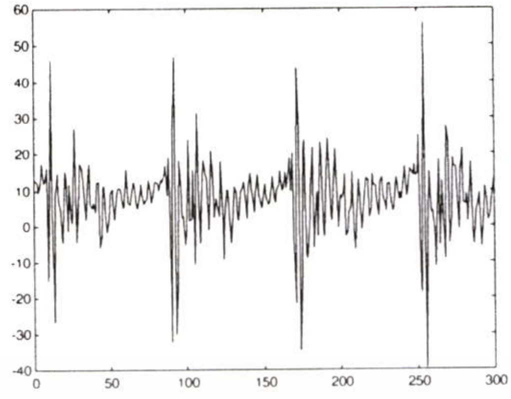
ผู้พูด	สระเสียงสั้น		สระเสียงยาว		จำนวนเสียงทั้งหมด	%ความถูกต้อง
	ค่าที่แยกถูกต้อง	ค่าที่แยกไม่ถูกต้อง	ค่าที่แยกถูกต้อง	ค่าที่แยกไม่ถูกต้อง		
ชาย 3 คน	680	76	715	41	1512	92.26
หญิง 3 คน	694	62	721	35	1512	93.58
ชาย 1 หญิง 1	476	28	484	20	1008	95.24

จากตารางที่ 4.1 แสดงเปอร์เซ็นต์ความถูกต้องในการแยกแยะสระเสียงสั้น และสระเสียงยาวของเสียงผู้พูดเพศชาย 3 คน ผู้พูดเพศหญิง 3 คน และเสียงผู้พูดเพศชาย 1 คนเพศหญิง 1 คน ซึ่งผลที่ได้จากการแยกแยะสระเสียงสั้น และสระเสียงยาวนี้ จะนำ ไปวิเคราะห์และทดสอบในขั้นตอนของการรู้จำเสียงสระ 24 เสียง โดยจะนำเฉพาะค่าที่แยกแยะเสียงสระได้ถูกต้องไปทดสอบ

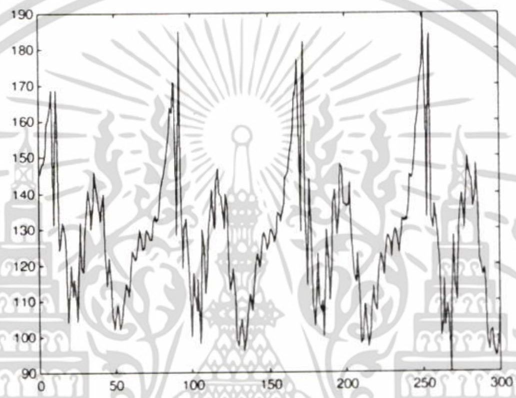
4.5 การหาค่าพารามิเตอร์ที่ใช้ในวิเคราะห์การแยกหน่วยเสียงสระเสียงเดี่ยว

4.5 .1 การหาสเปกตรัมแอลพีซี

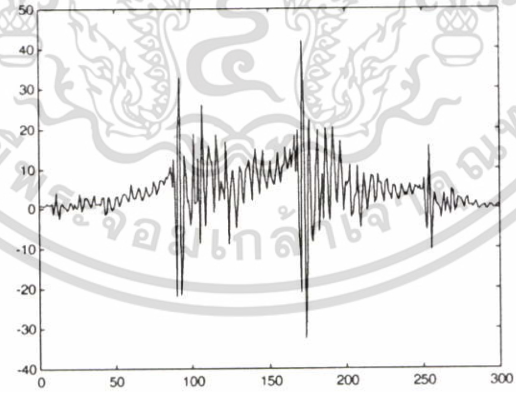
วิธีการหาสเปกตรัมแอลพีซีนั้นได้กล่าวมาแล้วในบทที่3 แสดงได้ดังรูปที่4. 15 (a)-(e) ซึ่งจะเป็นการนำสัญญาณตัวอย่างที่ผ่านกระบวนการตัดเสียงแล้วมาคำนวณหาค่าตามขั้นตอนต่างๆ เพื่อให้ได้ค่าสเปกตรัมแอลพีซี



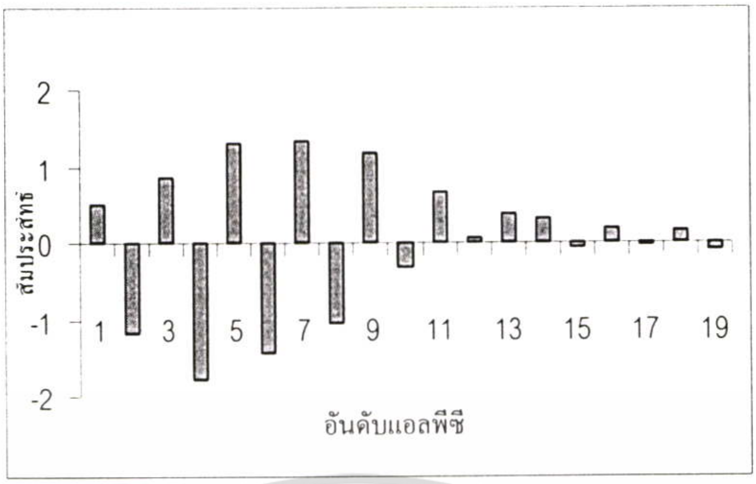
(a) ตัวอย่างสัญญาณที่ผ่านการพรีเอมฟาซีส



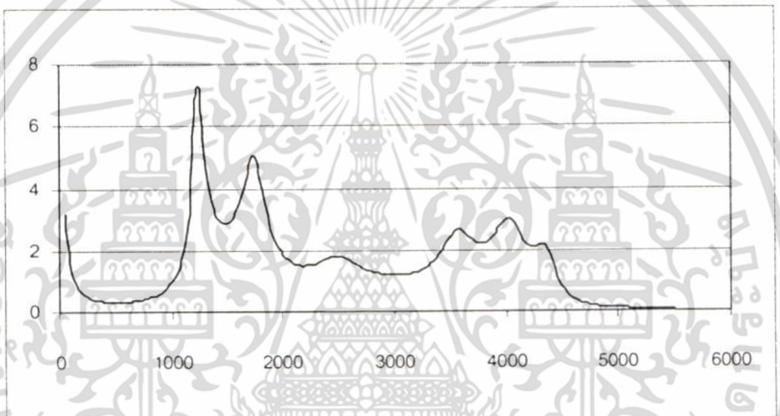
(b) ตัวอย่างสัญญาณที่ถูกแบ่งออกเป็นเฟรม



(c) ตัวอย่างสัญญาณที่ผ่านการแฮมมิ่งวินโดว์



(d) ค่าสัมประสิทธิ์ที่ได้จากการคำนวณแอลพีซีอันดับที่19



(e) สเปกตรัมแอลพีซี

รูปที่ 4.15 ขั้นตอนต่างๆ ในการคำนวณหาค่าสเปกตรัมแอลพีซี

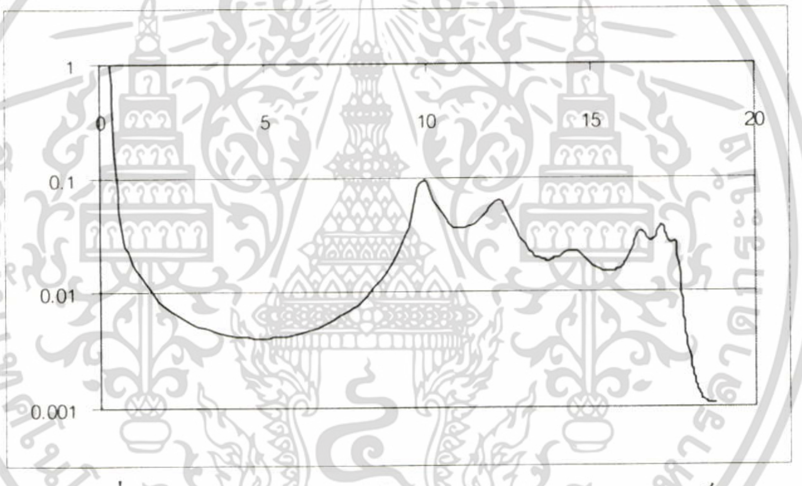
จากรูปที่ 4.15 แสดงให้เห็นถึงขั้นตอนต่างๆ ในการคำนวณหาค่าแอลพีซีสเปกตรัมของสัญญาณเสียง โดยเริ่มจากสัญญาณเสียงที่ผ่านการตัดเสียงแล้วนำมาทำการพรีเอมฟาซีส เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวนมีค่าค่อนข้างคงที่ตลอดช่วงความถี่ โดยเน้นความถี่สูงให้มีขนาดสูงขึ้นดังรูปที่ 4.15 (a) สัญญาณที่ผ่านการพรีเอมฟาซีสแล้วจะถูกแบ่งออกเป็นเฟรม เฟรมละ 300 แซมเปิ้ล การวิเคราะห์สัญญาณเสียงจะวิเคราะห์ทีละเฟรมโดยแต่ละเฟรมจะถูกเลื่อนไปทีละ 100 แซมเปิ้ล ดังรูปที่ 4.15 (b) จากนั้นนำสัญญาณแต่ละเฟรมมาคูณด้วยฟังก์ชันวินโดว์ เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไป ดังรูปที่ 4.15 (c) แล้วนำสัญญาณที่ผ่านการวินโดว์ไปคำนวณหาสัมประสิทธิ์แอลพีซี ได้ดังรูปที่ 4.15 (d) เมื่อได้ค่าสัมประสิทธิ์แอลพีซีแล้วก็สามารถคำนวณหาสเปกตรัมแอลพีซีได้ ดังรูปที่ 4.15 (e)

การคำนวณหาค่าสัมประสิทธ์แอลพีซีนั้นจำเป็นต้องเลือกอันดับของแอลพีซีให้เหมาะสมถ้าอันดับแอลพีซีมีค่าต่ำจะทำให้การคำนวณเร็วขึ้น แต่ค่าสเปกตรัมแอลพีซีที่คำนวณได้มีรายละเอียด

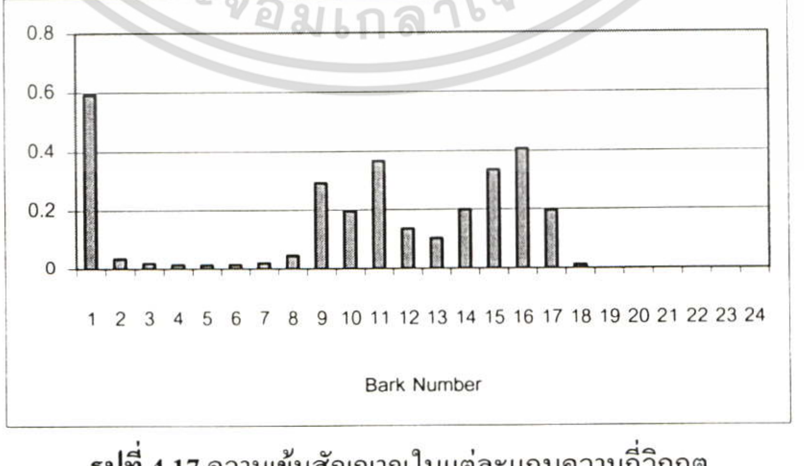
น้อยส่งผลให้การคำนวณหา Critical Band Intensity เกิดความผิดพลาด แต่ถ้าค่าอันดับแอลพีซีมีค่าสูงจะใช้เวลานาน ทำให้การคำนวณยุ่งยากและซับซ้อนเกินความจำเป็น ดังนั้นในวิทยานิพนธ์นี้จึงเลือกใช้แอลพีซีอันดับที่ 19 [5] ซึ่งให้ประสิทธิภาพสูงสุด

4.5.2 การแปลงจากสเกลความถี่ ไปเป็นสเกลบาร์ก หรือ Critical Band Rate Scale และการคำนวณหา Critical Band Intensity

สเกลบาร์กหรือ Critical Band Rate Scale นั้นช่วยให้ความสามารถในการแสดงคุณสมบัติของการตอบสนองต่อความถี่ต่างๆ ของการได้ยินของมนุษย์เป็นเชิงเส้นตลอดช่วงความถี่ที่มนุษย์สามารถได้ยิน ดังนั้นจึงทำการแปลงสเปกตรัมแอลพีซีที่เป็นความถี่ไปบนสเกลบาร์กได้ดังรูปที่ 4.16 แล้วนำค่าสเปกตรัมแอลพีซีที่อยู่บนสเกลบาร์ก ไปคำนวณหาค่าความเข้มของสัญญาณ ในแต่ละความถี่วิกฤต ซึ่งค่าที่ได้นี้จะนำไปใช้เป็นองค์ประกอบของเวกเตอร์แทนหน่วยเสียงต่างๆ ดังแสดงในรูปที่ 4.17



รูปที่ 4.16 สเปกตรัมแอลพีซีของรูป 4.15(e) บนสเกลบาร์ก

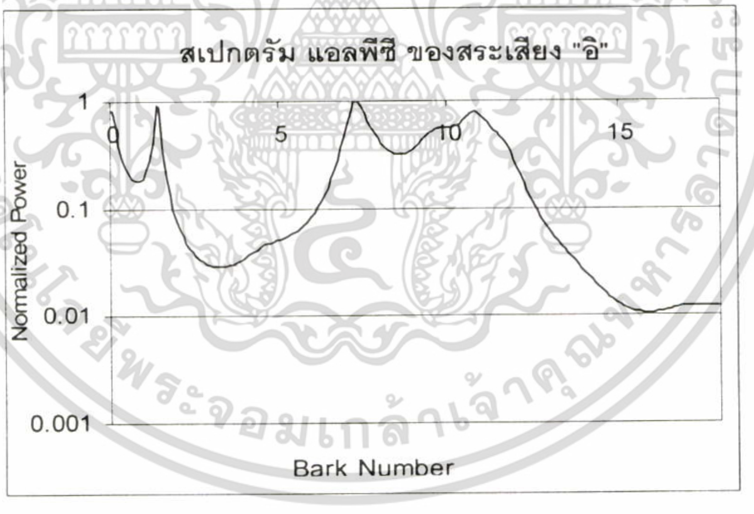


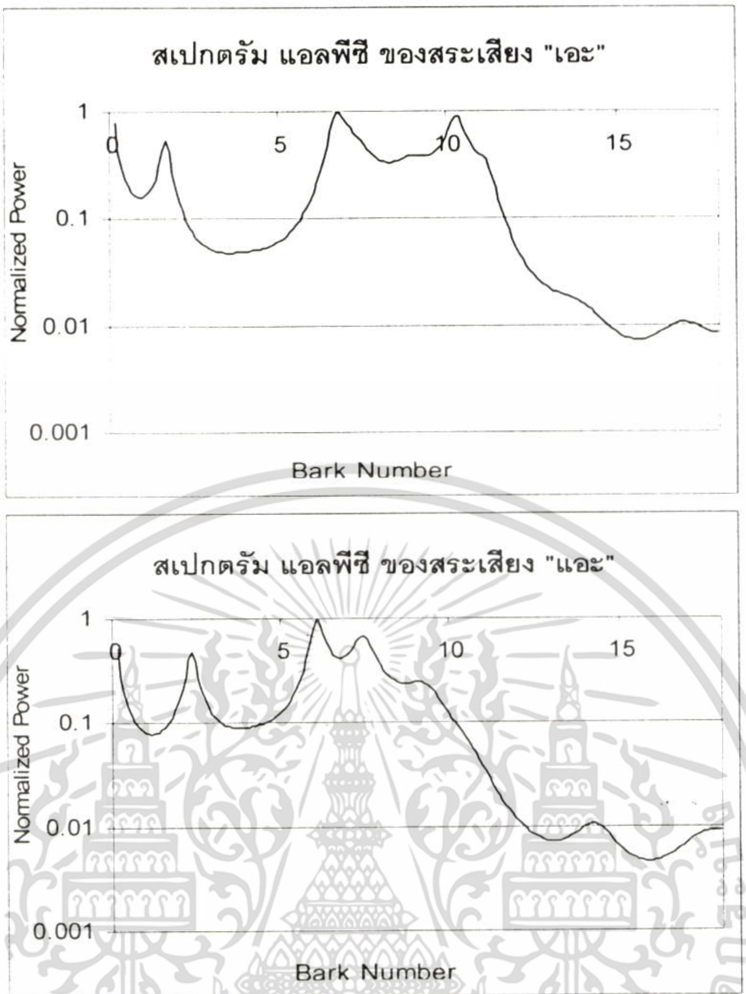
รูปที่ 4.17 ความเข้มสัญญาณในแต่ละแถบความถี่วิกฤต

จากรูปที่ 4.16 และรูปที่ 4.17 เมื่อนำมาเปรียบเทียบกันจะเห็นว่าที่ความถี่ 5.5 kHz นั้น จะมีค่าตรงกับความเข้มสัญญาณในแต่ละความถี่วิกฤตบาร์กที่ 18 ซึ่งแสดงว่าค่าความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตที่จะใช้เป็นองค์ประกอบของเวกเตอร์แทนหน่วยเสียงต่างๆ จะมี 18 ค่าสำหรับสัญญาณเสียงแต่ละเฟรม ค่าองค์ประกอบของเวกเตอร์แทนหน่วยเสียงต่างๆ ที่ได้จะนำไปใช้สร้างแบบอ้างอิงและแบบทดสอบของการรู้จำเสียงสระเสียงเดี่ยวแต่ละเสียง

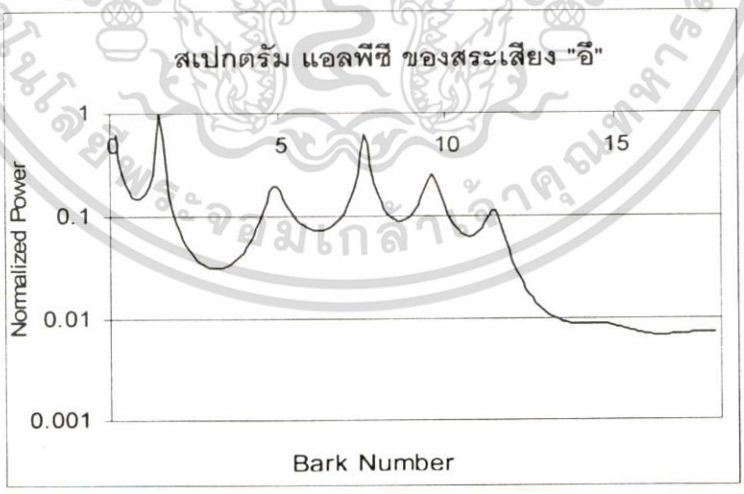
4.6 การสร้างแบบอ้างอิงและแบบทดสอบ

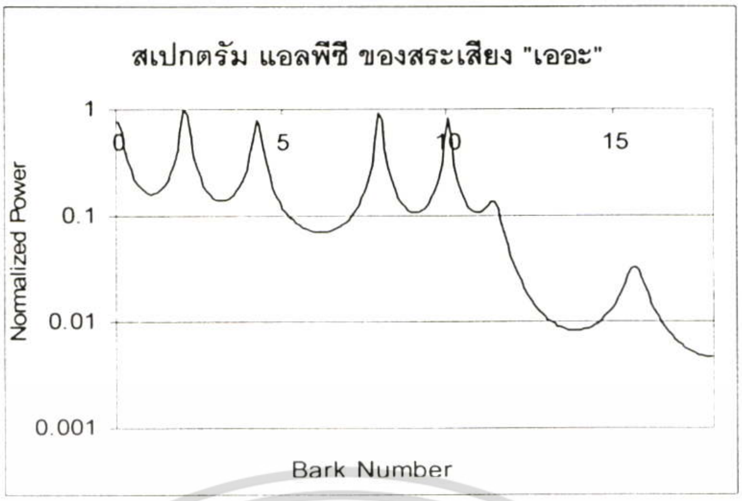
การทดลองนี้จะใช้เสียงของผู้พูด 8 คน (ผู้ชาย 4 คน และ ผู้หญิง 4 คน) โดยจะให้แต่ละคนพูดคนละ 504 คำ 2 รอบ โดยใช้เสียงพยัญชนะ 21 เสียง รวมกับเสียงสระเดี่ยว 18 เสียง และสระผสม 6 เสียง จะได้ตัวอย่างเสียงทั้งหมด 8,064 คำ จะแบ่งออกเป็น 2 ส่วนคือ ส่วนที่นำไปใช้สร้างแบบอ้างอิง 4,032 คำ และส่วนที่นำไปใช้ทดสอบอีก 4,032 คำ จากนั้นนำเสียงพูดแต่ละคำไปคำนวณหาสเปกตรัม แอลพีซี โดยใช้แอลพีซี อันดับที่ 19 จะได้สเปกตรัมแอลพีซี ของแต่ละเฟรมใน 1 คำ แล้วนำค่าสเปกตรัมแอลพีซีที่ได้เป็นความถี่แปลงให้อยู่บนสเกลบาร์กของสระเสียงเดี่ยวและสระเสียงผสม ดังรูปที่ 4.18 , 4.19 , 4.20 และ 4.21 ตามลำดับ




สเปกตรัม แอลพีซี ของสระเสียง "เอะ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number
สเปกตรัม แอลพีซี ของสระเสียง "แอะ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number

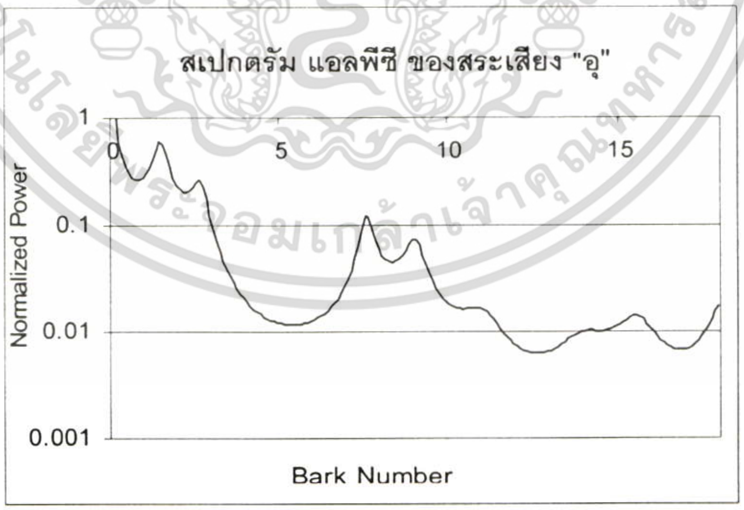
(a) สระหน้า

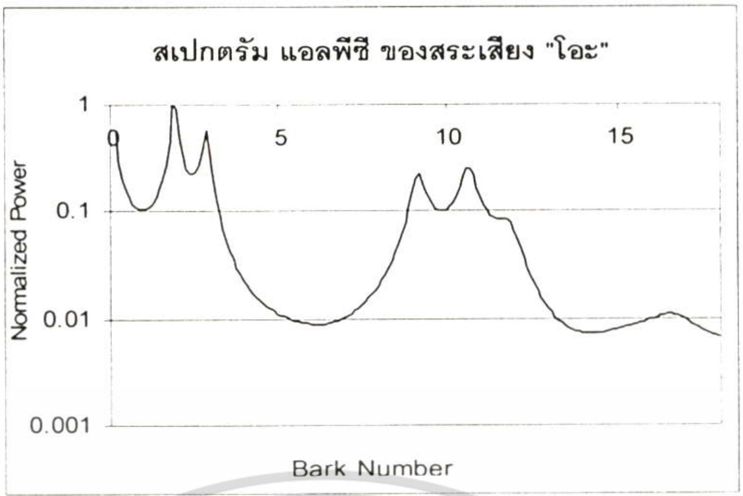

สเปกตรัม แอลพีซี ของสระเสียง "อึ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number


สเปกตรัม แอลพีซี ของสระเสียง "เออะ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number


สเปกตรัม แอลพีซี ของสระเสียง "อะ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number

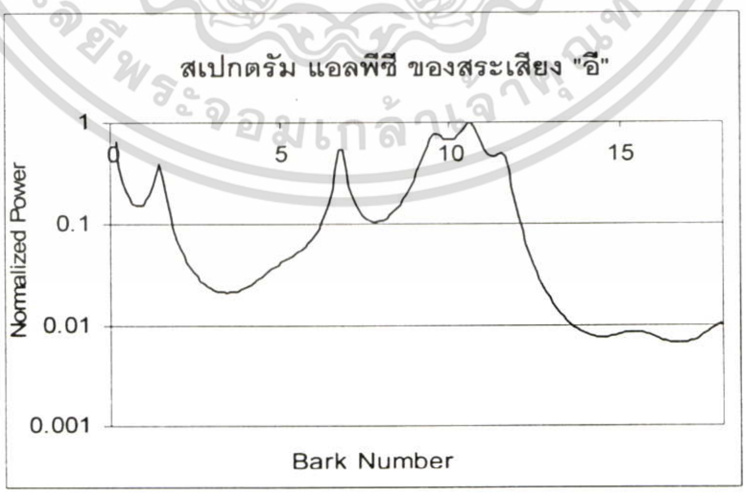
(b) สระกลาง

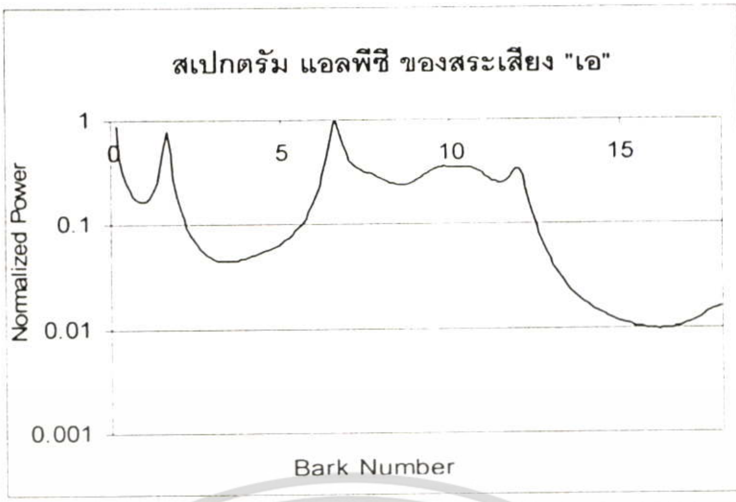

สเปกตรัม แอลพีซี ของสระเสียง "อุ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number

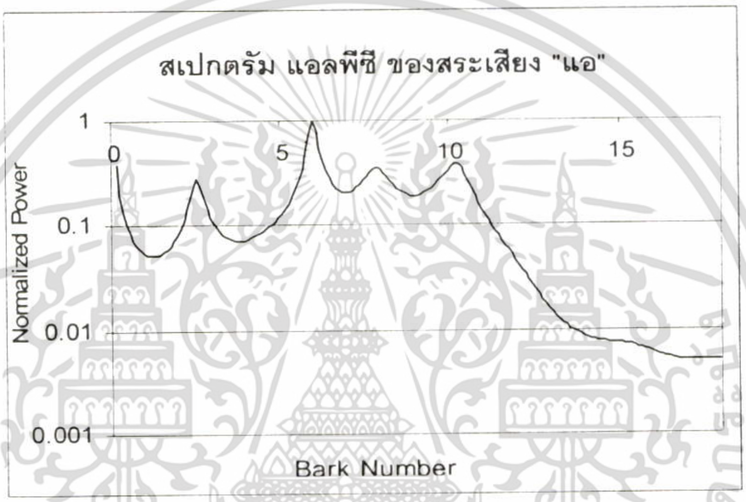


(c) สระหลัง

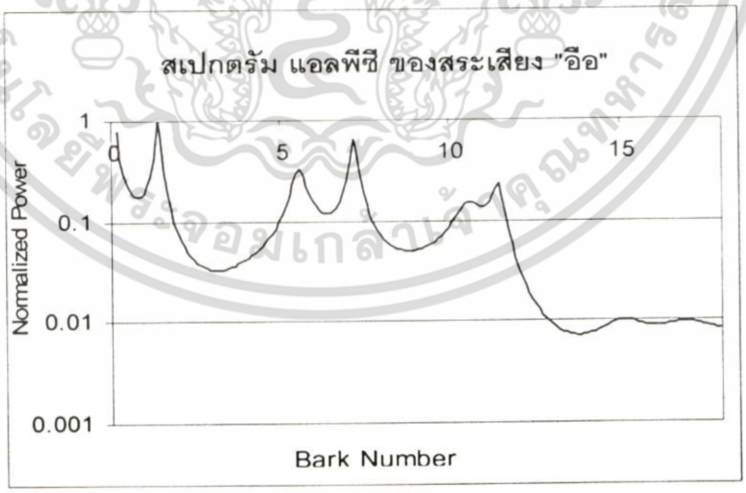
รูปที่ 4.18 สเปกตรัมแอลพีซีของเสียงสระเดี่ยวเสียงสั้นทั้ง 9 เสียง

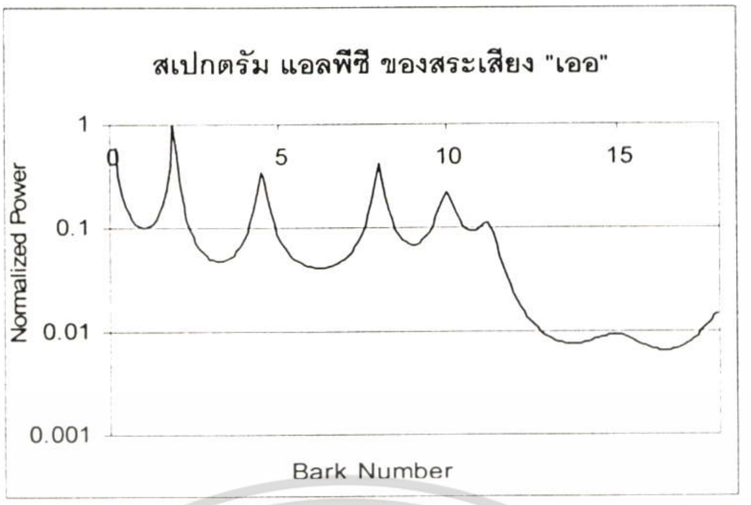



สเปกตรัม แอลพีซี ของสระเสียง "เอ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number


สเปกตรัม แอลพีซี ของสระเสียง "แอ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number

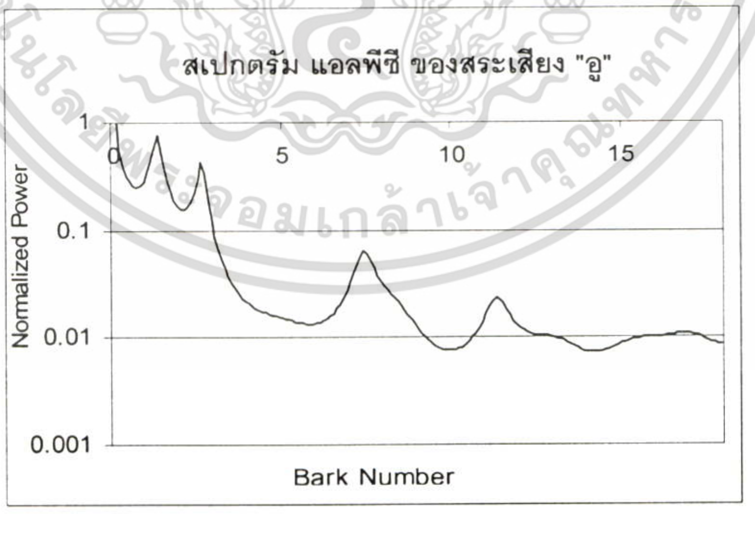
(a) สระหน้า

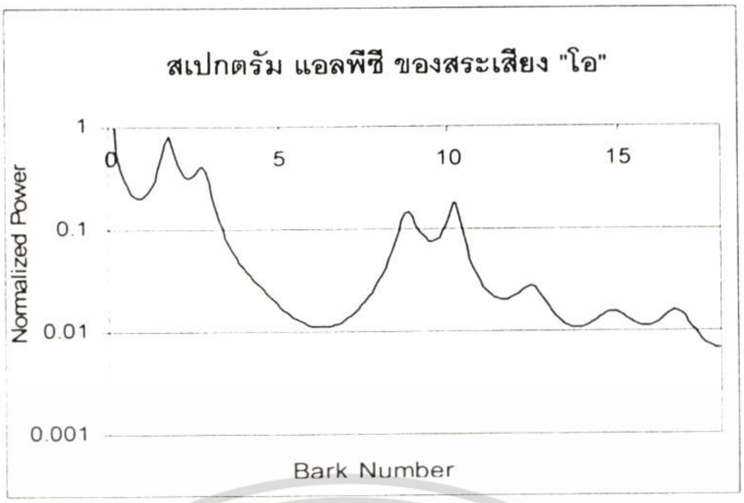

สเปกตรัม แอลพีซี ของสระเสียง "อือ"
Normalized Power
1
0.1
0.01
0.001
0
5
10
15
Bark Number

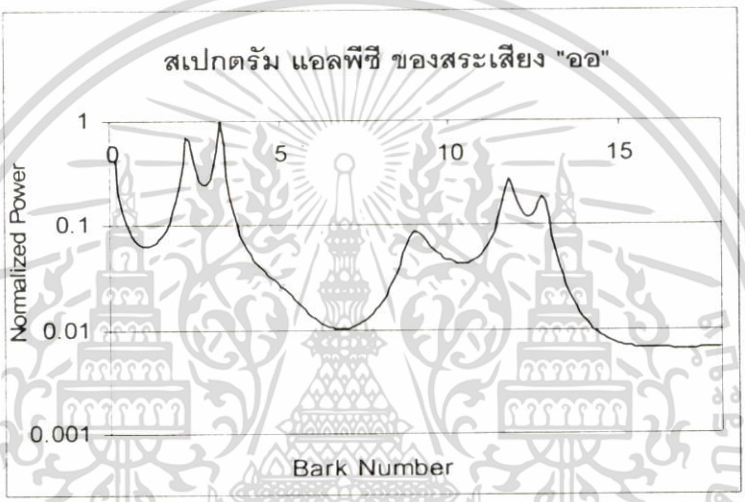

สเปกตรัม แอลพีซี ของสระเสียง "เออ"
1
0.1
0.01
0.001
0
5
10
15
Normalized Power
Bark Number


สเปกตรัม แอลพีซี ของสระเสียง "อา"
1
0.1
0.01
0.001
0
5
10
15
Normalized Power
Bark Number

(b) สระกลาง

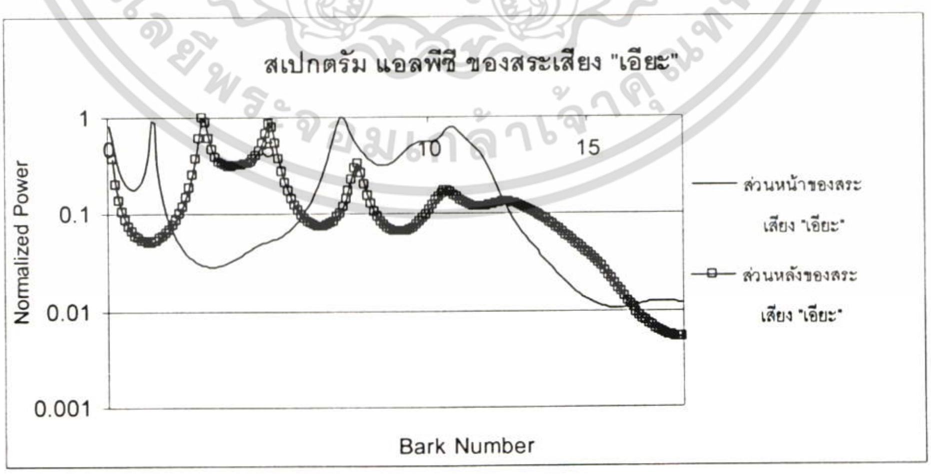

สเปกตรัม แอลพีซี ของสระเสียง "อู"
1
0.1
0.01
0.001
0
5
10
15
Normalized Power
Bark Number



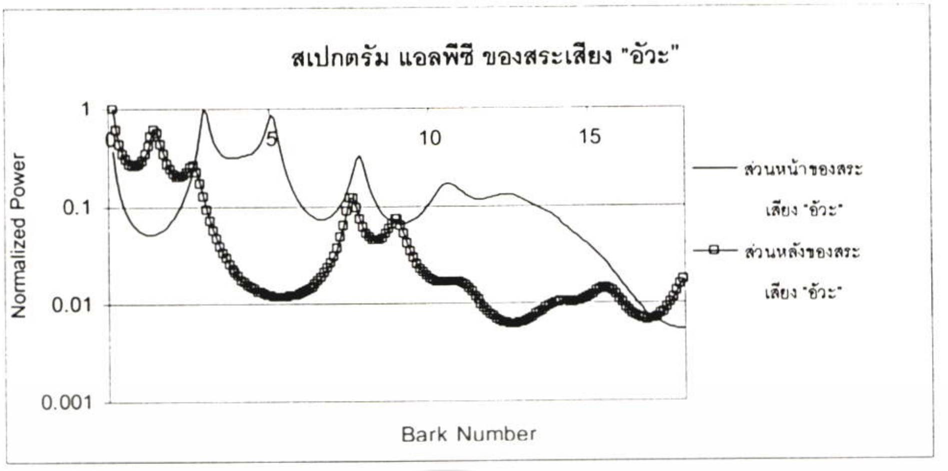


(c) สระหลัง

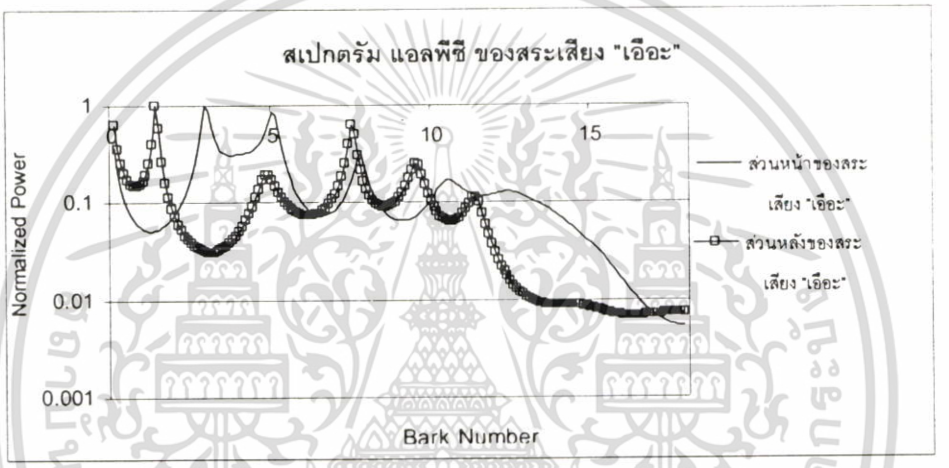
รูปที่ 4.19 สเปกตรัมแอลพีซีของเสียงสระเดี่ยวเสียงยาวทั้ง 9 เสียง



(a) สระ “เอียะ” (“อิ” + “อะ”)

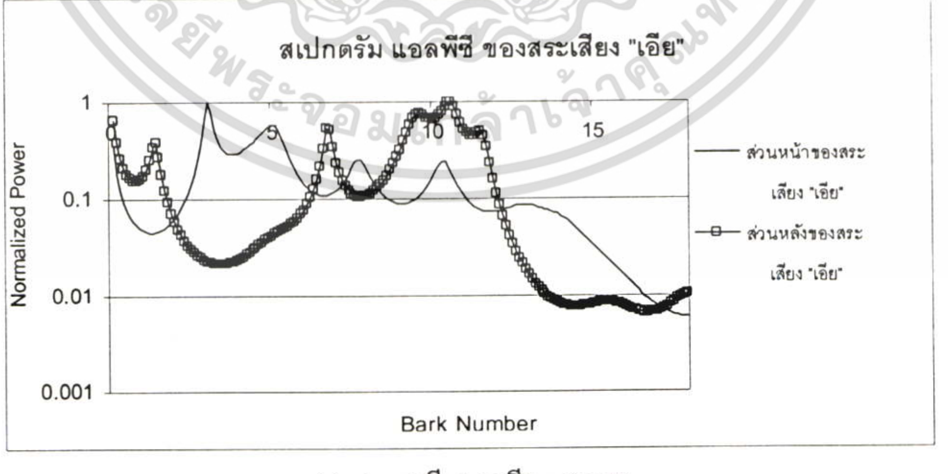


(b) สระ “อัวะ” (“อุ” + “อะ”)

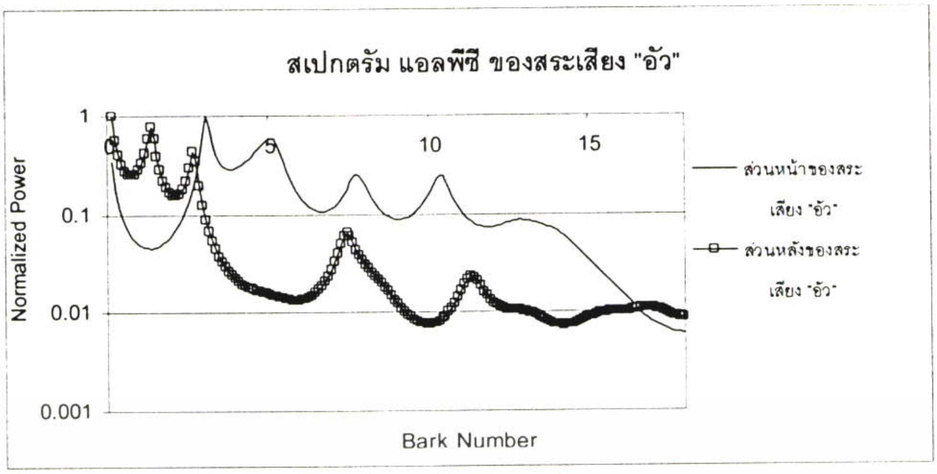


(c) สระ “เอือะ” (“อึ” + “อะ”)

รูปที่ 4.20 สเปกตรัมแอลพีซีของเสียงสระผสมเสียงสั้น ทั้ง 3 เสียง



(a) สระ “เอีย” (“อี” + “อา”)



(b) สระ “อัว” (“อู” + “อา”)

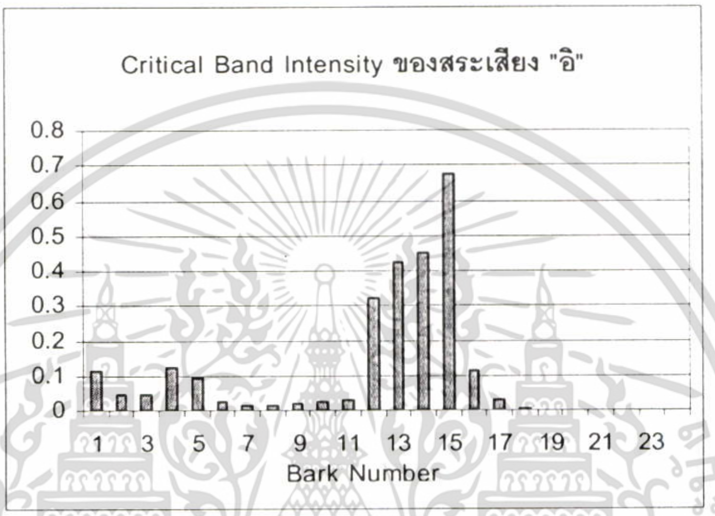
(c) สระ “เอือ” (“อือ” + “อา”)

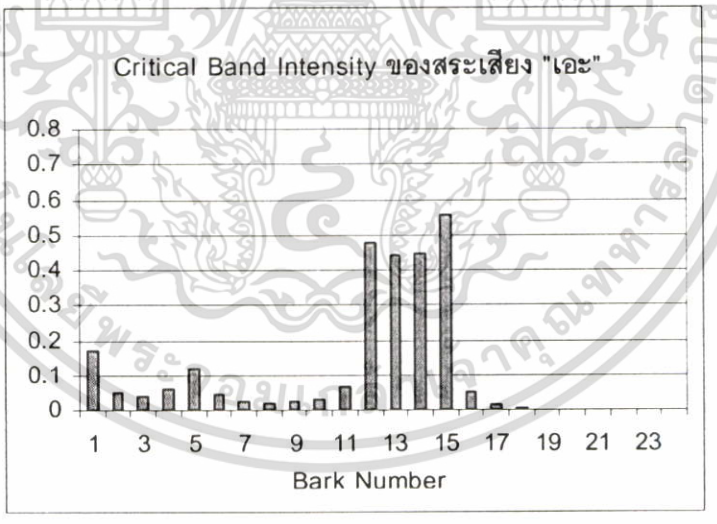
รูปที่ 4.21 สเปกตรัมแอลพีซีของเสียงสระผสมเสียงยาว ทั้ง 3 เสียง

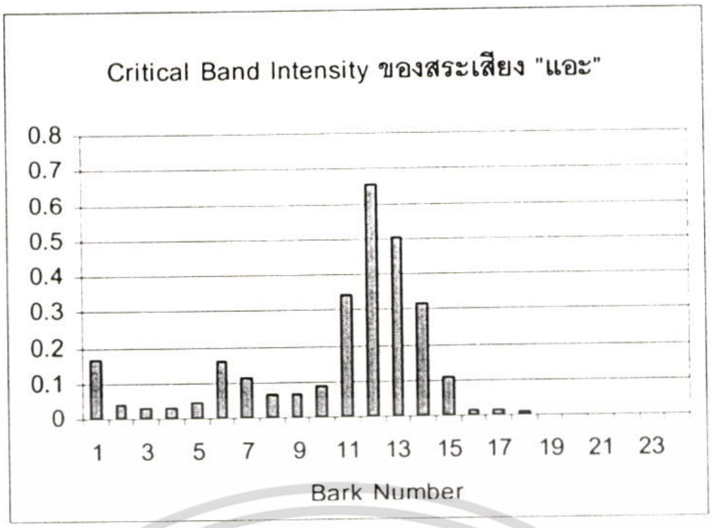
จากรูปที่ 4.18 และ 4.19 จะแสดงค่าสเปกตรัมแอลพีซีของสระเดี่ยวเสียงสั้น 9 เสียง และสระเดี่ยวเสียงยาว 9 เสียง นั้นจะแบ่งออกเป็น 3 กลุ่มด้วยกัน คือ กลุ่มของสระหน้าเสียงสั้นและเสียงยาวจะประกอบด้วยเสียงของสระ “อิ เอะ แอะ” และ “อี เอ แอ” ดังรูปที่ 4.18(a) และ 4.19(a) กลุ่มของสระกลางเสียงสั้นและเสียงยาวประกอบด้วยเสียงสระ “อึ เออะ อะ”และ “อือ เออ อา” ดังรูปที่ 4.18(b) และ 4.19(b) และกลุ่มของสระหลังเสียงสั้นและเสียงยาวประกอบด้วยเสียงสระ “อุ โอะ เอาะ” และ “อู โอ ออ” ดังรูปที่ 4.18(c) และ 4.19(c) ตามลำดับ เมื่อเปรียบเทียบกันระหว่างเสียงสระในกลุ่มเดียวกันและเสียงสระต่างกลุ่ม จะเห็นว่าเสียงสระในกลุ่มเดียวกันสเปกตรัมแอลพีซีจะมีลักษณะแตกต่างกันน้อยกว่าเสียงสระต่างกลุ่ม

จากรูปที่ 4.20 และ 4.21 จะแสดงค่าสเปกตรัมแอลพีซีของสระผสมเสียงสั้น 3 เสียง และสระผสมเสียงยาว 3 เสียง ซึ่งจะเห็นว่าสระเสียงผสมนั้นเกิดจากสระเสียงเดี่ยว 2 เสียงมารวมกัน ดังรูปที่ 4.20(a) และ 4.21(a) แสดงค่าสเปกตรัมแอลพีซีของสระเสียงผสมสั้นและยาว “เอียะ” และ

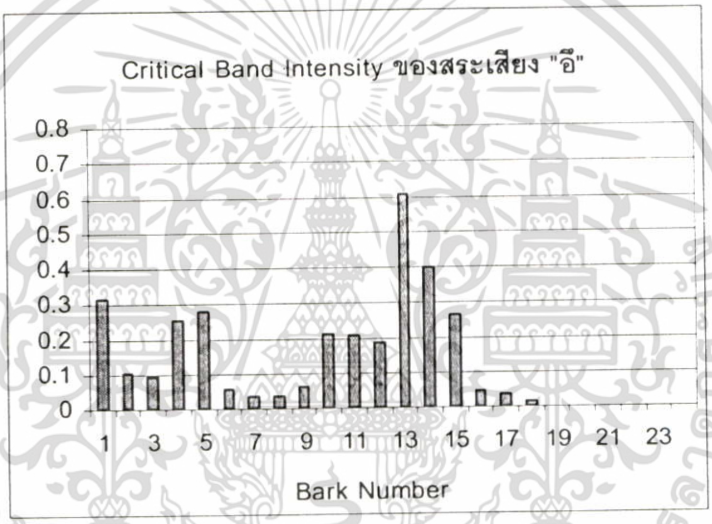
"เอีย" ซึ่งเกิดจากการรวมกันระหว่างสระเสียงเดี่ยว "อิ+อะ" และ "อี+อา" และสระเสียงผสมสั้นและยาวของสระ "อัวะ" และ "อัว" ที่เกิดจากการรวมกันระหว่างสระเสียงเดี่ยว "อุ+อะ" และ "อู+อา" ดังรูปที่ 4.20(b) และ 4.21(b) และสระเสียงผสมสั้นและยาวของสระ "เอือะ" และ "เอือ" ที่เกิดจากการรวมกันระหว่างสระเสียงเดี่ยว "อึ+อะ" และ "อือ+อา" ดังรูปที่ 4.20(c) และ 4.21(c) จากนั้นนำค่าสเปกตรัมแอลพีซีที่ได้ไปคำนวณหาค่าความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตได้ดังรูปที่ 4.22 , 4.23, 4.24 และ 4.25 ตามลำดับ

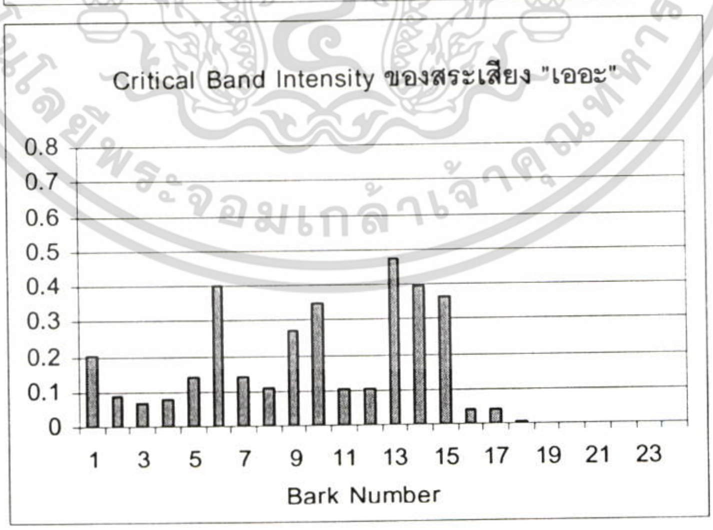


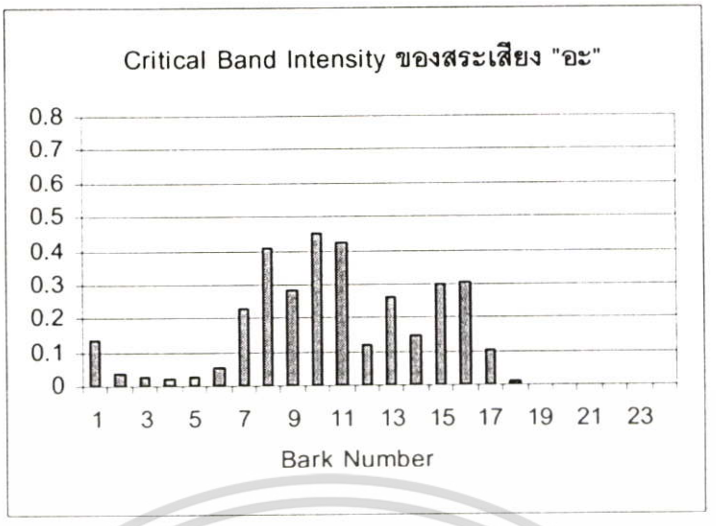



Critical Band Intensity ของสระเสียง "แอะ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

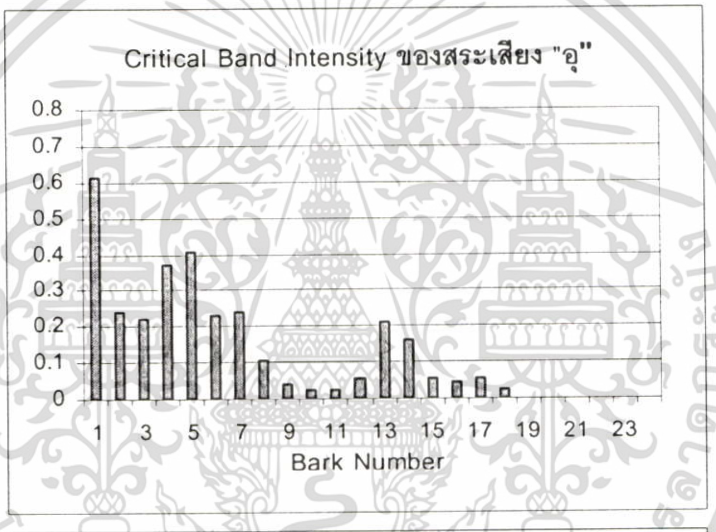
(a) สระหน้า

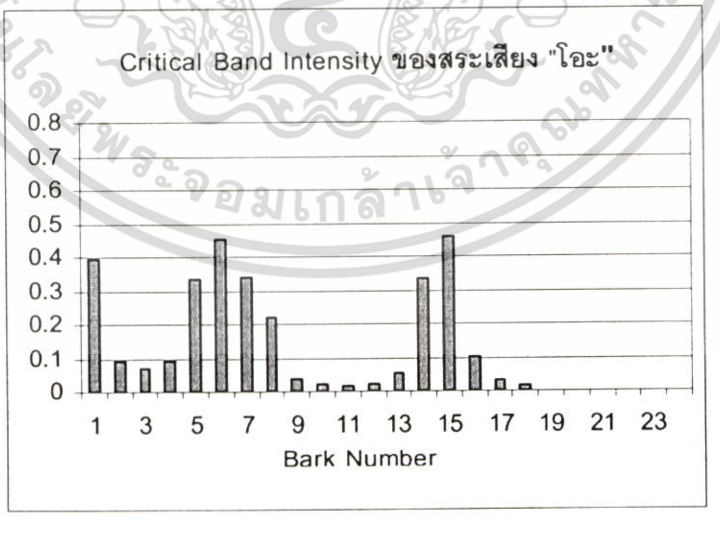

Critical Band Intensity ของสระเสียง "อึ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

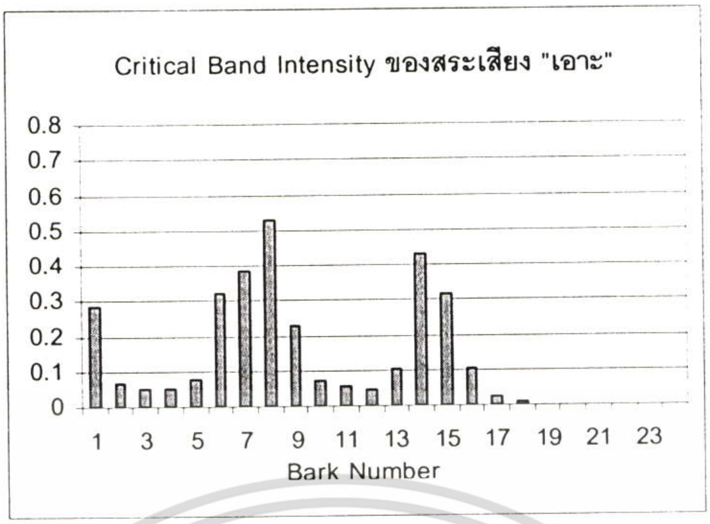

Critical Band Intensity ของสระเสียง "เออะ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number


Critical Band Intensity ของสระเสียง "อะ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

(b) สระกลาง

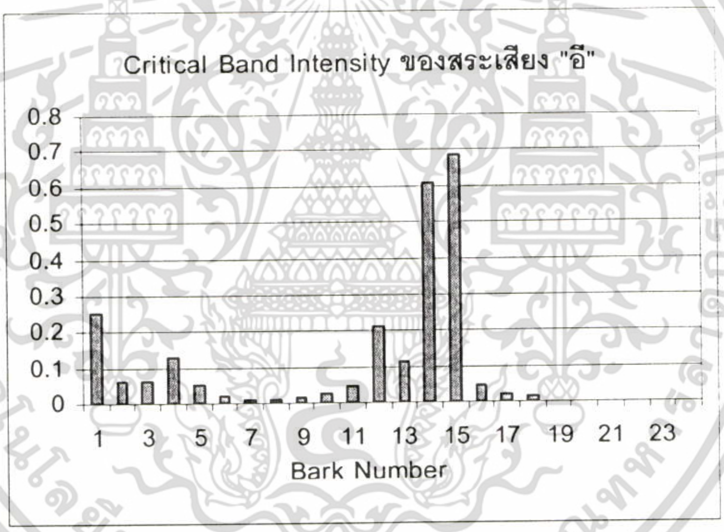

Critical Band Intensity ของสระเสียง "อุ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

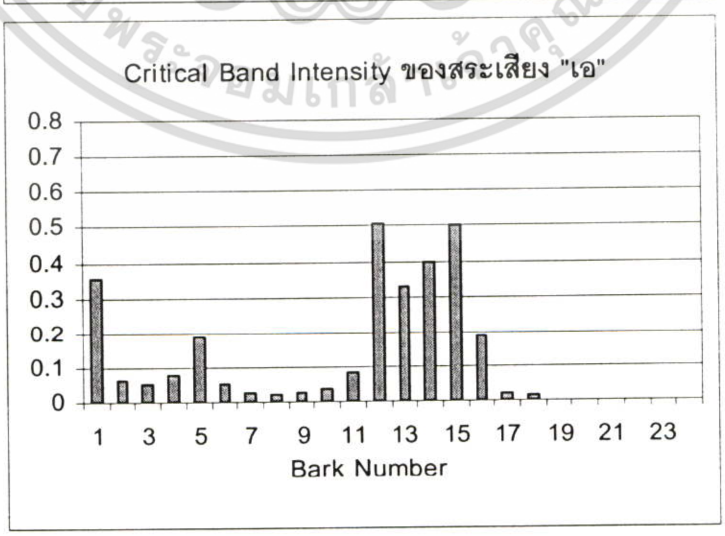

Critical Band Intensity ของสระเสียง "โอะ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

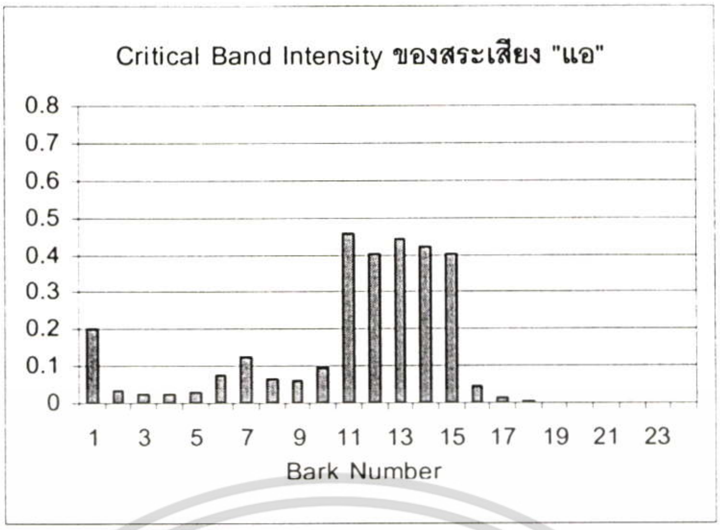


(c) สระหลัง

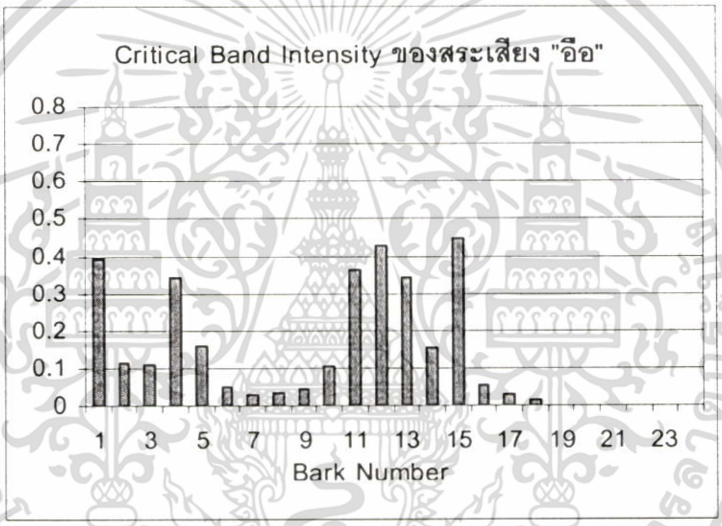
รูปที่ 4.22 ความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระเดี่ยวเสียงสั้น 9 เสียง

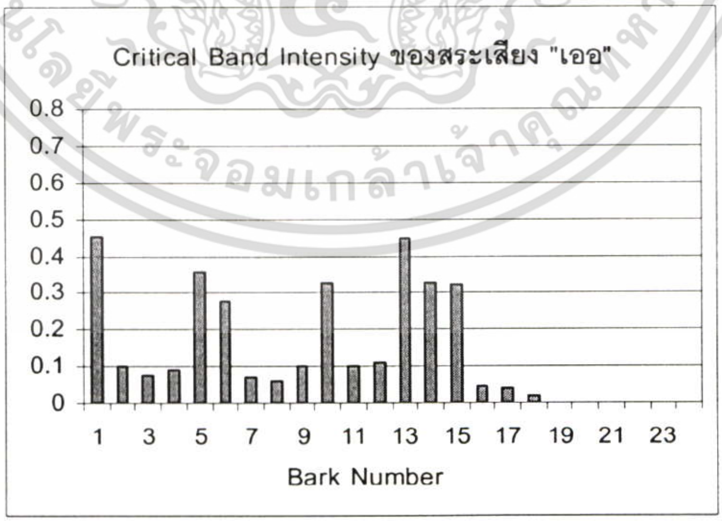


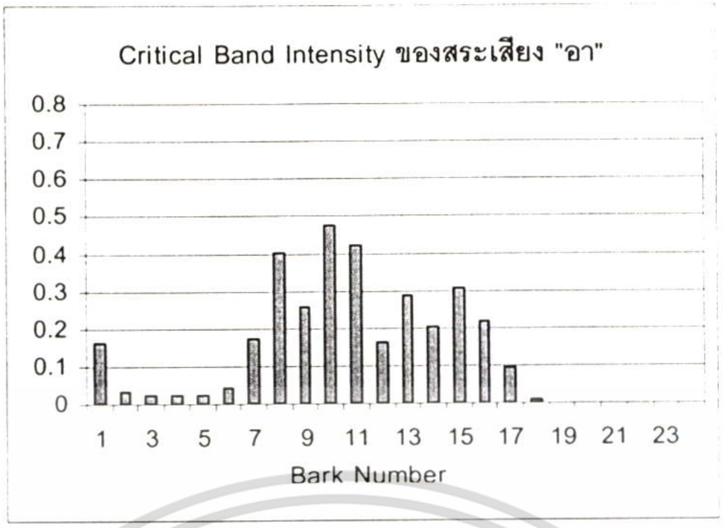



Critical Band Intensity ของสระเสียง "แอ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

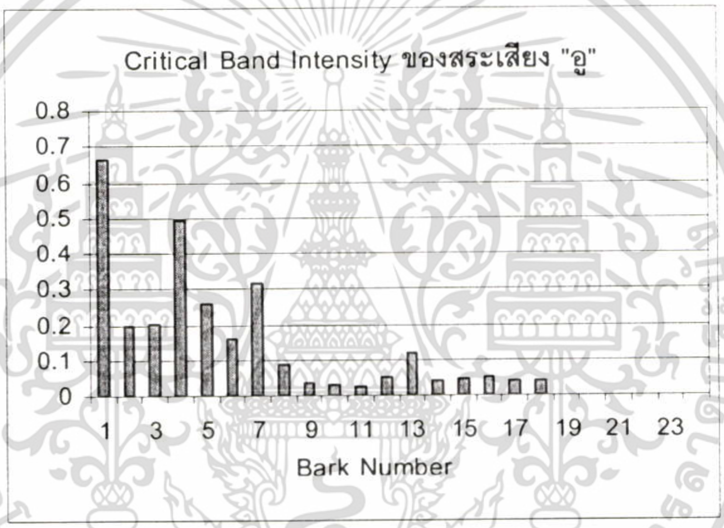
(a) สระหน้า

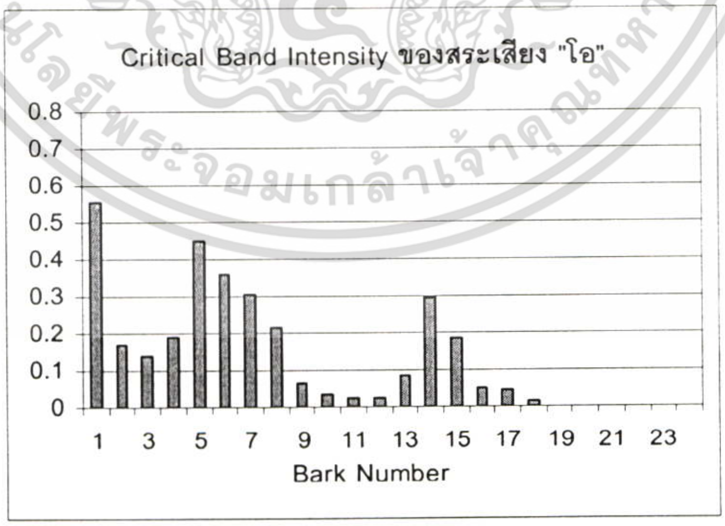

Critical Band Intensity ของสระเสียง "อือ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

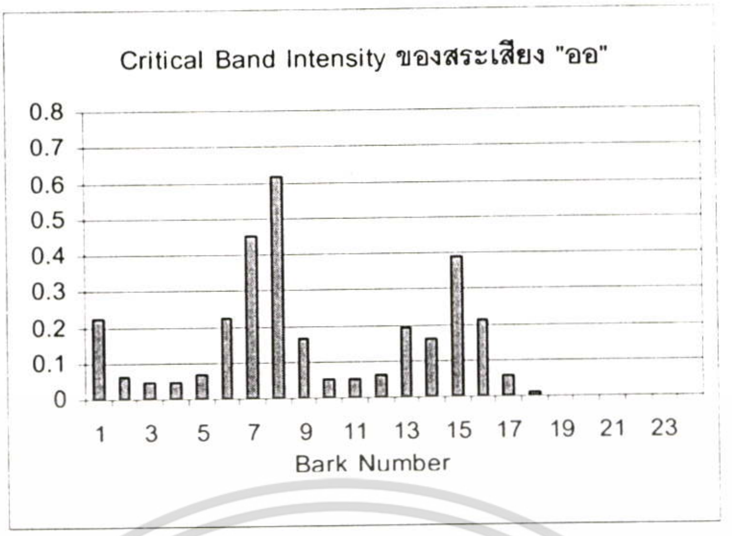

Critical Band Intensity ของสระเสียง "เออ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number


Critical Band Intensity ของสระเสียง "อา"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

(b) สระกลาง

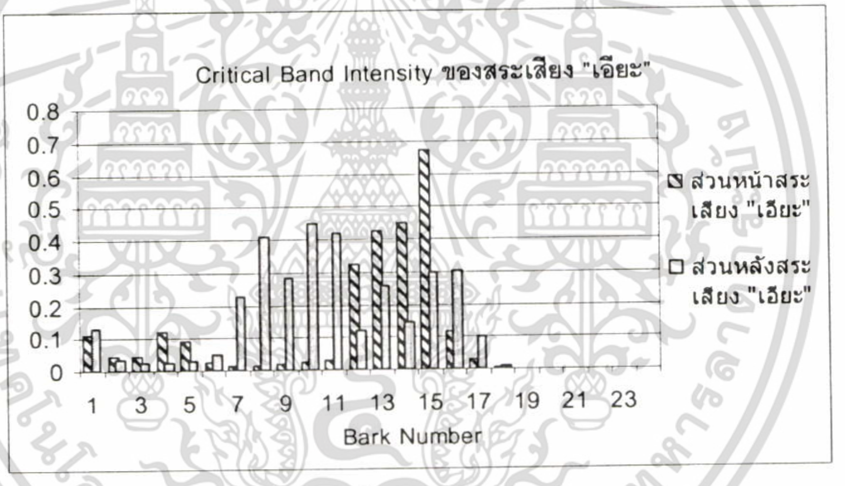

Critical Band Intensity ของสระเสียง "อู"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number


Critical Band Intensity ของสระเสียง "โอ"
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23
Bark Number

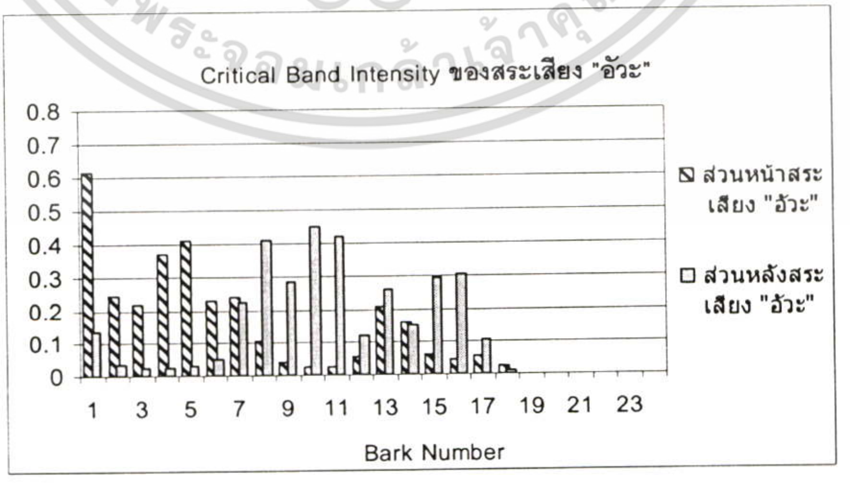


(c) สระหลัง

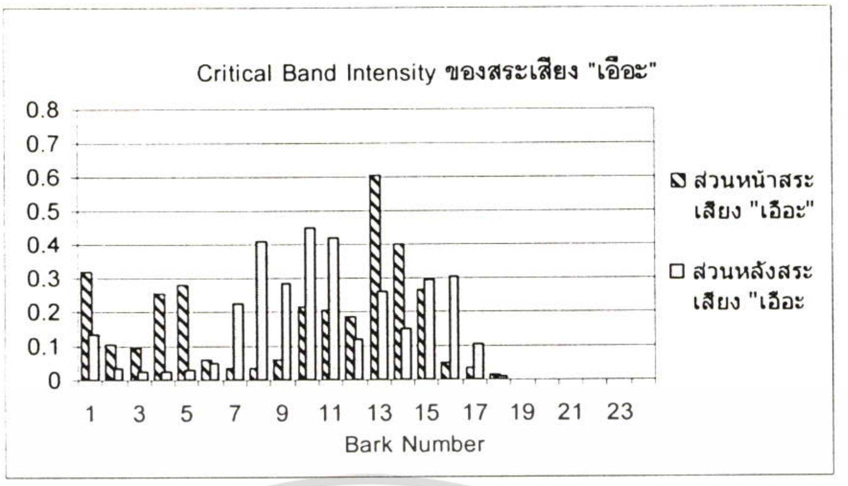
รูปที่ 4.23 ความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระเดี่ยวเสียงยาว 9 เสียง



(a) สระ "เอียะ" (" อิ " + " อะ ")

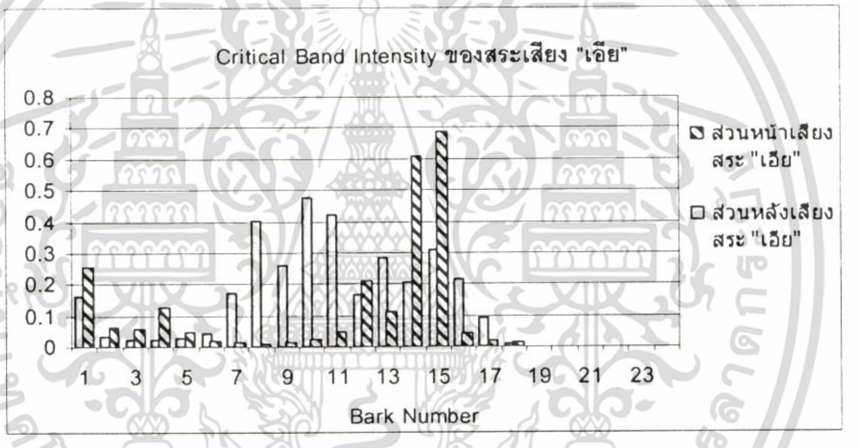


(b) สระ "อัวะ" (" อุ " + " อะ ")

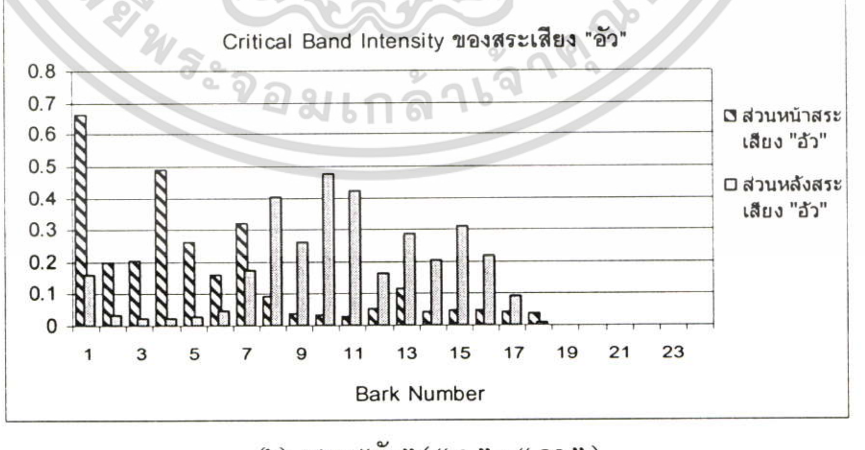


(c) สระ "เอือะ" (" อึ " + " อะ ")

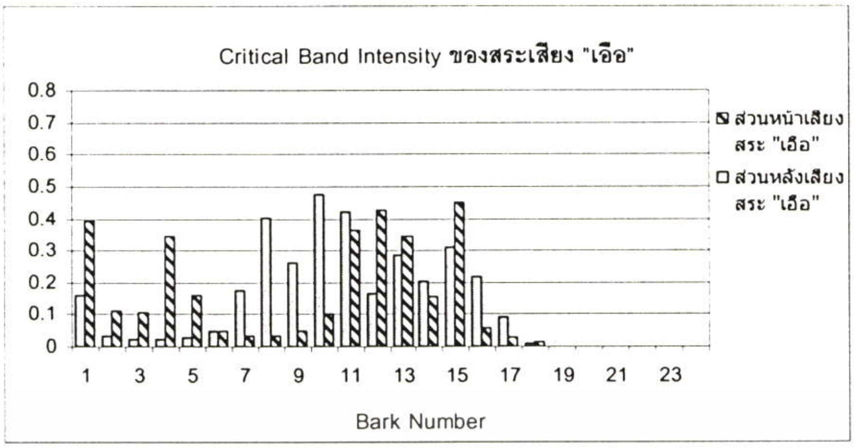
รูปที่ 4.24 ความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระผสมเสียงสั้น 3 เสียง



(a) สระ "เอีย" (" อี " + " อา ")



(b) สระ "อัว" (" อู " + " อา ")



(c) สระ "เอือ" (" อือ " + " อา ")

รูปที่ 4.25 ความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระผสมเสียงยาว 3 เสียง

ค่าความเข้มสัญญาณในแต่ละแถบความถี่วิกฤต เป็นองค์ประกอบของเวกเตอร์ที่เราใช้เป็นตัวแทนของแต่ละเฟรมของสัญญาณเสียงสระในการสร้างแบบอ้างอิงและการรู้จำจากรูปที่4.22 และ 4.23 แสดงความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระเดี่ยวเสียงสั้นและเสียงยาว จะเห็นได้ว่ามีความแตกต่างกันระหว่างเสียงสระ ซึ่งจะทำให้สามารถแยกแยะเสียงสระทั้งหมดออกจากกันได้ และจากรูป 4.24 และ 4.25 แสดงความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระผสมเสียงสั้นและเสียงยาว ซึ่งมีความแตกต่างกันเนื่องจากความเข้มสัญญาณในแต่ละแถบความถี่วิกฤตของเสียงสระผสมที่เกิดจากสระเสียงเดี่ยว 2 เสียง ทำให้สามารถนำลักษณะความแตกต่างของสระเดี่ยวและสระผสม มาใช้ในการแยกแยะเสียงสระทั้งหมดออกจากกัน จากนั้นนำเวกเตอร์แทนหน่วยเสียงของแต่ละเฟรมในแต่ละคำที่คำนวณได้ของสระเดี่ยวเสียงสั้น 9 เสียงและสระเดี่ยวเสียงยาว 9 เสียงมาจัดกลุ่มโดยใช้เสียงพูดของผู้พูด 8 คน (ผู้ชาย 3 คน และ ผู้หญิง 3 คน) ที่พูดเสียงสระเสียงเดี่ยวจำนวน 2,268 คำและใช้เสียงพูดของผู้พูด 2 คน(ผู้ชาย 1 คน และ ผู้หญิง 1 คน) ที่พูดเสียงสระเสียงเดี่ยวจำนวน 756 คำ มาสร้างโมเดลเสียงได้ทั้งหมด 18 โมเดลเสียงแบ่งเป็น สระหน้าเสียงสั้นประกอบด้วย " อิ เอะ แอะ" สระกลางเสียงสั้น " อะ อึ เออะ" สระหลังเสียงสั้น "อุ โอะ เอาะ" และสระหน้าเสียงยาวประกอบด้วย " อี เอ แอ" สระกลางเสียงยาว " อา อือ เออ" สระหลังเสียงยาว "อู โอ ออ" ตามลำดับ จากนั้นนำเสียงพูดสระเสียงเดี่ยวในส่วนที่เหลือจากการสร้างโมเดลมาทดสอบ

สำหรับการทดสอบแบบอ้างอิงนั้นจะใช้การค้นหา K-Nearest Neighbor ค่า k ที่ใช้หากมีค่าน้อยเกินไปอาจจะทำให้ผลการตัดสินใจของ K-Nearest Neighbor ผิดพลาดได้มากกว่าการใช้ค่า k ที่เหมาะสม ในงานวิจัยนี้จะใช้ค่า k=9 สำหรับการตัดสินใจโดยวิธี K-Nearest Neighbor [10]

เมื่อได้ค่าพารามิเตอร์ต่างๆ ที่ใช้ในการรู้จำเสียงสระเสียงเดี่ยวและสระเสียงผสมแล้ว จึงนำค่าที่ได้มาทำการสร้างแบบอ้างอิงและแบบทดสอบ การทดลองจะทดลองแบบอ้างอิงสำหรับผู้พูด 6

คน(Gender Dependent Reference model) จะใช้เสียงตัวอย่างจากผู้พูด 6 คน เพศชาย 3 คน เพศหญิง 3 คน และสำหรับผู้พูด 2 คน (Speaker Independent Reference model) จะใช้เสียงตัวอย่างจากผู้พูด 2 คน เพศชาย 1 คน เพศหญิง 1 คน ซึ่งจะใช้เสียงทั้งหมดจำนวน 8,064 เสียง โดยจะนำเสียงไปใช้ในการสร้างแบบอ้างอิงจำนวน 4,032 เสียง และนำเสียงไปใช้ในการทดสอบจำนวน 4,032 เสียง

4.7 ขั้นตอนการในทดสอบการรู้จำสระเสียงเดี่ยว แบบเฟรมโดยเทคนิค KNN

การทดสอบสระเสียงเดี่ยว จะนำแบบทดสอบที่ได้จากการหาค่าความเข้มสัญญาณในแต่ละแถบความถี่วิกฤต ที่ได้จากการค่าพารามิเตอร์ต่างๆดังที่ได้กล่าวมาแล้วในข้างต้น มาทดสอบกับแบบอ้างอิงที่ได้สร้างขึ้น โดยการทดสอบสระเสียงเดี่ยวนั้นจะเป็นการทดสอบการรู้จำแบบเฟรม และหาค่าความถูกต้องของสระทั้ง 24 เสียง จากจำนวนเฟรมในแต่ละเสียง เพื่อวิเคราะห์ออกมาเป็นคำ ในการทดสอบจะใช้เสียงของผู้พูดเพศชาย 3 คน และผู้พูดเพศหญิง 3 คน ในการทดสอบกับแบบอ้างอิงแบบขึ้นกับเพศ และจะใช้เสียงของผู้พูดเพศชาย 1 คน และผู้พูดเพศหญิง 1 คน ในการทดสอบอ้างอิงแบบไม่ขึ้นกับเพศ

โดยนำสระเดี่ยวเสียงสั้นและสระเดี่ยวเสียงยาว ทั้ง 18 เสียง (อะ อิ อึ อุ เอะ แอะ เอาะ เออะ โอะ อา อี อือ อู เอ แอ ออ เออ โอ) มาทำการทดสอบแบบอ้างอิงสระเสียงเดี่ยวเสียงสั้น 9 และสระเดี่ยวเสียงยาว 9 โมเดล โดยแบ่งการทดสอบออกเป็น 2 ขั้นตอน คือ

4.7.1 การทดสอบแบบอ้างอิงสระเสียงสั้น

4.7.2 การทดสอบแบบอ้างอิงสระเสียงยาว

โดยจะทำการทดสอบเฉพาะเสียงสระที่แยกแยะเสียงสั้นและเสียงยาวได้ถูกต้อง ซึ่งจะนำผลการทดสอบการแยกแยะสระเสียงสั้นและสระเสียงยาว จากตารางที่ 4.1 มาทำการทดสอบแบบอ้างอิงสระเสียงเดี่ยว ทั้ง 18 โมเดล

สำหรับการทดสอบแบบอ้างอิงของสระเดี่ยวเสียงสั้นและสระเดี่ยวเสียงยาวแบบเฟรม ของผู้พูดเพศชาย 3 คน จะให้ผลของการทดสอบค่าเปอร์เซ็นต์ความถูกต้อง แสดงในตารางที่ 4.2 และ 4.3 ตามลำดับ

ตารางที่ 4.2 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงสั้นแบบเฟรม สำหรับการทดสอบแบบอ้างอิง แบบขึ้นกับผู้พูดเพศชาย 3 คน

แบบอ้างอิง / แบบทดสอบ		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	
สระหน้า	อิ	**1374**	127	20	2	12	9	2	5	3	88.42
	เอะ	124	**1386**	63	10	25	22	2	5	3	84.51
	แอะ	11	40	**1550**	49	5	13	1	6		92.54
สระกลาง	อะ	15	11	30	**1372**	4	10	3	12	25	92.58
	อึ	46	17	5	8	**1156**	78	44	13	6	84.20
	เออะ	3	19	13	17	42	**1444**	7	33	55	88.43
สระหลัง	อุ	13	11		4	27	9	**1236**	72	16	89.05
	โอะ	4	3	1	27	12	24	35	**1179**	147	82.33
	เอาะ		1	2	24	9	38	11	99	**1390**	88.31
ค่าเฉลี่ย											**87.82**

ตารางที่ 4.3 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงยาวแบบเฟรม สำหรับการทดสอบแบบอ้างอิง แบบขึ้นกับผู้พูดเพศชาย 3 คน

แบบอ้างอิง / แบบทดสอบ		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	
สระหน้า	อี	**3867**	99	11		42	1	44			95.15
	เอ	61	**3696**	171	8	11	16	10	1	1	92.98
	แอ	25	149	**3594**	50	10	20	8	2		93.16
สระกลาง	อา	1	6	48	**2990**	3	18	13	6	2	96.86
	อือ	105	25	15	3	**3608**	101	88	8	4	91.18
	เออ	3	14	14	8	32	**3559**	7	38	33	95.98
สระหลัง	อู	23	11	4	1	17	3	**3870**	70		96.77
	โอ	14	5		10	12	36	125	**3547**	112	91.87
	ออ	3			46	2	25	5	187	**3230**	92.34
ค่าเฉลี่ย											**94.03**

ตารางที่ 4.4 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงสั้นแบบเฟรม สำหรับการทดสอบแบบอ้างอิง แบบขึ้นกับผู้พูดเพศหญิง 3 คน

แบบทดสอบ \ แบบอ้างอิง		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	
สระหน้า	อิ	**1318**	79	7	8	17	4	8	5		91.15
	เอะ	65	**1080**	191	7	3	26		2	7	78.20
	แอะ	29	29	**1310**	44	1	15	1	7	10	90.59
สระกลาง	อะ	3	7	42	**1362**	3	9	10	11	17	93.03
	อึ	38	6	4	12	**1258**	57	88	5	3	85.52
	เออะ		10	11	4	28	**1152**	1	26	32	91.14
สระหลัง	อุ	9		2	5	40	3	**1333**	85	12	89.52
	โอะ		4	4	19	6	79	25	**1124**	115	81.69
	เอาะ	2	6	5	35	2	40	3	117	**1278**	85.89
ค่าเฉลี่ย											**87.42**

ตารางที่ 4.5 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงยาวแบบเฟรม สำหรับการทดสอบแบบอ้างอิง แบบขึ้นกับผู้พูดเพศหญิง 3 คน

แบบทดสอบ \ แบบอ้างอิง		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	
สระหน้า	อี	**3009**	130	20	13	55	5	12	5	3	92.53
	เอ	63	**2662**	337	8	3	34	4	17	11	84.80
	แอ	7	231	**2755**	66	3	42		8	20	87.96
สระกลาง	อา	26	15	26	**3179**	9	20	11	9	16	96.01
	อือ	73	7	51	34	**3012**	56	131	13	2	89.14
	เออ	6	18	14	25	50	**2986**	4	34	72	93.05
สระหลัง	อู	8	3	1	6	95	6	**2891**	197		90.15
	โอ	5	17	1	9		85	83	**2734**	327	83.84
	ออ	2	7	2	17	2	25	5	61	**2852**	95.93
ค่าเฉลี่ย											**90.38**

ตารางที่ 4.6 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงสั้นแบบเฟรม สำหรับการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน

แบบอ้างอิง / แบบทดสอบ		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	
สระหน้า	อิ	**808**	213	37	2	12	9	1	5		74.33
	เอะ	63	**912**	131	2	12	17		1	2	80.00
	แอะ	6	43	**1021**	33	4	5	1	4	5	91.00
สระกลาง	อะ	7		53	**901**	2	6	5	3	7	91.57
	อึ	35	25	6	11	**714**	83	50	5	6	76.36
	เออะ	2	16	15	6	23	**1053**	3	29	21	90.15
สระหลัง	อุ	3	1	1	5	8	17	**771**	130	84	75.59
	โอะ	1	1	3	1	2	35	23	**943**	101	84.95
	เอาะ	1	1	4	25	3	18	1	33	**925**	91.49
ค่าเฉลี่ย											83.94

ตารางที่ 4.7 ผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงยาวแบบเฟรม สำหรับการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน

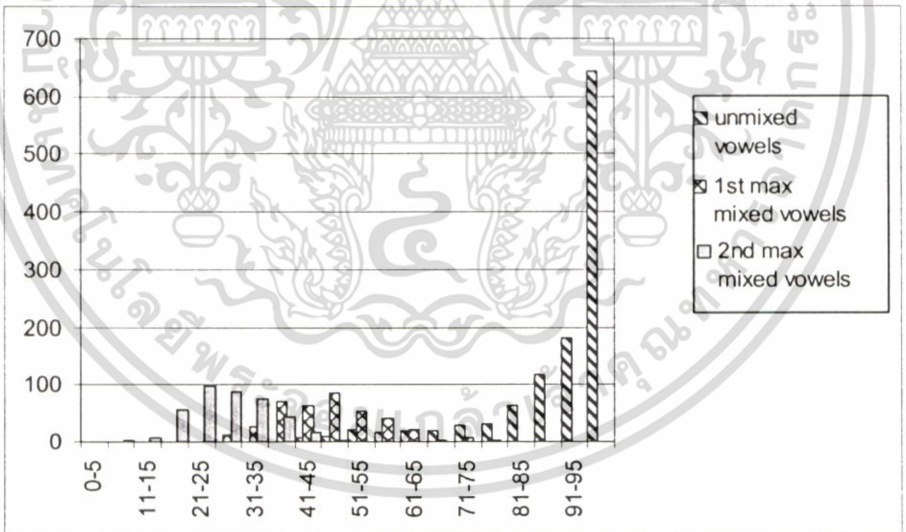
แบบอ้างอิง / แบบทดสอบ		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	
สระหน้า	อี	**2540**	35	3		31		1			97.32
	เอ	41	**2359**	54	6	8	4				95.43
	แอ	10	67	**2225**		7	13				95.82
สระกลาง	อา			10	**2250**				4		99.38
	อือ	36	9	2		**2563**	17	10			97.19
	เออ		8	5		19	**2204**		16		97.87
สระหลัง	อู	4				8		**2405**	34		98.12
	โอ		6		5		89	11	**2110**	54	92.75
	ออ		8		2		14			**1948**	98.78
ค่าเฉลี่ย											**96.96**

จากตารางที่ 4.2 และ 4.3 แสดงค่าเปอร์เซ็นต์ความถูกต้องแบบเฟรมของสระเดี่ยวเสียงสั้นและสระเดี่ยวเสียงยาวของผู้พูดเพศชาย 3 คน มีค่าประมาณ 88% และ 94% จากตารางที่ 4.4 และ 4.5 แสดงค่าเปอร์เซ็นต์ความถูกต้องแบบเฟรมของสระเดี่ยวเสียงสั้นและสระเดี่ยวเสียงยาวของผู้พูดเพศหญิง 3 คน มีค่าประมาณ 87% และ 90% ตามลำดับ และจากตารางที่ 4.6 และ 4.7 แสดงค่าเปอร์เซ็นต์ความถูกต้องแบบเฟรมของสระเดี่ยวเสียงสั้นและสระเดี่ยวเสียงยาวของผู้พูดเพศชาย 1 คน

ผู้พูดเพศหญิง 1 คน มีค่าประมาณ 84% และ 97% ตามลำดับ จะเห็นได้ว่าค่าความผิดพลาดที่เกิดขึ้นนั้นส่วนใหญ่จะเกิดขึ้นในกลุ่มสระเดียวกัน คือกลุ่มสระหน้า สระกลาง และสระหลัง เนื่องจากในค่าสเปกตรัมแอลพีซีละความเข้มสัญญาณในแต่ละแถบความถี่ในกลุ่มสระเดียวกันมีค่าใกล้เคียงกัน ส่วนค่าความผิดพลาดที่เกิดขึ้นกับสระที่อยู่ต่างกลุ่มกันจะมีค่าน้อยกว่าสระที่อยู่ในกลุ่มเดียวกัน เนื่องจากค่าสเปกตรัมแอลพีซี และค่าความเข้มของสัญญาณในแต่ละแถบความถี่ ในสระต่างกลุ่มมีค่าแตกต่างกันจึงทำให้ความผิดพลาดที่เกิดขึ้นมีค่าน้อย

4.8 การทดสอบการรู้จำสระเสียงเดี่ยว และ สระเสียงผสม แบบคำโดยใช้กฎการตัดสินใจอย่างง่าย

นำสระผสมเสียงสั้น และสระผสมเสียงยาว 6 เสียง (อัวะ เอียะ เอือะ อัว เอีย เอือ) ที่ผ่านการวิเคราะห์เสียงแล้วมาทดสอบกับโมเดลของเสียงสระเสียงเดี่ยวทั้ง 18 โมเดลเสียง ทำการทดสอบเช่นเดียวกับการทดสอบสระเสียงเดี่ยว แล้วใช้สถิติความถูกต้องจากเซตของแบบทดสอบในการหาค่าเปอร์เซ็นต์ของค่าที่มากที่สุดจากสระเสียงเดี่ยวและสระเสียงผสม แล้วนำมาพล็อตฮิสโตรแกรมดังแสดงในรูปที่ 4.26

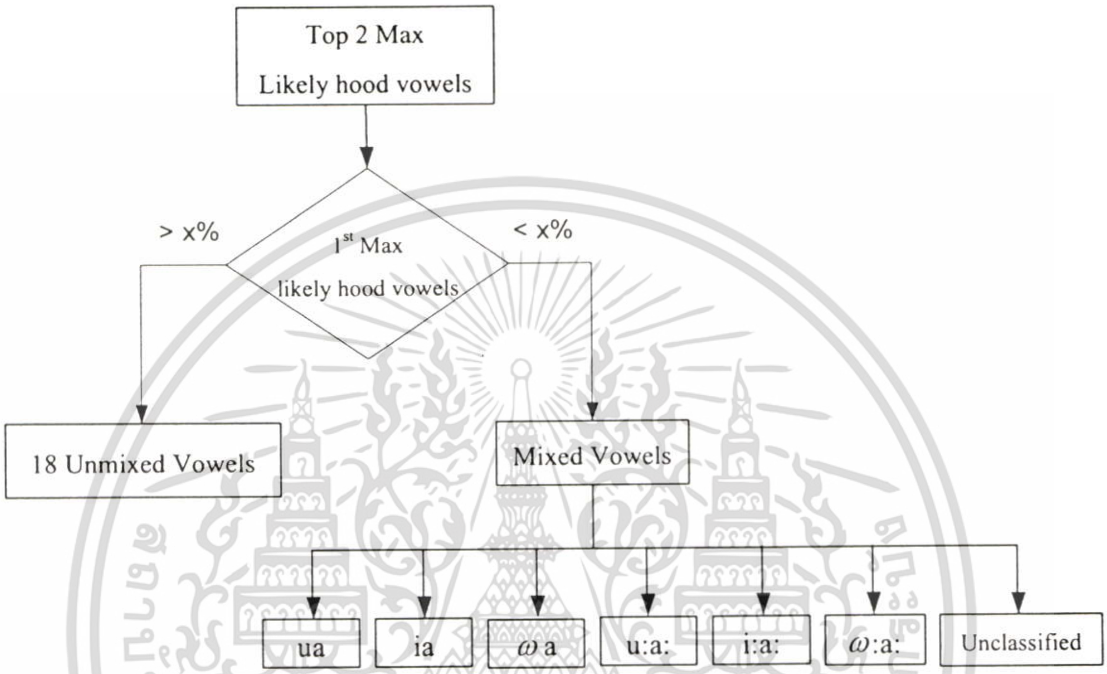


รูปที่ 4.26 แสดงสระเสียงเดี่ยว ที่มีค่ามากที่สุดเมื่อเทียบกับทุกสระ โดยที่ค่าเฉลี่ยเปอร์เซ็นต์ความถูกต้องมีค่ามากกว่า 70 % ใช้แยกสระเสียงเดี่ยวกับสระเสียงผสม ของผู้พูดหลายคน

จากรูปที่ 4.26 จะเห็นว่ากราฟประกอบด้วย 3 ส่วนคือ กราฟของสระเสียงเดี่ยว กราฟของสระเสียงผสมที่มีค่ามากที่สุด และสระเสียงผสมที่มีค่ารองลงมาเมื่อเทียบกับเสียงสระตัวอื่นๆ ลักษณะของกราฟสระเสียงเดี่ยว จะมีค่ามากอยู่ในช่วง 70-100 % ส่วนลักษณะของกราฟสระเสียงผสมจะประกอบ

ด้วย 2 ส่วน ส่วนแรกคือค่าที่มากที่สุดจะมีค่าอยู่ในช่วง 35-75 % ขึ้นไป ส่วนหลังคือค่าที่มีค่ามากที่สุดรองลงมาจะมีค่าอยู่ในช่วง 15-55%

ค่าที่ได้จากฮิสโตรแกรมสามารถนำไปแบ่งแยกระหว่างสระเสียงเดี่ยวและสระเสียงผสมได้ โดยเราสามารถแบ่งแยกสระเสียงเดี่ยวและสระเสียงผสมได้ที่ 70% ดังแสดงในรูปที่ 4.26



รูปที่ 4.27 แสดงกฎการตัดสินใจอย่างง่ายที่ใช้ในการวิเคราะห์ระหว่างสระเสียงเดี่ยวกับสระเสียงผสม

จากรูปที่ 4.27 แสดงวิธีการใช้กฎการตัดสินใจอย่างง่ายเพื่อนำไปวิเคราะห์แยกแยะสระเสียงเดี่ยวและสระเสียงผสมออกจากกัน โดยจะนำผลจากการรู้จำแบบเฟรม ของสระเสียงเดี่ยวและสระเสียงผสมทั้งหมดมาวิเคราะห์ หาค่าเปอร์เซ็นต์ความถูกต้องสูงสุดของจำนวนเฟรมในหนึ่งคำของการรู้จำในอันดับที่หนึ่ง และอันดับที่สอง ของแต่ละเสียงสระที่ทำการทดสอบแบบอ้างอิงมาแล้ว โดยจะถูกนำมาเปรียบเทียบกับค่ามาตรฐานในการแบ่งแยกสระเสียงเดี่ยวและเสียงผสม โดยเราจะเลือกค่ามาตรฐานจากเปอร์เซ็นต์ความถูกต้องในการรู้จำเสียงสระที่ 70% ในการแยกสระเสียงเดี่ยวและสระเสียงผสมออกจากกัน โดยเสียงสระที่มีค่าเปอร์เซ็นต์ความถูกต้องในการรู้จำแบบเฟรมที่มีค่าตั้งแต่ 70% ขึ้นไปเมื่อเทียบกับเสียงสระอื่นๆในกลุ่มสระเสียงเดี่ยว จะสามารถระบุได้ว่าเป็นสระเสียงเดี่ยวแบบคำที่ถูกต้อง ส่วนสระเสียงผสมจะใช้ค่าเปอร์เซ็นต์ความถูกต้องที่มากที่สุดจากผลการรู้จำแบบเฟรมอันดับหนึ่ง และค่าที่มากที่สุดอันดับสองมาเป็นตัวเลือก โดยที่ผลของการรู้จำแบบเฟรมในอันดับหนึ่งต้องมีค่าน้อยกว่า 70% ลงไป จึงจะเข้ากลุ่มในการวิเคราะห์สระเสียงผสม โดยจะแสดงวิธีการรู้จำสระเสียงผสมได้ดังตารางที่ 4.8 ดังนี้คือ

ตารางที่ 4. 8 วิธีการรู้จำสระเสียงผสมแบบคำ

	สระเสียงผสมแบบคำ	ผลการรู้จำสระเสียงผสมแบบเฟรม < 70%	
		ผลการรู้จำแบบเฟรมมากสุดอันดับหนึ่ง	ผลการรู้จำแบบเฟรมมากสุดอันดับสอง
สระผสมเสียงสั้น	อัวะ	อุ	อะ
		อะ	อุ
	เอียะ	อิ	อะ
		อะ	อิ
	เอือะ	อึ	อะ
		อะ	อึ
สระผสมเสียงยาว	อัว	อู	อา
		อา	อู
	เอีย	อี	อา
		อา	อี
	เอือ	อือ	อา
		อา	อือ

ส่วนผลของการรู้จำแบบเฟรมที่มีค่ามากที่สุดอันดับหนึ่งและค่ามากที่สุดอันดับสองที่เป็นเสียงสระอื่นนอกเหนือจาก 6 ตัวเลือกนี้ จะถือว่าคำนั้นไม่สามารถจำแนกเสียงสระได้

จากนั้นนำสระเสียงเดี่ยวและสระเสียงผสมที่แยกออกจากกันมาคำนวณหาค่าเปอร์เซ็นต์ความถูกต้องได้ดังตารางที่ 4.9

ตารางที่ 4.9 แสดงค่าเปอร์เซ็นต์ความถูกต้องของการแยกสระเสียงเดี่ยวและเสียงผสม

ผู้พูด	เสียงสระ	ค่าหลังจากแยกสระ	ค่าหลังจากแยกสระ		%ความถูกต้อง	%ค่าเฉลี่ย
			ค่าที่แยกถูกต้อง	ค่าที่แยกไม่ถูกต้อง		
ชาย 3 คน	เสียงเดี่ยว	1072	1033	39	96.36	90.24
	เสียงผสม	323	293	30	90.71	
หญิง 3 คน	เสียงเดี่ยว	1094	1028	66	93.97	
	เสียงผสม	321	284	37	88.47	
ชาย 1 หญิง 1	เสียงเดี่ยว	733	679	54	92.63	
	เสียงผสม	227	180	47	79.30	

จากตารางที่ 4.9 จะเห็นว่าค่าเปอร์เซ็นต์ความถูกต้องในการรู้จำเสียงสระที่ 70% ของค่าที่มากที่สุดอันดับหนึ่งเมื่อใช้เปรียบเทียบกับเฟรมเสียงสระอื่นๆ เพื่อแยกสระเสียงเดี่ยวและสระเสียงผสมออกจากกัน ให้ค่าความถูกต้องที่ 90.24% และมีบางส่วนที่แยกเสียงออกมายังไม่ถูกต้องมีค่าเท่ากับ 9.76% ของการระบุเสียงสระแบบคำ

นำผลการวิเคราะห์การแยกแยะสระเสียงเดี่ยว และสระเสียงผสม ที่ใช้ค่าเปอร์เซ็นต์ความถูกต้อง ในการรู้จำเสียงสระแบบเฟรมที่ 70% นำไปประบุเสียงสระในภาษาไทย ทั้ง 24 เสียงสระ ดังแสดงในตารางที่ 4.10 4.11 และ 4.12 สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศชาย 3 คน การทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศหญิง 3 คน และการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน ตามลำดับ

ตารางที่ 4.10 แสดงผลของการรู้จำเสียงสระ 24 เสียงในภาษาไทย แบบคำ สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศชาย 3 คน

เสียงสระ 24 เสียง		สระเสียงสั้น												สระเสียงยาว												ไม่สามารถจำแนกได้	% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	อัวะ	เอียะ	เอือะ	อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	อัว	เอีย	เอือ		
สระเสียงสั้น	อิ	55																								5	91.67
	เอะ	1	59																							2	95.16
	แอะ			61																						1	98.39
	อะ				61																						100
	อึ					49																				6	89.09
	เออะ						57																			2	96.61
	อุ							50																		2	96.15
	โอะ								50	2																2	92.59
	เอาะ									60																1	98.36
	อัวะ										55		2													4	90.16
	เอียะ											45	1													2	93.75
	เอือะ											1	42													2	93.33
สระเสียงยาว	อี													59												1	98.33
	เอ														60											3	95.24
	แอ															60										2	96.77
	อา																52										100
	อือ																	58								3	95.08
	เออ																		61							1	98.39
	อู																			62							100
	โอ																				61	1				1	96.83
	ออ																					58				3	95.08
	อัว																						47			3	94.00
	เอีย													2										52		4	89.66
	เอือ																	2						3	52	4	85.25
ค่าเฉลี่ย																											**95.00**

ตารางที่ 4.11 แสดงผลของการรู้จำเสียงสระ 24 เสียงในภาษาไทย แบบคำ สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศหญิง 3 คน

เสียงสระ 24 เสียง		สระเสียงสั้น												สระเสียงยาว												ไม่สามารถจำแนกได้	% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	อัวะ	เอียะ	เอือะ	อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	อัว	เอีย	เอือ		
สระเสียงสั้น	อิ	57	1									1														1	95.00
	เอะ	1	49	2																						5	85.96
	แอะ	1	1	55																						4	90.16
	อะ			1	61								1														96.83
	อึ					54																				4	93.10
	เออะ						50																			2	96.15
	อุ							58	1																	2	95.08
	โอะ							1	51	3																4	86.44
	เอาะ								1	57																4	91.94
	อัวะ										52															6	89.66
	เอียะ											46														6	88.46
	เอือะ											3	43													5	84.31
สระเสียงยาว	อี													61												1	98.39
	เอ														58	2										3	92.06
	แอ														2	59										1	95.16
	อา																61										100
	อือ																	61								2	96.83
	เออ																		62							1	98.41
	อู																	1		59	1					1	95
	โอ																				58	3				2	92.06
	ออ																					62					100.00
	อัว																1						43			4	89.58
	เอีย																							52		4	92.86
	เอือ																						1	1	48	6	85.71
ค่าเฉลี่ย																											**92.89**

ตารางที่ 4.12 แสดงผลของการรู้จำเสียงสระ 24 เสียงในภาษาไทย แบบคำ สำหรับการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน

เสียงสระ 24 เสียง		สระเสียงสั้น												สระเสียงยาว												ไม่สามารถจำแนกได้	% ความถูกต้อง
		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	อัวะ	เอียะ	เอือะ	อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	อัว	เอีย	เอือ		
สระเสียงสั้น	อิ	32	6																							2	80.00
	เอะ		35	2																						5	83.33
	แอะ			39																						2	95.12
	อะ				38						1	1															95.00
	อึ					29		1					1													6	78.38
	เออะ						41																			1	97.62
	อุ							28																		9	75.68
	โอะ								33	1																6	82.50
	เอาะ									39																2	95.12
	อัวะ				1						34															7	80.95
	เอียะ											28	2													7	75.68
	เอือะ											3	29													6	76.32
สระเสียงยาว	อี													41												1	97.62
	เอ														40											2	95.24
	แอ															40										1	97.56
	อา																40										100.00
	อือ																	41								1	97.62
	เออ																		42								100.00
	อู																			42							100.00
	โอ																				39					3	92.86
	ออ																					40					100.00
	อัว																1						29			3	87.88
	เอีย																							29	1	6	80.56
	เอือ																							2	31	8	75.61
ค่าเฉลี่ย																											**89.19**

จากตารางที่ 4.10 4.11 และ 4.12 จะแสดงค่าเปอร์เซ็นต์ความถูกต้องของเสียงสระ 24 เสียงในภาษาไทยแบบคำ สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูดเพศชาย 3 คน มีค่าประมาณ 95% ค่าเป็นเซ็นต์ความถูกต้องของเสียงสระ 24 เสียงในภาษาไทยแบบคำ สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูดเพศหญิง 3 คน มีค่าประมาณ 93% และแสดงค่าเปอร์เซ็นต์ความถูกต้องของเสียงสระ

24 เสียง ในภาษาไทยแบบคำ สำหรับการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน มีค่าประมาณ 89% ตามลำดับ ซึ่งความผิดพลาดของการรู้จำเสียงสระที่เกิดขึ้น เกิดจากการออกเสียงสระที่มีเสียงใกล้เคียงกับเสียงสระอื่นๆ ทำให้ผลการรู้จำเสียงนั้นผิดพลาดได้ ค่าความผิดพลาดของสระเสียงเดี่ยวที่เกิดขึ้นนั้น ส่วนมากจะเกิดจากความผิดพลาดของสระที่มีลักษณะเสียงใกล้เคียงกันมากกว่า ที่เกิดจากความผิดพลาดของเสียงสระที่อยู่ต่างกลุ่มสระ ส่วนค่าความผิดของสระเสียงผสมที่เกิดขึ้นนั้นส่วนมากจะเกิดจากการออกเสียงสระผสม ที่มีลักษณะเสียงใกล้เคียงกับเสียงสระผสมในกลุ่มเดียวกัน และเกิดจากการออกเสียงที่มีลักษณะเสียงใกล้เคียงกับสระเสียงเดี่ยว ทำให้มีค่าเปอร์เซ็นต์ความถูกต้องมากกว่า 70% และถูกระบุเป็นสระเสียงเดี่ยวนั่นเอง

บทที่ 5
สรุปผลและอภิปรายผล

5.1 สรุปผลการทดลอง

วิทยานิพนธ์นี้จะนำเสนอการรู้จำหน่วยเสียงสระ 24 เสียงในภาษาไทย โดยการใช้คุณลักษณะทางพลังงานของเสียงที่เปล่งออกมา โดยพิจารณาที่ลักษณะความแตกต่างการหยุดเสียงของสระเสียงสั้นและสระเสียงยาว แล้วนำมาหาค่าสัมประสิทธิ์การฟิตเคิร์ฟแบบพหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียง เพื่อนำไปสร้างเป็นแบบอ้างอิง และแบบทดสอบ ในการแยกแยะสระเสียงสั้น และสระสระเสียงยาวออกจากกัน ด้วยเทคนิค K-Nearest Neighbor จากนั้นนำสระที่แยกแยะเสียงสั้น และเสียงยาวได้แล้ว ไปวิเคราะห์หาค่าสเปกตรัมแอลพีซี แล้วแปลงให้อยู่บนสเกลบาร์ก จากนั้นนำไปคำนวณหาค่าความเข้มสัญญาณบนแถบความถี่วิกฤต เพื่อนำไปสร้างแบบอ้างอิงและแบบทดสอบ โดยใช้เทคนิค K-Nearest Neighbor ในการรู้จำหน่วยเสียงสระแบบเฟรม แล้วใช้กฎการตัดสินใจอย่างง่ายในการรู้จำเสียงสระ 24 เสียงแบบคำ ผลที่ได้จากการทำวิจัยสามารถสรุปได้ดังนี้

5.1.1 การแยกแยะสระเสียงสั้น และสระเสียงยาว

จากการทดสอบการแยกแยะสระเสียงสั้น และสระเสียงยาวในภาษาไทยโดยใช้สัมประสิทธิ์การฟิตเคิร์ฟแบบพหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียงแล้วใช้เทคนิค K-Nearest Neighbor ในการแยกแยะเสียงสระ ให้ค่าเปอร์เซ็นต์ความถูกต้องในการแยกแยะสระเสียงสั้น ดังนี้คือ

5.1.11 สำหรับการทดสอบแบบขึ้นกับผู้พูด เพศชาย 3 คน จะให้ผลของเปอร์เซ็นต์ความถูกต้องในการแยกแยะสระเสียงสั้นและสระเสียงยาว ประมาณ 92%

5.1.1.2 สำหรับการทดสอบแบบขึ้นกับผู้พูด เพศหญิง 3 คนจะให้ผลของเปอร์เซ็นต์ความถูกต้องในการแยกแยะสระเสียงสั้นและสระเสียงยาว ประมาณ 94%

5.1.1.3 สำหรับการทดสอบแบบไม่ขึ้นกับผู้พูด เพศชาย 1คน เพศหญิง 1 คนจะให้ผลของเปอร์เซ็นต์ความถูกต้องในการแยกแยะสระเสียงสั้นและสระเสียงยาว ประมาณ 95%

โดยการแยกแยะเสียงสระนั้น เลือกใช้ค่า k= 5 ในการทดสอบ

5.1.2 การรู้จำสระเสียงเดี่ยวแบบเฟรม

จากการทดสอบการรู้จำสระเสียงเดี่ยวแบบเฟรม ด้วยวิธี K-Nearest Neighbor ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบเฟรม ของสระเดี่ยวเสียงสั้น และสระเดี่ยวเสียงยาว ดังนี้คือ

5.1.2.1 สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศชาย 3 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบเฟรม ของสระเดี่ยวเสียงสั้น ประมาณ 88% และสระเดี่ยวเสียงยาว ประมาณ 94%

5.1.2.2 สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศหญิง 3 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบเฟรม ของสระเดี่ยวเสียงสั้น ประมาณ 87% และสระเดี่ยวเสียงยาว ประมาณ 90%

5.1.2.3 สำหรับการทดสอบแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบเฟรม ของสระเดี่ยวเสียงสั้น ประมาณ 84% และสระเดี่ยวเสียงยาวประมาณ 97%

โดยการรู้จำสระเสียงเดี่ยวแบบเฟรมนั้น เลือกใช้ค่า k= 9 ในการทดสอบ

5.1.3 การรู้สระเสียงเดี่ยว และสระเสียงผสมแบบคำ

จากการทดสอบ โดยการนำค่าที่ได้จากการรู้จำแบบเฟรมมาคำนวณหาค่าเปอร์เซ็นต์ความถูกต้อง ของเฟรมที่มีค่ามากที่สุดอันดับหนึ่ง และอันดับสองของทั้งผู้เพศพูดชาย และผู้พูดเพศหญิงมาทำการวิเคราะห์หาค่าที่เหมาะสมในการแยกแยะสระเสียงเดี่ยวและสระเสียงผสมออกจากกัน โดยวิธีการพล็อตฮิสโตรแกรม ได้ค่ามาตรฐานในการแยกแยะสระเสียงเดี่ยว และเสียงผสมที่ 70% ของค่าเปอร์เซ็นต์ความถูกต้องแบบเฟรม และเมื่อนำมาทดสอบการแยกแยะสระเสียงเดี่ยว และสระเสียงผสม ได้ผลความถูกต้องในการแยกแยะสระเสียงเดี่ยว และเสียงผสม แบบขึ้นกับผู้พูด เพศชาย 3 คน แบบขึ้นกับผู้พูด เพศหญิง 3 คน และแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน ประมาณ 90%

จากนั้น นำค่ามาตรฐานที่ได้ ไปทดสอบการรู้จำหน่วยเสียงสระทั้ง 24 เสียง แบบคำโดยเสียงสระที่มีค่าความถูกต้องในการรู้จำแบบเฟรมที่มีค่าตั้งแต่ 70% ขึ้นไปเมื่อเทียบกับเสียงสระอื่นๆในกลุ่มสระเสียงเดี่ยวแล้ว จะสามารถรู้จำได้ว่าเป็นสระเสียงเดี่ยวแบบคำที่ถูกต้อง ส่วนสระเสียงผสมจะใช้ค่าที่มากที่สุดจากผลการรู้จำแบบเฟรมอันดับหนึ่ง และอันดับสองมาเป็นตัวเลือก โดยผลของการรู้จำแบบเฟรมในอันดับหนึ่งต้องมีค่าน้อยกว่า 70% ลงไป ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบคำดังนี้คือ

5.1.3.1 สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศชาย 3 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบคำของหน่วยเสียงสระ 24 เสียง ประมาณ 95%

5.1.3.2 สำหรับการทดสอบแบบอ้างอิงแบบขึ้นกับผู้พูด เพศหญิง 3 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบคำของหน่วยเสียงสระ 24 เสียง ประมาณ 93%

5.1.3.3 สำหรับการทดสอบแบบอ้างอิงแบบไม่ขึ้นกับผู้พูด เพศชาย 1 คน เพศหญิง 1 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องการรู้จำแบบคำของหน่วยเสียงสระ 24 เสียง ประมาณ 89%

5.2 อภิปรายผลความผิดพลาด

จากผลการทดลองจะเห็นว่าค่าความผิดพลาดที่เกิดขึ้นในขั้นตอนการแยกแยะสระเสียงสั้นและสระเสียงยาว ส่วนใหญ่จะเกิดจากกลุ่มของสระเสียงยาวที่ผิดมาอยู่ในกลุ่มสระเสียงสั้น เนื่องจากการออกเสียงสระเสียงยาวในแต่ละเสียงจะให้ค่าพลังงานที่แตกต่างกันมากกว่าสระเสียงสั้น จึงทำให้เกิดความผิดพลาดในการออกเสียงได้มากกว่า โดยผลความผิดพลาดในการออกเสียงนั้นจะส่งผลกระทบต่อการรู้จำเสียงสระแบบเฟรม เนื่องจากเสียงสระที่แยกแยะเสียงสั้นและเสียงยาวผิดนั้น มีผลต่อการทดสอบแบบอ้างอิงของสระ ดังแสดงในตารางที่ 5.1

ตารางที่ 5.1 แสดงผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงยาวแบบเฟรม สำหรับผู้พูดเพศหญิง 3 คน (ทดสอบเฉพาะสระเสียงสั้นที่ผิดมาอยู่ในกลุ่มสระเสียงยาว)

แบบทดสอบ \ แบบอ้างอิง		สระหน้า			สระกลาง			สระหลัง			% ความถูกต้อง
		อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	
สระหน้า	อิ	46	12			5	17	10			51.11
	เอะ	3	69				92				42.07
	แอะ			37			38			3	47.44
สระกลาง	อะ										
	อึ		4			79	8	61			51.97
	เออะ	1		1	9	3	276			1	94.85
สระหลัง	อุ						1	56			98.25
	โอะ	1			1	1	31	37	4	39	3.51
	เอาะ							2	1	27	90.00
ค่าเฉลี่ย											59.90

จากตารางที่ 5.1 แสดงผลการทดสอบแบบอ้างอิงสระเดี่ยวเสียงยาวแบบเฟรม สำหรับผู้พูดเพศหญิง 3 คน ที่ทำการทดสอบเฉพาะสระเสียงสั้นที่ผิดมาอยู่ในกลุ่มสระเสียงยาว เพื่อศึกษาถึงแนวโน้ม ของเสียงสระที่ทดสอบกับแบบอ้างอิงที่แตกต่างกันในลักษณะของค่าพลังงาน แต่จะมีค่าสเปกตรัมแอลพีซีของเสียงสระที่ใกล้เคียงกัน จากการทดสอบได้ผลความถูกต้องในการรู้จำเสียงสระแบบเฟรมของสระเสียงเดี่ยว ประมาณ 60% ซึ่งพอที่จะรู้จำเสียงสระบางสระได้ ด้วยวิธีการเปรียบเทียบคู่เสียงสระเดี่ยว 9 คู่เสียง แต่ยังไม่สามารถรู้จำได้ทุกเสียงสระ ซึ่งจะต้องศึกษาเพื่อหาวิธีในการวิเคราะห์และพัฒนาต่อไป

ส่วนผลจากการทดสอบการรู้จำเสียงสระ 24 เสียง ทั้งแบบเฟรม และแบบคำนั้น ค่าความผิดพลาดที่เกิดขึ้นนั้นส่วนใหญ่จะเกิดจากกลุ่มสระเดียวกัน เนื่องจากเสียงสระในกลุ่มเดียวกันมีลักษณะของ สเปกตรัมแอลพีซีและค่าความเข้มสัญญาณในแต่ละความถี่วิกฤตที่ใกล้เคียงกันจึงทำให้การรู้จำเกิดความผิดพลาดได้

5.3 ข้อเสนอแนะ

5.3.1 การรู้จำเสียงสระ 24 เสียงในภาษาไทย นี้เป็นแนวทางในการวิจัยการรู้จำแบบแยกแยะรายละเอียดของหน่วยเสียงที่สามารถรู้จำหน่วยเสียงสระในภาษาไทยได้ทั้ง 24 หน่วยเสียง ซึ่งสามารถพัฒนาไปสู่ระบบการรู้จำคำพูดในภาษาไทยได้ แต่เนื่องด้วยโปรแกรมที่ใช้ในการคำนวณยังใช้เวลานาน เนื่องจากต้องใช้ข้อมูลจำนวนมากในการทดลอง และโปรแกรมที่เขียนขึ้นยังเป็นการสร้างแบบจำลองของสระเสียงเดี่ยวและสระเสียงผสมอยู่ จึงเป็นเรื่องที่น่าสนใจที่จะเขียนโปรแกรมเพื่อพัฒนาให้สามารถ นำหลักการรู้จำนี้ไปใช้งานได้จริง

5.3.2 เสียงที่อัดจากไมโครโฟน มีสัญญาณรบกวนจากภายนอก ทำให้เสียงที่นำมาวิเคราะห์มีลักษณะผิดเพี้ยนไปบ้าง โดยถ้าเพิ่มขั้นตอนในการลดสัญญาณรบกวน จะทำให้ได้เสียงที่นำไปวิเคราะห์มีความสมบูรณ์มากยิ่งขึ้นก่อนที่จะนำไปวิเคราะห์ต่อไป

5.4 การประยุกต์ใช้งาน

5.4.1 นำหลักการรู้จำเสียงสระ 24 เสียงในภาษาไทย ไปประยุกต์ใช้กับการรู้จำระบบคำพูดในภาษาไทยแบบแยกแยะรายละเอียดของหน่วยเสียงได้

5.4.2 นำไปประยุกต์ใช้กับงาน ที่พัฒนาทางด้านการควบคุมสั่งงานด้วยเสียงได้

เอกสารอ้างอิง

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

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

[3] ณัฏฐชา จิตติวรางกูล. "ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย" วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 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 ลิ่ววิระยะ และวิภาวดี อ่วมเจริญ. "การประมาณค่าด้วยสมการพหุนาม." วิทยานิพนธ์วิทยาศาสตรบัณฑิต คณะวิทยาศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2541.

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

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

[10] ปราโมทย์ เดชะอำไพ. **ระเบียบวิธีเชิงตัวเลขในงานวิศวกรรม.** กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2538.

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

[12] กัญณ์ สิทธิประเสริฐ, ไกรสิน ส่งวัฒนา. "การแยกสระเสียงสั้น และ สระเสียงยาวในภาษาไทยโดยใช้การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง" **วิศวกรรมสารลาดกระบัง,** ปีที่ 21, ฉบับ 2, มิถุนายน 2547. หน้า 43-48

[13] SongWatana K., Suktangman N., "Mixed and Unmixed vowels recognition in Thai spoken language by classifying frames of The Transfer Function on Bark Scale using KNN technique" **Proc. The 1st KMITL International Conference Integration of Science and Technology for Sustainable Development Bangkok,** Thailand, August, 2004. Vol.1. Pp. 251-154

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

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

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

[17] David M. Haward, James Angus. **Acoustic and Psychoacoustic.** Prentice Hall, 1996

ภาคผนวก ก.
ผลงานวิจัยที่ได้รับการตีพิมพ์

1. **พีรพัฒน์ กุลทัพ**, ไกรสิน ส่งวัฒนา. “การรู้จำสระเสียงสั้น และสระเสียงยาวในภาษาไทยโดยใช้การฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสาม บนฟังก์ชันพลังงานเสียง ด้วยเทคนิค KNN” วิศวกรรมสาร มก, ฉบับที่ 53 , สิงหาคม – พฤศจิกายน 2547.
2. **พีรพัฒน์ กุลทัพ** , ไกรสิน ส่งวัฒนา. “การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียง และสเปกตรัมแอลพีซี บนสเกลบาร์กโดยใช้เทคนิค KNN” วิศวสารลาดกระบัง, ปีที่ 21, ฉบับที่ 4, ธันวาคม 2547.

การรู้จำสระเสียงสั้น และ สระเสียงยาวในภาษาไทย โดยใช้การฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสาม บนฟังก์ชันพลังงานเสียง ด้วยเทคนิค KNN

Short and Long Vowels Recognition in Thai Spoken Language Using 3rd Order Polynomial Coefficient Curve Fitting on the Voice Energy Function using KNN technique

พีรพัฒน์ กุลทัพ ไกรสิน ส่งวัฒนา

ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

บทความนี้เสนอ วิธีการรู้จำสระเสียงสั้น และ สระเสียงยาวในภาษาไทย โดยใช้ ลักษณะของพลังงานเสียง เป็นตัวบ่งชี้ โดยนำสัญญาณเสียงแต่ละเสียงมาคำนวณหาค่าพลังงานของเสียง และนำลักษณะของพลังงานเสียง มาสร้างเป็นฟังก์ชันพหุนามอันดับสาม เพื่อนำค่าสัมประสิทธิ์ ของฟังก์ชั่นพหุนาม ไปแยกแยะ สระเสียงสั้น และ สระเสียงยาว โดยใช้เทคนิคของ KNN

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) โดยใช้ข้อมูลเสียงจากผู้พูดคนไทย ทั้งหมด 10 คน แบ่งเป็นเพศชาย 5 คน และ เพศหญิง 5 คน รวมทั้งสิ้น 5,040 เสียง จากผลการทดลองให้เปอร์เซ็นต์ความถูกต้อง 94.05%

Abstract

This paper presents method short and long vowels recognition in Thai spoken language by using voice energy feature. The feature of energy is fitted by using 3^{rd} order polynomial regression function. The coefficient of polynomial regression is used to classify the short and long vowels by K-Nearest Neighbor technique.

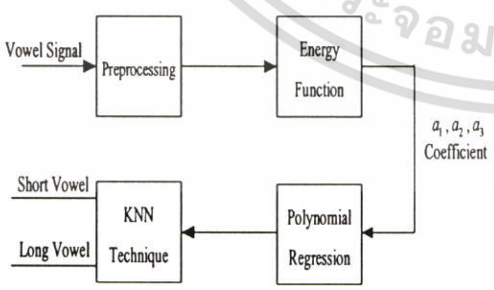
The vowels in Thai spoken language consist of 9 unmixed short vowels (a, i, ω, u, o, e, ε, γ, ɔ̇), 9 unmixed long vowels (aa, ii, ωω, uu, oo, ee, εε, γγ, ɔ̇ɔ̇), 3 mixed short vowels (ia, ωa, ua) and 3 mixed long vowels (iia, ωωa, uua). The voice data samples consist of 5,040 words are collected from 10 speakers, 5-male and 5-female Thai subjects. The result of percentage accuracy 94.05%

1.บทนำ

ในระบบการรู้จำเสียงพูดภาษาไทยนั้น มีความซับซ้อนมาก เนื่องจากในหนึ่งคำ หรือ พยางค์ ประกอบด้วยเสียงวรรณยุกต์ เสียงพยัญชนะ และเสียงสระ ซึ่งในส่วนของ การรู้จำเสียงสระนั้น ยังไม่สามารถระบุเสียงสั้น และเสียงยาวได้ เนื่องจากความแตกต่างของสระเสียงสั้น และสระเสียงยาวที่เห็นได้ชัดที่สุดคือ ระยะเวลาในการออกเสียง แต่ความแตกต่างนี้ไม่ทำให้สามารถแยกแยะ เสียงที่สั้นและยาวระหว่างบุคคลได้ โดยในบทความนี้จะเสนอวิธีการรู้จำ สระเสียงสั้น และสระเสียงยาว โดยขั้นตอนจะเริ่มจาก นำเสียงสระจากตัวอย่างเสียง ซึ่งประกอบด้วยสระเสียงเดี่ยว 9 คู่ และ สระผสม 3 คู่ ที่ผันตามเสียงพยัญชนะ 21 หน่วยเสียง มาหาระดับพลังงานของเสียงสระโดยใช้พลังงานกำลังสอง [1] จากนั้นนำลักษณะพลังงานเสียง มาสร้างฟังก์ชันการถดถอยพหุนามกำลังสาม และใช้ค่าสัมประสิทธิ์ฟังก์ชัน การถดถอยพหุนาม a_3, a_2 และ a_1 มาใช้ในการแยกแยะ และ จัดกลุ่มสระเสียงสั้น และ สระเสียง ยาว โดยใช้วิธีการของ K-Nearest Neighbor

2. ขั้นตอนการแยกสระเสียงสั้น และ สระเสียงยาว

ในการแยกแยะเสียงสระเสียงสั้น และ เสียงสระเสียงยาวนั้นแบ่งออกเป็นแบ่งออกเป็น 4 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 [2]

2.1.1 การปรับบรรทัดฐานแอมปลิจูด (Amplitude Normalization)

การปรับบรรทัดฐานของสัญญาณเสียงเพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน ดังสมการที่ (1)

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

เมื่อ $S[i]$ คือ สัญญาณเสียง

$\widetilde{S}[i]$ คือ สัญญาณเสียงที่ผ่านการปรับบรรทัดฐานแอมปลิจูด

N คือจำนวนบิต(Bit) ที่ใช้แทนค่า ของสัญญาณ เสียงพูด

2.1.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis)

กรรมวิธีการเน้นล่วงหน้าจะทำให้ความลาดเอียงในเชิงความถี่แบนราบลงและทำให้อัตราส่วนของสัญญาณเสียงพูดต่อสัญญาณรบกวนมีค่าสูงขึ้น ดังสมการที่ (2)

$$S'[i] = \widetilde{S}[i] - a\widetilde{S}(i-1) \tag{2}$$

เมื่อ $S'[i]$ คือ สัญญาณที่ผ่านการพรีเอมฟาซีส

a คือสัมประสิทธิ์ของตัวกรอง (0.95)

2.1.3 กรรมวิธีการวางกรอบหน้าต่าง (Windowing)

กรรมวิธีการวางกรอบหน้าต่าง เป็นการป้องกันการเปลี่ยนแปลง อย่างทันทีทันใดที่บริเวณปลายของกรอบเสียงพูด ดังสมการที่ (3) และ สมการที่ (4)

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

$$S''_n[i] = S'_n[i] * W_n \qquad (4)$$

เมื่อ $S''_n[i]$ คือสัญญาณเสียงพูดที่ i ในกรอบเสียงพูด n

n= 0,1,…..,N-1

2.2 การวิเคราะห์หาพลังงานเสียงพูด (Energy Function)

การคำนวณหาค่าพลังงานเสียง จะทำทีละกรอบเสียงพูด ซึ่งการคำนวณหาค่าพลังงานของสัญญาณเสียงพูดจะใช้ลักษณะพลังงานกำลังสอง (Square Energy) ซึ่งเป็นการวัดค่าพลังงาน จากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ ดังสมการที่ (5)

$$En = \left[\sum_{i=1}^{K} S''^{2}_n[i] \right] \qquad (5)$$

เมื่อ En คือ ค่าพลังงานของกรอบเสียงพูดที่ n

K คือจำนวนสัญญาณเสียงในแต่ละกรอบเสียงพูด

ข้อมูลพลังงานของแต่ละเสียงสระ จะถูกนำมา Normalize ให้ค่าพื้นที่ใต้ function เป็น 1 ด้วยสมการที่ (6) ก่อนที่จะนำไปทำการทดสอบ

$$Q_n = AE_n \qquad (6)$$

$$\sum_{N}^{M} AE_n = 1 \qquad (7)$$

$$A = \frac{1}{\sum_{N}^{M} E_n} \qquad (8)$$

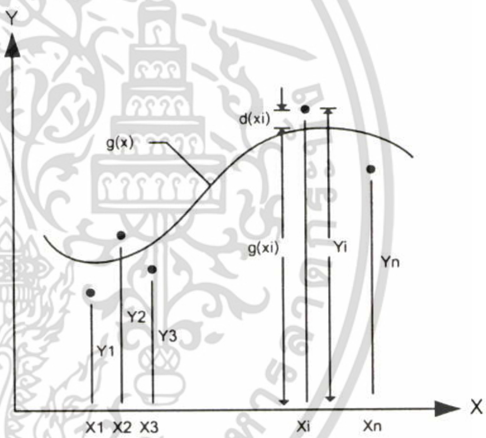
เมื่อ M คือจำนวนแซมปลิ้งสัญญาณเสียงในหนึ่งเฟรม

A คือ Normalize factor

Q_n คือค่า E_n ที่ ผ่านการ Normalize

2.3 การถดถอยแบบพหุนาม

การถดถอยแบบพหุนาม (polynomial regression) เป็นระเบียบวิธี ที่ใช้สร้างฟังก์ชั่นพหุนามสำหรับข้อมูลที่มีการกระจายโดยทั่วไป ที่ไม่อยู่ในรูปของเชิงเส้น (nonlinear) ซึ่งชุดข้อมูลดังรูปที่ 2 เป็นชุดของข้อมูลที่ประกอบด้วย $x_i, y_i, i = 1, 2, \ldots, n$ กล่าวคือ มีจำนวนข้อมูลทั้งสิ้น n ข้อมูล และ สามารถสร้างฟังก์ชั่นพหุนามอันดับ m ได้ดังสมการที่ (9)



รูปที่ 2 ชุดข้อมูลการสร้างฟังก์ชันพหุนาม

$$g(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \qquad (9)$$

โดย $a_0, a_1, \ldots, a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่สร้างขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุด

ซึ่งฟังก์ชั่นการถดถอยพหุนาม เขียนให้อยู่ในรูปแบบของระบบสมการที่ประกอบด้วย m+1 สมการย่อยได้ดังนี้

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

โดยในบทความนี้ เลือกใช้การถดถอยพหุนามอันดับสาม เพราะลักษณะการกระจายของข้อมูลพลังงานเสียง ไม่เป็นเชิงเส้นที่มีลักษณะคล้ายรูปพาราโบล่าร์คว่ำ ที่มีการเปลี่ยนแปลงระดับพลังงานเสียง ระหว่างช่วงที่ออกเสียง เมื่อนำฟังก์ชั่นพหุนามอันดับสาม มาประมาณค่า จะมีค่าใกล้เคียงกับ ฟังก์ชั่นพลังงานเสียงมากที่สุด และ ง่ายต่อการนำฟังก์ชั่นประมาณค่าที่ได้ ไปวิเคราะห์เพื่อแยกเสียงสระเสียงสั้น และเสียงสระเสียงยาว

โดยการถดถอยพหุนามอันดับสาม จะทำให้ได้ค่าสัมประสิทธิ์เป็น $a_0, a_1, a_2,$ และ a_3 แต่เนื่องจาก a_0 เป็นส่วนตัดแกน Y ซึ่งคือจุดเริ่มต้นของพลังงานเสียง ไม่เป็นส่วนสำคัญของการบ่งชี้ลักษณะฟังก์ชั่นพลังงานเสียง

ดังนั้นจึงนำเฉพาะค่า $a_1, a_2,$ และ a_3 มาเป็นตัวแยกแยะเสียง สระเสียงสั้น และ เสียงสระเสียงยาว

2.4 การแยกแยะโดยใช้เทคนิค KNN (K-Nearest Neighbor)

KNN เป็นวิธีที่ดัดแปลงมาจาก วิธี NN และให้ค่าความถูกต้องมากที่สุดเมื่อเทียบกับวิธีอื่น โดยในบทความนี้จะใช้การแยกแยะโดยมีเซตของ Feature Vector 3 มิติ ของค่าสัมประสิทธิ์การถดถอยพหุนาม a_3, a_2 และ a_1 ซึ่งอธิบาย ได้โดยเซตของเวกเตอร์ 3 มิติคือ $\mathrm{Im} = [(a_1)(a_2)(a_3)]_{j=1}^{N}$ โดยที่ j อยู่ในช่วง 1 ถึง N และเมื่อมีเวกเตอร์ของเสียงใดๆที่ไม่ทราบกลุ่ม เป็น $\mathrm{I_P} = [(a_1)(a_2)(a_3)]_{j=1}^{N}$ และต้องการที่จะจัดเข้ากลุ่มใดกลุ่มหนึ่งใน 2 กลุ่ม ซึ่งสามารถที่จะทำการคำนวณระยะห่างเวกเตอร์ ที่ไม่ทราบกลุ่มกับเวกเตอร์ทุกตัวในเซต $(\mathrm{I_m})_{j=1}^{N}$ โดยที่ j=1…N ระยะห่างที่ใช้คือระยะทางแบบ Euclidean หาได้โดยสมการ

$$\left| I_P - I_M \right| = \sqrt{\sum_{g=1}^{K} \left[(I_P)_g - (I_m)_g \right]^2} \tag{10}$$

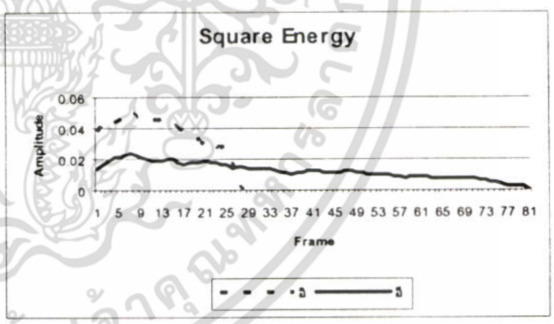
โดยที่ $(I_P)_g$ เป็นองค์ประกอบในมิติหนึ่งของเวกเตอร์ I_P

I_P เป็นเวกเตอร์เป้าหมายที่สนใจและต้องการหา

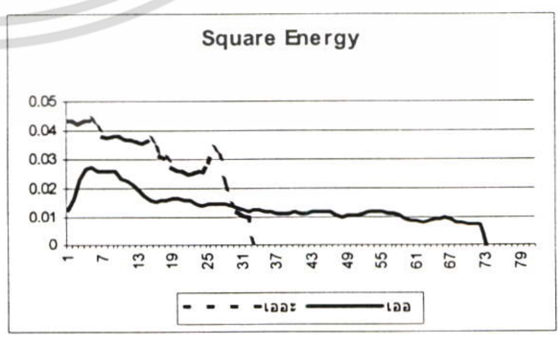
$(I_m)_g$ เป็น Nearest Neighbor

3. ผลการทดลอง

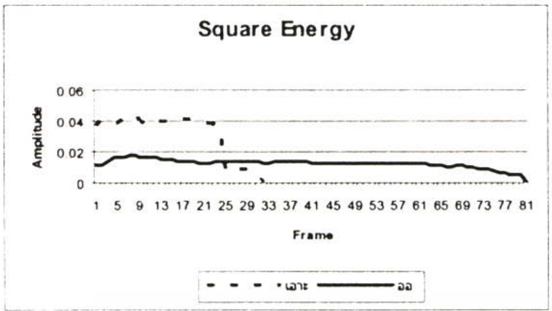
เมื่อนำสัญญาณเสียงสระ มาผ่านการเตรียมข้อมูลสัญญาณเสียงเพื่อนำไปคำนวณหาค่าพลังงานเสียงในรูปแบบพลังงานกำลังสอง จากนั้นทำการ Normalize ให้พื้นที่ function เป็น 1 จะมีลักษณะของพลังงานเสียงดังนี้



(a) เสียง อิ และ เสียง อี



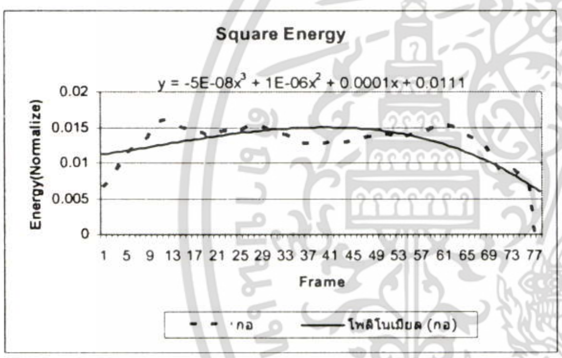
(b) เสียง เออะ และ เสียง เออ



(c) เสียง เอาะ และ เสียง ออ

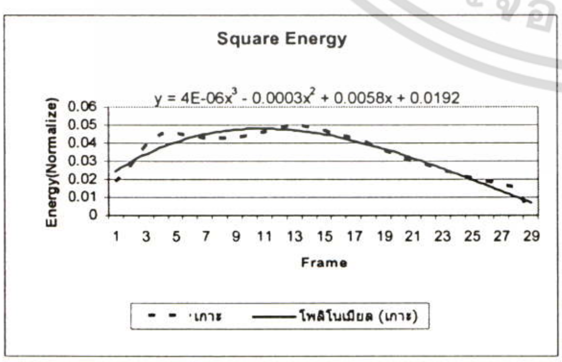
รูปที่ 3 ลักษณะพลังงานเสียงแบบ Square Energy

จากรูปที่ 3 จะเห็นได้ว่า ลักษณะการหยุดเสียงของสระเสียงสั้นนั้น จะเป็นการหยุดแบบทันทีทันใด ส่วนสระเสียงยาวจะมีลักษณะการหยุดเสียงแบบค่อยๆลดระดับลงช้าๆ



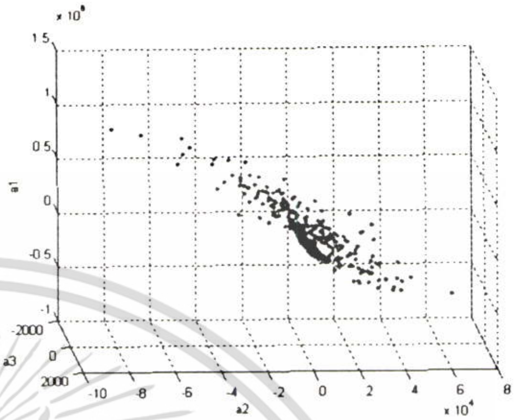
รูปที่ 4 สัมประสิทธิ์การถดถอยพหุนามอันดับสามของสระเสียงยาว

จากรูปที่ 4 จะได้ a_0 =0.0111, a_1 = 0.0001 , a_2 = 1E-06, a_3 = -5E-08



รูปที่ 5 สัมประสิทธิ์การถดถอยพหุนามอันดับสามของสระเสียงสั้น

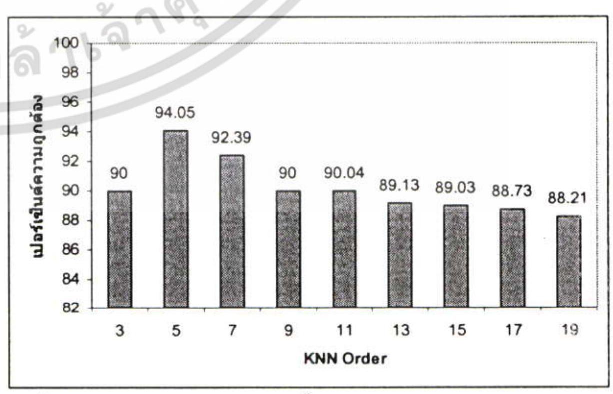
จากรูปที่5 จะได้ a_0 =0.0192, a_1 = 0.0058 , a_2 = -0.0003, a_3 = 4 E-06



รูปที่ 6 การกระจายข้อมูลสัมประสิทธิ์ a_1 , a_2 , a_3

จากรูปที่ 6 แสดงการกระจายข้อมูล a_1, a_2 และ a_3 นำค่าสัมประสิทธิ์ a_1, a_2 และ a_3 ไปจัดกลุ่มสระเสียงสั้น และ สระเสียงยาว เป็น Vector 3 มิติ จากสัมประสิทธิ์การถดถอยพหุนาม a_1, a_2 และ a_3 ของตัวอย่างเสียง 5,040 เสียง (สระเสียงสั้น 2,520 เสียง, สระเสียงยาว 2,520 เสียง)

จากนั้นทำการทดสอบโดยนำค่าสัมประสิทธิ์การถดถอยพหุนาม a_1, a_2 และ a_3 ของตัวอย่างเสียงอีก 5,040 เสียง (สระเสียงสั้น 2,520 เสียง, สระเสียงยาว 2,520 เสียง) มาทดสอบ ด้วยวิธี KNN โดยเลือกค่า K ค่าต่างๆ ซึ่งผลการทดสอบได้ดังนี้



รูปที่ 7 อัตราการแยกสระเสียงสั้นและสระเสียงยาวของเสียงทดสอบที่ค่า K ต่างๆ

จากรูปจะพบว่าค่า K ที่ให้อัตราการแยกแยะสระเสียงสั้นและสระเสียงยาวมากที่สุดคือ K= 5 ซึ่งให้อัตราการแยกแยะดังนี้

ตารางที่ 1 ผลการทดสอบการรู้จำสระเสียงสั้นและสระเสียง ยาว

Energy Algorithm	สระเสียงสั้น	สระเสียงยาว	เปอร์เซ็นต์
Square	2310	2430	94.05%

4. สรุป

ในผลการทดลอง[1] จะใช้วิธีการหาค่าสัมประสิทธิ์ a_2 ค่าหนึ่ง สำหรับ แยก สระเสียงสั้นและสระเสียงยาว ซึ่งจะได้ค่าที่ไม่แน่นอน และ เปลี่ยนแปลงตาม ชุดเสียงที่นำมาใช้หาค่า a_2 และค่าความถูกต้องก็เปลี่ยนแปลงตามค่า a_2 ที่เลือกใช้ โดยในบทความนี้ ได้นำวิธี KNN มาวิเคราะห์โดยสร้างเป็นแบบอ้างอิงเสียงจากค่าสัมประสิทธิ์ a_1, a_2, และ a_3 และนำไปใช้กับ เสียงทดสอบ ทำให้เกิดความแน่นอนในการรู้จำสระเสียงสั้นและเสียงยาวมากยิ่งขึ้นกว่าที่จะเลือกใช้เฉพาะ a_2 ค่าเดียว

บทความนี้ ให้เปอร์เซ็นต์ความถูกต้องอยู่ที่ 94.05 % จากการทดสอบสัญญาณเสียงทั้งสิ้น 5,040 เสียงทดสอบ ใช้เทคนิค KNN โดยเลือกใช้ค่า K = 5 ในการทดสอบ

5. เอกสารอ้างอิง

[1] กัญน์ สิทธิประเสริฐ , ไกรสิน ส่งวัฒนา, "การแยกสระเสียงสั้นและสระเสียงยาวในภาษาไทยโดยใช้ การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง" , วิศวสารลาดกระบัง , ฉบับที่ 2 , 2547.

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

[3] เบญจวรรณ อินทรารักษ์สกุล , " การประมาณค่าด้วยสมการพหุนาม" , บัณฑิต , สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,2541

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

[5] K. Songwatana and W. Kongkavitool " Unmixed Vowels Recognition in Thai spoken language using Vocal Tract Transfer functions on Bark Scale", Conference Proceddings WPMC'00 ,3 rd,vol.1,pp224-227,2000.

การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสามบนฟังก์ชั่นพลังงานเสียงและสเปกตรัมแอลพีซีบนสเกลบาร์กโดยใช้เทคนิค KNN

24 Vowels Recognition in Thai Spoken Language Using 3rd Order Polynomial Coefficient Curve Fitting on the Voice Energy and Spectrum LPC on Bark Scale by KNN Technique

พีรพัฒน์ กุลทัพ ไกรสิน ส่งวัฒนา

ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง

บทคัดย่อ

บทความนี้จะนำเสนอ วิธีการรู้จำหน่วยเสียงสระ 24 เสียงในภาษาไทย ซึ่งประกอบด้วยสระเดี่ยว เสียงสั้นจำนวน 9 เสียง (อะ อิ อึ อุ โอะ เอะ แอะ เออะ เอาะ) สระเดี่ยวเสียงยาวจำนวน 9 เสียง (อา อี อือ อู โอ เอ แอ เออ ออ) สระผสมเสียงสั้นจำนวน 3เสียง (เอียะ เอือะ อัวะ) และ สระผสมเสียงยาวจำนวน 3 เสียง (เอีย เอือ อัว)

สัญญาณเสียงจะถูกนำมาคำนวณหาค่าสัมประสิทธิ์การฟิตเคิร์ฟแบบพหุนามอันดับสาม บนฟังก์ชั่นพลังงานเสียง แล้วใช้วิธีการ KNN เพื่อแยกสระเสียงสั้น สระเสียงยาว จากนั้นจะถูกนำมาคำนวณหาค่าสเปกตรัมแอลพีซีแล้วแปลงให้อยู่บนสเกลบาร์ก นำค่าสเกลบาร์กที่ได้มาคำนวณหาค่าความเข้มในสัญญาณแต่ละความถี่วิกฤต เพื่อนำไปใช้สร้างแบบอ้างอิงการรู้จำสระเดี่ยวเสียงสั้น 9 โมเดล และสระเดี่ยวเสียงยาว 9 โมเดล แล้วนำเสียงทดสอบที่เหลือมาทดสอบกับแบบจำลองใช้กฎการตัดสินใจอย่างง่าย ในการบ่งชี้เสียงสระแต่ละเสียงโดยได้ผลความถูกต้องในการรู้จำ เสียงสระ 24 เสียงใน ผู้พูดคนเดียว เพศชาย 95% เพศหญิง 92% และ ผู้พูดหลายคนเพศชาย 94% เพศหญิง 91%

ABSTRACT

This paper presents 24 vowels recognition in Thai spoken language consisting of the 9 short unmixed vowels (a, i, ω, u, o, e, ε, γ, ɔ) , the 9 long unmixed vowels (aa, ii, ωω, uu, oo, ee, εε, γγ, ɔɔ) , the 3 short mixed vowels (ia, ωa, ua) and the 3 long mixed vowels (i:a, ω:a, u:a)

The coefficient of 3rd Order polynomial curve fitting is calculated on the voice energy. It is then used to classify short and long vowels by KNN technique. The LPC spectrum for the each frame of the audio signal are then calculated transformed on to the Bark Scale to determine the Critical Band Intensity. The 9 reference models short unmixed and the 9 reference models long unmixed recognition are created. A decision rule is applied to identify mixed and unmixed vowels in Thai language. The results of recognition a speaker dependent 24 vowels in Thai spoken language average percentage accuracy in male 95% female 92% and the results of recognition gender dependent 24 vowels in Thai spoken language average percentage accuracy in male 95% female 91%.

1.บทนำ

การรู้จำเสียงพูดในภาษาไทยนั้นมีความซับซ้อนมาก เนื่องจากในหนึ่งคำหรือพยางค์ประกอบด้วยเสียงวรรณยุกต์ เสียงพยัญชนะ และเสียงสระ ซึ่งในส่วนของการรู้จำเสียงสระนั้น ยังไม่สามารถระบุเสียงสั้น เสียงยาว และไม่สามารถรู้จำเสียงสระในภาษาไทยได้ทั้งหมด การรู้จำในอดีตที่ผ่านมาส่วนมากเป็นการรู้จำแบบเสียงที่เปล่งออกมาทั้งชุดมีทั้งคำพูดเดี่ยวและคำพูดต่อเนื่อง ทำให้ต้องใช้เนื้อที่จำนวนมากในการจัดเก็บแบบจำลองอ้างอิง ซึ่งในบทความนี้จะนำเสนอวิธีการรู้จำสระทั้ง 24 เสียงแบบแยกแยะรายละเอียดของหน่วยเสียงทำให้ไม่จำเป็นต้องใช้เนื้อที่ในการจัดเก็บแบบอ้างอิงมากและวิธีเหมาะสำหรับการพัฒนาไปสู่ระบบการรู้จำคำเป็นจำนวนมาก โดยขั้นตอนจะเริ่มจากการหาค่าพลังงานกำลังสอง จากนั้นนำค่าพลังงานเสียงที่ได้มาประดิษฐ์ฟังก์ชันการถดถอยพหุนามกำลังสามและใช้ค่าสัมประสิทธิ์ a_3 , a_2 และ a_1 มาใช้แยกแยะเสียงที่ได้จากการจัดกลุ่มสระเสียงสั้น และสระเสียงยาว โดยวิธี KNN เสียงที่ได้จากการจัดกลุ่มจะนำมาหาค่า LPC แล้วแปลงให้อยู่บนสเกลบาร์กเพื่อใช้ในการคำนวณหาค่าความเข้มสัญญาณในแต่ละแถบความถี่วิกฤต เพื่อนำไปใช้ในการสร้างแบบอ้างอิงของการรู้จำเสียงสระเสียงสั้น 9 โมเดล และสระเสียงยาว 9 โมเดล แล้วใช้กฎการตัดสินใจอย่างง่ายในการรู้จำเสียงสระผสมและสระเสียงเดี่ยวทั้ง 24 เสียง

2. ขั้นตอนการรู้จำเสียงสระในภาษาไทย

ในการวิเคราะห์เสียงสระเดี่ยวและสระเสียงผสมนั้นแบ่งออกเป็น 6 ขั้นตอน ดังแสดงในรูปที่ 1

2.1 การวิเคราะห์หาเสียงพูด(Energy Function)[1]

การคำนวณหาค่าพลังงานของสัญญาณเสียงพูดจะใช้ลักษณะพลังงานกำลังสอง (Square Energy) ซึ่งเป็นการวัดค่าพลังงาน จากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ ดังสมการที่ (1)

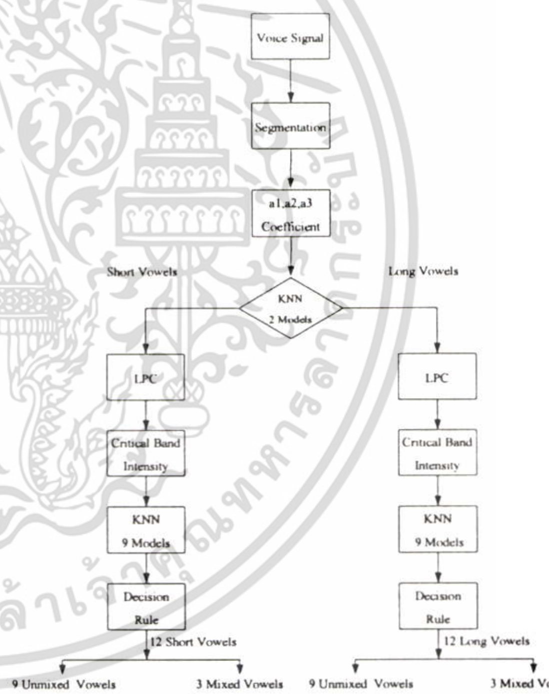
$$En = \left[\sum_{i=1}^{K} S_n^2[i] \right] \qquad (1)$$

นำค่าพลังงานของสัญญาณเสียงพูดแบบกำลังสองในกรอบสัญญาณเสียงพูดไปทำการฟิตเคิร์ฟ ด้วยวิธีการถดถอยแบบพหุนาม โดยใช้ค่าสัมประสิทธิ์อันดับสามในการคำนวณ

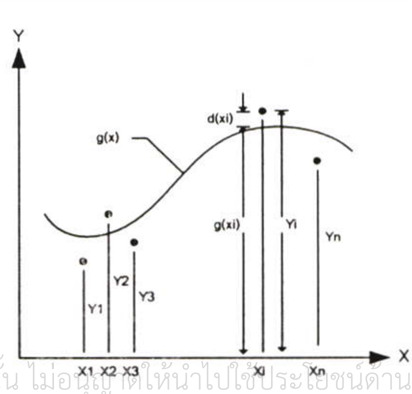
2.2 การถดถอยแบบพหุนาม [2]

ในผลการทดลอง[1] จะใช้วิธีการหาค่าสัมประสิทธิ์ a_2 ค่าหนึ่ง สำหรับ แยก สระเสียงสั้นและสระเสียงยาว ซึ่งจะได้ค่าที่ไม่แน่นอน และ เปลี่ยนแปลงตาม ชุดเสียงที่นำมาใช้หาค่า a_2 และค่าความถูกต้องก็เปลี่ยนแปลงตามค่า a_2 ที่เลือกใช้ โดยในบทความนี้ ได้นำวิธี KNN มาวิเคราะห์ โดยสร้างเป็นแบบอ้างอิงเสียงจากค่าสัมประสิทธิ์ $a_1, a_2,$ และ a_3 และนำไปใช้กับ เสียงทดสอบ ทำให้เกิดความแน่นอนในการรู้จำสระเสียงสั้นและเสียงยาวมากยิ่งขึ้นกว่าที่จะเลือกใช้เฉพาะ a_2 ค่าเดียว จากรูปที่ 2 เป็นชุดของข้อมูลที่ประกอบด้วย $x_i, y_i, i = 1, 2, ..., n$ กล่าวคือ มีจำนวนข้อมูลทั้งสิ้น n ข้อมูล และ สามารถประดิษฐ์ฟังก์ชั่นพหุนามอันดับ m ได้ดังสมการที่ (2)

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



รูปที่ 1 แสดงขั้นตอนการรู้จำเสียงสระในภาษาไทย



รูปที่ 2 ชุดข้อมูลการประดิษฐ์ฟังก์ชันพหุนาม

โดย $a_0, a_1, ..., a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่จะประดิษฐ์ขึ้นมานี้ จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุด จากนั้นนำค่า $a_0, a_1, ..., a_m$ ไปสร้างโมเดลในการแยกแยะสระเสียงสั้นและสระเสียงยาว โดยใช้เทคนิค KNN ซึ่งในบทความนี้ใช้ค่า K=5 ในการแยกแยะเสียงสระสั้นและเสียงสระยาวเพื่อนำไปใช้ในการรู้จำเสียงสระทั้ง 24 เสียง

ซึ่งฟังก์ชั่นการถดถอยพหุนาม เขียนให้อยู่ในรูปแบบของระบบสมการที่ประกอบด้วย m+1 สมการย่อย[1]

2.3 การคำนวณ Transfer function ของอวัยวะกำทอนเสียงจากสัมประสิทธิ์ที่ได้จากการทำ LPC

หลักพื้นฐานของการประมาณเชิงเส้นคือการประมาณค่าเชิงเส้นสัญญาณที่ผ่านมาจำนวน P ค่า ด้วยสมการ

$$s(n) \approx \sum_{k=1}^{P} a_k s(n-k) \qquad (3)$$

เมื่อ a_k เป็นสัมประสิทธิ์ของ LPC ที่อันดับ k โดยมีเงื่อนไขว่าค่าที่ใช้ในการประมาณต้องทำให้ผลรวมของกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด เราสามารถแปลงการประมาณ (3) ให้เป็นสมการได้โดยการรวมเอาสัญญาณการกระตุ้น Gu(n) เข้าไปด้วยจะได้

$$s(n) = \sum_{k=1}^{P} a_k s(n-k) + Gu(n) \qquad (4)$$

เมื่อแปลงสมการ (2) ให้มาอยู่ใน z-domain จะได้เป็น

$$S(z) = \sum_{k=1}^{p} a_k z^{-1} S(z) + GU(z) \qquad (5)$$

จัดรูปให้สามารถคำนวณ Transfer Function ของอวัยวะกำทอนเสียงได้ดังนี้

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k}^{P} a_k z^{-1}} \qquad (6)$$

2.4 Critical Band Rate Scale หรือสเกลบาร์ก

Critical Band Rate Scale หรือสเกลบาร์กนั้น เป็นสเกลที่ใช้ในการวัดสเปกตรัมในเชิงของ Psychoacoustics ซึ่งจะมีความใกล้เคียงกับระบบการได้ยินของมนุษย์มาก โดยค่าความเข้มพลังงาน ที่แสดงอยู่ในสเปกตรัมบนสเกลบาร์กนั้น จะแสดงให้เห็นถึงระดับการตอบสนองต่อเสียงที่ความถี่นั้นๆ ของระบบการได้ยินของมนุษย์ สมการต่อไปนี้เป็นสมการแสดงความสัมพันธ์ ระหว่างความถี่บนสเกลปกติเป็น Hz และสเกลบาร์กตลอดช่วงความถี่

เสียง

$$m(Bark) = 13\arctan(0.76f / KHz) + 3.5\arctan(f / 7.5KHz)^2 \qquad (7)$$

2.5 การใช้ Critical Band Intensity เป็น Feature Vector[3,4]

ระดับการตอบสนองต่อเสียงที่ความถี่ใดๆของระบบการได้ยินเสียงของมนุษย์นั้นสามารถคาดคะเนได้จากการหาความเข้มของสัญญาณเสียงในแต่ละ Critical Band โดยสามารถคำนวณได้จาก

$$I_m = \int_{m-0.5Bark}^{m+0.5Bark} \frac{dI}{dm} dm \qquad (8)$$

m เป็นค่า Critical Band Rate โดยที่ m=1,2,3…18

I_m เป็นค่าความเข้มของสัญญาณในแต่ละ Critical Band

2.6. การแยกแยะโดยใช้เทคนิค K-Nearest Neighbor (KNN)

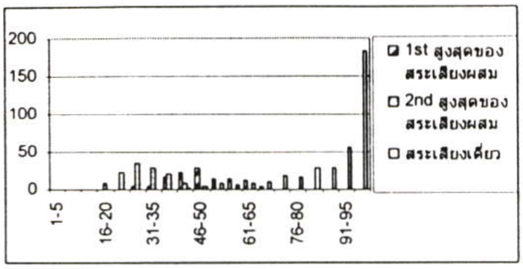
KNN เป็นกฎที่ดัดแปลงมาจากกฎ NN เพื่อให้เหมาะสมกับการรับรู้เสียงพูดแบบต่างบุคคล เทคนิค KNN นี้นิยมใช้ในการแยกแยะ โดยมีเซตของ Feature Vector 18 มิติ ของ Critical Band Intensity เสียงตัวอย่าง N เสียง มีเสียงสระ 12 รูปแบบ ซึ่งอธิบายได้โดยเซตของเวกเตอร์ 18 มิติ $I_m = [(I_1)(I_2)\ldots(I_{18})]_{j=1}^{N}$ และต้องการที่จะจัดเข้ากลุ่มใดกลุ่มหนึ่งใน 9 กลุ่ม ซึ่งสามารถ ที่จะทำการคำนวณระยะห่างเวกเตอร์ที่ไม่ทราบกลุ่ม $I_p = [(I_1)(I_2)\ldots(I_{18})]_{j=1}^{N}$ กับเวกเตอร์ทุกตัวในเซต $(I_m)_{i=1}^{N}$ โดยที่ i=1...N ระยะห่างที่ใช้คือระยะทางแบบ Euclidean หาได้โดยสมการ

$$\left|\underline{I_p} - \underline{I_m}\right| = \sqrt{\sum_{i=1}^{K}\left[(I_p)_i - (I_m)_i\right]^2} \qquad (9)$$

แล้วคำนวณหาเวกเตอร์ที่ใกล้ที่สุด K ตัว กับเวกเตอร์ที่ต้องการจัดกลุ่มและสามารถทำการจัดกลุ่มเวกเตอร์นั้นโดยตรวจดูว่าเวกเตอร์ทั้ง K ตัว ที่ใกล้เคียงที่สุดตรงกับกลุ่มใดมากที่สุดก็จะจัดให้เวกเตอร์ที่ต้องการทราบเข้ากลุ่มนั้น ซึ่งบทความนี้ใช้ค่า K=9

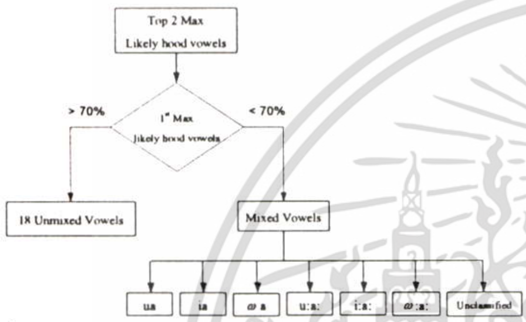
3. กฎการตัดสินใจอย่างง่าย (Decision Rule)

การวิเคราะห์เสียงสระที่ผ่านมาในแต่ละเฟรมของคำนำมาหาสถิติของคำโดยใช้ค่าความเปอร์เซ็นต์ความถูกต้อง จากนำค่าที่ได้มาพล็อตฮิสโตรแกรมได้ดังรูปที่ 3 โดยที่ค่าเฉลี่ยเปอร์เซ็นต์ความถูกต้องมีค่ามากกว่า 70% ใช้แยกเสียงสระเดี่ยวกับสระเสียงผสม



รูปที่ 3 แสดงสระเสียงเดี่ยวที่มีค่ามากที่สุดเมื่อเทียบกับทุกสระใช้แยกเสียงสระเดี่ยวกับสระเสียงผสม

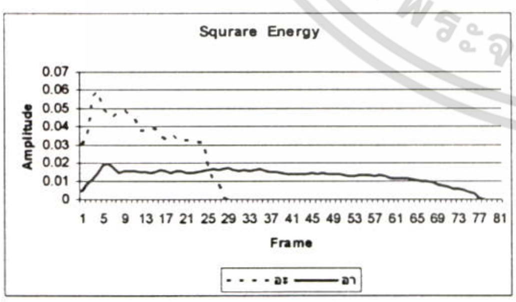
จากนั้นนำค่าสถิติของคำที่ได้จากฮิสโตรแกรมมาใช้ในการตัดสินใจอย่างง่ายแยกระหว่างสระเสียงเดี่ยวและสระเสียงผสมโดยใช้ค่ามากที่สุดสองอันดับแรก ดังรูปที่ 4



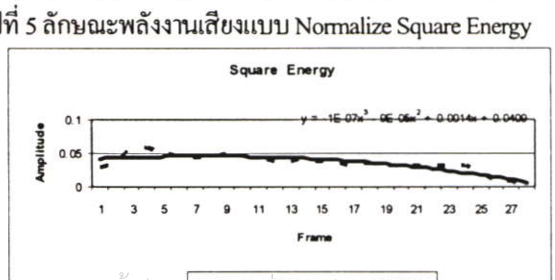
รูปที่ 4 แสดงกฎการตัดสินใจอย่างง่ายที่ใช้ในการแยกระหว่างสระเสียงเดี่ยวและสระเสียงผสม

4. ผลการทดลอง

สัญญาณเสียงสระที่ผ่านการเตรียมข้อมูลสัญญาณเสียงเพื่อนำไปคำนวณหาค่าพลังงานเสียงในรูปแบบ พลังงานกำลังสอง จากนั้นทำการ Normalize ให้พื้นที่ function เป็น 1 จะมีลักษณะของพลังงานเสียงดังนี้



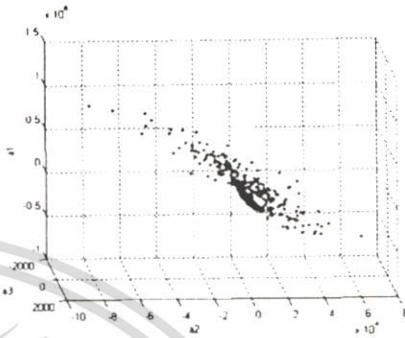
รูปที่ 5 ลักษณะพลังงานเสียงแบบ Normalize Square Energy



รูปที่ 6 สัมประสิทธิ์การถดถอยพหุนามอันดับสามของสระเสียงสั้น

จากรูปที่ 5 จะเห็นได้ว่า ลักษณะการหยุดเสียงของสระเสียงสั้นนั้น จะเป็นการหยุดแบบทันทีทันใด ส่วนสระเสียงยาวจะมีลักษณะการหยุดเสียงแบบค่อยๆลดระดับลงช้าๆ ดังรูปที่ 6

จากรูปที่ 6 จะได้ a_0 =0.0409, a_1 =0.0014, a_2 = -9E-05, a_3 = -1E-07



รูปที่ 7 การกระจายข้อมูลสัมประสิทธิ์ a_1, a_2, a_3

จากรูปที่ 7 แสดงการกระจายข้อมูลของค่า a_1, a_2 และ a_3 ซึ่งนำไปจัดกลุ่มสระเสียงสั้น และ สระเสียงยาว เป็น เวคเตอร์ 3 มิติ โดยใช้ค่าสัมประสิทธิ์ a_1, a_2 และ a_3 ของตัวอย่างเสียงทั้งหมด 504 เสียง จากนั้นนำมาสร้างแบบอ้างอิงสำหรับแยกแยะเสียงสั้นและเสียงยาวเพื่อทดสอบหาค่าเปอร์เซ็นต์ความถูกต้องได้ดังตารางที่ 1 และ 2 ตามลำดับ

ตารางที่ 1 ผลการทดสอบแบบอ้างอิงสำหรับแยกแยะสระเสียงสั้นและสระเสียงยาวของผู้พูดคนเดียว

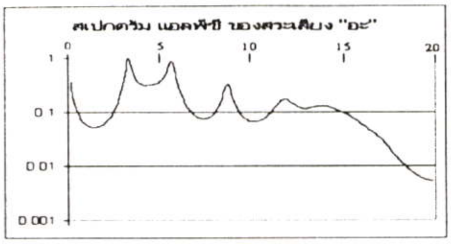
ผู้พูด	แบบอ้างอิง / แบบทดสอบ	สระเสียงสั้น	สระเสียงยาว	%ความถูกต้อง
คนเดียวชาย	สระเสียงสั้น	236	16	93.66
	สระเสียงยาว	5	247	98.02
คนเดียวหญิง	สระเสียงสั้น	239	13	94.84
	สระเสียงยาว	8	236	96.83
ค่าเฉลี่ย				95.84

ตารางที่ 2 ผลการทดสอบแบบอ้างอิงสำหรับแยกแยะสระเสียงสั้นและสระเสียงยาวของผู้พูดคนหลายคน

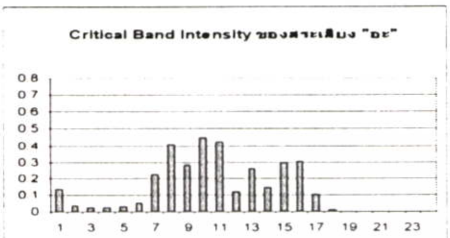
ผู้พูด	แบบอ้างอิง / แบบทดสอบ	สระเสียงสั้น	สระเสียงยาว	%ความถูกต้อง
หลายคนชาย	สระเสียงสั้น	679	77	89.81
	สระเสียงยาว	42	714	94.44
หลายคนหญิง	สระเสียงสั้น	694	62	91.8
	สระเสียงยาว	33	723	95.63
ค่าเฉลี่ย				92.92

เสียงที่ผ่านทดสอบแบบอ้างอิงแล้ว จะถูกนำมาใช้ ในการหาสเปกตรัมแอลพีซี เพื่อนำไปคำนวณหาค่าความเข้มของสัญญาณ ในแต่ละแถบความถี่วิกฤต ดังรูปที่ 7และ รูปที่ 8

ค่าความเข้มของสัญญาณในแต่ละแถบความถี่วิกฤตจะถูกนำไปใช้ในการสร้างแบบอ้างอิงเพื่อใช้ในการรู้จำเสียงสระเสียงเดี่ยว



รูปที่ 7 สเปกตรัมแอลพีซีบนสเกลบาร์กของสระเดี่ยวเสียงสั้น



รูปที่ 8 ค่าความเข้มของสัญญาณในแต่ละแถบความถี่วิกฤต

โดยจะแบ่งออกเป็นแบบอ้างอิงของสระเสียงสั้น 9 โมเดล และแบบอ้างอิงของสระเสียงยาว 9 โมเดล จากนั้นนำเสียงของสระเดี่ยวที่เหลือมาทดสอบแบบอ้างอิงสระเดี่ยวเสียงสั้นและแบบอ้างสระเดี่ยวเสียงยาวได้ดังตารางที่ 3 4 5 และ 6 ได้ค่าเปอร์เซ็นความถูกต้องประมาณ 95% 87% 90% และ 87%ตามลำดับ

ตารางที่ 3 ผลการทดสอบการรู้จำสระเสียงเดี่ยวแบบเฟรมของผู้พูดคนเดียวเพศชาย

สระเสียงสั้น	คิ	เคะ	แคะ	คะ	คึ	เคอะ	คุ	โคะ	เคาะ	% ความถูกต้อง
คิ	505	64	1		4	7		3		85.9
เคะ	48	528	9		10	9			1	87.1
แคะ	2	17	554	35	4	3	1			90.6
คะ			2	527	1	7	1	1	3	97.6
คึ	16	6	4	8	450	1	7		3	90.9
เคอะ		30	7	1	5	542		11	5	90.2
คุ	2	1			2	6	608	4		97.6
โคะ	1	1	2			4	10	578	10	95.4
เคาะ			2	4	5	8	1	5	505	95.6
สระเสียงยาว	คี	เค	แค	คา	คือ	เคอ	คู	โค	คอ	% ความถูกต้อง
คี	1409	36	3		23		1			96.6
เค	31	1409	33	1	3	2				95.3
แค	5	56	1303			4	7			95.1
คา			5	1305				1		99.5
คือ	16	9	1		1414	7	1			97
เคอ		4	1		9	1256		6		98.4
คู			1		6		1336	34		97.1
โค				1		4		1334		99.6
คอ						1			1136	99.9
ค่าเฉลี่ย										94.9

ตารางที่ 4 ผลการทดสอบการรู้จำสระเสียงเดี่ยวแบบเฟรมของผู้พูดคนเดียวเพศหญิง

สระเสียงสั้น	คิ	เคะ	แคะ	คะ	คึ	เคอะ	คุ	โคะ	เคาะ	% ความถูกต้อง
คิ	549	9	1	2	15	2	12			93.1
เคะ		322	95			20		2	5	72.5
แคะ	1	84	376	15		7	1	3	9	75.8
คะ	14	6		475	1		3	4	4	93.7
คึ	27	1		14	495	4	17	2		88.4
เคอะ		3	6		1	381		67	4	82.5
คุ	8			4	26	4	441	2	1	90.7
โคะ		1	8			55		300	38	74.6
เคาะ		3	5	13	1	23	2	29	358	82.5
สระเสียงยาว	คี	เค	แค	คา	คือ	เคอ	คู	โค	คอ	% ความถูกต้อง
คี	1075	7	1	8	16	2	7	8	4	95.3
เค	1	752	223	1		21		16	2	74.0
แค		40	875		1	30		4	11	90.3
คา	1			1053		1	3	3	1	99.2
คือ	15			2	940	4	63	3		91.6
เคอ		21	42	3	6	970		43	31	86.8
คู	3			4	16	2	990			97.3
โค		13	3	3		53		935	20	90.3
คอ		6		1		11		29	965	94.8
ค่าเฉลี่ย										87.4

ตารางที่ 5 ผลการทดสอบการรู้จำสระเสียงเดี่ยวแบบเฟรมของผู้พูดหลายคนเพศชาย

สระเสียงสั้น	คิ	เคะ	แคะ	คะ	คึ	เคอะ	คุ	โคะ	เคาะ	% ความถูกต้อง
คิ	1374	127	20	2	12	9	2	3	3	88.4
เคะ	124	1306	61	10	25	22	2	3	3	84.5
แคะ	11	40	1550	49	5	13	1	6		92.5
คะ	15	11	30	1372	4	10	3	12	23	92.6
คึ	46	17	5	8	1156	78	64	13	6	84.2
เคอะ	3	19	13	17	42	1444	7	33	33	88.4
คุ	13	11		4	27	9	1236	72	16	89.0
โคะ	4	3	1	27	12	24	35	1179	147	82.5
เคาะ		1	2	24	9	38	11	99	1390	88.3
สระเสียงยาว	คี	เค	แค	คา	คือ	เคอ	คู	โค	คอ	% ความถูกต้อง
คี	3867	99	11		42	1	44			95.2
เค	61	3496	171	8	11	16	10	1	1	93.0
แค	25	149	3594	50	10	20	8	1		93.2
คา	1	6	40	3990	3	18	13	6	2	96.9
คือ	105	25	15	3	3600	101	80	8	4	91.2
เคอ	3	14	14	8	32	3509	7	38	33	96.0
คู	23	11	6	1	17	3	3870	70		96.8
โค	14	5		10	12	36	125	3547	112	91.9
คอ	3			66	1	25	5	197	3230	92.3
ค่าเฉลี่ย										90.9

ตารางที่ 6 ผลการทดสอบการรู้จำสระเสียงเดี่ยวแบบเฟรมของผู้พูดหลายคนเพศหญิง

สระเสียงสั้น	คิ	เคะ	แคะ	คะ	คึ	เคอะ	คุ	โคะ	เคาะ	% ความถูกต้อง
คิ	1318	89	7	8	27	14	8	5		89.3
เคะ	73	1080	205	36	1	26	2	4	7	75.2
แคะ	35	31	1210	64	4	15	3	7	20	88.4
คะ	13	3	42	1190	9	11	10	11	17	90.8
คึ	38	6	4	12	1258	57	88	5	3	85.5
เคอะ	2	10	31	7	28	1152	11	26	32	88.7
คุ	9		7	3	40	3	1233	85	12	88.5
โคะ		4	4	19	6	79	25	1124	175	78.3
เคาะ	2	6	5	35	2	40	3	117	1278	85.9
สระเสียงยาว	คี	เค	แค	คา	คือ	เคอ	คู	โค	คอ	% ความถูกต้อง
คี	2983	130	20	13	65	5	22	5	3	91.9
เค	63	2662	341	8	3	34	4	17	11	84.7
แค	7	334	2755	76	3	42		8	20	84.9
คา	36	25	26	2902	19	20	21	9	16	94.4
คือ	83	7	51	34	2894	95	141	13	2	87.1
เคอ	16	18	34	25	50	2806	14	34	72	91.9
คู	8	3	1	6	95	26	2891	197	12	89.3
โค	15	17	6	11	4	85	117	2734	317	82.4
คอ	9	18	13	27	20	31	15	71	2752	92.8
ค่าเฉลี่ย										87.2

จากนั้นนำสระเสียงผสมมาทดสอบกับแบบอ้างอิงของสระเดี่ยวเสียงสั้นและแบบอ้างอิงสระเดี่ยวเสียงยาวเพื่อหาค่าเปอร์เซ็นความถูกต้อง แล้วใช้กฎการตัดสินใจอย่างง่ายในการวิเคราะห์โดยจะพิจารณาจากค่าเปอร์เซ็นต์ความถูกต้องของเสียงสระค่าที่มากที่สุดกับค่าที่มากรองลงมาเมื่อเทียบกับค่าอื่นๆของสระเสียงผสมในแต่ละคำ ค่าเปอร์เซ็นต์ความถูกต้องของสระเสียงเดี่ยวและสระเสียงผสมจะถูกนำมาหาค่าสถิติโดยใช้ฮิสโตรแกรมจะได้ค่าที่ใช้ในการแบ่งแยกระหว่างสระเสียงเดี่ยวและสระเสียงผสม โดยค่าเปอร์เซ็นต์ความถูกต้องที่ได้ประมาณ 70% [3] ผลจากการทดสอบการรู้จำสระเสียงเดี่ยวและสระเสียงผสม ของผู้พูดคนเดียวและผู้พูดหลายคนแบบต่างเพศ ได้ดังตารางที่ 7 8 9 และ10 ค่าเปอร์เซ็นต์ความถูกต้องประมาณ 95% 92% 94% และ 91% ตามลำดับ

ตารางที่ 7 ผลการทดสอบการรู้จำสระเสียงเดี่ยวและสระเสียงผสมของผู้พูดคนเดียวเพศชาย

สระเสียงสั้น		กิ	เกะ	แกะ	กะ	กึ	เกอะ	กุ	โกะ	เกาะ	กัวะ	เกียะ	เกือะ	Unclassify	% Accuracy
สระเสียงสั้น	กิ	19	1									1			90.5
	เกะ	1	19												95.0
	แกะ			19	1										95.0
	กะ				20										100.0
	กึ					20							1		95.2
	เกอะ		1				19								95.0
	กุ							18							100.0
	โกะ								21						100.0
	เกาะ									21					100.0
	กัวะ								2		19				90.5
	เกียะ		1	1								14			87.5
	เกือะ	1					1		1				15		83.3
สระเสียงยาว		กี	เก	แก	กา	กือ	เกอ	กู	โก	กอ	กัว	เกีย	เกือ		
สระเสียงยาว	กี	21													100.0
	เก		21												100.0
	แก		1	20											95.2
	กา				21										100.0
	กือ					21									100.0
	เกอ						21								100.0
	กู							21							100.0
	โก								21						100.0
	กอ									21					100.0
	กัว				1				1		14		1		82.4
	เกีย				2							19			90.5
	เกือ					2						1	17		85.0
ค่าเฉลี่ย															95.2

ตารางที่ 8 ผลการทดสอบการรู้จำสระเสียงเดี่ยวและสระเสียงผสมของผู้พูดคนเดียวเพศหญิง

สระเสียงสั้น		กิ	เกะ	แกะ	กะ	กึ	เกอะ	กุ	โกะ	เกาะ	กัวะ	เกียะ	เกือะ	Unclassify	% Accuracy
สระเสียงสั้น	กิ	19										1			95.0
	เกะ		19	1								1			90.5
	แกะ		3	16								2			84.2
	กะ				20										100.0
	กึ					19							1		95.0
	เกอะ						16		2				3		76.2
	กุ							21							100.0
	โกะ								18		1		1		90.0
	เกาะ				1				1	17	1		1		81.0
	กัวะ										20				95.2
	เกียะ											15	1		93.8
	เกือะ					1					1		15		88.2
สระเสียงยาว		กี	เก	แก	กา	กือ	เกอ	กู	โก	กอ	กัว	เกีย	เกือ		
สระเสียงยาว	กี	21													100.0
	เก		17	2							1	1			81.0
	แก		1	20											95.2
	กา				21										100.0
	กือ					20							1		95.2
	เกอ						19						2		90.5
	กู							18			3				85.7
	โก								19		2				90.5
	กอ									21					100.0
	กัว				1						20				95.2
	เกีย											17	1		94.4
	เกือ											2	14		87.5
ค่าเฉลี่ย															91.9

ตารางที่ 9 ผลการทดสอบการรู้จำสระเสียงเดี่ยวและสระเสียงผสมของผู้พูดหลายคนเพศชาย

สระเสียงสั้น		อิ	เอะ	แอะ	อะ	อึ	เออะ	อุ	โอะ	เอาะ	อัวะ	เอียะ	เอือะ	Unclassify	% Accuracy
สระเสียงสั้น	อิ	55	1									1		3	91.7
	เอะ	1	59											2	95.2
	แอะ			60	1									1	96.8
	อะ				61										100.0
	อึ					49							1	5	89.1
	เออะ		2				56							1	94.9
	อุ							50						2	96.2
	โอะ								50	2				2	92.6
	เอาะ									60				1	98.4
	อัวะ								2		55		1	3	90.2
	เอียะ		1									44	1	2	91.7
	เอือะ						2		1			1	40	1	88.9
สระเสียงยาว		อี	เอ	แอ	อา	อือ	เออ	อู	โอ	ออ	อัว	เอีย	เอือ		
สระเสียงยาว	อี	59												1	98.3
	เอ		60											3	95.2
	แอ		2	59										1	95.2
	อา				52										100.0
	อือ					58								3	95.1
	เออ						61							1	98.4
	อู							62							100.0
	โอ								61	1				1	96.8
	ออ									58				3	95.1
	อัว				2				1		46			1	92.0
	เอีย	1			3							52		2	89.7
	เอือ					2						3	52	4	85.2
ค่าเฉลี่ย															94.4

ตารางที่ 10 ผลการทดสอบการรู้จำสระเดี่ยวและสระผสมเสียงยาวของผู้พูดคนเดียวหญิง

สระเสียงสั้น		กิ	เกะ	แกะ	กะ	กึ	เกอะ	กุ	โกะ	เกาะ	กัวะ	เกียะ	เกือะ	Unclassify	% Accuracy
สระเสียงสั้น	กิ	57	1									1		1	95.0
	เกะ	1	49	2										5	86.0
	แกะ	1	3	54										3	88.5
	กะ			1	61								1		96.8
	กึ					54								4	93.1
	เกอะ		1				46		3					4	86.8
	กุ							58	1					2	95.1
	โกะ							1	51	3				4	86.4
	เกาะ				2				2	54	1			3	87.1
	กัวะ										52			6	89.7
	เกียะ											46	2	5	86.8
	เกือะ					2						3	42	4	82.4
สระเสียงยาว		กี	เก	แก	กา	กือ	เกอ	กู	โก	กอ	กัว	เกีย	เกือ		
สระเสียงยาว	กี	60												2	96.8
	เก		55	2								1		4	88.7
	แก		2	58										2	93.5
	กา				61										100.0
	กือ					59	1							3	93.7
	เกอ					2	58							3	92.1
	กู					1		56	2					2	90.3
	โก								58	3				2	92.1
	กอ									60				2	96.8
	กัว				1						43			4	89.6
	เกีย											52		4	92.9
	เกือ										1	2	47	6	82.9
ค่าเฉลี่ย															91.0

5. สรุป

การรู้จำเสียงสระ 24 เสียงในภาษาไทย ด้วยวิธีการฟิตเคิร์ฟแบบสัมประสิทธิ์พหุนามอันดับสามบนฟังก์ชั่นพลังงานเสียงและสเปกตรัมแอลพีซีบนสเกลบาร์กโดยใช้เทคนิค KNN นั้นจะให้ค่าเปอร์เซ็นต์ความถูกต้องในการรู้จำเสียงสระของผู้พูดคนเดียวเพศชาย 95% เพศหญิง 92% และผู้พูดหลายคนเพศชาย 94% เพศหญิง 91% จากการใช้กฎการตัดสินใจอย่างง่ายในการบ่งชี้สระแต่ละเสียง ซึ่งในการรู้จำแบบผู้พูดคนเดียวกับหลายผู้พูดนั้นมีค่าความถูกต้องใกล้เคียงกัน เนื่องจากเสียงสระที่ผิดจะอยู่ในกลุ่มเสียงเดียวกัน

6. เอกสารอ้างอิง

[1] กัญญ์ สิทธิประเสริฐ , ไกรสิน ส่งวัฒนา, "การแยกสระเสียงสั้นและสระเสียงยาวในภาษาไทยโดยใช้ การฟิตเคิร์ฟแบบพหุนามอันดับสอง บนฟังก์ชันพลังงานเสียง" , วิศวสารลาดกระบัง , ฉบับที่ 2 , 2547.

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

[3] K.Songwatana and N.Suktangman. "Mixed and Unmixed vowels recognition in Thai spoken language by classifying frames of The Transfer Function on Bark Scale using KNN technique" The 1st KMITL International Conference Integration of Science and Technology for Sustainable Development., 2004.

[4] K.Songwatana, W. Kongkavitool " Unmixed Vowels Recognition in Thai spoken language using Vocal Tract Transfer functions on Bark Scale" Conference Proceddings WPMC'00 ,3 rd,vol.1,pp224-227,2000.

ประวัติผู้เขียน

นายพีรพัฒน์ กุลทัพ เกิดเมื่อวันที่ 15 พฤศจิกายน พ.ศ. 2518 ที่จังหวัดสุพรรณบุรี สำเร็จการศึกษาระดับปริญญาตรี คณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมโทรคมนาคม จากสถาบันเทคโนโลยีราชมงคล ปีการศึกษา 2542