สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์

VOICE RECOGNITION

ปรางค์ทอง นุ่มเกลี้ยง
ปิยธิดา ไพรระหง
ศิริพร มีวุฒิดี

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2542

เลขหมู่.....
เลขทะเบียน 36152
วัน, เดือน, ปี 11 ก.ค. 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VOICE RECOGNITION

Prangthong Numklieng

Piyathida Prairahong

Siriporn Meevudhidee

A SPECIAL PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF SCIENCE

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCES

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 1999

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ Voice Recognition
ชื่อนักศึกษา	นางสาวปรางค์ทอง นุ่มเกลี้ยง 39054636
	นางสาวปิยธิดา ไพรระหง 39054639
	นางสาวศิริพร มีวุฒิดี 39054666
ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ธีรวัฒน์ ประกอบผล

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง อนุมัติให้นับปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ ประจำปีการศึกษา 2542

คณะกรรมการสอบ	ลายมือชื่อ
ประธานกรรมการ อาจารย์วีระชัย ตันยะสิทธิ์	
กรรมการ อาจารย์วิสันต์ ตั้งวงษ์เจริญ	
กรรมการและอาจารย์ที่ปรึกษา อาจารย์ธีรวัฒน์ ประกอบผล	

(อาจารย์ไพรบูลย์ พันธรักษ์พงษ์)
หัวหน้าภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

ลิขสิทธิ์ของภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์
นักศึกษา	นางสาวปรางค์ทอง นุ่มเกลี้ยง 39054636
	นางสาวปิยธิดา ไพรระหง 39054639
	นางสาวศิริพร มีวุฒิดี 39054666
ปริญญา	วิทยาศาสตรบัณฑิต
ภาควิชา	คณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
สาขาวิชา	วิทยาการคอมพิวเตอร์
ปีการศึกษา	2542
อาจารย์ที่ปรึกษา	อาจารย์ธีรวัฒน์ ประกอบผล

บทคัดย่อ

ปัญหาพิเศษฉบับนี้ เป็นการวิเคราะห์เสียงเพื่อให้มนุษย์ สามารถสื่อสารกับคอมพิวเตอร์ได้ โดยที่ ผู้พูดจะพูดคำสั่ง 1 คำสั่งผ่านทางไมโครโฟน คำสั่งที่เป็นเสียงนี้จะถูกนำมาวิเคราะห์ แล้วแสดงผลในรูปของภาพกราฟฟิกสามมิติซึ่งจะเคลื่อนไหวสัมพันธ์กับคำสั่งเสียงที่ผู้พูดได้ทำการพูดผ่านเข้ามา

วัตถุประสงค์สำหรับปัญหาพิเศษนี้ คือการเพิ่มความสามารถในการติดต่อสื่อสารกับคอมพิวเตอร์ที่นอกเหนือไปจากอุปกรณ์อินพุทที่มีอยู่เดิม นอกจากนี้ยังเป็นแนวทางในการศึกษาเพื่อนำไปประยุกต์ใช้กับแอพพลิเคชันอื่นๆ และเป็นแนวทางในการสร้างเทคโนโลยีใหม่ๆ ที่เกี่ยวข้องกับการรู้จำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Special Project Title	Voice Recognition
Students	Miss Prangthong Numklieng 39054636
	Miss Piyathida Prairahong 39054639
	Miss Siriporn Meevudhidee 39054666
Degree	Bachelor's Degree of Science
Department	Mathematics and Computer Sciences, Faculty of Science
Programme	Computer Sciences
Academic Year	1999
Special Project Advisor	Lecturer Teerawat Prakobphon

ABSTRACT

This project presents Voice Recognition. As man speaks one instruction into microphone, his voice instruction will be synthesized. Then animation of an 3D graphic picture will show, it relate to the man's voice instruction.

The objectives of this project are development about communication with computer, using this project as guide to develop another application and to create the new technology about Voice Recognition.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ในการทำปัญหาพิเศษเรื่อง การวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ (Voice Recognition) สามารถสำเร็จลงไปได้ด้วยดี คณะผู้จัดทำต้องขอขอบพระคุณ อาจารย์ธีรวัฒน์ ประกอบผล อาจารย์ผู้รับผิดชอบปัญหาพิเศษฉบับนี้ ที่กรุณาให้คำแนะนำ และเป็นที่ปรึกษา ในเรื่องแนวทางการพัฒนาโปรแกรม การแก้ไขปัญหาต่าง ๆ รวมทั้งเป็นผู้ตรวจสอบความถูกต้องของปัญหาพิเศษฉบับนี้

นอกจากนี้คณะผู้จัดทำขอขอบพระคุณ บิดา มารดา ที่ได้ให้ความสนับสนุนทางด้านกำลังใจและทุนทรัพย์ จนการทำปัญหาพิเศษครั้งนี้สำเร็จด้วยดี รวมทั้งเพื่อน ๆ ทุกคนที่ให้ความช่วยเหลือในด้านต่าง ๆเกี่ยวกับปัญหาพิเศษไว้ ณ ที่นี้

คณะผู้จัดทำ

มีนาคม 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

กิตติกรรมประกาศ.....III

สารบัญ.....IV

สารบัญตาราง.....VII

สารบัญภาพ.....VIII

ตารางสัญลักษณ์.....XII

บทที่ 1 บทนำ.....1

1.1 ที่มาของปัญหาพิเศษ.....1

1.2 วัตถุประสงค์ของปัญหาพิเศษ.....1

1.3 ขอบเขตปัญหาพิเศษ.....1

1.4 ระยะเวลาในการดำเนินงานปัญหาพิเศษ.....2

1.5 ขั้นตอนการดำเนินงาน.....2

1.6 ประโยชน์ที่คาดว่าจะได้รับ.....3

1.7 ข้อตกลงเบื้องต้น.....3

1.8 อุปกรณ์ที่ใช้ในการทำปัญหาพิเศษ.....4

บทที่ 2 ทฤษฎีและหลักการที่เกี่ยวข้อง.....5

2.1 ลักษณะของเสียงพูด.....5

2.1.1อวัยวะที่ใช้ในการออกเสียง.....1

2.1.2การเกิดเสียง.....6

2.2 การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้น ๆ.....7

2.3 การหาค่าพลังงานของสัญญาณเสียง.....8

2.4 การหาค่าอัตราตัดศูนย์(Zero Crossing).....9

2.5 การประมาณเชิงเส้น(Linear Predictive Coding).....10

2.6 เวคเตอร์ควอนไตซ์เซชัน (Vector Quantization).....14

2.6.1 เวคเตอร์ควอนไตซ์เซชัน.....14

2.6.2 ส่วนประกอบของเวคเตอร์ควอนไตซ์เซชัน.....16

2.6.3 หลักการของเวคเตอร์ควอนไตซ์เซชัน.....17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.4 การคำนวณเวคเตอร์ควอนไตซ์เซชัน.....17
2.6.5 การวัดความคลาดเคลื่อน.....18
2.6.6 การจัดกลุ่มเพื่อสร้างแบบอ้างอิงโดยใช้ K-means Algorithm.....18
2.6.7 การคำนวณจุดศูนย์กลางสำหรับการออกแบบ
เวคเตอร์ควอนไตซ์เซชันโค้ดบุค.....19
2.6.8 กฎการตัดสินใจ (Decision Rules) 20
2.6.9 เวคเตอร์ควอนไตซ์เซชันสำหรับระบบการรับรู้เสียงพูด.....21
2.7 รูปแบบของไฟล์เวฟ.....22
2.7.1 โครงสร้างไฟล์เวฟ.....22
2.7.2 Sample Points และ Sample Frames.....23
2.7.3 Format Chunk.....25
2.7.4 Data Chunk.....26
2.8 ข้อมูลเกี่ยวกับซอฟต์แวร์.....27
2.8.1 Microsoft Visual C++.....27
2.8.2 Multimedia File I/O Services.....28
2.8.3 Microsoft Visual Basic.....31
2.8.4 Poser 4.....31
บทที่ 3 วิธีดำเนินการวิจัย 32
3.1 ส่วนของพรีโพรเซสซิ่ง.....32
3.1.1 การหาขอบเขตของคำโดยใช้การวิเคราะห์หาค่าพลังงานของสัญญาณ.....32
3.1.2 การหาขอบเขตของคำโดยใช้การวิเคราะห์หาค่าอัตราตัดศูนย์.....38
3.1.3 การนอร์มอลไลซ์.....44
3.2 การหาค่าพารามิเตอร์.....48
3.2.1 การพรีเอมฟาซิส.....48
3.2.2 การแบ่งช่วงสัญญาณ.....49
3.2.3 การวินโดว์.....49
3.2.4 การหาออโตคอร์รีเลชัน.....50
3.2.5 การหาค่าสัมประสิทธิ์แอลพีซี.....50
3.2.6 การหาอัตราการขยาย G 51
3.2.7 การเปลี่ยนพารามิเตอร์แอลพีซีเป็นสัมประสิทธิ์เซปสตรัม.....51
3.2.8 การเวทค่าพารามิเตอร์.....52
3.3 เวคเตอร์ควอนไตซ์เซชัน.....55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 ค่าการตัดสินใจ.....59
บทที่ 4 การทดลองและผลการทดลอง.....60
4.1 การเปรียบเทียบความคล้ายคลึงของเสียง.....61
4.1.1 เปรียบเทียบโดยใช้ฐานเสียงที่ 1(กลุ่ม A) 61
4.1.2 เปรียบเทียบโดยใช้ฐานเสียงที่ 2(กลุ่ม B) 72
4.2 การทดสอบการจดจำเสียงพูด.....103
4.3 การแสดงหน้าจอการทำงานและวิธีการใช้งานโปรแกรม.....108
4.3.1 การเรียกเข้าโปรแกรม.....108
4.3.2 การใช้งานโปรแกรม.....109
4.3.3 ข้อจำกัดในการใช้งานโปรแกรม.....113
บทที่ 5 บทวิจารณ์และบทสรุป.....114
5.1 บทสรุป.....114
5.2 บทการวิจารณ์และแนวการพัฒนา.....114
ภาคผนวก ก คำชี้แจงการติดตั้งโปรแกรม.....116
บรรณานุกรม.....118

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

ตารางสัญลักษณ์.....XII

1.1 ระยะเวลาในการดำเนินงานปัญหาพิเศษ.....2

4.1 แสดงการจดจำเสียง “เดิน”, “วิ่ง”, “ยิ้ม”, “หาว” โดยใช้สัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักเปรียบเทียบเมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม A และนำเสียงผู้พูดอ้างอิงมาเป็นแบบทดสอบ.....103

4.2 แสดงการจดจำเสียง “เดิน”, “วิ่ง”, “ยิ้ม”, “หาว” โดยใช้สัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักเปรียบเทียบเมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม A และนำเสียงที่ไม่ได้เป็นแบบอ้างอิงมาเป็นแบบทดสอบ.....104

4.3 แสดงการจดจำเสียง “เดิน”, “วิ่ง”, “ยิ้ม”, “หาว” โดยใช้สัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักเปรียบเทียบเมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม B และนำเสียงผู้พูดอ้างอิงมาเป็นแบบทดสอบ.....105

4.4 แสดงการจดจำเสียง “เดิน”, “วิ่ง”, “ยิ้ม”, “หาว” โดยใช้สัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักเปรียบเทียบเมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม B และนำเสียงที่ไม่ได้เป็นแบบอ้างอิงมาเป็นแบบทดสอบ.....106

4.5 แสดงความถูกต้องเมื่อใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักโดยใช้เสียงผู้ทำแบบอ้างอิงในกลุ่ม A และกลุ่ม B และใช้เสียงของผู้เป็นแแบบอ้างอิงมาเป็นแบบทดสอบ.....106

4.6 แสดงความถูกต้องเมื่อใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักโดยใช้เสียงผู้ทำแบบอ้างอิงในกลุ่ม A และกลุ่ม B และใช้เสียงของผู้ที่ไม่ได้นำเสียงมาเป็นแบบอ้างอิงมาเป็นแบบทดสอบ.....107

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

รูปที่ หน้า

2.1 แสดงบล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย.....10
2.2 แสดงการกระจายเฟรมของเสียงพูดแต่ละจุดแทนเฟรมของเสียง.....15
2.3 การรวมกลุ่มของเฟรมของเสียงเพื่อไปสร้างโค้ดบุค.....15
2.4 แสดงบล็อกไดอะแกรมของเวคเตอร์ควอนไตซ์เซชัน.....16
2.5 แสดงเวคเตอร์ควอนไตซ์เซอร์ที่ใช้ในระบบการรับรู้เสียงพูด.....21
2.6 ตัวอย่างไฟล์เวฟขนาดเล็ก.....23
2.7 แสดงลักษณะการจัดบิตกรณี sample point 12 บิต ถูกจัดเป็นแบบ 16 บิต.....24
2.8 แสดงการจัดเก็บ LSB.....24
2.9 แสดงรูปแบบของ sample frame ของ waveform แบบ 2 แชนแนล.....24
2.10 แสดงลำดับการจัดเก็บ sample point ของแต่ละแชนแนลภายใน sample frame.....25
3.1 แสดงบล็อกไดอะแกรมการจดจำเสียงพูดโดยทั่วไป.....31
3.2 แสดงขั้นตอนการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง
โดยวิธีการหาพลังงานของสัญญาณ.....34
3.3 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณ โดย
การเปรียบเทียบค่าพลังงาน..... 37
3.4 แสดงขั้นตอนการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง
โดยวิธีการเปรียบเทียบอัตราการผ่านศูนย์.....39
3.5 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง
โดยวิธีการหาอัตราการตัดศูนย์.....42
3.6 แสดงผังงานของฟังก์ชันสำหรับการพิจารณาการตัดกับแกนเวลาของรูปคลื่นสัญญาณ....43
3.7 แสดงขั้นตอนการนอร์มอลไลซ์สัญญาณ.....44
3.8 แสดงผังงานการนอร์มอลไลซ์สัญญาณ.....45
3.9.ก แสดงสัญญาณเสียงพูดเดิม.....46
3.9.ข แสดงการประมาณค่าในช่วงเชิงเส้นสัญญาณนอร์มอลไลซ์จุด..... 47
3.9.ค แสดงสัญญาณนอร์มอลไลซ์..... 47
3.10 แสดงการแบ่งช่วงที่ใช้ในการวิเคราะห์.....49
3.11 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์.....50
3.12 แสดงขั้นตอนการเตรียมสัญญาณในการวิเคราะห์.....53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.13 แสดงผังงานของโปรแกรม LPC.....54
3.14 แสดงบล็อกไดอะแกรมของเวคเตอร์ควอนไตซ์เซชัน.....55
3.15 Flow Chart แสดงการทำงานของโปรแกรมส่วนโค้ดบุค.....57
3.16 Flow Chart แสดงการทำงานของโปรแกรมส่วน Compair.....58
4.1 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน.....61
4.1 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 62
4.2 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง 63
4.2 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 64
4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม.....64
4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 65
4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 66
4.4 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว.....66
4.4 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 67
4.5 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า เดิน.....68
4.6 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า เดิน กับ ค่าสัมประสิทธิ์ เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง.....68
4.7 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง.....69
4.8 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า วิ่ง กับ ค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของแบบอ้างอิง.....69
4.9 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม.....70
4.10 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า ยิ้ม กับ ค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของแบบอ้างอิง.....70
4.11 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า หาว 71
4.12 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า หาว กับ ค่าสัมประสิทธิ์เซปสตรัม ถ่วงน้ำหนักของแบบอ้างอิง.....71
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน.....72
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 73
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 74

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 75
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 76
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 77
4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ) 78
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง 78
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 79
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 80
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 81
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 82
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 83
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 84
4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ) 85
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม.....85
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 86
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 87
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 88
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 89
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 90
4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ) 91
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว.....92
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 93
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 94
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 95
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 96
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 97
4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ) 98
4.17 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า เดิน.....98
4.18 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า เดิน กับ ค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของแบบอ้างอิง.....99
4.19 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง.....99
4.20 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า วิ่ง กับ ค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของแบบอ้างอิง.....100
4.21 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม.....100
4.22 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า ยิ้ม กับ ค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของแบบอ้างอิง.....101
4.23 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า หาว.....101
4.24 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า หาว กับ ค่าสัมประสิทธิ์เซปสตรัม
ถ่วงน้ำหนักของแบบอ้างอิง.....102
4.25 หน้าจอโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์....108
4.26 แสดงส่วนหน้าจอของโปรแกรมที่ใช้ในการบันทึกเสียง.....109
4.27 แสดงส่วนหน้าจอของโปรแกรมที่ใช้แสดงผลภาพกราฟฟิก.....109
4.28 แสดงส่วนหน้าจอของโปรแกรมที่ใช้ในการฟังเสียงที่ทำการบันทึกเสียงไว้และ
แสดงผลข้อมูลของเสียงนั้น.....110
4.29 แสดงหน้าจอของโปรแกรมที่ใช้ในการแสดงผลข้อมูลเสียง.....110
4.30 แสดงหน้าจอของส่วนให้ความช่วยเหลือในการใช้โปรแกรม.....111
4.31 แสดงส่วนการเรียกโปรแกรมที่ใช้ในการสร้างฐานเสียง.....112
4.32 แสดงโปรแกรมที่ใช้สร้างฐานเสียง.....112
4.33 แสดงโปรแกรมที่ใช้เก็บข้อมูลเสียงเพื่อนำมาสร้างฐานเสียง.....112
4.34 แสดงปุ่มปิดโปรแกรม.....113

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางสัญลักษณ์

สัญลักษณ์	คำอธิบาย	หน่วย
e(n)	ความคลาดเคลื่อนระหว่างสัญญาณเดิมกับสัญญาณที่ได้จากการประมาณลำดับที่	เมตร
En	ค่าผลรวมของกำลังสองของความคลาดเคลื่อนลำดับที่ n	ตารางเมตร
Energy	ค่าพลังงานของเสียงของแต่ละเฟรม	เมตร
G	อัตราการขยายสัญญาณ	-
L	ลำดับของเฟรมข้อมูลเสียง	-
N	ลำดับของข้อมูลเสียง	-
N	จำนวนข้อมูลเสียงในแต่ละเฟรม	ตัวอย่างสัญญาณ
P	ลำดับของการประมาณเชิงเส้น	-
S(n)	สัญญาณเสียงเดิมที่มาจากต้นกำเนิดเสียง	-
S´(n)	สัญญาณเสียงที่ได้จากการประมาณ	-
Sign(S(n))	ฟังก์ชันที่ใช้ทำการพิจารณาการตัดกับแกนเวลาของสัญญาณ	-
w(n)	ฟังก์ชันวินโดว์	-
Z	อัตราการตัดศูนย์ หน่วยเป็นเฮิร์ต	เฮิร์ต
α_k	ค่าคงที่อันดับที่ k ที่ใช้สำหรับประมาณสัญญาณ	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ที่มาของปัญหาพิเศษ

ปัจจุบันได้มีการใช้คอมพิวเตอร์อย่างแพร่หลายมากขึ้น โดยปกติการติดต่อสื่อสารระหว่างคอมพิวเตอร์กับมนุษย์จะทำโดย เครื่องคอมพิวเตอร์จะรับคำสั่งเข้าทางอุปกรณ์อินพุทซึ่งได้แก่ คีย์บอร์ด, เมาส์ , คันโยก(joy stick) เป็นต้น จากนั้นเครื่องคอมพิวเตอร์จะทำการประมวลผลและแสดงผลออกมาทางจอภาพหรือเครื่องพิมพ์ การที่จะเพิ่มความสามารถของมนุษย์ในการติดต่อสื่อสารกับเครื่องคอมพิวเตอร์ จึงเป็นที่มาของการพัฒนาการติดต่อกับเครื่องคอมพิวเตอร์โดยใช้เสียงพูด ในการนี้จะต้องศึกษาครอบคลุมถึงลักษณะของเสียงพูดในหลายๆ ด้าน และศึกษาการวิเคราะห์เสียงพูดโดยใช้กระบวนการคำนวณทางคณิตศาสตร์เข้าช่วย

1.2 วัตถุประสงค์ของปัญหาพิเศษ

1.2.1 เพื่อพัฒนาระบบให้สามารถวิเคราะห์ และรู้จำเสียงพูดทางด้าน Word based speech recognition

1.2.2 เพื่อหา อัลกอริทึมที่เหมาะสม ในการเพิ่มอัตราความถูกต้องของการรู้จำ

1.2.3 เพื่อเป็นแนวทางในการศึกษาและพัฒนาแอพพลิเคชันซอฟแวร์ที่ประมวลผล โดยการรับคำสั่งอินพุทที่เป็นเสียงจากไมโครโฟน

1.3 ขอบเขตปัญหาพิเศษ

ปัญหาพิเศษนี้จะสร้างภาพกราฟฟิกขึ้นมา จากนั้นผู้ใช้คอมพิวเตอร์สามารถใช้เสียงสั่งให้ภาพกราฟฟิกเคลื่อนไหวโดยคอมพิวเตอร์จะทำการประมวลผลตามคำสั่งที่ผู้ใช้คอมพิวเตอร์พูดโดยคำสั่งที่ใช้จะมีลักษณะเป็นคำสั้นๆ เช่น เดิน วิ่ง ยิ้ม หาว เป็นต้น คำสั่งที่ใช้สั่งงานในปัญหาพิเศษนี้จะมีจำนวนอย่างน้อย 4 คำสั่ง เมื่อผู้ใช้คอมพิวเตอร์(User) ทำการพูดคำสั่งใด คำสั่งนั้นก็จะถูกนำไปวิเคราะห์และแสดงผลเป็นภาพเคลื่อนไหวของภาพกราฟฟิกที่สร้างขึ้น โดยภาพจะเคลื่อนไหวสัมพันธ์กับคำสั่งที่ผู้ใช้คอมพิวเตอร์พูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ระยะเวลาในการดำเนินงานปัญหาพิเศษ

ระยะเวลาในการดำเนินงานปัญหาพิเศษ มีรายละเอียด ดังนี้

วัน/เดือน/ปี	การทำงาน
1 มิ.ย. - 25 มิ.ย. 42	ศึกษาปัญหาและที่มาของหัวข้อปัญหาพิเศษ
26 มิ.ย. - 5 ก.ค. 42	รวบรวมข้อมูลในเรื่องของการรู้จำเสียง
6 ก.ค. - 25 ก.ค. 42	ศึกษาทฤษฎีการสุ่มสัญญาณ ทฤษฎีแอลพีซี ศึกษาทฤษฎีการแปลงสัญญาณอนาลอกเป็นดิจิตอล
26 ก.ค. - 20 ก.ย. 42	ศึกษาทฤษฎีการวิเคราะห์เสียง
21 ก.ย. - 4 ต.ค. 42	ศึกษาเครื่องมือ(Tool)ต่าง ๆ ที่ต้องใช้
5 ต.ค. - 20 ต.ค. 42	จัดทำเอกสารประกอบโครงงานพิเศษ
21 ต.ค. - 10 ม.ค. 43	ออกแบบโปรแกรมและทำการเขียนโปรแกรม
11 ม.ค. - 25 ม.ค. 43	ออกแบบและสร้างภาพกราฟฟิก
26 ม.ค. - 10 มี.ค. 43	ทดสอบและปรับปรุงโปรแกรม
11 มี.ค. - 20 มี.ค. 43	จัดทำเอกสารประกอบโครงงานพิเศษ

ตารางที่ 1.1 ระยะเวลาในการดำเนินงานปัญหาพิเศษ

1.5 ขั้นตอนการดำเนินงาน

1.5.1 ศึกษาระบบงานของการรู้จำเสียง เพื่อเป็นแนวทางในการศึกษาทฤษฎีต่างๆ ที่เกี่ยวข้องกับการรู้จำเสียงต่อไป

1.5.2 ศึกษาทฤษฎีต่างๆที่เกี่ยวข้องกับการรู้จำเสียง เช่น ทฤษฎีการสุ่ม สัญญาณ การประมาณเชิงเส้น การแปลงสัญญาณจากสัญญาณอนาลอกเป็นสัญญาณดิจิตอล เป็นต้น ซึ่งในขั้นตอนนี้จะศึกษาถึงวิธีการในการวิเคราะห์เสียงเพื่อใช้ในการเปรียบเทียบเสียงด้วย

1.5.3 ศึกษาการทำงานของการ์ดเสียง โดยนำเอาทฤษฎีที่กล่าวมาข้างต้นไปทำ การเปรียบเทียบกับการทำงานในแต่ละขั้นตอนของการ์ดเสียง เพื่อให้เกิดความเข้าใจและทราบถึงกระบวนการเกิดข้อมูลทางเสียงซึ่งเป็นข้อมูลดิจิตอล เพื่อจะนำไปใช้งานในปัญหาพิเศษนี้ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5.4 ศึกษาโครงสร้างรูปแบบของเวฟไฟล์ และ พลัซ์ โค้ด มอดดูเลชั่นฟอร์แมต(Pulse Code Modulation Format) เพื่อจะนำไปใช้ประโยชน์ในการพัฒนาโปรแกรมในการเขียนโปรแกรม Voice Recognition

1.5.5 ศึกษาภาษาทางคอมพิวเตอร์ที่ใช้ในการเขียนโปรแกรม ปัจจุบันภาษาทางคอมพิวเตอร์มีอยู่เป็นจำนวนมากและแต่ละภาษาล้วนมีข้อดีข้อเสียแตกต่างกันออกไป จึงจำเป็นต้องศึกษาเพื่อให้เหมาะกับระบบงานของปัญหาพิเศษนี้

1.5.6 ศึกษาเครื่องมือต่างๆที่จะต้องใช้ ในขั้นตอนนี้จะศึกษาเครื่องมือต่างๆที่มีอยู่ รวมถึงข้อดีและข้อเสียของเครื่องมือแต่ละชนิดด้วย เพื่อนำไปใช้ในระบบงานไม่ว่าจะเป็นโปรแกรมทาง มัลติมีเดียซึ่งเป็นเครื่องมือที่สามารถนำมาใช้ในการพัฒนา ซอฟต์แวร์ ทางการรู้จำเสียง โปรแกรมทาง 3D Studio Max , Poser 3D ซึ่งจะใช้ในการสร้างภาพกราฟฟิก

1.5.7 ออกแบบระบบการทำงาน

1.5.8 ทำการเขียนโปรแกรมซึ่งจะใช้ภาษาที่ได้เลือกไว้แล้วจากขั้นตอนในการศึกษาภาษาทางคอมพิวเตอร์

1.5.9 ทดสอบระบบงานจริง และปรับปรุงระบบงานเป็นการทดลองใช้กับระบบงานจริง เพื่อดูความเหมาะสม และทำการปรับปรุงเพื่อให้ตรงกับความต้องการ

1.5.10 จัดทำเอกสารการวิจัยสรุปผลงานทดสอบ ทำการสร้างเอกสารประกอบ และเอกสารอ้างอิงในการทำปัญหาพิเศษ

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 เพื่อเป็นแนวทางในการพัฒนาโปรแกรมทางด้านการรู้จำเสียง

1.6.2 พัฒนาการเขียนโปรแกรมทางมัลติมีเดีย

1.7 ข้อตกลงเบื้องต้น

1.7.1 ระบบรู้จำเสียงที่สร้างขึ้นนั้น เจาะจงคุณสมบัติของผู้ใช้งาน คือ เพศหญิงและเพศชายที่มีอายุระหว่าง 20 – 22 ปี

1.7.2 ระบบรู้จำเสียงที่พัฒนาขึ้นนั้น จะสามารถทำงานได้ดีสำหรับผู้ใช้ที่ทำการฝึกฝนการออกเสียงคำสั่งแต่สะคำสั่งเพื่อใช้งานโปรแกรมเท่านั้น

1.7.3 ระบบรู้จำเสียงที่พัฒนาขึ้น ไม่สามารถทำงานในภาวะที่มีเสียงรบกวนจากสภาพแวดล้อมได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7.4 โปรแกรมมีการป้องกันความผิดพลาดที่อาจเกิดขึ้นได้ เนื่องจากความไม่เพียงพอของทรัพยากรคือ ความผิดพลาดที่อาจเกิดขึ้นเนื่องจาก ความไม่เพียงพอของหน่วยความจำเท่านั้น

1.8 อุปกรณ์ที่ใช้ในการทำปัญหาพิเศษ

1.8.1 เครื่องคอมพิวเตอร์ Pentium II 350 MHz

1.8.2 หน่วยความจำ 64 MB

1.8.3 ฮาร์ดดิสก์ ความจุ 4.3 GB

1.8.4 การ์ดเสียง YAMAHA OPL3-SAx WDM

1.8.5 ไดร์ฟซีดีรอม 36 x

1.8.6 ไมโครโฟน

1.8.7 ลำโพง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1 ลักษณะของเสียงพูด

คนเราเปล่งเสียงพูดด้วยอวัยวะที่ใช้ในการออกเสียง (organs of speech) ทำตามเสียงที่มีในระบบภาษาของตน แม้ว่าคนที่อยู่ในสังคมเดียวกันจะใช้ภาษาเดียวกันแต่ถ้าพิจารณาเสียงที่เปล่งออกมาจริงๆแล้วแต่ละครั้งก็อาจจะส่งเกตลักษณะที่แตกต่างกันได้ เราจึงสามารถจำเสียง จำวิธีพูดของคนที่เราคุ้นเคยได้ เสียงพูดที่จะอธิบายด้วยหลักเกณฑ์ทางวิทยาศาสตร์แม้ว่าในภาษาหนึ่งๆ จะมีเสียงต่างกันมากบ้างน้อยบ้าง แต่ละเสียงก็สามารถนำมาพิจารณาและอธิบายให้รู้ลักษณะการออกเสียงและตำแหน่งที่เกิดเสียงได้คำอธิบายนี้จะทำให้เข้าใจลักษณะเสียงทุกเสียง วิชาที่ว่าด้วยเสียงพูดเรียกว่า วิชาสัทศาสตร์ (Phonetics)

ในการศึกษาเสียงพูดแบ่งได้เป็น 2 ลักษณะคือ

ก. สรีรสัทศาสตร์ (Articulatory) เป็นการศึกษาเสียงพูดจากอวัยวะและการเคลื่อนไหวอวัยวะที่ทำให้เกิดเสียงพูด การอธิบายนี้ก็จะอธิบายโดยอาศัยลักษณะและอาการเคลื่อนไหวของอวัยวะที่เกี่ยวข้องในการเปล่งเสียงพูดนั้น

ข. กลสัทศาสตร์ (Acoustic Phonetics) เป็นการศึกษาเสียงพูดจากลักษณะคลื่นเสียงที่ผู้พูดเปล่งออกมาแล้ว และผู้ฟังได้ยินว่ามีลักษณะทางกลศาสตร์อย่างไร การศึกษาตามแนวนี้ต้องอาศัยความรู้ทางฟิสิกส์และคณิตศาสตร์ช่วยอธิบายลักษณะของคลื่นเสียง

2.1.1 อวัยวะที่ใช้ในการออกเสียง

อวัยวะที่ใช้ในการออกเสียงมีอยู่หลายส่วน แต่ละส่วนสามารถทำให้เสียงพูดแตกต่างกันไปได้ อวัยวะเหล่านี้มีปากและส่วนต่างๆ ในปาก ช่องคอ กล่องเสียง ช่องว่างในปากและช่องว่างในจมูก อวัยวะที่ใช้ในการเปล่งเสียงพูดแบ่งออกเป็น 2 ประเภทคือ

ก. อวัยวะที่ใช้ในการทำอาการ (Articulator) คืออวัยวะที่เคลื่อนไหวผลักลมไปยังส่วนต่างๆ อวัยวะที่สำคัญคือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

ข. อวัยวะซึ่งเป็นตำแหน่งที่เกิดเสียงต่างๆ (Point Of Articulator) คือตำแหน่งที่เกิดเสียงต่างๆ เช่น ริมฝีปาก ฟัน เพดาน ส่วนต่างๆเป็นต้น

อวัยวะส่วนที่มีหน้าที่ในการออกเสียงโดยตรงมีดังนี้

2.1.1.1 ริมฝีปาก เป็นอวัยวะส่วนที่เคลื่อนไหวได้มากและทำให้เสียงแตกต่างกันได้มากเราอาจบังคับให้ริมฝีปากอยู่ชิดกัน ห่างกัน ยื่นออกมา เป็นต้น ก็ได้ ลักษณะริมฝีปากแบบต่างๆนี้ล้วนมีอิทธิพลต่อการออกเสียงและทำให้เสียงแตกต่างกันไปทั้งสิ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1.2 ฟันเป็นอวัยวะที่ทำให้เกิดเสียงหลายชนิด เช่นเมื่อฟันกดลงบนริมฝีปากล่างลมที่ผ่านออกมาโดยแรงจะลอดช่องออกมาซึ่งทำให้เกิดเสียงได้

2.1.1.3 ปุ่มเหงือก เป็นส่วนนูนออกมาอยู่หลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่าเป็นคลื่น ปุ่มเหงือกเป็นบริเวณที่ทำให้เกิดเสียงปุ่มเหงือก

2.1.1.4 เพดานแข็งหรือเพดานอ่อน คือส่วนที่เป็นกระดูกแข็ง

2.1.1.5 เพดานอ่อน คือ ส่วนเพดานที่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับ ขึ้นลงได้เวลาหายใจ เพดานและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนลดระดับลงมา เปิดช่องว่างให้ลมออกทางจมูกเวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดหลังคอนอกจากเวลาออกเสียงนาสิกเท่านั้น

2.1.1.6 ลิ้นไก่ เป็นก้อนเนื้อเล็กๆ อยู่ปลายเพดานอ่อนตรงกลางปาก ลิ้นรัวได้

2.1.1.7 ลิ้น ลิ้นเป็นส่วนที่เคลื่อนที่มากที่สุดในการออกเสียง ซึ่งสามารถแบ่งออกได้เป็น 3 ส่วนคือ

1.ปลายลิ้น คือส่วนของปลายลิ้นซึ่งสามารถยกขึ้นไปแตะกับอวัยวะส่วนต่างๆ ได้

2.หน้าลิ้น คือส่วนที่อยู่ตรงข้ามกับเพดานแข็งถ้าวางลิ้นราบกับปากเช่นเดียวกับเวลาที่ไม่ได้พูด

3. ลิ้น ถ้าวางลิ้นทาบกับปาก ลิ้นส่วนนี้จะอยู่ตรงข้ามกับเพดานอ่อน

2.1.1.8 แผ่นเนื้อปากหลอดลม เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ตรงโคนลิ้นลงไปในคอ มีหน้าที่ปิดช่องลมในขณะรับประทานอาหารและเปิดช่องลมเพื่อพูด

2.1.1.9 ช่องคอ อยู่ถัดจากช่องปากไปจนถึงเส้นเสียง

2.1.1.10 เส้นเสียง เป็นอวัยวะสำคัญที่ทำให้เกิดเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ แผ่นภายในกล่องเสียง ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างซึ่งเป็นช่องผ่านให้ลมไปถึงปอดและออกมาจากปอดได้ ช่องนี้เรียกว่าช่องว่างระหว่างเส้นเสียง

2.1.1.11 ช่องจมูก หมายถึงโพรงในช่องจมูกซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก

เสียงที่เกิดขึ้นนั้นไม่ว่าจะเป็นเสียงประเภทใด จะผ่านไปตามทางเดินของเสียง โดยจะเริ่มตั้งแต่ช่องว่างระหว่างเส้นเสียงถึงริมฝีปาก ในทางวิทยาศาสตร์สรุปได้ว่าทางเดินของเสียงคือท่อนำเสียงที่มีรูปร่างไม่แน่นอน

2.1.2 การเกิดเสียง (Speech Production) การเกิดของเสียงแบ่งออกเป็น 3 ขั้นตอนคือ

ขั้นตอนที่ 1 จุดเริ่มต้น เป็นขั้นตอนที่ลมเริ่มถูกขับออกจากปอด ผ่านเข้าไปสู่ตอนที่ 2

ขั้นตอนที่ 2 การดัดแปลงลมที่เส้นเสียง อวัยวะที่ใช้ในขั้นตอนนี้คือส่วนที่ต่อจากปอดขึ้นมาถึงกล่องเสียงและที่กล่องเสียงเส้นเสียงจะทำหน้าที่เป็นลิ้นปิดเปิดทำให้เกิดเสียง 2 ชนิดคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. เสียงก้อง (Voiced) เกิดจากเส้นเสียงปิดกั้นลมไว้ ลมที่ผ่านออกมาจะเพิ่มแรงดันมากขึ้นจนเส้นเสียงปิดเปิดสลับกันไป ทำให้เกิดเสียงก้องขึ้นมา ซึ่งสามารถเรียกความถี่ในการปิดเปิดของเส้นเสียงว่า ความถี่มูลฐาน

2. เสียงไม่ก้อง (Unvoiced) เสียงชนิดนี้เส้นเสียงจะเปิดตลอดเวลา ลมจึงผ่านออกมาได้สะดวกทำให้เกิดเสียงไม่ก้องขึ้น

ขั้นตอนที่ 3 การเปลี่ยนแปลงลักษณะเส้นเสียง อวัยวะที่ใช้คือส่วนที่ต่อจากกล่องเสียงจนถึงริมฝีปากโดยลมที่ผ่านออกจากกล่องเสียงจะทำให้เกิดเสียงในลักษณะต่างๆ ซึ่งจะเกิดจากการเปลี่ยนแปลงของอวัยวะที่กล่าวไว้ในหัวข้อที่แล้ว

2.2 การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้นๆ

เนื่องจากว่าสัญญาณเสียงเป็นสัญญาณที่แปรตามเวลา มีการแปรเปลี่ยนที่ไม่แน่นอน เช่น ในขณะที่พูดช้าๆ รูปร่างของโพรงเสียงรวมทั้งลักษณะรูปแบบของการกระตุ้นอาจจะไม่มีการเปลี่ยนแปลงในช่วงเวลาที่นานที่สุดประมาณ 200 มิลลิวินาที แต่ในขณะที่พูดอย่างรวดเร็ว อาจจะมีช่วงที่ไม่มีการเปลี่ยนแปลงสั้นมากคือประมาณ 80 มิลลิวินาที ก็ได้

ดังนั้นเทคนิคในการวิเคราะห์เสียงพูดส่วนใหญ่แล้ว จะสมมติให้สัญญาณเสียงมีคุณสมบัติที่เปลี่ยนแปลงสัมพันธ์กับเวลาอย่างเชื่องช้า นั่นก็คือเราต้องแบ่งทำการวิเคราะห์หาพารามิเตอร์ของสัญญาณเสียงพูดในช่วงเวลาสั้นๆ เหมือนมองผ่านช่องแคบๆ ที่เรียกว่า ช็อตไทม์วินโดว์ (short-time window) เมื่อเทียบตามเวลาที่เสียงจะมีการเปลี่ยนแปลงได้ เพื่อจะได้มองเห็นเหมือนกับว่า เราหาพารามิเตอร์นั้นๆได้มาจากสัญญาณเสียงที่อยู่ภายในช่องแคบๆ และมีความเสถียรในช่วงเวลาสั้นๆ

เทคนิคส่วนใหญ่จะกำหนดให้พารามิเตอร์ได้มาจากค่าเฉลี่ยจากการวิเคราะห์ในช่องเวลาแคบๆนั้น สำหรับกรณีที่ต้องพิจารณาพารามิเตอร์ต่างๆที่มีการเปลี่ยนแปลง ก็จะทำการแบ่งสัญญาณเสียงออกเป็นช่องหลายๆ ช่องหรืออาจจะเรียก กรอบการวิเคราะห์ (analysis frame) ดังนั้นพารามิเตอร์ต่างๆ จะสามารถหาได้ทันเพียงพอที่จะติดตามการเปลี่ยนแปลงของสัญญาณ สำหรับในช่วงที่สัญญาณมีการเปลี่ยนแปลงช้า อาจจะกำหนดให้ช่องแคบมีขนาดใหญ่ประมาณ 100 มิลลิวินาที แต่ในทางตรงข้ามถ้าสัญญาณมีการเปลี่ยนแปลงเร็ว ก็ต้องใช้ช่องแคบที่มีขนาดเล็กมากๆ ประมาณ 5-10 มิลลิวินาที เพื่อป้องกันการสูญหายของรายละเอียดของสัญญาณถัดไป

2.2.1 รูปแบบของช่องแคบ (Windows)

การกำหนดขนาดของช่องแคบที่ใช้ขึ้นอยู่กับ

2.2.1.1 ช่องจะต้องสั้นพอที่จะทำให้คุณสมบัติของเสียงที่กำลังพิจารณาไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญในช่องแคบนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1.2 ช่องแคบจะต้องยาวพอที่จะทำให้การจัดเตรียมตัวอย่างของเสียงเพื่อจะนำตัวอย่างไปคำนวณหาพารามิเตอร์ให้ได้ตามต้องการเช่น ในกรณีที่มีสัญญาณรบกวนเข้ามาแทรกอยู่บางช่วงในสัญญาณเสียงด้วย ถ้าเราเลือกใช้ช่องแคบที่มีขนาดใหญ่กว่า เมื่อทำการหาค่าพารามิเตอร์โดยเฉลี่ย ก็จะทำให้ส่วนประกอบของสัญญาณรบกวนถูกตัดทิ้งหรือมองข้ามไป

2.2.1.3 ช่องแคบที่เหมาะสม ไม่ควรสั้นเกินกว่าช่วงหนึ่งคาบของสัญญาณเสียงในช่วงที่กำลังวิเคราะห์ เงื่อนไขนี้จะมีผลต่อค่าเฟรมเรท (frame rate) ซึ่งก็คือจำนวนครั้งต่อวินาทีที่ทำการวิเคราะห์สัญญาณเสียง โดยการขยับช่องแคบไปเป็นคาบๆ ตามแกนเวลา ตามปกติเฟรมเรทจะมีค่าประมาณ 2 เท่าของส่วนกลับของขนาดช่องแคบ นั่นก็คือช่องแคบถัดๆ กันไปจะมีการซ้อนทับกัน 50 เปอร์เซ็นต์

การนำฟังก์ชันของช่องแคบที่มีช่วงขนาดจำกัด w(n) มาคูณเข้ากับสัญญาณ s(n) จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของช่องแคบ รูปแบบของช่องแคบที่ง่ายที่สุดคือ กรอบแบบสี่เหลี่ยม (rectangular window) ซึ่งมีนิยามดังนี้

$$w(n) = \begin{cases} 1, & n=0,1,\ldots,N-1 \\ 0, & n \text{ อื่นๆ} \end{cases} \qquad (2.2.1)$$

ในสมการนี้ คือการกำหนดช่องของการวิเคราะห์ให้มีจำนวนตัวอย่าง N ตัวอย่าง

รูปแบบของช่องแคบในลักษณะนี้ ก็มีฟังก์ชันของช่องแคบหลายลักษณะด้วยกันตัวอย่างเช่นBlackman ,Barlett ,Hamming,Hanning เป็นต้น โดยฟังก์ชันของช่องแคบที่นิยมใช้กันมากในวิเคราะห์สัญญาณเสียงก็คือ ฟังก์ชัน Hamming ซึ่งมีรูปร่างตามลักษณะของ (cosine pulse) มีนิยามดังนี้

$$w(n) = \begin{cases} 0.54-0.46\cos(2\pi n)/(N-1), & n=0,1,\ldots,N-1 \\ 0 & n \text{ อื่นๆ} \end{cases} \qquad (2.2.2)$$

2.3 การหาค่าพลังงานของสัญญาณเสียง

พลังงานของสัญญาณเป็นตัวแทนอันหนึ่งที่เรามักจะนำมาใช้ในการวิเคราะห์ลักษณะต่างๆ ของสัญญาณทั่วๆไป โดยพลังงานของสัญญาณ s(n) ใดๆ ที่แปรตามเวลาสามารถนิยามได้ว่า

$$Energy = \sum_{n=-\infty}^{\infty} s^2 \qquad (2.3.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่สำหรับสัญญาณเสียงซึ่งเป็นสัญญาณที่แปรเปลี่ยนอยู่ตลอดเวลา ไม่มีเสถียรภาพตามเวลา เราจะต้องแบ่งสัญญาณออกมาเป็นช่วงเล็กๆ ตามแกนเวลา เรียกว่าแบ่งออกเป็นเฟรม เช่น เฟรมละประมาณ 10-30 วินาที หรือเฟรมละ 100 ตัวอย่าง เป็นต้นดังนั้นก็จะสามารถหาพลังงานของเสียงของแต่ละเฟรมได้เป็น

$$Energy_{l}(m) = \sum_{n=0}^{N-1} s^{2}(n) \qquad (2.3.2)$$

โดยที่ l แทนลำดับของเฟรมข้อมูลเสียง, $l = 0,1,2,3,\ldots,L$

N จำนวนข้อมูลเสียงในแต่ละเฟรม

การวัดค่าพลังงานดังในสมการที่ 2.3.2 นั้นมีข้อจำกัดตรงที่ว่ามันจะมีความไวต่อสัญญาณที่มีขนาดใหญ่ๆ เนื่องมาจากเราใช้วิธียกกำลังสองค่าของสัญญาณอินพุท ดังนั้นการแก้ปัญหาอย่างหนึ่งก็คือ วัดพลังงานของเสียงโดยใช้สมการดังนี้

$$Energy_{l}(m) = \sum_{n=0}^{N-1} |s(n)| \qquad (2.3.3)$$

2.4 การหาค่าอัตราการตัดศูนย์ (Zero Crossing)

การเกิดอัตราการตัดศูนย์จะเกิดขึ้นจากการที่รูปคลื่นของสัญญาณมีการตัดกับแกนเวลานั่นคือค่าของสัญญาณจะมีการเปลี่ยนสัญลักษณ์ทางคณิตศาสตร์ นั่นเอง อัตราการเกิดอัตราการตัดศูนย์เป็นเครื่องมืออย่างง่ายที่ใช้ในอธิบายการเปลี่ยนแปลงข้อมูลของสัญญาณค่าอัตราการตัดศูนย์นั้นสามารถนำมาใช้ในการตัดสินว่าสัญญาณนั้นเป็นเสียงก้อง (Voiced) หรือเสียงไม่ก้อง (Unvoiced) เนื่องจากเสียงก้องส่วนใหญ่จะมีค่าพลังงานอยู่ในช่วงความถี่ต่ำ ส่วนเสียงไม่ก้องจะมีพลังงานอยู่ในช่วงความถี่สูง และค่าอัตราการตัดศูนย์ก็มีความสัมพันธ์โดยตรงกับค่าความถี่ของสัญญาณ ดังนั้นจึงอาจสรุปได้ว่า สัญญาณเสียงที่มีค่าอัตราการตัดศูนย์สูงจะเป็นเสียงไม่ก้องและสัญญาณที่มีค่าอัตราการตัดศูนย์ต่ำจะเป็นเสียงก้อง แต่อย่างไรก็ตามการกำหนดขนาดของค่าอัตราการตัดศูนย์ที่แน่นอนเพื่อจำแนกชนิดของเสียงนั้น จะต้องอาศัยผลจากการทดลองเป็นหลัก

ในการหาค่าอัตราการตัดศูนย์สามารถทำได้โดยสมการดังนี้

$$z = \frac{1}{2N} * \sum_{n=1}^{N} |Sign\,(s(n)) - Sign\,(s(n-1))| \qquad (2.4.1)$$

$$\text{Sign}(s(n)) = \begin{cases} 1, \text{เมื่อ } s(n)>0 \\ 0, \text{ที่อื่นๆ} \end{cases}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

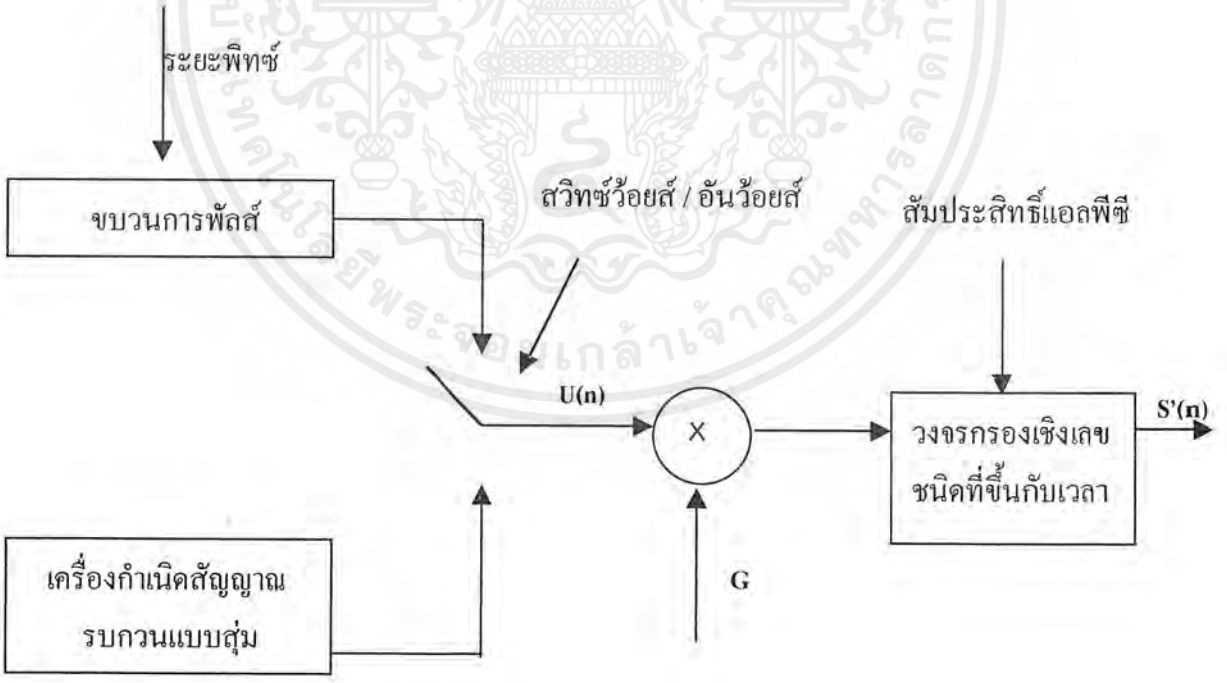
2.5 การประมาณเชิงเส้น (Linear Predictive Coding)

จากหลักการพื้นฐานของการประมาณเชิงเส้นคือ การประมาณค่าสัญญาณจากผลรวมเชิงเส้นของสัญญาณก่อนหน้านี้ สมมติว่าสัญญาณเดิม s(n) การประมาณค่าสัญญาณเป็น s'(n) ดังนั้นสามารถอธิบายการประมาณเชิงเส้นได้ด้วยสมการต่อไป

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \qquad (2.5.1)$$

เมื่อ α_k เป็นค่าคงที่ เรียกวิธีการนี้ว่าการประมาณเชิงเส้นอันดับ p โดยมีเงื่อนไขว่า ค่า α_k ที่ใช้ในการประมาณจะต้องทำให้ ผลรวมของกำลังสองของความคลาดเคลื่อน $\{s(n) - s'(n)\}^2$ มีค่าน้อยที่สุด นั่นคือ $\sum e^2(n) = \sum \{s(n) - s'(n)\}^2$ มีค่าต่ำสุด ซึ่งจะใช้การประมาณเชิงเส้นวิธีออโตคอร์รีเลชั่น หรือวิธีอัตตสัมพันธ์

จากหลักการพื้นฐานของการประมาณเชิงเส้นและแบบจำลองระบบสร้างสัญญาณเสียงเราสามารถเขียน บล็อกไดอะแกรมการทำการประมาณเชิงเส้น มาสร้างสัญญาณเสียงพูดได้ดังรูปที่ 2.1 (Rebiner and Levinson,1981)



รูปที่ 2.1 แสดงบล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.1 สามารถเขียนสมการได้เป็น

$$s(n) = G * u(n) + \sum_{k=1}^{p} a_k s(n-k) \tag{2.5.2}$$

การประมาณเชิงเส้นโดยใช้สัมประสิทธิ์ $\{\alpha_k\}$ คือ

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.5.3}$$

ดังนั้นความคลาดเคลื่อนคือ

$$e(n) = s(n) - s'(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.5.4}$$

ฟังก์ชันถ่ายโอนระหว่าง e(n) และ s(n) คือ

$$A(z) = E(z)/S(z) = 1\sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.5.5}$$

จากสมการ (2.5.2)-(2.5.4) จะเห็นได้ $\{\alpha_k\} = \{a_k\}$ แล้ว

$$e(n) = G * u(n) \tag{2.5.6}$$

ดังนั้น ค่าผลรวมของกำลังสองของความคลาดเคลื่อน

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

$$E_n = \sum_m [s(m) - s'(m)]^2 \tag{2.5.7}$$

โดยที่ n คือ ช่วงที่ n ของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า E_n ต่ำที่สุดจะต้องมีเงื่อนไขว่า

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการ (2.5.7)

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

เมื่อ i=1,2,3,...,p

$$= -2\left[\sum_m s(m)s(m-i) - \sum_{k=1}^{p} \sum_m \alpha_k s(m-k)s(m-i) \right] \qquad (2.5.8)$$

$$\frac{\partial E_n}{\partial \alpha_i} = 0 \qquad \text{ก็ต่อเมื่อ} \qquad (2.5.9)$$

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

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

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,j) = \phi_n(i,0) \qquad (2.5.10)$$

โดยสมการ (2.5.8)-(2.5.9) จะเห็นได้ว่า

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

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

$$E_n = \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \qquad (2.5.11)$$

สมมติว่าใน 1 เฟรม ของสัญญาณ ที่ตัดมาคำนวณมี N ตัวอย่าง คือ $s_n(0), s_n(1), s_n(2), \ldots, s_n(N-1)$ ในที่นี้เราให้ $s_n(m)=0$ เมื่อ $m<0$ หรือ $m>N-1$ เพราะฉะนั้น

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

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

ให้

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m)s_n(m+k) \qquad \text{เมื่อ } k=0,1,2,3,\ldots,p \qquad (2.5.12)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น จากสมการ (2.5.11) และ (2.5.12) จะได้ว่า

$$R_n(k) = \sum_{m=0}^{N-1-k} x'(m)x'(m+k) \qquad (2.5.13)$$

จากสมการที่ (2.5.12) จะได้ว่า

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \qquad \text{เมื่อ } i=1,2,3,\ldots,p \qquad (2.5.14)$$

จากสมการที่ 2.5.14 เขียนให้อยู่ในรูปของเมตริกซ์ได้เป็น

$$\begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & \cdots & R_n(p-2) \\ \cdots & \cdots & \cdots & \cdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \cdots \\ R_n(p) \end{bmatrix} \qquad (2.5.15)$$

หรือ $R_n.\alpha = r_n$

เมื่อ

$$R_n = \begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & \cdots & R_n(p-2) \\ \cdots & \cdots & \cdots & \cdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_p \end{bmatrix} \qquad (2.5.16)$$

และ

$$r_n = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \cdots \\ R_n(p) \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 เวคเตอร์ ควอนไตซ์เซชัน (Vector Quantization)

การรับรู้เสียงพูด เป็นลักษณะของการรับรู้รูปแบบ (Pattern Recognition) คือจะเป็นการเปรียบเทียบระหว่าง แบบทดสอบ (Test Pattern) กับ แบบอ้างอิง (Reference Pattern) ซึ่งเป็นรูปแบบที่เราทราบและเก็บไว้ล่วงหน้า

ขั้นตอนในการรับรู้แบ่งเป็น 2 ขั้นตอน ดังนี้

ก.ขั้นเรียนรู้ (Learning) จะเป็นการสร้างกลุ่มของแบบอ้างอิง ในการรับรู้เสียงพูด ในขั้นนี้จะทำการวิเคราะห์เสียงพูดก่อน แล้วเก็บลักษณะของเสียงในรูปของพารามิเตอร์ ดังได้กล่าวไว้แล้วในบทที่ 3 และใช้เวคเตอร์ ควอนไตซ์เซชัน ในการสร้างแบบอ้างอิง เพื่อใช้เปรียบเทียบในขั้นตอนต่อไป

ข.ขั้นรับรู้ (Recognition) จะเป็นการทดสอบการรับรู้ระหว่างแบบอ้างอิง กับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์แบบทดสอบ กับแบบอ้างอิงทั้งหมด แบบอ้างอิงที่เลือกคือ แบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

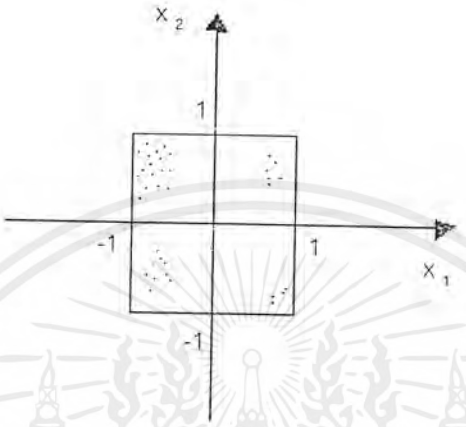
2.6.1 เวคเตอร์ ควอนไตซ์เซชัน

ในทฤษฎีของการสื่อสารนั้น แหล่งเข้ารหัสจะใช้เทคนิคการเปลี่ยนค่าสัญญาณ เอาท์พุท ไปเป็นรหัสแบบไบนารี ซึ่งเป้าหมายของเครื่องเข้ารหัสคือ การเป็นไปได้ที่จะมีการผิดเพี้ยนของข้อมูลน้อยสุด ด้วยอัตราความความเร็ว (bit rate) ที่กำหนดเวคเตอร์ ควอนไตซ์เซชัน เป็นเทคนิคเข้ารหัสเทคนิคหนึ่งที่มีคุณภาพ มันจะทำการเข้ารหัสเวคเตอร์พารามิเตอร์ของคลื่นที่แทนสเปกตรัมของเสียง ซึ่งเป็นเวคเตอร์อินพุทให้เป็นตัวเลขจำนวนเต็ม (index) โดยนำไปเปรียบเทียบกับโค้ดบุคที่ได้สร้างเก็บไว้ก่อนแล้ว และโค้ดบุคที่ถูกเลือกนั้นจะเป็นตัวที่ใกล้เคียงที่สุดกับเวคเตอร์อินพุท ผลที่ได้จากเวคเตอร์ควอนไตซ์เซชัน ที่แสดงออกมานั้นขึ้นกับเซตของโคดบุคที่สร้างเก็บไว้ก่อนแล้วนั่นเอง

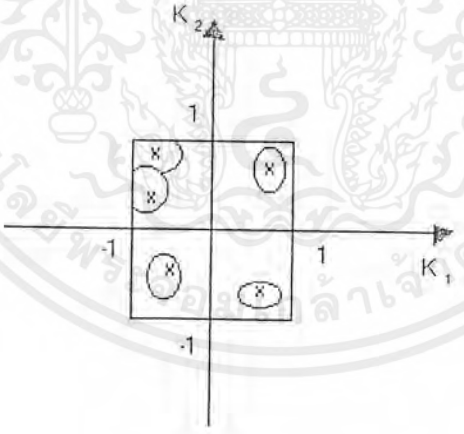
ถ้าเราเปรียบเทียบอัตราข้อมูลของการแทนเวคเตอร์สำหรับเสียงที่ยังไม่ได้เข้ารหัส เราจะเป็นว่าการวิเคราะห์เสียงนั้นต้องการการลดลงของอัตราข้อมูล พิจารณาดังนี้ เสียงพูดที่แซมเปิลด้วยความถี่ 10 kHz และมีขนาด 16 บิต จะต้องการอัตราข้อมูล 160,000 bps ในการจัดเก็บซึ่งเป็นรูปแบบที่ยังไม่ได้มีการบีบอัดข้อมูล สำหรับการวิเคราะห์เสียงพูด พิจารณาเวคเตอร์ที่มี 10 มิติ ใช้ 100 สเปกตรัมต่อวินาที ถ้าเราใช้ 16 บิตแทนแต่ละเวคเตอร์ อัตราที่ใช้ในการเก็บคือ 100*10*6 bps หรือ 16,000 bps สามารถลดไปได้ถึง 10 เท่าของสัญญาณที่ยังไม่ได้บีบอัด ยิ่งไปกว่านั้นวิธีการเวคเตอร์ควอนไตซ์เซชันยังสามารถลดอัตราข้อมูลลงไปได้อีก โดยการสร้างโค้ดบุค เพื่อใช้ในการชี้ว่าเวคเตอร์ควอนไตซ์เซชันยังสามารถลดอัตราข้อมูลลงไปได้อีก โดยการสร้างโค้ดบุค เพื่อใช้ในการชี้ว่าเวคเตอร์โค้ดบุคไหนที่เหมือนกับเวคเตอร์เสียงอินพุทที่เข้ามามากที่สุด สมมุติโค้ดบุคแทนด้วยตัวเลข 10 บิต และมีอัตราข้อมูล 100 สเปกตรัมต่อวินาที เราจะได้ว่าอัตราข้อมูล 1,000 bps ซึ่งสามารถลดอัตราข้อมูลลงไปได้ถึง 16 เท่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เวคเตอร์สเปกตรัมที่จะนำไปทำการควอนไตซ์เป็นเวคเตอร์ที่ได้จากการประมาณเชิงเส้น และผ่านการปรับปรุงให้ดีขึ้นโดยการทำเซปสตรัมและการเวทค่าพารามิเตอร์ ซึ่งจะแสดงโดย V ซึ่งแต่ละเวคเตอร์เป็นเวคเตอร์ที่มี p มิติ (p คือ order หลังจากผ่านการเวทค่าเรียบร้อยแล้ว)
เช่นในรูปตัวอย่างนี้ คือ p = 2 order = 2 ของ LPC Model



รูปที่ 2.2 แสดงการกระจายเฟรมของเสียงพูดแต่ละจุดแทนเฟรมของเสียง



รูปที่ 2.3 การรวมกลุ่มของเฟรมของเสียงเพื่อไปสร้างโค้ดบุค

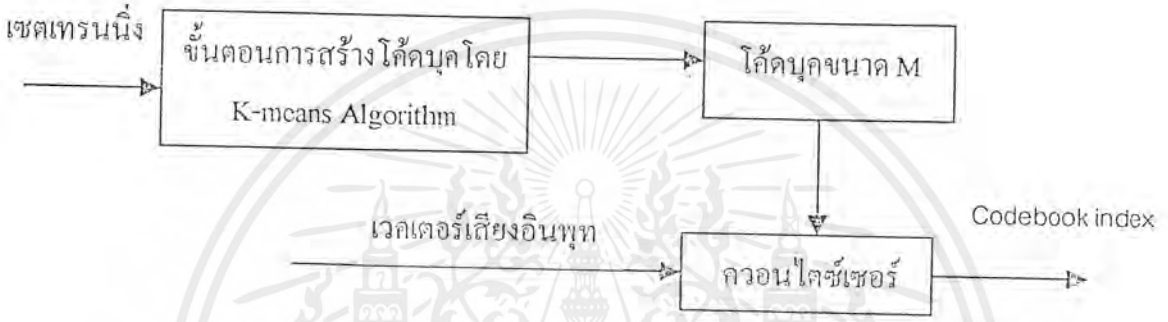
ซึ่ง x แทนเวคเตอร์ศูนย์กลางเพื่อให้เวคเตอร์ทั้งหมดในกลุ่มสามารถแบ่งแยกได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อดีของวิธีการเวคเตอร์ควอนไตซ์เซชัน คือ

1. ลดความจุในการเก็บข้อมูลสเปกตรัม ที่ได้จากการวิเคราะห์ลง
2. ลดการคำนวณในการเปรียบเทียบความเหมือนของเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ลงซึ่งเป็นการคำนวณหลักการที่ใช้ในการรับรู้เสียงพูด
3. กระบวนการการเลือกโค้ดบุคที่ดีที่สุดเพื่อแสดงแทนเวคเตอร์สเปกตรัมของเสียงพูดที่เข้ามาเป็นไปด้วยความรอบคอบ

2.6.2 ส่วนประกอบของเวคเตอร์ควอนไตซ์เซชัน



รูปที่ 2.4 แสดงบล็อกไดอะแกรมของเวคเตอร์ควอนไตซ์เซชัน

2.6.2.1 การทำงานของเวคเตอร์ควอนไตซ์เซชัน

1. เซตของเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ $v_1, v_2, v_3, \ldots, v_L$ ซึ่งเป็นเซตขนาดใหญ่จะถูกนำมาใช้ในการสร้างเซตของโค้ดบุค เพื่อใช้เป็นตัวแทนเวคเตอร์สเปกตรัมของเสียงที่สังเกตได้ ขนาดโค้ดบุคของเวคเตอร์ควอนไตซ์เซชันคือ $M = 2^B$ เวคเตอร์ (เราเรียกว่า B – bit โค้ดบุค) ซึ่งเราต้องการให้ $L>>M$ เพื่อที่จะหาเซตของ M โค้ดบุคที่ดีที่สุด ในทางปฏิบัตินั้นเพื่อให้เวคเตอร์ ควอนไตซ์เซชันโค้ดบุค สามารถทำงานได้ดีนั้น ขนาด L อย่างน้อยที่สุดที่เราสามารถใช้ได้ คือ 10M

2. การวัดความเหมือนหรือระยะทาง ระหว่างเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ 2 เวคเตอร์ ดังนั้นสามารถจัดกลุ่มเซตของเวคเตอร์เทรนนิ่ง เพื่อสร้างโค้ดบุค เราแสดงระยะทางสเปกตรัม $d(v_i, v_j)$ ระหว่าง 2 เวคเตอร์ v_i และ v_j คือ d_{ij}

3.ฟังก์ชันย่อยในการคำนวณหาจุดศูนย์กลางเพื่อแบ่งแยกเซตของเทรนนิ่งจำนวนLเวคเตอร์ไปเป็น M กลุ่ม และ M โค้ดบุค ที่ได้คือจุดศูนย์กลางของแต่ละกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

4. ฟังก์ชันย่อยในการแบ่งแยกสำหรับเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ ซึ่งจะเลือกโค้ดบุคที่ใกล้เคียงที่สุดให้กับเวคเตอร์อินพุท และผลลัพธ์ที่ได้คือ ดัชนีโค้ดบุคซึ่งเป็นตัวแทนของสเปกตรัม มักจะใช้กฎ Nearest Neighbour ในการพิจารณา ส่วนนี้เป็นหัวใจหลักของควอนไตซ์เซอร์ ซึ่งอินพุท คือ เวคเตอร์สเปกตรัมของเสียงและเอาท์พุท คือ ดัชนีโค้ดบุคที่มีความเหมือนกับเวคเตอร์เสียงอินพุทมากที่สุด

2.6.2.2 เซต เทรนนิ่ง (Training) ของเวคเตอร์ควอนไตซ์เซชัน

เซตเทรนนิ่งของเวคเตอร์ควอนไตซ์เซชันซึ่งเป็นเสียงพูดที่จะนำมาใช้ในการสร้างโค้ดบุค ซึ่งเซตเทรนนิ่งจะต้องมีขนาดใหญ่มาก ๆ ถึงจะสามารถสร้างโค้ดบุคที่ดีได้ เซตเทรนนิ่งจะแปรตามสภาวะต่าง ๆ ดังนี้

- ผู้พูด รวมทั้งช่วงของอายุ เพศ ความเร็ว ช้าในการพูด ระดับเสียง ฯลฯ
- สภาวะแวดล้อมขณะพูด เช่น ในห้องเงียบหรือในรถ
- ระบบการแปลงการส่งรวมทั้งไมโครโฟน หูโทรศัพท์
- หน่วยของเสียงรวมถึงคำศัพท์ต่าง ๆ และข้อความที่สนทนา

2.6.3 หลักการของเวคเตอร์ควอนไตซ์เซชัน

เวคเตอร์ควอนไตซ์เซชันเป็นวิธีการควอนไตซ์คลื่นเสียง ในการควอนไตซ์แบบเวคเตอร์มีวิธีการควอนไตซ์ดังนี้ คือ จะมีเวคเตอร์โค้ด หรือโค้ดบุคเก็บไว้ อินพุทที่เข้ามาจะถูกทำการเปรียบเทียบกับโค้ดบุคที่มีอยู่ โดยจะพิจารณาว่าอินพุทที่เข้ามานั้นห่างจากโค้ดบุคใดน้อยที่สุด อินพุทดังกล่าวจะถูกแทนด้วยเวคเตอร์โค้ดนั้น โค้ดบุค ที่ใช้จะเหมาะสมเพียงใด ขึ้นกับความคลาดเคลื่อนรวมทั้งหมดของต้นแบบที่ใช้หา เมื่อต้นแบบที่พิจารณามีจำนวนจำกัด โค้ดบุคที่มีความคลาดเคลื่อนน้อยที่สุด จะได้จากการทำซ้ำจนกว่าจะได้ค่าที่เหมาะสม โดยเหตุนี้จึงขึ้นกับค่าแรนดอมที่สุ่มขึ้นมาจำนวนเท่ากับขนาดของโค้ดบุคเพื่อใช้เป็นค่าเริ่มต้น ในวิธีการนี้ข้อมูลต้นแบบที่ต้องการหาถูกแยกไปอยู่ในแต่ละกลุ่มของค่าเริ่มต้น จะทำการหาค่าเฉลี่ยของแต่ละกลุ่มเพื่อเป็นค่ากลางของกลุ่ม ทำวิธีการดังกล่าวซ้ำจนกว่าความคลาดเคลื่อนรวมจะต่ำกว่าค่า ๆ หนึ่ง หรือการลดลงของความคลาดเคลื่อนรวมน้อยกว่าค่า ๆ หนึ่ง โดยค่ากลางดังกล่าวของกลุ่มจะถูกเก็บเป็นเวคเตอร์โค้ด แต่แน่นอนว่าค่าความคลาดเคลื่อนรวมจะลดลงทุกครั้งที่มีการคำนวณซ้ำใหม่ จึงขึ้นกับค่าที่กำหนดว่าต้องการให้ความคลาดเคลื่อนรวมน้อยเพียงใด

2.6.4 การคำนวณเวคเตอร์ควอนไตซ์เซชัน

ถ้าสมมุติ $x = [x_1, x_2, \ldots, x_p]$ มี p มิติ เราจะทำการหาควอนไตซ์ของเวคเตอร์ x กับโค้ดบุค เราสามารถเขียนว่า y เป็นควอนไตซ์ของค่า x

$$y = q(x) \qquad (2.6.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

36152

โดย q( ) เป็นโอเปอร์เรเตอร์ของควอนไตซ์ y ถูกเรียกว่าเอาท์พุทเวคเตอร์ของค่า x โดย y เป็นค่าใดค่าหนึ่งใน $Y = \{ y_i , 1 \leq i \leq M\}$ $y_i = [y_{i1} , y_{i2} , \dots , y_{ip}]$ Y เป็นเซ็ตของโค้ดบุค M เป็นขนาดชองโค้ดบุค และ $\{y_i\}$ เป็นเซ็ตของเวคเตอร์โค้ด y_i อาจเรียกว่าเป็นโค้ดอ้างอิง และ M อาจเรียกว่าจำนวนระดับขั้น จะทำการแบ่งเวคเตอร์ x ไปใน M เซล $\{C_i , 1 \leq i \leq M \}$ เมื่อ ป อยู่ในเซล c_i

$$q(x) = Y_i \quad \text{ถ้า } x \in C_i$$

ถ้า x ถูกควอนไตซ์ได้ค่า y ค่าความคลาดเคลื่อนจากการควอนไตซ์สามารถแสดงได้โดยระยะห่างของทั้งสอง d(x,y) โดยความคลาดเคลื่อนรวมคือ

$$D = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} d [x (n), y (n)] \qquad (2.6.2)$$

การคำนวณหาโค้ดบุคมีหลายวิธี แต่ที่จะนำมาใช้คือ Lloyd Algorithm (K-means Algorithm)

2.6.5 การวัดความคลาดเคลื่อน

การวัดความคลาดเคลื่อนเป็นส่วนจำเป็นและเป็นประโยชน์ต่อการออกแบบโค้ดบุค ซึ่งเราวัดโดยการวัดระยะทางซึ่งสมการพีชคณิตที่ใช้ในการหาระยะทางในปัจจุบันมีวิธีการหลายแบบ แต่ที่จะนำมาใช้ในคือ square error distortion

square error distortion :

ถ้าสัญญาณที่เข้ามามี P มิติ เราสามารถหาระยะห่างระหว่างสัญญาณเข้ากับเวคเตอร์โค้ด

$$d (v_1 , v_2) = \| v_1 - v_2 \|^2 = \sum_{i=0}^{k-1} (x_i - y_i)^2 \qquad (2.6.3)$$

ซึ่งเป็นวิธีการคำนวณที่ง่าย และรวดเร็วในการวัดความคลาดเคลื่อน จึงนิยมใช้วิธีนี้ แต่ก็ยังไม่เหมาะสมเพียงพอ ควรใช้ Itakura distance มากกว่า แต่เป็นวิธีที่มีความซับซ้อนมาก และยุ่งยากในการคำนวณ

2.6.6 การจัดกลุ่มเพื่อสร้างแบบอ้างอิงโดยใช้ K-means Algorithm

การสร้างกลุ่มของแบบอ้างอิงเป็นขั้นตอนในส่วนของการเรียนรู้ โดยเริ่มต้นจากการนำข้อมูลสัญญาณเสียงที่ได้จากส่วนของการวิเคราะห์มาทำการสร้างเป็นกลุ่มของแบบอ้างอิง ในระบบการรับรู้เสียงพูดแบบต่างบุคคล จะใช้แบบอ้างอิงของคำหนึ่ง ๆ จากผู้พูดจำนวนมาก เพื่อที่จะได้ครอบคลุมถึงความแปรปรวนต่าง ๆ ที่จะเกิดขึ้นระหว่างผู้พูดแต่ละคน แต่อย่างไรก็ตามผลที่เกิดขึ้นตามมา เมื่อจำเป็นจะต้องมีข้อมูลจำนวนมากในกลุ่มของแบบอ้างอิง คือ

ก. เวลาที่ใช้ในการตอบสนอง ในการรับรู้เสียงพูด ขั้นตอนการเปรียบเทียบแบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทดสอบกับแบบอ้างอิงทั้งหมดที่มีเป็นขั้นตอนที่สำคัญในการรับรู้ เมื่อแบบอ้างอิงมีจำนวนมาก การเปรียบเทียบจึงจำเป็นที่จะต้องอาศัยเวลาเพิ่มขึ้นด้วย

ข. เนื้อที่ในหน่วยความจำสำรองที่ใช้ในการเก็บแบบอ้างอิง เมื่อมีแบบอ้างอิงจำนวนมากที่จำเป็นต้องใช้ในการเปรียบเทียบ เนื้อที่ในหน่วยความจำสำรองก็จำเป็นต้องมีเพิ่มขึ้นด้วย

ค. ความถูกต้องในการรับรู้ เนื่องจากจำนวนแบบอ้างอิงในแต่ละคำไม่ได้เป็นสัดส่วนโดยตรงกับความถูกต้องในการรับรู้ กล่าวคือเมื่อเราเพิ่มจำนวนแบบอ้างอิงในแต่ละคำไปจนถึงระดับหนึ่ง ความถูกต้องในการรับรู้จะเริ่มคงที่ และในบางครั้งอาจจะมีค่าลดลงด้วย

จากเหตุผลดังกล่าว จึงทำให้มีความพยายามที่จะจัดกลุ่มของแบบอ้างอิงในแต่ละคำใหม่เพื่อให้ได้แบบอ้างอิงที่จำนวนพอเหมาะ และสามารถใช้เป็นตัวแทนของแบบอ้างอิงที่มีอยู่ทั้งหมดได้ อัลกอลิทึมที่ใช้ในการแบ่งกลุ่มมีอยู่เป็นจำนวนมาก สำหรับอัลกอลิทึมที่จะกล่าวถึงคือ K-means Algorithm ตามขั้นตอนโดยสรุป ดังนี้

2.6.6.1 การหาจุดศูนย์กลางของกลุ่ม (Center Cluster)

ในขั้นตอนนี้จะทำการกำหนดจำนวนกลุ่มของแบบอ้างอิงให้มีค่าเท่ากับ M และเริ่มต้นจากการแรนดอมค่ามาจำนวนเท่ากับขนาดของโค้ดบุค เพื่อเป็นค่าจุดศูนย์กลางของแต่ละกลุ่มก่อน จากนั้นจะแบ่งเวคเตอร์ต่าง ๆ เข้าไปตามกลุ่มต่าง ๆ โดยพิจารณาจากระยะทาง หรือความต่างกันน้อยที่สุดกับกลุ่มไหนก็จะให้อยู่กลุ่มนั้น จากนั้นก็จะทำการเฉลี่ยค่าของเวคเตอร์ที่อยู่ในกลุ่มเพื่อให้เป็นจุดศูนย์กลางใหม่

2.6.2.2 การจัดกลุ่มใหม่ (Reclassification)

แบบอ้างอิงที่เหลืออยู่ในแต่ละคำ จะถูกจัดให้เข้าไปในแต่ละกลุ่ม โดยมีเงื่อนไขคือระยะทางของแบบอ้างอิงนี้กับแบบอ้างอิงในกลุ่มจะมีค่าเฉลี่ยน้อยที่สุด หลังจากที่แบบอ้างอิงทุกแบบถูกจัดเข้ากลุ่มเรียบร้อยแล้ว จะทำการหาจุดศูนย์กลางของแต่ละกลุ่มใหม่ ขั้นตอนนี้จะทำซ้ำจนกว่าจุดศูนย์กลางของแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลง อย่างไรก็ตามจากขั้นตอนดังกล่าวจะเห็นว่า การกำหนดค่าจำนวนกลุ่มที่เหมาะสมจะทำได้ลำบาก ดังนั้นจึงได้มีการปรับปรุงเทคนิคนี้เป็นการกำหนดค่าระยะทางสูงสุดที่จะทำให้เกิดกลุ่มใหม่ขึ้น คือเมื่อแบบอ้างอิงใดที่ถูกเลือกให้เข้ากลุ่มด้วยระยะทางที่น้อยที่สุดแล้ว แต่ระยะทางนี้ยังคงมีค่ามากกว่าระดับที่กำหนด แบบอ้างอิงนี้จะถูกนำไปสร้างเป็นจุดศูนย์กลางของกลุ่มใหม่ทันที

2.6.7 การคำนวณจุดศูนย์กลางสำหรับการออกแบบเวคเตอร์ควอนไตซ์เซชันโค้ดบุค

การสร้างโค้ดบุคโดยใช้ K-means สามารถสรุปสั้น ๆ ได้ดังนี้

ก) ความเพี้ยนน้อยสุด สำหรับแต่ละเวคเตอร์อินพุท xt เมื่อเปรียบเทียบกับเวคเตอร์ yi ซึ่งเป็นโค้ดบุค C ที่เข้าไป กลุ่มของเวคเตอร์เทรนนิ่งจะเป็นไปตาม Code words

ข) การคำนวณหาจุดศูนย์กลาง แต่ละกลุ่มของเวคเตอร์ จะคำนวณหาจุดศูนย์กลางใหม่ ซึ่งค่าความผิดเพี้ยนเฉลี่ยต่ำ สำหรับสมาชิกกลุ่มที่ถูกแทนโดยเวคเตอร์จุดศูนย์กลาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำ 2 ขั้นตอนซ้ำจนกว่าจะเกิดการลู่เข้า (convergent) ขั้นแรกนั้นเป็นการทำโดยตรง คือนำเข้าไปเพื่อหาความผิดเพี้ยน ส่วนขั้นที่ 2 นั้นเป็นการหาจุดศูนย์กลาง ซึ่งเป็นขั้นตอนที่สำคัญพิจารณา $\{x_i\}^L_{i=1}$ และค่าความเพี้ยน d(x,y) โดยทั่วไป เราสมมุติว่าเวคเตอร์เล่านี้ถูกกำหนดไว้ในกลุ่มป้ายเดียวกัน (หรือ Code word) จุดศูนย์กลางของ $\{x_i\}^L_{i=1}$ ถูกจำกัดความเป็นเหมือนเวคเตอร์ $\overline{Y}$ ซึ่งเป็นค่าความเพี้ยนเฉลี่ยน้อยที่สุด

$$\overline{Y} = \underline{\Delta} \arg \quad \min \quad \frac{1}{L}\sum_{i=1}^{L} d\,(x_i, y) \tag{2.6.4}$$

เมื่อ x_i และ y เป็นเวคเตอร์โดยที่ $x_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ik})$ $(y_1, y_2, \ldots, y_k)$ วัดในพื้นที่ k มิติ กับระยะทางตามแบบยูคลิด จุดศูนย์กลางคือค่าเฉลี่ยของเซตหนึ่งของ เทรนนิ่ง

$$\overline{Y} = \frac{1}{L}\sum_{i=0}^{L} x_i \tag{2.6.5}$$

$\overline{Y}$ เป็นจุดศูนย์กลางซึ่งเป็นเวคเตอร์ที่อยู่ตรงกลางของ $\{x_i\}^L_{i=1}$ ซึ่งแต่ละมิติจะไม่ขึ้นแก่กันหมายความว่า แต่ละ y_k เป็นค่ากลางของ $\{x_i\}^L_{i=1}$

2.6.8 กฎการตัดสินใจ (Decision Rules)

ผลที่ได้จากการเปรียบเทียบระยะทางระหว่างแบบทดสอบกับแบบอ้างอิงแต่ละแบบ ระยะทางที่ได้ทั้งหมดจะนำมาผ่านขั้นตอนการตัดสินใจ เพื่อหาผลลัพธ์ของการรับรู้ กฎการตัดสินใจ (Decision Rules) ที่ใช้ในการรับรู้เสียงพูด มีดังนี้ คือ

2.6.8.1 กฎ Nearest Neighbor (NN)

ผลลัพธ์ของการรับรู้จากกฎการตัดสินใจนี้ ได้แก่ แบบอ้างอิงที่มีระยะทางจากแบบทดสอบน้อยที่สุด กฎนี้เหมาะสำหรับระบบการรู้เสียงพูดแบบบุคคลเดี่ยว

2.6.8.2 กฎ K- Nearest Neighbor (KNN)

เป็นกฎที่ดัดแปลงมาจากกฎ NN อีกทีหนึ่ง เพื่อให้เหมาะสมกับการรับรู้เสียงพูดแบบต่างบุคคล ซึ่งมีแบบอ้างอิงในแต่ละคำมากกว่า 1 แบบ กฎนี้แบ่งออกเป็น 2 แบบ คือ

ก. แบบที่ 1 จะทำการหาแบบอ้างอิงจำนวน K แบบ ที่มีระยะทางห่างจากแบบทดสอบน้อยที่สุด ผลลัพธ์ของการรับรู้จะได้จาก แบบอ้างอิงที่มีจำนวนมากที่สุดในจำนวน K แบบ ที่ได้รับการเลือกมา

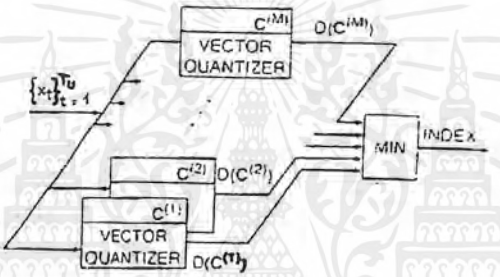
ข. แบบที่ 2 จะทำการหาแบบอ้างอิงจำนวน K แบบ ในแต่ละคำ โดยที่ทั้ง K แบบมีระยะทางห่างจากแบบทดสอบน้อยที่สุด ผลลัพธ์ของการเรียนรู้จะได้จาก แบบอ้างอิงที่ค่าเฉลี่ยของระยะทางในจำนวน K แบบ ที่ได้รับการเลือกมาน้อยที่สุด วิธีนี้จะให้ผลลัพธ์ที่ถูกต้องมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยกำหนดค่า K เป็น 2 หรือ 3 เมื่อเทียบกับกฎ NN อย่างไรก็ตามวิธีนี้จะทำให้เวลาที่ใช้ในการรับรู้เพิ่มขึ้นด้วย

ในระบบเสียงพูดทั่วไป จะมีการกำหนดค่าระยะทางมากที่สุดที่ยอมรับได้ (Reject Value) เพื่อเพิ่มความถูกต้องของการรับรู้ในกรณีที่ผลลัพธ์ที่ได้จากการตัดสินใจมีค่าใกล้เคียงกันมาก วิธีการกำหนดค่านี้ต้องใช้ความระมัดระวังเป็นอย่างมาก เนื่องจากถ้าเรากำหนดค่านี้สูงเกินไป หรือไม่ได้กำหนด โอกาสที่จะทำให้ผลลัพธ์ของการรับรู้ผิดพลาดไปจะมีมาก แต่ถ้าเรากำหนดค่านี้ต่ำเกินไปจะทำให้ผลลัพธ์ของการรับรู้ สำหรับแบบอ้างอิงที่ถูกต้องเกิดขึ้นได้ยาก ดังนั้นโดยทั่วไปค่าระยะทางนี้มักจะได้มาจากการทดลอง เพื่อหาค่าที่เหมาะสมสำหรับกลุ่มของแบบอ้างอิงหนึ่ง ๆ เท่านั้น

2.6.9 เวคเตอร์ควอนไตซ์เซชันสำหรับระบบการรับรู้เสียงพูด



รูปที่ 2.5 แสดงเวคเตอร์ควอนไตซ์เซอร์ที่ใช้ในระบบการรับรู้เสียงพูด

เพื่อที่จะสร้างโค้ดบุค เราต้องการเวคเตอร์ทดสอบเป็นจำนวนมาก ซึ่งจะจัดเวคเตอร์ออกเป็นกลุ่ม ๆ และแต่ละกลุ่ม i แทนจุดศูนย์กลางด้วย y_i ซึ่งหาจุดศูนย์กลางได้จากการเฉลี่ยค่าเวคเตอร์ทั้งหมดในกลุ่ม ซึ่ง y_i จะเป็นเวคเตอร์โค้ดที่เก็บไว้ หรือเป็นโค้ดบุคเก็บไว้สำหรับการควอนไตซ์เวคเตอร์เสียงต่อไป เวคเตอร์เสียงอินพุทของควอนไตซ์เซอร์ = x_i จะถูกเปรียบเทียบระยะทางกับ y_i d (x_i,y_i) ค่า y_m ที่เราเลือกคือ ระยะทาง d น้อยที่สุด ซึ่ง y_m จะแทน x_i ที่เข้ามา จากนั้นจะส่งดัชนี m ไปยังเครื่องรับจากเมตริกซ์ระยะทาง เราได้ว่า ถ้า x ใกล้ y_1 มากกว่า y_2 สรุปได้ว่า x เสียงเหมือน y_1 มากกว่า y_2

เวคเตอร์ควอนไตซ์เซชันที่ใช้เพื่อการออกแบบการรับรู้เสียงพูดนั้น มีจำนวนควอนไตซ์เซอร์ M ตัวนั่นหมายถึง มี M ระดับของเสียงเพื่อการรับรู้ แต่ละระดับเสียงสามารถพิจารณามาจากเซตของข้อมูลเทรนนิ่ง $\{x_t^{(i)}\}$ เมื่อ i = 1 , 2 , ... , M ซึ่ง M เป็นดัชนีระดับ แต่ละเซตของเทรนนิ่งในแต่ละระดับจะเก็บเสียงที่อยู่ในระดับเดียวกัน เมื่อมีเสียงที่เราไม่รู้ $\{x_t\}^{Tu}$ เข้ามา จะเป็นอินพุทเข้าไปยัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทุก ๆ ควอนไตซ์เซชัน ค่าดัชนีระดับ (Index) ที่เราเลือก จะเป็นระดับที่มีความผิดเพี้ยนเฉลี่ยน้อยที่สุด D(c (i)) เมื่อ i = 1 , 2 , … , M ซึ่งความผิดเพี้ยนเฉลี่ยนี้เราดูได้จากการวัดระยะทาง

2.7 รูปแบบของไฟล์เวฟ

รูปแบบของไฟล์ wave เป็นรูปแบบไฟล์อย่างหนึ่งสำหรับการเก็บข้อมูลดิจิตอลออดิโอ (digital audio data) ซึ่งจะรองรับการจัดเก็บได้ทั้งแบบ 8 บิต , 16 บิต และหลายอัตราการสุ่ม รวมทั้งสามารถใช้ได้กับทั้งแบบ 1 แชนแนล และ 2 แชนแนล ของออดิโอ (1 แชนแนลเป็นแบบ mono และ 2 แชนแนล เป็นแบบสเตริโอ) รูปแบบนี้เป็นที่นิยมมากกับแพลตฟอร์ม IBM PC และถูกใช้อย่างกว้างขวางในโปรแกรมมืออาชีพซึ่งประมวลผลข้อมูลดิจิตอลออดิโอหรือเวฟฟอร์ม

รูปแบบนี้ใช้เวอร์ชันของ Electronic Arts Interchange File Format (IFF) ของไมโครซอฟท์สำหรับการจัดเก็บข้อมูลใน "chunks"

2.7.1 โครงสร้างไฟล์เวฟ

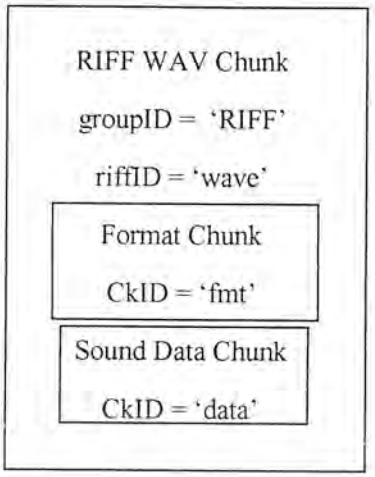
ไฟล์ wave เป็นการรวบรวมชนิดที่แตกต่างกันของ chunks จำนวนหนึ่ง ซึ่งมี Format ("fmt") chunk อันหนึ่งที่ถูกต้องการ ซึ่งจะเก็บพารามิเตอร์ที่สำคัญสำหรับการอธิบาย waveform เช่น อัตราการสุ่ม และมี Data chunk ซึ่งเก็บข้อมูลที่แท้จริงของ waveform ที่ถูกต้องการ ส่วน chunks ชนิดอื่น ๆ ก็ตามแต่ความเหมาะสม ในจำนวน chunks ชนิดอื่น ๆ นี้ อย่างเช่น cue chunk , list chunk เป็นต้น

แอพพลิเคชันซึ่งใช้ wave ต้องสามารถอ่าน 2 chunks ที่ถูกต้องการได้ คือ Format chunk และ Data chunk และสามารถเลือกที่จะปฏิเสธ chunk อื่น ๆ ได้ โปรแกรม ๆ หนึ่ง ซึ่งทำการสำเนา wave จะสำเนา chunks ทั้งหมดใน wave และทำการเลือกที่จะไม่ตีความ

ไม่มีข้อบังคับเกี่ยวกับลำดับของ chunks ภายในไฟล์ wave แต่มีข้อยกเว้นว่า Format chunk ต้องอยู่ก่อน Data chunk โปรแกรมที่ถูกเขียนขึ้นโดยขาดความยืดหยุ่นคาดหวังว่า Format chunk เป็น chunk แรก (หลังจากส่วนหัวของ RIFF) แม้ว่าโปรแกรมเหล่านั้นไม่ควรจะทำเช่นนั้นเพราะว่าการเจาะจงเช่นนั้นจะไม่ถูกกระทำในไฟล์ wave

ตัวอย่างของไฟล์ wave ขนาดเล็ก มันจะประกอบไปด้วย wave หนึ่ง wave ซึ่งบรรจุ 2 chunk นั่นคือ Format Chunk และ Data chunk

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 ตัวอย่างของไฟล์ wave ขนาดเล็ก

2.7.2 Sample Points และ Sample Frames

ส่วนของการตีความไฟล์ wave จะอ้างอิงถึง 2 คอนเซ็ปต์ คือ Sample points และ Sample frames โดยที่ sample point เป็นค่า ๆ หนึ่งซึ่งแทนตัวอย่างของเสียง ๆ หนึ่งที่เวลาใด ๆ สำหรับ wave form ที่มีการจัดเก็บมากกว่า 8 บิต แต่ละ sample point จะถูกเก็บเป็นแนวยาว ซึ่งเป็นค่าเลขฐานสอง อาจจะเป็นได้ตั้งแต่ 9 ถึง 32 บิต ขึ้นอยู่กับฟิลด์ wBitsPerSample ใน Format chunk โดยสมมุติว่ารูปแบบ PCM เป็นรูปแบบที่ไม่มีการบีบอัด ตัวอย่างเช่น แต่ละ sample point ของ waveform ที่มีการจัดเก็บแบบ 16 บิต จะเป็นหนึ่ง word หรือ 2 ไบต์ ซึ่ง 32767(0x7FFF) เป็นค่าสูงสุด และ –32768 (0x8000) เป็นค่าต่ำสุด สำหรับ waveforms แบบ 8 บิต หรือน้อยกว่า แต่ละ sample point เป็น 1 ไบต์ ซึ่งไม่คิดเครื่องหมายดังนั้น 255 เป็นค่าสูงสุด และ 0 เป็นค่าต่ำสุด

เนื่องจากการทำการอ่านและเขียนของ CPU ส่วนใหญ่กระทำกับ 8 บิต จึงถูกตัดสินว่า หนึ่ง sample point จะมีขนาดได้เป็นจำนวนเท่าของ 8 เมื่อถูกเก็บในไฟล์ wave ซึ่งจะทำให้ wave นั้นง่ายขึ้นในการอ่านเข้าสู่ memory ถ้าขบวนการในการแปลงอนาลอกเป็นดิจิตอลให้ sample point ขนาด 1 บิต ถึง 8 บิต sample point จะถูกเก็บในไฟล์ wave เป็น 8 บิต แต่ถ้าขบวนการในการแปลงอนาลอกเป็นดิจิตอลให้ sample point ขนาด 9 บิต ถึง 16 บิต sample point จะถูกเก็บในไฟล์ wave เป็น 16 บิต และถ้าขบวนการในการแปลงอนาลอกเป็นดิจิตอลให้ sample point ขนาด 17 ถึง 24 บิต sample point จะถูกเก็บในไฟล์ wave เป็น 3 ไบต์ และถ้าขบวนการแปล'อนาลอกเป็นดิจิตอลให้ sample point ขนาด 25 ถึง 32 บิต sample point จะถูกเก็บในไฟล์ wave เป็น 32 บิต หรือ double word เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยิ่งไปกว่านั้น บิตข้อมูลจะถูกจัดจากซ้ายและบิตที่เหลือจะเป็น 0 ตัวอย่างเช่น กรณี sample point มีขนาด 12 บิต sample point ต้องถูกบันทึกเป็น 16 บิต โดยที่ 12 บิตจะถูกจัดจากทางซ้าย ดังนั้นจะเป็นว่าทั้ง 12 บิตจะอยู่บิต 4 ถึง 15 และบิต 0 ถึง 3 ถูกกำหนดเป็น 0 ดังแสดงในรูปข้างล่าง

1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0

รูปที่ 2.7 แสดงลักษณะการจัดบิตกรณี sample point 12 บิต ถูกจัดเป็นแบบ 16 บิต

แต่สังเกตว่าเนื่องจากรูปแบบ wave ใช้ Intel little endian byte order ดังนั้น LSB จะถูกเก็บเป็นอันดับแรกใน ไฟล์ wave ดังรูป

0	1	1	1	0	0	0	0

1	0	1	0	0	0	0	1

รูปที่ 2.8 แสดงการจัดเก็บ LSB

สำหรับเสียงที่มีหลายแชนแนล เช่น waveform แบบ stereo sample point จากแต่ละแชนแนลจะถูกนำมาแทรก ตัวอย่างเช่น สมมุติให้เป็น waveform แบบ stereo ซึ่งเป็นแบบ 2 แชนแนล แทนการเก็บ sample points สำหรับแชนแนลซ้ายเป็นอันดับแรก และเก็บ sample points สำหรับแชนแนลขวาต่อไป ให้ทำการผสม sample points ของทั้งสองแชนแนลเข้าด้วยกัน โดยจะทำการเก็บ sample points แรกของแชนแนลซ้ายก่อน จากนั้นจึงทำการเก็บ sample point แรกของแชนแนลขวา ต่อไปจึงทำการเก็บ sample point อันดับที่สองของแชนแนลซ้าย จากนั้นจึงทำการเก็บ sample point อันสองของแชนแนลขวา ทำเช่นนี้ไปเรื่อย ๆ ดังนั้น sample points ซึ่งถูกทำให้เกิดขึ้นที่ถูกเก็บติดกันจะถูกเล่นพร้อมกัน sample points ซึ่งถูกทำให้เกิดขึ้นที่ถูกเก็บติดกันถูกเล่นพร้อมกัน (ส่งไปยัง D/A converter) จะถูกเรียกว่า sample frame เช่นตัวอย่าง waveform แบบ stereo ทุก ๆ 2 sample points จะถูกทำเป็น sample frame ใด ๆ ดังรูปข้างล่าง

Ch1	Ch2	Ch1	Ch2	---	Ch1	Ch2

= 1 sample point

รูปที่ 2.9 แสดงรูปแบบของ sample frame ของ waveform แบบ 2 แชนแนล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับ waveform แบบโมโนโฟนิค sample frame ก็คือ sample point สำหรับ waveform แบบมัลติแชนแนล จะแสดงถึงลำดับการจัดเก็บ sample points ของแต่ละแชนแนลภายใน sample frame ดังข้างล่าง

channels

รูปที่ 2.10 แสดงลำดับการจัดเก็บ sample points ของแต่ละแชนแนลภายใน sample frame

sample points ภายใน sample frame จะถูกรวมเข้าด้วยกัน ไม่มี bytes ที่ไม่ถูกใช้ใน sample frame คล้ายกับว่า sample frame ถูกรวมเข้าด้วยกันโดยไม่มีส่วนที่เป็น pad bytes (bit ที่เกิดแล้วถูกกำหนดเป็น 0) สังเกตว่า การพิจารณาข้างต้นรูปแบบของข้อมูลภายใน data chunk ไม่มีการบีบอัดข้อมูล แต่มีบางเทคนิคจะทำการจัดเก็บข้อมูลที่ถูกบีบอัดใน data chunk และเป็นที่สังเกตว่าข้อมูลจะต้องการที่จะไม่มีการบีบอัด

2.7.3 Format chunk

format (fmt) chunk จะทำการอธิบายพารามิเตอร์พื้นฐานของข้อมูล waveform เช่น อัตราการสุ่ม , จำนวนบิตในการเก็บ และจำนวนแชนแนลของดิจิตอลออดิโอ ที่ถูกเก็บในไฟล์ wave

```
# define FormatID'fmt'   /*chunk ID สำหรับ Format Chunk */
typedef struct {
ID        chunkID;
Long      chunkSize;
Short     wFormatTag;
Unsigned short wChannels;
Unsigned long dwSamplesPerSec;
Unsigned long dwAvgBytesPerSec;
Unsigned short wBlock Align;
Unsigned short wBitsPerSample;
} Format chunk;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ID เป็น "fmt" เสมอ ฟิลด์ chunkSize เป็นจำนวนไบต์ใน chunk ซึ่งไม่รวม 8 ไบต์ที่ถูกใช้โดย ID และฟิลด์ size สำหรับ Format chunk chunksize จะเปลี่ยนแปลงไปขึ้นอยู่แบรูปแบบของไฟล์ wave ที่ถูกระบุว่าคืออะไร (ขึ้นกับค่าของ wFormatTag)

ข้อมูล wave จะถูกเก็บโดยปราศจากการบีบอัด ในกรณีซึ่ง sample points ถูกจัดเก็บดังที่อธิบายใน sample points และ sample frames ในทางตรงกันข้าม รูปแบบที่แตกต่างกันของการบีบอัดจะถูกใช้เมื่อทำการเก็บข้อมูลเสียงใน Data chunk โดยการบีบอัดนั้น แต่ละ sample point จะใช้จำนวนไบต์ต่างกันในการเก็บ wFormatTag จะบอกว่าการบีบอัดถูกใช้เพื่อการเก็บข้อมูลหรือไม่ ถ้าใช้การบีบอัด (wFormatTag จะมีค่าที่ไม่ใช่ 1) แล้วนั่นจะเป็นฟิลด์ที่เพิ่มเข้าไปต่อท้าย Format chunk ซึ่งจะให้ข้อมูลที่ถูกต้องการสำหรับโปรแกรมที่ต้องการจะนำข้อมูลออกมาและทำการคลายการบีบอัดข้อมูลที่ถูกเก็บไว้ ฟิลด์ที่เพิ่มเข้าไปฟิลด์แรกจะเป็น unsigned short ซึ่งระบุจำนวนไบต์ที่เพิ่มขึ้นยิ่งไปกว่านั้น รูปแบบที่ถูกบีบอัดต้องมี Fact chunk ซึ่งเก็บ unsingned long เพื่อเพิ่มขนาดของ waveform หลังจากคลายการบีบอัด แต่ถ้าไม่มีการบีบอัด (wFormatTag = 1) จะไม่มีฟิลด์เพิ่ม

ฟิลด์ wChannels จะเก็บจำนวนของแชนแนลออดิโอสำหรับเสียง ค่า 1 หมายถึงเสียงแบบโมโนโฟนิค, 2 หมายถึง สเตริโอ, 4 หมายถึง เสียง 4 แชนแนล เป็นต้น สำหรับเสียงมัลติแชนแนล sample point จากแต่ละแชนแนลจะถูกนำมาแทรกเข้าด้วยกัน กลุ่มของ sample point ที่ถูกนำมาแทรกจะเรียกว่า sample frame

ฟิลด์ dwSamplesPerSec เป็นอัตราสุ่มซึ่งเสียงจะถูก playback ในหน่วย sample frame ต่อวินาที ซึ่งมีอัตรามาตรฐาน MPC อยู่ 3 อัตราคือ 11025,22050 และ 44100 กิโลเฮิร์ต และอัตราอื่นๆก็สามารถใช้ได้

ฟิลด์ dwAvgBytesPerSec จะบอกจำนวนไบต์ที่ play ทุกๆวินาที dwAvgBytesPerSec จะถูกใช้โดยแอพพลิเคชันเพื่อประมาณขนาดของ RAM บัฟเฟอร์ที่ถูกต้องการเพื่อ playback wave โดยปราศจากการล่าช้า ค่าของมันจะเท่ากับสูตรข้างล่าง

dwSamplesPerSec*wBlockAlign

ฟิลด์ wBlockAlign จะเท่ากับสูตรข้างล่าง

wChannels*(wBitsPerSample%8)

wBlockAlign เป็นขนาดของ sample fame ในหน่วยไบต์ ฟิลด์ wBitsPerSample บอกถึงจำนวนบิทของ sample point

2.7.4 Data Chunk

Data chunk จะเก็บ sample frame จริงๆ

```
# define DataID 'data' /* chunk สำหรับ data */
typedef    struct{
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
ID   chunkID;
Long  chunkSize;
Unsigned char waveformData[]
}DataChunk;
```

ID เป็น'data'เสมอ ChunkSize เป็นจำนวนของไบต์ใน chunk ไม่นับ 8 ไบต์ที่ใช้โดย ID และฟิลด์ Size และ pad bytes

การพิจารณาต่อไปนี้สมมติข้อมูลไม่ถูกบีบอัด

อาร์เรย์ WaveFormatData จะเก็บข้อมูล waveform จริงๆ และข้อมูลจะถูกจัดเข้าสู่สิ่งที่เรียกว่า sample frames เราสามารถที่จะทำการพิจารณาจำนวนไบต์ของข้อมูล waveform จริงๆ จากฟิลด์ chunkSize ของ Data chunk จำนวนของ sample frames ใน waveformData ถูกพิจารณาโดยการหาร chunkSize โดย wBlockAlign ของ Format chunk

2.8 ข้อมูลเกี่ยวกับซอฟต์แวร์

2.8.1 Microsoft Visual C++

Microsoft Visual C++ เป็นโปรแกรมประเภทวิชวล อีกตัวหนึ่งจากบริษัทไมโครซอฟต์ ผู้ผลิตระบบปฏิบัติการวินโดวส์ เป็นเครื่องมือพัฒนาโปรแกรมที่มีความสามารถสูงในยุคนี้ Microsoft Visual C++ ได้รับการพัฒนาให้มีความยืดหยุ่นและมีประสิทธิภาพสูงขึ้นมาจากภาษา C++ และได้สนับสนุนการพัฒนาโปรแกรมในหลาย ๆ ด้านไม่ว่าจะเป็นการสร้างโปรแกรมทั่วไป, การสร้างโปรแกรมการจัดการฐานข้อมูล, การสร้างโปรแกรมบนระบบเครือข่าย หรือ มัลติมีเดีย อย่างครบครัน

ในปัจจุบัน Microsoft Visual C++ ได้รับการพัฒนาจนถึงเวอร์ชันที่ 6 มีลักษณะเป็น IDE (Integrated Development Environment) คือเป็นโปรแกรมซึ่งมีไว้ใช้สำหรับเพิ่มความสะดวกในการสร้างและแก้ไขโปรเจ็กต์ให้ง่ายขึ้นโดยจะมีเครื่องมืออำนวยความสะดวกในการพัฒนาโปรแกรมต่าง ๆ ให้เรียกใช้งานใน IDE เช่น โปรแกรมที่ใช้ในการดีบัก, คอมไพเลอร์และลิงค์เกอร์ เป็นต้น นอกจากนี้ยังมีส่วนรองรับการพัฒนาโปรแกรมบนวินโดวส์ โดยมี MFC (Microsoft Foundation Class) เป็นไลบรารีที่จะช่วยอำนวยความสะดวกในการพัฒนาโปรแกรมบนวินโดวส์

นอกจากนี้ Microsoft Visual C++ ยังมีส่วนที่สนับสนุนทางด้านของมัลติมีเดีย เกี่ยวกับรูปภาพ, การเล่นไฟล์ภาพเคลื่อนไหวรวมถึงทางด้านของเสียง (Sound) ทั้งยังมีส่วนของเครื่องมือที่สนับสนุนการใช้งานที่เกี่ยวข้องกับ Buffer I/O ,RIFF files ที่เรียกกันว่า Multimedia file I/O services อีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8.2 Multimedia file I/O services

Multimedia file I/O services ได้จัดเตรียมฟังก์ชันในการใช้งานมากกว่าการบริการพื้นฐานของ Operating System รวมทั้งสนับสนุนการใช้งานที่เกี่ยวข้องกับ Buffer I/O ,RIFF files ,memory files และ custom storage systems

ตัวอย่างของฟังก์ชันที่มีไว้บริการใน Multimedia file I/O มีดังนี้

1. mmioOpen จะทำการเปิดไฟล์ ไฟล์ที่เปิดสามารถเป็นได้ทั้ง ไฟล์มาตรฐาน memory file ฟังก์ชัน mmioOpen จะทำการส่งค่า handle กลับคืนแต่ไม่ใช่ standard file handle ซึ่ง mmioOpen มีโครงสร้างดังนี้

```
HMMIO mmioOpen(
   LPSTR szFilename,
   LPMMIOINFO lpmmioinfo,
   DWORD  dwOpenFlags
);
```

พารามิเตอร์

SzFilename : ตำแหน่งของ string ซึ่งเก็บชื่อไฟล์ที่ต้องการจะเปิด

Lpmmioinfo : ตำแหน่งของโครงสร้าง MMIOINFO ซึ่งเก็บพารามิเตอร์ที่ใช้โดย lpmmioinfo

dwOpenFlags : แฟลกสำหรับการเปิดไฟล์เช่น MMIO_READ, MMIO_WRITE และ MMIO_READWRITE

ค่าที่ส่งกลับ : ถ้าไฟล์ไม่สามารถเปิดได้จะส่งค่ากลับเป็น NULL

2. mmioRead ใช้เพื่อทำการอ่านไบต์ข้อมูลจำนวนหนึ่งจากไฟล์ที่เปิดโดยใช้ฟังก์ชัน mmioOpen และมีโครงสร้างดังนี้

```
LONG mmioRead(
   HMMIO  hmmio,
   HPSTR  pch,
   LONG  cch
);
```

พารามิเตอร์

hmmio : ไฟล์ handle ของไฟล์ที่จะอ่าน

pch : ตำแหน่งของบัฟเฟอร์เพื่อที่จะเก็บข้อมูลที่อ่านมาจากไฟล์

cch : จำนวนไบต์ที่อ่านจากไฟล์

ค่าที่ส่งกลับ : ค่าที่ส่งกลับคือ จำนวนไบต์ที่อ่านจริง ถ้าอ่านถึงท้ายไฟล์แล้วไม่มีไบต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่สามารถอ่านได้ จะส่งค่า 0 กลับมา ถ้ามีข้อผิดพลาดที่เกิดขึ้นจากการอ่านไฟล์ จะทำการส่งค่า –1 กลับไป

3. mmioSetBuffer ใช้เพื่อระบุให้บัฟเฟอร์ I/O ทำงานได้หรือไม่ได้ หรือ ทำการเปลี่ยนแปลงบัฟเฟอร์ หรือขนาดบัฟเฟอร์ สำหรับไฟล์ที่ถูกเปิด โดยใช้ฟังก์ชัน mmioOpen ซึ่งมีโครงสร้างดังนี้

```
MMRESULT mmioSetBuffer(
    HMMIO hmmio,
    LPSTR pchBuffer,
    UINT wFlags
);
```

พารามิเตอร์

hmmio : ไฟล์ handle ของไฟล์ที่ถูกเปิดขึ้นมาเพื่อจะทำงานกับบัฟเฟอร์

pchBuffer : ตำแหน่งของบัฟเฟอร์ที่นิยามโดยแอปพลิเคชั่น เพื่อใช้สำหรับบัฟเฟอร์ I/O ซึ่งถ้าค่านี้เป็น NULL ฟังก์ชัน mmioSetBuffer จะทำการจองพื้นที่ภายในสำหรับบัฟเฟอร์ I/O

cchBuffer : ขนาดของบัฟเฟอร์ที่นิยามโดยแอปพลิเคชั่น หรือขนาดของบัฟเฟอร์สำหรับ mmioSetBuffer ทำการจองพื้นที่

wFlags : ต้องเป็น 0

ค่าที่ส่งกลับ : ค่าที่ส่งกลับจะเป็น 0 ถ้าสามารถกระทำได้สำเร็จหรือเป็นค่าอื่นในกรณีเกิดข้อผิดพลาด ค่าที่สามารถทำการนิยามได้เช่น

MMIOERR_CANNOTWRITE : สิ่งที่มีอยู่ภายในของบัฟเฟอร์เก่าไม่สามารถถูกเขียนลงสู่ดิสก์ได้ ดังนั้นการทำงานจึงถูกยกเลิก

MMIOERR_OUTOFMEMORY : ไม่สามารถทำการจองพื้นที่ในหน่วยความจำให้แก่บัฟเฟอร์ได้ เนื่องจากพื้นที่ในหน่วยความจำไม่เพียงพอ

4. mmioWrite ใช้ทำการเขียนไบต์ข้อมูลจำนวนหนึ่งลงสู่ไฟล์ที่เปิดโดยใช้ mmioOpen ซึ่งมีโครงสร้างดังนี้

```
LONG mmioWrite(
HMMIO hmmio,
    char _huge* pch,
    LONG cch
);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พารามิเตอร์

hmmio : ไฟล์ handle ของไฟล์ที่จะทำการเขียนข้อมูล

pch : ตำแหน่งของบัฟเฟอร์ที่เก็บข้อมูลที่จะถูกเขียนลงไบต์

cch : จำนวนไบต์ที่จะทำการเขียนลงไฟล์

ค่าที่ส่งกลับ : ค่าที่ส่งกลับคือจำนวนไบต์ที่ทำการเขียนลงไฟล์จริง ๆ ถ้ามีข้อผิดพลาดเกิดขึ้น ค่าที่ส่งกลับจะเป็น –1

5. mmioClose ใช้เมื่อต้องการปิดไฟล์ซึ่งถูกเปิดโดย mmioOpen ซึ่งมีโครงสร้างดังนี้

```
MMRESULT mmioClose(
    HMMIO hmmio,
    UNIT wFlags
);
```

พารามิเตอร์

hmmio : ไฟล์ handle ของไฟล์ที่จะทำการปิด

wFlags : แฟลก สำหรับกระทำการปิดไฟล์สามารถทำการนิยามได้ดังค่าต่อไปนี้

MMIO_FHOPEN : ถ้าไฟล์ถูกเปิดโดยการส่งไฟล์ handle ซึ่งไม่ ได้มีชนิดเป็น HMMIO การใช้ flag นี้เพื่อบอกให้ฟังก์ชัน mmioClose ทำการปิด multimedia file handle ที่ไม่ใช่ไฟล์ handle มาตรฐาน

ค่าที่ส่งกลับ : ถ้าสามารถทำได้สำเร็จจะส่งค่า 0 กลับคืน หรือถ้ามีข้อผิดพลาดมากขึ้นจะส่งค่าอื่นกลับคืน ค่าของข้อผิดพลาดสามารถเกิดได้จากฟังก์ชัน mmioFlush ค่าที่สามารถเป็นไปได้จะรวมถึงค่าต่อไปนี้ด้วย

MMIOERR_CANNOTWRITE : ส่วนที่อยู่ภายในบัฟเฟอร์ไม่สามารถถูกเขียนลงสู่ดิสก์

6. mmioCreateChunk ฟังก์ชันนี้จะทำการสร้าง chunk ในไฟล์ RIFF ซึ่งถูกเปิดโดยการใช้ฟังก์ชัน mmioOpen chunk ใหม่จะถูกสร้างขึ้น ณ ตำแหน่งไฟล์ปัจจุบัน หลังจาก chunk ใหม่ถูกสร้างขึ้น ตำแหน่งไฟล์ปัจจุบันจะเป็นจุดเริ่มต้นของส่วนข้อมูลของ chunk ใหม่ ซึ่งมีโครงสร้างดังนี้

```
MMRESULT mmioCreateChunk (
    HMMIO hmmio,
    LPMMCRINFO lpck,
    UINT wFlags
);
```

พารามิเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hmmio : ไฟล์ handle ของไฟล์ RIFF ที่ถูกเปิด

lpck : ตำแหน่งของโครงสร้าง MMCKINFO ที่ถูกนิยามขึ้นโดยแอปพลิเคชั่น ซึ่งจะเก็บข้อมูลเกี่ยวกับ chunk ที่ถูกสร้าง

wFlags : ระบุ ค่าของ chunk ที่จะทำการสร้าง ซึ่งสามารถนิยามได้ดังนี้

MMIO_CREATELIST : "LIST"chunk

MMIO_CREATERIFF : "RIFF"chunk

ค่าที่ส่งกลับ : ส่งค่า MMSYSERR_NOERROR กลับ ถ้าสามารถทำได้สำเร็จหรือถ้ามีข้อผิดพลาดเกิดขึ้นจะส่งค่าความผิดพลาดกลับคืน ซึ่งค่าความผิดพลาดมีดังต่อไปนี้

MMIOERR_CANNOTSEEK : เมื่อไม่สามารถพิจารณาออฟเซตของส่วนข้อมูลของ chunk

MMIOERR_CANNOTWRITE : เมื่อไม่สามารถเขียนส่วนหัวของ chunk ได้

2.8.3 Microsoft Visual Basic

Microsoft Visual Basic เป็นภาษาคอมพิวเตอร์ที่ได้รับความนิยมนำมาใช้พัฒนาโปรแกรมบนระบบปฏิบัติการวินโดว์ เนื่องจากเป็นภาษาคอมพิวเตอร์ที่ใช้เทคโนโลยีในลักษณะวิชวล ซึ่งเพียงแต่เลือกคอนโทรล ที่เหมาะสมแล้ววาดลงบนฟอร์ม ก็สามารถสร้างจอภาพที่ใช้สำหรับติดต่อกับผู้ใช้ได้ รวมถึงการใช้เทคนิคการเขียนโปรแกรมแบบอีเวนต์ ไดรฟเวน ซึ่งเป็นการเขียนโปรแกรมเพื่อกำหนดขั้นตอนการทำงานต่าง ๆ ให้กับคอนโทรล ที่สร้างขึ้นตามเหตุการณ์ต่าง ๆ ที่เกิดขึ้น เช่น การเลื่อนเมาส์ หรือการรับข้อมูลจากคีย์บอร์ด เป็นต้น

ในส่วนของการวาดภาพ หรือ การวาดกราฟ Microsoft Visual Basic ได้จัดให้มีส่วนของ พิคเจอร์ บ็อกซ์คอนโทรล นอกจากนี้ ยังมีฟังก์ชันที่ใช้ในการวาดจุด เส้น วงกลม อีกด้วย

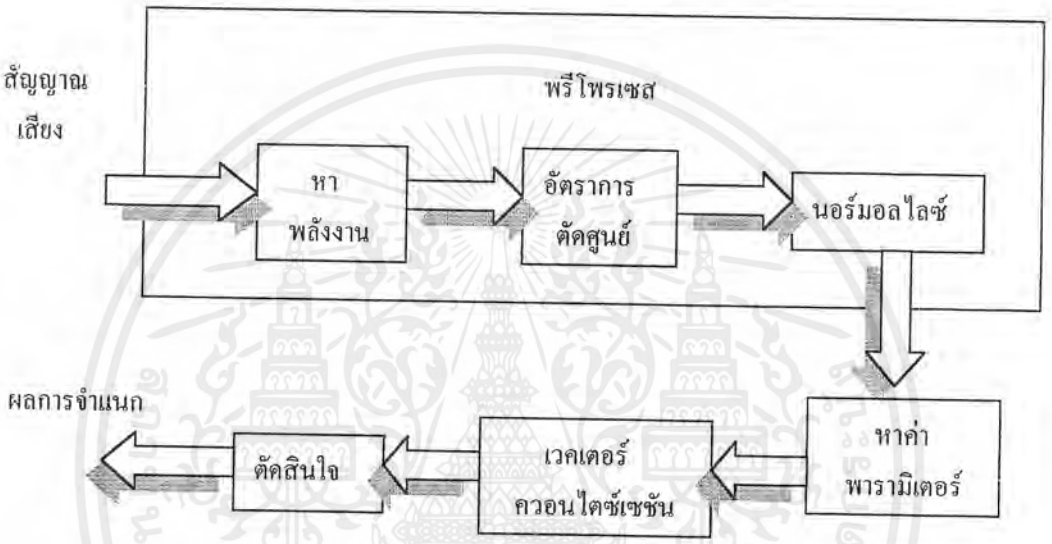
2.8.4 Poser 4

Poser 4 เป็นโปรแกรมสำหรับสร้างภาพกราฟฟิก โดยในตัวโปรแกรมจะมีเครื่องมือสำหรับการสร้างร่างกาย การแสดงท่าทางต่างๆรวมทั้งเคลื่อนไหวได้ในลักษณะที่เป็นภาพประกอบ หรือภาพเคลื่อนไหว 3 มิติ ตัวโปรแกรมจะมีเวิร์กสเปซซึ่งเป็นหน้าต่างที่ปรากฏขึ้นหลังจากที่ทำการเปิดโปรแกรม เป็นที่รวมของสิ่งที่ใช้ควบคุมส่วนต่างๆที่ใช้ใน Poser 4 เวิร์กสเปซสามารถเคลื่อนย้ายไปทุกๆที่และตำแหน่งที่เคลื่อนที่ไปก็จะถูกบันทึกไว้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การคำนวณและการสร้าง

ระบบการจดจำเสียงพูดสำหรับปริญญานิพนธ์นี้สามารถแบ่งออกเป็นส่วนต่างๆ ดังนี้คือ ส่วนของพรีโพรเซสซิ่ง ซึ่งจะเป็นการหาขอบเขตคำและการนอร์มอลไลซ์สัญญาณ เพื่อเตรียมนำสัญญาณไปผ่านขั้นการหาค่าพารามิเตอร์ ขั้นเปรียบเทียบความคล้ายคลึงซึ่งกระทำโดยอาศัยหลักการของ



รูปที่ 3.1 แสดงบล็อกไดอะแกรมการจดจำเสียงพูดโดยทั่วไป

จากรูปที่ 3.1 สามารถอธิบายการทำงานได้เป็นส่วนๆ ดังนี้

3.1 ส่วนของพรีโพรเซสซิ่ง ซึ่งแบ่งออกเป็นส่วนต่างๆ ดังนี้

3.1.1 การหาขอบเขตของคำโดยใช้การวิเคราะห์หาค่าพลังงานของสัญญาณ ในการหาขอบเขตของคำ ซึ่งมีวิธีการดังนี้

1. ก่อนเข้าสู่การหาค่าพลังงานของสัญญาณสำหรับปริญญานิพนธ์นี้จะทำการตัดตัวอย่างสัญญาณ 15 ตัวอย่างสัญญาณแรก เนื่องจากพบว่า 15 ตัวอย่างสัญญาณแรกของทุกสัญญาณมีค่าเท่ากัน

2. คำนวณหาค่าพลังงานของสัญญาณในช่วงเวลาที่เหลือทั้งหมด โดยการคำนวณหาค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พลังงานของสัญญาณซึ่งจะทำการคำนวณเป็นเฟรม กำหนดในเฟรมหนึ่งมีจำนวนตัวอย่างสัญญาณ 100 ตัวอย่างสัญญาณ โดยจะทำการคำนวณเป็นเฟรมต่อไปเรื่อย ๆ จนครบทุกตัวอย่างสัญญาณของสัญญาณ

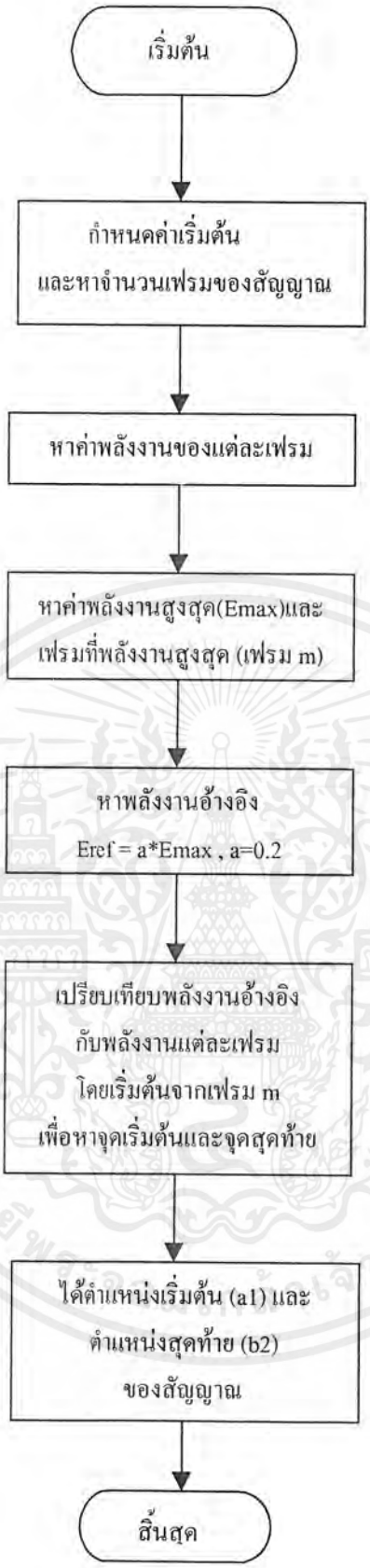
3. พิจารณาหาค่าพลังงานสูงสุดของสัญญาณ (E_{max}) และดูว่าที่ค่าพลังงานสูงสุดนี้อยู่ที่เฟรมใด แล้วกำหนดให้เป็นเฟรมที่ m

4. ทำการกำหนดพลังงานอ้างอิง (E_ref) ซึ่งพลังงานอ้างอิงสามารถคำนวณได้จาก a เท่าของพลังงานสูงสุด ($a*E_{max}$) ซึ่งในที่นี้ค่า a จะมีค่าเท่ากับ 0.2 พลังงานอ้างอิงนี้จะเป็นตัวบอกว่า จุดเริ่มตนและจุดสุดท้ายของคำจะอยู่นอกช่วงนี้

5. เมื่อกำหนดพลังงานอ้างอิง เป็นที่เรียบร้อยแล้ว ทำการหาจุดเริ่มต้นของคำ โดยเฟรมที่มีค่าพลังงานสูงสุด จะถูกพิจารณาเป็นเฟรมแรก (เฟรมที่ m) ซึ่งนำค่าพลังงานของสัญญาณเฟรมที่ m ไปเปรียบเทียบกับพลังงานอ้างอิง แล้วเลื่อนเฟรมที่จะเปรียบเทียงพลังงานกับพลังงานอ้างอิงไปด้านหลังเรื่อย ๆ (เฟรมที่ m-1 , m-2 ,... ,2 , 1) จนกว่าจะพบเฟรมที่มีค่าพลังงานน้อยกว่าค่าพลังงานอ้างอิงเป็นเฟรมแรก ก็จะได้จุดเริ่มต้นของสัญญาณ (a1) โดยถือว่าค่าพลังงานของสัญญาณค่าแรกที่มีค่าน้อยกว่าพลังงานอ้างอิง (E_ref) เป็นจุดเริ่มต้นของคำ

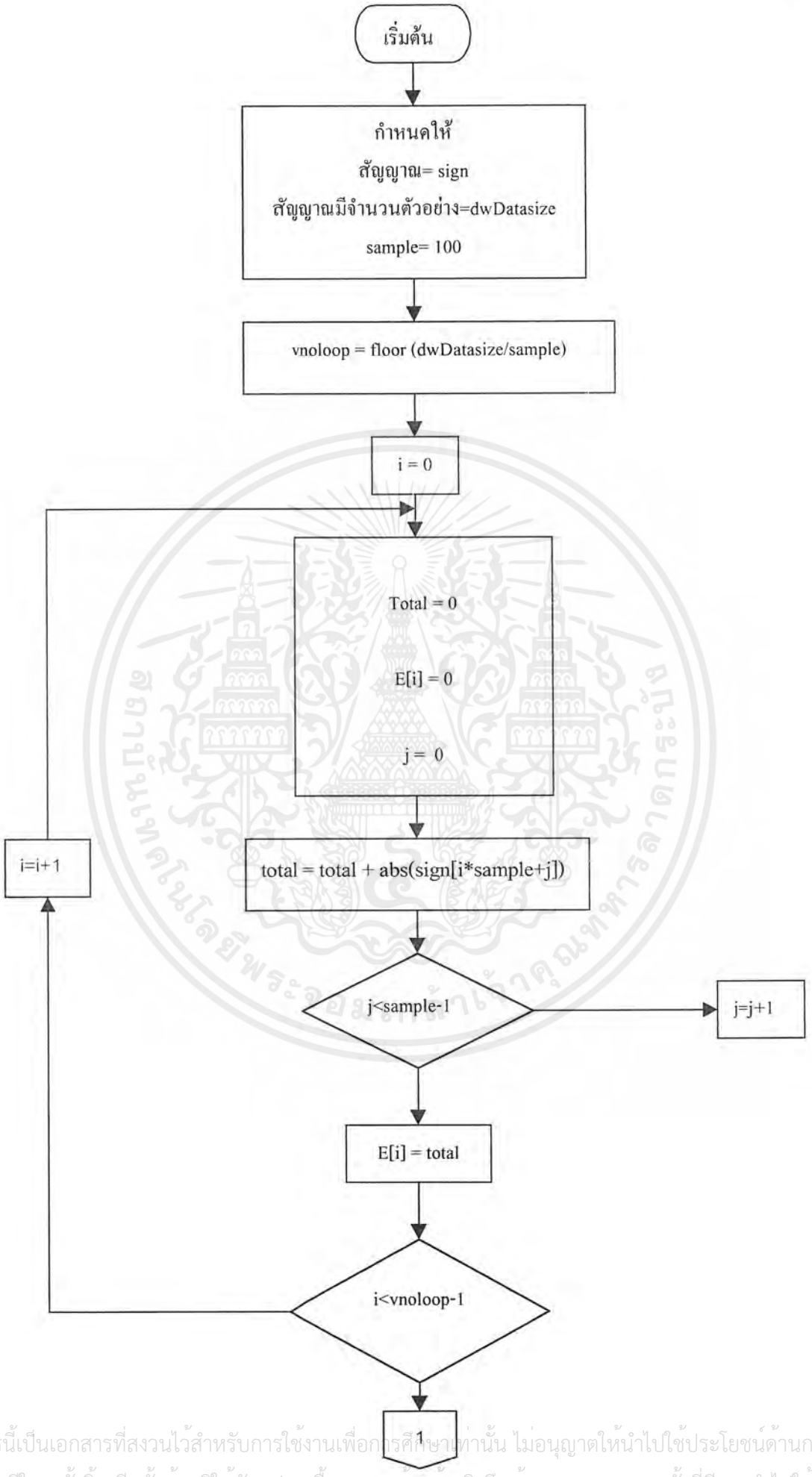
6. การหาจุดสุดท้ายของคำ จะทำการพิจารณาเช่นเดียวกับการจุดเริ่มต้นของคำ คือเริ่มเปรียบเทียบจากเฟรมที่มีพลังงานสูงสุด (เฟรมที่ m) เช่นกัน แต่การเลื่อนเฟรมไปทางด้านหน้าแทน (เฟรมที่ m+1 , m+2 ...) จนกว่าจะพบเฟรมแรกที่มีพลังงานน้อยกว่าค่าพลังงานอ้างอิง ก็จะได้จุดสุดท้ายของสัญญา (b1) โดยถือว่าค่าพลังงานของสัญญาณค่าแรกที่มีค่าน้อยกว่าพลังงานอ้างอิง เป็นจุดสุดท้ายของคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

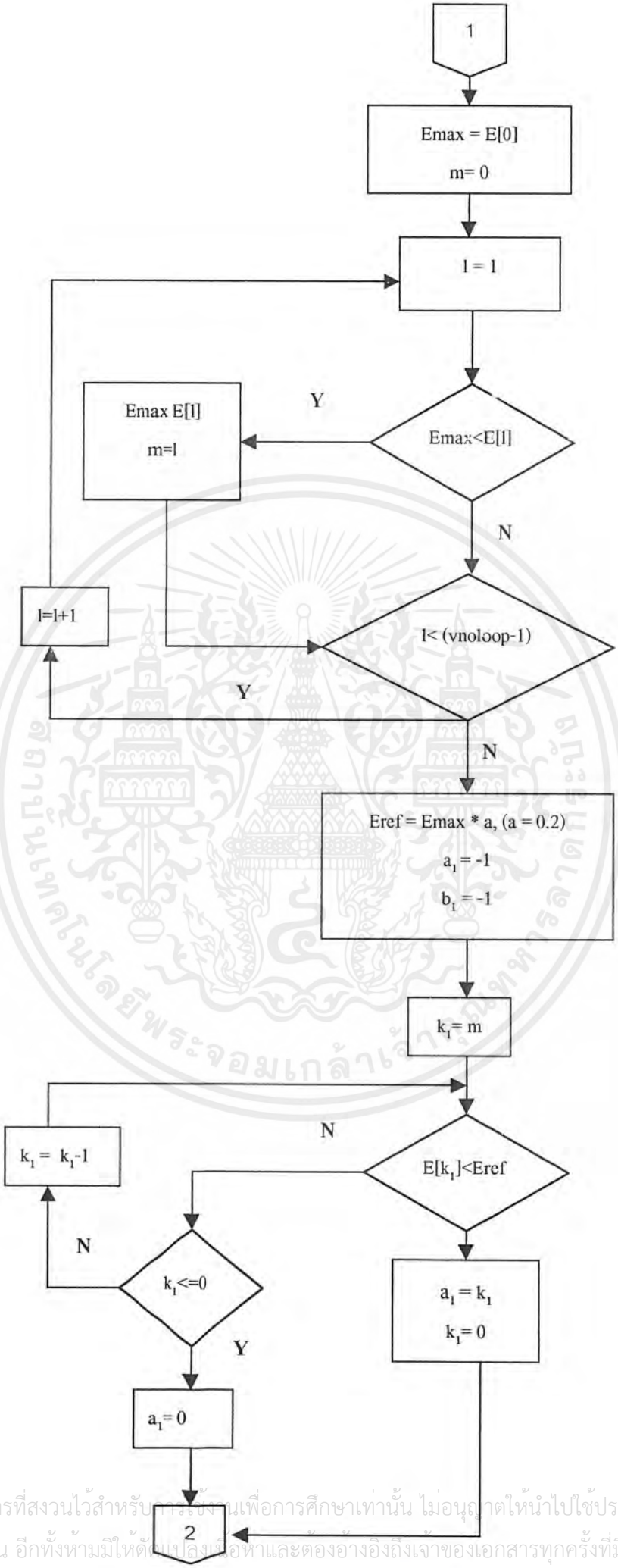


รูปที่ 3.2 แสดงขั้นตอนการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียงโดยวิธีการหาพลังงานของสัญญาณ

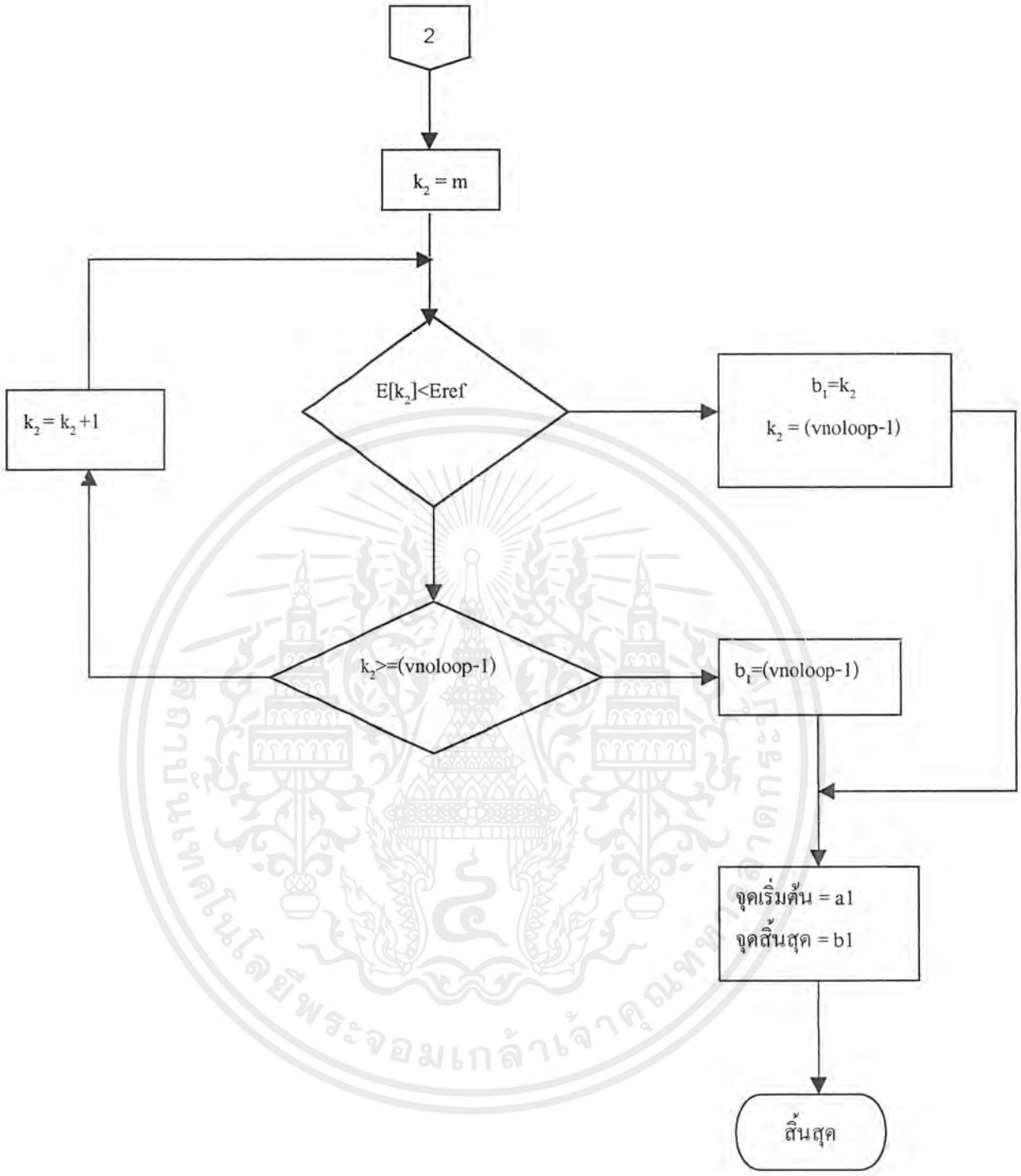
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณ
โดยการเปรียบเทียบค่าพลังงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 การหาขอบเขตคำโดยใช้การวิเคราะห์หาค่าอัตราการตัดศูนย์ สาเหตุที่ใช้วิธีนี้ช่วยเนื่องจากการหาขอบเขตของคำโดยการเปรียบเทียบค่าพลังงานอาจมีข้อผิดพลาดได้ จึงต้องใช้วิธีหาค่าอัตราการตัดศูนย์ (Zero crossing) เข้าช่วยพิจารณา ซึ่งมีวิธีการดังนี้

1. คำนวณหาค่าอัตราการตัดศูนย์ในช่วงเวลาที่เหลือจากการตัดสัญญาณ 15 ตัวอย่างสัญญาณแรกทั้งหมด โดยการคำนวณค่าอัตราการตัดศูนย์จะคำนวณเป็นเฟรม กำหนดให้แต่ละเฟรมมีตัวอย่าง 100 ตัวอย่าง และทำการคำนวณจนครบช่วงเวลาของเสียงที่เหลือทั้งหมด

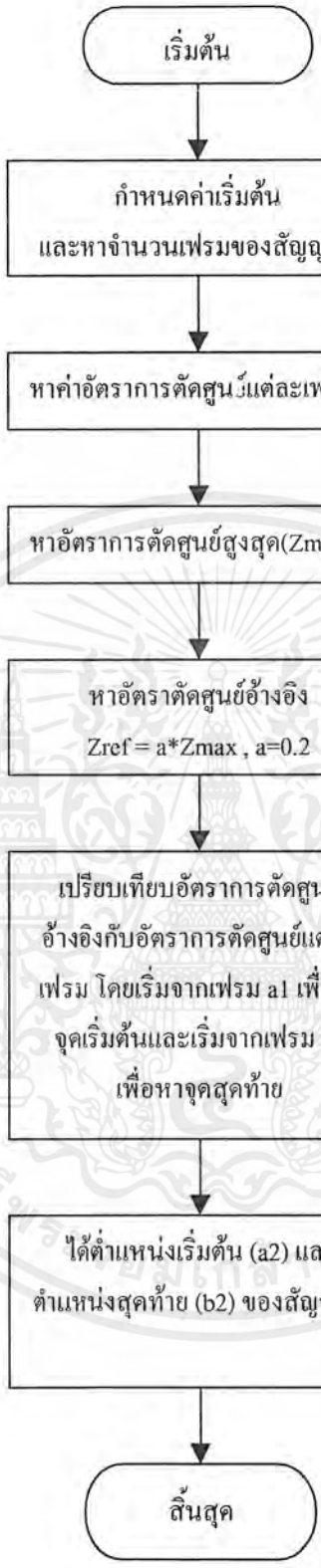
2. พิจารณาค่าอัตราการตัดศูนย์สูงสุด (Z_{max}) กำหนดให้เป็นเฟรมที่ p

3. กำหนดค่าอัตราการตัดศูนย์อ้างอิง (Z_{ref}) ซึ่งจะมีค่าเป็น a เท่าของค่าอัตราการตัดศูนย์สูงสุด ($a^* Z_{ref}$) โดยค่า a จะเท่ากับ 0.2 เช่นกัน

4. จากจุดเริ่มต้น (a1) ที่หาได้จากขั้นตอนที่ 1 จะเป็นเฟรมแรกที่จะพิจารณา โดยจะทำการเปรียบเทียบอัตราการตัดศูนย์ของเฟรมดังกล่าวกับอัตราการตัดศูนย์อ้างอิง ถ้ามีค่ามากกว่า จุดเริ่มต้นจะถูกเปลี่ยนตำแหน่ง โดยจะเลื่อนไปยังจุดแรกที่มีค่าน้อยกว่าอัตราการตัดศูนย์อ้างอิง (a2) และจะถือว่าจุดนี้เป็นจุดเริ่มต้นที่แท้จริง ซึ่งการเลื่อนเฟรมจะเลื่อนไปด้านหลัง (เฟรมที่ a1-1 , a1-2 ,... ,2 , 1) แต่ในกรณีที่ a1 น้อยกว่าอัตราการตัดศูนย์อ้างอิง จะถือว่า a1 ที่หาได้จากขั้นตอนที่ 1 เป็นจุดเริ่มต้นที่ถูกต้องอยู่แล้ว

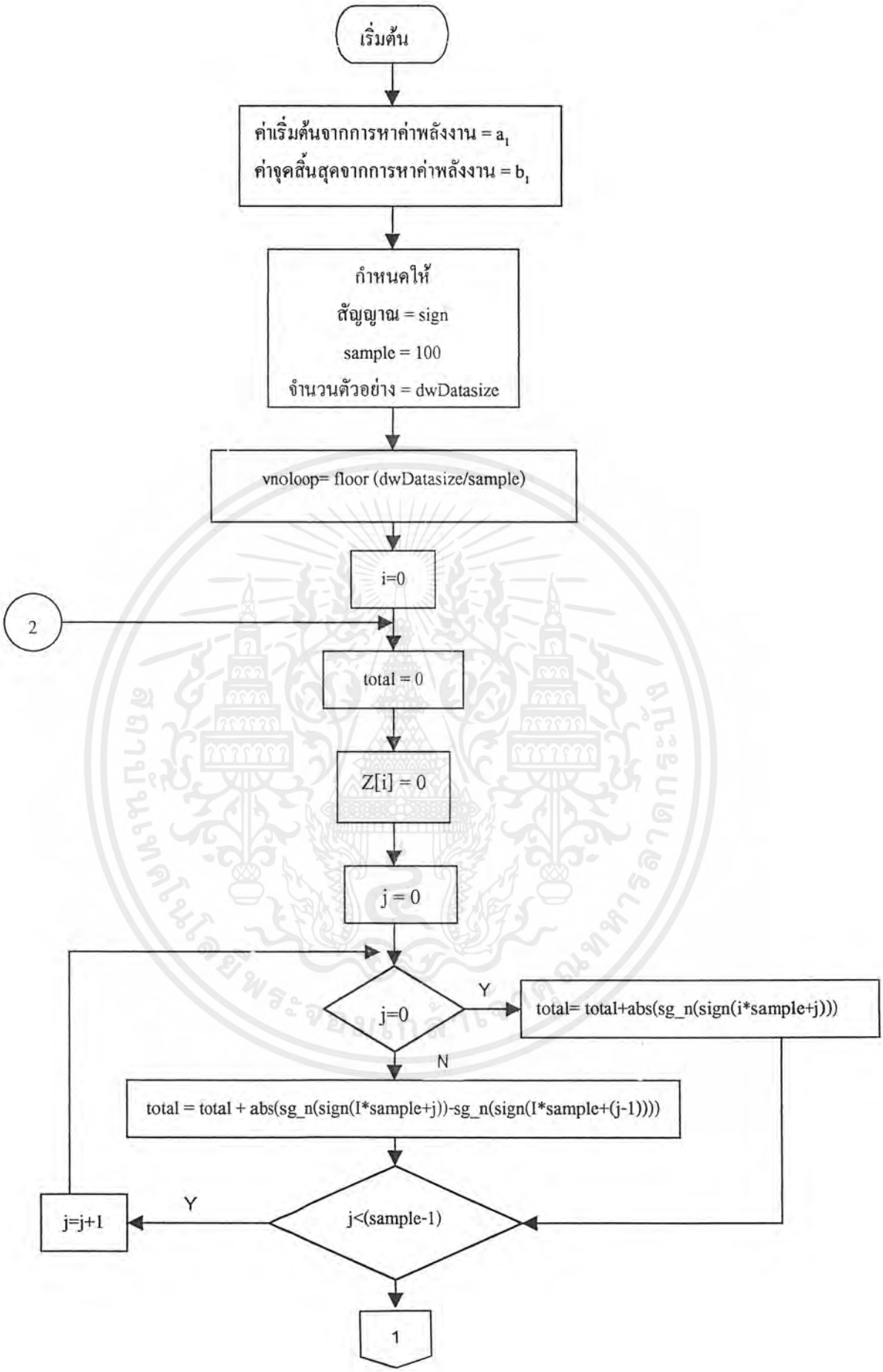
5. การหาจุดสุดท้าย จะหาจากสุดท้าย (b1) ที่ได้จากขั้นตอนที่ 1 พิจารณาเหมือนข้อ 3 คือจะเป็นเฟรมแรกที่จะเปรียบเทียบอัตราการตัดศูนย์กับอัตราการตัดศูนย์อ้างอิง ถ้ามากกว่า จุดสุดท้ายจะเปลี่ยนตำแหน่ง โดยจะเลื่อนเฟรมไปทางด้านหน้า (เฟรมที่ b1+1 , b1+2) จนกว่าจะเจอจุดแรกที่มีค่าน้อยกว่าอัตราการตัดศูนย์อ้างอิง (b2) และจะถือว่าจุดนี้เป็นจุดสุดท้ายที่แท้จริง แต่ถ้าอัตราการตัดศูนย์ที่ b1 น้อยกว่าอัตราการตัดศูนย์อ้างอิง ก็กล่าวได้ว่า b1 เป็นจุดสุดท้ายที่ถูกต้องอยู่แล้วเช่นกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

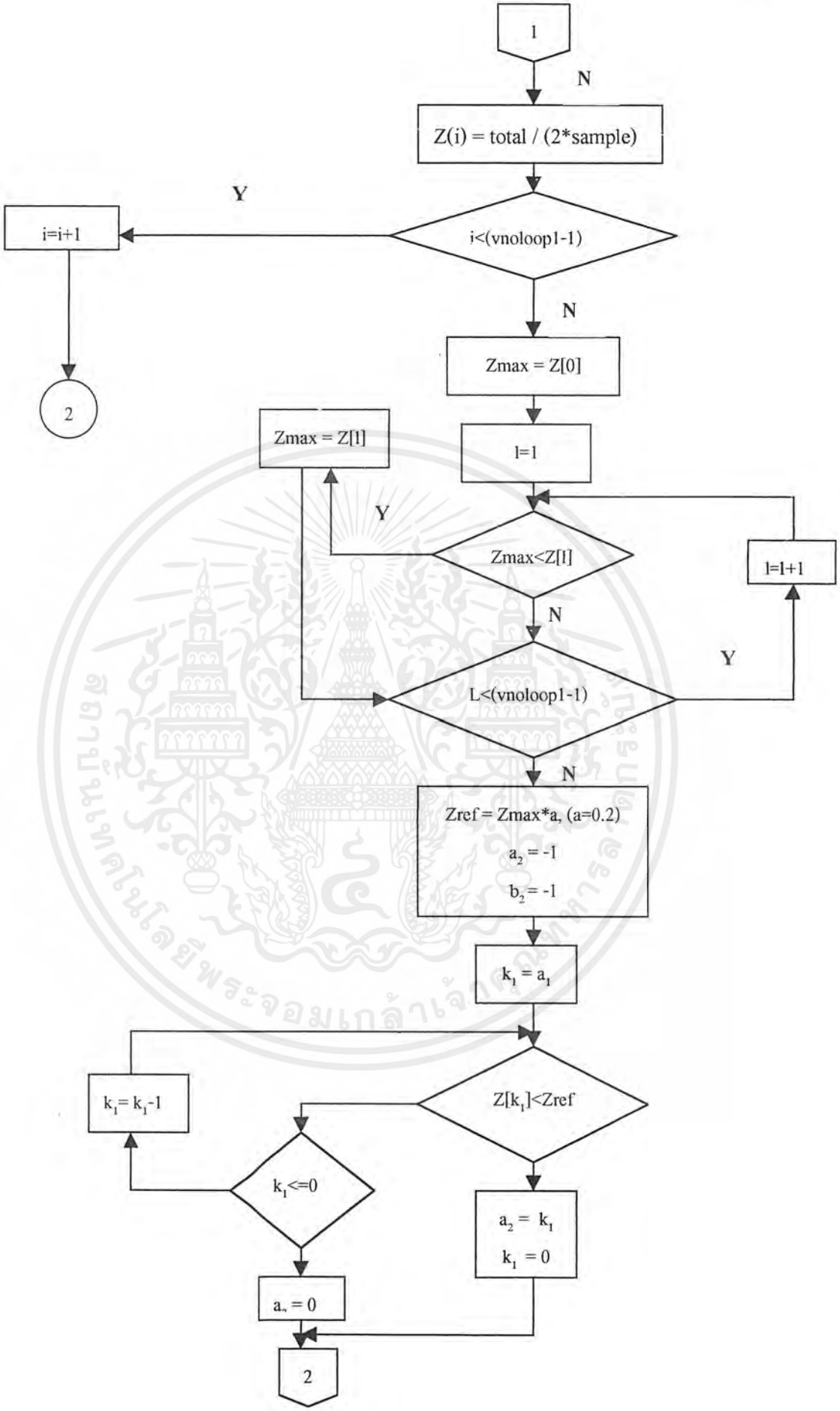


รูปที่ 3.4 แสดงขั้นตอนการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง โดยวิธีการเปรียบเทียบอัตราการผ่านศูนย์

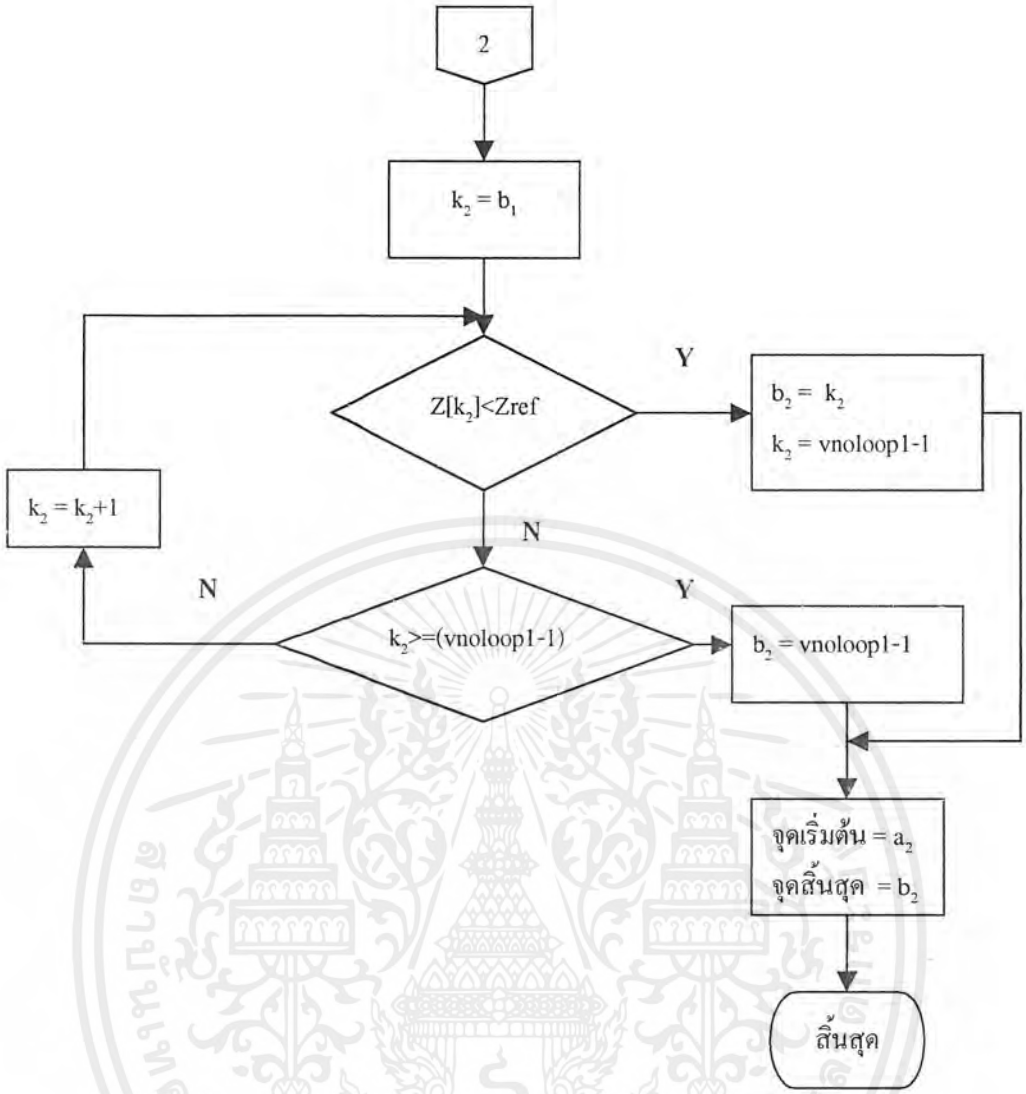
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

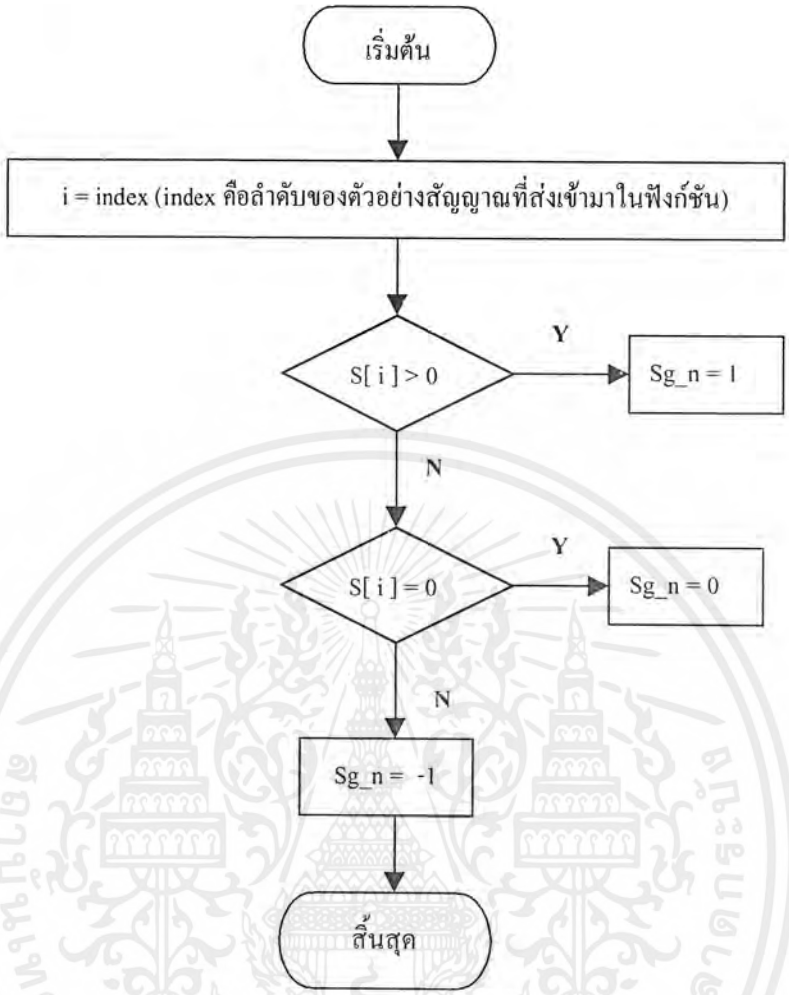


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของ สัญญาณเสียงด้วยวิธีการหาอัตราการตัดศูนย์

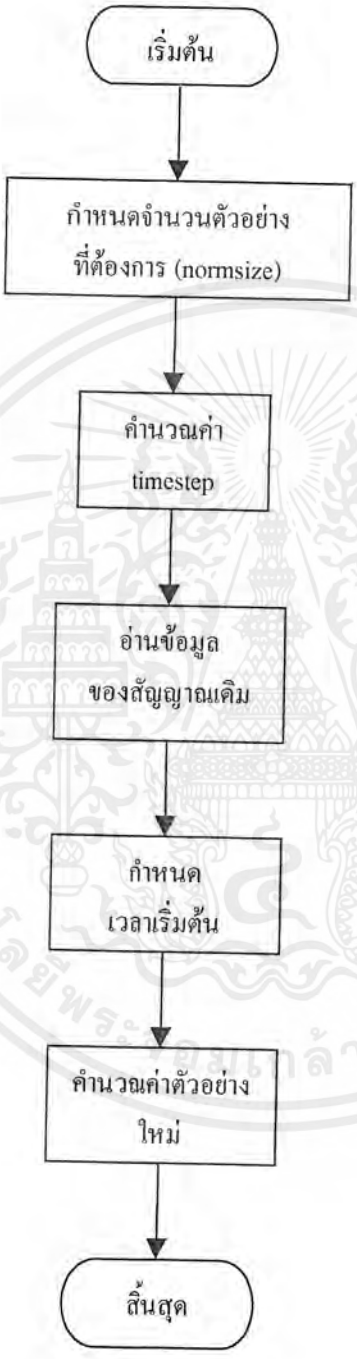
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 แสดงผังงานของฟังก์ชัน สำหรับการพิจารณาตัดกับแกนเวลาของรูปคลื่นสัญญาณ

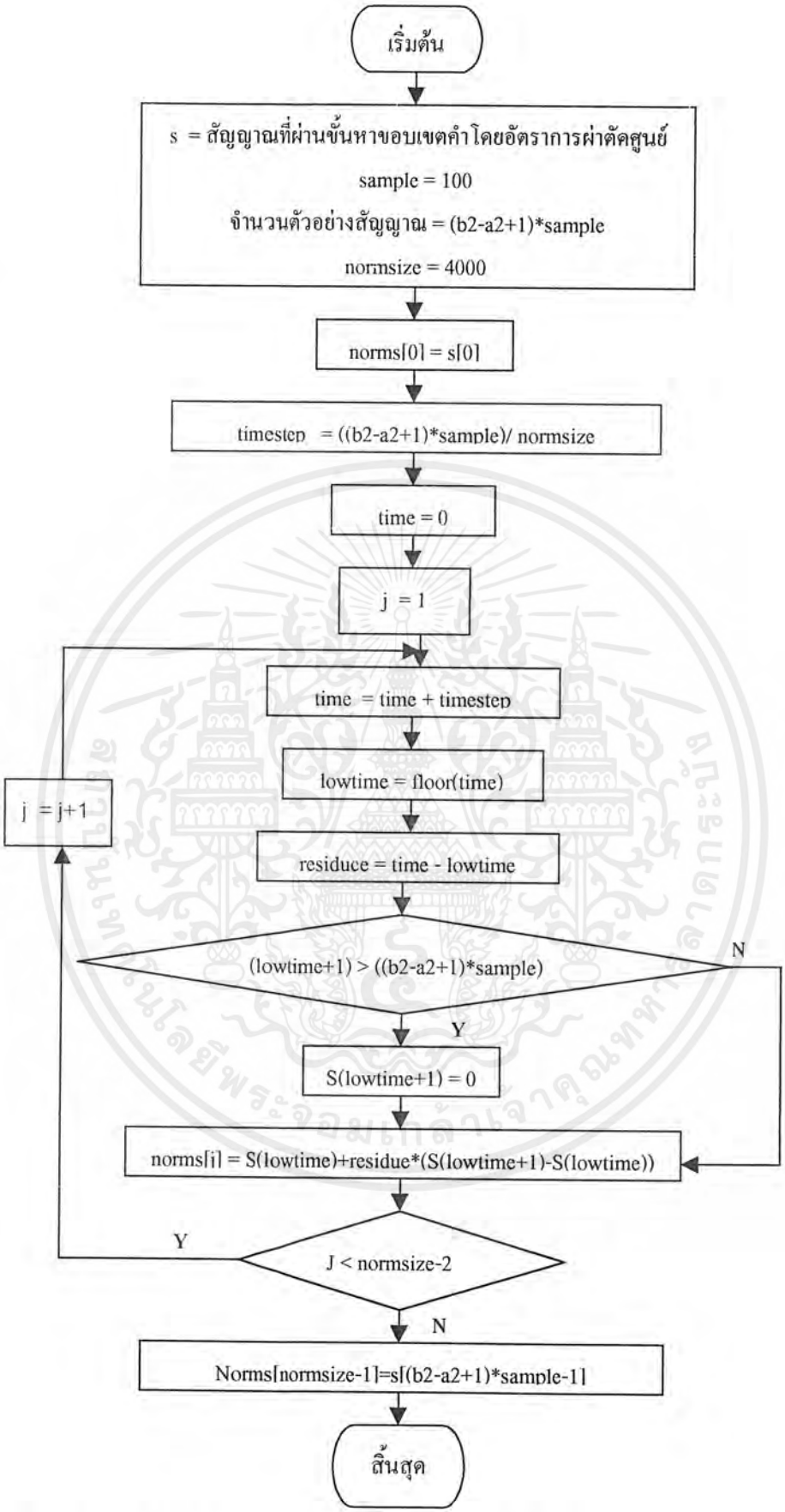
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 การนอร์มอลไลซ์ (Normalization) เนื่องจากสัญญาณเสียงพูดแต่ละคำมีความยาวไม่เท่ากัน จึงต้องมีการนอร์มอลไลซ์สัญญาณเสียงให้ยาวเท่ากัน ขั้นตอนของการนอร์มอลไลซ์สัญญาณเสียงพูดแสดงดังรูปที่ 3.7 และรูปที่ 3.8



รูปที่ 3.7 แสดงขั้นตอนการนอร์มอลไลซ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 แสดงผังงานของการนอร์มอลไลซ์สัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ length(S) คือ จำนวนตัวอย่างทั้งหมดของสัญญาก่อนนอร์มอลไลซ์

normsize คือ จำนวนตัวอย่างที่ต้องการ

S คือ ข้อมูลเสียงพูดเดิม

normS คือ ข้อมูลของเสียงพูดที่ผ่านการนอร์มอลไลซ์แล้ว

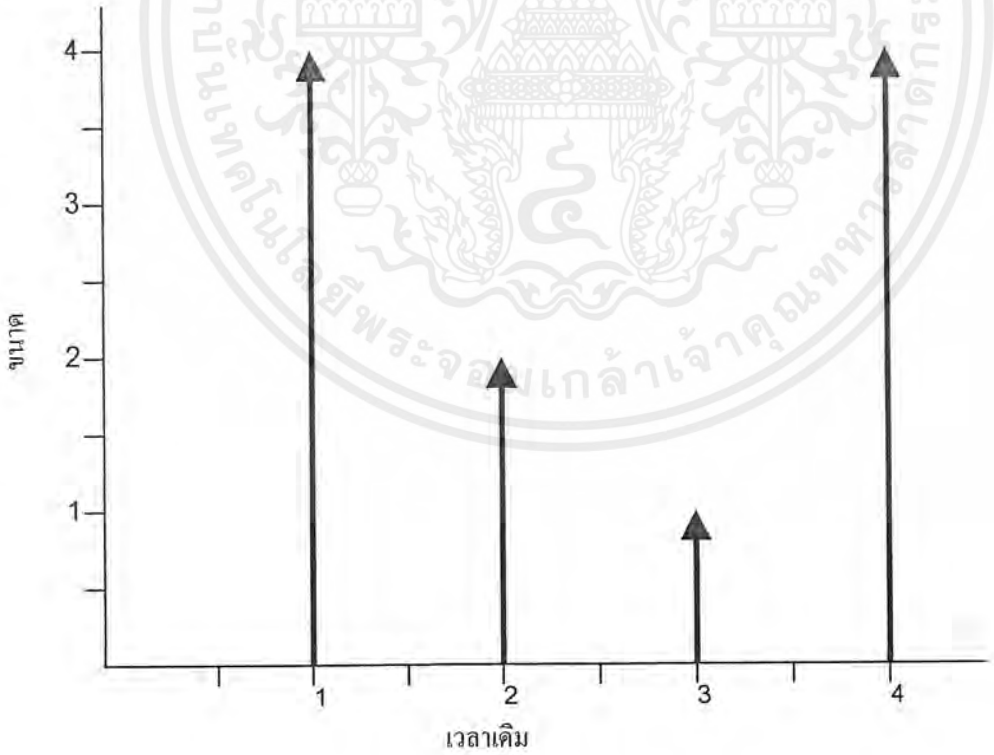
ค่าต่าง ๆ ในรูปที่ 3.8 คำนวณได้ดังนี้

timestep = (จำนวนตัวอย่างเดิม) / (จำนวนตัวอย่างที่ต้องการ)

normS(j) = S(lowmime)+residue* (S(lowtime+1)-S(lowtime))

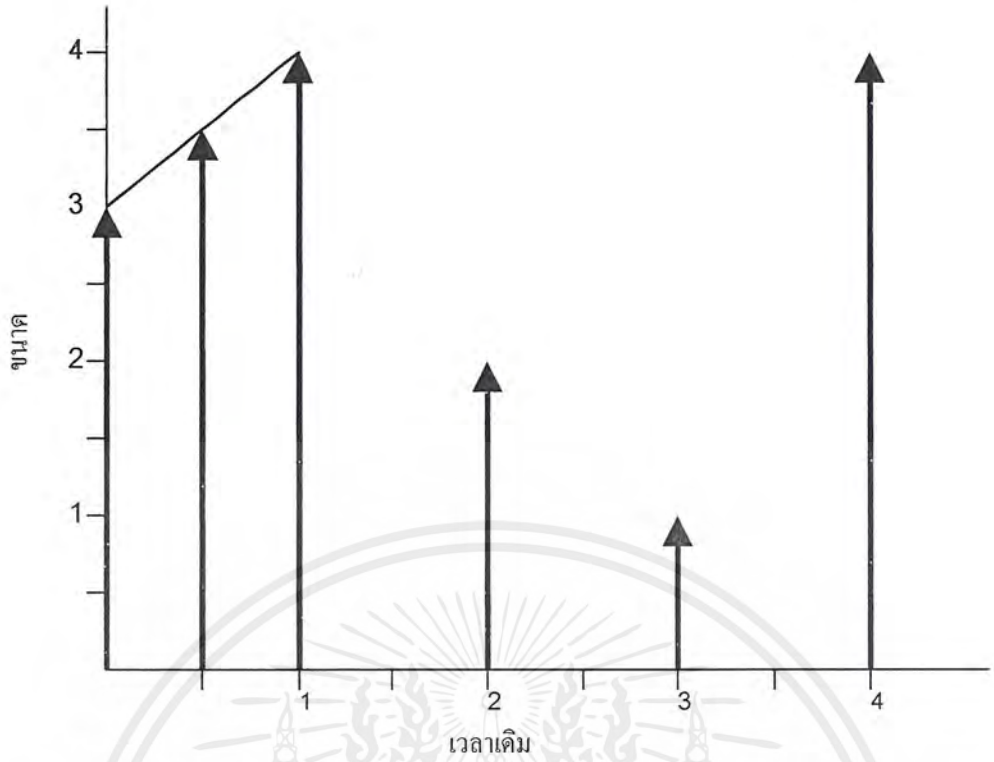
เมื่อ lowtime , residue คือจำนวนเต็มและเศษเหลือของค่า time ในแต่ละรอบของ j

ตัวอย่างของการนอร์มอลไลซ์สัญญาณเสียงพูด แสดงดังรูป 3.9 ซึ่งแสดงการนอร์มอลไลซ์สัญญาณขนาด 5 จุดข้อมูลในรูปที่ 3.9.ก เป็นสัญญาณขนาด 9 จุดข้อมูลในรูปที่ 3.9.ค ในขั้นแรกคำนวณค่า timestep ได้เท่ากับ 0.5 สัญญาณนอร์มอลไลซ์จุดแรกมีค่าเท่ากับสัญญาณเสียงพูดเดิมบนแกนเวลาเดิมที่เวลาเท่ากับ timestep เนื่องจากสัญญาณเสียงพูดเดิมเป็นสัญญาณไม่ต่อเนื่อง ทำให้ไม่ทราบค่าสัญญาณที่แท้จริงที่เวลาไม่เท่ากับจำนวนเต็ม จึงทำการประมาณค่าในช่วงเชิงเส้น (linear interpolation) จากสัญญาณเสียงพูดเดิม 2 จุด ที่อยู่ใกล้ที่สุดที่เวลาเป็นจำนวนเต็ม คือสัญญาณเสียงพูดที่เวลา 0 และที่เวลา 1 ดังแสดงในรูปที่ 3.9.ข สัญญาณนอร์มอลไลซ์ที่เวลาอื่นคำนวณจากวิธีเดียวกันนี้

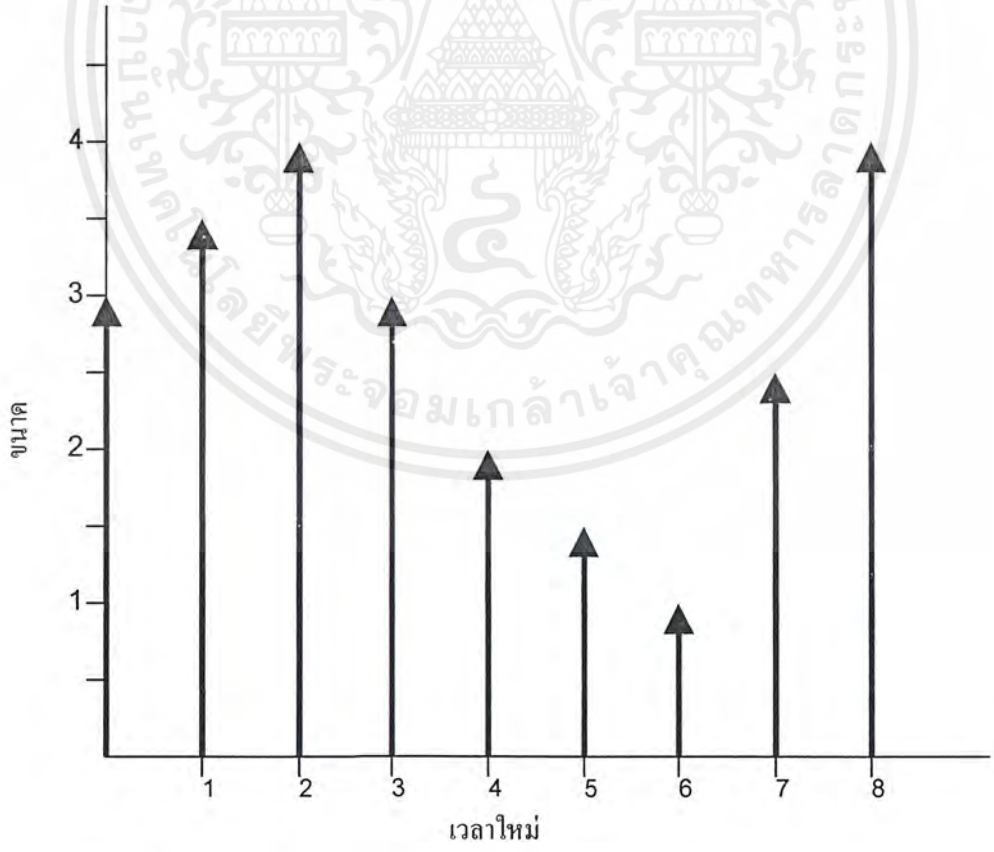


รูปที่ 3.9.ก แสดงสัญญาณเสียงพูดเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9.ข แสดงการประมาณค่าในช่วงเชิงเส้นสัญญาณนอร์มอลไลซ์จุด



รูปที่ 3.9.ค แสดงสัญญาณนอร์มอลไลซ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การหาค่าพารามิเตอร์ ซึ่งเป็นขั้นตอนการหาค่าพารามิเตอร์ สำหรับปริญญานิพนธ์นี้จะทำการหาค่าพารามิเตอร์ที่สำคัญของสัญญาณเสียง โดยจะใช้วิธีลิเนียร์พรีดิกทีฟ (Linera Predictive Coding - LPC) หาค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักต่อไป

การวิเคราะห์แบบลิเนียร์พรีดิกทีฟเป็นการวิเคราะห์เสียงพูดที่ดีวิธีหนึ่ง โดยการวิเคราะห์นี้ใช้ในการหาค่าพารามิเตอร์พื้นฐานของเสียง ซึ่งพารามิเตอร์ที่หาได้จากวิธีนี้มีความถูกต้องและใช้เวลาในการคำนวณไม่มากนัก

การหาค่าพารามิเตอร์ของเสียงนี้สำหรับลิเนียร์พรีดิกทีฟจะทำการหาค่าสัมประสิทธิ์แอลพีซี (a) โดยที่เราจะต้องเตรียมสัญญาณในการวิเคราะห์ก่อน ซึ่งขั้นตอนในการวิเคราะห์แสดงดังรูปที่ 3.12

การวิเคราะห์สามารถแบ่งเป็นขั้นตอนต่าง ๆ ดังนี้

3.2.1 การพรีเอมฟาซีส (Preemphasis)

3.2.2 การแบ่งช่วงสัญญาณ (Frame Blocking)

3.2.3 การวินโดว์ (Windowing)

3.2.4 การหาออโตคอร์รีเลชั่น (Autocorrelation Analysis)

3.2.5 การหาสัมประสิทธิ์แอลพีซี a (LPC Analysis)

3.2.6 การหาอัตราการขยาย G

3.2.7 การเปลี่ยนพารามิเตอร์แอลพีซี เป็นสัมประสิทธิ์เซปสตรัม

3.2.8 การเวทค่าพารามิเตอร์ (Parameter Weighting)

3.2.1 การพรีเอมฟาซีส

เนื่องจากสัญญาณเสียงพูด จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ ดังนั้นเพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน มีค่าค่อนข้างคงที่ เราจึงต้องมีการพรีเอมฟาซีส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้น นั่นคือ การพรีเอมฟาซีสเป็นการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (High pass filter) ซึ่งมักจะนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนดังนี้

$$H(z) = 1-az^{-1} \quad 0.9 \leq a \leq 1.0 \qquad (3.2.1)$$

สำหรับสัญญาณที่ผ่านการพรีเอมฟาซีสนี้ จะมีเอาท์พุทของสัญญาณ s'(n) ซึ่งมีความสัมพันธ์กับสัญญาณอินพุท s(n) โดยสามารถเขียบสมการได้ดังนี้

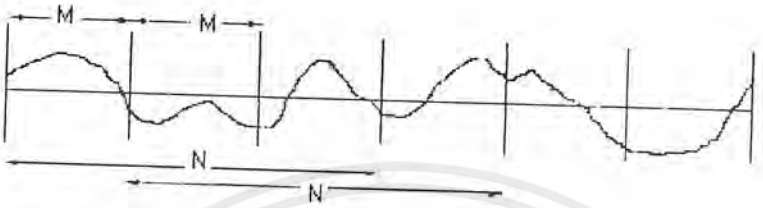
$$s'(n) = s(n)-as(n-1) \qquad (3.2.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยิ่งค่า a มีค่าใกล้ 1 เท่าใด ความถี่สูงก็จะถูกขยายมากขึ้นเท่านั้น ค่า a ที่นิยมใช้ในการหาค่าสัมประสิทธิ์แอลพีซี คือ a = 15/16 = 0.9375

3.2.2 การแบ่งช่วงสัญญาณ

สัญญาณที่ผ่านการพรีเอมฟาซีสแล้ว s'(n) จะถูกแบ่งออกเป็นช่วง ๆ หรือเฟรมโดยมีช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างของสัญญาณ ดังรูปที่ 3.10



รูป 3.10 แสดงการแบ่งช่วงที่ใช้ในการวิเคราะห์

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า M ใหญ่กว่า N ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N จะทำให้ตัวอย่างของสัญญาณทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่จะทำให้การคำนวณช้าลง

ในการวิเคราะห์หาค่าสัมประสิทธิ์แอลพีซี ถ้าทำการกำหนดให้สัญญาณเสียงที่นำมาวิเคราะห์เมื่อผ่านการพรีเอมฟาซีสแล้ว มีการแบ่งช่วงเป็นช่วง ๆ โดยมีอันดับของช่วงเป็น l^{th} โดยสัญญาณในแต่ละเฟรมเป็น $x_l(n)$ และมีจำนวนช่วงทั้งหมดเป็น L เฟรม ดังนั้น

$$x_l(n) = s'(Ml+n) \qquad (3.2.3)$$

$$n = 0, 1, \ldots, N-1$$

$$l = 0, 1, \ldots, L-1$$

และทำการกำหนดให้ค่า N = 240 ตัวอย่าง และ M = 80 ตัวอย่าง ตามลำดับ เมื่อมีการสุ่มตัวอย่างของสัญญาณที่ 8 กิโลเฮริตซ์

3.2.3 การวินโดว์

พิจารณาช่วงสัญญาณ N ตัวอย่าง ของช่วงใด ๆ ที่ตัดมาวิเคราะห์ (รูปที่ 3.11) จะเห็นว่าที่ขอบของเฟรมนี้มีความไม่ต่อเนื่องของสัญญาณ ถ้าพิจารณาในโดเมนความถี่สูง ก็จะมีความถี่สูงเกิดขึ้น ดังนั้นเพื่อที่ลดองค์ประกอบทางความถี่สูงเหล่านี้ เราจะคูณด้วยฟังก์ชันวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบและไม่ทำให้สเปคตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ถ้ากำหนดให้วินโดว์เป็น $w(n)$, $0 \leq n \leq N-1$ แล้วสัญญาณที่ผ่านการวินโดว์แล้วจะได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต่อเนื่องของสัญญาณที่ขอบและไม่ทำให้สเปคตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ถ้ากำหนดให้วินโดว์เป็น w(n) , $0 \leq n \leq N-1$ แล้วสัญญาณที่ผ่านการวินโดว์แล้วจะได้

$$x_l'(n) = x_l(n)w(n), \qquad 0 \leq n \leq N-1 \tag{3.2.4}$$

รูปที่ 3.11 แสดงส่วนของสัญญาที่ตัดมาวิเคราะห์

ในที่นี้จะใช้ฟังก์ชันวินโดว์แฮมมิ่ง (Hamming window function) ซึ่งนิยามโดยสมการดังนี้ คือ

$$w(n) = 0.54-0.46\cos(2\pi n/N-1), \quad 0 \leq n \leq N-1 \tag{3.2.5}$$

3.2.4 การหาออโตคอร์รีเลชั่น

สัญญาณใดในแต่ละเฟรมที่ผ่านวินโดว์แล้ว จะถูกนำมาผ่านการหาค่าออโตคอร์รีเลชั่น

$$R(m) = \sum_{m=0}^{N-1-m} x'(n)x'(n+m) \tag{3.2.6}$$

เรียกวิธีการนี้ว่าการประมาณเชิงเส้นอันดับ p โดยเรียกอันดับ p ว่าอันดับของการวิเคราะห์แอลพีซี โดยทั่วไปค่า p จะมีค่าระหว่าง 8 ถึง16

3.2.5 การหาค่าสัมประสิทธิ์แอลพีซี

ในการวิเคราะห์หาค่า LPC parameters ที่นิยมใช้กันได้รับการพัฒนาโดย Levinson และ Durbin ซึ่งเรียกว่า Levinson-Durbin recursive method มีขั้นตอนดังนี้

$$E_0 = R(0)$$

เมื่อ m=1,2,…,p

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$
\begin{aligned}
K_m &= -\left[R(m) + \sum_{k=1}^{m-1} a_{m-1}(k) R(m-k) \right] \Big/ E_{m-1} \\
a_m(m) &= K_m \\
a_m(k) &= a_{m-1}(k) + K_m a_{m-1}(m-k), 1 \le k \le m-1 \\
E_m &= \left(1 - K_m^2\right) E_{m-1}
\end{aligned}
\tag{3.2.7}
$$

สัมประสิทธิ์ $\{a_m(k), 1 \le k \le m\}$ ก็คือค่าสัมประสิทธิ์แอลพีซี (LPC parameter) ที่อันดับ m ซึ่งผลลัพธ์สุดท้ายคือ $\{a_m(k), 1 \le k \le p\}$ โดยที่ p คือดันดับของสัมประสิทธิ์ที่ต้องการ

3.2.6 การหาอัตราการขยาย G

จากสมการ 2.5.6 จะได้ว่า

$$
\begin{aligned}
e(n) &= G * u(n) \\
E_n &= \sum_{m=0}^{N-1} e^2(m) = G \sum_{m=0}^{N-1} u^2(m)
\end{aligned}
$$

จากสมการ 2.5.11 จะได้ว่า

$$
\begin{aligned}
E_n &= \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \\
&= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)
\end{aligned}
\tag{3.2.8}
$$

และจากสมการ 3.3.8 เราสามารถหาค่า G โดยตรงจาก

$$
G^2 = \frac{R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)}{\sum_{m=0}^{N-1} u^2(m)}
\tag{3.2.9}
$$

3.2.7 การเปลี่ยนพารามิเตอร์แอลพีซี เป็นสัมประสิทธิ์เซปสตรัม

ในการรู้จำเสียงพูดนั้น สัมประสิทธิ์เซปสตรัมนี้เป็นพารามิเตอร์ที่มีลักษณะน่าเชื่อถือได้ดีว่าสัมประสิทธิ์ LPC ทั้งยังมีความสัมพันธ์ใกล้ชิดกับการรับรู้เสียง ตามความรู้สึกของมนุษย์โดยแท้จริง สัมประสิทธิ์เซปสตรัมสามารถหาได้โดยตรงจากสัมประสิทธิ์แอลพีซี ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$
\begin{aligned}
C_0 &= \ln G \\
C_m &= a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)C_k a_{m-k}, 1 \le m \le p \\
C_m &= \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)C_k a_{m-k}, m > p \\
Q &\approx \frac{3}{2}p
\end{aligned}
\qquad (3.2.10)
$$

3.2.8 การเวทค่าพารามิเตอร์ (Parameter Weighting)

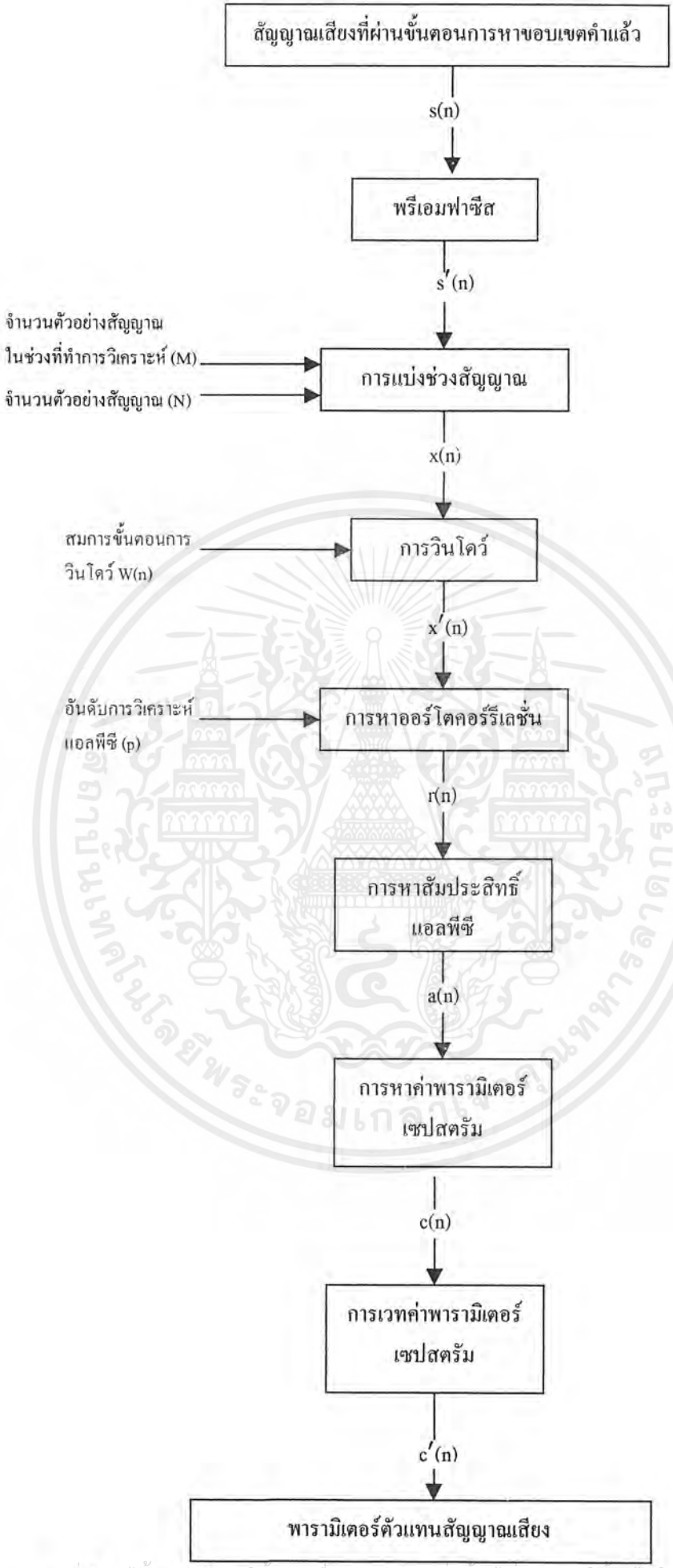
เนื่องจากสัมประสิทธิ์เซปสตรัมที่ได้นั้น ช่วงลำดับต้น ๆ และลำดับท้าย ๆ ของเฟรมที่นำมาวิเคราะห์จะเกิดความคลาดเคลื่อนมากกว่าบริเวณส่วนอื่น เพราะฉะนั้น จึงทำการถ่วงน้ำหนัก เพื่อลดค่าความคลาดเคลื่อนดังกล่าวนี้ ด้วยฟังก์ชันเวทติ่ง ดังนี้คือ

$$
W_m = \left[1 + \frac{Q}{2}\sin\left(\frac{\pi m}{Q}\right)\right], 1 \le m \le Q \qquad (3.2.11)
$$

จะได้

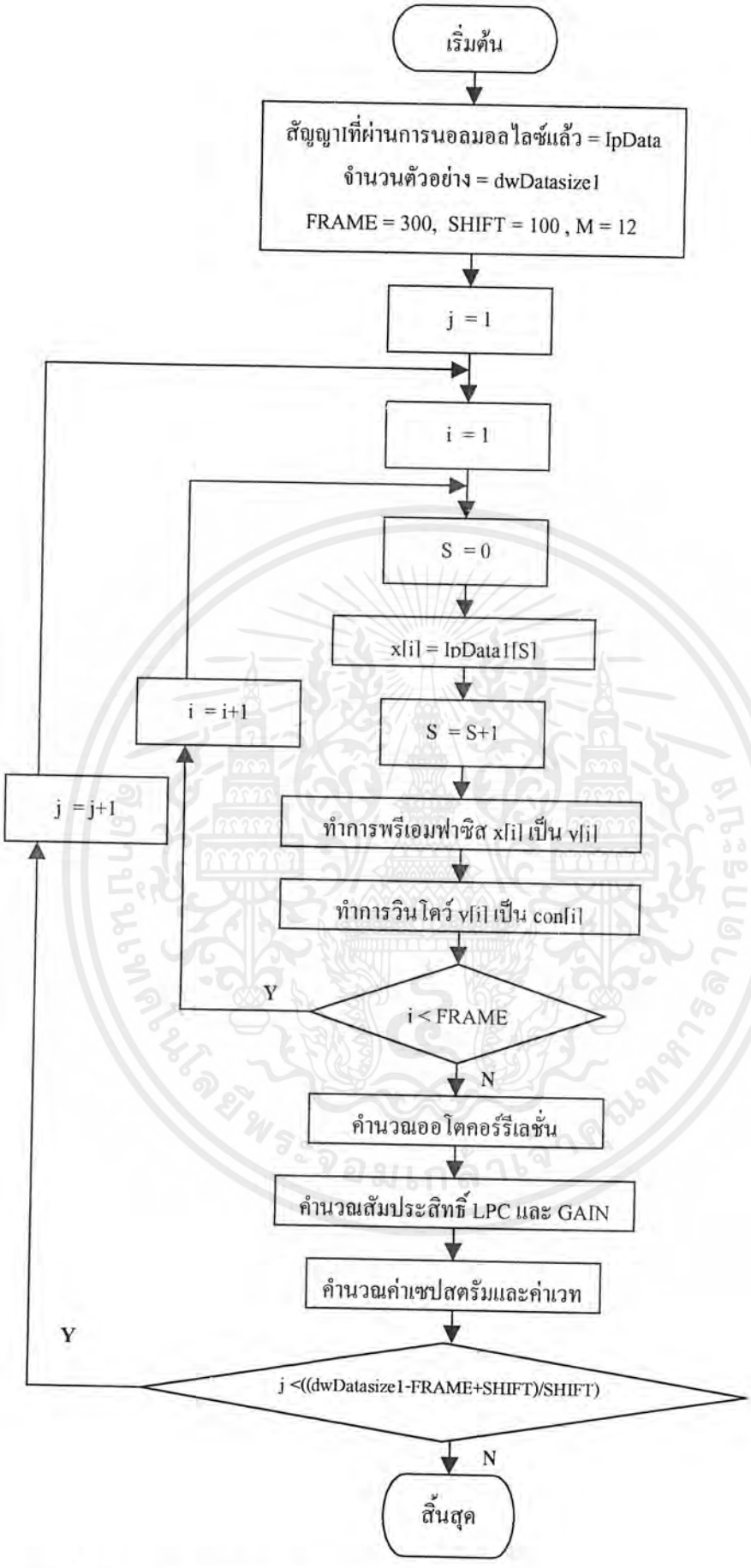
$$
C_m = C_m * W_m
$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.12 แสดงขั้นตอนการเตรียมสัญญาณในการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.13 แสดงผังงานของโปรแกรม LPC

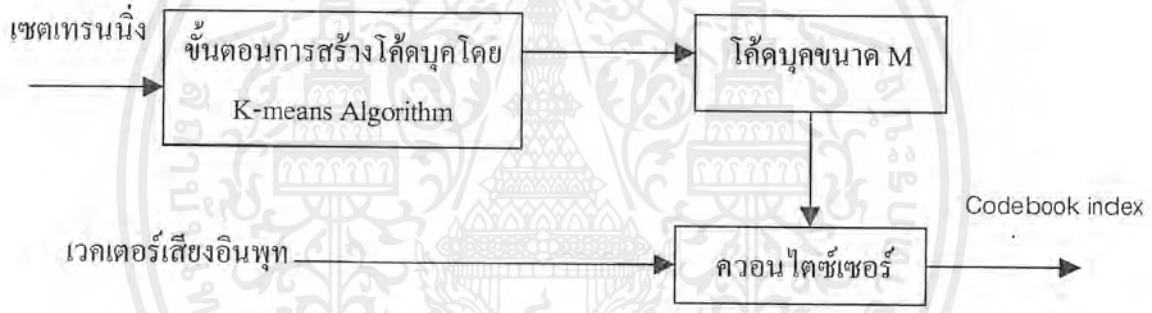
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 เวคเตอร์ ควอนไตซ์เซชัน การรับรู้เสียงพูด เป็นลักษณะหนึ่งของการรับรู้รูปแบบ (Pattren Recognition) คือจะเป็นการเปรียบเทียบระหว่าง แบบทดสอบ (Test Pattern) กับ แบบอ้างอิง (Reference Pattern) ซึ่งเป็นรูปแบบที่เราทราบและเก็บไว้ล่วงหน้า

ขั้นตอนในการรับรู้แบ่งเป็น 2 ขั้นตอน ดังนี้

ก.ขั้นเรียนรู้ (Learning) จะเป็นการสร้างกลุ่มของแบบอ้างอิง ในการรับรู้เสียงพูด ในขั้นนี้จะทำการวิเคราะห์เสียงพูดก่อน แล้วเก็บลักษณะของเสียงในรูปของพารามิเตอร์ ดังได้กล่าวไว้แล้วในบทที่ 3 และใช้เวคเตอร์ ควอนไตซ์เซชัน ในการสร้างแบบอ้างอิง เพื่อใช้เปรียบเทียบในขั้นตอนต่อไป

ข.ขั้นรับรู้ (Recognition) จะเป็นการทดสอบการรับรู้ระหว่างแบบอ้างอิง กับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์แบบทดสอบ กับแบบอ้างอิงทั้งหมด แบบอ้างอิงที่เลือกคือแบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด



รูปที่ 3.1.4 แสดงบล็อกไดอะแกรมของเวคเตอร์ ควอนไตซ์เซชัน

การทำงานของเวคเตอร์ควอนไตซ์เซชัน

1. เซตของเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ $v_1, v_2, v_3, \ldots, v_L$ ซึ่งเป็นเซตขนาดใหญ่จะถูกนำมาใช้ในการสร้างเซตของโค้ดบุค เพื่อใช้เป็นตัวแทนเวคเตอร์สเปกตรัมของเสียงที่สังเกตได้ ขนาดโค้ดบุคของเวคเตอร์ควอนไตซ์เซชันคือ $M = 2^B$ เวคเตอร์ (เราเรียกว่า B – bit โค้ดบุค) ซึ่งเราต้องการให้ L>>M เพื่อที่จะหาเซตของ M โค้ดบุคที่ดีที่สุด ในทางปฏิบัตินั้นเพื่อให้เวคเตอร์ ควอนไตซ์เซชันโค้ดบุค สามารถทำงานได้ดีนั้น ขนาด L อย่างน้อยที่สุดที่เราสามารถใช้ได้ คือ 10M

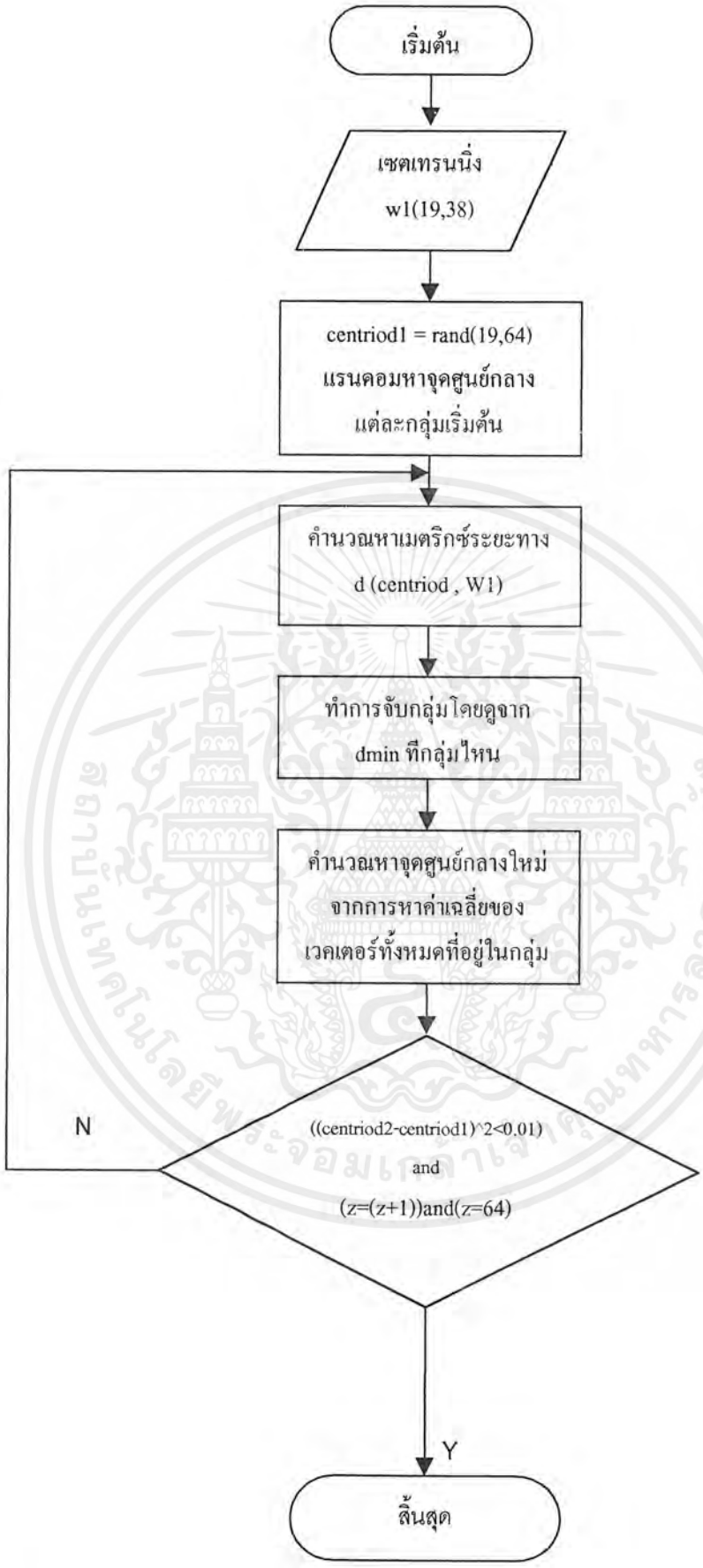
2. การวัดความเหมือนหรือระยะทางระหว่างเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ 2 เวคเตอร์ ดังนั้นสามารถจัดกลุ่มเซตของเวคเตอร์เทรนนิ่ง เพื่อสร้างโค้ดบุค เราแสดงระยะทางสเปกตรัม $d(v_i, v_j)$ ระหว่าง 2 เวคเตอร์ v_i และ v_j คือ d_{ij}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ฟังก์ชันย่อยในการคำนวณหาจุดศูนย์กลาง เพื่อแบ่งแยกเซตของเทรนนิ่งจำนวน L เวคเตอร์ ไปเป็น M กลุ่ม และ M โค้ดบุค ที่ได้คือจุดศูนย์กลางของแต่ละกลุ่ม

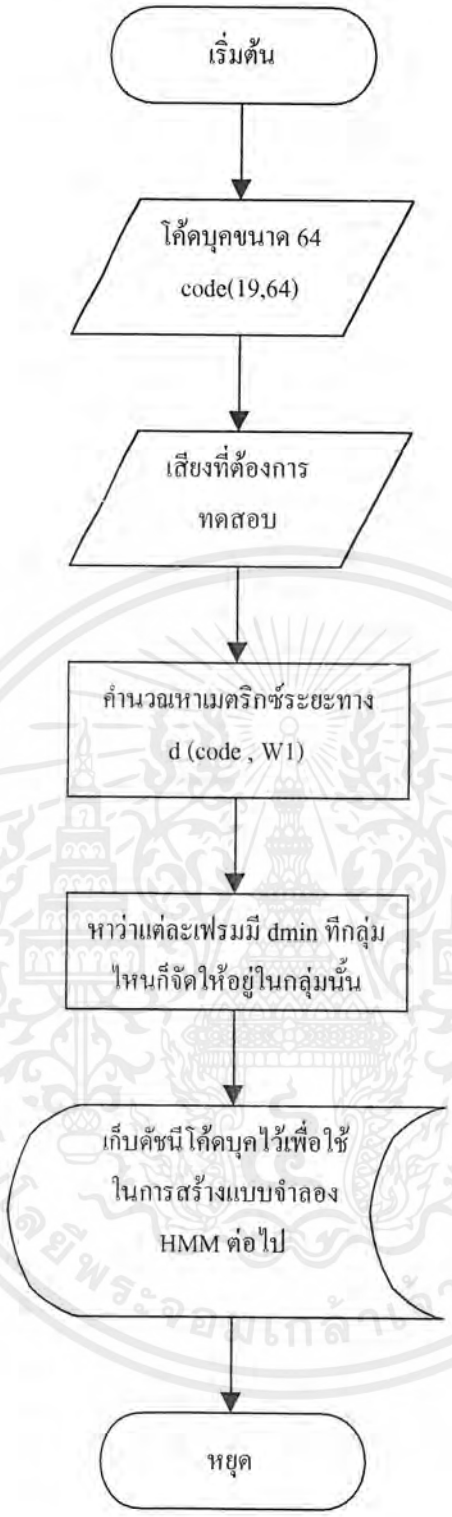
4. ฟังก์ชันย่อยในการแบ่งแยกสำหรับเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ ซึ่งจะเลือกโค้ดบุคที่ใกล้เคียงที่สุดให้กับเวคเตอร์อินพุท และผลลัพธ์ที่ได้คือ ดัชนีโค้ดบุคซึ่งเป็นตัวแทนของสเปกตรัม มักจะใช้กฎ Nearest Neighbour ในการพิจารณษ ส่วนนี้เป็นหัวใจหลักของควอนไตซ์เซอร์ ซึ่งอินพุท คือ เวคเตอร์สเปกตรัมของเสียงและเอาท์พุท คือ ดัชนีโค้ดบุคที่มีความเหมือนกับเวคเตอร์เสียงอินพุทมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.15 Flow Chart แสดงการทำงานของโปรแกรม ส่วนโค้ดบุค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.16 Flow Chart แสดงการทำงานของโปรแกรม ส่วน Compair

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 ภาคการตัดสินใจ เมื่อทำการเปรียบเทียบความคล้ายคลึงระหว่างสัญญาณแบบทดสอบกับสัญญาณแบบอ้างอิง “เดิน”,”วิ่ง”,”ยิ้ม”,”หาว” แล้ว หลังจากนั้นก็ทำการเลือกผลที่ได้จากการเปรียบเทียบในภาคที่แล้วที่น้อยที่สุด ตามกฎการตัดสินใจ นั่นคือการติดสินใจในการจำแนกเสียงพูดว่าเสียงที่นำมาทดสอบเป็นคำว่าอะไรนั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

การทดลองจะแบ่งเป็น 3 ส่วนดังต่อไปนี้

4.1 เปรียบเทียบความคล้ายคลึงของเสียง โดยค่าพารามิเตอร์ ซึ่งได้แก่สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก และค่าความผิดพลาด

4.2 การทดสอบการจดจำเสียงพูดโดยแยกไปตามเงื่อนไขต่าง ๆ

4.3 การแสดงหน้าจอ และวิธีการใช้งานโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์

ในส่วนที่ 4.1 แสดงผลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก และค่าความผิดพลาดในรูปของกราฟโดยมีรายละเอียดดังต่อไปนี้

กราฟแสดงค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก

- แกน x แทน Weight Cepstrum Number
- แกน y แทน Weight Cepstrum

กราฟแสดงค่าความผิดพลาด

- แกน x แทน ลำดับของแบบอ้างอิง
- แกน y แทน ค่าความผิดพลาด

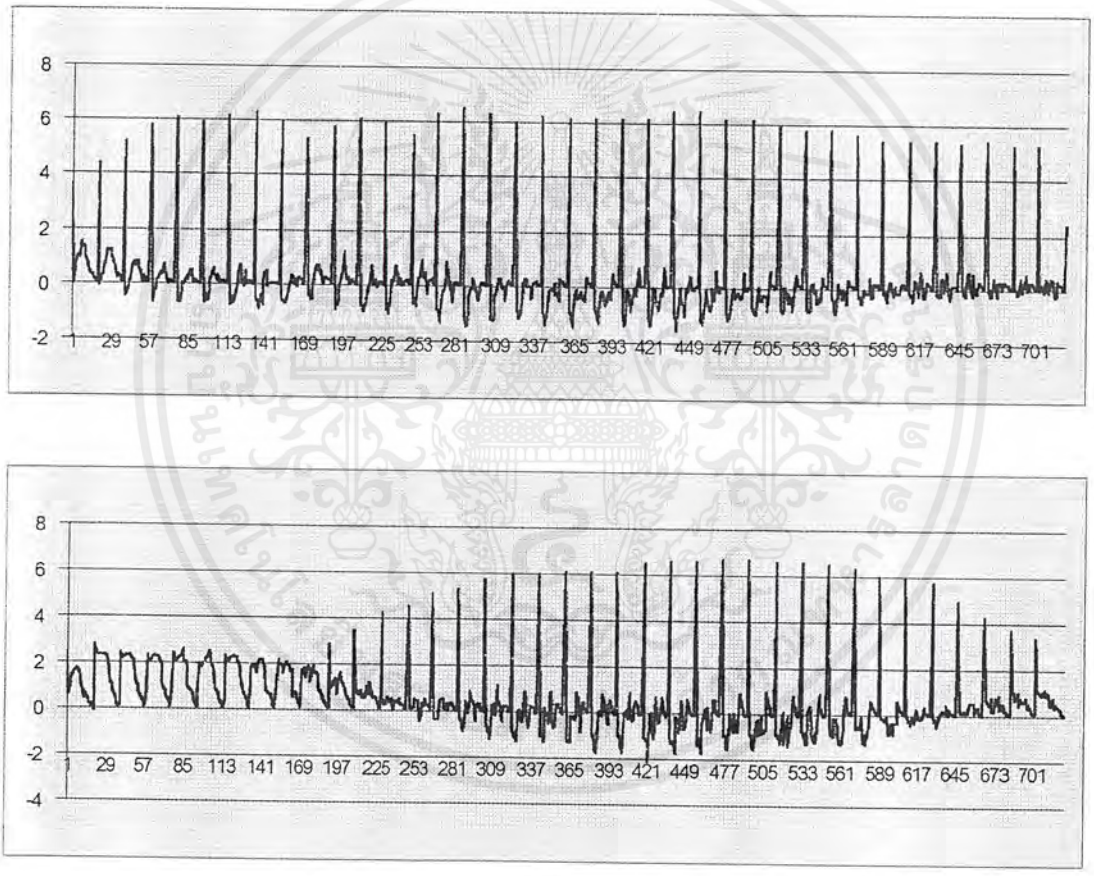
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1 การเปรียบเทียบค่าความคล้ายคลึงของเสียง

การเปรียบเทียบค่าความคล้ายคลึงของเสียงโดยค่าพารามิเตอร์ซึ่งได้แก่ สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักและค่าความผิดพลาด แบ่งออกเป็น 2 ส่วน ได้แก่

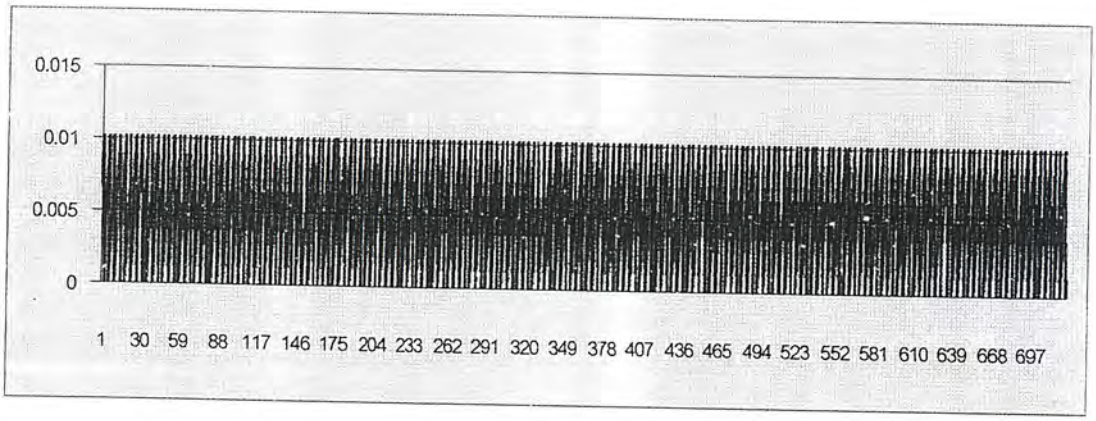
4.1.1 เป็นการเปรียบเทียบค่าความคล้ายของเสียงโดยใช้ค่าพารามิเตอร์ ซึ่งได้แก่สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักและค่าความผิดพลาด โดยแบบอ้างอิงที่ 1 (กลุ่ม A) ประกอบด้วย ข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ที่ได้มาจาก จำนวนแบบอ้างอิงที่สร้างจากนักศึกษาชายจำนวน 2 คน และนักศึกษาหญิงจำนวน 4 คน โดยที่แต่ละคนได้ทำการพูด คำว่า เดิน,วิ่ง,ยิ้ม,หาว อย่างละ 5 ครั้ง แสดงได้ดังกราฟต่อไปนี้

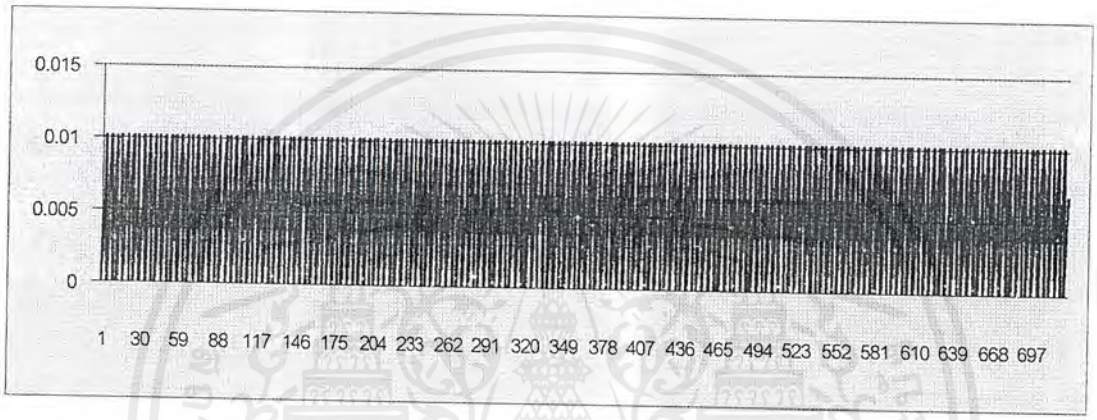
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า เดิน มีดังนี้



รูปที่ 4.1 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

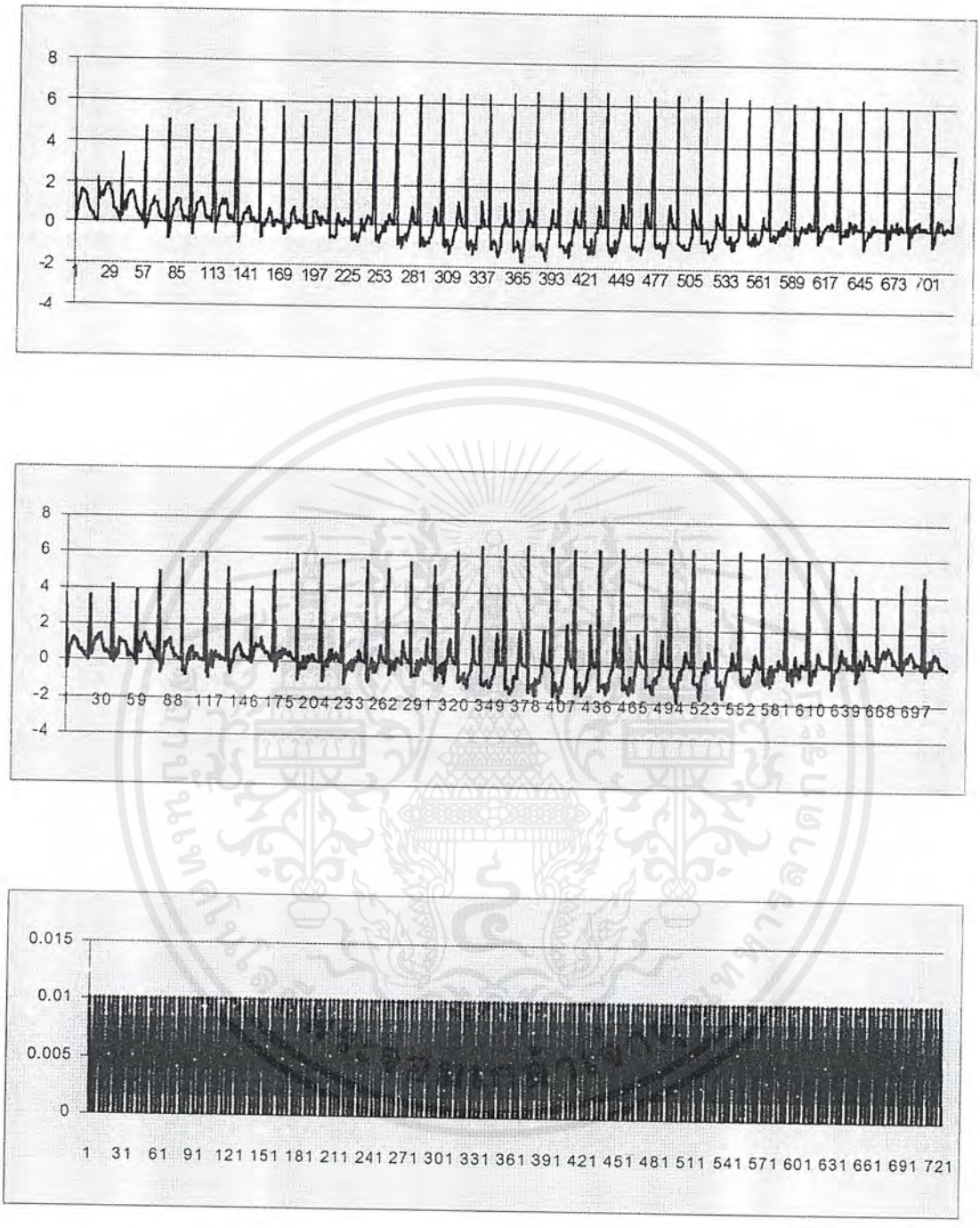




รูปที่ 4.1 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

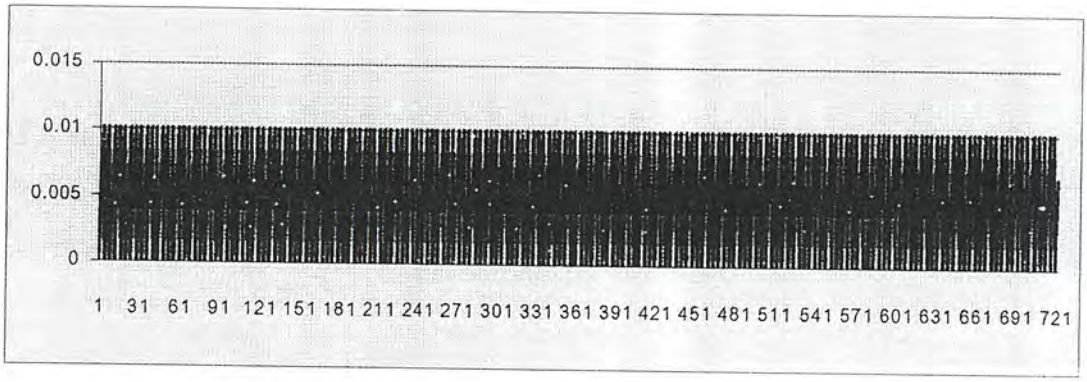
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

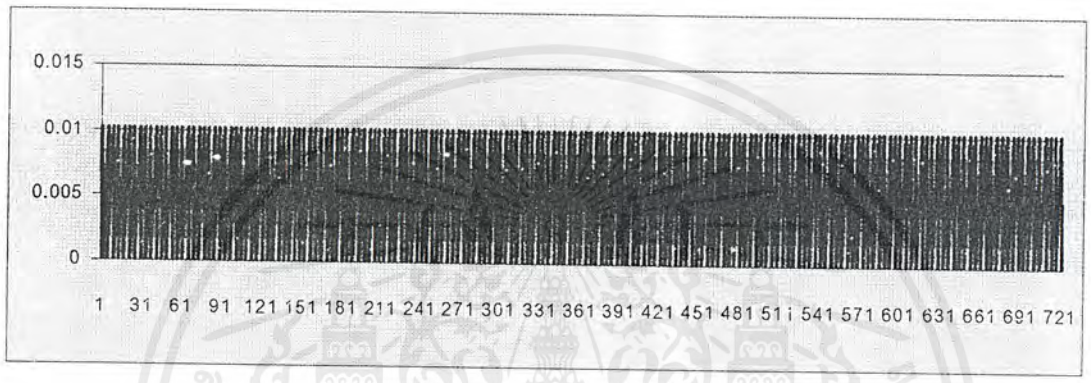
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า วิ่ง



รูปที่ 4.2 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง

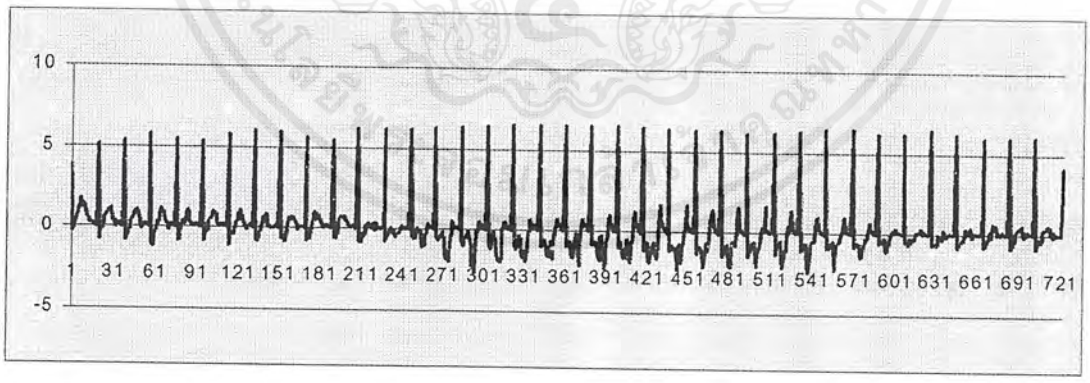
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





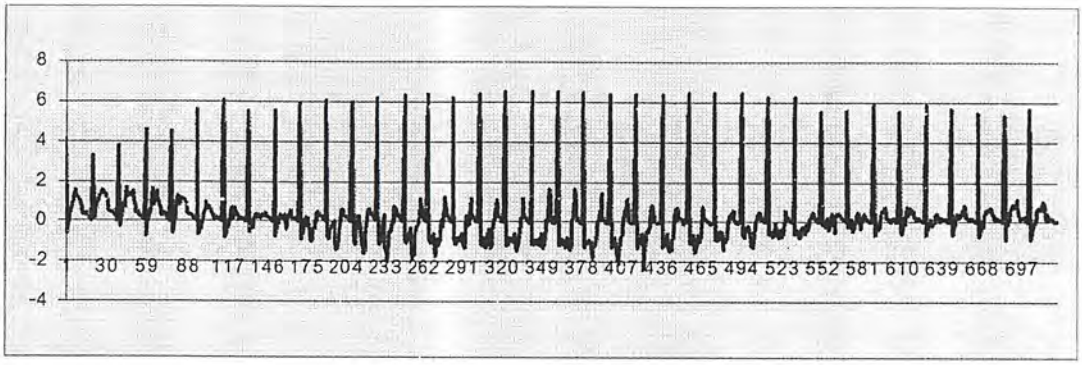
รูปที่ 4.2 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

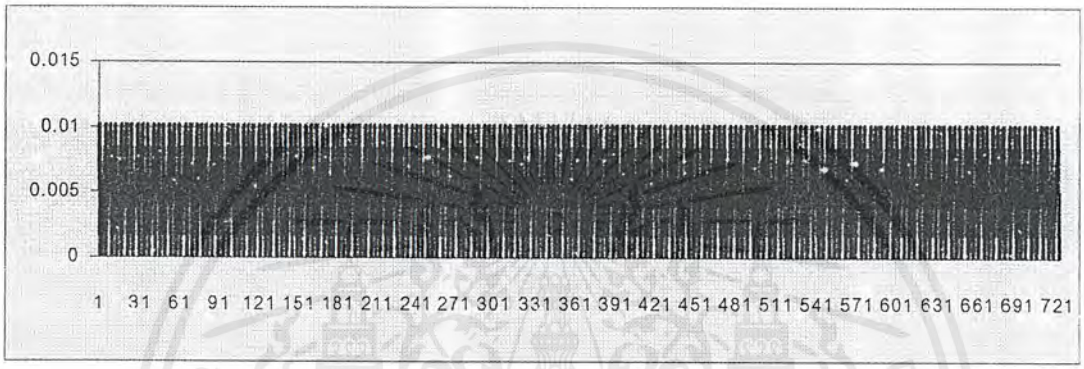
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า ยิ้ม

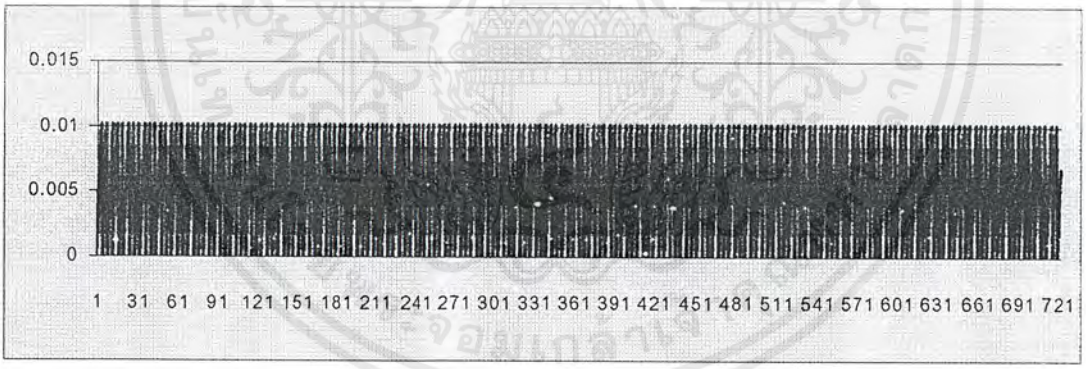


รูปที่ 4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

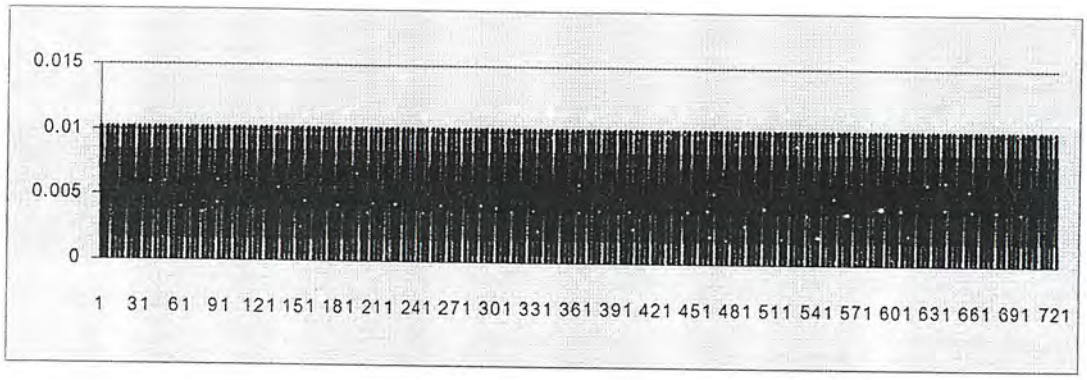






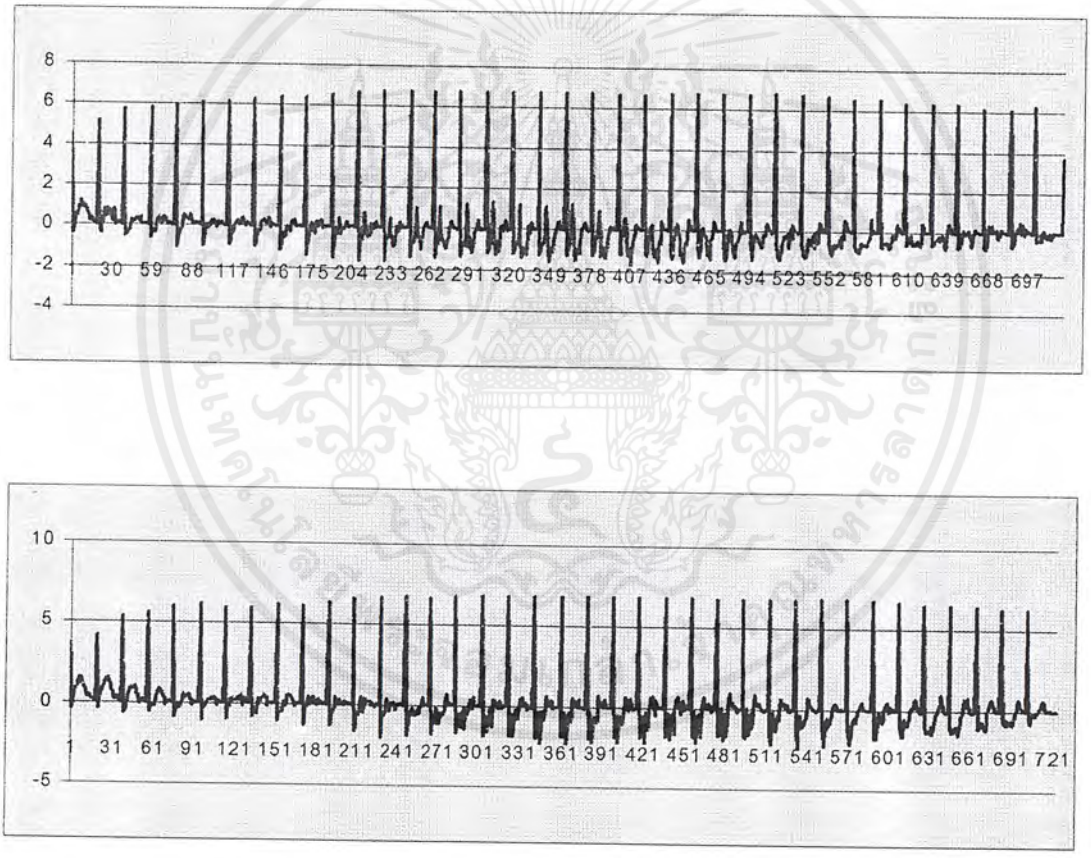
รูปที่ 4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



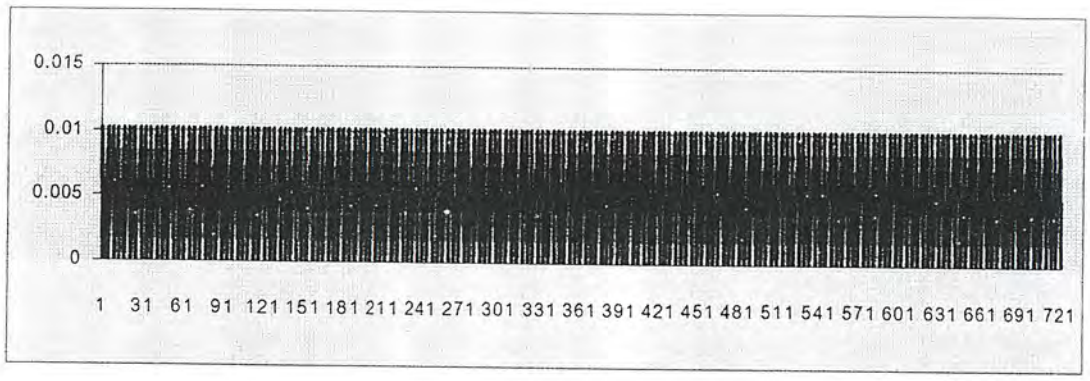
รูปที่ 4.3 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

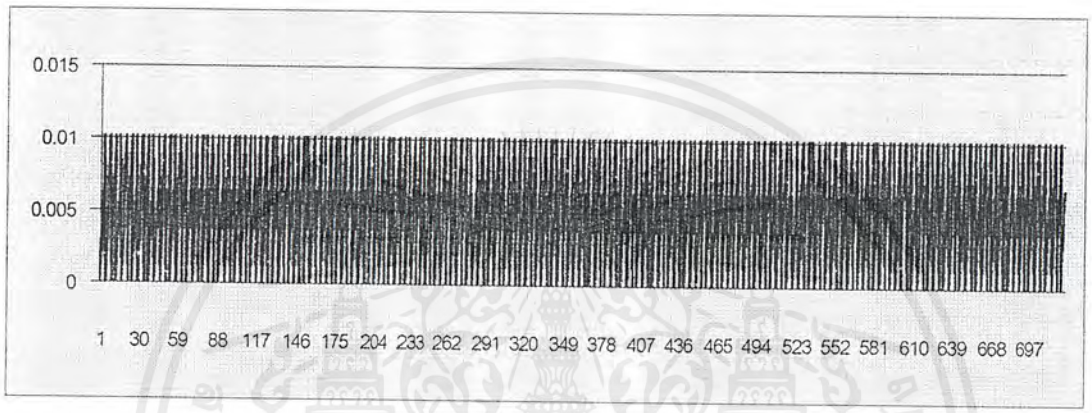
กราฟแลดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า หาว

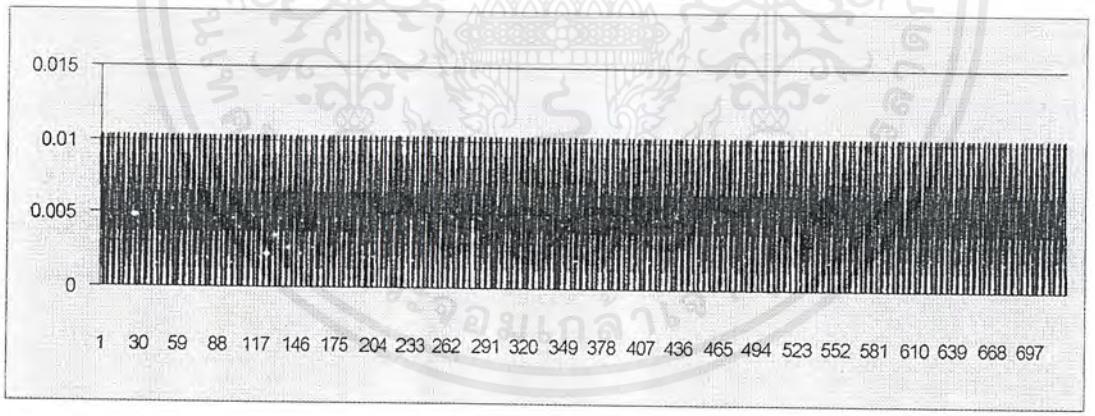


รูปที่ 4.4 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปปสตรัมถ่วงน้ำหนักของคำว่า หาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



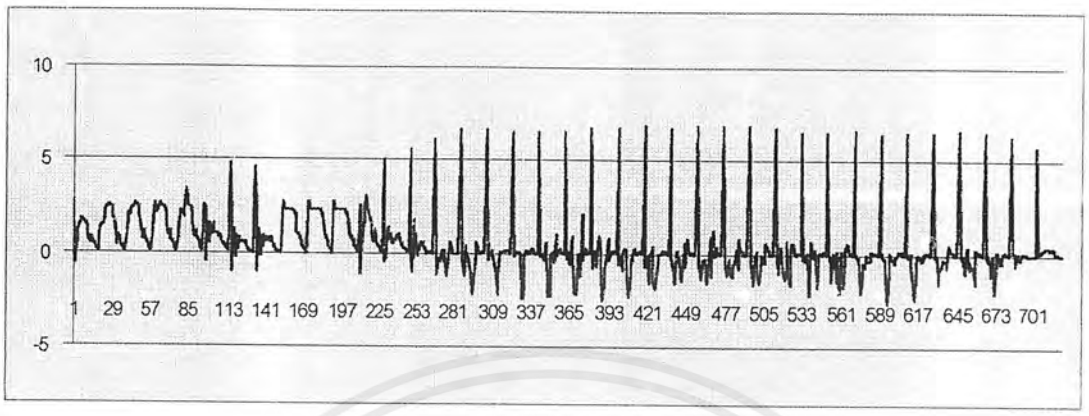




รูปที่ 4.4 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

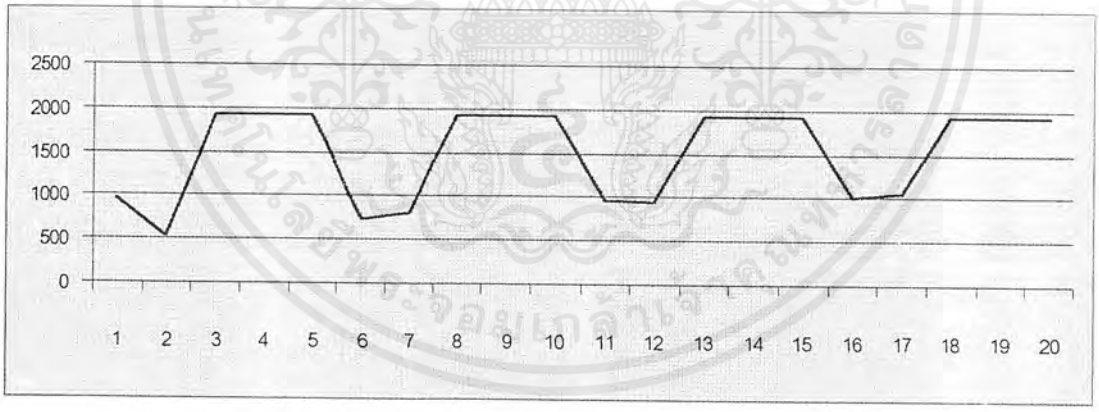
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **เดิน** แสดงได้ดังนี้



รูปที่ 4.5 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า เดิน

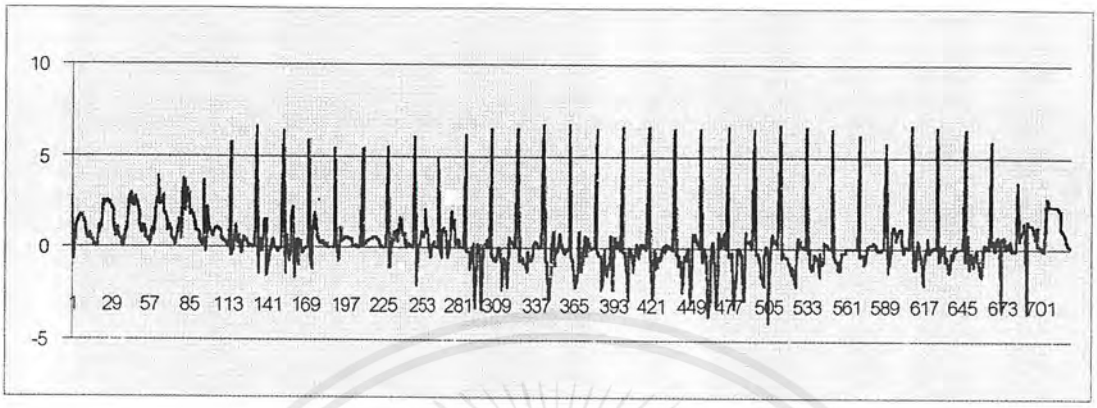
เมื่อนำ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า เดิน มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ



รูปที่ 4.6 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า เดิน กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

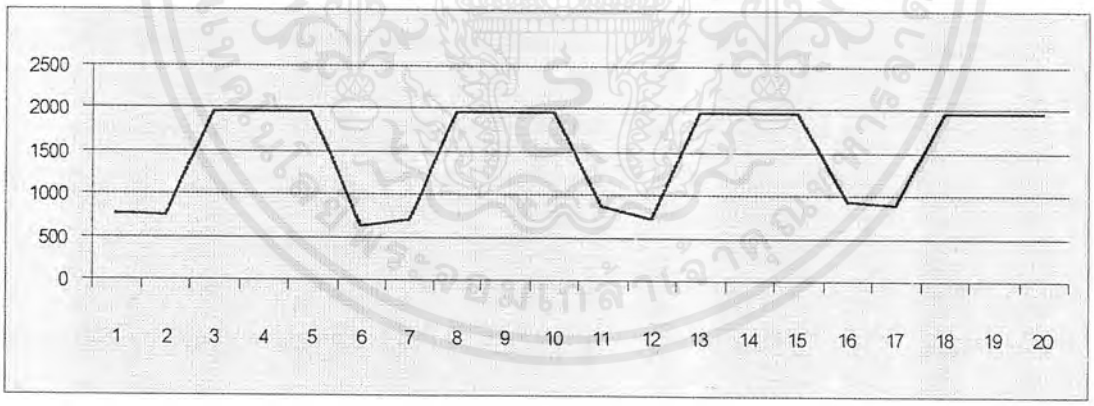
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง แสดงได้ดังนี้



รูปที่ 4.7 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง

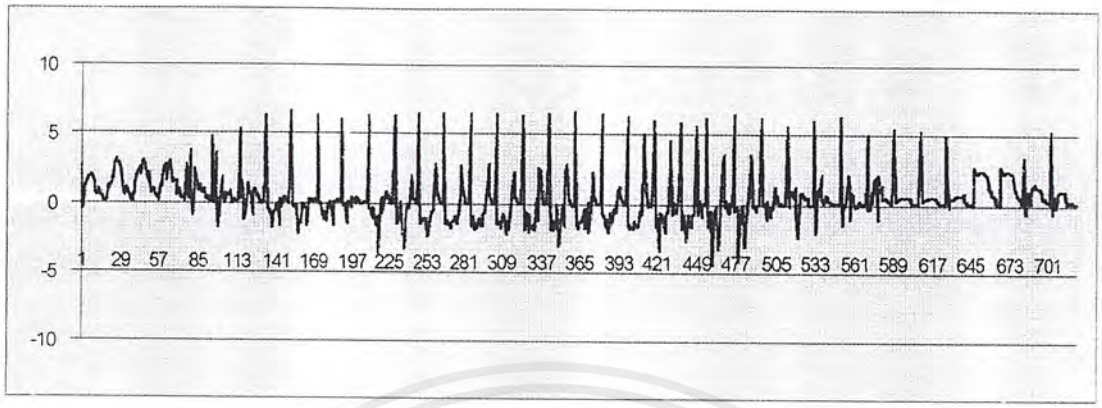
เมื่อนำค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ



รูปที่ 4.8 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า วิ่ง กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

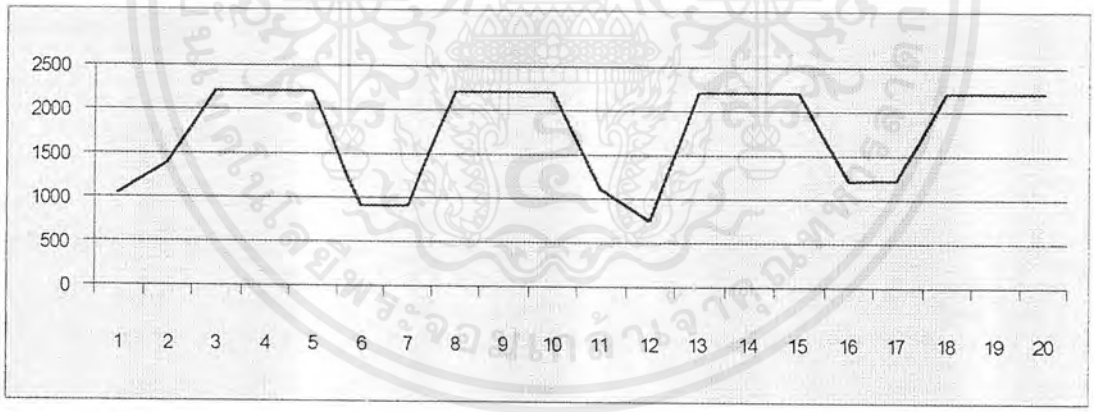
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม แสดงได้ดังนี้



รูปที่ 4.9 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม

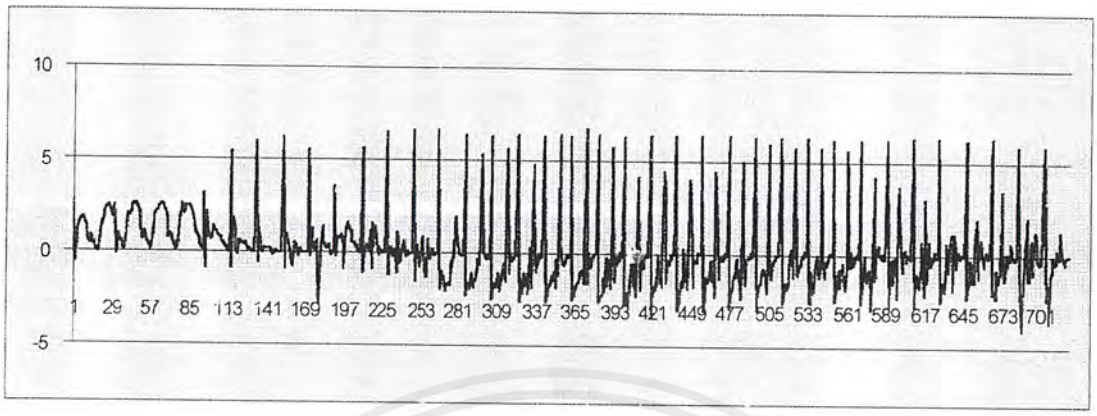
เมื่อนำ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ



รูปที่ 4.10 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า ยิ้ม กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

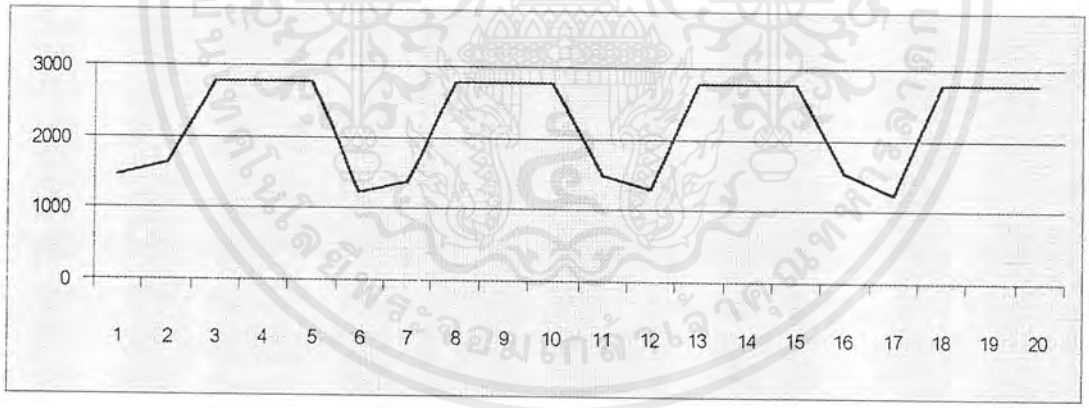
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า หาว แสดงได้ดังนี้



รูปที่ 4.11 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า หาว

เมื่อนำ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า หาว มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ

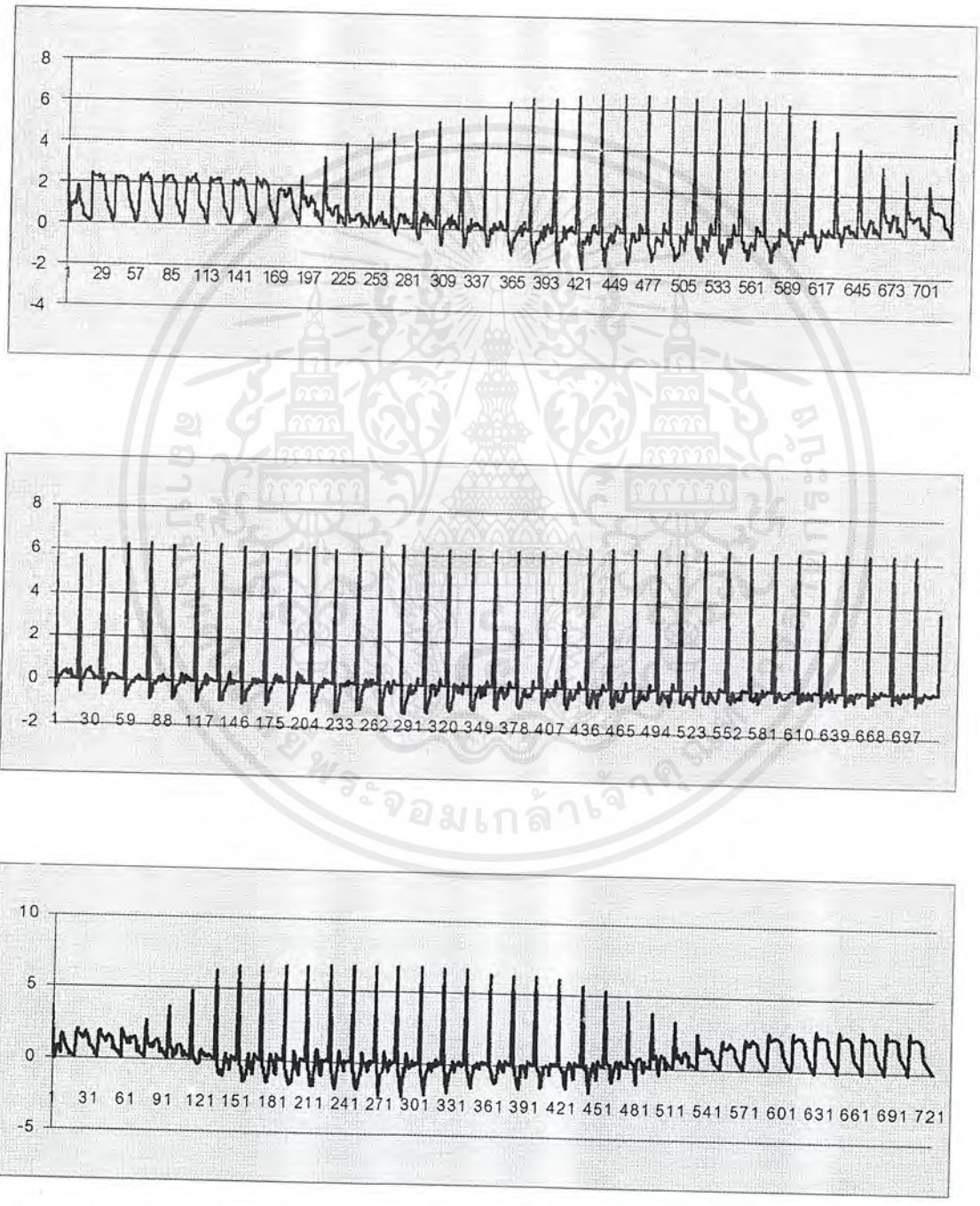


รูปที่ 4.12 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า หาว กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

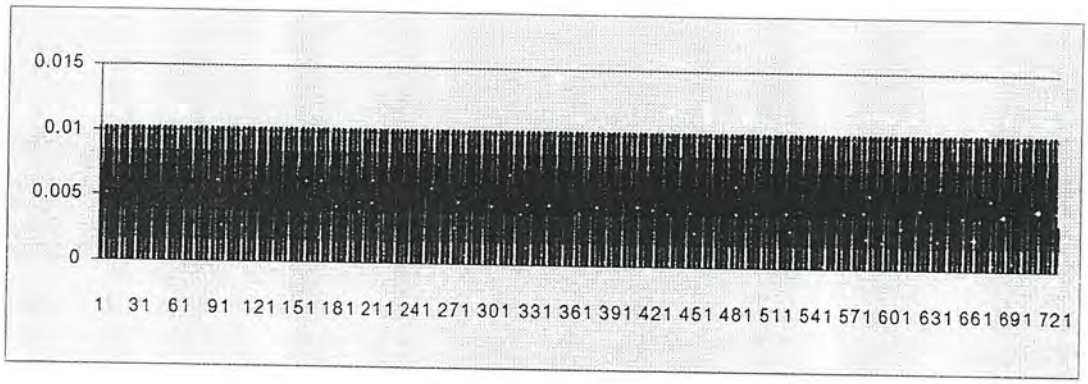
4.1.2 เป็นการเปรียบเทียบค่าความคล้ายคลึงของเสียงโดยค่าพารามิเตอร์ซึ่งได้แก่ สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักและค่าความผิดพลาด โดยแบบอ้างอิงที่ 2 (กลุ่ม B) นี้ ประกอบด้วย ข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ที่ได้มาจาก จำนวนแบบอ้างอิงที่สร้างจากนักศึกษาชาย จำนวน 9 คน และนักศึกษาหญิงจำนวน 9 คน โดยที่แต่ละคนได้ทำการพูด คำว่า เดิน,วิ่ง,ยิ้ม,หาว อย่างละ 5 ครั้ง แสดงได้ดังกราฟต่อไปนี้

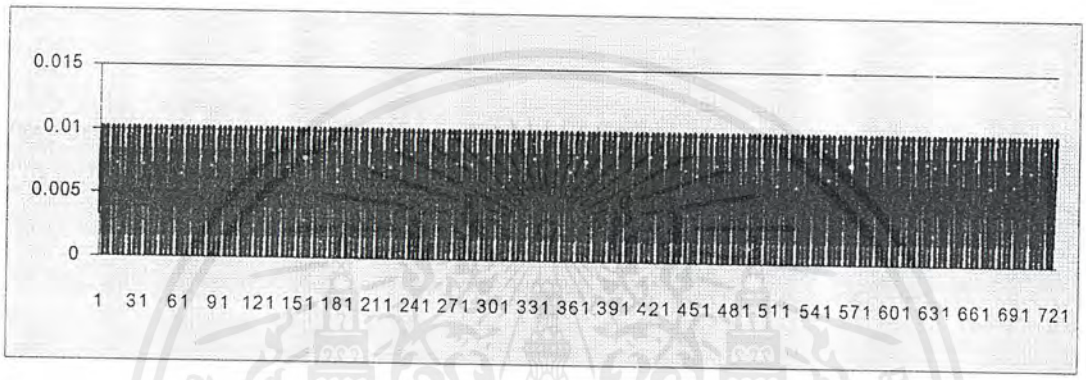
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า **เดิน** มีดังนี้

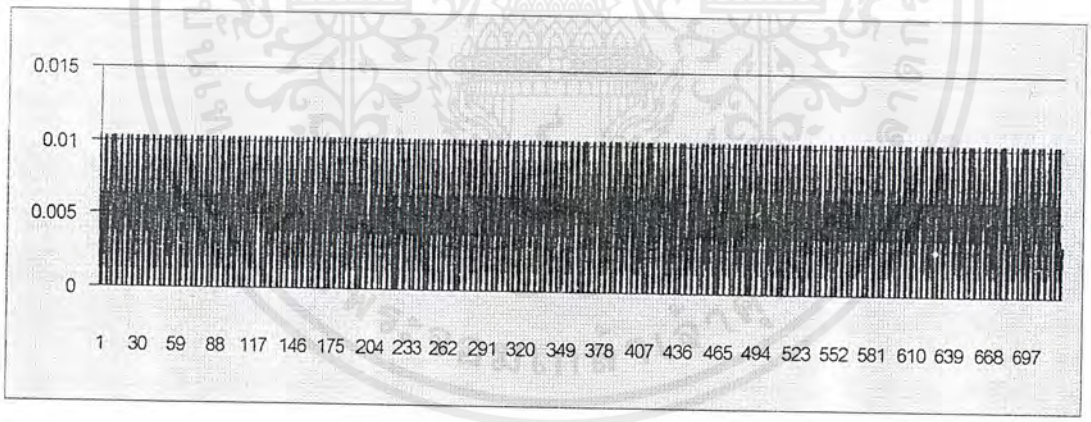


รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

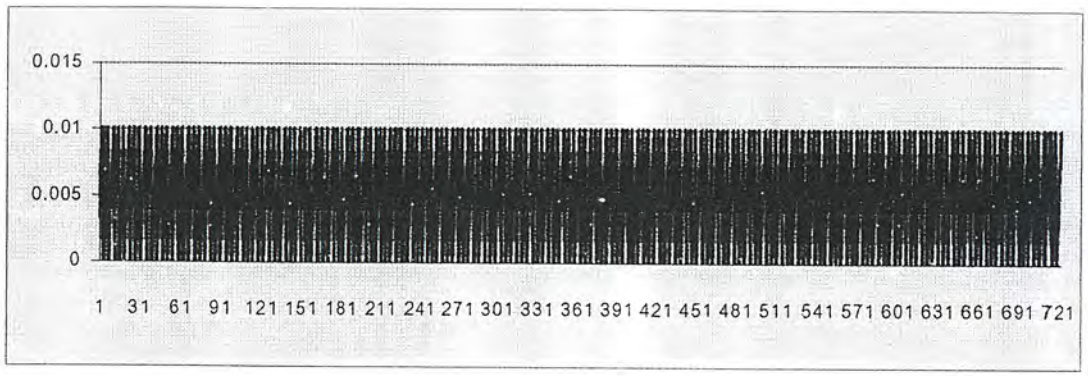


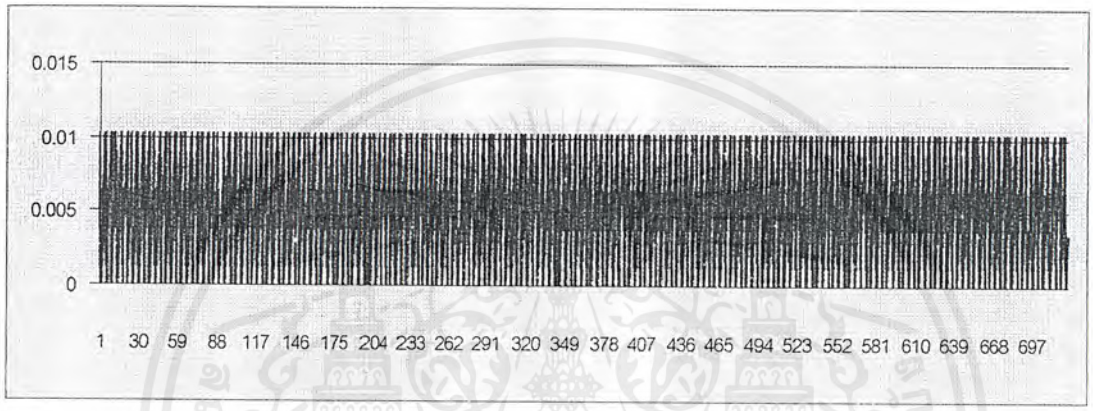




รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

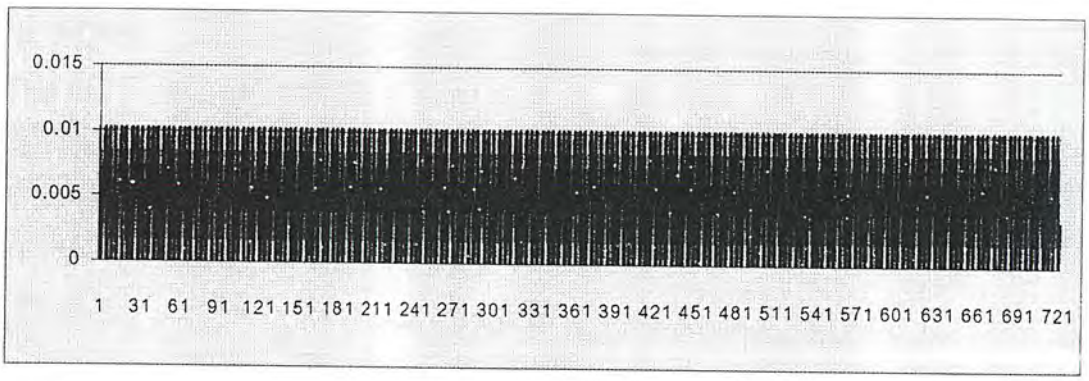
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

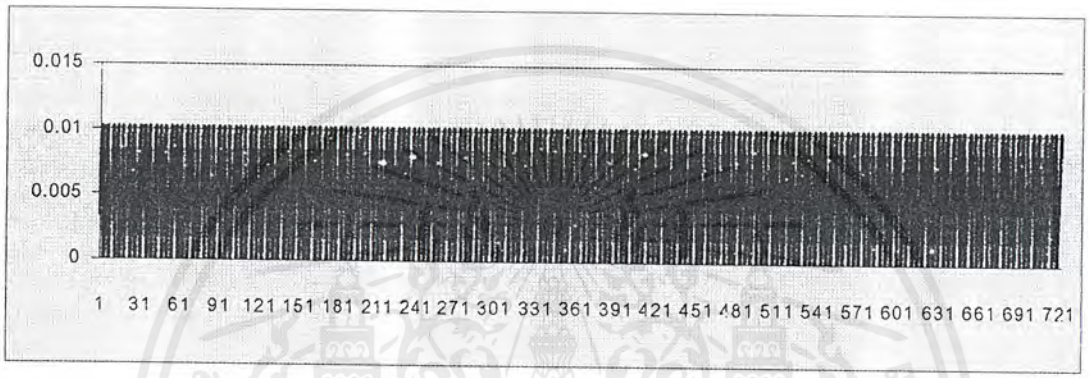


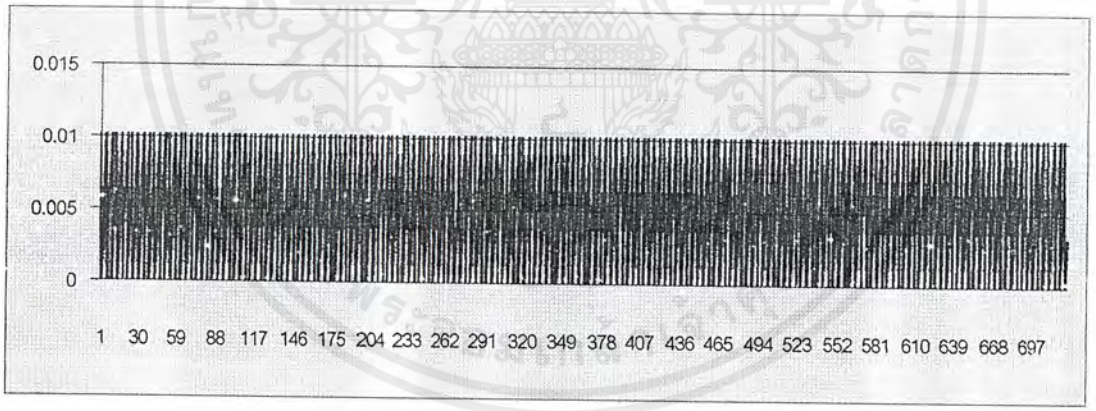


รูปที่4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

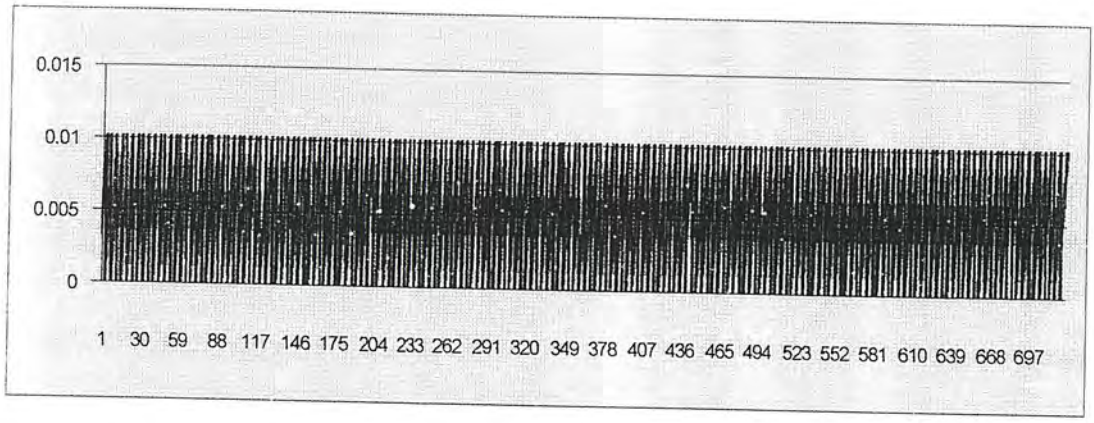


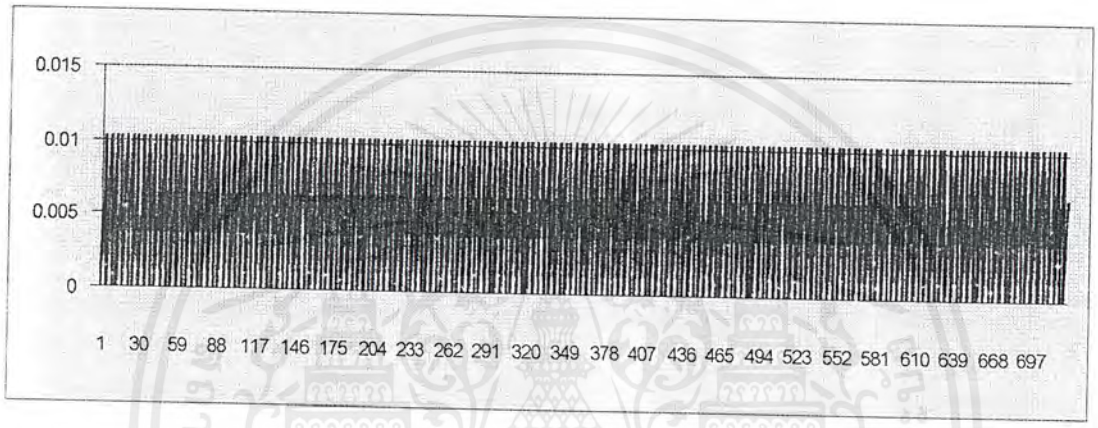


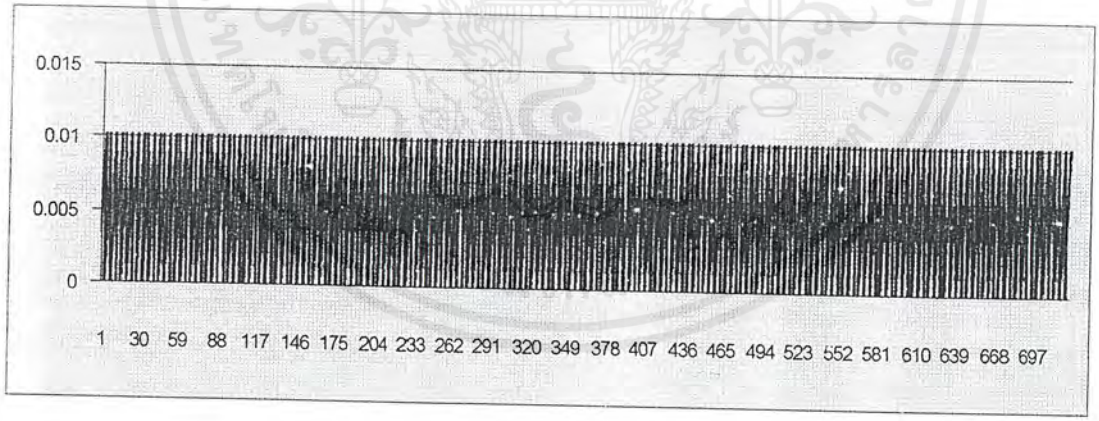


รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

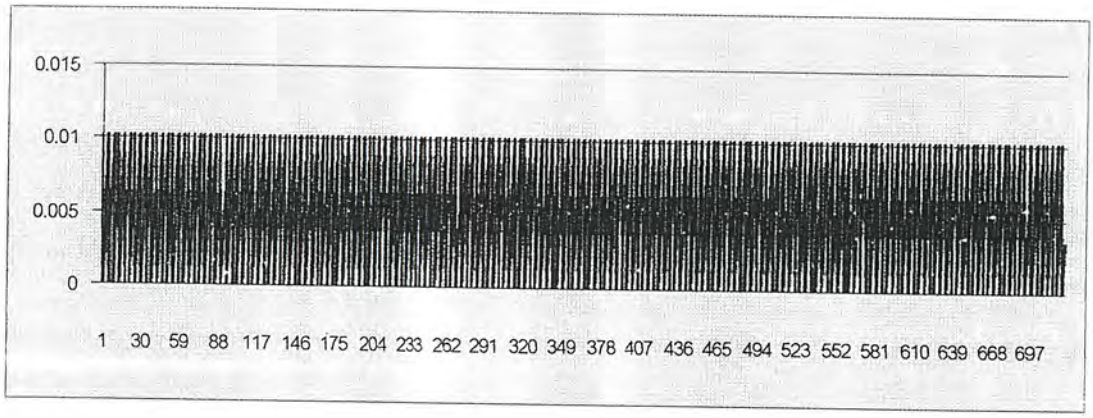


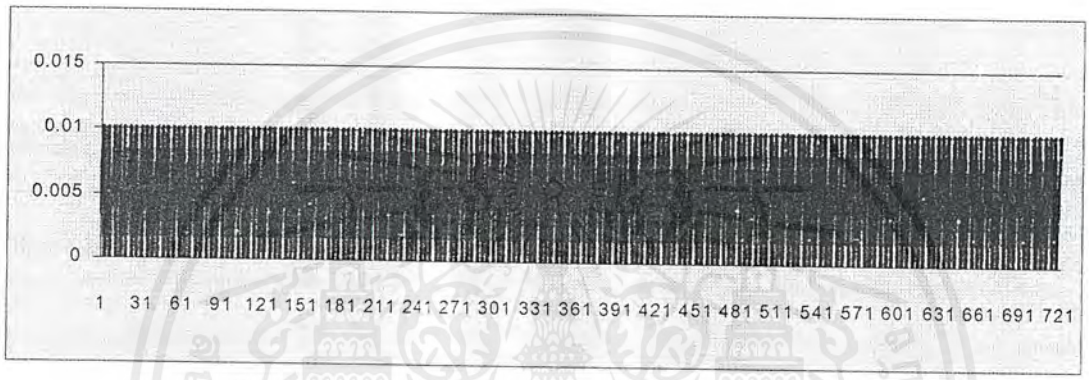


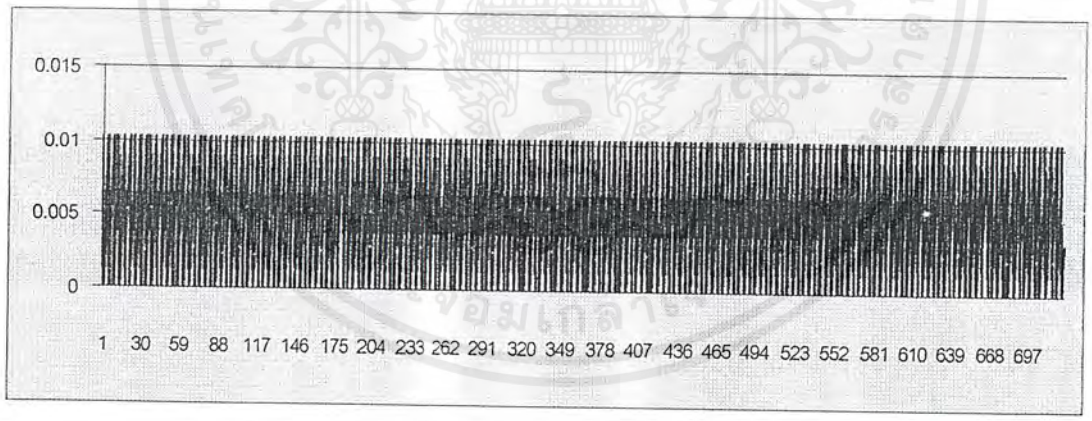


รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

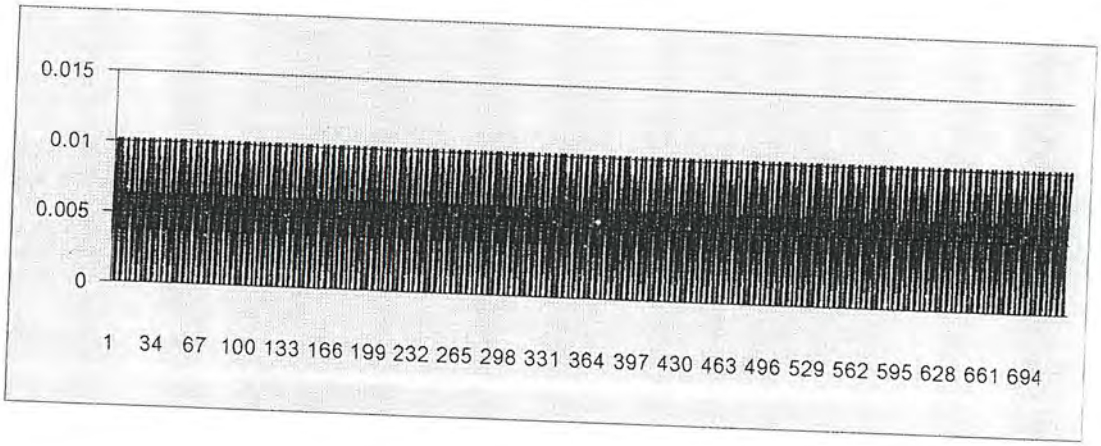


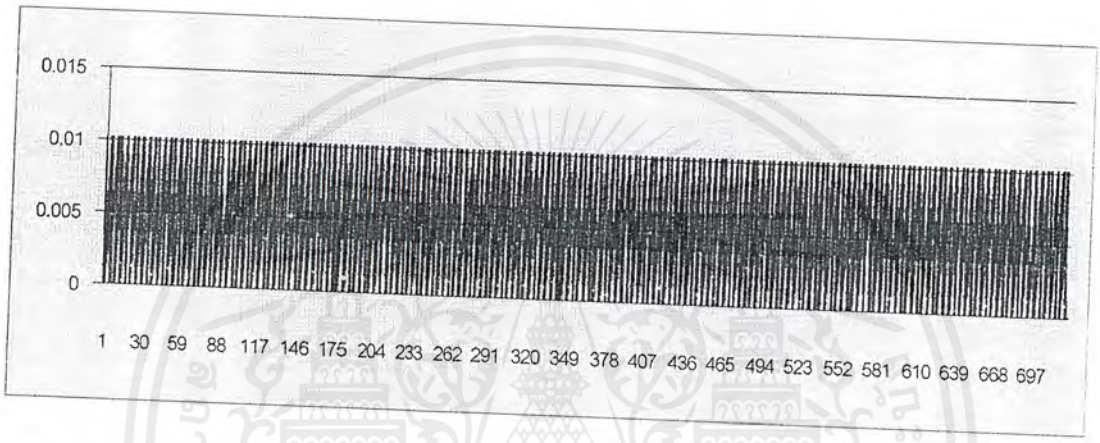




รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

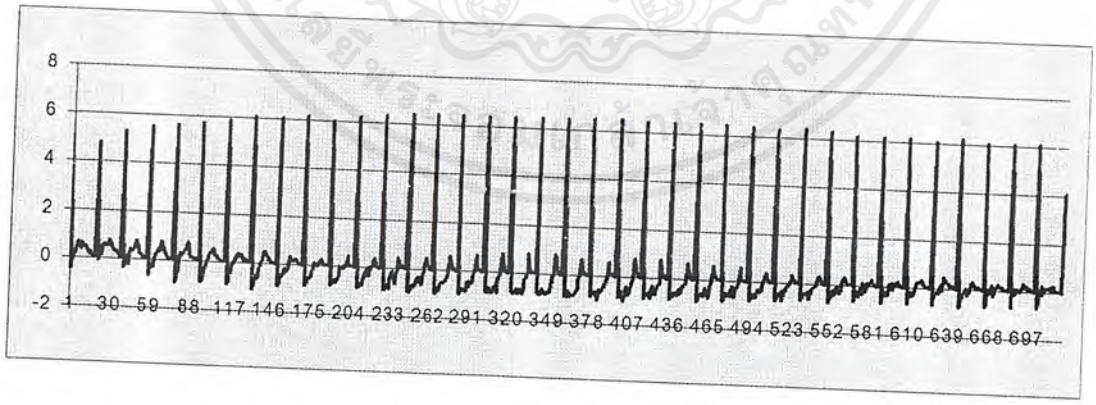
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





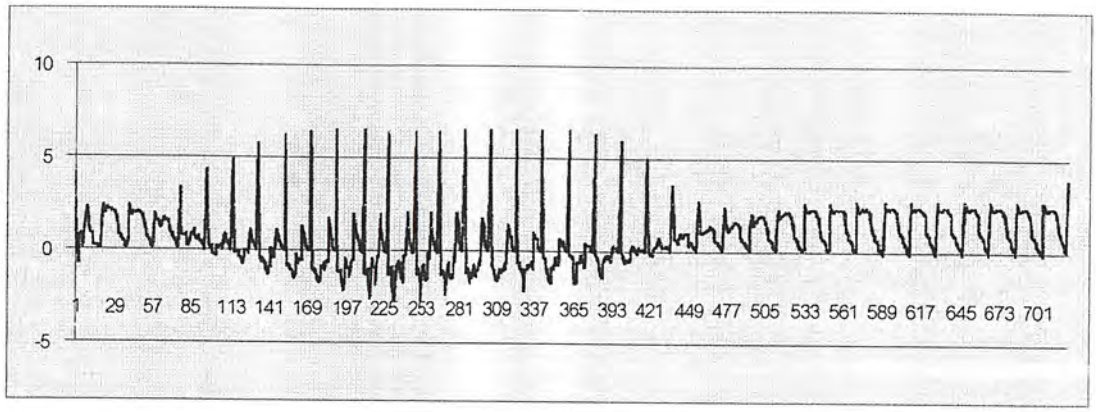
รูปที่ 4.13 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า เดิน (ต่อ)

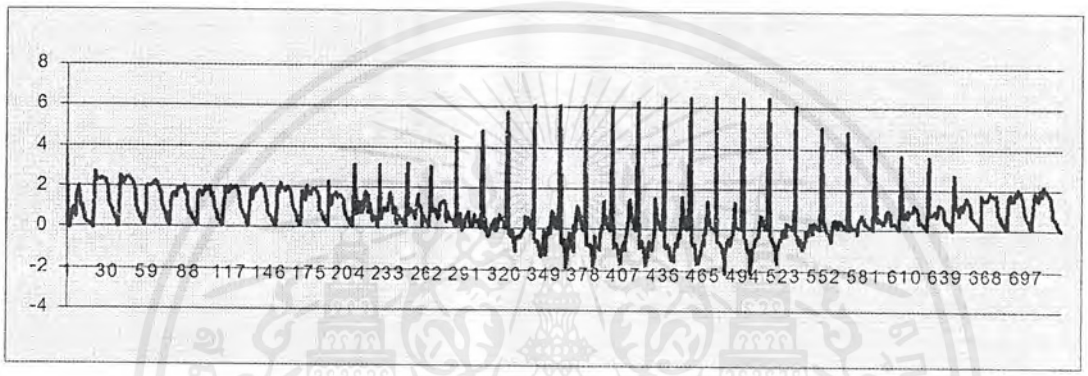
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า วิ่ง มีดังนี้

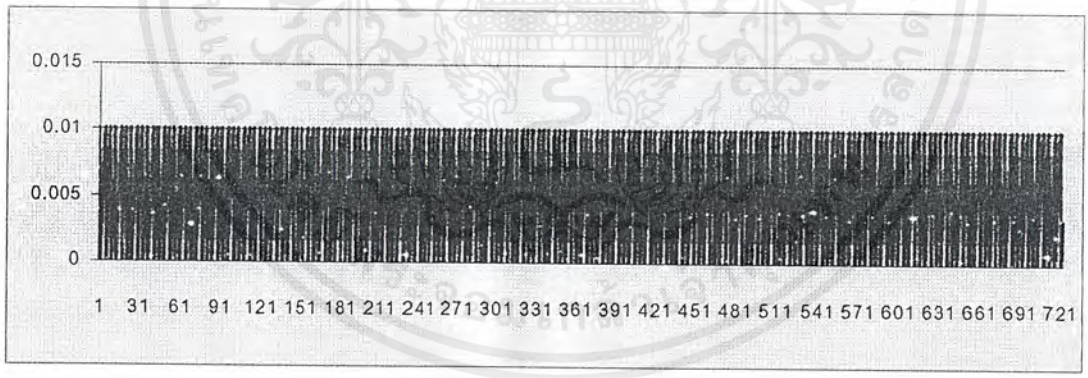


รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

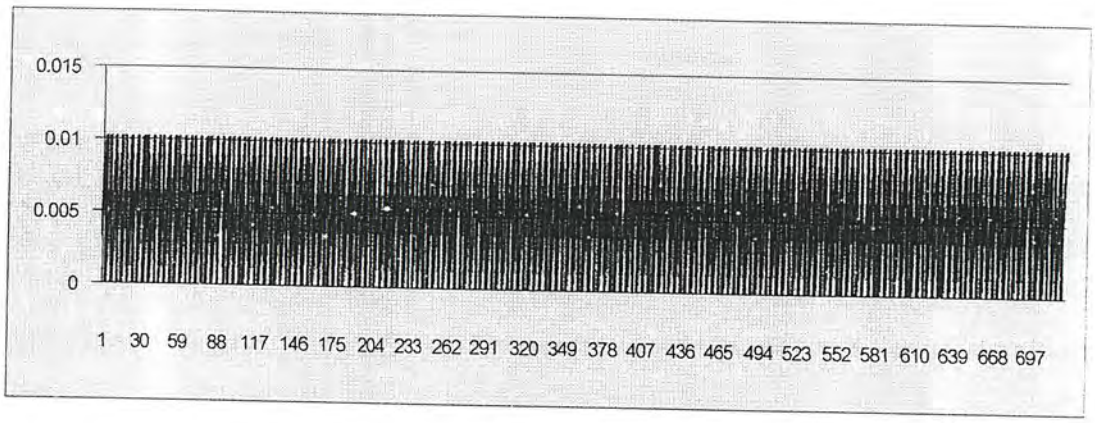


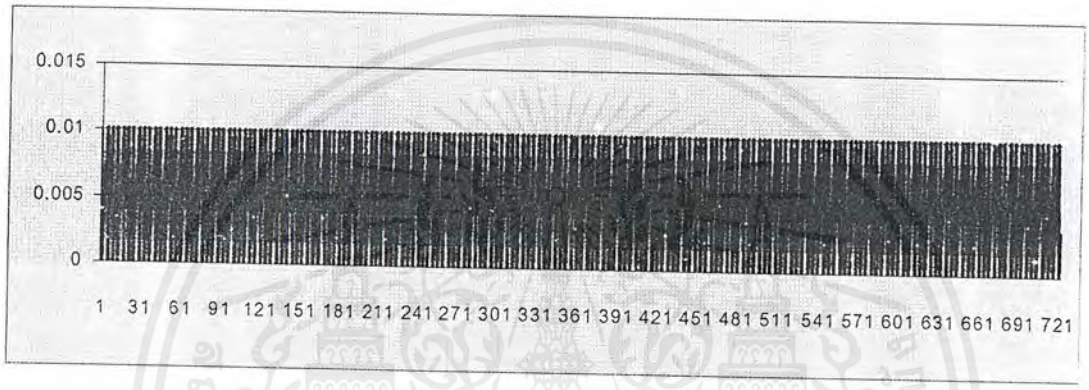




รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

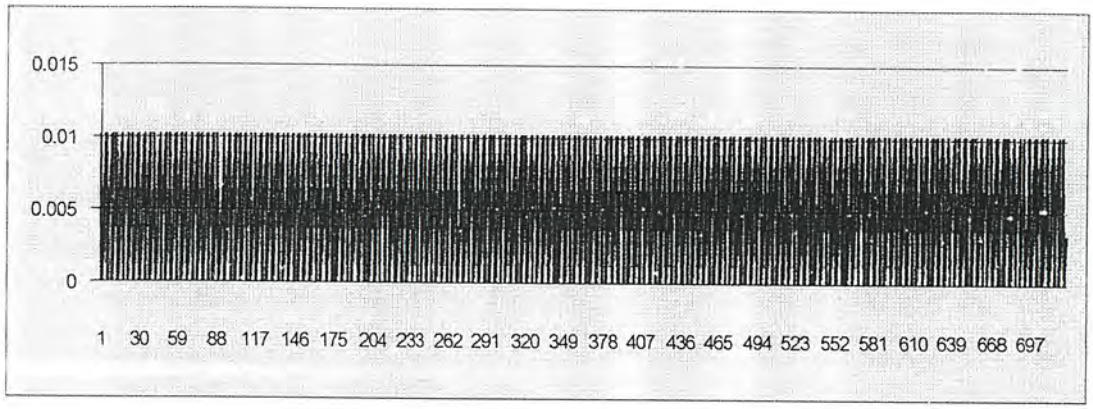
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

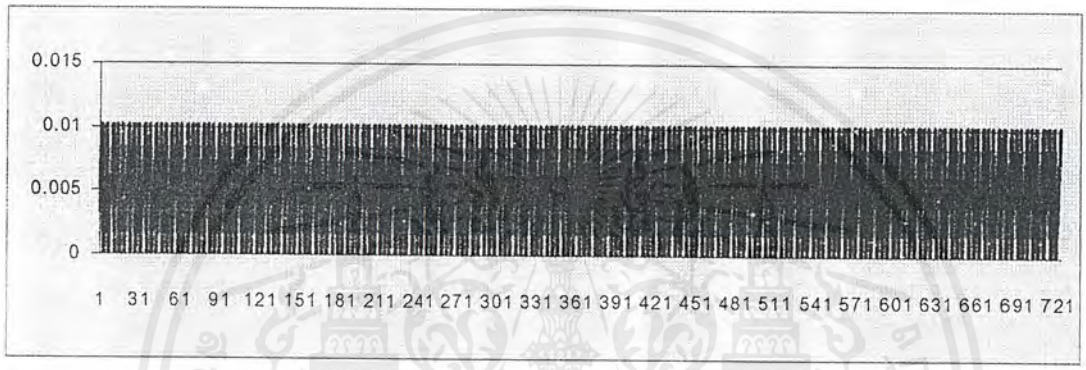


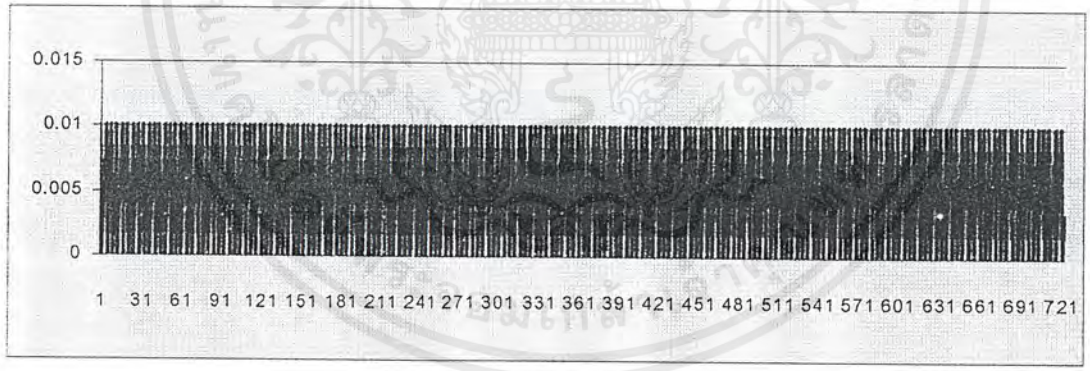


รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

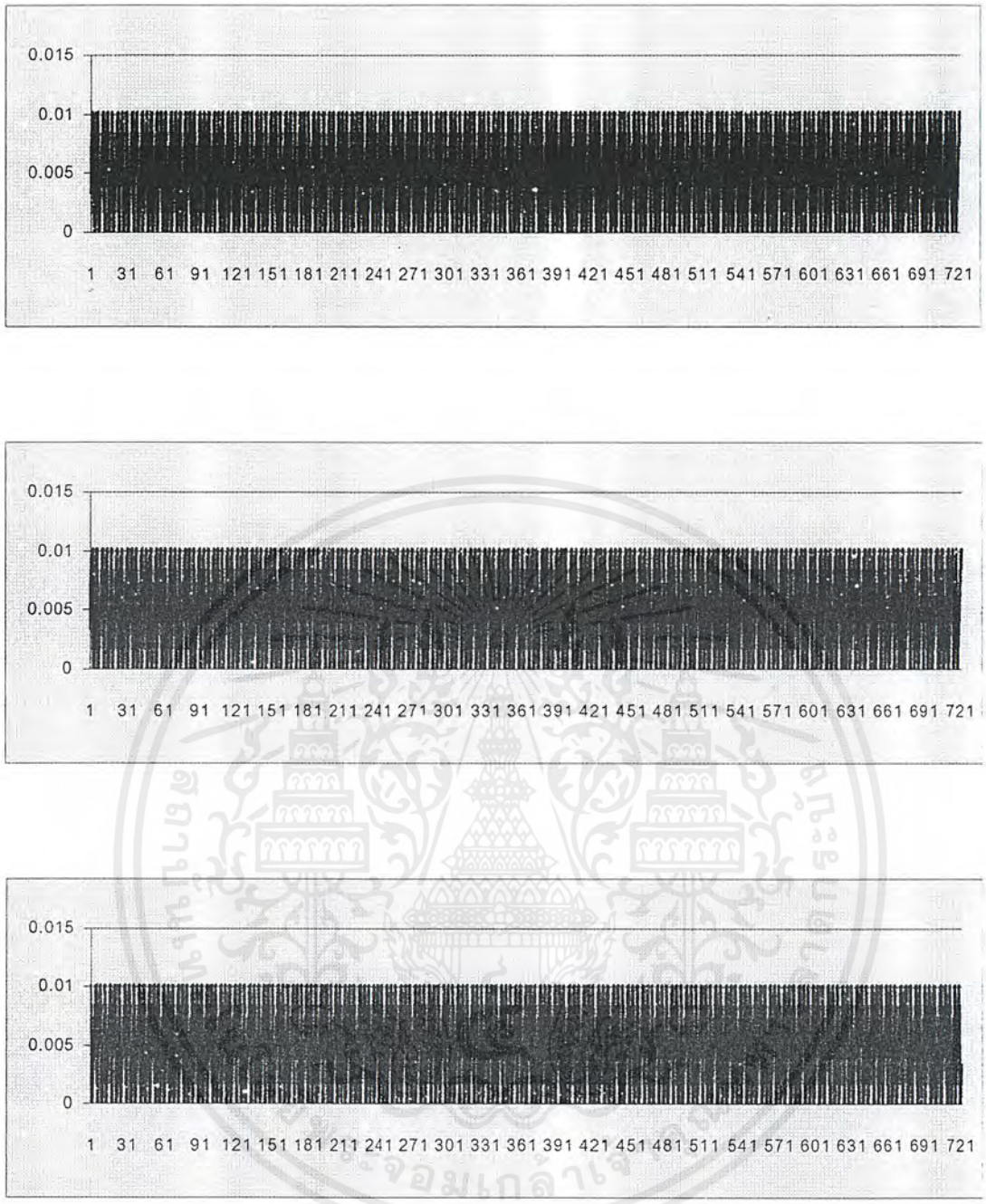






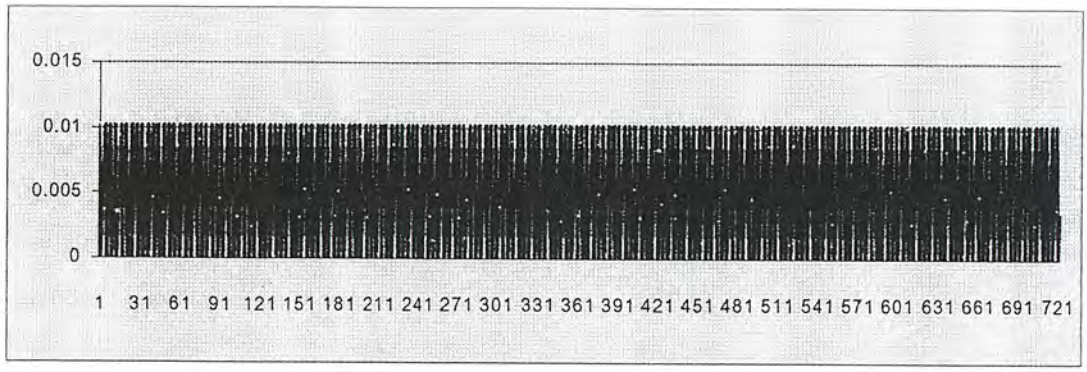
รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

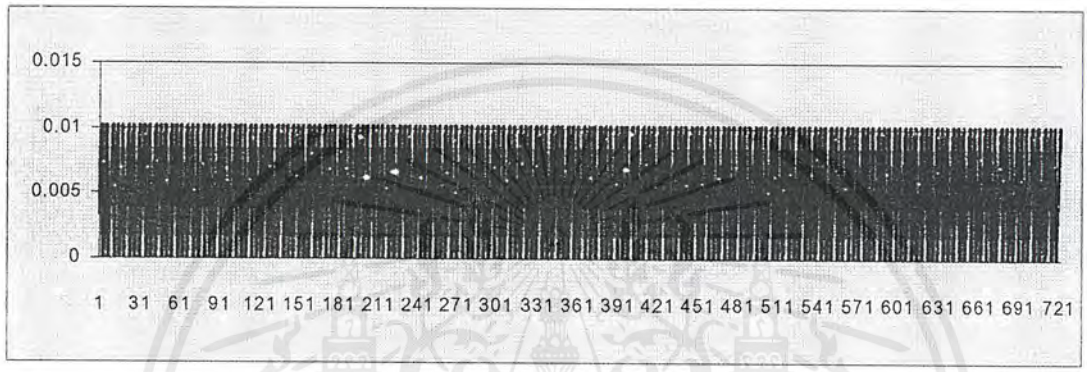
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

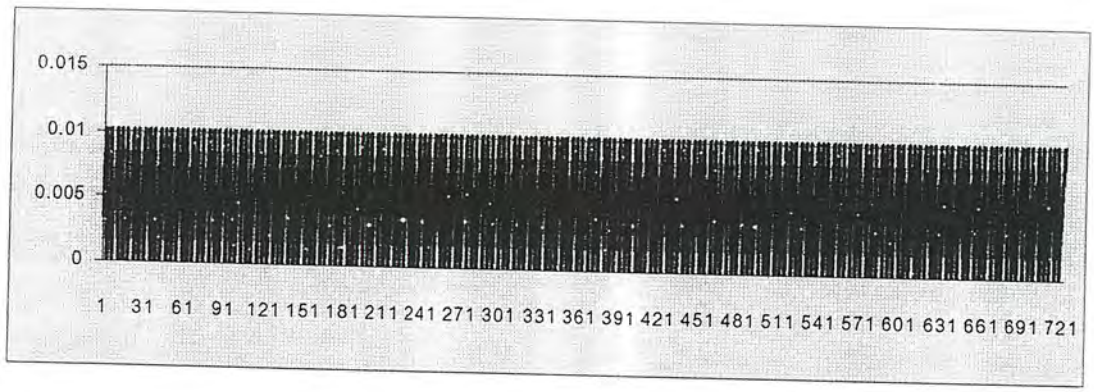
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

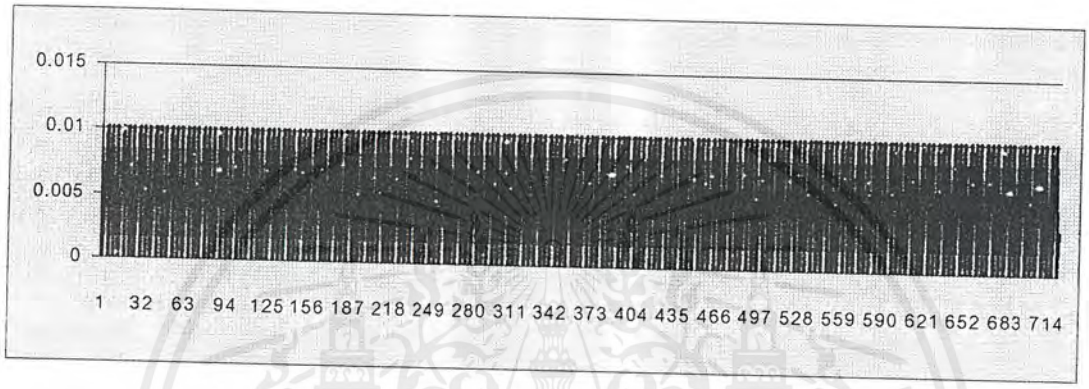




รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

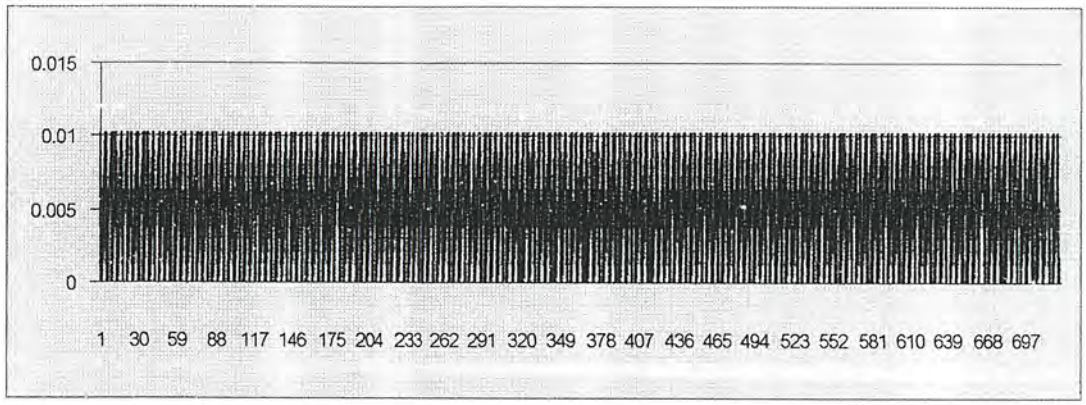
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





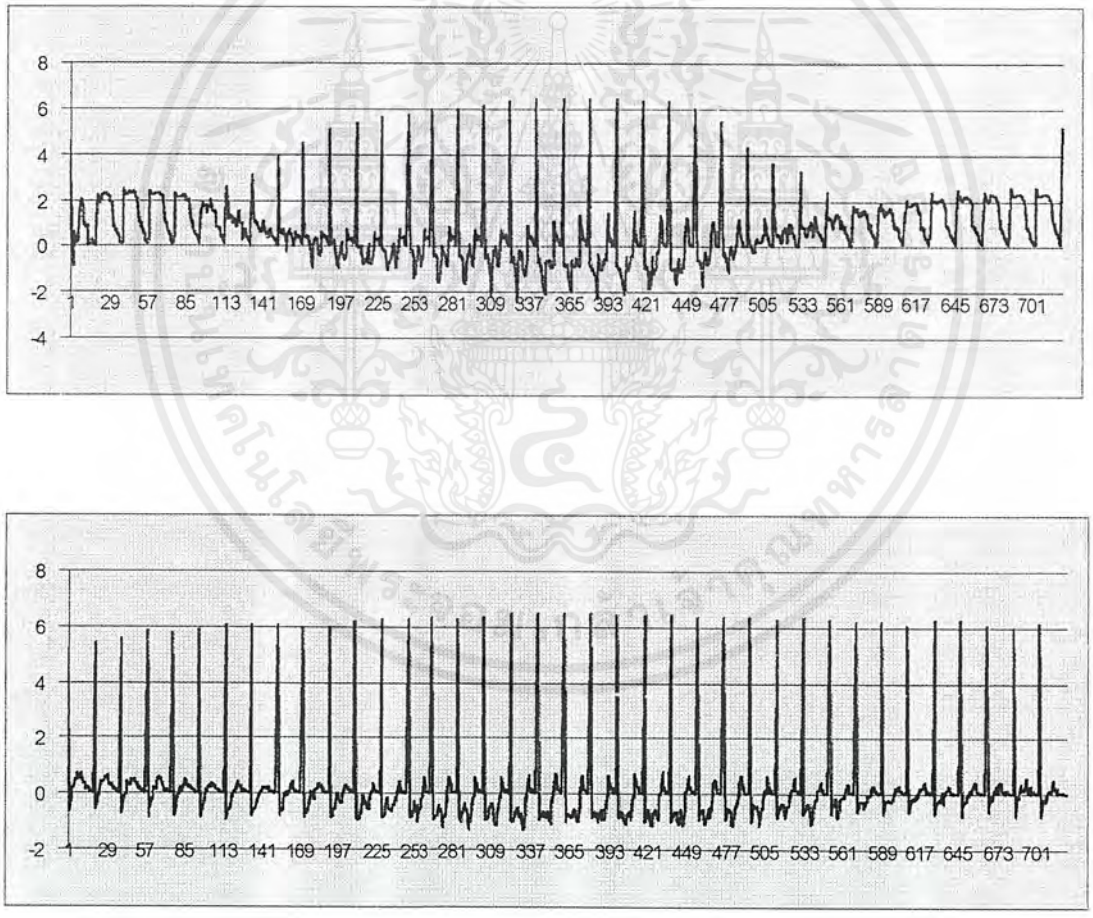
รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



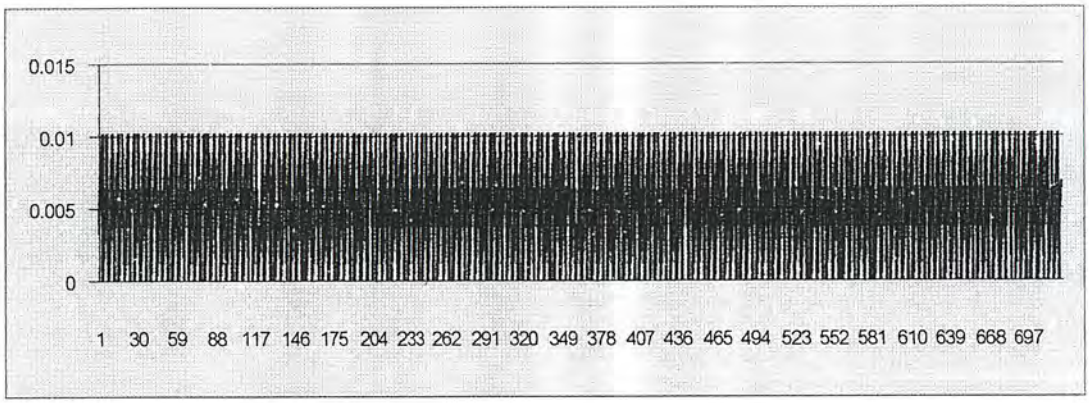
รูปที่ 4.14 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า วิ่ง (ต่อ)

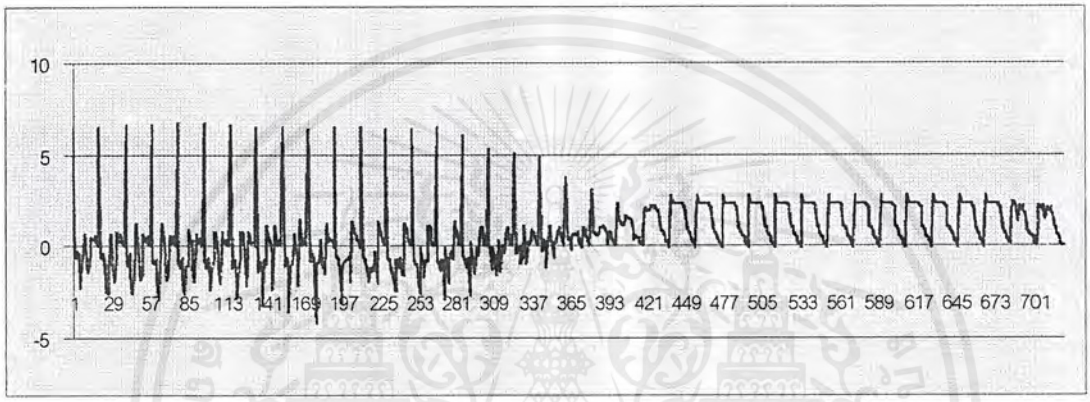
กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า **ยิ้ม** มีดังนี้

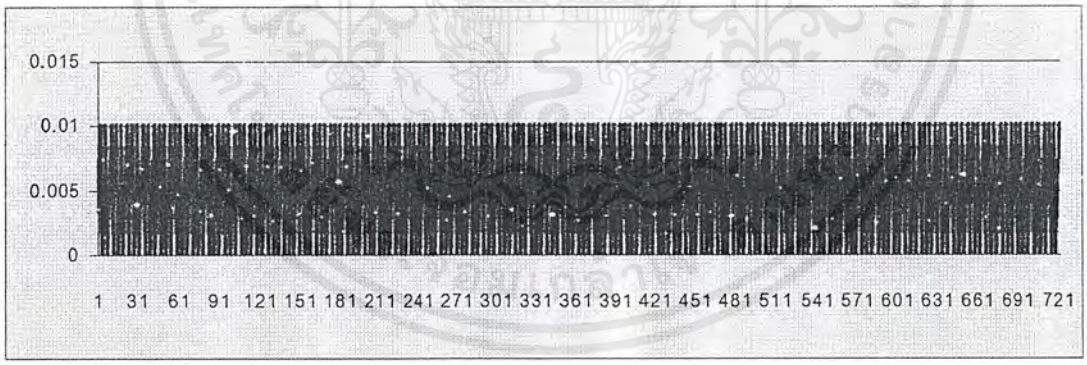


รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

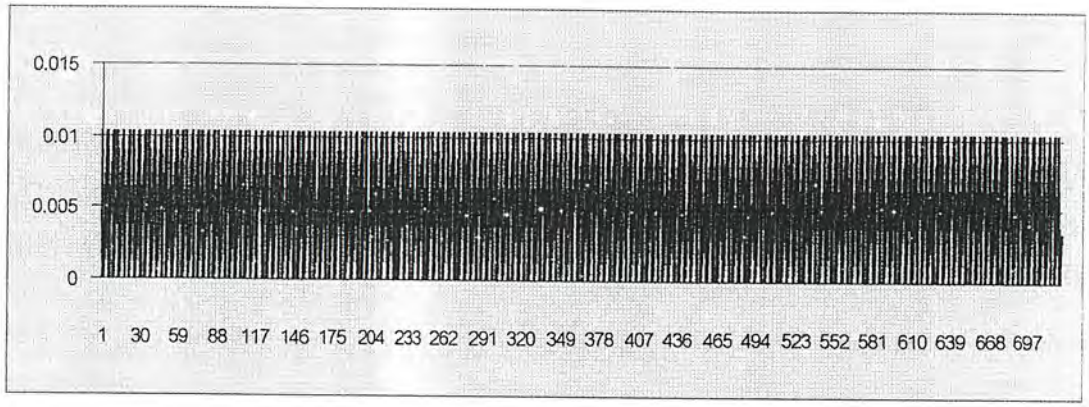


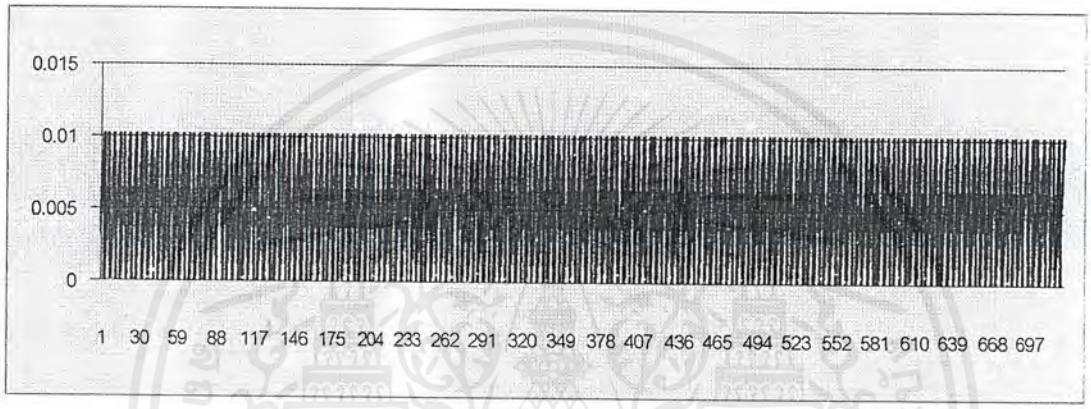


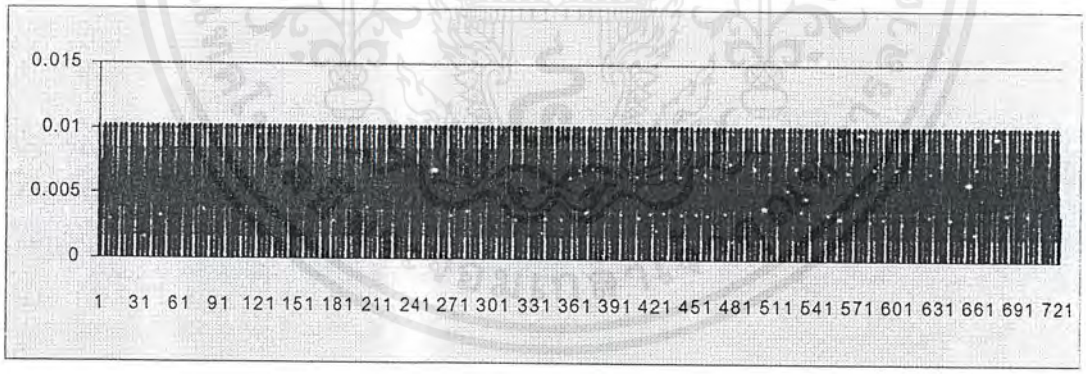


รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

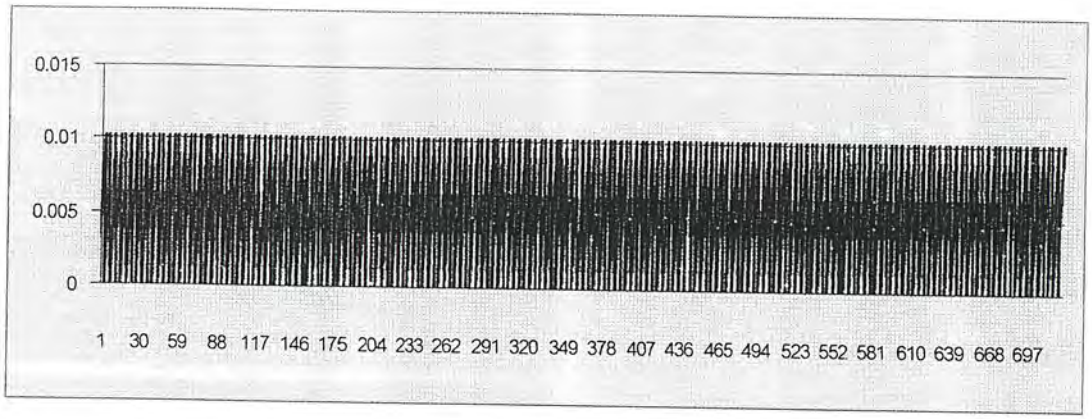


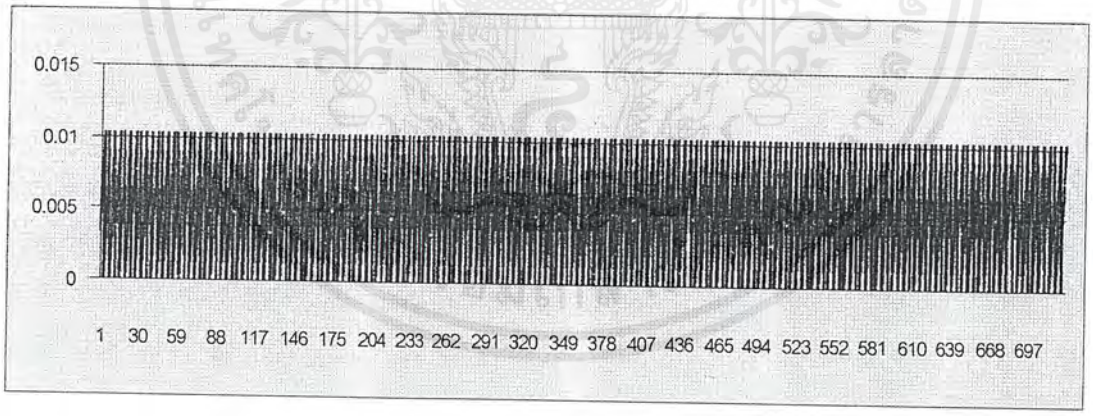




รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

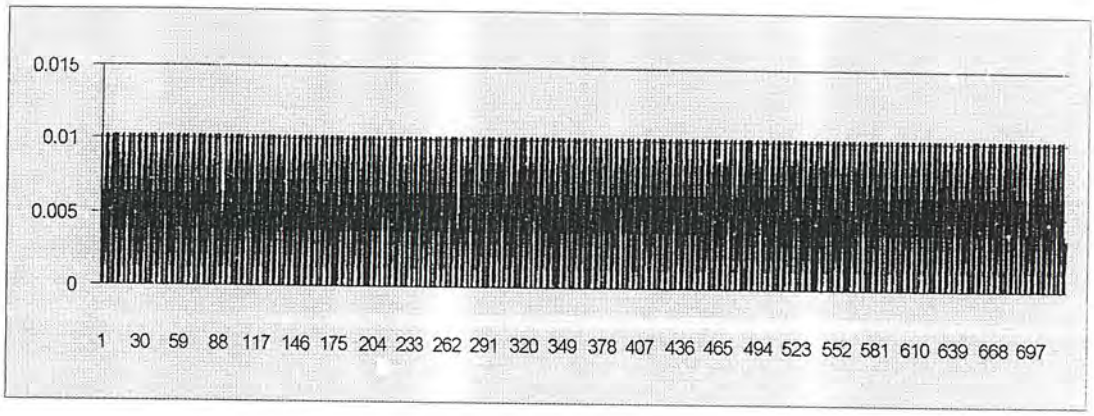
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

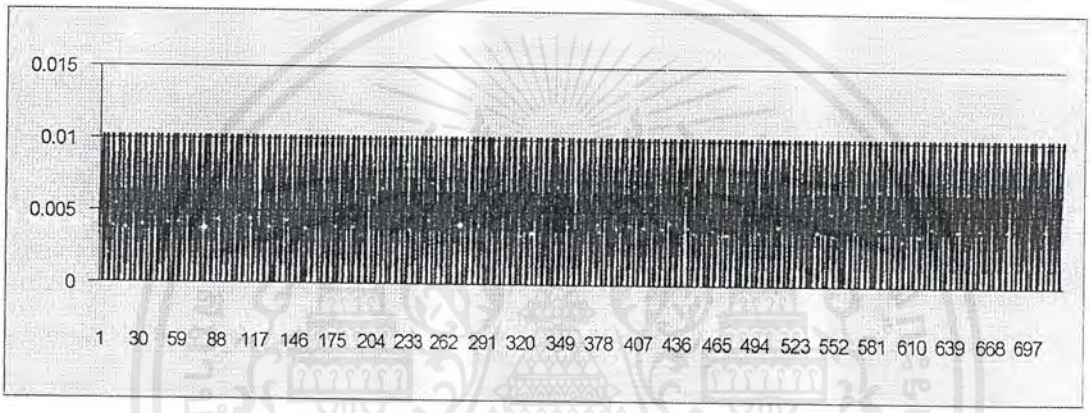


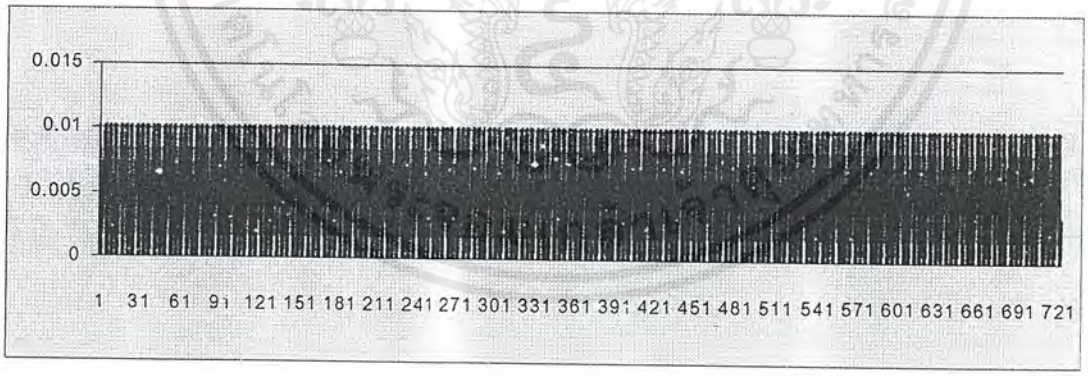


รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

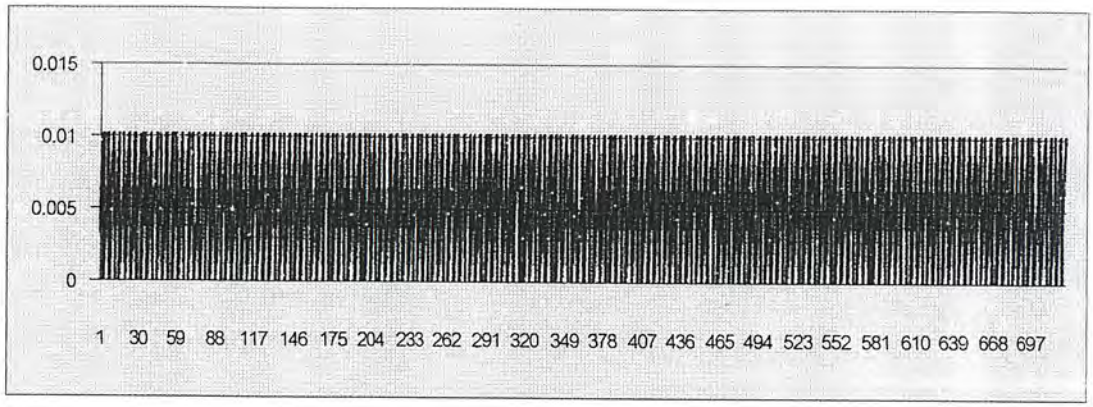


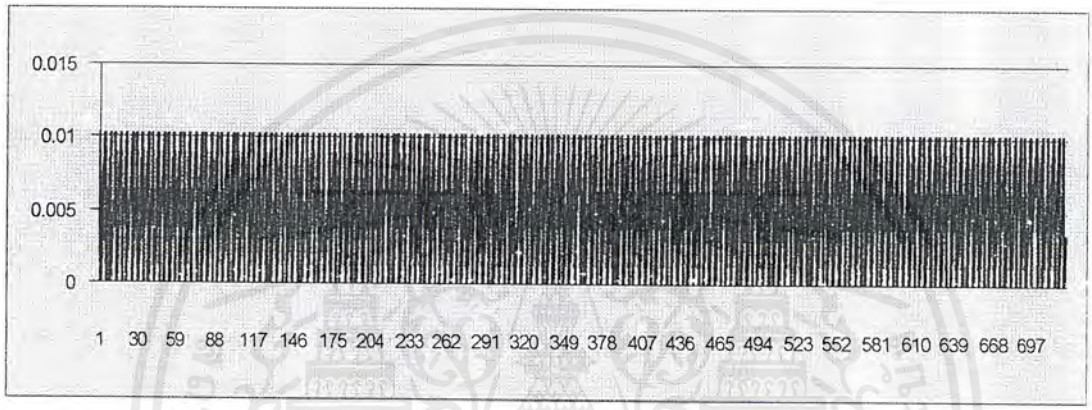




รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

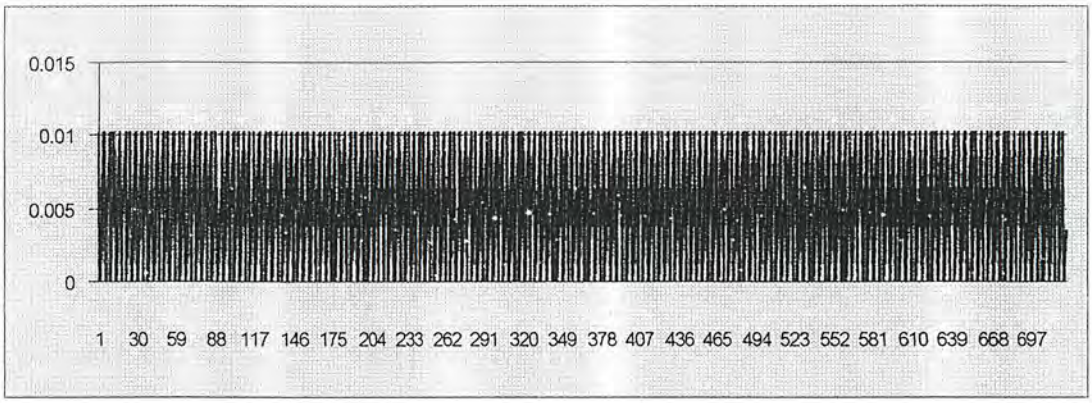
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

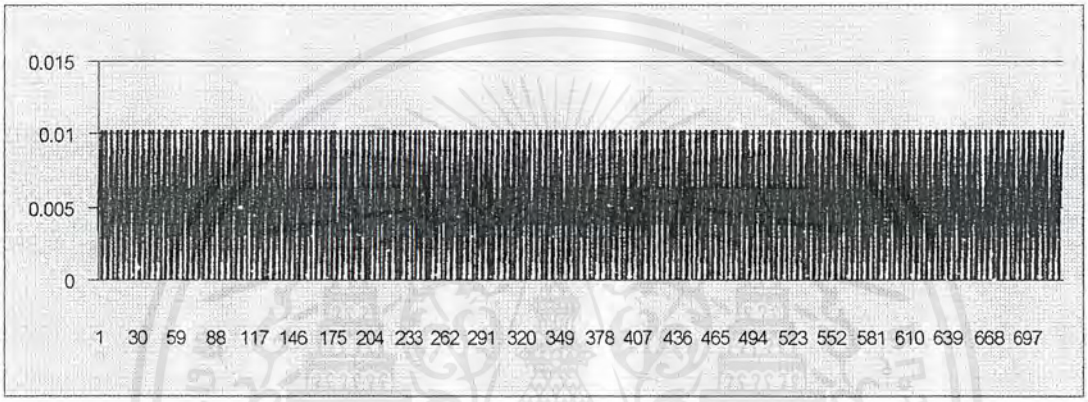


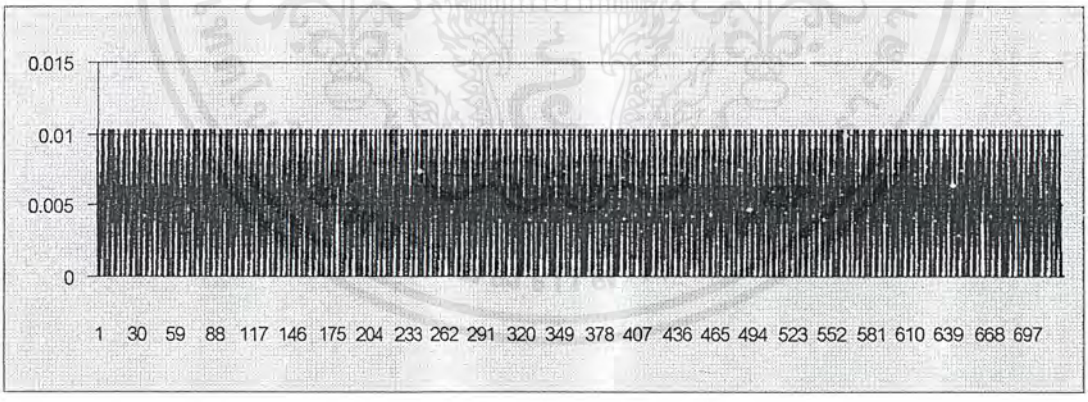


รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



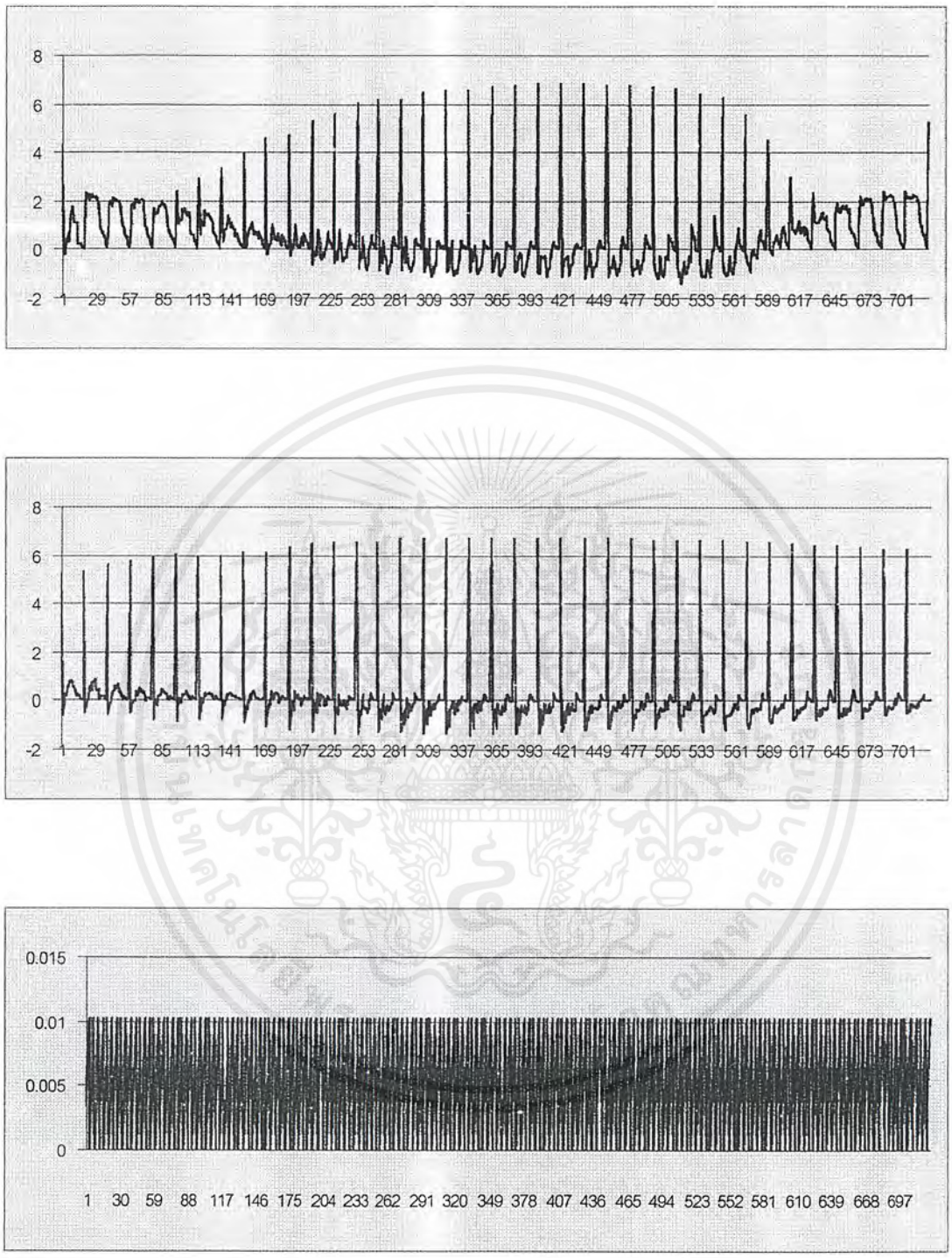




รูปที่ 4.15 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า ยิ้ม (ต่อ)

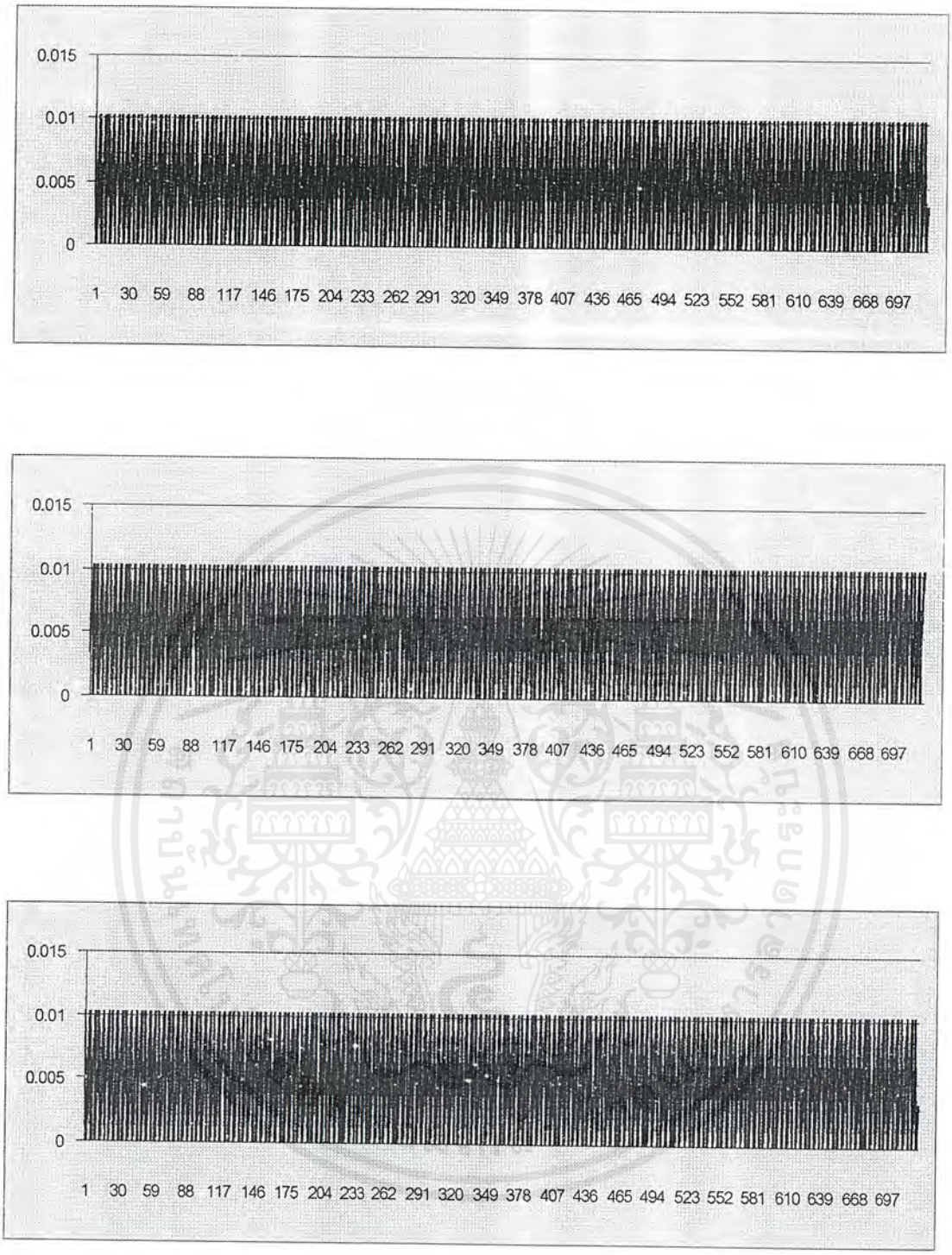
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของคำว่า **หาว** มีดังนี้



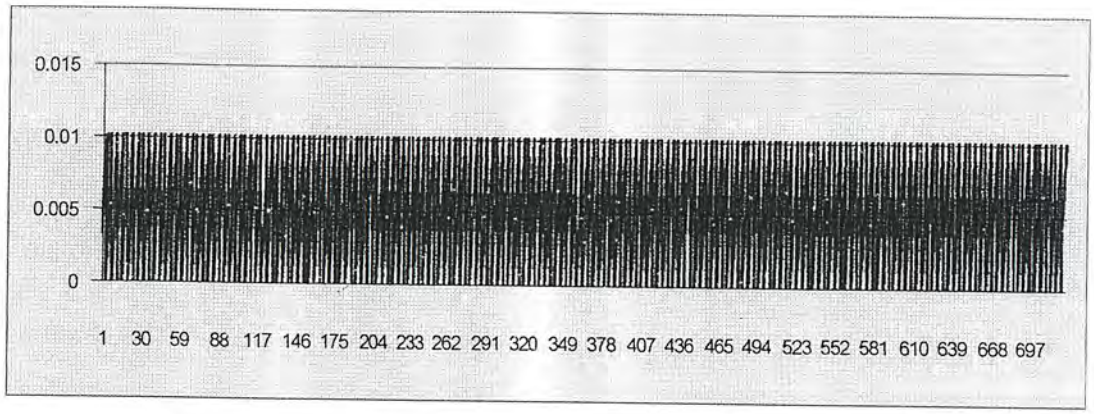
รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว

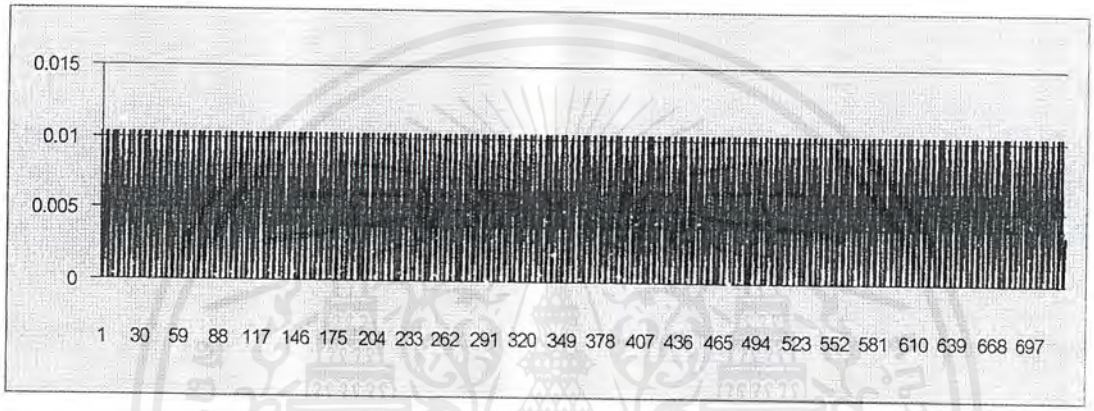
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

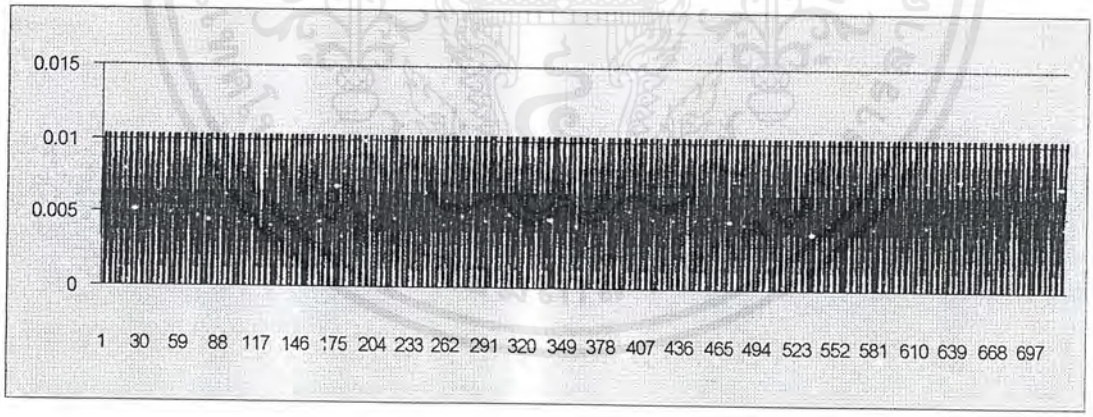


รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

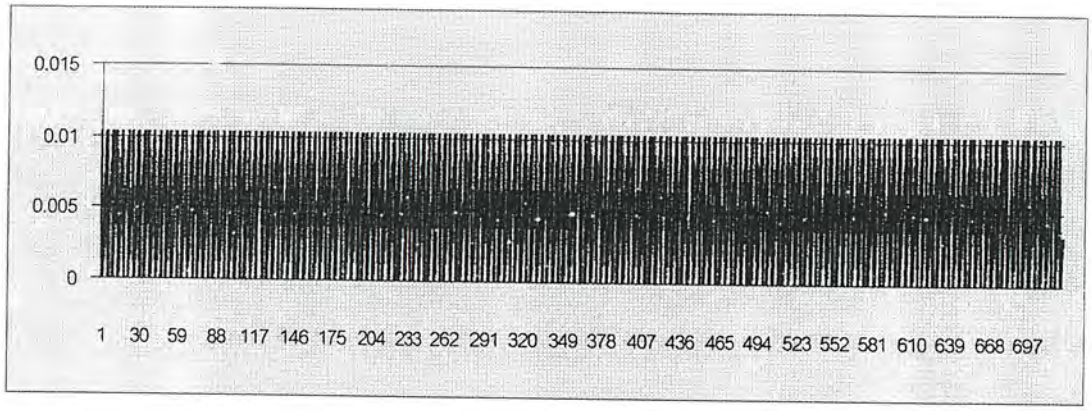


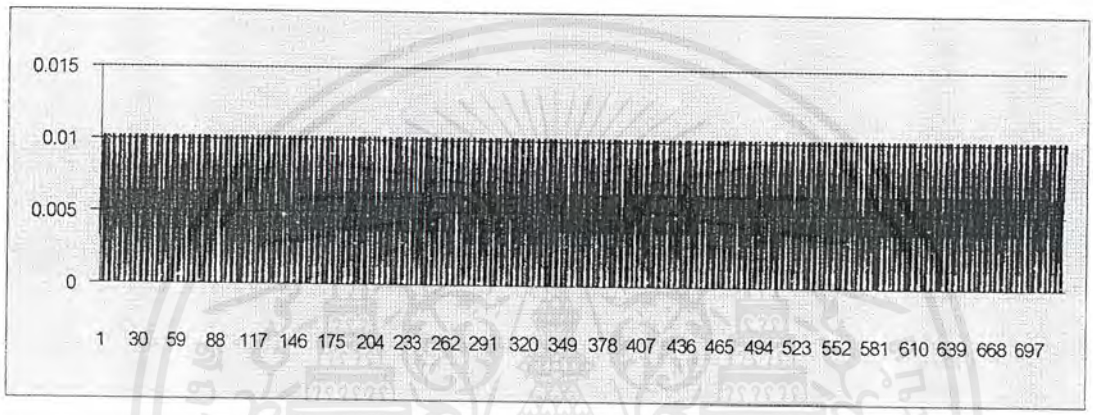


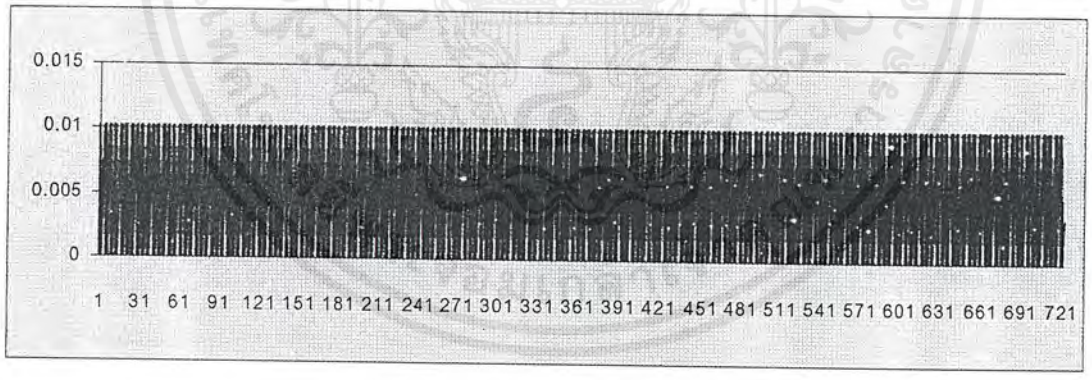


รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

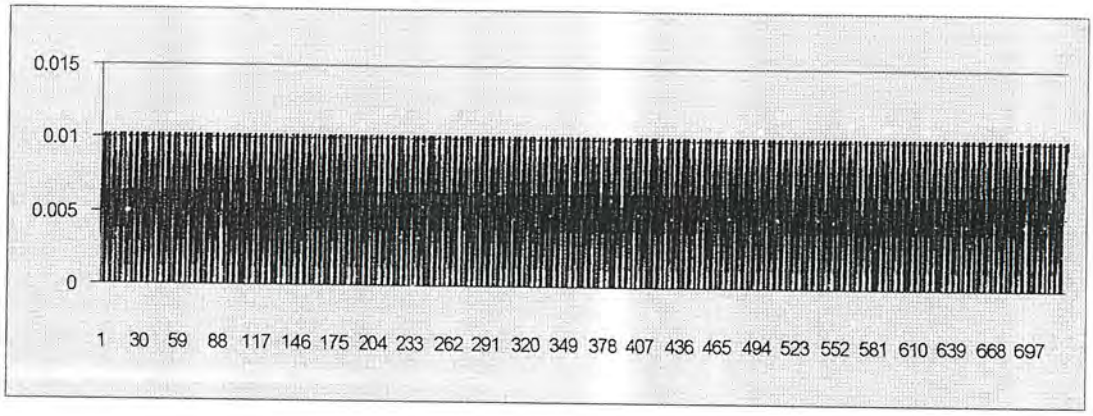


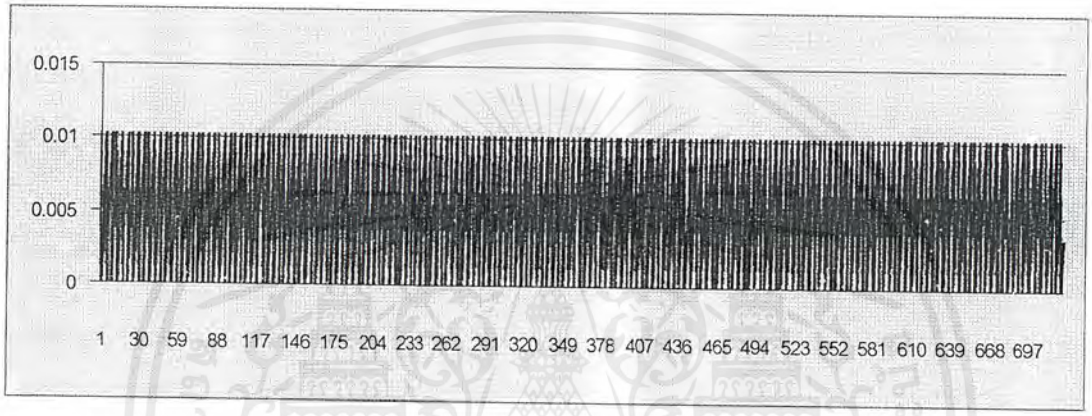


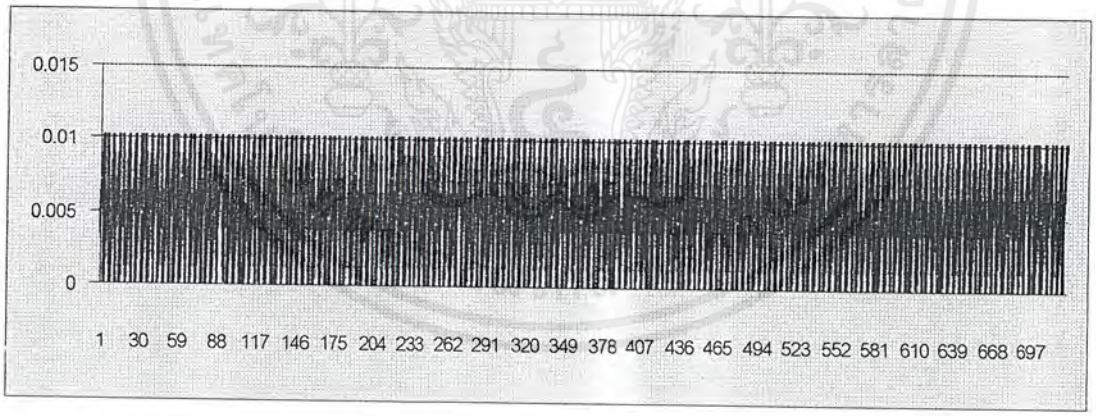


รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

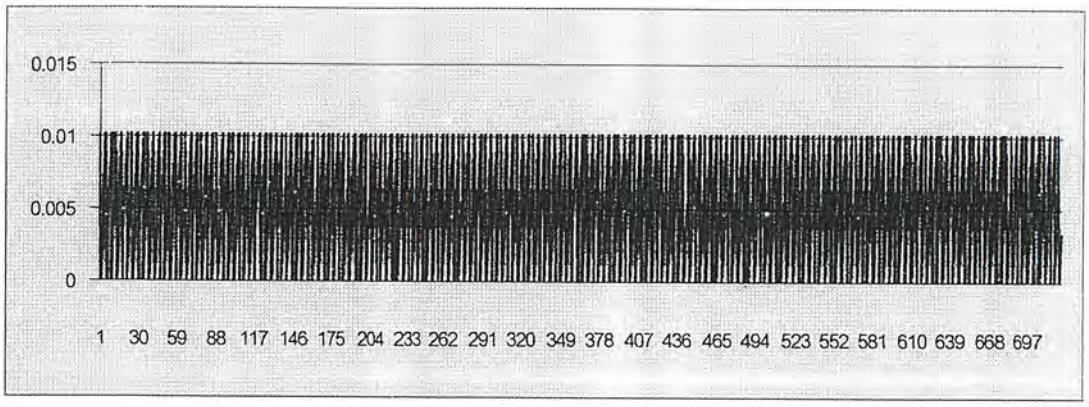


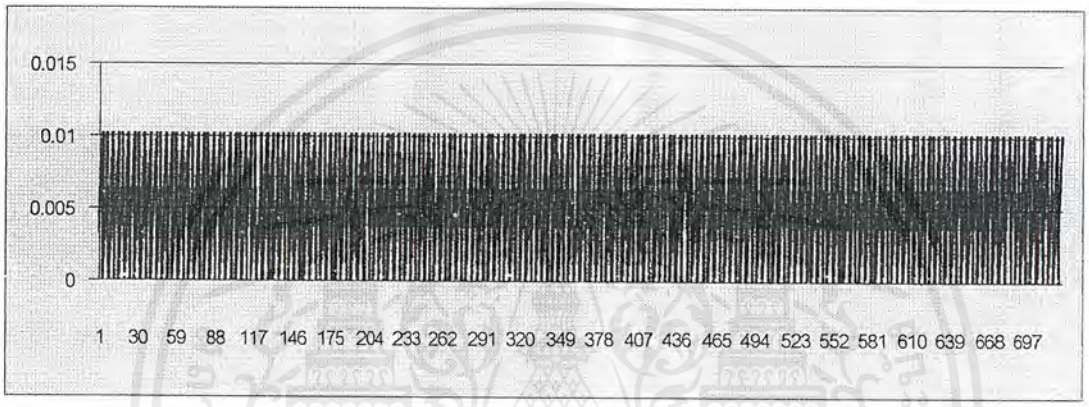


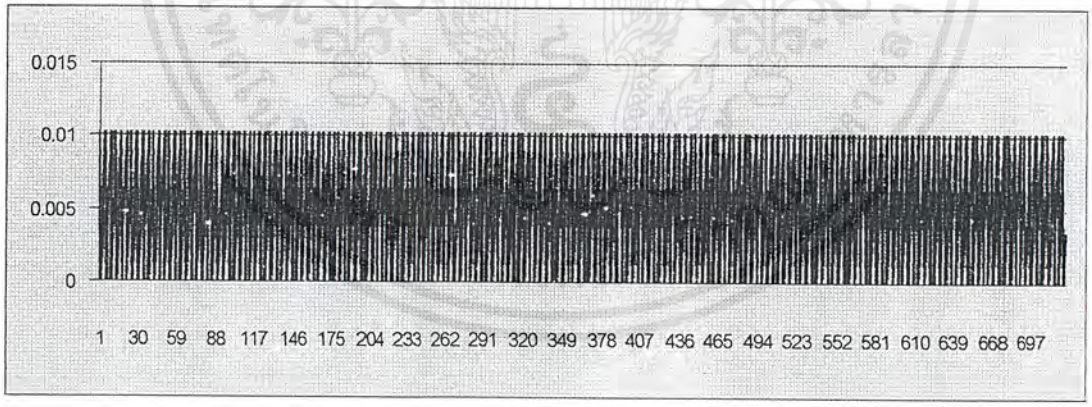


รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

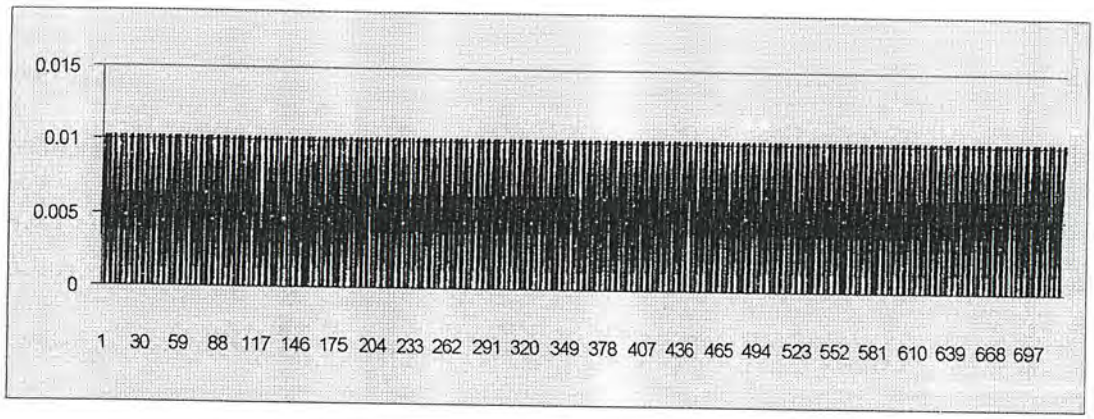






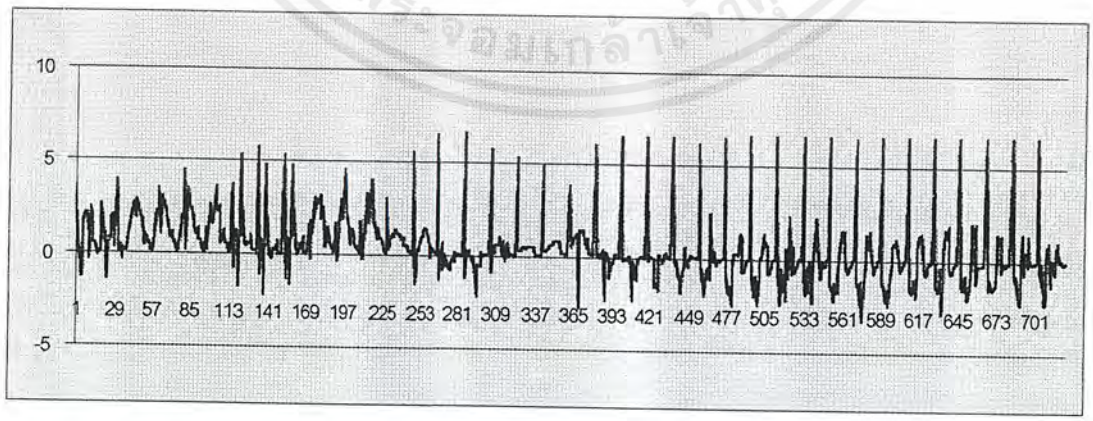
รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.16 กราฟแสดงข้อมูลของค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของคำว่า หาว (ต่อ)

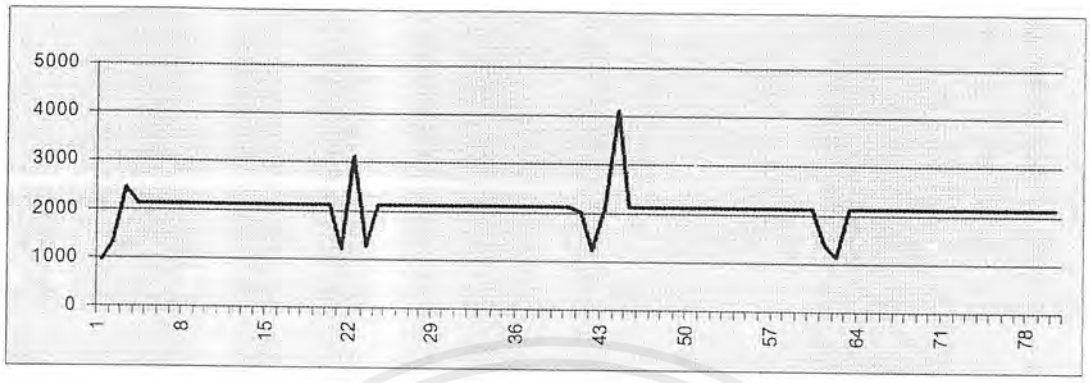
เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **เดิน** แสดงได้ดังนี้



รูปที่ 4.17 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า เดิน

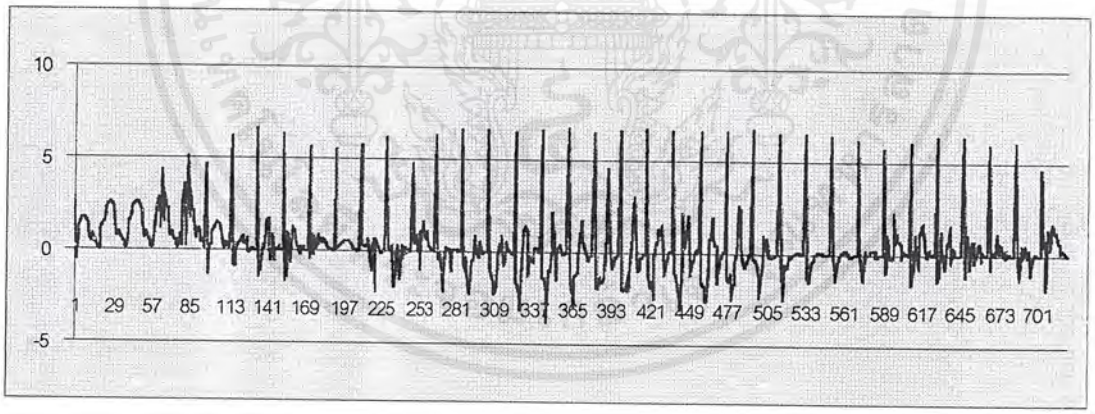
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อนำ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **เดิน** มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ



รูปที่ 4.18 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า เดิน กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

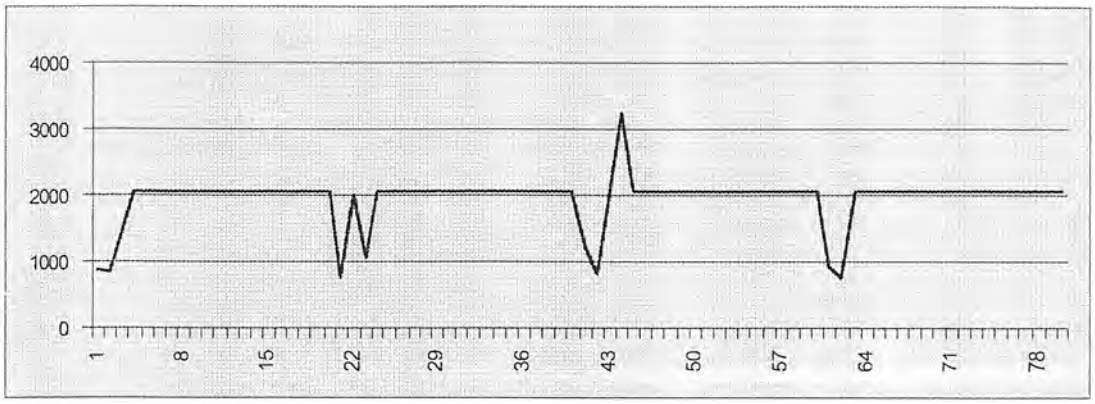
เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง แสดงได้ดังนี้



รูปที่ 4.19 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง

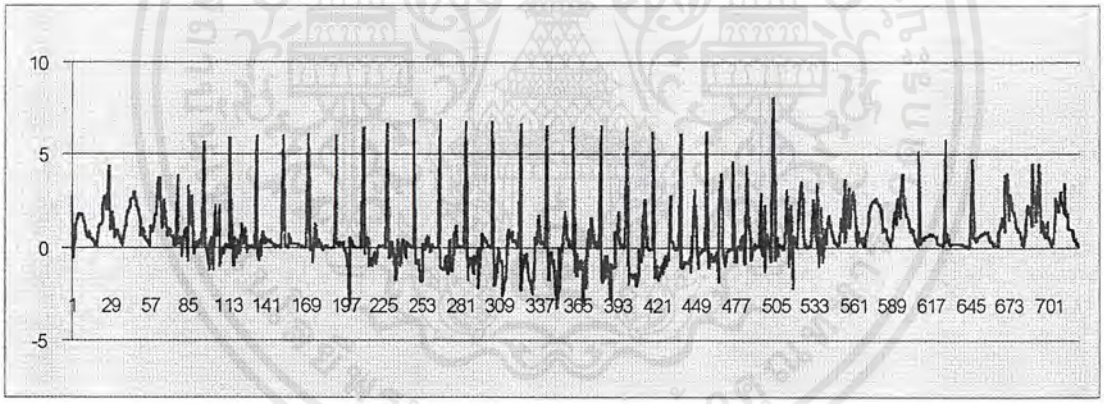
เมื่อนำค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า วิ่ง มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า วิ่ง กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

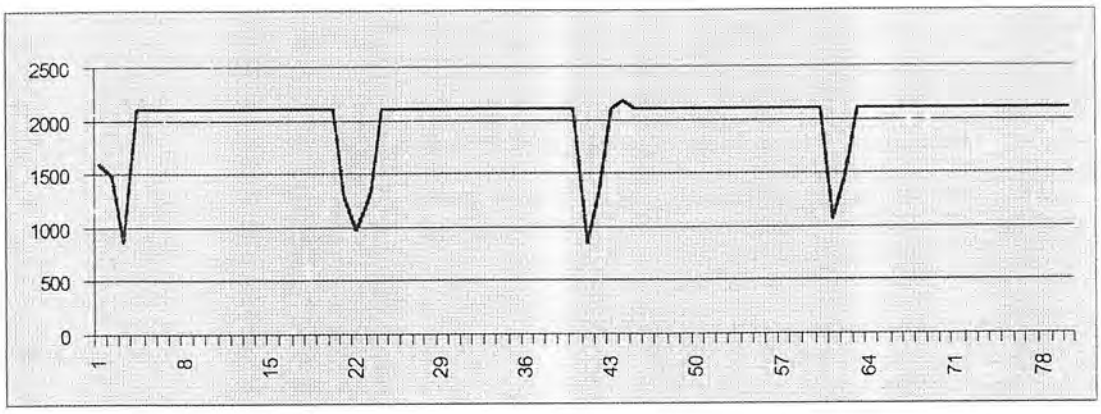
เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **ยิ้ม** แสดงได้ดังนี้



รูปที่ 4.21 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า ยิ้ม

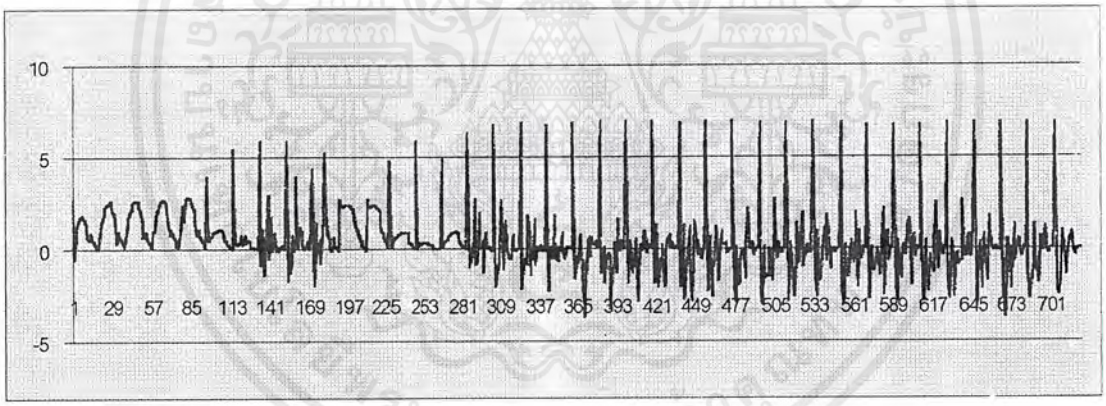
เมื่อนำค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **ยิ้ม** มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.22 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า ยิ้ม กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

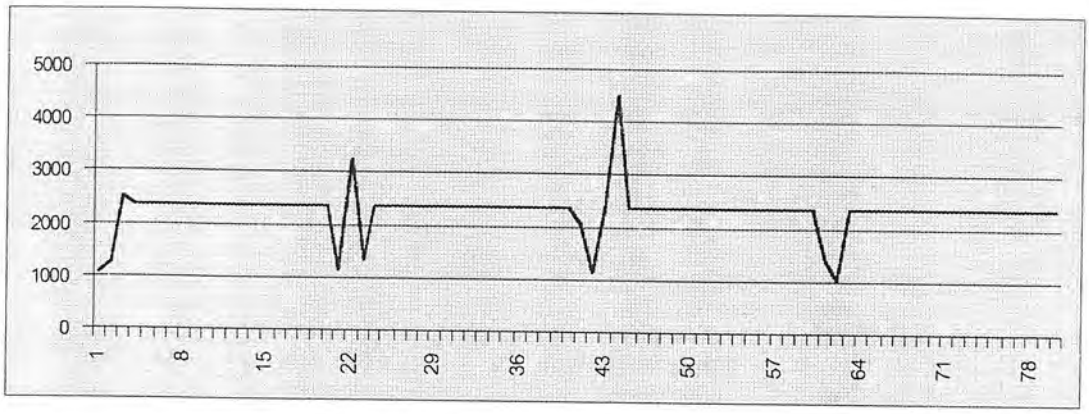
เมื่อทำการทดสอบ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า หาว แสดงได้ดังนี้



รูปที่ 4.23 ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบ คำว่า หาว

เมื่อนำ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก ของเสียงที่เป็นแบบทดสอบ คำว่า **หาว** มาเปรียบเทียบกับค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง จะได้ค่าของความคลาดเคลื่อนซึ่งแสดงได้ดังกราฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.24 ค่าของความคลาดเคลื่อนที่ได้จากการเปรียบเทียบค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของเสียงที่เป็นแบบทดสอบคำว่า หาว กับ ค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักของแบบอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การทดสอบการจดจำเสียงพูด

โดยทำการเก็บเสียงพูดเพื่อนำมาใช้ในการทดลอง โดยจะทำการบันทึกจากนักศึกษาชายจำนวน 9 คน นักศึกษาหญิงจำนวน 9 คน ซึ่งในจำนวนนี้จะทำการแบ่งกลุ่มออกเป็น 2 กลุ่ม A (นักศึกษาชาย 2 คน,นักศึกษาหญิง 2 คน) โดยจะทำการบันทึกเสียง "เดิน", "วิ่ง", "ยิ้ม", "หาว" รวม 4 คำ โดยจะทำการบันทึกคำละ 5 ครั้งต่อคน เพื่อนำเสียงมาเป็นแบบอ้างอิง และกลุ่ม B (นักศึกษาชาย 9 คน, นักศึกษาหญิง 9 คน) โดยทำการบันทึกเสียงเพื่อนำมาเป็นแบบอ้างอิงเช่นเดียวกับกลุ่ม A

การทดลองจะแยกกลุ่มพิจารณาเป็น 2 กลุ่มคือ กลุ่ม A และ กลุ่ม B โดยแต่ละกลุ่มจะนำมาใช้เป็นแบบอ้างอิงโดยมีแบบทดสอบ 2 กลุ่ม เช่นเดียวกัน คือ กลุ่มคนที่นำเสียงมาเป็นแบบอ้างอิงและกลุ่มที่ไม่ได้เป็นกลุ่มที่ทำแบบอ้างอิง ว่าผลการจดจำจะเป็นอย่างไร ซึ่งในการจดจำนี้จะเปรียบเทียบโดยใช้หลักการของเวคเตอร์ควอนไตซ์เซชัน โดยนำสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก มาเป็นพารามิเตอร์ที่ใช้ในการเปรียบเทียบ

การทดสอบที่กล่าวมานั้นสามารถแสดงผลการทดสอบแบ่งเป็นกรณีต่างๆได้ดังตารางต่อไปนี้

แบบอ้างอิง	ถูกต้อง	ผิดพลาด	คิดเป็นร้อยละ
เดิน	6	4	60
วิ่ง	6	4	60
ยิ้ม	3	7	70
หาว	3	7	70
เดิน	6	4	60
วิ่ง	2	8	20
ยิ้ม	8	2	80
หาว	5	5	50
เดิน	7	3	70
วิ่ง	2	8	20
ยิ้ม	5	5	50
หาว	0	10	0

ตารางที่ 4.1 แสดงการจดจำเสียง "เดิน", "วิ่ง", "ยิ้ม", "หาว" โดยใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักเปรียบเทียบ เมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม A และนำเสียงผู้พูดอ้างอิงมาเป็นแบบทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบอ้างอิง	ถูกต้อง	ผิดพลาด	คิดเป็นร้อยละ
เดิน	3	7	30
วิ่ง	2	8	20
ยิ้ม	7	3	70
หาว	4	6	40
เดิน	3	7	30
วิ่ง	2	8	20
ยิ้ม	7	3	70
หาว	2	8	20
เดิน	8	2	80
วิ่ง	5	5	50
ยิ้ม	1	9	10
หาว	2	8	20

ตารางที่ 4.2 แสดงการจดจำเสียง "เดิน", "วิ่ง", "ยิ้ม", "หาว" โดยใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักเปรียบเทียบและใช้จำนวนผู้พูดอ้างอิงในกลุ่ม A และนำเสียงที่ไม่ได้เป็นแบบอ้างอิงมาเป็นแบบทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบอ้างอิง	ถูกต้อง	ผิดพลาด	คิดเป็นร้อยละ
เดิน	6	4	60
วิ่ง	2	8	20
ยิ้ม	7	3	70
หาว	7	3	70
เดิน	6	4	60
วิ่ง	3	7	30
ยิ้ม	6	4	60
หาว	0	10	0
เดิน	6	4	60
วิ่ง	1	9	10
ยิ้ม	2	8	20
หาว	4	6	40

ตารางที่ 4.3 แสดงการจดจำเสียง "เดิน", "วิ่ง", "ยิ้ม", "หาว" โดยใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักเปรียบเทียบ เมื่อใช้จำนวนผู้พูดอ้างอิงในกลุ่ม B และนำเสียงผู้พูดอ้างอิงมาเป็นแบบทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบอ้างอิง	ถูกต้อง	ผิดพลาด	คิดเป็นร้อยละ
เดิน	5	5	50
วิ่ง	6	4	60
ยิ้ม	7	3	70
หาว	10	0	100
เดิน	7	3	70
วิ่ง	5	5	50
ยิ้ม	6	4	60
หาว	8	2	80
เดิน	4	6	40
วิ่ง	5	5	70
ยิ้ม	7	3	50
หาว	6	4	60

ตารางที่ 4.4 แสดงการจดจำเสียง "เดิน", "วิ่ง", "ยิ้ม", "หาว" โดยใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักเปรียบเทียบ และใช้จำนวนผู้พูดอ้างอิงในกลุ่ม B และนำเสียงที่ไม่ได้เป็นแบบอ้างอิงมาเป็นแบบทดสอบ

ตารางที่ 4.1 ถึงตารางที่ 4.4 แสดงการจดจำเสียงตามเงื่อนไขต่างๆที่กำหนดขึ้น ซึ่งสามารถสรุปความถูกต้องได้ดังตารางที่ 4.5 และตารางที่4.6 ดังนี้

ร้อยละความถูกต้อง	ผู้ทำแบบอ้างอิง (กลุ่ม A)	ผู้ทำแบบอ้างอิง (กลุ่ม B)
สูงสุด	80	70
ต่ำสุด	0	0
เฉลี่ย	40	35

ตารางที่ 4.5 แสดงความถูกต้องเมื่อใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก โดยใช้เสียงของผู้ทำแบบอ้างอิงในกลุ่ม A และกลุ่ม B และใช้เสียงของผู้เป็นแบบอ้างอิงมาเป็นแบบทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ร้อยละความถูกต้อง	ผู้ทำแบบอ้างอิง (กลุ่ม A)	ผู้ทำแบบอ้างอิง (กลุ่ม B)
สูงสุด	80	100
ต่ำสุด	10	40
เฉลี่ย	45	70

ตารางที่ 4.6 แสดงความถูกต้องเมื่อใช้สัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก โดยใช้เสียงของผู้ทำแบบอ้างอิงในกลุ่ม A และกลุ่ม B และใช้เสียงของผู้ที่ไม่ได้นำเสียงมาเป็นแบบอ้างอิงมาเป็นแบบทดสอบ

ผลสรุปจาก 2 ตารางข้างต้นได้ดังนี้

จากผลสรุปความถูกต้องดังตารางที่ 4.5 และ ตารางที่ 4.6 พบว่า เมื่อทำการเพิ่มจำนวนผู้พูดเพื่อสร้างแบบอ้างอิงนั้น มีผลต่อความถูกต้องของการรู้จำเสียงพูด โดยจะทำให้ค่าความถูกต้องของผู้ที่ไม่ได้มีเสียงในแบบอ้างอิงนั้นเพิ่มสูงขึ้น และจากผลดังตาราง ในส่วนของความถูกต้องของผู้ที่มีเสียงอยู่ในแบบอ้างอิงนั้นจะเพิ่มสูงขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การแสดงหน้าจอการทำงานและวิธีการใช้งานโปรแกรม

การแสดงหน้าจอการทำงาน และวิธีการใช้งานโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ แบ่งเป็น

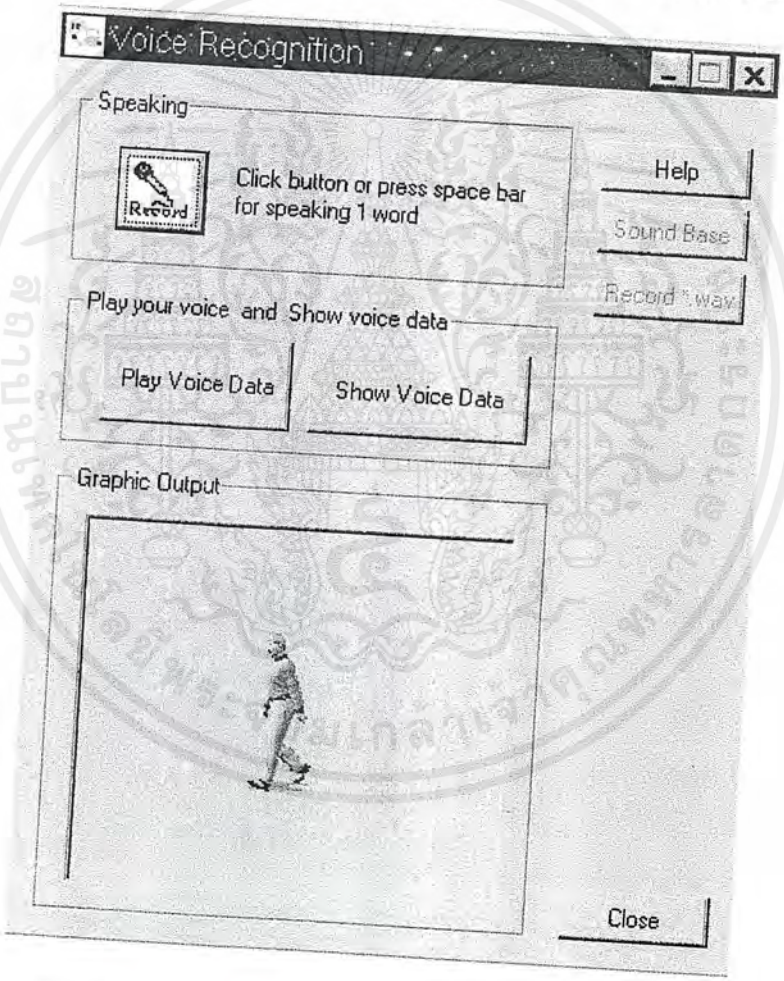
4.3.1 การเรียกเข้าโปรแกรม

4.3.2 การใช้งานโปรแกรม

4.3.3 ข้อจำกัดในการใช้งานโปรแกรม

4.3.1 การเรียกเข้าโปรแกรม

ขั้นตอนในการเรียกเข้าโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ ให้ ทำการเอ็กซีคิวท์ ไฟล์ ที่ชื่อ Speech.exe จะปรากฏหน้าจอโปรแกรมดังภาพ

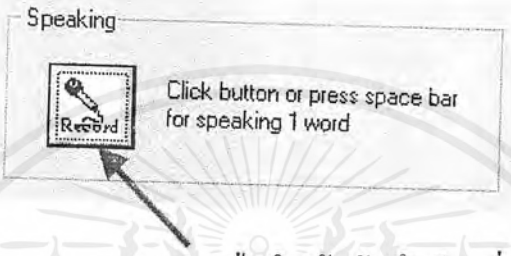


รูปที่ 4.25 หน้าจอโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 การใช้งานโปรแกรม

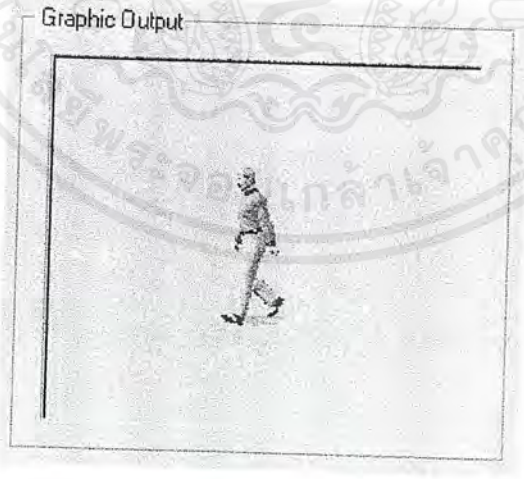
4.3.2.1 หลังจากทำการเข้าสู่โปรแกรมเรียบร้อยแล้ว เมื่อต้องการจะทำการสั่งให้ภาพกราฟฟิกเคลื่อนไหว ให้ทำการกดที่ สเปซบาร์ คีย์ บนคีย์บอร์ด หรือ คลิกเมาส์ ที่ปุ่มรูปไมโครโฟน แล้วทำการพูดคำสั่งทางไมโครโฟนทันที ที่ปล่อย เมาส์หรือ สเปซบาร์คีย์ มีเวลาในการพูดคำสั่ง 1 วินาทีและคำที่ใช้ในการพูดเพื่อสั่งภาพกราฟฟิกให้เคลื่อนไหว ได้แก่ คำว่า เดิน, วิ่ง , ยิ้ม และคำว่า หาว



ปุ่มสำหรับบันทึกเสียงเพื่อ
ให้ภาพกราฟฟิกเคลื่อนไหว

รูปที่ 4.26 แสดงส่วนหน้าจอของโปรแกรมที่ใช้ในการบันทึกเสียง

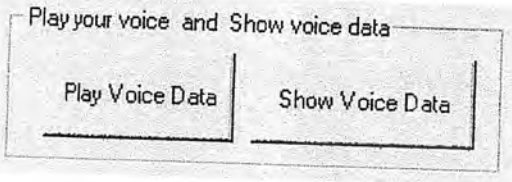
4.3.2.2 ภาพกราฟิกจะแสดงอยู่ในส่วนของ กราฟฟิก เอาท์พุท บนหน้าจอโปรแกรม



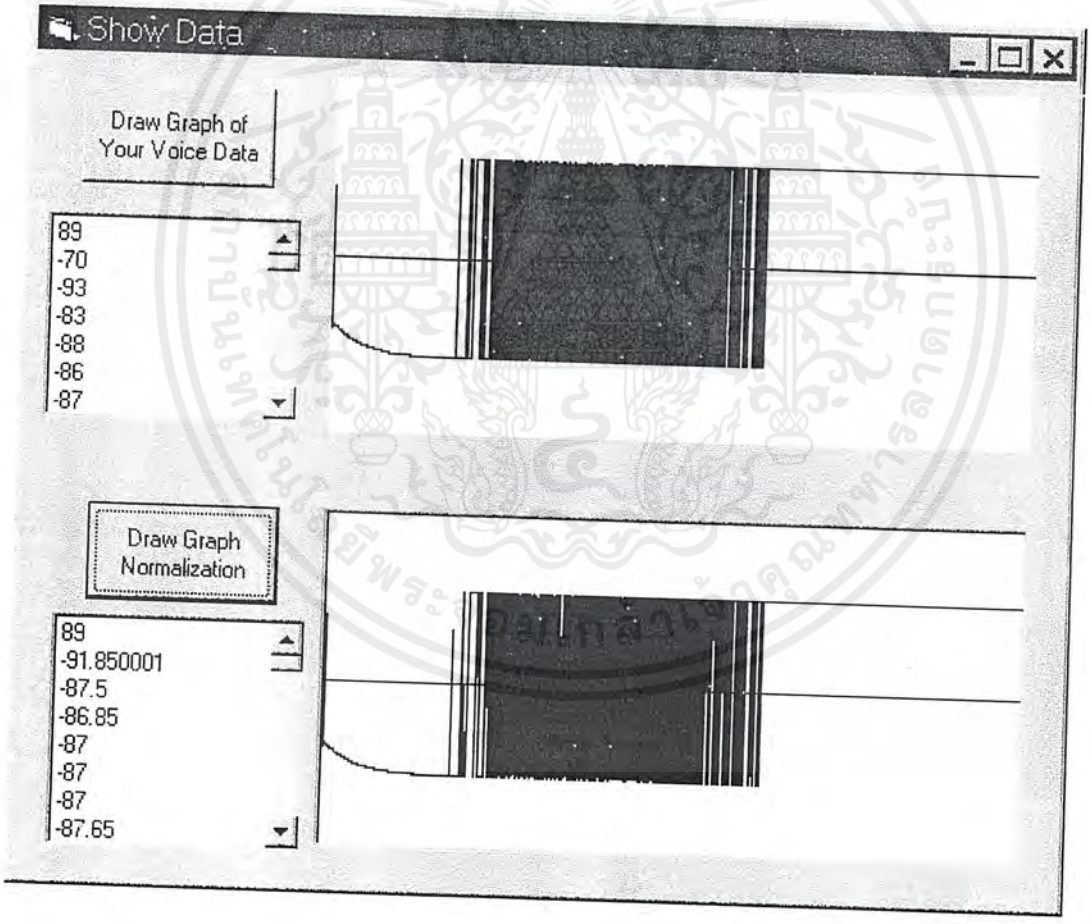
รูปที่ 4.27 แสดงส่วนหน้าจอของโปรแกรมที่ใช้แสดงผลภาพกราฟฟิก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2.3 หากต้องการที่จะฟังเสียงที่ถูกบันทึกเข้าไป ให้ใช้เมาส์ คลิกที่ปุ่ม "Play Voice Data" และหากต้องการดูค่าข้อมูลของเสียง ให้ใช้เมาส์คลิกที่ ปุ่ม "Show Voice Data" จะปรากฏ หน้าต่างย่อยของโปรแกรมซึ่งทำงานในส่วนของการแสดงผลข้อมูลเสียงที่ผู้ใช้ทำการบันทึกเสียงไว้ขึ้นมาดังภาพ



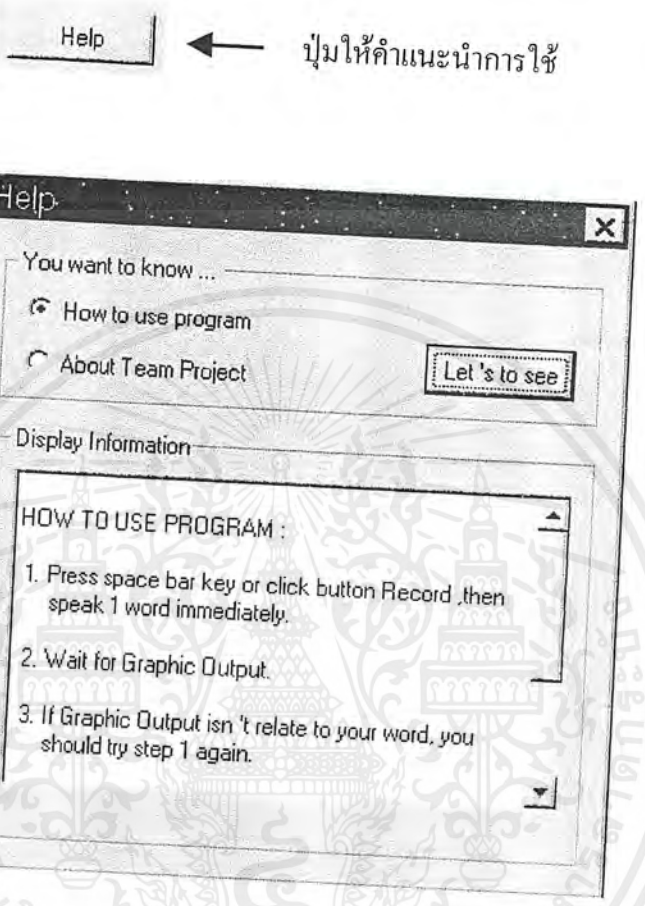
รูปที่ 4.28 แสดงส่วนหน้าจอของโปรแกรมที่ใช้ในการฟังเสียงที่ทำการบันทึกเสียงไว้และแสดงผลข้อมูลของเสียงนั้น



รูปที่ 4.29 แสดงหน้าจอของโปรแกรมที่ใช้ในการแสดงผลข้อมูลเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

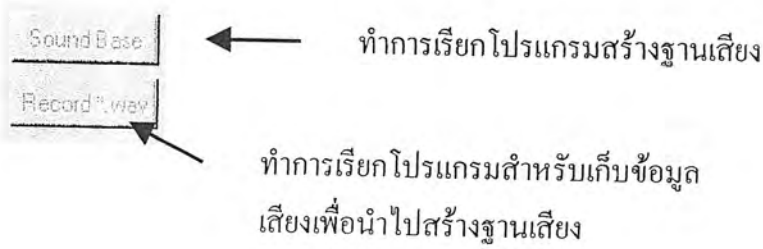
4.3.2.4 หากต้องการทราบรายละเอียดเกี่ยวกับการใช้งานโปรแกรม หรือวิธีการใช้โปรแกรม ให้ใช้เมาส์คลิกที่ ปุ่ม "Help" จะปรากฏหน้าจอของส่วนให้ให้คำแนะนำในการใช้โปรแกรมนี้ขึ้นมา



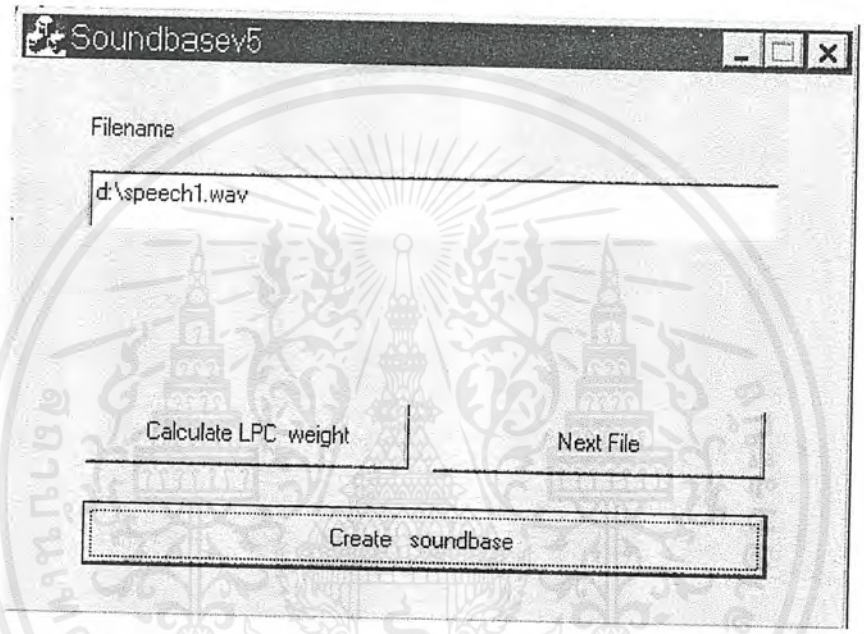
รูปที่ 4.30 แสดงหน้าจอของส่วนให้ความช่วยเหลือในการใช้โปรแกรม

4.3.2.5 ปุ่มบนหน้าจอ 2 ปุ่ม คือปุ่ม "Sound Base" และ ปุ่ม "Record *.wav" ผู้ใช้ไม่สามารถใช้งานได้นั้น เป็นปุ่มที่ใช้เรียกโปรแกรมที่ใช้ในการสร้างฐานเสียง โดยที่ "Sound Base" ทำการเรียกโปรแกรมสร้างฐานเสียง ปุ่ม "Record *.wav" ทำการเรียกโปรแกรมสำหรับเก็บข้อมูลเสียงเพื่อนำไปสร้างฐานเสียงต่อไป ซึ่งทั้ง 2 ปุ่มนี้ ไม่อนุญาตให้ ผู้ใช้ ใช้งานได้ มีไว้สำหรับผู้ทำการพัฒนาโปรแกรมนี้เท่านั้น

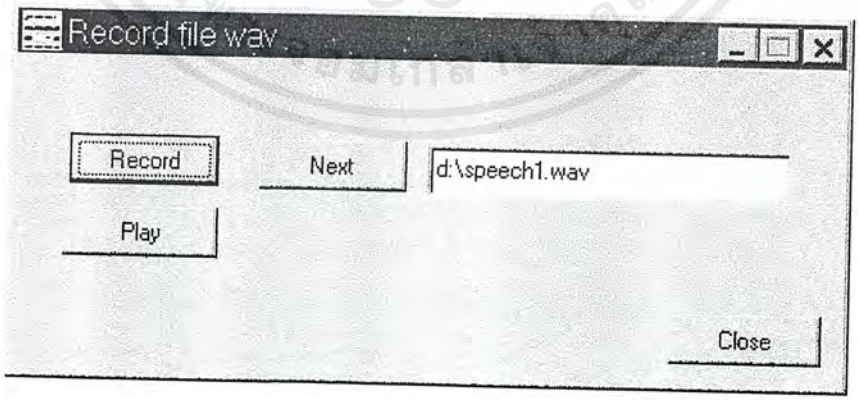
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.31 แสดงส่วนการเรียกโปรแกรมที่ใช้ในการสร้างฐานเสียง



รูปที่ 4.32 แสดงโปรแกรมที่ใช้สร้างฐานเสียง



รูปที่ 4.33 แสดงโปรแกรมที่ใช้เก็บข้อมูลเสียงเพื่อนำมาสร้างฐานเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2.6 ปุ่ม "Close" ใช้สำหรับปิดโปรแกรม

รูปที่ 4.34 แสดงปุ่มปิดโปรแกรม

4.3.3 ข้อจำกัดในการใช้งานโปรแกรม

การใช้งานโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์มีข้อจำกัดทางด้านฮาร์ดแวร์และซอฟแวร์ดังต่อไปนี้

4.3.1.3 โปรแกรมทำงานบนเครื่อง PC Stand Alone

4.3.1.4 ระบบปฏิบัติการ วินโดว์ 98

4.3.1.5 ต้องมีอุปกรณ์ การ์ดเสียงและลำโพง

4.3.1.6 ต้องมีอุปกรณ์ ไมโครโฟน

4.3.1.7 หน่วยความจำอย่างน้อย 32 MB

4.3.1.8 CPU Pemtium II ขึ้นไป

4.3.1.9 ความละเอียดของจอภาพอย่างต่ำ 800 x 600 จุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทวิจารณ์และบทสรุป

5.1 บทสรุป

จากผลการทดลองระบบการจดจำของเสียง "เดิน, "วิ่ง", "ยิ้ม", "หาว" โดยใช้สัญญาณเสียงที่สุ่มด้วยความถี่ 8 กิโลเฮิร์ต ซึ่งใช้แบบอ้างอิงแบบต่างๆ โดยใช้ผู้พูดจำนวน 6 คน, 18 คน คนละ 5 เสียงต่อคำ ตามลำดับ ในการสร้างแบบอ้างอิง

ผลการทดลองสรุปได้ดังนี้

5.1.1 เสียงพูดของคนๆเดียวกัน เมื่อได้พูดแต่ละครั้ง จะพบว่าเสียงที่พูดจะไม่เหมือนกัน

5.1.2 การหาขอบเขตของคำ มีผลต่อความถูกต้องของการจดจำเสียง การหาขอบเขตของคำที่ถูกต้องแน่นอน จะเพิ่มอัตราการจดจำเสียงให้สูงขึ้น จากการศึกษาปัญหาพิเศษที่ได้ทำการทดสอบแล้วได้เลือกใช้ค่าคงที่ในการพิจารณาหาขอบเขตของคำ = 0.3

5.1.3 เนื่องจากสัญญาณที่ผ่านการหาขอบเขตของคำ มีความยาวของเสียงที่แตกต่างกัน ทำให้การหาค่าเฉลี่ยของสัญญาณที่นำมาสร้างแบบอ้างอิงทำได้ไม่แน่นอน เพื่อความถูกต้องของการสร้างแบบอ้างอิง จึงทำการปรับสัญญาณก่อนที่จะมาทำการหาค่าพารามิเตอร์ให้มีความยาวเท่ากันเสียก่อน

5.1.4 การหาค่าพารามิเตอร์ในปัญหาพิเศษนี้ได้เลือกใช้ การถ่วงน้ำหนักของค่าพารามิเตอร์เซปสตรัม และจำนวนออเดอร์ในการหาค่าพารามิเตอร์ ในปัญหาพิเศษนี้ได้เลือกใช้จำนวนออร์เดอร์ เท่ากับ 12

5.1.5 จำนวนคนที่ใช้เสียงมาทำเป็นแบบอ้างอิงมีผลต่อการจดจำ โดยจำนวนคนยิ่งมาก จะให้ผลการจดจำดีขึ้นเมื่อใช้เสียงของผู้ที่ไม่ได้ใช้เป็นแบบอ้างอิงมาทดสอบ แต่ความถูกต้องของการจดจำจะลดลงถ้าเพิ่มจำนวนผู้ที่นำเสียงมาทำแบบอ้างอิง เมื่อใช้เสียงของผู้ที่นำมาทำแบบอ้างอิงทดสอบ

5.1.6 ระบบการรู้จำเฉพาะบุคคลให้ความผิดพลาดที่น้อยกว่าระบบการรู้จำที่ไม่เฉพาะเจาะจงบุคคล และสามารถทำได้ง่ายกว่า ไม่จำเป็นต้องใช้เสียงจำนวนมากเพื่อนำมา สร้างแบบอ้างอิง

5.2 บทการวิจารณ์และแนวการพัฒนา

5.2.1 การทดสอบนี้ได้ทำการทดสอบเฉพาะเสียงของบุคคลในวัย 20-22 ปีเท่านั้น จึงควรพัฒนาต่อไปให้สามารถจดจำได้ในช่วงวัย 18-30 ปี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.2 เนื่องจากการทดสอบที่ใช้นี้ เป็นการทดสอบที่ให้รู้ว่า วิธีการที่ทดสอบนี้สามารถที่จะนำมาใช้ในการจดจำเสียงได้ดีในระดับพอใช้ ซึ่งได้ทำการเขียนโปรแกรมบน Visual C++ Version 6.0 และพบว่าภาษาโปรแกรมนี้มีข้อจำกัดสูงสำหรับปัญหาพิเศษนี้ และมีฟังก์ชันที่สนับสนุนสำหรับการทำปัญหาพิเศษนี้น้อย จึงควรใช้โปรแกรมอื่นๆ ช่วยในการทำปัญหาพิเศษเพื่อการพัฒนาต่อไป

5.2.3 การหาค่าพารามิเตอร์ของสัญญาณเสียงนั้น ได้ทำการหาในโดเมนเวลา เนื่องจากวิธีการไม่ยุ่งยากแต่เมื่อเทียบกับการหาค่าพารามิเตอร์ในโดเมนความถี่ ซึ่งมีวิธีการที่ซับซ้อนกว่า ความถูกต้องของการจดจำจะน้อยกว่าดังนั้นจึงควรพัฒนาโดยการหาค่าพารามิเตอร์ในโดเมนความถี่ เพื่อการจดจำที่ดีขึ้น

5.2.4 ในโครงงานนี้ไม่ได้ทำการทดลองเพื่อ

- หาค่าคงที่ a ที่เหมาะสมในการหาขอบเขตของคำ
  a คือ อัตราส่วนระหว่างค่าพลังงานหรืออัตราตัดศูนย์ของข้อมูลเสียงเฟรมหนึ่งเมื่อเทียบกับค่าพลังงานหรืออัตราตัดศูนย์สูงสุดของข้อมูลเสียงชุดนั้นที่น้อยที่สุดที่ยอมรับว่าข้อมูลเสียง ณ เฟรมนั้น เป็นส่วนที่อยู่ในขอบเขตของคำ
- พิจารณาค่าสัมประสิทธิ์แบบต่างๆ ที่จะใช้ในการเปรียบเทียบคำที่เหมะสม
- หาจำนวนออเดอร์ในการหาค่าพารามิเตอร์ที่เหมาะสม

ดังนั้นเพื่อประโยชน์ในการพัฒนาต่อไป จึงแนะนำให้ทำการทดลองเพื่อพิจารณาค่าหรือปัจจัยข้างต้นให้เหมาะสมที่สุดสำหรับปัญหาพิเศษ เพื่อเพิ่มความถูกต้องในการรู้จำเสียง และลดความผิดพลาดที่มีผลต่อการจดจำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

คำชี้แจงการติดตั้งโปรแกรม

ข้อจำกัดในการใช้งานโปรแกรม

การใช้งานโปรแกรมการวิเคราะห์เสียงเพื่อการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์มีข้อจำกัดทางด้านฮาร์ดแวร์และซอฟแวร์ดังต่อไปนี้

1. โปรแกรมทำงานบนเครื่อง PC Stand Alone
2. ระบบปฏิบัติการ วินโดว์ 98
3. ต้องมีอุปกรณ์ การ์ดเสียงและลำโพง
4. ต้องมีอุปกรณ์ ไมโครโฟน
5. หน่วยความจำอย่างน้อย 32 MB
6. CPU Pemtium II ขึ้นไป
7. ความละเอียดของจอภาพอย่างต่ำ 800 x 600 จุด

ในแผ่น CD-ROM ประกอบไปด้วยรายชื่อไฟล์ดังต่อไปนี้

- Setup.exe
- _inst32i.ex_
- _isdel.exe
- _sys1.cap
- _user1.cap
- Data.tag
- data1.cap
- lang.dat
- layout.bin
- os.dat
- setup.bmp
- Setup.ini
- Setup.ins
- Setup.lid
- Special Project Voice Recognition.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การติดตั้งโปรแกรม

ทำการติดตั้งโปรแกรมโดย ดับเบิ้ลคลิกที่ไฟล์ Setup.exe เพื่อคัดลอกไฟล์ของโปรแกรม เมื่อการคัดลอกไฟล์สิ้นสุดลง จะปรากฏไฟล์ที่มีรายชื่อดังต่อไปนี้อยู่ในโฟล์เดอร์ Voice1.0 ในพาท C:\Voice1.0\

- Speech.exe
- soundbasev5.exe
- Record File.exe
- DrawGraph_Sound.exe
- Help1.txt
- Help2.txt
- sound_norm.txt
- sound_ori.txt
- codebase.dat
- codebase1.dat
- codebase2.dat
- codebase3.dat
- sps.dat
- LPC_Weight.dat
- walk.avi
- run.avi
- yawn1.avi
- smile.avi
- Girl.avi
- speech.wav
- Speech.ico

การเรียกใช้งานโปรแกรม

ผู้ใช้สามารถใช้งานโปรแกรมได้โดยดับเบิ้ลคลิกที่ไฟล์ Speech.exe เพื่อเปิดโปรแกรม จากนั้นสามารถใช้งานโปรแกรมตามวิธีการใช้งานโปรแกรมในบทที่ 4 ตอนที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

ประทวน ล้อมกลาง. "มาสั่งคอมพิวเตอร์ด้วยเสียงกันเถอะ".นิตยสารไมโครคอมพิวเตอร์. ปีที่ 17, ฉบับ 170, กันยายน 2542 หน้า 108-112

นางสาวกรุณา แก้วสมศรี และคณะ.2539. " การต่อหมายเลขโทรศัพท์โดยใช้เสียง. " วิทยานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมโทรคมนาคม, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

นายอภิชัย เอกผาชัยสวัสดิ์ และนายอานนท์ ดารามาลย์. 2541. " การจดจำเสียงพูด." วิทยานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมโทรคมนาคม, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

Thomus W.Parsons. 1987 . Voice and Speech Processing ; McGraw-Hill Book Company.

Paul M.Emble and Bruce Kimble. 1991.C Language Algorithm for Digital Signal Processing : New York, Prentic Hall

Peter Aitken and Scott Jarol. 1995. Visual C++ Multimedia Adventure Set ; Copyright 1995 By The Coriolis Group,Inc.

Microsoft Corporation 1991-1998. "MSDN LIBRALY" .[Online]. Available. http:///www.microsoft .com/msdn/

Texas Instruments. "Speech Synthesis Processors" [Online]. Available. http:// www.ti.com/sc/docs/msp/index.htm

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้