สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การตระหนักรู้เสียงพูด

(VOICE RECOGNITION)

นายดำรัส โสมภีร์

นางสาวฤทัย รุ้งสีทอง

ปพ.
ด5117
2549

เลขหมู่.....
เลขทะเบียน 72891
วัน,เดือน,ปี 25 มิ.ย. 2550

b 11774113
i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมบัณฑิต

สาขาวิศวกรรมระบบควบคุม

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2549

ภาควิชาวิศวกรรมระบบควบคุม คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การตระหนักรู้เสียงพูด

VOICE RECOGNITION

ผู้จัดทำ นายดำรัส โสมภีร์ รหัสประจำตัว 46010238

นางสาวฤทัย รุ้งสีทอง รหัสประจำตัว 46010639

อาจารย์ที่ปรึกษา .....

(ผศ. ถาวร เบญจนราสุทธิ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การตระหนักรู้เสียงพูด

โดย
นายดำรัส โสมภีร์
นางสาวฤทัย รุ้งสีทอง

อาจารย์ที่ปรึกษา
ผศ. ถาวร เบญจนราสุทธิ์

ปีการศึกษา 2549

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้ นำเสนอการศึกษาการวิเคราะห์สัญญาณเสียงพูดของบุคคลในการตระหนักรู้เสียงพูด โดยเป็นการเปรียบเทียบสัญญาณเสียงที่ต้องการทดสอบกับสัญญาณเสียงต้นแบบว่ามีลักษณะสัญญาณแบบเดียวกันหรือไม่ ขอบเขตการศึกษาในปริญญานิพนธ์ฉบับนี้เป็นการออกแบบและสร้างระบบตระหนักรู้จำเสียงพูดแบบคำโดด ซึ่งมีหลักการทำงานสำคัญคือ ระบบจะทำการหาขอบเขตของคำโดยการหาค่าพลังงานและการตัดศูนย์ หลังจากนั้นจึงนำสัญญาณที่ผ่านการตัดศูนย์แล้วมาหาลักษณะเด่นของคำด้วยวิธีการหาค่าสัมประสิทธิ์แอลพีซี และใช้แบบจำลองอ้างอิงฮิดเดนท์มาร์คอฟในการตระหนักรู้เสียงพูด

การทดสอบสมรรถนะของระบบทำโดยการใช้ผู้ทดสอบ ชายและหญิง กลุ่มละ 10 คน เสียงของผู้ทดสอบแต่ละคนจะถูกนำไปฝึกระบบด้วยเสียงอ่านตัวเลขภาษาไทยจากศูนย์ถึงเก้า จากนั้นทำการทดสอบระบบโดยเปรียบเทียบสัญญาณที่ต้องการทดสอบกับสัญญาณต้นแบบ จากผลการทดสอบพบว่า ประสิทธิภาพการตระหนักรู้เสียงพูดเฉลี่ยเสียงผู้ชายและผู้หญิงเป็น 82.5% และ 88% ตามลำดับ ระบบที่สร้างยังมีข้อจำกัดคือระบบจะตระหนักรู้เสียงพูดได้เฉพาะคำที่มีอยู่ในฐานข้อมูลเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VOICE RECOGNITION

By

Mr.Damrat Sommapee

MissRuthai Rungseethong

Advisor

Asst. Prof. Taworn Benjanarasuth

Academic Year 2006

Abstract

In this thesis, a speech signal processing for voice recognition is studied. A comparison between the model signal and the tested signal is conducted to find out whether both are alike. The scope of the study in this thesis is to design and construct the monosyllabic word recognition system. The several important principles are employed. Firstly, energy detection and zero-crossing detection are utilized to search the bound of each word. Then the characteristic of each word is identified by using LPC method. The Hidden Markov Model of each word is finally constructed and used for voice comparison and recognition

The efficiency of the recognition system is evaluated based on 10 males and 10 females voices. Thai words "zoon" to "kao" for 0 to 9 of each person are used to trained the system. After that, tested voices are acquired and compared with the trained model signals. The results show 82.5% success for male voices and 88% success for female voices. However, the restriction of the studied voice recognition technique is that it will work with the words in database only.

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตกรรมประกาศ

การทำปริญญานิพนธ์นี้ผู้วิจัยขอขอบพระคุณ ผศ.ถาวร เบญจนราสุทธิ์ ซึ่งเป็นอาจารย์ควบคุมการวิจัย ที่กรุณาให้แนวคิด คำแนะนำ คำปรึกษา ในการปรับปรุงแก้ไข และเป็นผู้ประสิทธิ์ประสาทวิชา ให้กับผู้วิจัยจนผู้วิจัยสามารถทำการประยุกต์ใช้ทฤษฎีต่างๆเข้ากับโครงงานวิจัย จนทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี

ขอขอบพระคุณเพื่อนๆ และอาจารย์ ในภาควิชาวิศวกรรมระบบควบคุม ที่คอยให้กำลังใจ คำแนะนำ ตลอดจนผู้ให้ความสนใจในปริญญานิพนธ์นี้ ทำให้ผู้วิจัยเกิดแรงบันดาลใจในการทำโครงงานวิจัยจนสำเร็จลุล่วงไปได้อย่างดี

สุดท้ายนี้ผู้วิจัยขอมอบคุณงามความดีของปริญญานิพนธ์ฉบับนี้ให้แด่ บิดา มารดา ครอบครัว ครูบาอาจารย์ และผู้มีพระคุณทุกท่าน

ผู้จัดทำ
นายดำรัส โสมภีร์
นางสาวฤทัย รุ้งสีทอง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 บทนำ

1.1 ความสำคัญของโครงงาน 1

1.2 วัตถุประสงค์ของการทำโครงงาน 1

1.3 ขอบเขตของปริญญานิพนธ์ 2

1.4 วิธีการดำเนินงาน 2

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการทำโครงงาน 2

บทที่ 2 ทฤษฎีและหลักการ

2.1 การเปล่งเสียงของมนุษย์ 3

2.2 กระบวนการผลิตเสียงพูด 4

2.3 เสียงพูดของมนุษย์ 5

2.4 การศึกษาการวิเคราะห์เสียงในช่วงเวลาสั้นๆ 6

2.5 รูปแบบของช่องแคบ 6

2.6 การหาค่าพลังงานของสัญญาณเสียง 7

2.7 การหาค่าอัตราการตัดศูนย์ 8

2.8 การประมาณเชิงเส้น 9

2.9 การหาค่าพารามิเตอร์ 13

2.9.1 การพรีเอมฟาซีส 13

2.9.2 การแบ่งช่วงสัญญาณ 14

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

2.9.3 การวิเคราะห์หาคุณลักษณะของเสียง....15
2.10 การเวทค่าพารามิเตอร์....16
2.11 การจัดระดับเวคเตอร์....17
2.12 ฮิดเดนมาร์คอฟโมเดล....19
2.12.1 สมมุติฐานสำหรับทฤษฎีฮิดเดนมาร์คอฟ....22
2.12.2 ปัญหาพื้นฐานของฮิดเดนมาร์คอฟ....23
2.12.3 การหาค่าปรากฏ....24
2.12.4 การหาลำดับของค่าปรากฏ....27
2.12.5 การหาลำดับสเตทที่ดีที่สุด....29
2.12.6 ระบบการจำลองมาร์คอฟ....31

บทที่ 3 การคำนวณและการสร้าง

3.1 การเตรียมสัญญาณในการวิเคราะห์....34
3.2 การวิเคราะห์สัญญาณขั้นหนึ่ง....36
3.3 การวิเคราะห์สัญญาณขั้นที่สอง....38
3.4 การวิเคราะห์สัญญาณขั้นที่สาม....39
3.5 การเวทค่าพารามิเตอร์....40
3.6 การจัดระดับเวคเตอร์....40
3.6.1. การสร้างโค้ดบุค....40
3.6.2. การเปรียบเทียบ....41
3.7 แบบจำลองฮิดเดนมาร์คอฟ....42
3.8 การตัดสินใจ....44

บทที่ 4 การทดลองและผลการทดลอง

4.1 การทดลอง....46
4.2 ขั้นตอนการเรียนรู้....67
4.3 ขั้นตอนการตระหนักรู้เสียง....67

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

4.3.1การทดสอบการตระหนักรู้เสียงพูด....68
4.3.2 การเปรียบเทียบจำนวนเสียงในแบบอ้างอิง....76

บทที่ 5 สรุปและวิเคราะห์ผลการทดลอง

5.1 สรุปผลการทดลอง....79
5.2 วิจารณ์ผลการทดลอง....81

เอกสารอ้างอิง....82

ภาคผนวก....83

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง **หน้า**

4.1 การตระหนักรู้สียงต้นแบบของผู้ชายโดยใช้ โค้ดบุค 32....68
4.2 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 32....69
4.3 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 32....69
4.4 การตระหนักเสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 32....70
4.5 การตระหนักรู้สียงต้นแบบของผู้ชายโดยใช้ โค้ดบุค 64....70
4.6 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 64....71
4.7 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 64....71
4.8 การตระหนักเสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 64....72
4.9 การตระหนักรู้สียงต้นแบบของผู้ชายโดยใช้ โค้ดบุค 128....72
4.10 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 128....73
4.11 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 128....73
4.12 การตระหนักเสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 128....74
4.13 ค่าความถูกต้องการตระหนักรู้เสียง....75
4.14 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน คนละ 1 ครั้ง....76
4.15 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน คนละ 2 ครั้ง....77
4.16 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน คนละ 3 ครั้ง....77
4.17 ผลการเปรียบเทียบจำนวนเสียงในแบบอ้างอิง....78

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูป **หน้า**

2.1 อวัยวะระบบการพูด....3

2.2 แผนภาพระบบการพูด....5

2.3 บล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย....9

2.4 แสดงส่วนสัญญาณที่ตัดมาวิเคราะห์....14

2.5 การกระจายเฟรมของเสียงพูด....17

2.6 การรวมกลุ่มของเสียงเพื่อสร้างโค้ด....18

2.7 ส่วนประกอบต่างๆ ของฮิดเดนมาร์คอฟ....21

2.8 กระบวนการไปด้านหน้า....26

2.9 กระบวนการถอยหลัง....27

2.10 ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่เสตท i 30

2.11 การตระหนักรู้โค้ดแบบมาร์คอฟ....31

3.1 บล็อกไดอะแกรมขั้นตอนการตระหนักรู้เสียงพูด....33

3.2 แผนผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง....35
โดยการเปรียบเทียบพลังงาน

3.3 แผนผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง....37
โดยการหาอัตราการตัดศูนย์

3.4 ขั้นตอนการนอร์มอลไลซ์สัญญาณ....38

3.5 ขั้นตอนการสร้างแบบจำลองอ้างอิง....43

3.6 ขั้นตอนการตระหนักรู้เสียง....45

4.1 การเตรียมสัญญาณเสียงศูนย์....47

4.2 การเตรียมสัญญาณเสียงหนึ่ง....48

4.3 การเตรียมสัญญาณเสียงสอง....49

4.4 การเตรียมสัญญาณเสียงสาม.... 50

4.5 การเตรียมสัญญาณเสียงสี่....51

4.6 การเตรียมสัญญาณเสียงห้า....52

4.7 การเตรียมสัญญาณเสียงหก....53

4.8 การเตรียมสัญญาณเสียงเจ็ด....54

4.9 การเตรียมสัญญาณเสียงแปด....55

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ(ต่อ)

รูป **หน้า**

4.10 การหาขอบเขตของคำเลขเก้า....56
4.11 การเตรียมสัญญาณเสียงศูนย์....57
4.11 การเตรียมสัญญาณเสียงหนึ่ง....58
4.13 การเตรียมสัญญาณเสียงสอง....59
4.14 การเตรียมสัญญาณเสียงสาม....60
4.15 การเตรียมสัญญาณเสียงสี่....61
4.16 การเตรียมสัญญาณเสียงห้า....62
4.17 การเตรียมสัญญาณเสียงหก....63
4.18 การเตรียมสัญญาณเสียง....64
4.19 การเตรียมสัญญาณเสียงแปด....65
4.20 การเตรียมสัญญาณเสียงเก้า....66
5.1 การหาขอบเขตของสัญญาณเลขสาม ที่ไม่ถูกต้อง....79
5.2 การหาขอบเขตของสัญญาณเลขสาม ที่ถูกต้อง....80

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญของโครงงาน

ปัจจุบันนี้การนำเสียงมาประยุกต์ใช้กับอุปกรณ์อิเล็กทรอนิกส์กำลังได้รับความนิยมอย่างแพร่หลาย เนื่องจากการติดต่อสื่อสารระหว่างมนุษย์กับอุปกรณ์อิเล็กทรอนิกส์นั้น ส่วนใหญ่จะทำได้โดยการป้อนคำสั่งผ่านทางแป้นพิมพ์ (keyboard) หรือเมาส์ (mouse) ซึ่งในทางปฏิบัติพบว่าเป็นวิธีที่ได้รับความนิยมจากบุคคลทั่วไป แต่อาจเป็นวิธีที่ไม่เหมาะสมหรือไม่สะดวกกับบุคคลบางประเภท เช่น คนพิการ จึงได้มีการพัฒนาให้อุปกรณ์อิเล็กทรอนิกส์สามารถรับรู้คำสั่งจากเสียงพูดของมนุษย์ได้ ซึ่งจะทำให้การติดต่อสื่อสารระหว่างมนุษย์และอุปกรณ์อิเล็กทรอนิกส์เป็นไปอย่างสะดวกและง่ายขึ้น

ปริญญานิพนธ์นี้เป็นการศึกษาเพื่อวิเคราะห์สัญญาณเสียง โดยจะเป็นการเปรียบเทียบระหว่างสัญญาณเสียงที่ต้องการทดสอบกับสัญญาณเสียงต้นแบบว่ามีลักษณะสัญญาณแบบเดียวกันหรือไม่ โดยจะแบ่งเป็น 7 ขั้นตอน ขั้นแรกจะเริ่มจากการนำสัญญาณเสียงมาทำการหาขอบเขตของคำ ขั้นที่ 2 เป็นการหาอัตราการตัดศูนย์เป็นการตรวจสอบว่าหาขอบเขตของคำถูกต้องหรือไม่ ขั้นที่ 3 การนอร์มอลไลซ์เป็นการทำให้สัญญาณเสียงที่นำมาวิเคราะห์มีขนาดเท่ากัน ขั้นที่ 4 การหาสัมประสิทธิ์แอลพีซีเป็นขั้นตอนการดึงลักษณะเด่นของสัญญาณ ขั้นที่ 5 การทำดัชนีโค้ดบุคเป็นการลดจำนวนข้อมูลจากขั้นตอนที่แล้วเพื่อที่จะให้เหลือลักษณะเด่นที่แท้จริง ขั้นที่ 6 คือการทำแบบจำลองฮิดเดนมาร์คอฟ เป็นการสร้างแบบจำลองเพื่อนำไปเปรียบเทียบกับสัญญาณเสียงที่ต้องการทดสอบและในขั้นตอนที่ 7 การตัดสินใจ เป็นการนำสัญญาณเสียงที่เป็นแบบจำลองและสัญญาณเสียงที่ต้องการทดสอบมาเปรียบเทียบกัน

1.2 วัตถุประสงค์ของการทำโครงงาน

1. เพื่อศึกษาลักษณะและรูปแบบของสัญญาณเสียง ว่าสัญญาณเสียงในลักษณะหรือรูปแบบใดที่เหมาะสมจะนำมาทำการวิเคราะห์
2. เพื่อศึกษาและวิเคราะห์รูปแบบของการรู้จำเสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตของโครงงาน

1. ทำการวิเคราะห์สัญญาณเสียงโดยใช้โปรแกรมแมทแลป (Matlab)
2. ทำการแทนค่าของสัญญาณเสียงในแต่ละช่วงด้วยสัมประสิทธิ์แอลพีซี (LPC: Linear Predictive Coding) ได้อย่างเสถียร
3. ทำการวิเคราะห์เสียงโดยใช้หลักการ ฮิดเดนมาร์คอฟ (Hidden Markov Model)

1.4 วิธีการดำเนินงาน

1. ศึกษาทฤษฎีพร้อมทั้งค้นคว้าหาข้อมูลเพิ่มเติมที่เกี่ยวกับการรู้จำเสียง
2. ทำการจัดเตรียมและรวบรวมเสียงเพื่อนำมาใช้ในการวิเคราะห์สัญญาณ
3. ทดสอบสัญญาณเสียงที่ได้จากการเก็บรวบรวม เพื่อหาสัญญาณเสียงที่มีประสิทธิภาพมากที่สุด
4. นำสัญญาณเสียงที่เลือกมาทำการวิเคราะห์หาแบบจำลองตามทฤษฎีของการรู้จำเสียง เพื่อให้ได้แบบจำลองที่มีความเสถียรที่สุด
5. ทำการเปรียบเทียบแบบจำลองที่ได้กับสัญญาณเสียงจากตัวอย่าง ที่ได้จากการเก็บรวบรวมเพื่อทดสอบความถูกต้องของแบบจำลอง
6. นำข้อมูลที่ได้จากการวิเคราะห์มาทำการสรุปผล

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการทำโครงงาน

1. ได้รับความรู้เพิ่มเติมจากการศึกษาเกี่ยวกับการรู้จำเสียงและวิธีการใช้โปรแกรมแมทแลปเพื่อนำมาวิเคราะห์ตามขั้นตอนต่างๆ
2. สามารถนำไปประยุกต์ใช้เป็นแนวทางในด้านการใช้เสียง เพื่อทำการติดต่อกับอุปกรณ์อิเล็กทรอนิกส์
3. เพื่อเป็นแนวทางในการศึกษาและค้นคว้าเพิ่มเติมในเรื่องที่เกี่ยวข้องต่อไป

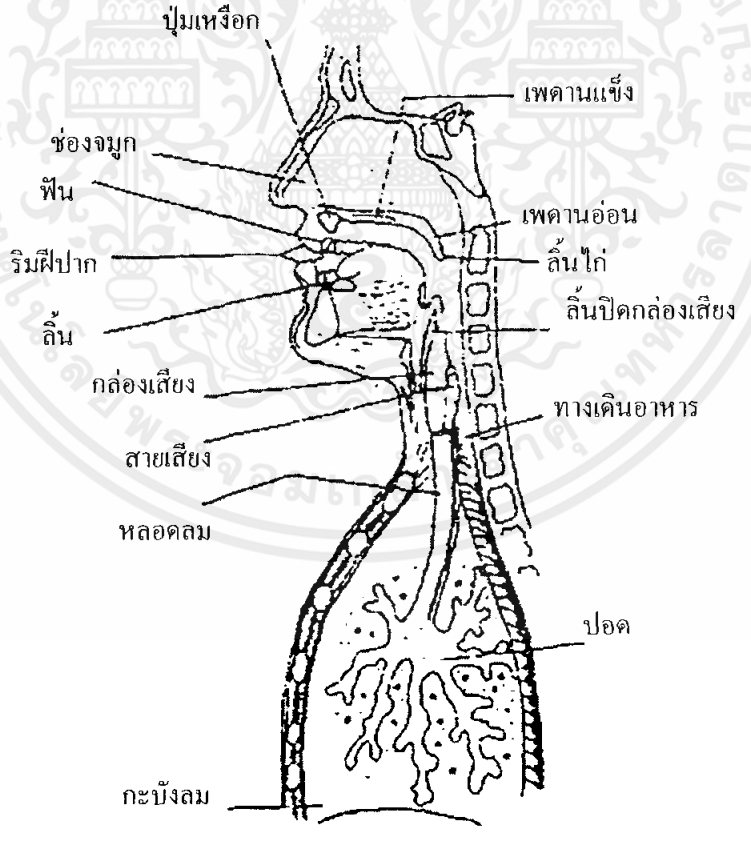
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและหลักการ

ในบทที่ 2 นี้จะกล่าวถึง หลักการในการสร้างระบบการตระหนักรู้เสียงพูด โดยเริ่มจากการศึกษาที่มาของการกำเนิดเสียง หลังจากนั้นก็ทำการศึกษาการวิเคราะห์สัญญาณของเสียงตามลำดับขั้นตอนต่างๆ ดังต่อไปนี้

2.1 การเปล่งเสียงของมนุษย์

จากการศึกษาด้านกายวิภาคศาสตร์ของมนุษย์ วิชาที่ว่าด้วยเสียงของภาษา (Phonetics) และศาสตร์ทางด้านเสียง (Acoustics) ช่วยให้เข้าใจขั้นตอนการทำงานร่วมกันของอวัยวะต่างๆ ในการเปล่งเสียงพูดตลอดจนลักษณะทางกายภาพของเสียงพูดเพื่อนำมาวิเคราะห์ และสร้างแบบจำลองเลียนแบบเสียงพูดของมนุษย์ อวัยวะในการเปล่งเสียงดังรูปที่ 2.1



รูปที่ 2.1 อวัยวะที่เกี่ยวข้องกับระบบการพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำให้เกิดเสียงเป็นหน้าที่หนึ่งของระบบหายใจ การออกเสียงหรือการพูดของมนุษย์แต่ละครั้ง จะต้องมีการทำงานร่วมกันของอวัยวะต่างๆ ของร่างกาย

1. ปอดและกระบังลม ทำหน้าที่สำคัญในการหายใจ และเป็นต้นกำเนิดการไหลของอากาศในกระบวนการผลิตเสียง

2. หลอดลมทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียงและเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

3. กล่องเสียง เป็นอวัยวะพิเศษที่ทำหน้าที่เป็นทางเดินอากาศเวลาหายใจ และเป็นตัวผลิตพัลส์ (Pulse) ของอากาศขณะเปล่งเสียง ซึ่งประกอบด้วยเส้นเสียง และช่องสายเสียง และเส้นเสียงในลักษณะต่างๆ

4. ช่องปากและส่วนของหลอดอาหารตอนต้นอวัยวะกลุ่มนี้ต่อจากกล่องเสียง อาจเรียกว่าอวัยวะกำทอนเสียง ทำหน้าที่กำทอนโดยให้กำทอนทั้งเสียงเกิดจากกล่องเสียง และเสียงที่เกิดภายในช่องปากขนาดของอวัยวะกำทอนเสียงขึ้นอยู่กับตำแหน่งของลิ้น ริมฝีปาก ขากรรไกร และเพดานอ่อน และเปลี่ยนแปลงไปตามการออกเสียง

5. โพรงจมูก เริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสองด้าน ทำหน้าที่กำทอนเสียงร่วมกับช่องปาก เมื่อมีการเปล่งเสียงที่ออกจากจมูก

2.2 กระบวนการผลิตเสียงพูด

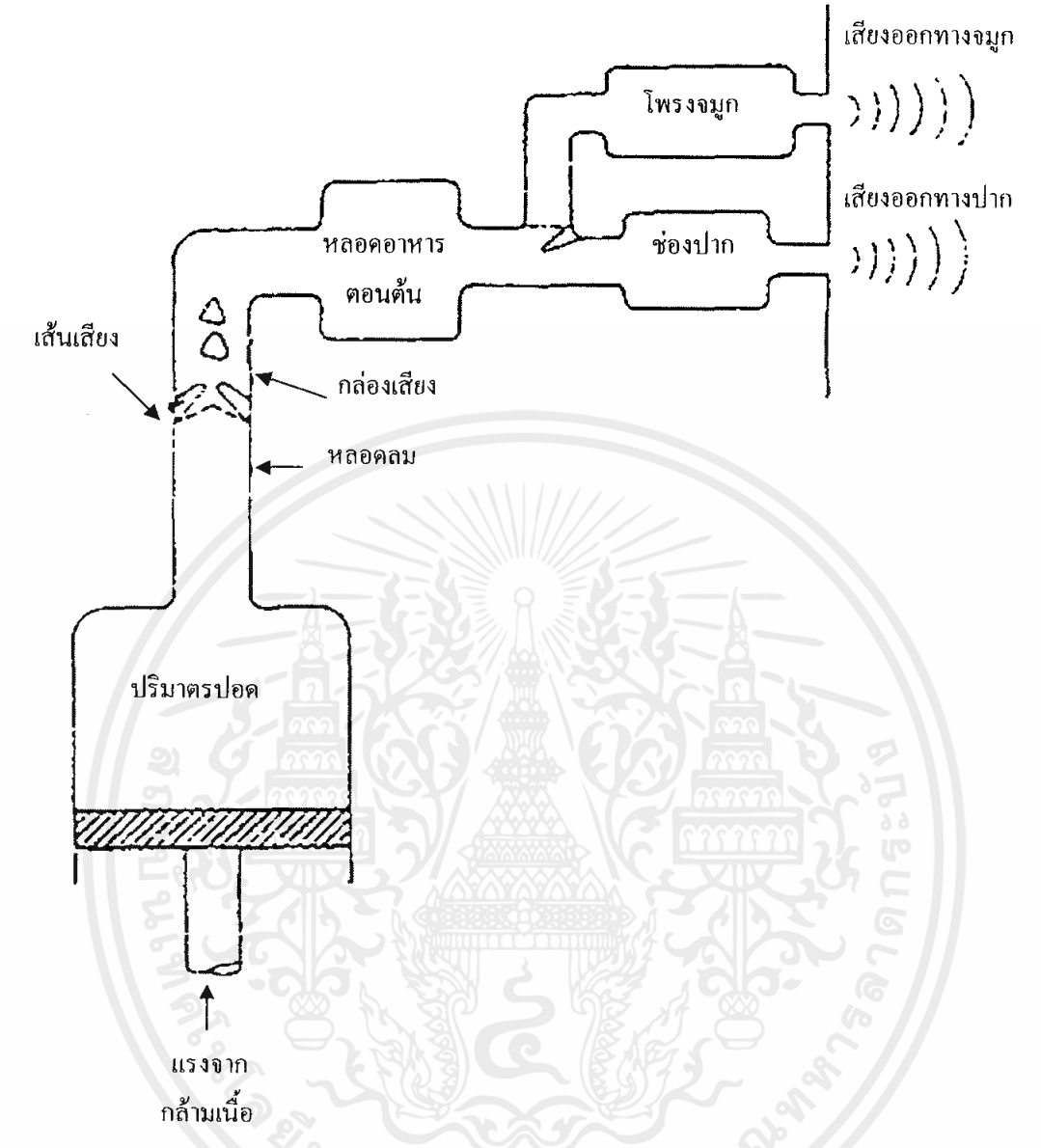
จากระบบเสียงพูด สามารถแสดงเป็นแผนภาพ ของการของระบบกำเนิดเสียงดังรูปที่ 2.2 ซึ่งสามารถจำแนกกลไกการสร้างเสียงพูดของมนุษย์ได้ 3 แบบ ดังนี้

1. อากาศไหลจากปอดจะถูกมอดูเลท (Modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (Quasi–Periodic Pulse–Like Excitation)

2. อากาศที่ไหลออกจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (Noise–Like Excitation)

3. อากาศที่ไหลถูกกัก และกดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศที่มีแรงดันพุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 แผนภาพระบบการพูด

2.3 เสียงพูดของมนุษย์

เสียงพูดเป็นคลื่นตามยาว เกิดจากการสั่นของอนุภาคตัวกลางนั้นคือ อากาศ และทิศทาง การสั่นของอนุภาค จะอยู่ในทิศเดียวกันกับทิศทางของการเคลื่อนที่คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงตามเวลา เสียงพูดแบ่งออกเป็น 2 ชนิด ตามการกำเนิดเสียง หรือโหมด (Mode) การกระตุ้น คือ

1. เสียงก้อง (Voiced) เกิดจากการบังคับอากาศให้ผ่านช่องสายเสียงทำให้มีการเปลี่ยนแปลงความดังหย่อนของเสียง โดยเสียงจะสั่นเกิดเป็นพัลส์ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงก้อง ตัวอย่างเสียงก้องได้แก่ เสียงสระ เสียงพยัญชนะ ที่ต้องออกเสียงจากลำคอ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เสียงไม่ก้อง (Unvoiced) เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูกโดยอวัยวะภายในช่องปาก ริมฝีปาก ขวางการไหลของอากาศให้ผ่านได้เป็นช่องเล็กๆ อากาศจึงไหลผ่านได้อย่างรวดเร็วและปั่นป่วนจนกระทั่งสร้างเป็นเสียงรบกวนช่องความถี่กว้าง ตัวอย่างเสียงที่ไม่ก้อง ได้แก่ เสียงพยัญชนะที่ไม่ได้ออกจากลำคอ

2.4 การศึกษาการวิเคราะห์เสียงในช่วงเวลาสั้นๆ

เนื่องจากสัญญาณเสียงเป็นสัญญาณที่แปรตามเวลา มีการแปรเปลี่ยนได้ไม่แน่นอน เช่น ในขณะที่พูดช้าๆ รูปร่างของโพรงเสียงหรือองค์ประกอบต่างๆ ของเสียงอาจจะไม่เปลี่ยนแปลงในช่วงเวลาที่นานที่สุด 200 มิลลิวินาที แต่ถ้าพูดเร็ว อาจจะมีช่วงที่ไม่มีการเปลี่ยนแปลงสั้นมากคือ 80 มิลลิวินาที

ดังนั้นในการวิเคราะห์เสียงจะให้สัญญาณเสียงมีคุณสมบัติที่เปลี่ยนแปลงสัมพันธ์กับเวลาอย่างช้า เพราะต้องการทำการวิเคราะห์หาพารามิเตอร์ของสัญญาณเสียงพูดในช่วงเวลาสั้นๆ เมื่อมองผ่านช่วงที่เรียกว่า ช๊อตไทม์วินโดว์ (Short–Time Window) เมื่อต้องการเทียบตามเวลาที่เสียงจะมีการเปลี่ยนแปลง เพื่อจะได้มองเห็นเหมือนกับว่า หาพารามิเตอร์นั้นๆ ได้มาจากสัญญาณเสียงที่อยู่ภายในช่องแคบๆ และมีความเสถียรภาพในช่วงเวลาสั้นๆ

โดยกำหนดให้พารามิเตอร์ได้มาจากค่าเฉลี่ยจากการวิเคราะห์ภายในช่องเวลาแคบๆ สำหรับกรณีที่ต้องพิจารณาพารามิเตอร์ต่างๆ ที่มีการเปลี่ยนแปลง ก็จะทำการแบ่งสัญญาณเสียงออกเป็นหลายๆ ช่องหรืออาจเรียกว่า เฟรมหรือกรอบการวิเคราะห์ (Frame Analysis) เพราะฉะนั้นค่าพารามิเตอร์ต่างๆ ที่ต้องการจะสามารถหาได้ทันที เมื่อมีการเปลี่ยนแปลงของสัญญาณเกิดขึ้น

สำหรับในช่วงที่สัญญาณมีการเปลี่ยนแปลงช้า อาจจะให้ช่องแคบมีขนาดใหญ่ประมาณ 100 มิลลิวินาที แต่ถ้าสัญญาณมีการเปลี่ยนแปลงรวดเร็ว ก็ต้องใช้ช่องแคบขนาดเล็กมากๆ ประมาณ 5 ถึง 10 มิลลิวินาที เพื่อมิให้รายละเอียดของสัญญาณในช่องถัดไปสูญหาย

2.5 รูปแบบของช่องแคบ

คุณสมบัติของช่องแคบ (Window) มีดังต่อไปนี้

1. ช่องจะต้องสั้นพอที่จะทำให้คุณสมบัติของเสียงที่กำลังพิจารณาไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญในช่องแคบนี้

2. ช่องแคบจะต้องยาวพอ ที่จะทำให้การจัดเตรียมตัวอย่างของเสียงเพื่อจะนำไปคำนวณหาค่าพารามิเตอร์ให้ได้ตามต้องการอย่าง เช่น ในกรณีที่มีสัญญาณรบกวนเข้ามาแทรกอยู่บางช่วงในสัญญาณเสียงด้วย เลือกใช้ช่องแคบที่มีขนาดใหญ่กว่า เมื่อทำการหาค่าพารามิเตอร์โดยเฉลี่ยจะทำให้ส่วนประกอบของสัญญาณรบกวนถูกตัดทิ้งหรือมองข้าม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ช่องแคบที่เหมาะสม ไม่ควรสั้นเกินกว่าหนึ่งช่วงคาบของสัญญาณเสียงในช่วงที่กำลังวิเคราะห์เงื่อนไขนี้ จะมีผลต่อค่าเฟรมเรท (Frame Rate) ซึ่งก็คือจำนวนครั้งต่อวินาทีที่ทำการวิเคราะห์สัญญาณเสียง โดยการขยับช่องแคบไปเป็นคาบๆ ตามแกนเวลาปกติเฟรมเรทจะมีค่าประมาณ 2 เท่า ของส่วนกลับของขนาดช่องแคบ นั่นคือช่องแคบถัดๆ กันไปจะมีการซ้อนทับกัน 50 เปอร์เซ็นต์

การนำฟังก์ชันของช่องแคบที่มีช่องแคบที่มีช่วงขนาดจำกัด $w(n)$ มาคูณเข้ากับสัญญาณ $s(n)$ จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของช่องแคบ รูปแบบง่ายที่สุดคือ กรอบแบบสี่เหลี่ยม (Rectangular Window) ซึ่งสามารถเขียนสมการได้ดังนี้

$$w(n) = \begin{cases} 1, n = 0,1,..., N-1 \\ 0, n = other \end{cases} \tag{2.1}$$

เมื่อช่องของการวิเคราะห์มีตัวอย่าง N ตัวอย่าง โดยฟังก์ชั่นช่องแคบมีหลายลักษณะแต่ที่นิยมใช้กันมากในการวิเคราะห์สัญญาณเสียงก็ คือ ช่องแคบแบบแฮมมิ่ง (Hamming Window) ซึ่งมีสมการดังนี้

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(2\pi \frac{n}{N}\right), n = 0,1,...N-1 \\ 0, n = other \end{cases} \tag{2.2}$$

2.6 การหาค่าพลังงานของสัญญาณเสียง

พลังงานของสัญญาณ เป็นตัวแทนอันหนึ่ง ที่มักจะนำมาใช้ในการวิเคราะห์ลักษณะต่างๆของสัญญาณทั่วไป โดยพลังงานของสัญญาณที่แปรตามเวลา $s(n)$ ใดๆ สามารถหาได้ดังนี้

$$\text{Energy} = \sum_{n=-\infty}^{\infty} s^2(n) \tag{2.3}$$

แต่สำหรับสัญญาณเสียงซึ่งแปรเปลี่ยนอยู่ตลอด จะต้องแบ่งสัญญาณออกมาพิจารณาเป็นช่วงเล็กๆ ตามแกนเวลา เรียกว่าแบ่งออกเป็นเฟรม เช่น เฟรมละประมาณ 10-30 มิลลิวินาที หรือเฟรมละ 100 ตัวอย่าง เป็นต้น ดังนั้นจะสามารถหาพลังงานของเสียงแต่ละเฟรมได้เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Energy}_l(m) = \sum_{n=0}^{N-1} s^2(n) \tag{2.4}$$

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

N คือจำนวนข้อมูลเสียงในแต่ละเฟรม

การวัดค่าพลังงานในสมการที่ (2.4) นั้นมีข้อจำกัดตรงที่จะมีความไวต่อสัญญาณที่มีขนาดใหญ่ๆ เนื่องจากใช้วิธียกกำลังสองค่าของสัญญาณอินพุท ดังนั้นวิธีการแก้ปัญหาวิธีหนึ่งคือการวัดพลังงานของเสียงโดยใช้ผลรวมค่าสัมบูรณ์ ดังนี้

$$\text{Energy}_l(m) = \sum_{n=0}^{N-1} |s(n)| \tag{2.5}$$

2.7 การหาค่าอัตราการตัดศูนย์

การเกิดอัตราการตัดศูนย์ (Zero-Crossing) จะเกิดขึ้นจากการที่รูปคลื่นของสัญญาณ มีการตัดกับแกนเวลานั่น คือ ค่าของสัญญาณจะมีการเปลี่ยนสัญลักษณ์ทางคณิตศาสตร์นั่นเอง อัตราการเกิดอัตราการตัดศูนย์เป็นเครื่องมืออย่างง่ายที่ใช้อธิบายการเปลี่ยนแปลงของสัญญาณ โดยที่อัตราการตัดศูนย์ของสัญญาณนั้นสามารถนำมาใช้ในการตัดสินว่าสัญญาณเสียงเป็นเสียงก้อง หรือ ไม่ก้อง เนื่องจากเสียงก้องส่วนใหญ่จะมีค่าพลังงานอยู่ในช่วงความถี่ต่ำ เสียงไม่ก้องจะมีพลังงานอยู่ในช่วงความถี่สูงโ ดยที่ค่าอัตราการตัดศูนย์จะมีความสัมพันธ์โดยตรงกับค่าความถี่ของสัญญาณ สามารถสรุปได้ว่าสัญญาณเสียงที่มีค่าอัตราการตัดศูนย์สูงจะเป็นเสียงไม่ก้องและสัญญาณเสียงที่มีค่าอัตราการตัดศูนย์ต่ำจะเป็นเสียงก้อง อย่างไรก็ตามการกำหนดขนาดของค่าอัตราการตัดศูนย์ที่แน่นอนเพื่อจำแนกชนิดของเสียงนั้นจะต้องอาศัยผลจากการทดลองเป็นหลัก การหาอัตราการตัดศูนย์ทำได้โดยสมการ

$$z = \left(\frac{1}{2N}\right) \sum_{n=1}^{N} |Sign(s(n)) - Sign(s(n-1))| \tag{2.6}$$

เมื่อ

$$Sign(s(n)) = \begin{cases} 1, s(n) > 0 \\ 0, s(n) = 0 \\ -1, s(n) < 0 \end{cases}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

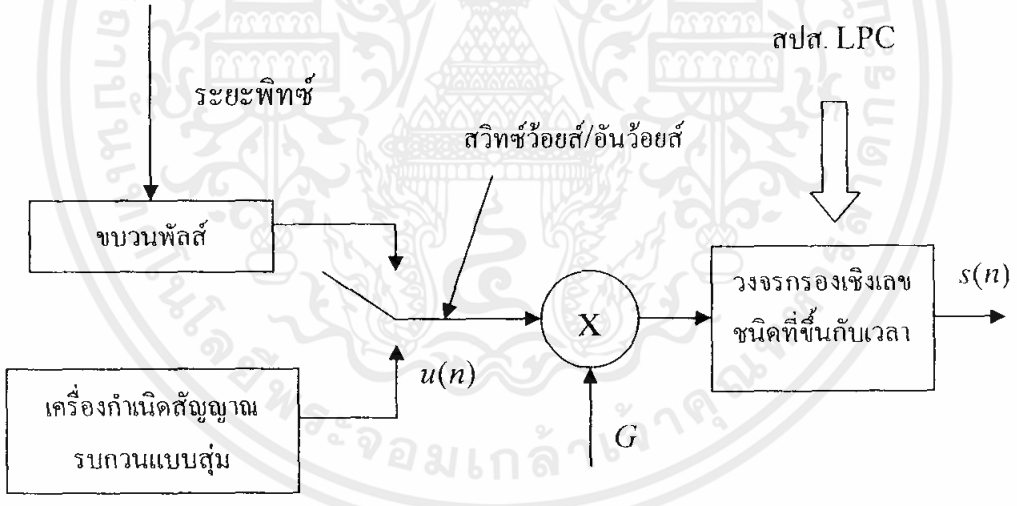
2.8 การประมาณเชิงเส้น

จากหลักการพื้นฐานของการประมาณเชิงเส้น (Linear Prediction Coding: LPC) คือการประมาณค่าสัญญาณจากผลรวมเชิงเส้นของสัญญาณก่อนหน้านี้ สมมติว่าสัญญาณเดิม $s(n)$ การประมาณค่าสัญญาณเป็น $s'(n)$ ดังนั้นสามารถอธิบายการประมาณเชิงเส้นได้ด้วยสมการ

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.7}$$

เมื่อ α_k เป็นค่าคงที่ เรียกวิธีนี้ว่าการประมาณเชิงเส้นอันดับ p โดยมีเงื่อนไขว่าค่า α_k ที่ใช้ในการประมาณจะต้องทำให้ ผลรวมของกำลังสองของความคลาดเคลื่อน หรือ $\{s(n)-s'(n)\}^2$ มีค่าน้อยที่สุด นั่นคือ $\sum e^2(n) = \sum \{s(n)-s'(n)\}^2$ มีค่าต่ำที่สุด ซึ่งในการศึกษานี้จะใช้การประมาณเชิงเส้นด้วยวิธีอัตตสัมพันธ์หรือออโตคอร์รีเลชั่น (Autocorrelation Method)

หลักการพื้นฐานของการประมาณเชิงเส้นและแบบจำลองระบบสร้างสัญญาณเสียง สามารถอธิบายการดำเนินการด้วยบล็อกไดอะแกรม ดังรูปที่ 2.3



รูปที่ 2.3 บล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย

โดยจากรูปที่ 2.3 อธิบายด้วยสมการความสัมพันธ์ ดังนี้

$$s(n) = G \cdot u(n) + \sum_{k=1}^{p} a_k s(n-k) \tag{2.8}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประมาณเชิงเส้นโดยใช้สัมประสิทธิ์ $\{\alpha_k\}$ คือ

$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.9}$$

ดังนั้น ความคลาดเคลื่อนจากการการประมาณเชิงเส้น คือ

$$e(n) = s(n) - s'(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.10}$$

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

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \tag{2.11}$$

จากสมการที่ (2.8) - (2.10) จะเห็นได้ว่าถ้า $\{\alpha_k = a_k\}$ แล้ว

$$e(n) = G \cdot u(n) \tag{2.12}$$

ดังนั้น ค่าผลรวมของกำลังสองของความคลาดเคลื่อน

$$E_n = \sum_{m} e_n^2(m)$$

$$E_n = \sum_{m} [s(m) - s'(m)]^2 \tag{2.13}$$

โดย n คือ ช่วงของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า E_n ต่ำที่สุดจะต้องมีเงื่อนไขว่า

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

จากสมการที่ (2.13)

$$\frac{\partial E_n}{\partial \alpha_i} = -2 \sum_{m} [s(m) - \sum_{k=1}^{p} \alpha_k s(m-k)] s(m-i)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$= -2[\sum_{m} s(m)s(m-i) - \sum_{k=1}^{p} \sum_{m} \alpha_k s(m-k)s(m-i)] \quad (2.14)$$

เมื่อ $i = 1, 2, 3, \ldots, p$

ดังนั้น $\frac{\partial E_n}{\partial \alpha_i} = 0$ ก็ต่อเมื่อ

$$\sum_{k=1}^{p} \alpha_k \sum_{m} s(m-k)s(m-i) = \sum_{m} s(m)s(m-i) \quad (2.15)$$

เมื่อ $i = 1, 2, 3, \ldots, p$

ถ้ากำหนดให้ $\phi_n(i,k) = \sum s(m-k)s(m-i)$ สามารถเขียนสมการที่ (2.15) ได้เป็น

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,k) = \phi_n(i,0) \quad (2.16)$$

จากสมการที่ (2.14) - (2.15) จะเห็นได้ว่า

$$E_n = \sum_{m} s^2(m) - \sum_{k=1}^{p} \alpha_k \sum_{m} s(m)s(m-k) \quad (2.17)$$

และจาก $\phi_n(i,k) = \sum s(m-k)s(m-i)$ จะสามารถเขียนสมการข้างต้นได้เป็น

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

สมมติว่าในหนึ่งเฟรมของสัญญาณที่ตัดมาคำนวณมี N ตัวอย่าง คือ $s(0), s(1), s(2), \ldots, s(N-1)$ ในที่นี้ให้ $s(m) = 0$ เมื่อ $m < 0$ หรือ $m > N-1$ เพราะฉะนั้น

$$\phi(i,k) = \sum_{m} s(m-k)s(m-i)$$

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

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากนิยามอัตตสัมพันธ์หรือออโตคอร์รีเลชั่น

$$R(k) = \sum_{m}^{N-1-k} s(m)s(m+k) \quad , k \;=\; 0,1,2,\ldots,p \tag{2.19}$$

ดังนั้นหากต้องการทำให้ $E_n \;=\; 0$ จากสมการที่ (2.17) และ (2.18) สามารถเขียนใหม่ได้เป็น

$$\sum_{k=1}^{p} \alpha_k R_n(|\, i-k \,|) = R_n(i), \quad i \;=\; 1,2,3,\ldots,p \tag{2.20}$$

จากสมการที่(2.20) สามารถเขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

$$\begin{bmatrix} R_n(0) & R_n(1) & \ldots & R_n(p-1) \\ R_n(1) & R_n(0) & \ldots & R_n(p-2) \\ \ldots & \ldots & \ldots & \ldots \\ R_n(p-1) & R_n(p-2) & \ldots & R_n(p-p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \ldots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \ldots \\ R_n(p) \end{bmatrix} \tag{2.21}$$

หรือ

$$R_n \alpha = r_n \tag{2.22}$$

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

การหาอัตราขยาย G

เมื่อได้ค่า $R_n(0), R_n(1), R_n(2),\ldots, R_n(p)$ แล้ว สามารถหา α จากสมการ(2.22) นั่นคือ

$$\alpha = R_n^{-1} r_n \tag{2.23}$$

จากสมการ (2.12) และ (2.13) จะได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

เพราะฉะนั้น
$$E_n = \sum_{m=0}^{N-1} e^2(m) = G \sum_{m=0}^{N-1} u^2(m) \tag{2.24}$$

จากสมการ(2.18) จะได้

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

$$= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \tag{2.25}$$

จากสมการ(2.25) จะสามารถหาค่า G โดยตรงจาก

$$G^2 = \frac{R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)}{\sum_{m=0}^{N-1} u^2(m)} \tag{2.26}$$

2.9 การหาค่าพารามิเตอร์

ในการหาค่าพารามิเตอร์ของเสียงที่สนใจนั้น ประกอบด้วยขั้นตอนต่างๆ ที่จำเป็นต้องใช้ในการคำนวณ ดังนี้

2.9.1 การพรีเอมฟาซีส

การพรีเอมฟาซีส (Pre-Emphasis) ใช้วงจรอันดับหนึ่ง มีฟังก์ชันถ่ายโอนเป็น

$$H(z) = 1 - az^{-1} \quad , 0.9 \le a \le 1.0 \tag{2.27}$$

นั่นคือ หากให้สัญญาณอินพุทที่เข้ามาเป็น $s(n)$ และสัญญาณที่ผ่านการพรีเอมฟาซีสแล้วเป็น $s'(n)$ สามารถเขียนความสัมพันธ์ของทั้งสองสัญญาณโดยสมการ

$$s'(n) = s(n) - as(n-1) \tag{2.28}$$

หากค่า a ยิ่งมีค่าเข้าใกล้ 1 เท่าใด ความถี่สูงจะถูกขยายมากขึ้นเท่านั้น ค่า a ที่นิยมใช้สำหรับการหาพารามิเตอร์แอลพีซี คือค่า $15/16 = 0.9375$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9.2 การแบ่งช่วงสัญญาณ

การแบ่งช่วงของสัญญาณ (Frame Blocking) ของแต่ละช่วงสัญญาณนั้น เป็นการพิจารณาสัญญาณออกเป็นทีละช่วงหรือเฟรม หากสมมติว่าแต่ละเฟรมมีขนาด N จุดข้อมูลและมีช่วงข้อมูลเหลื่อมกับเฟรมก่อนหน้า M จุดข้อมูล ดังนั้นเงื่อนไขของการกำหนดขนาดของการแบ่งช่วงของสัญญาณจะต้องทำให้ค่า M ต้องมีค่าน้อยกว่า N และหากการแบ่งช่วงสัญญาณนี้กำหนดให้ลำดับของช่วงหรือเฟรมเป็น l^{th} สัญญาณในแต่ละเฟรมเป็น $x_l(n)$ และมีจำนวนเฟรมทั้งหมดเป็น L ดังนั้นค่า $x_l(n)$ ในแต่ละเฟรมของสัญญาณเสียง $s'(n)$ สามารถเขียนได้ดังสมการ

$$x_1(n) = s'(Ml+n), \quad n = 0,1,...,N-1 \quad \text{และ} \quad l = 0,1,...,L-1 \tag{2.29}$$

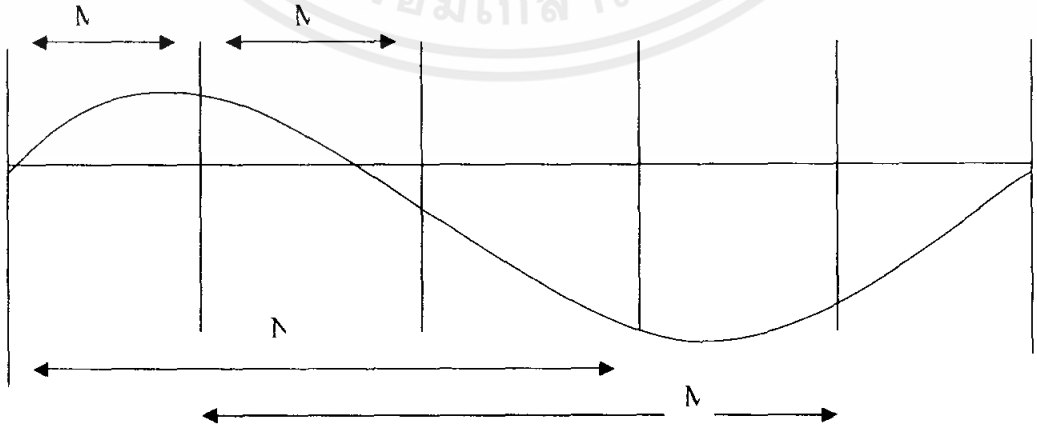
จากการแบ่งสัญญาณออกเป็นช่วงในการวิเคราะห์นั้น มักเกิดความไม่ต่อเนื่องตรงขอบของเฟรมที่ถูกแบ่งช่วงสัญญาณแล้ว ดังตัวอย่างซึ่งแสดงในรูปที่ 2.4 ดังนั้นจึงต้องทำการลดความไม่ต่อเนื่องดังกล่าวโดยอาศัยช่องแคบแบบแฮมมิ่ง ซึ่งนิยามโดยสมการ

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right), \quad 0 \le n \le N-1 \tag{2.30}$$

ซึ่งฟังก์ชันวินโดว์แฮมมิ่งนี้มีคุณลักษณะของสเปกตรัมที่เหมาะสมในการวิเคราะห์เสียง คือ มีความถี่รีโซลูชันสูง (High Frequency Resolution) เพราะมีโลบหลัก (Main Lobe) แคบ และมีการลดทอน (Attenuation) นอกช่วงความถี่ผ่านต่ำสูง เพราะขนาดของไซด์โลบ (Side Lobe) มีค่าน้อย

การนำสัญญาณ $x_l(n)$ ผ่านการวินโดว์ $w(n)$ สามารถคำนวณได้โดยสมการ

$$x_l'(n) = x_l(n)w(n) \quad , 0 \le n \le N-1 \tag{2.31}$$



รูปที่2.4 แสดงส่วนสัญญาณที่ตัดมาวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9.3 การวิเคราะห์หาคุณลักษณะของเสียง

การหาคุณลักษณะของเสียง ซึ่งเป็นขั้นตอนหลักในการหาค่าพารามิเตอร์ จะอาศัยวิธีลิเนียร์พรีดิกทีฟ หรือแอลพีซี การวิเคราะห์แบบลิเนียร์พรีดิกทีฟนี้ เป็นการวิเคราะห์เสียงพูดที่ดีวิธีหนึ่งในการใช้หาค่าพารามิเตอร์พื้นฐานของเสียง ซึ่งพารามิเตอร์ที่ได้เหล่านี้มีความถูกต้องและให้เวลาในการคำนวณไม่มากนัก โดยที่จะต้องเตรียมสัญญาณในการวิเคราะห์ก่อน แล้วจึงนำสัญญาณนี้ไปใช้ในวิธีลิเนียร์พรีดิกทีฟ ขั้นตอนในการวิเคราะห์สามารถแบ่งได้ดังนี้

1. การหาออโตคอรีเลชั่น สัญญาณแต่ละเฟรมที่ผ่านช่องแคบมาแล้วนั้น จะถูกนำมาหาค่าออโตคอรีเลชั่นดังสมการ

$$R(m) = \sum_{n=0}^{N-1-m} x'(n)x'(n+m), \quad m = 0,1,\ldots, p \tag{2.32}$$

เนื่องจากการวิเคราะห์หาออโตคอรีเลชั่นสำหรับอันดับ p ที่มีค่ามาก จะทำให้การประมาณเสียงมีความใกล้เคียงมากยิ่งขึ้น แต่ถ้าอันดับ p มีค่ามากเกินไปจะทำให้การคำนวณมีความยุ่งยากและใช้เวลานาน โดยทั่วไปนิยมให้ค่า p มีค่าระหว่าง 8 ถึง 16

อนึ่งอันดับ p ที่ใช้งานนี้เรียกว่าอันดับการวิเคราะห์แอลพีซี และการดำเนินการในการหาคุณลักษณะของเสียงนี้ว่าการประมาณเชิงเส้นอันดับ p

2. การหาสัมประสิทธิ์แอลพีซี วิธีการวิเคราะห์หาค่าพารามิเตอร์แอลพีซีที่นิยมใช้กันวิธีหนึ่ง ได้ถูกพัฒนาขึ้นโดยเลวินสัน (Levinson) และ เดอร์บิน (Durbin) ซึ่งเรียกว่าวิธีการทำซ้ำของเลวินสัน-เดอร์บิน (Levinson-Durbin Recursive Method) ซึ่งสามารถสรุปขั้นตอนการคำนวณได้ดังนี้

$$E(0) = R(0) \tag{2.33}$$

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} R(|i-j|)}{E(i-1)}, \qquad 1 \le i \le p \tag{2.34}$$

$$\alpha_j^i = k_i \tag{2.35}$$

$$\alpha_j^i = \alpha_j^{(i-1)} - k_i \alpha_{(i-j)}^{(i-1)}, \quad 1 \le j \le i-1 \tag{2.36}$$

$$E(i) = (1 - k_i^2) E(i-1) \tag{2.37}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมการจะถูกคำนวณแบบย้อนกลับ สำหรับ $i = 1,2,..., p$ และผลลัพธ์สุดท้ายจะได้ว่า พารามิเตอร์ $\{a_m(k), 1 \le k \le m\}$ ก็คือค่าพารามิเตอร์แอลพีซี หรือ α_j' ที่อันดับ p ซึ่งผลลัพธ์สุดท้าย คือ $\{a_m(k), 1 \le k \le p\}$ โดยที่ p คืออันดับของพารามิเตอร์ที่ต้องการ

3. สัมประสิทธิ์เซปสตรัม สัมประสิทธิ์เซปสตรัมนี้เป็นพารามิเตอร์ที่มีความน่าเชื่อถือกว่าสัมประสิทธิ์แอลพีซี ทั้งยังมีความสัมพันธ์ใกล้ชิดกับการรับรู้เสียงตามความรู้สึกของมนุษย์โดยแท้จริง ในขั้นตอนนี้จะสามารถหาสัมประสิทธิ์เซปสตรัมได้โดยตรงจากพารามิเตอร์แอลพีซี ดังขั้นตอนต่อไปนี้

$$C_o = \ln G \tag{2.38}$$

$$C_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) C_k a_{m-k} \quad , 1 \le m \le p \tag{2.39}$$

$$C_m = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) C_k a_{m-k} \quad , m > p \tag{2.40}$$

$$Q \approx \left(\frac{3}{2} \right) P \tag{2.41}$$

2.10 การเวทค่าพารามิเตอร์

การเวท ค่าพารามิเตอร์ (Parameter Weighting) หรือถ่วงน้ำหนักสัมประสิทธิ์เซปสตรัมที่ได้ช่วงลำดับต้นๆและท้ายๆของเฟรมที่นำมาวิเคราะห์จะเกิดความคลาดเคลื่อนมากกว่าบริเวณส่วนอื่น เพราะฉะนั้น จะทำการลดความคลาดเคลื่อนเหล่านั้นด้วยฟังก์ชันเวทติ้ง จะได้พารามิเตอร์สุดท้าย คือ

$$C_m' = C_m * W_m \tag{2.42}$$

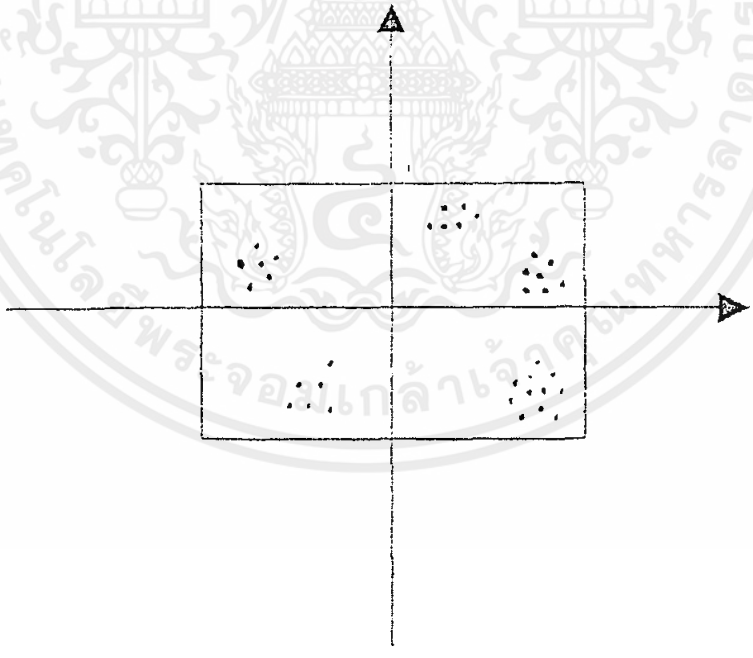
เมื่อ

$$W_m = \left[1 + \frac{Q}{2} \sin\left(\frac{\pi m}{Q} \right) \right], \quad 1 \le m \le Q \tag{2.43}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.11 การจัดระดับเวคเตอร์

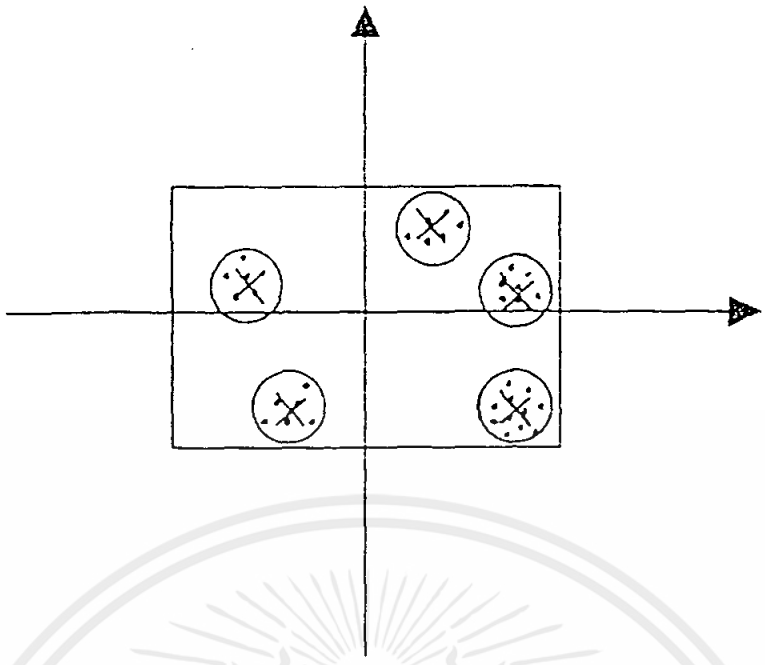
เวคเตอร์ควอนไตเซชั่น (Vector Quantization) เป็นวิธีการลดมิติ หรือจำนวนข้อมูลเวคเตอร์อินพุท หรือเซตเทรนนิ่ง หรือ พารามิเตอร์ที่ได้จากขั้นการหาสัมประสิทธิ์แอลพีซี จำนวนข้อมูลเวคเตอร์จะถูกเลือกมากลุ่มหนึ่งซึ่งใช้เป็นตัวแทนของข้อมูลจำนวนหนึ่ง หรือเรียกว่า การหาโค้ดบุค (Codebook) อินพุทที่เข้ามาจะถูกทำการเปรียบเทียบกับโค้ดบุค ที่มีอยู่โดยจะพิจารณาว่าอินพุทที่เข้ามานั้นห่างจากโค้ดบุคใดน้อยที่สุด อินพุทดังกล่าวจะถูกแทนด้วยเวคเตอร์โค้ดนั้น อินพุททุกตัวเป็นสมาชิกของเวคเตอร์ใดๆ จะถูกนำมาหาจุดศูนย์กลางร่วมใหม่ และนำจุดศูนย์กลางที่ได้นี้ไปทำการหาความคลาดเคลื่อนกับสมาชิกทุกตัว ถ้าค่าที่ได้มีค่ามากกว่าค่าที่กำหนดไว้ค่าหนึ่งหรือค่าที่ยอมรับได้ ก็จะนำศูนย์กลางใหม่นั้นไปเป็นโค้ดบุคแทน และจะทำการจัดกลุ่มอินพุทเข้ากับโค้ดบุคใหม่ที่ได้และหาค่าความคลาดเคลื่อนกับสมาชิกทุกตัว และจะทำการจัดกลุ่มอินพุทเข้ากับโค้ดบุคใหม่ที่ได้ และหาค่าความคลาดเคลื่อนอีกครั้งทำอย่างนี้ซ้ำๆ จนกระทั้งค่าความคลาดเคลื่อนที่มีค่าน้อยถึงค่าที่ยอมรับได้ก็จะถือได้ว่าโค้ดบุคที่ดีที่สุดจะเป็นตัวแทนของอินพุททั้งหมด จะสังเกตได้ว่าทุกครั้งที่มีการหาโค้ดบุคใหม่นั้น ค่าความคลาดเคลื่อนที่ได้จะมีค่าลดลงทุกครั้งด้วย โดยการกระจายเสียงเฟรมของเสียงพูด และการรวมกลุ่มของเสียงอธิบายได้ตามรูปที่ 2.5 และ2.6



รูปที่ 2.5 การกระจายเฟรมของเสียงพูด

72891

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 การรวมกลุ่มของเสียงเพื่อสร้างโค้ด

ตัวอย่างเวคเตอร์ควอนไตซ์เซชั่น สมมติให้เวคเตอร์แต่ละตัวมี 2 มิติ และทำการหาค่า โค้ดบุค ขนาด 8 เวคเตอร์ ทั้งหมดจะถูกจัดเข้ากลุ่มของโค้ดบุคต่างๆ แล้วทำการหาจุดศูนย์กลางใหม่โดยการเฉลี่ยค่าเวคเตอร์สมาชิกทุกตัวที่อยู่ในกลุ่มเดียวกัน ผลที่ได้คือ โค้ดบุค 8 ตัว เป็นตัวแทนของเวคเตอร์ทั้งหมด การทำงานของการควอนไตซ์แบบเวคเตอร์ แบ่งเป็น 2 ขั้นตอนคือการสร้างโค้ดบุคและขั้นตอนการเปรียบเทียบ

1. **การสร้างโค้ดบุค**

จากขั้นตอนการประมาณเชิงเส้นของเสียงตัวอย่างจำนวนมากจะได้เทรนนิ่งเซตซึ่งประกอบด้วยเวกเตอร์สเปคตรัมจำนวน L เฟรม; $x=\{x_i;1\le i\le L\}$ เฟรมละ P มิติ $x=[x_1,x_2,...,x_p]$ แล้วนำข้อมูลที่ได้มาทำการสร้างกลุ่มของแบบอ้างอิง

ในระบบการรับรู้เสียงพูดแบบต่างๆจะใช้อ้างอิงคำหนึ่งๆจากผู้พูดจำนวนมากเพื่อที่จะได้ครอบคลุมถึงความแปรปรวนต่างๆ ที่เกิดขึ้นระหว่างผู้พูดแต่ละคน เนื่องจากถ้าใช้แบบจำลองอ้างอิงจำนวนมากเวลาที่ใช้ในการตอบสนองจะมาก เนื้อที่ความจำสำรองที่ใช้เก็บแบบอ้างอิงจะเพิ่ม และเมื่อเพิ่มแบบอ้างอิงไประดับหนึ่งความถูกต้องในการรับรู้จะเริ่มคงที่ ดังนั้นการจัดกลุ่มแบบอ้างอิงใหม่ เพื่อให้ได้แบบอ้างอิงที่เหมาะสม และสามารถใช้เป็นตัวแทนของแบบอ้างอิงที่มีอยู่ทั้งหมดได้ อัลกอริทึ่มที่ใช้ได้แก่ K-Means Algorithm (Lloyd Algorithm)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ขั้นตอนการเปรียบเทียบ

เวคเตอร์ควอนไตซ์เซชั่น ที่ใช้เพื่อออกแบบการรับรู้เสียงนั้น มีจำนวนควอนไตเซอร์ M ตัว ซึ่งหมายถึงมี M ระดับเสียงเพื่อการรับรู้ แต่ละระดับเสียงพิจารณามาจากเซตของข้อมูลการเทรนนิ่งซึ่ง M เป็นดัชนีวัดระดับ แต่ละเซตของการเทรนนิ่งในแต่ละระดับจะเก็บเสียงที่อยู่ในระดับเดียวกันเมื่อมีเสียงที่ไม่ทราบ x_i จะเป็นอินพุทเข้าไปยังทุกๆควอนไตเซอร์ ค่าดัชนีระดับ (Index) ที่ถูกเลือก จะเป็นระดับที่มีความคลาดเคลื่อนน้อยที่สุด ซึ่งความคลาดเคลื่อนเฉลี่ยนั้นหาได้จากการวัดระยะทางโดยใช้วิธีการหาควาคลาดเคลื่อนกำลังสองรวม

$$d(v_1, v_2) = \|v_1 - v_2\| = \sum_{i=0}^{k-1} (x_i - y_i)^2 \tag{2.44}$$

2.12 ฮิดเดนมาร์คอฟโมเดล

นิยามและหลักการพื้นฐาน

ฮิดเดนมาร์คอฟโมเดล(Hidden Markov Model: HMM) คือ เซตของสเตท (State) ที่มีจำนวนจำกัด โดยในแต่ละสเตท จะเกี่ยวเนื่องกับฟังก์ชันการแจกแจงความน่าจะเป็น ในส่วนของการเปลี่ยนสเตทจะถูกกำกับโดยชุดข้อมูลที่เรียกว่าความน่าจะเป็นในการเปลี่ยนสเตท แต่ละสเตทจะสามารถสร้างผลลัพธ์หรือที่เรียกว่าค่าสังเกต หรือค่าปรากฏ (Observation) ขึ้นมาตามการแจกแจงความน่าจะเป็น ซึ่งผลที่ได้นั้นเป็นเพียงแค่ผลลัพธ์ที่สร้างขึ้นมา โดยแบบจำลองเท่านั้น ไม่ได้มาจากการมองเห็นสเตทของระบบภายในแต่อย่างใด ดังนั้นจะพบว่าสเตทเหล่านี้จะถูกซ่อนจากผู้สังเกต

ในการที่จะให้นิยามอย่างสมบูรณ์ จำเป็นต้องกล่าวถึงส่วนประกอบของฮิดเดนมาร์คอฟดังต่อไปนี้

1. จำนวนสเตทของแบบจำลอง N
2. จำนวนสัญลักษณ์ในการสังเกต M แต่ถ้าเป็นกรณีที่ผลจากการสังเกตมีลักษณะต่อเนื่อง M จะมีค่าอนันต์
3. เซตของความน่าจะเป็นในการเปลี่ยนสเตท $A = \{a_{ij}\}$

$$a_{ij} = \Pr\{q_{t+1} = j \mid q_t = i\}, \quad 1 \le i, j \le N \tag{2.45}$$

เมื่อ q_t คือ สเตทปัจจุบันของระบบ

q_{t+1} คือ สเตทถัดไปของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความน่าจะเป็นในการเปลี่ยนสเตทจะต้องมีคุณสมบัติตามข้อกำหนดทางความน่าจะเป็นโดยทั่วไป คือ

$$a_{ij} \geq 0, \quad 1 \leq i, j \leq N \tag{2.46}$$

และ

$$\sum_{j=1}^{N} a_{ij} = 1, \quad 1 \leq i, j \leq N \tag{2.47}$$

4. การแจกแจงความน่าจะเป็นในแต่ละ สเตท $B = \{b_{ij}(k)\}$

$$b_j(k) = \Pr\{o_t = v_k \mid q_t = j\}, \quad 1 \leq j \leq N, \ 1 \leq k \leq M \tag{2.48}$$

เมื่อ v_k คือ สัญลักษณ์ผลลัพธ์ที่สังเกตเห็นในตัวที่ k

o_t คือ ตัวแปรปัจจุบันที่ทำการสังเกต

q_t คือ สเตทปัจจุบันของระบบ

การแจกแจงความน่าจะเป็นในแต่ละ State จะต้องมีคุณสมบัติตามข้อกำหนด คือ

$$b_j(k) \geq 0, \quad 1 \leq j \leq N, \ 1 \leq k \leq M \tag{2.49}$$

และ

$$\sum_{k=1}^{M} b_j(k) = 1, \quad 1 \leq j \leq N \tag{2.50}$$

ในกรณีที่จะใช้ค่าจากการสังเกตเป็นค่าต่อเนื่อง จะใช้ฟังก์ชันความหนาแน่นความน่าจะเป็นแบบต่อเนื่อง แทนที่การใช้ความน่าจะเป็นแบบไม่ต่อเนื่อง ซึ่งโดยปกติแล้วสามารถประมาณค่าความหนาแน่นความน่าจะเป็นได้จากสมการต่อไปนี้

$$b_j(O) = \sum_{m=1}^{M} c_{jm} \eta(\mu_{jm}, U_{jm}, O) \tag{2.51}$$

เมื่อ O คือ เวกเตอร์ที่จะทำการสร้างแบบจำลองการทำงานของ

c_{jm} คือ สัมประสิทธิ์ส่วนผสมของส่วนผสมที่ m ในสเตทj

η คือ ฟังก์ชันความหนาแน่นความน่าจะเป็นที่มีลักษณะเป็นรูปครึ่งวงรี

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมาตรที่มีเวกเตอร์ค่าเฉลี่ย μ_{jm} และมีเมตริกซ์ความแปรปรวนร่วม U_{jm} โดยสามารถหาได้ตามสมการที่(2.52)

$$\eta(\mu,U,O)=\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\det(U)|}}\exp\left\{-\frac{1}{2}(O-\mu)^{T}U^{-1}(O-\mu)\right\} \tag{2.52}$$

โดยปกติแล้วจะใช้ฟังก์ชันความหนาแน่นแบบเกาส์เซี่ยน (Gaussian) สำหรับ η ได้ ส่วนผสม c_{jm} จะต้องเป็นไปตามข้อกำหนดทางความน่าจะเป็นดังนี้

$$c_{jm}\geq 0,\quad 1\leq j\leq N, 1\leq m\leq M \tag{2.53}$$

และ

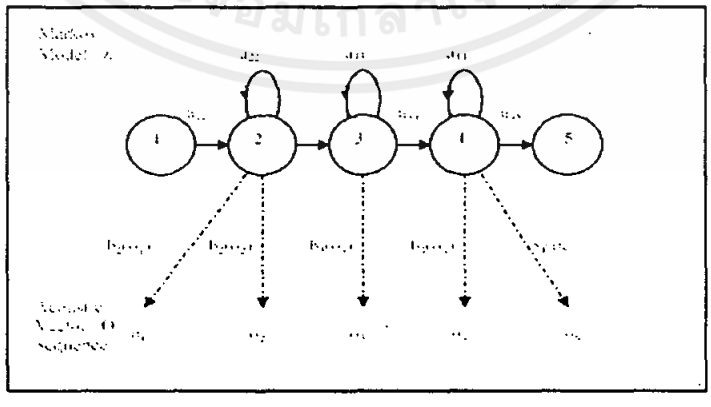
$$\sum_{m=1}^{M}c_{jm},\quad 1\leq j\leq N \tag{2.54}$$

5. การแจกแจงความน่าจะเป็นของสเตทเริ่มต้น $\pi=\{\pi_i\}$

$$\pi_i=\Pr\{q_1=i\},\quad 1\leq i\leq N \tag{2.55}$$

เมื่อ q_1 คือ สเตทแรกของระบบ

ดังนั้นจะใช้สัญลักษณ์ $\lambda=(A,B,\pi)$ แทนฮิดเดนมาร์คอฟที่มีการแจกแจงความหนาแน่นเป็นแบบไม่ต่อเนื่อง และ $\lambda=(A,c_{jm},\mu_{jm},U_{jm},\pi)$ แทนฮิดเดนมาร์คอฟที่มีความหนาแน่นเป็นแบบต่อเนื่อง



รูปที่ 2.7 ส่วนประกอบต่างๆ ของฮิดเดนมาร์คอฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.12.1 สมมติฐานสำหรับทฤษฎีฮิดเดนมาร์คอฟ

การที่จะนำฮิดเดนมาร์คอฟมาใช้ในงานจริงจำเป็นจะต้องมีการตั้งสมมติฐานเหล่านี้ขึ้นมาสำหรับทฤษฎีดังกล่าว เพื่อลดภาระทางด้านการคำนวณและการวิเคราะห์ สมมติฐานกล่าวได้ดังต่อไปนี้

สมมติฐานของกระบวนการมาร์คอฟ

ดังที่ได้กล่าวมาแล้วในนิยามของฮิดเดนมาร์คอฟความน่าจะเป็นในการเปลี่ยนสเตทคือ

$$a_{ij} = \Pr\{q_{t+1} = j \mid q_t = i\} \tag{2.56}$$

จากสมการดังกล่าว จะพบว่าสเตทถัดไปจะเป็นสเตทใดขึ้นอยู่กับสเตทปัจจุบัน ซึ่งจะเรียกสิ่งนี้ว่าสมมติฐานของมาร์คอฟ (The Markov Assumption) และจะทำให้ได้ผลลัพธ์เป็นแบบจำลองฮิดเดนมาร์คอฟอันดับที่หนึ่ง อย่างไรก็ตามโดยทั่วไปแล้วสามารถกำหนดให้การพิจารณาสเตทถัดไปขึ้นอยู่กับ สเตทก่อนหน้ามากกว่าหนึ่งสเตทซึ่งถ้าพิจารณาสเตทถัดไป โดยคำนึงถึงสเตทก่อนหน้าจำนวน k สเตทจะเรียกแบบจำลองนี้ว่าสเตทอันดับที่ k แต่การทำเช่นนี้ก็จะส่งผลให้การคำนวณมีความซับซ้อนมากขึ้นเช่นกัน

สมมติฐานของความไม่เปลี่ยนแปลงไปตามเวลา

ความน่าจะเป็นในการเปลี่ยนสเตทจะเป็นอิสระต่อเวลาจริงๆ (The Stationary Assumption) ที่การเปลี่ยนสเตทนี้เกิดขึ้นโดยที่สามารถอธิบายให้อยู่ในเชิงคณิตศาสตร์ดังนี้

$$\Pr(q_{t_1+1} = j \mid q_{t_1} = i) = \Pr(q_{t_2+1} = j \mid q_{t_2} = i) \tag{2.57}$$

สำหรับทุกค่าของ t_1 และ t_2

สมมติฐานของความอิสระของผลลัพธ์

สมมติฐานข้อนี้กล่าวว่าผลลัพธ์จากการสังเกตในตำแหน่งปัจจุบันจะมีค่าเป็นอิสระในทางสถิติกับผลลัพธ์ (The Output Independence Assumption) ตัวก่อนหน้า โดยถ้าหากชุดของค่าจากการสังเกตคือ $O = O_1, O_2, ..., O_T$ โดยถ้าพิจารณาตามสมมติฐานที่ได้กล่าวมาในข้างต้นจะได้ว่าสำหรับ λ

$$\Pr(O \mid q_1, q_2, ..., q_T, \lambda) = \prod_{t=1}^{T} \Pr(o_t \mid q_t, \lambda) \tag{2.58}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อย่างไรก็ตามสมมติฐานข้อนี้ต่างจาก 2 ข้อที่ได้กล่าวมาในข้างต้น เพราะความจำกัดในการใช้งาน ซึ่งบางกรณีแล้วสมมติฐานนี้อาจจะไม่เหมาะสม ซึ่งจะกลายเป็นข้อด้อยอย่างหนึ่งของฮิดเดนมาร์คอฟ

2.12.2 ปัญหาพื้นฐานของฮิดเดนมาร์คอฟ

จากที่กล่าวมาในข้างต้นทั้งหมด สามาระสรุปปัญหาพื้นฐานที่จะต้องมีการแก้เพื่อนำแบบจำลองนี้ไปทำงานจริงได้ทั้งหมด 3 ข้อด้วยกัน ได้แก่

1. ทำอย่างไรจึงจะสามารถคำนวณค่า $\Pr(O \mid \lambda)$ ได้อย่างมีประสิทธิภาพ โดยกำหนดชุดข้อมูลจากการสังเกตเป็น $O = O_1, O_2, ..., O_T$ และแบบจำลองฮิดเดนมาร์คอฟเป็น $\lambda = (A, B, \pi)$
2. ทำอย่างไรจึงจะสามารถเลือกชุดของสเตท $Q = q_1, q_2, ..., q_T$ ที่มีความสอดคล้องกับแบบจำลองมากที่สุด โดยกำหนดชุดข้อมูลจากการสังเกตเป็น $O = O_1, O_2, ..., O_T$ และแบบจำลอง ฮิดเดนมาร์คอฟเป็น $\lambda = (A, B, \pi)$
3. ทำอย่างไรจึงจะสามารถปรับค่าตัวแปรต่างๆ ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อทำให้ค่า $\Pr(O \mid \lambda)$ มีค่าสูงที่สุด

<u>ปัญหาข้อที่หนึ่ง</u> คือ ปัญหาในการหาลำดับของค่าปรากฏ (Evaluation Problem) ซึ่งจะเกี่ยวกับการหาค่าความน่าจะเป็นของชุดข้อมูลจากการสังเกตเมื่อเทียบกับแบบจำลอง หรืออาจกล่าวได้อีกมุมมองหนึ่งคือ เป็นปัญหาในการให้คะแนนในการเปรียบเทียบความเหมือนของแบบจำลองกับชุดข้อมูลจากการสังเกตที่ป้อนเข้ามา ซึ่งจะพบว่ามุมมองหลังที่ได้กล่าวถึงนั้นมีประโยชน์อย่างมากในการนำแบบจำลองนี้ไปใช้งาน ยกตัวอย่างเช่น ถ้าพยายามที่จะเลือกแบบจำลองที่มีความเหมือนมากที่สุด เทียบกับชุดข้อมูลจากการสังเกตออกมาจากชุดของแบบจำลอง คำตอบจากการแก้ปัญหาข้อแรกนี้จะสามารถทำให้เลือกแบบจำลองที่ถูกต้องออกมาได้อย่างมีประสิทธิภาพ

<u>ปัญหาข้อที่ 2</u> เรียกอีกอย่างหนึ่งว่า การหาลำดับสเตทที่เหมาะสม (Decoding Problem) ซึ่งจะเป็นปัญหาเกี่ยวกับการที่พยายามเปิดเผยส่วนที่ซ่อนอยู่ในแบบจำลอง กล่าวคือ เป็นการหาลำดับของสเตทที่มีความเหมาะสมกับแบบจำลองนั้นๆ มากที่สุด โดยในการหาสเตทดังกล่าวจะขึ้นอยู่กับข้อกำหนด ซึ่งในความเป็นจริงแล้วสามารถเลือกใช้งานข้อกำหนดได้ในหลายรูปแบบ ในแต่ละรูปแบบก็จะมีความแตกต่างกันไป ซึ่งจุดประสงค์ของการแก้ปัญหาข้อนี้มีหลายข้อ เช่น เพื่อเรียนรู้โครงสร้างของแบบจำลอง เพื่อหาลำดับสเตทที่เหมาะสมกับการรู้จำเสียงพูดแบบต่อเนื่องหรือเพื่อหาค่าทางสถิติของแต่ละแบบจำลอง เป็นต้น

<u>ปัญหาสุดท้าย</u> เรียกว่าการหาลำดับสเตทที่ดีที่สุด(Learning Problem) จะเกี่ยวกับการที่พยายามที่จะทำให้แบบจำลองมีความเหมาะสมกับชุดข้อมูลจากการสังเกตที่ให้เข้ามามากที่สุด ซึ่งจะเรียกชุดข้อมูลที่ใช้ในการปรับค่าตัวแปรของแบบจำลองนี้ว่า ชุดข้อมูลฝึกหัด (Training

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sequence) เพราะว่าชุดข้อมูลเหล่านี้จะถูกนำไปฝึกแบบจำลองให้รู้จำข้อมูลเหล่านั้น ปัญหาการฝึกหัดนี้เป็นปัญหาที่มีความสำคัญมากอย่างหนึ่งในการประยุกต์ใช้งานแบบจำลองในหลายๆ ด้าน เนื่องจากการฝึกหัดนี้จะทำให้สามารถปรับค่าของแบบจำลอง ให้เหมาะสมกับชุดข้อมูลฝึกหัดจากการสังเกตได้ กล่าวอีกนัยหนึ่งคือ เพื่อสร้างแบบจำลองของปรากฏการณ์ที่ดีที่สุดนั่นเอง

2.12.3 การหาลำดับของค่าปรากฏ

ในการที่จะหาค่า $\Pr(O|\lambda)$ เมื่อกำหนดแบบจำลอง $\lambda=(A,B,\pi)$ และชุดข้อมูลจากการสังเกต $O=O_1,O_2,...,O_T$ สามารถทำได้โดยการคำนวณความน่าจะเป็นสำหรับทุกๆ ชุดของ สเตท ความยาว T ที่เป็นไปได้ของแบบจำลองนั้นๆ ซึ่งสำหรับแต่ละลำดับของสเตท $Q=q_1,q_2,...,q_T$ เมื่อ q_1 คือ สเตทเริ่มต้นจะได้ความน่าจะเป็นของชุดข้อมูลที่สังเกตได้ O สำหรับแต่ละลำดับของสเตทคือ

$$\Pr(O|Q,\lambda)=\prod_{t=1}^{T}\Pr(O_t|q_t,\lambda) \tag{2.59}$$

และจากสมมติฐานความเป็นอิสระของผลลัพธ์ จะได้

$$\Pr(O|Q,\lambda)=b_{q_1}(O_1)\cdot b_{q_2}(O_2)\cdot\ldots\cdot b_{q_T}(O_T) \tag{2.60}$$

สามารถเขียนความน่าจะเป็นของลำดับสเตท Q ได้คือ

$$\Pr(Q|\lambda)=\pi_{q_1}a_{q_1q_2}a_{q_2q_3}\ldots a_{q_{T-1}q_T} \tag{2.61}$$

ในส่วนของความน่าจะเป็นร่วมของ O และ Q ซึ่งก็คือความน่าจะเป็นที่ทั้ง O และ Q เกิดขึ้นพร้อมกัน สามารถหาได้ง่ายๆ โดยการคูณทั้งสองสมการข้างบนเข้าด้วยกัน

$$\Pr(O,Q|\lambda)=\Pr(O|Q,\lambda)\cdot\Pr(Q|\lambda) \tag{2.62}$$

ดังนั้นสามารถหาความน่าจะเป็นของ O เมื่อให้แบบจำลอง λ สามารถหาได้จากการรวมความน่าจะเป็นร่วมด้านบนสำหรับทุกๆ ชุดของสเตท ที่เป็นไปได้ทั้งหมดเข้าด้วยกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{aligned}\Pr(O|\lambda) &= \sum_{all\ Q} \Pr(O|Q,\lambda)\cdot\Pr(Q|\lambda) \\ &= \sum_{q_1,q_2,\ldots,q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2)\ldots a_{q_{T-2}q_T} b_{q_T}(O_T)\end{aligned} \tag{2.63}$$

เมื่อทำการวิเคราะห์ความซับซ้อนในการคำนวณ พบว่าการหาค่าความน่าจะเป็นโดยตรงในแต่ละครั้ง จะเกี่ยวข้องกับการคำนวณถึง $2T\cdot N^T$ ครั้ง ซึ่งในการคำนวณแต่ละครั้งสำหรับแต่ละสเตทจะมีการคำนวณ $2T$ ครั้ง ส่วน N^T ก็คือจำนวนสเตทที่เป็นไปได้ทั้งหมด การคำนวณดังกล่าวไม่สามารถที่จะนำมาใช้งานได้ในความเป็นจริง แม้จะมีสเตทในจำนวนไม่มากนักก็ตาม เช่น ถ้าแบบจำลองมี $N=5$, $T=100$ จำนวนครั้งในการคำนวณจะมีสูงถึง $2\cdot 100\cdot 5^{100}\approx 10^{72}$ ครั้ง

เพราะฉะนั้นการแก้ปัญหาในข้อนี้จะต้องใช้วิธีอื่นแทนที่การคำนวณโดยตรง วิธีการหนึ่งที่สามารถนำมาใช้แก้ปัญหาข้อนี้มีชื่อเรียกว่ากระบวนการไปข้างหน้า(Forward Backward Procedure) ซึ่งเป็นการใช้ตัวแปรช่วย ซึ่งมีชื่อเรียกว่า ตัวแปรก้าวหน้า (Forward Variable) โดยการนำตัวแปรดังกล่าวมาใช้จะช่วยลดความซับซ้อนในการคำนวณลงไปอย่างมาก

สามารถให้คำนิยามแก่ตัวแปรก้าวหน้า ได้ว่า ตัวแปรก้าวหน้า $\alpha_t(i)$ คือ ความน่าจะเป็นของชุดข้อมูลจากการสังเกตบางส่วน $O_1,O_2,\ldots,O_t$ (ตั้งแต่เริ่มต้นจนถึงเวลา t และออกจาก สเตท i ที่เวลาดังกล่าว สามารถแสดงตัวแปรดังกล่าวในทางคณิตศาสตร์ได้ดังนี้คือ

$$\alpha_t(i) = \Pr(O_1,O_2,\ldots,O_t,q_t=i|\lambda) \tag{2.64}$$

สามารถหาค่าตัวแปรดังกล่าวได้จากการพิสูจน์ ดังต่อไปนี้

1. การเริ่มต้น (Initialization)

$$\alpha_1(i) = \pi_i b_i(O_1),\ 1\le i\le N \tag{2.65}$$

2. การเหนี่ยวนำ (Induction)

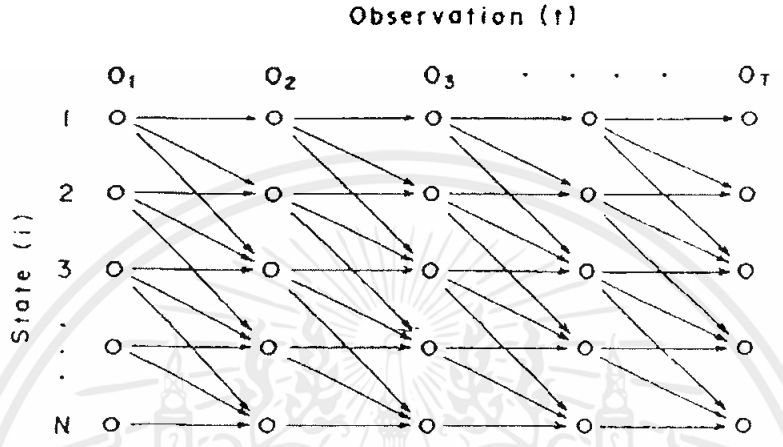
$$\alpha_{t+1}(i) = \left[\sum_{i=1}^{N} a_t(i)a_{ij}\right] b_j(O_{t+1}),\ 1\le t\le T-1,\ 1\le j\le N \tag{2.66}$$

3. การสิ้นสุด (Termination)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\Pr(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.67}$$

ถ้าพิจารณาความซับซ้อนของการคำนวณด้วยวิธีนี้พบว่าจะต้องมีการคำนวณเพียง N^2T ครั้ง แทนที่จะเป็น $2T \cdot N^T$ อย่างในการคำนวณโดยตรง



รูปที่ 2.8 กระบวนการไปด้านหน้า

ในทำนองเดียวกัน สามารถนิยามตัวแปรถอยหลัง (Backward Variable) ซึ่งก็คือ ความน่าจะเป็นของชุดข้อมูลจากการสังเกตบางส่วน $O_{t+1}, O_{t+2}, ..., O_T$ เมื่อกำหนดสเตทเท่ากับ i อธิบายตัวแปรดังกล่าวในทางคณิตศาสตร์ได้ดังนี้

$$\beta_t(i) = \Pr(O_{t+1}, O_{t+2}, ..., O_T \mid q_t = i, \lambda) \tag{2.68}$$

ซึ่งสามารถหาค่าตัวแปรดังกล่าวได้ดังนี้

1. การเริ่มต้น (Initialization)

$$\beta_t(i) = 1, \quad 1 \le i \le N \tag{2.69}$$

2. การเหนี่ยวนำ (Induction)

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \ t = T-1, T-2, ...1 \ \ 1 \le i \le N \tag{2.70}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

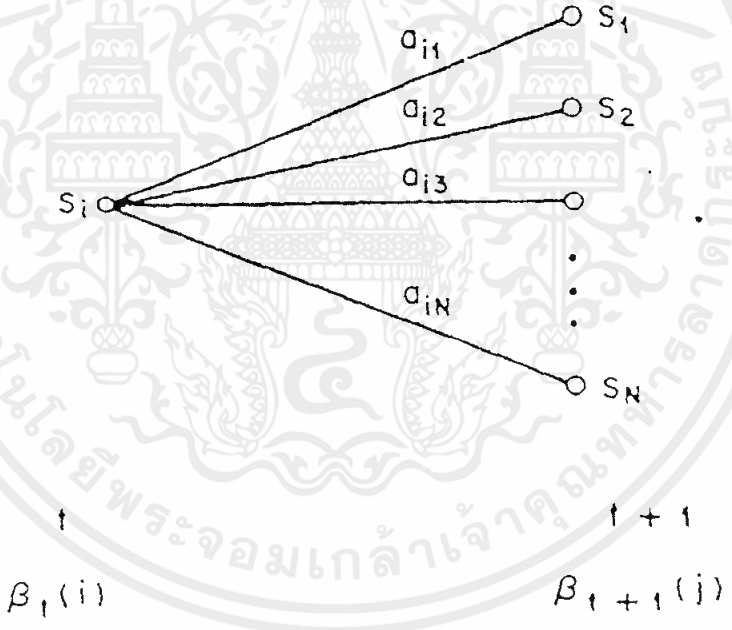
ในความเป็นจริงแล้วได้แก้ปัญหาข้อนี้ไปด้วยการใช้ตัวแปรก้าวหน้าเรียบร้อยแล้วอย่างไรก็ตาม สามารถแก้ปัญหาข้อนี้ได้อีกแบบโดยพิจารณาตัวแปรถอยหลังเข้ามาร่วมด้วยซึ่งจะถูกนำไปใช้ในขั้นตอนการฝึกหัดระบบ ซึ่งจะพบว่า

$$\alpha_t(i)\beta_t(i) = \Pr(O, q_t = i \mid \lambda), \quad 1 \le i \le N,\ 1 \le t \le T \tag{2.71}$$

ซึ่งจะคำนวณค่าความน่าจะเป็น $\Pr(O \mid \lambda)$ ได้จาก

$$\Pr(O \mid \lambda) = \sum_{i=1}^{N} \Pr(O, q_t = i \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) \tag{2.72}$$

ซึ่งสมการนี้จะถูกนำไปใช้อีกทีในการฝึกหัดระบบแบบ Gradient-Based



รูปที่ 2.9 กระบวนการถอยหลัง

2.12.4 การหาลำดับสเตทที่เหมาะสมที่สุด

ในกรณีนี้ต้องการที่จะหาลำดับของสเตทที่มีความเหมาะสมที่สุด เมื่อกำหนดชุดข้อมูลจากการสังเกต $O = O_1, O_2, ..., O_T$ และแบบจำลอง λ

การหาคำตอบในข้อนี้จะขึ้นอยู่กับมุมมองของความเหมาะสมของลำดับสเตทว่าถูกนิยามให้เป็นอย่างไร วิธีแรกอาจจะใช้วิธีการหาสเตทที่มีความเหมาะสมในช่วงเวลาต่างๆ กันมาต่อกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อให้ได้ลำดับทั้งหมด แต่ในบางครั้งวิธีการนี้ก็ไม่สามารถให้คำอธิบายที่ดีได้เท่าไรนัก ดังนั้นจะพิจารณาวิธีอื่นที่ไม่มีปัญหาดังกล่าว

วิทเทอร์บี อัลกอริทึม(Viterbi Algorithm) เป็นวิธีการในการหาลำดับของสเตท $Q = q_1, q_2, ..., q_T$ ที่ดีที่สุดเพียงทางเดินเดียว เมื่อกำหนดชุดข้อมูลจากการสังเกตเป็น $O = O_1, O_2, ..., O_T$ จำเป็นต้องกำหนดตัวแปรช่วยเพิ่มเติม

$$\delta_t(i) = \max_{q_1, q_2, ..., q_{t-1}} \Pr(q_1 q_2 ... q_t = i, O_1 O_2 ... O_t \mid \lambda) \tag{2.73}$$

กล่าวได้อีกนัยหนึ่งว่า $\delta_t(i)$ คือ คะแนนที่ดีที่สุด (ความน่าจะเป็นที่สูงที่สุด) ตามเส้นทางที่ค้นหามาที่เวลา t โดยพิจารณาข้อมูลจากการสังเกต t ข้อมูลแรกและจบลงที่สเตท i ซึ่งจากการพิสูจน์แล้วจะได้ว่า

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}) \tag{2.74}$$

ในการที่จะหาลำดับของสเตทที่ดีที่สุดจะต้องติดตามพิจารณาสเตทในทางเดินว่าสเตทใดบ้างที่ทำให้ค่าในสมการข้างต้นนี้มีค่ามากที่สุด สำหรับแต่ละค่า t และ j ซึ่งทำได้โดยกำหนดตัวแปรช่วยเพิ่มอีกตัวหนึ่งคือ $\psi_t(j)$ กระบวนการทั้งหมดสามารถแสดงได้ดังต่อไปนี้

1. การกำหนดค่าเริ่มต้น (Initialization)

$$\delta_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \tag{2.75}$$

$$\psi_1(i) = 0 \tag{2.76}$$

2. การทำซ้ำ (Recursion)

$$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \qquad 2 \le t \le T, \quad 1 \le j \le N \tag{2.77}$$

$$\psi_t(i) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}], \qquad 2 \le t \le T, \quad 1 \le j \le N \tag{2.78}$$

3. การสิ้นสุด (Termination)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P^* = \max_{1\le i\le N}[\delta_T(i)] \tag{2.79}$$

$$q_T^* = \arg\max_{1\le i\le N}[\delta_T(i)] \tag{2.80}$$

4. การหาเส้นทางย้อนกลับ (Path backtracking)

$$q_t^* = \psi_{t+1}[q_{t+1}^*], \quad t = T-1, T-2, ..., 1 \tag{2.81}$$

วิธีการที่ได้กล่าวมานี้จริงๆแล้วก็คือการค้นหาเส้นทางเดินที่ดีที่สุดบนกราฟ โดยที่ในแต่ละโหนดของกราฟจะเป็นสเตทของฮิดเดนมาร์คอฟในแต่ละช่วงเวลา $t, 1 \le t \le T$

2.12.5 การหาแบบจำลองที่ดีที่สุด

โดยทั่วไปแล้วปัญหาเกี่ยวกับการเรียนรู้ของฮิดเดนมาร์คอฟคือ จะทำอย่างไรจึงจะปรับค่าของตัวแปรของแบบจำลอง ให้สามารถอธิบายลักษณะของชุดข้อมูลฝึกหัดที่ได้ให้เข้าไปในแบบจำลองเพื่อฝึกหัดตามการประยุกต์ใช้งาน

โดยสำหรับผลลำดับค่าปรากฏหนึ่งๆ การเลือกค่าพารามิเตอร์ A, B, π ที่ดีที่สุดสามารถทำได้โดยใช้กระบวนการทำซ้ำ(Iterative) วิธีที่เลือกใช้คือวิธี บาม-เวลช์ (Baum-Welch) หรือ EM (Expectation-Maximization) โดยเมื่อนิยามให้

1. $\gamma_t(i) = P(q_t = i \mid O, \lambda)$

หมายถึงความน่าจะเป็นที่จะอยู่สเตท i ณ เวลา t โดยกำหนดลำดับดหตุการณ์ O และแบบจำลอง λ ให้สามารถแสดงค่า $\gamma_t(i)$ ได้ดังนี้

$$\gamma_t(i) = \frac{P(O, q_t = i \mid \lambda)}{P(O \mid \lambda)}$$

$$= \frac{P(O, q_t = i \mid \lambda)}{\sum_{i=1}^{N} P(O, q_t = i \mid \lambda)} \tag{2.82}$$

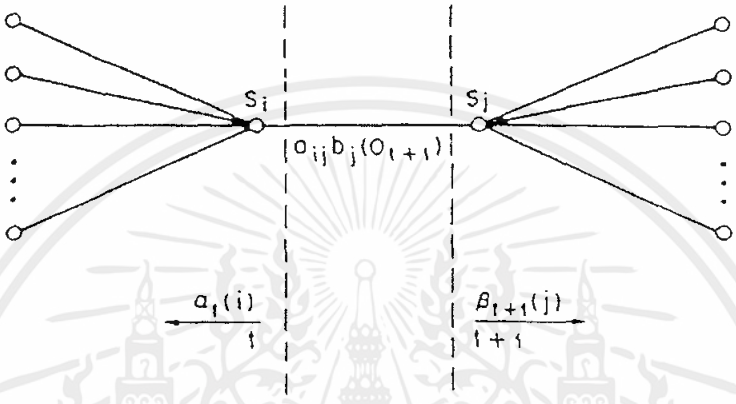
เนื่องจาก $P(O, q_t = i \mid \lambda)$ มีค่าเท่ากับ $\alpha_t(i)\beta_t(i)$ จึงทำให้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)} \tag{2.83}$$

2. $\varepsilon_t(i,j) = P(q_t = i, q_{t+1} j \mid O, \lambda)$

หมายถึง ความน่าจะเป็นที่จะอยู่ในสเตท i ที่เวลา t และ สเตท j ที่เวลา $t+1$ เมื่อกำหนดแบบจำลองและลำดับค่าปรากฏให้



รูปที่ 2.10 ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่เสตท i ที่เวลา t และอยู่ที่เสตท j ที่เวลา $t+1$

ซึ่งจากนิยามของตัวแปรไปข้างหน้า และตัวแปรย้อนกลับสามารถนำมาสัมผัสกับ $\varepsilon_t(i,j)$ ได้ดังนี้

$$\varepsilon_t(i,j) = \frac{P(q_t = i, q_{t+1} = j, O \mid \lambda)}{P(O \mid \lambda)} \tag{2.84}$$

และจะได้ความสัมพันธ์ของ $\gamma_t(i)$ กับ $\varepsilon_t(i,j)$ ดังนี้

$$\gamma_t(i) = \sum_{j=1}^{N}\varepsilon_t(i,j) \tag{2.85}$$

และ

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

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

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นสามารถหาค่าพารามิเตอร์ได้ดังนี้

$$\pi_i = \gamma_1(i) \qquad \text{เมื่อ} \qquad 1 \le i \le N$$

$$a_{ij} = \frac{\sum_{t=1}^{T} \varepsilon_{t-1}(i,j)}{\sum_{t=1}^{T} \gamma_{t-1}(i)} \tag{2.86}$$

$$b_j(k) = \frac{\sum_{t=1, O_t = v_k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)} \tag{2.87}$$

จากกระบวนการข้างต้น หากทำการคำนวณซ้ำๆ โดยให้ $\lambda' = (A', B', \pi')$ ซึ่งเป็นแบบจำลองเริ่มต้นแล้ว จะทำให้ความน่าจะเป็นของการเกิดลำดับค่าปรากฏ O ดีขึ้น จนกระทั่งถึงจุดวิกฤต เมื่อจุดวิกฤตของฟังก์ชั่นความน่าจะเป็นเกิดในกรณีที่ $\lambda' = \lambda$ หรือถ้า λ' มีความน่าจะเป็นมากกว่าแบบจำลอง λ ในลักษณะที่ $P(O|\lambda') > P(O|\lambda)$ นั่นก็คือ จะได้แบบจำลอง λ' ใหม่ ที่น่าจะทำให้เกิดลำดับค่าปรากฏ O ได้ดีกว่า

2.12.6 ระบบการจำลองมาร์คอฟ

เมื่อมีคำศัพท์อยู่ V คำในการทำการตระหนักรู้เสียง จะต้องสร้างแบบจำลองของคำแต่ละคำที่แตกต่างกัน คำแต่ละคำจะมีลำดับเทรนนิ่งที่ได้จากคุณลักษณะเฉพาะของคำศัพท์นั้นๆ จึงจะสามารถทำการตระหนักรู้เสียงพูด ได้มีขั้นตอนดังต่อไปนี้

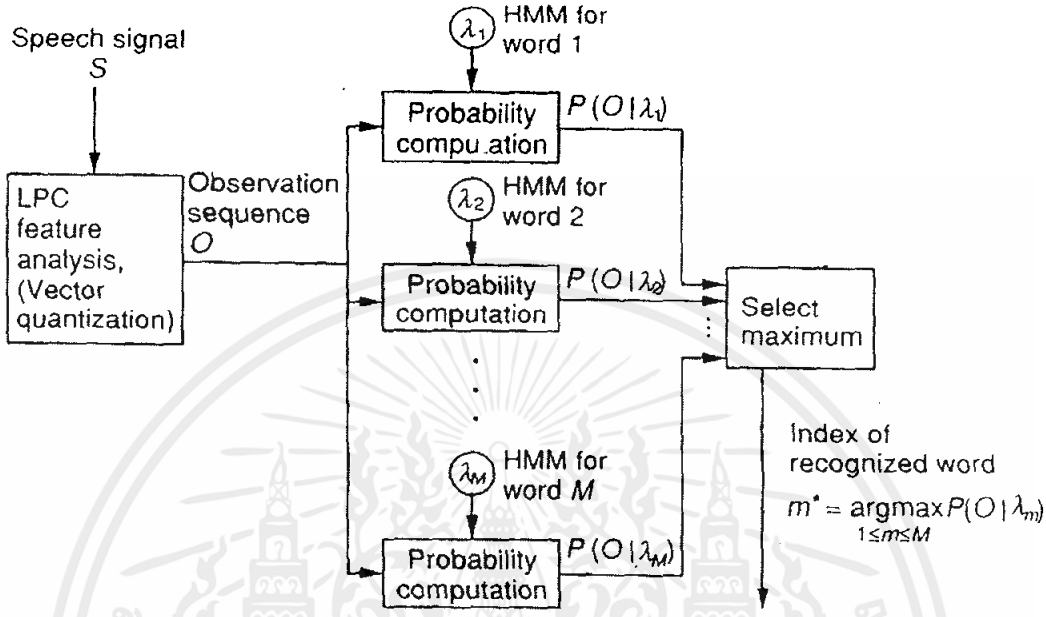
1. เมื่อมีคำศัพท์อยู่ V คำ ต้องสร้างแบบจำลองมาคอร์ฟ : λ_v ของแต่ละคำนั้นนั้นคือ การหาค่า A, B, λ ที่เหมาะสมกับลำดับเหตุการณ์เทรนนิ่งของเหตุการณ์นั้นๆ

2. ในการจะรู้คำศัพท์แต่ละคำ จะทำการหาค่า $P(O|\lambda)$ ของทุกๆแบบจำลอง แล้วเลือกแบบจำลองที่มีค่าความน่าจะเป็นในการเกิดค่าปรากฏสูงสุด คือ

$$V^* = \arg\max P(O|\lambda_v) \tag{2.88}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์ที่สอดคล้องกับกับแบบจะลองดังกล่าวนี้ จะเป็นคำเดียวกับคำศัพท์ที่ต้องการจะรู้จำนั้นเอง โดยขั้นตอนการคำนวณหาค่าความน่าจะเป็น จะใช้วิธีวิทเทอร์บี ดังที่ได้กล่าวแล้วในการหาลำดับสเตทที่เหมาะสมที่สุด ดังรูปที่ 2.11 เป็นขั้นตอนการตระหนักรู้โค้ดแบบมาร์คอฟ

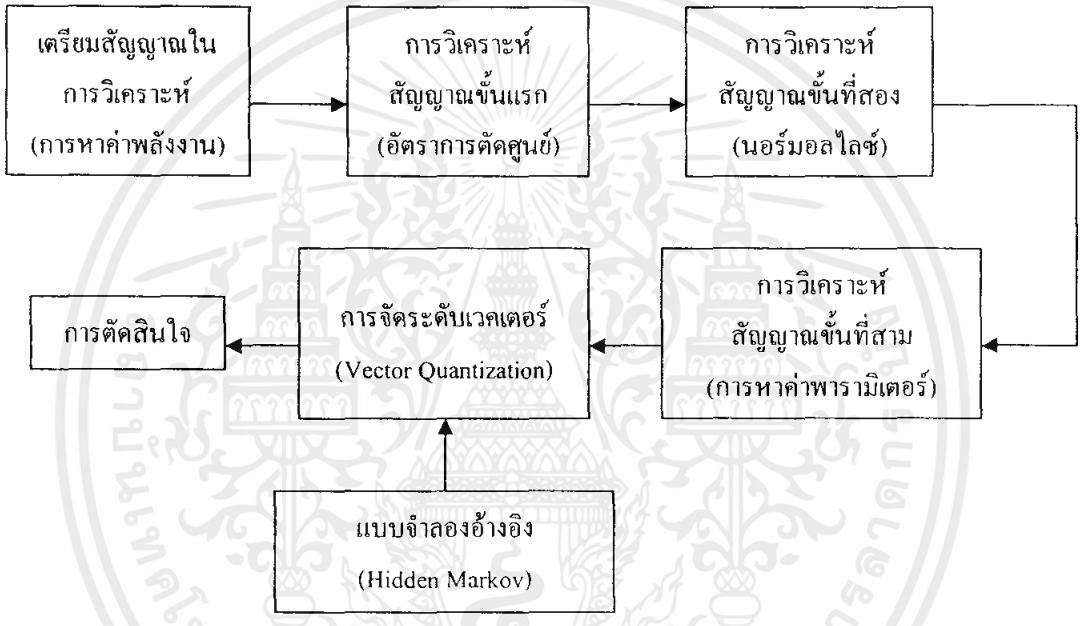


รูปที่ 2.11 การตระหนักรู้โค้ดแบบมาร์คอฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การคำนวณและการสร้าง

ระบบการตระหนักรู้เสียงพูด โดยทั่วไปสามารถแบ่งเป็นการดำเนินการย่อยต่างๆ ตามบล็อกไดอะแกรมดังรูปที่ 3.1 ซึ่งประกอบไปด้วย ภาคการวิเคราะห์สัญญาณ การหาค่าพารามิเตอร์ การจัดระดับเวคเตอร์ การสร้างแบบจำลอง การตัดสินใจ จนกระทั่งให้ผลลัพธ์ คือ การตระหนักรู้เสียงพูด



รูปที่3.1 บล็อกไดอะแกรมขั้นตอนการตระหนักรู้เสียงพูด

จากบล็อกไดอะแกรมขั้นตอนการตระหนักรู้เสียงพูด เราสามารถอธิบายเป็นส่วนๆได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 การเตรียมสัญญาณในการวิเคราะห์

เป็นขั้นตอนการเตรียมสัญญาณ เพื่อนำไปวิเคราะห์สัญญาณในขั้นต่อไปโดยในขั้นตอนนี้เป็นการหาขอบเขตของคำซึ่งเป็นส่วนที่แท้จริงที่จะนำไปใช้ต่อในส่วนอื่นๆ

การวิเคราะห์หาค่าพลังงาน

ใช้หลักการวิเคราะห์สัญญาณในโดเมนเวลาเข้ามาช่วย สำหรับขั้นตอนมีดังต่อไปนี้

1. คำนวณค่าพลังงานของสัญญาณในช่วงเวลาที่ถูกบันทึกทั้งหมด โดยการคำนวณหาค่าพลังงานของสัญญาณซึ่งจะทำการคำนวณเป็นเฟรม กำหนดให้เฟรมหนึ่งมีจำนวนแซมเปิล 100 แซมเปิล และทำการคำนวณเป็นเฟรมต่อไปเรื่อยๆจนครบทุกเฟรม

2. พิจารณาหาค่าพลังงานสูงสุดของสัญญาณ (Emax) และพิจารณาว่า ค่าพลังงานสูงสุดที่หาได้นั้นอยู่ที่เฟรมใด แล้วสมมติให้เป็นเฟรมที่ m

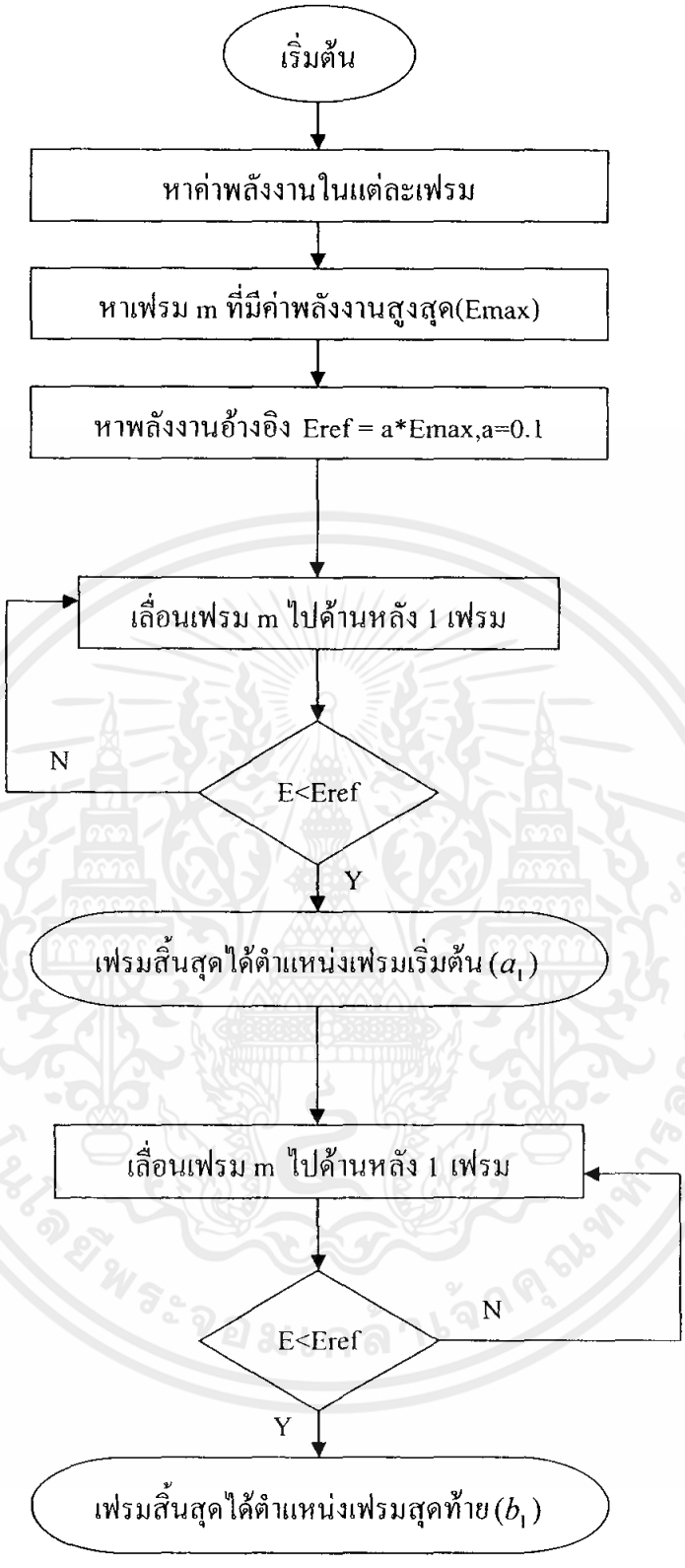
3. ทำการกำหนดพลังงานอ้างอิง (Eref) โดยพลังงานอ้างอิงสามารถคำนวณได้จาก a เท่าของพลังงานสูงสุด (a *Emax) ซึ่งในที่นี้เลือกใช้ค่า a = 0.1 จากการที่หาค่าพลังงานสูงสุดจะทำให้ทราบว่า เฟรมเริ่มต้นและเฟรมสุดท้ายของคำจะอยู่นอกช่วงที่เราหา a *Emax ได้

4. ทำการหาเฟรมเริ่มต้นของคำ โดยจะเริ่มพิจารณาที่ค่าพลังงานสูงสุดเป็นเฟรมแรก (เฟรมที่ m) ซึ่งนำค่าพลังงานของสัญญาณ เฟรมที่ m ไปเปรียบเทียบกับพลังงานอ้างอิง แล้วเลื่อนเฟรมที่จะเปรียบเทียบกับพลังงานอ้างอิงไปด้านหลังเรื่อยๆ (เฟรมที่ m-1 , m-2 ,..., 2 , 1) จนกว่าจะพบเฟรมที่มีค่าพลังงานน้อยกว่าค่าพลังงานอ้างอิงเป็นเฟรมแรก ก็จะได้เฟรมเริ่มต้นของสัญญาณ (a_1)

5. การหาเฟรมสุดท้ายของคำ จะทำการพิจารณาเช่นเดียวกับการหาเฟรมเริ่มต้นของคำ คือเริ่มเปรียบเทียบจากเฟรมที่มีพลังงานสูงสุด (เฟรมที่ m) เช่นกัน แต่การเลื่อนเฟรมจะเลื่อนไปทางด้านหน้าแทน (เฟรมที่ m+1, m+2, ...) จนกว่าจะพบเฟรมที่มีค่าพลังงานน้อยกว่าค่าพลังงานอ้างอิง ก็จะได้จุดสุดท้ายของสัญญาณ (b_1)

สามารถหาเฟรมเริ่มต้นและเฟรมสุดท้าย ได้ดังรูปที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 การหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียงโดยการเปรียบเทียบพลังงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การวิเคราะห์สัญญาณขั้นหนึ่ง

เป็นขั้นตอนการวิเคราะห์สัญญาณ เพื่อให้เป็นการแน่ใจว่า การหาขอบเขตของคำในขั้นตอนการหาค่าพลังงานนั้นถูกต้อง ทำให้การตระหนักรู้เสียงพูดมีความแม่นยำมากขึ้น

การวิเคราะห์หาค่าอัตราการตัดศูนย์

เนื่องจากการใช้การวิเคราะห์โดยการหาค่าพลังงานของเสียงอาจเกิดการผิดพลาด จึงนำเอาการวิเคราะห์หาค่าอัตราการตัดศูนย์มาร่วมพิจารณา โดยมีวิธีการดังนี้

1. คำนวณหาค่าอัตราการตัดศูนย์ในช่วงเวลาที่ถูกบันทึกทั้งหมด โดยการคำนวณหาค่าอัตราการตัดศูนย์จะคำนวณเป็นเฟรม กำหนดให้แต่ละเฟรมมีตัวอย่าง 100 ตัวอย่าง และทำการคำนวณจนครบช่วงเวลาของเสียงที่ถูกบันทึกทั้งหมด

2. กำหนดระดับสูงสุด (Zmax) ให้เป็นเฟรมที่ p

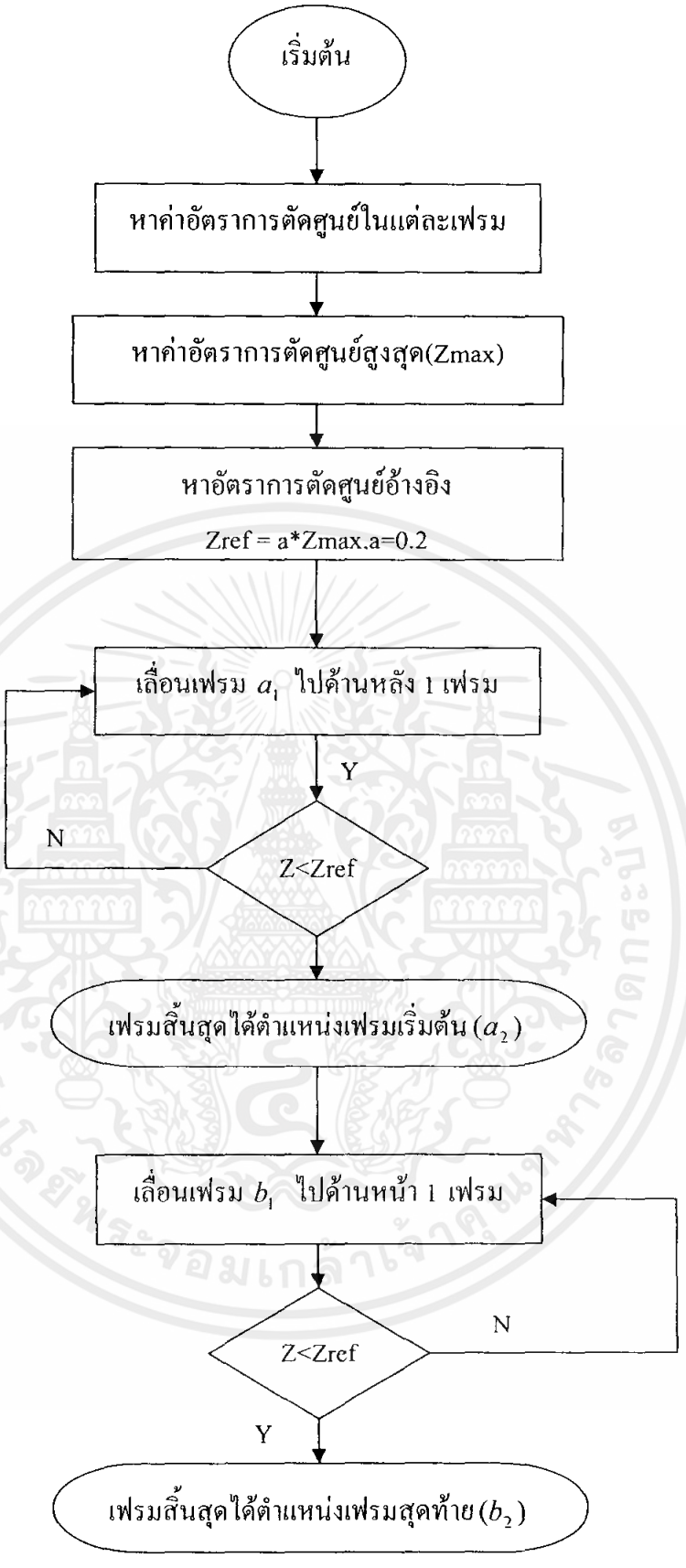
3. กำหนดระดับอ้างอิง (Zref) มีค่าเป็น a เท่าของอัตราการตัดศูนย์สูงสุด (a*Zref) โดยค่า a จะเท่ากับ 0.1 เช่นกัน

4. จากจุดเริ่มต้น (a_1) ที่หาได้จากขั้นตอนที่ 1 จะเป็นเฟรมแรกที่พิจารณาโดยจะทำการเปรียบเทียบอัตราการตัดศูนย์ของเฟรมดังกล่าวกับอัตราการตัดศูนย์อ้างอิง ถ้ามีค่ามากกว่า เฟรมเริ่มต้นจะถูกเปลี่ยนตำแหน่ง โดยจะเลื่อนไปยังเฟรมแรกที่มีค่าน้อยกว่าอัตราการตัดศูนย์อ้างอิง (a_2) และจะถือว่าจุดนี้เป็นเฟรมเริ่มต้นที่แท้จริง ซึ่งการเลื่อนเฟรมจะเลื่อนไปด้านหลัง (เฟรมที่ a_{1-1}, a_{1-2},.....) แต่ในกรณีที่ a_1 น้อยกว่าอัตราการตัดศูนย์อ้างอิง จะถือว่า a_1 ที่หาได้จากขั้นตอนที่ 1 เป็นเฟรมเริ่มต้นที่ถูกต้องอยู่แล้ว

5. การหาเฟรมสุดท้าย จากเฟรมสุดท้าย (b_2) ที่ได้จากขั้นตอนที่ 1 จะพิจารณาเหมือนข้อ 3 คือจะเป็นเฟรมแรกที่จะเปรียบเทียบอัตราการตัดศูนย์อ้างอิง ถ้ามีค่ามากกว่า เฟรมสุดท้ายจะถูกเปลี่ยนตำแหน่ง โดยจะเลื่อนเฟรมจะเลื่อนไปด้านหน้า (เฟรมที่ b_{1+1}, b_{1+2},.....) จนกว่าจะเจอเฟรมแรกที่มีค่าน้อยกว่าอัตราการตัดศูนย์อ้างอิง (b_2) และจะถือว่าเฟรมนี้เป็นเฟรมสุดท้ายที่แท้จริง แต่ถ้าอัตราการตัดศูนย์ที่ b_1 น้อยกว่าอัตราการตัดศูนย์อ้างอิงก็กล่าวได้ว่า b_1 เป็นเฟรมสุดท้ายที่ถูกต้องอยู่แล้วเช่นกัน

จากขั้นตอนทั้งหมดที่กล่าวมานี้สามารถดำเนินการได้ดังรูปที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 การหาเฟรมเริ่มต้นและเฟรมสุดท้ายของสัญญาณเสียงโดยการหาอัตราการตัดศูนย์

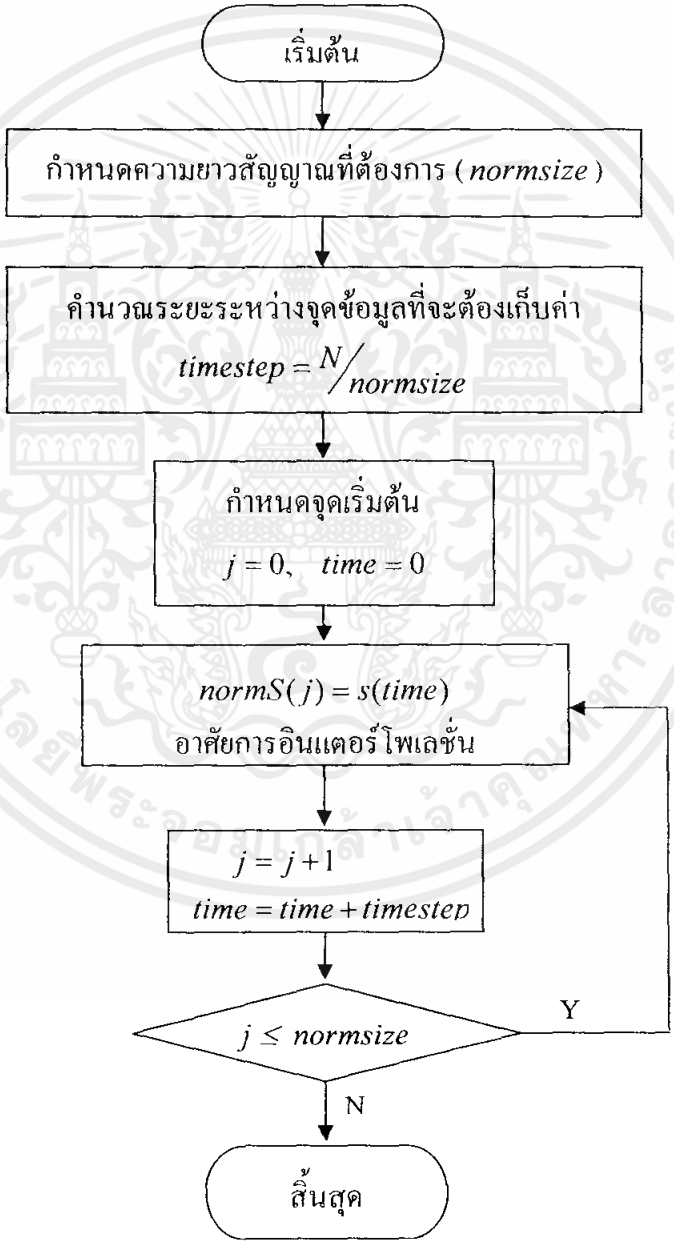
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การวิเคราะห์สัญญาณขั้นที่สอง

เนื่องจากสัญญาณเสียงพูดแต่ละคำมีความยาวไม่เท่ากันจึงต้องมีการนอร์มอลไลซ์สัญญาณเสียงพูดให้ยาวเท่ากัน

การนอร์มอลไลซ์

การนอร์มอลไลซ์ทำขึ้นเพื่อให้สัญญาณที่สนใจมีความยาวเท่าๆ กันโดยหลักการเปลี่ยนคาบการสุ่ม และใช้หลักการอินแตอร์โพเลชั่น(Interpolation) ในการประมาณค่าหากจุดที่ทำการสุ่มใหม่นั้นไม่ตรงกับจุดซึ่งมีข้อมูลอยู่ ขั้นตอนการนอร์มอลไลซ์สัญญาณเสียงพูดแสดงดังรูปที่ 3.4 เมื่อ สัญญาณอินพุทคือ $s(n)$ โดยมีความยาว N และเอาท์พุทคือสัญญาณ $normS(n)$



รูปที่ 3.4 ขั้นตอนการนอร์มอลไลซ์สัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การวิเคราะห์สัญญาณขั้นที่สาม

เป็นการหาค่าพารามิเตอร์แบ่งออกเป็น

การหาค่าพรีเอมฟาซิส

เป็นการนำสัญญาณที่ได้จากขั้นตอนที่สองมาผ่านระบบซึ่งมีฟังก์ชั่นถ่ายโอน

$$H(z) = 1 - az^{-1} \text{ เมื่อ } 0.9 \le a \le 1.0$$

โดยในโครงงานนี้เลือกใช้ $a = 15/16 = 0.937$

การแบ่งช่วงสัญญาณ

จากสัญญาณที่ผ่านการพรีเอมฟาซิสจะแบ่งพิจารณาเป็นเฟรมโดยแต่ละเฟรมมีขนาด N จุดข้อมูลและมีข้อมูลเหลื่อมกับเฟรมก่อนหน้า M จุดข้อมูล โดยต้องให้ค่า M มีค่าน้อยกว่า N การคำนวณค่า ณ เฟรมที่ l สามารถทำได้โดยอาศัยสมการ

$$x_l(n) = s'(Ml + n) \text{ เมื่อ } n = 0,1,\ldots,N-1$$

ค่าที่ใช้ในการทดลองคือ $M = 80$ และ $N = 240$ และในการทดลองนี้ทำการกำหนดให้มีการสุ่มตัวอย่างของสัญญาณด้วยความถี่ 22,050 เฮิรตซ์ ดังนั้นช่วงเวลาในการวิเคราะห์ของแต่ละเฟรม คือ $240/22{,}050 = 10ms$ และเวลาที่เหลื่อมกันในแต่ละเฟรมคือ $80/22{,}050 = 3.6ms$

และเพื่อลดความไม่ต่อเนื่องของเฟรม จะนำสัญญาณที่แบ่งเป็นช่วงนั้นผ่านเข้าช่องแคบแฮมมิ่งซึ่งมีสมการเป็น

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

การวิเคราะห์คุณลักษณะของเสียง

ในการวิเคราะห์ใช้การหาค่าสัมประสิทธิ์แอลพีซีเพราะเป็นวิธีที่ทำการคำนวณง่ายและใช้ข้อมูลน้อยกว่าวิธีอื่นๆ แต่ให้ผลการทดลองในระดับที่ยอมรับได้ (ดูขั้นตอนการดำเนินงานได้ในบทที่ 2) และค่าอันดับที่มีความเหมาะสมกับข้อมูลคือ p ที่อันดับ 12

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 การเวทค่าพารามิเตอร์

ทำการเวท หรือถ่วงน้ำหนักสัมประสิทธิ์เซปสตรัมที่ได้ ช่วงลำดับต้นๆและท้ายๆของเฟรมที่นำมาวิเคราะห์จะเกิดความคลาดเคลื่อนมากกว่าบริเวณส่วนอื่น เพราะฉะนั้น เราจะทำการลดความคลาดเคลื่อนเหล่านั้นด้วยฟังก์ชันเวทติ้ง (ดูได้จากบทที่ 2) จะได้พารามิเตอร์สุดท้ายคือ

$$C_m^{'} = C_m * W_m$$

3.6 การจัดระดับเวคเตอร์

การจัดระดับเวคเตอร์เป็นเทคนิคการเข้ารหัสเทคนิคหนึ่งที่มีคุณภาพ มันจะทำการเข้ารหัสเวคเตอร์ พารามิเตอร์ของคลื่น ที่แทนสเปกตรัมของเสียง ซึ่งเป็นเวคเตอร์อินพุทที่ตัวเลขเป็นจำนวนเต็ม โดยนำไปเปรียบเทียบกับโค้ดบุคที่ได้สร้างเก็บไว้ก่อนแล้ว โค้ดบุคที่ถูกเลือกนั้นจะเป็นตัวใกล้เคียงกับเวคเตอร์อินพุทผลที่ได้จากเวคเตอร์ควอนไตเซชั่น ที่แสดงออกมาขึ้นอยู่กับเซตของโค้ดบุคที่สร้างเก็บไว้ก่อนแล้วนั่นเอง ขั้นตอนการจัดระดับเวคเตอร์ มี 2 ขั้นตอน คือ การสร้างโค้ดบุค และการเปรียบเทียบ

3.6.1 การสร้างโค้ดบุค

สร้างโค้ดบุคขึ้นมาเป็นตัวแทนสเปคตรัมของเสียงที่สังเกตได้ เพื่อลดจำนวนข้อมูลของพารามิเตอร์ที่ได้มาจากขั้นตอนการวิเคราะห์สัญญาณขั้นที่ 3 การสร้างโค้ดบุคมีขั้นตอนดังนี้

นำเทรนนิ่งเซตมาใช้ในการสร้างโค้ดบุค

ขนาดโค้ดบุคของการควอนไตซ์ แบบเวคเตอร์ คือ $M = 2^B$ เวคเตอร์ และ เพื่อที่จะหาเซตของ M โค้ดบุคที่ดีที่สุด จำนวนเวคเตอร์อินพุทจะต้องมากกว่าโค้ดบุคมากๆ

การสุ่มค่าเริ่มต้น

การสุ่มค่าเริ่มต้น เป็นวิธีหนึ่งในการออกแบบโค้ดบุค ซึ่งคือการเลือกค่าเริ่มต้นของโค้ดบุค เรียกโค้ดบุคที่ได้จากการสุ่มค่าเริ่มต้นนี้ว่า แรนดอมโค้ดบุค (Random Codebook) ถึงแม้วิธีนี้จะไม่ใช่วิธีที่ดีนักแต่โค้ดบุคที่ได้จากการสุ่มก็ให้ผลเป็นที่ยอมรับ

การหาความคลาดเคลื่อน

การหาความคลาดเคลื่อนเป็นส่วนที่จำเป็นและเป็นประโยชน์ต่อการออกแบบโค้ดบุค สมการทางพีชคณิตที่ใช้ในการหาระยะทางมีหลายวิธี แต่ที่นำมาใช้คือ การหาความคลาดเคลื่อนกำลังสองรวม (Total Square Error) ซึ่งเป็นวิธีที่ง่ายและเร็ว ถ้าสัญญาณมี P มิติ สามารถหาระยะห่างระหว่างสัญญาณอินพุท (x) กับเวคเตอร์โค้ด (y) โดยสมการ

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

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจัดกลุ่มและการหาจุดศูนย์กลางของกลุ่ม

การจัดกลุ่มเป็นการจัดเวคเตอร์อินพุทเข้าไปตามกลุ่มต่างๆ ของแรนดอมโค้ดบุค โดยพิจารณาระยะทางหรือความคลาดเคลื่อนน้อยที่สุด ของเวคเตอร์อินพุท x กับเวคเตอร์โค้ดบุค y ซึ่งเป็นโค้ดบุค จากนั้นจะทำการหาค่าเฉลี่ยของแต่ละกลุ่ม เพื่อเป็นค่ากลางของกลุ่มนั้นๆจะได้

$$\bar{Y} = \frac{1}{L}\sum_{i=1}^{L} x_i$$

$\bar{Y}$ เป็นจุดศูนย์กลางซึ่งเป็นเวคเตอร์ที่อยู่ตรงกลางของ $\{x_i\}_{i=1}^{L}$ ซึ่ง แต่ละมิติจะไม่ขึ้นแก่กัน หมายความว่าแต่ละ y_k เป็นค่ากลางของ $\{x_i\}_{i=1}^{L}$ ทำ 2 ขั้นตอนนี้ซ้ำ จะเกิดการลู่เข้า โดยความคลาดเคลื่อนรวมจะต่ำกว่าค่าหนึ่งๆ ซึ่งค่าความคลาดเคลื่อนรวมจะลดลงทุกครั้งที่มีการคำนวณซ้ำใหม่ จึงขึ้นกับค่าที่กำหนดว่าจะต้องการให้ความคลาดเคลื่อนรวมน้อยกว่าค่ากลางดังกล่าวของแต่ละกลุ่ม จะถูกเก็บเป็นเวคเตอร์โค้ดจะได้ว่า y เป็นควอนไตซ์ของค่า x

โดย $q(.)$ เป็นดอเปอร์เรเตอร์ของการควอนไตซ์ y ถูกเรียกว่าเอาท์พุทเวกเตอร์ของค่า x โดย y เป็นค่าใดค่าหนึ่งใน $y_{il} = [y_{i1} y_{i2} \ldots y_{ip}]$

$$Y = \{y_i; 1 \le i \le M\} \quad \text{โดย} \quad y_{il} = [y_{i1} y_{i2} \ldots y_{ip}]$$

Y เป็นเซตของโค้ดบุค M เป็นขนาดของโค้ดบุคและ $\{y_i\}$ เป็นเซตของเวคเตอร์โค้ด y_i อาจเรียกว่าโค้ดอ้างอิงและ M อาจเรียกว่าจำนวนระดับขั้น จะทำการแบ่งเวคเตอร์ x ไปใน M เซล $\{C_i; 1 \le i \le M\}$ เมื่อ x อยู่ในเซล C_i

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

3.6.2 การเปรียบเทียบ

เป็นขั้นตอนการเปรียบเทียบพารามิเตอร์กับโค้ดบุคอ้างอิงที่จัดทำไว้แล้ว ผลลัพธ์ที่ได้จะเป็นเวคเตอร์โค้ดที่มีจำนวนข้อมูลลดลงจากขั้นตอนการวิเคราะห์ขั้นที่ 3 โดยเป็นการวัดระยะห่างระหว่างเวคเตอร์โค้ดกับพารามิเตอร์ที่ต้องการทดสอบ ด้วยวิธีการหาความคลาดเคลื่อนกำลัง 2 รวม (Total Square Error)

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7 แบบจำลองฮิดเดนมาร์คอฟ

ในการสร้างแบบจำลอง และการรู้จำเสียงเพื่อเปรียบเทียบความคล้ายคลึงของสัญญาณเสียงแบบ HMM ของเสียงมีขั้นตอนดังนี้

1. สุ่มค่าเริ่มต้น a,b และกำหนดให้ π = [1 0 0 0 0 0] ตามเงื่อนไขแบบจำลองแบบ Left-Right

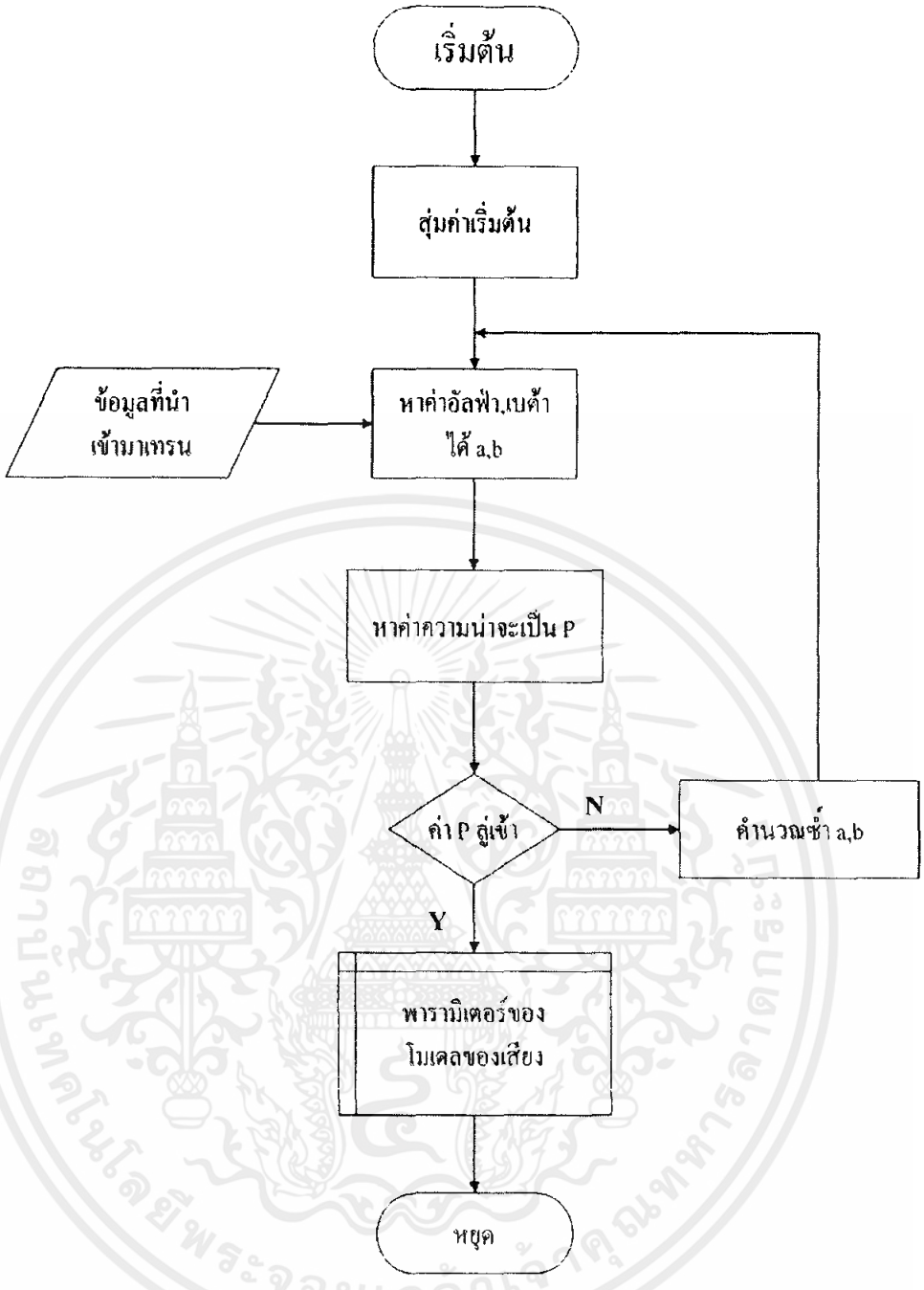
2. หาค่า α,β จากค่า a,b เริ่มต้น และลำดับค่าปรากฏ $O=O_1,O_2,...,O_T$ ซึ่งเรียกว่าลำดับเทรนนิ่งตามวิธีของกระบวนการไปข้างหน้าและย้อนกลับ โดยใช้ลำดับของค่าปรากฏหลายๆ ลำดับเข้ามาเทรนเพื่อความถูกต้อง

3. หาค่าพารามิเตอร์ a,b,π ที่ให้ความน่าจะเป็นสูงสุดที่จะเป็นแบบจำลอง λ ที่เหมาะสม

4. ตรวจสอบค่าพารามิเตอร์แบบจำลองที่ได้ว่าลู่เข้าหรือยัง โดยใช้วิธีการคำนวณค่า a , b ซ้ำ โดยในที่นี้ตั้งไว้ว่าเมื่อไม่มีการเปลี่ยนแปลงหรือค่าเท่าเดิมให้หยุดการทำซ้ำ ก็จะได้ค่าพารามิเตอร์ a,b,π ของแบบจำลองที่ต้องการ

5. เก็บค่าพารามิเตอร์ a,b,π ที่ได้จากข้อ 4. เป็นพารามิเตอร์ของแบบจำลองไว้ และหลังจากที่ได้แบบจำลองฮิดเดนมาร์คอฟของแต่ละคำศัพท์แล้วเมื่อมีลำดับค่าปรากฏ $O=O_1,O_2,...,O_T$ ของเสียงที่ไม่ทราบ ซึ่งเป็นเสียงที่ต้องการทดสอบเข้ามา เราจะทำการคำนวณหาความน่าจะเป็น $P(O|\lambda)$ ทุกแบบจำลองของแต่ละคำศัพท์โดยใช้วิธี วีเทอร์บีอัลกอริทึม แล้วเลือกเอาคำศัพท์ที่มีความน่าจะเป็นสูงสุด ซึ่งก็คือคำศัพท์ที่แบบจำลองได้นั่นเอง ขั้นตอนเป็นไปดังรูปที่ 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 ขั้นตอนการสร้างแบบจำลองอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.8 การตัดสินใจ

เป็นขั้นตอนการตัดสินใจว่าเสียงที่นำมาทดสอบกับระบบตระหนักรู้จำเสียงนั้นเป็นเสียงอะไร โดยเปรียบเทียบกับแบบจำลองอ้างอิง จากนั้นก็จะทำการเลือกค่าความน่าจะเป็นสูงสุด เพื่อตัดสินใจว่าเป็นเสียงใดในแบบจำลองอ้างอิง ซึ่งจะมีกระบวนการดังต่อไปนี้

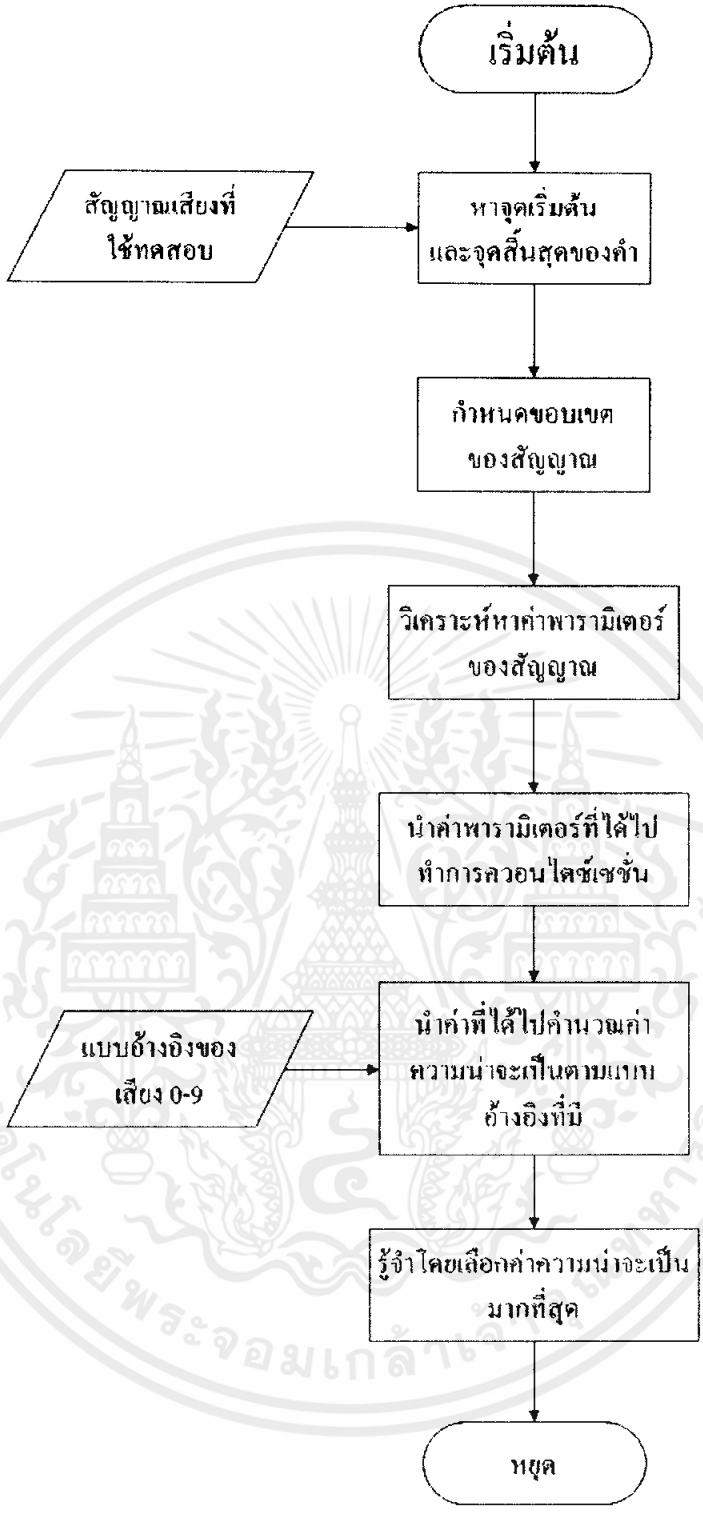
1. การหาดัชนีโค้ดบุค

นำค่าพารามิเตอร์ของเสียง นำมาเปรียบเทียบกับโค้ดบุค โดยวิธีความคลาดเคลื่อนกำลังสอง จนได้ดัชนีโค้ดบุคของเฟรมเสียงที่ต้องการทดสอบ และเก็บดัชนีโค้ดบุคของแต่ละเฟรมไว้เป็นลำดับค่าปรากฏ เพื่อนำไปคำนวณในขั้นตอนต่อไป

2. การตระหนักรู้

นำค่าปรากฏ $O=\{O_1O_2O_3...O_T\}$ จากขั้นตอนการหาดัชนีโค้ดบุค มาเข้ากระบวนการวิทเทอร์บี อัลกอริทึมในการคำนวนเปรียบเทียบบแบบจำลองอ้างอิงของแต่ละเสียง แล้วเลือกเอาค่าความน่าจะเป็นที่มีค่ามากที่สุด เพื่อบอกว่าเสียงนั้นเป็นเสียงใด

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 ขั้นตอนการตระหนักรู้เสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

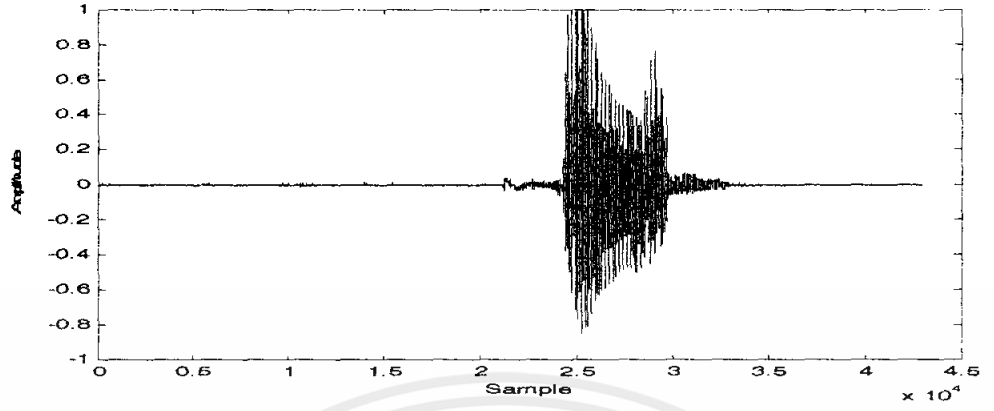
การทดลองในปริญญานิพนธ์ เป็นการทดลองระบบการตระหนักรู้เสียงพูด ซึ่งทำการทดลองโดยจะทำการเตรียมสัญญาณ หลังจากนั้นนำสัญญาณที่เตรียมไว้มาทำการสร้างแบบจำลอง ในที่นี้ใช้บุคคลมาทำแบบจำลองอ้างอิงชายจำนวน 6 คน คนละ 5 เสียง และหญิงจำนวน 6 คน คนละ 5 เสียงเช่นเดียวกัน หลังจากนั้นจึงจะทำการตระหนักรู้เสียงพูด ซึ่งจะเปรียบเทียบประสิทธิภาพระหว่างการตระหนักรู้เสียงพูดโดยเสียงของบุคคลในแบบอ้างอิง และบุคคลอื่น ๆ

4.1 การเตรียมสัญญาณ

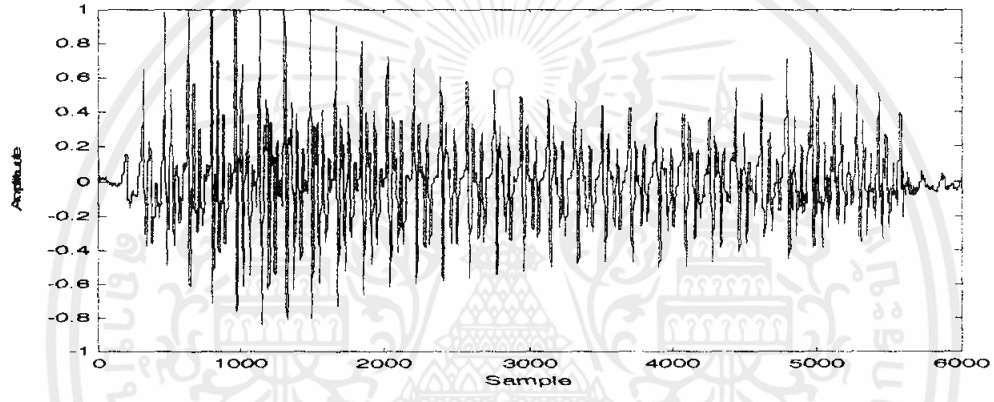
1. ทำการหาค่าพลังงานของสัญญาณเสียง 0-9
2. ทำการหาอัตราการตัดศูนย์ของสัญญาณเสียง 0-9
3. ทำการนอร์มอลไลซ์สัญญาณ ที่ได้จากการหาอัตราการตัดศูนย์
4. หาค่าพารามิเตอร์ของสัญญาณเสียง 0-9 โดยหาจากสัมประสิทธิ์แอลพีซี
5. ทำการถ่วงน้ำหนักค่าพารามิเตอร์

ผลจากการเตรียมสัญญาณจะแสดงจะดังรูป 4.1-4.20

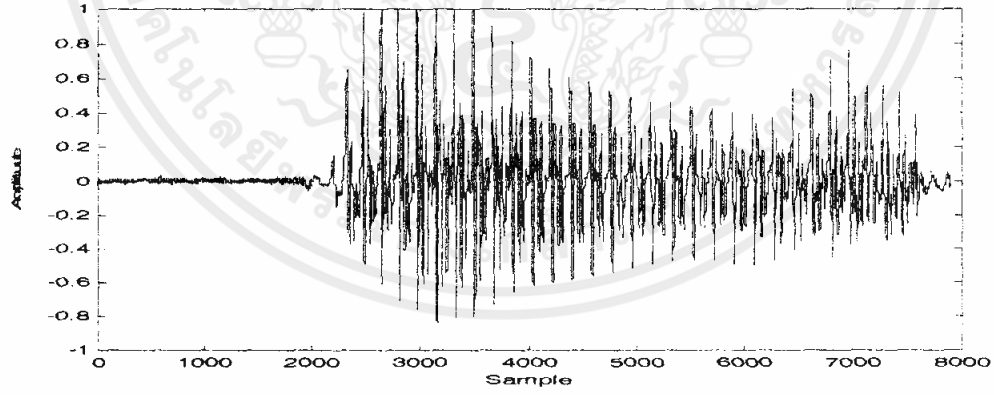
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

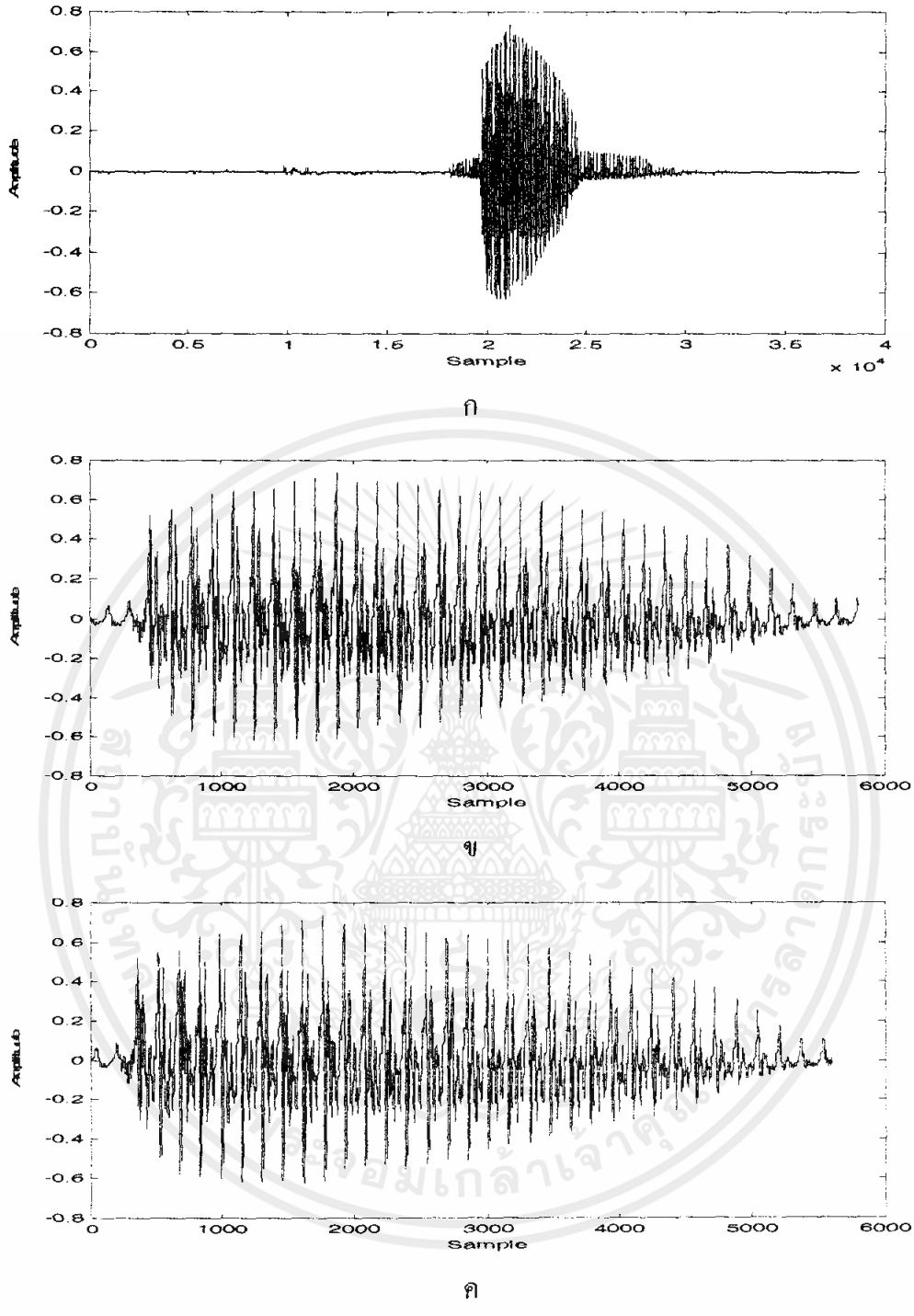
รูปที่ 4.1 การเตรียมสัญญาณเสียงศูนย์

ก เสียงตั้งต้น

ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



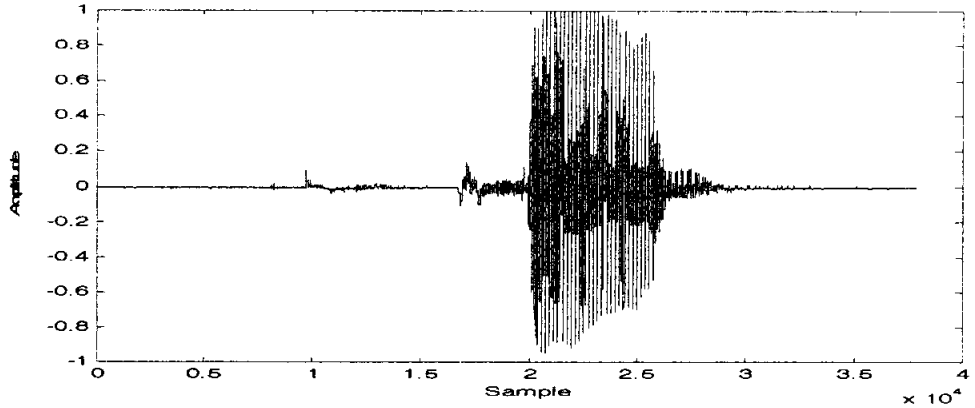
รูปที่ 4.2 การเตรียมสัญญาณเสียงหนึ่ง

ก เสียงตั้งต้น

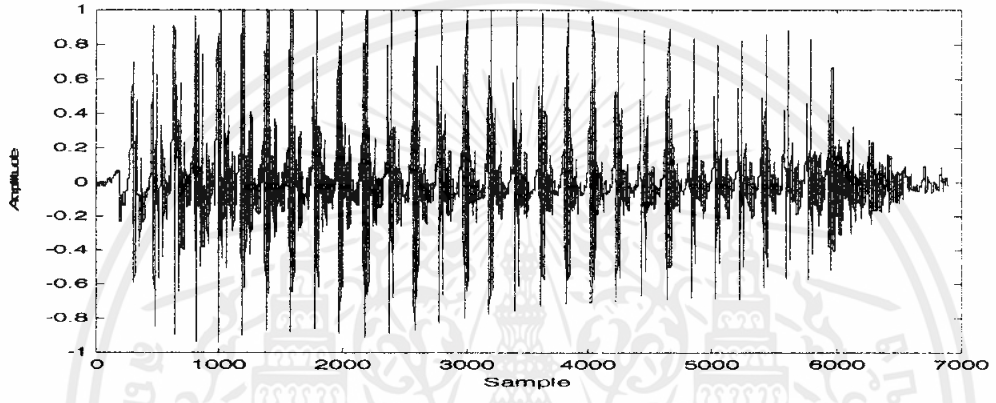
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

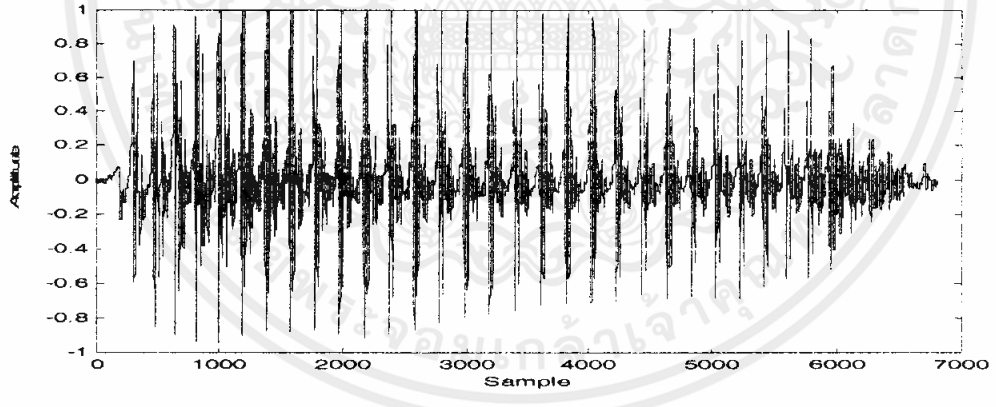
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

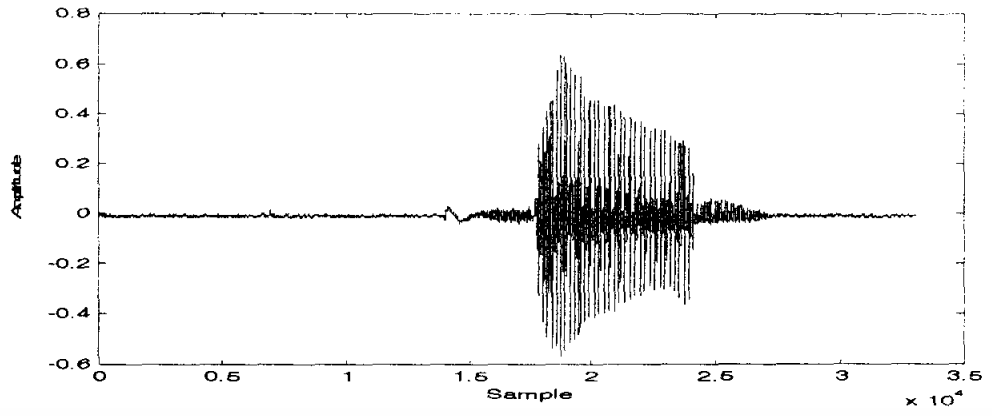
รูปที่ 4.3 การเตรียมสัญญาณเสียงสอง

ก เสียงตั้งต้น

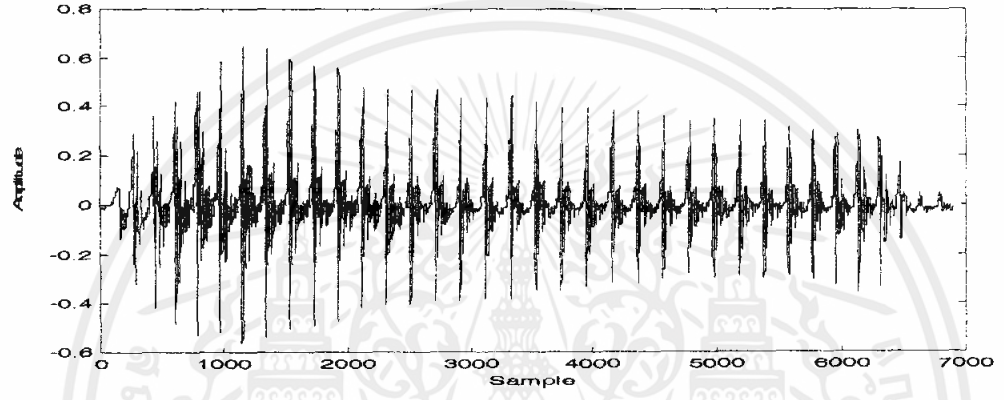
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

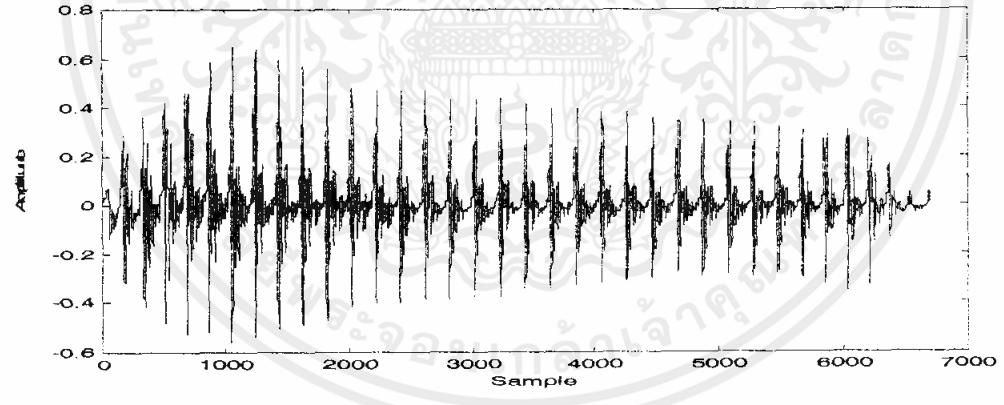
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

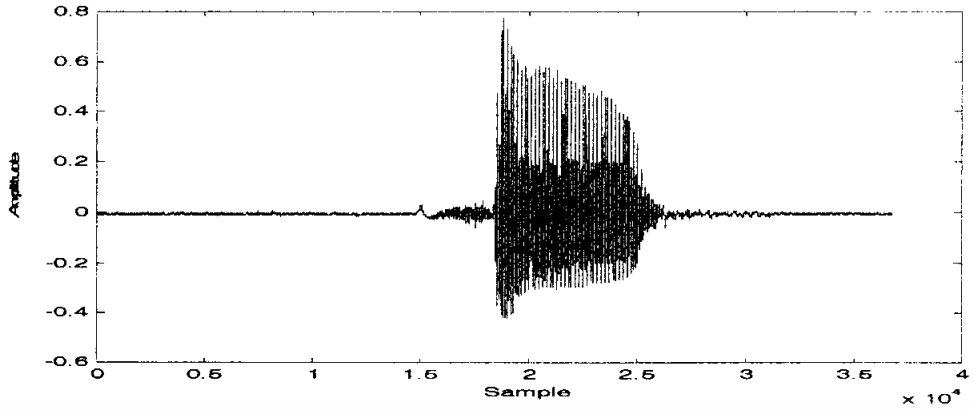
รูปที่ 4.4 การเตรียมสัญญาณเสียงสาม

ก เสียงตั้งต้น

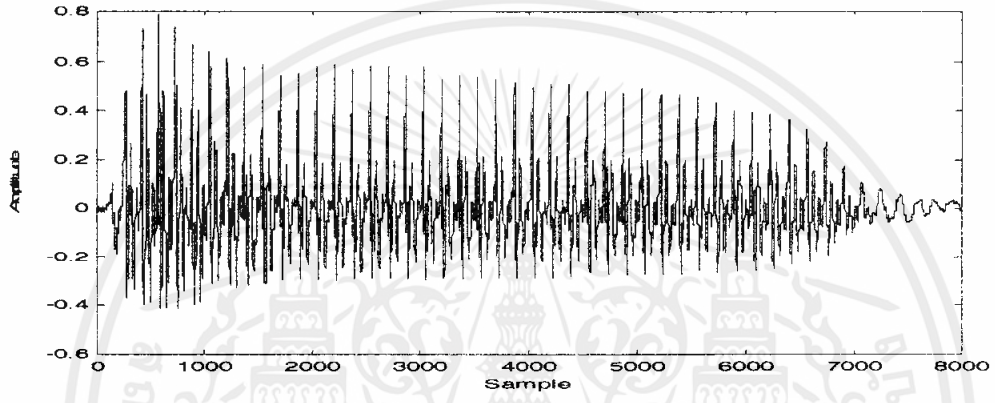
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

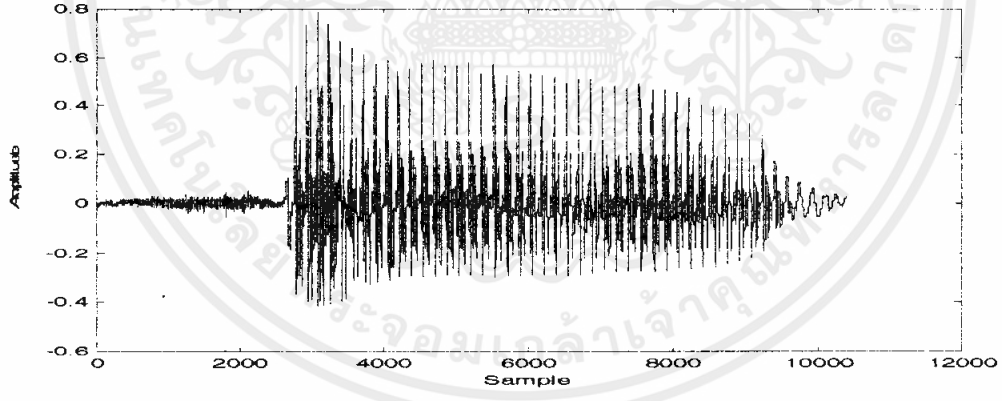
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

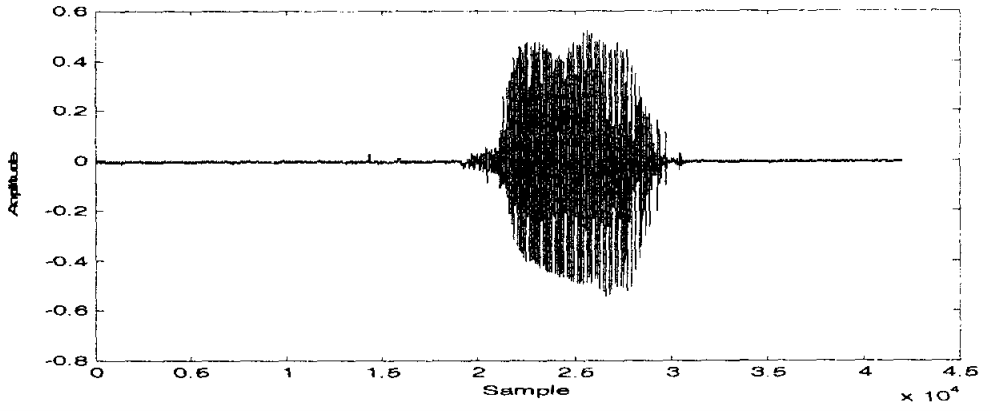
รูปที่ 4.5 การเตรียมสัญญาณเสียงสี่

ก เสียงตั้งต้น

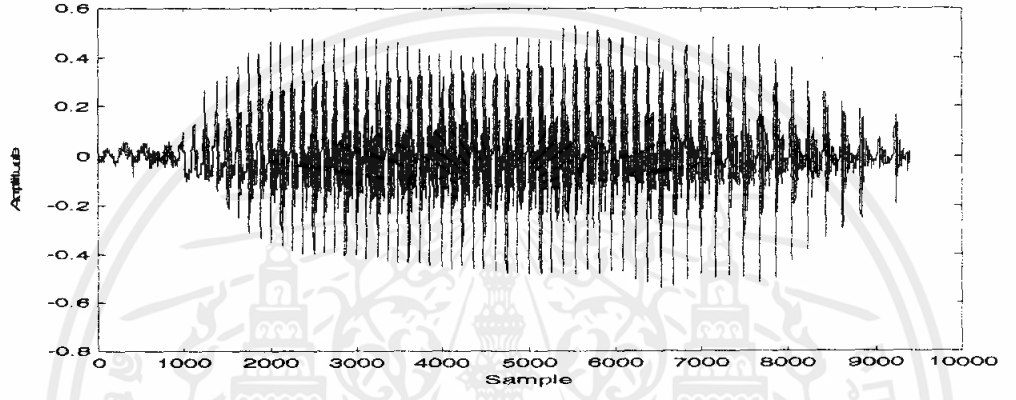
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

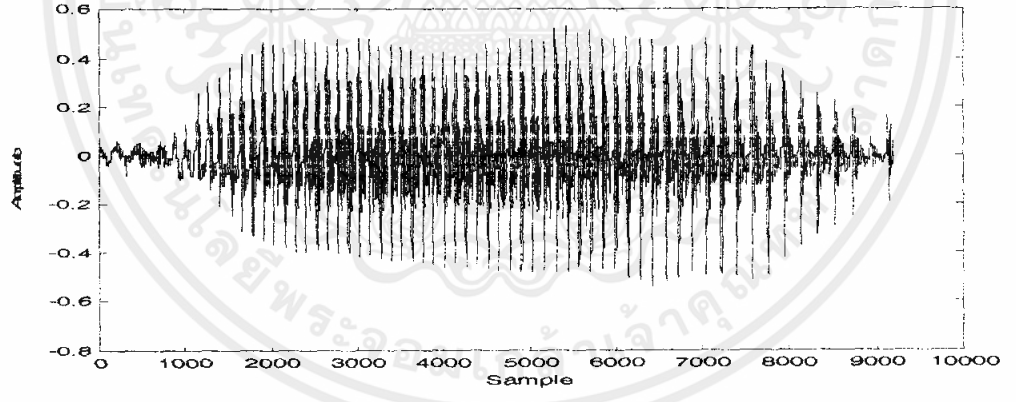
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

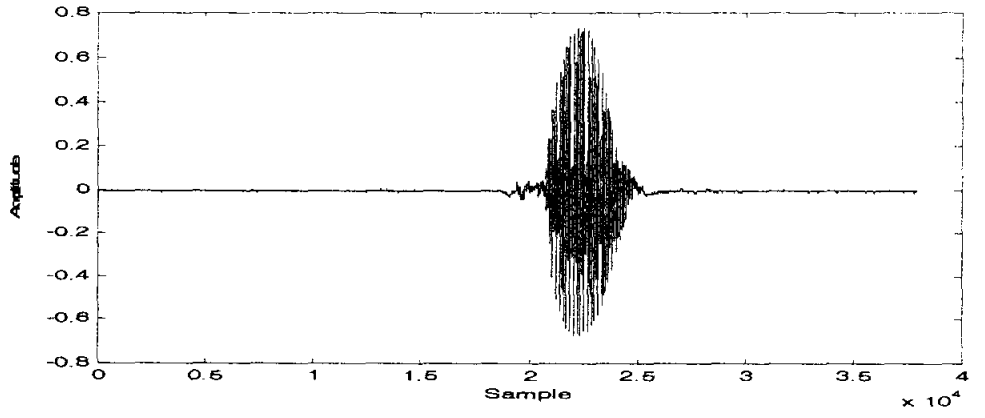
รูปที่ 4.6 การเตรียมสัญญาณเสียงห้า

ก เสียงตั้งต้น

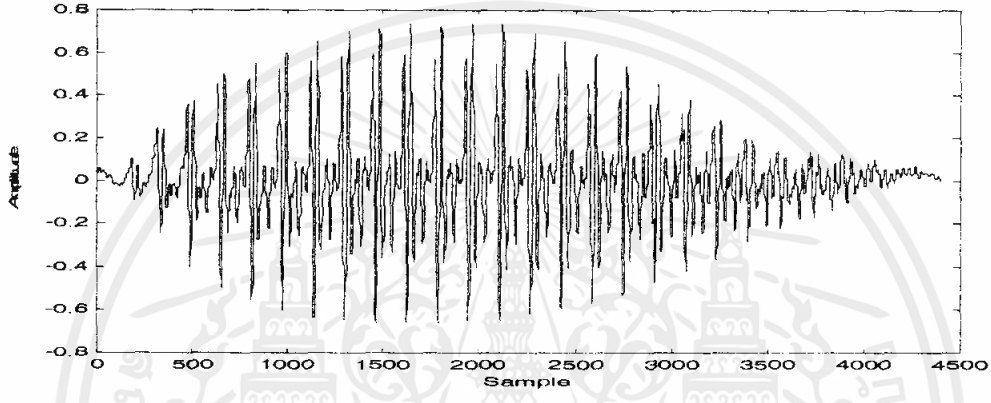
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

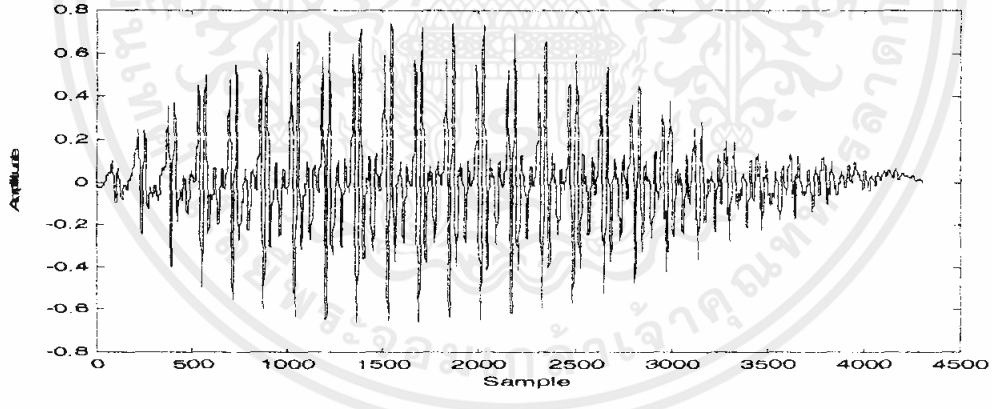
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

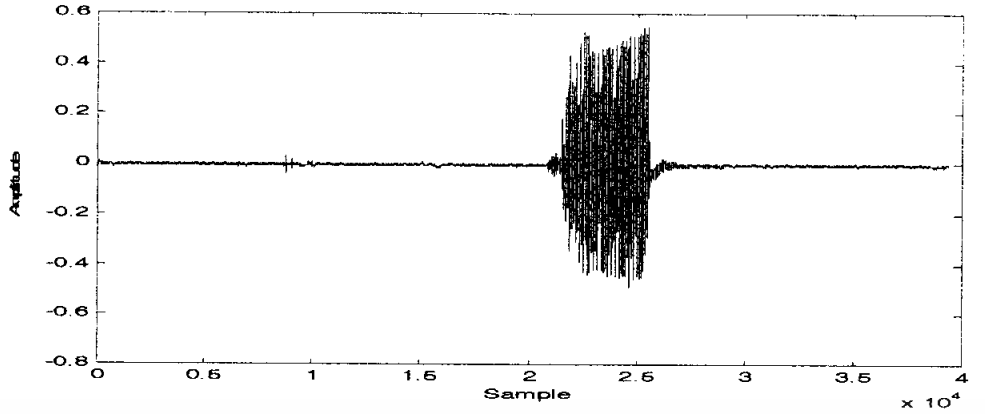
รูปที่ 4.7 การเตรียมสัญญาณเสียงหก

ก เสียงตั้งต้น

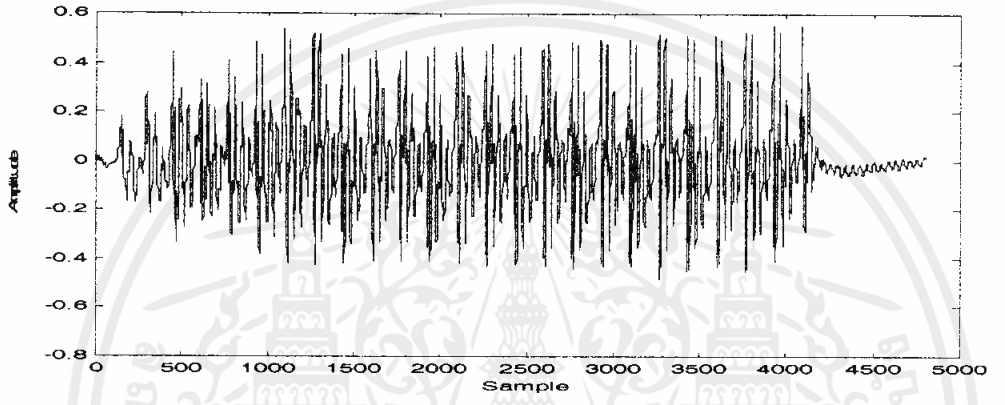
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

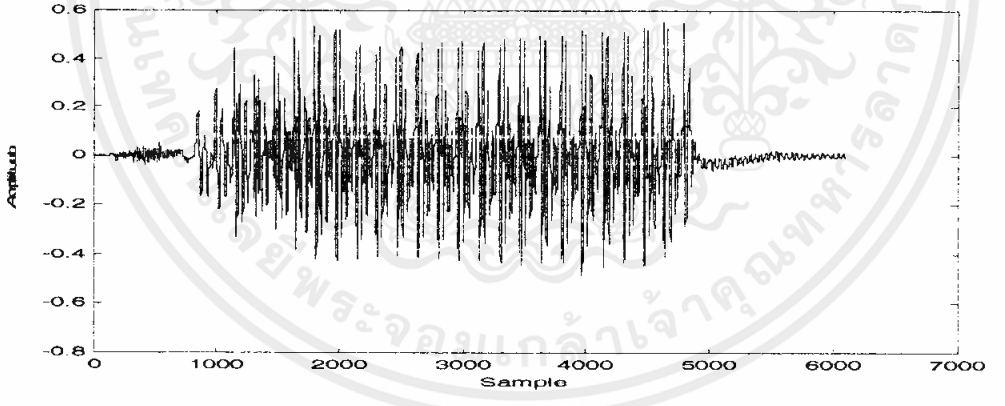
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

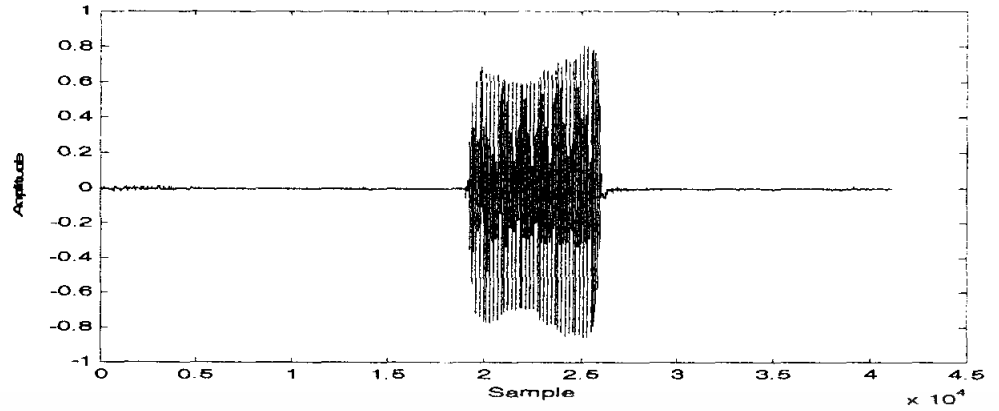
รูปที่ 4.8 การเตรียมสัญญาณเสียงเจ็ด

ก เสียงตั้งต้น

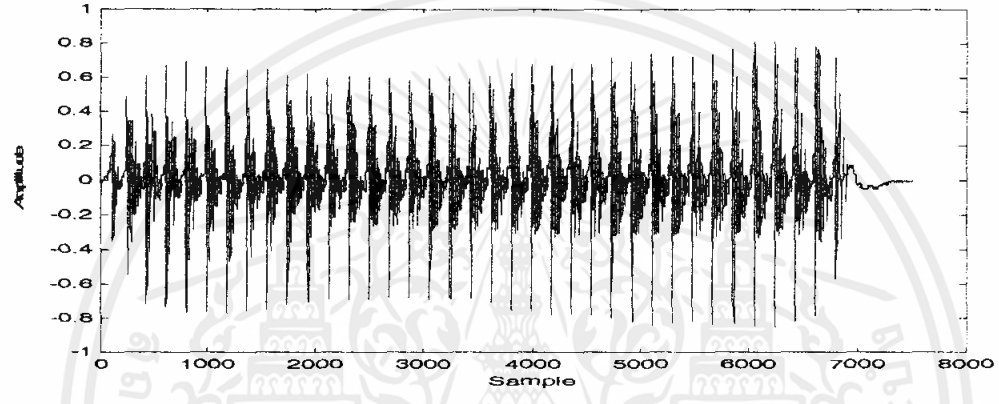
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

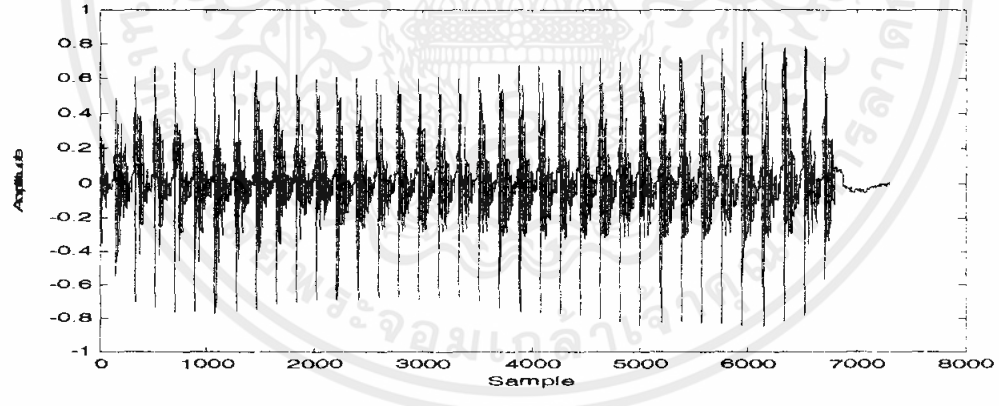
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

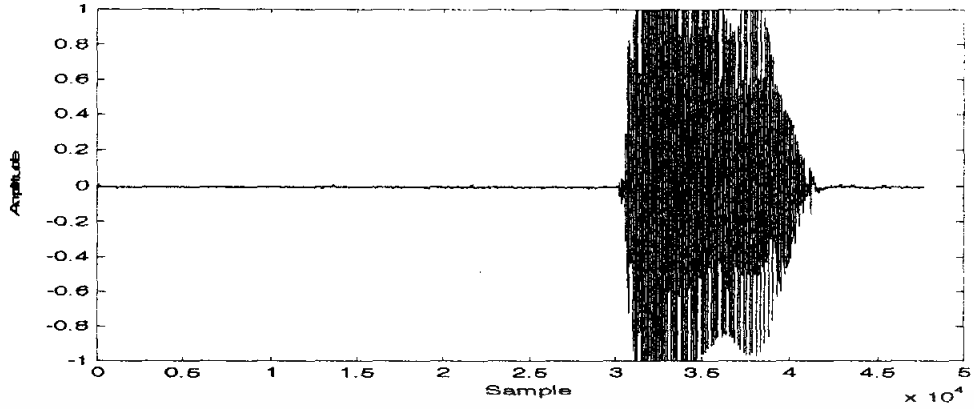
รูปที่ 4.9 การเตรียมสัญญาณเสียงแปด

ก เสียงตั้งต้น

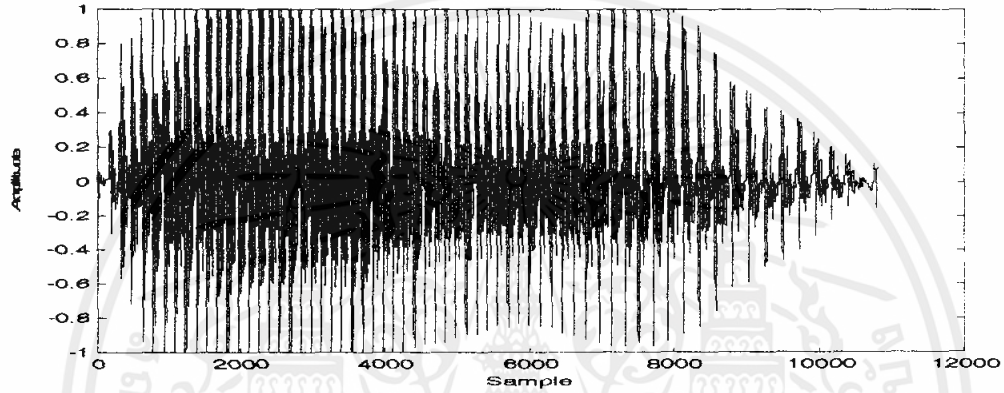
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

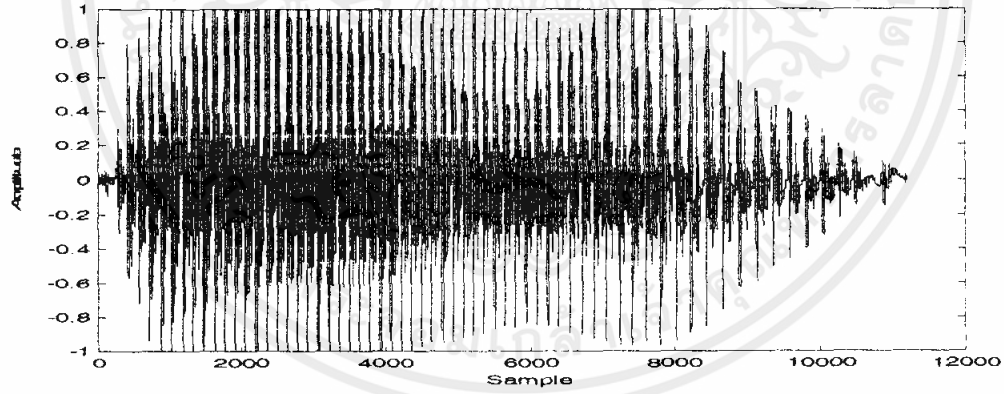
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข



ค

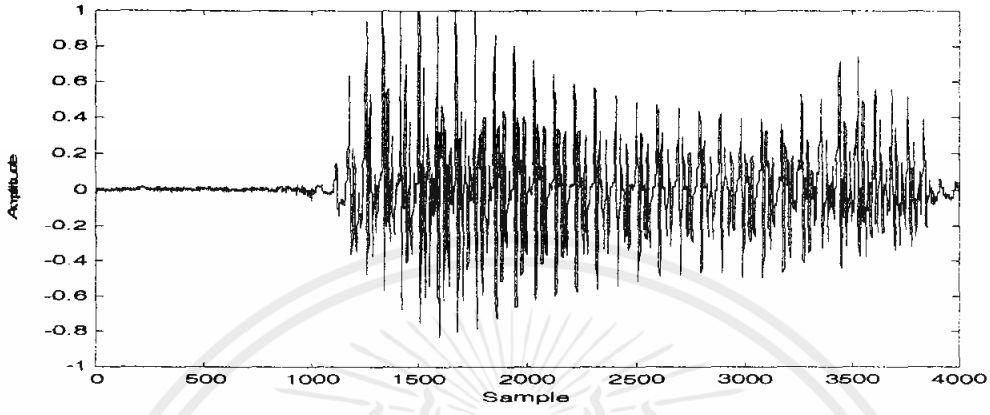
รูปที่ 4.10 การเตรียมสัญญาณเสียงเก้า

ก เสียงตั้งต้น

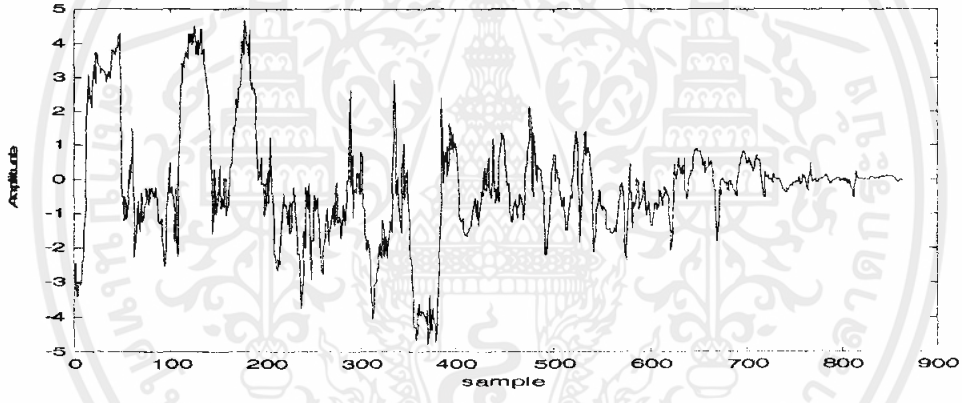
ข กราฟการหาขอบเขตของคำ

ค กราฟการหาค่าอัตราการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



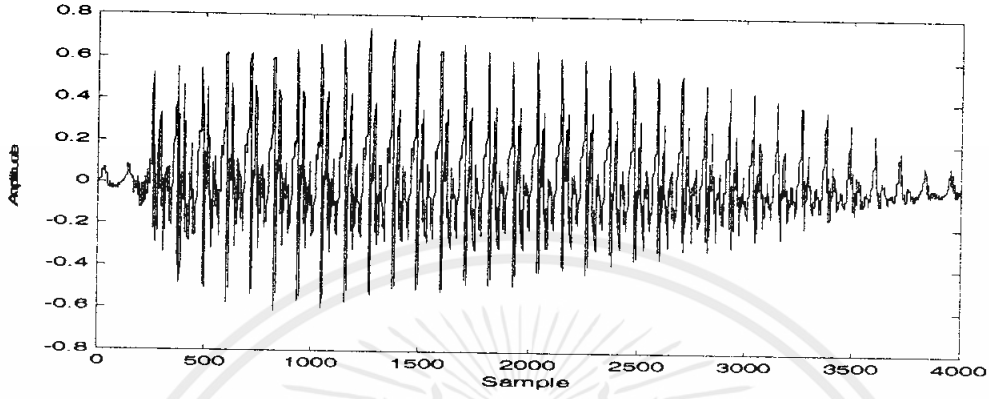
ก



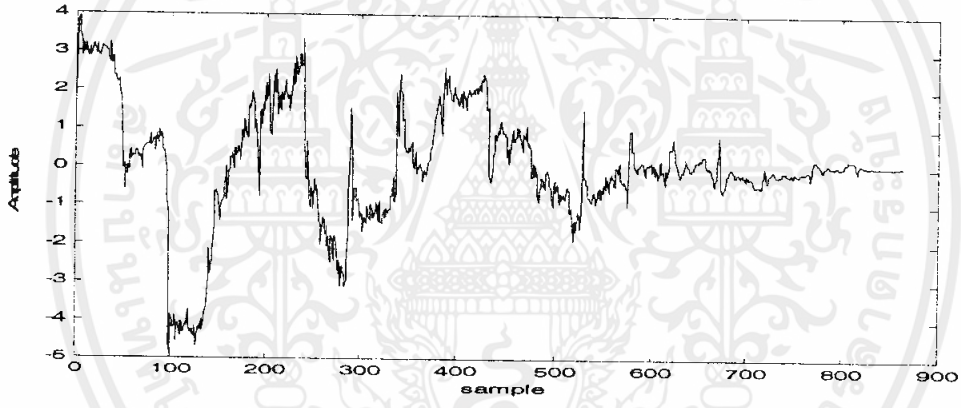
ข

รูปที่ 4.11 การเตรียมสัญญาณเสียงศูนย์
ก กราฟการนอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



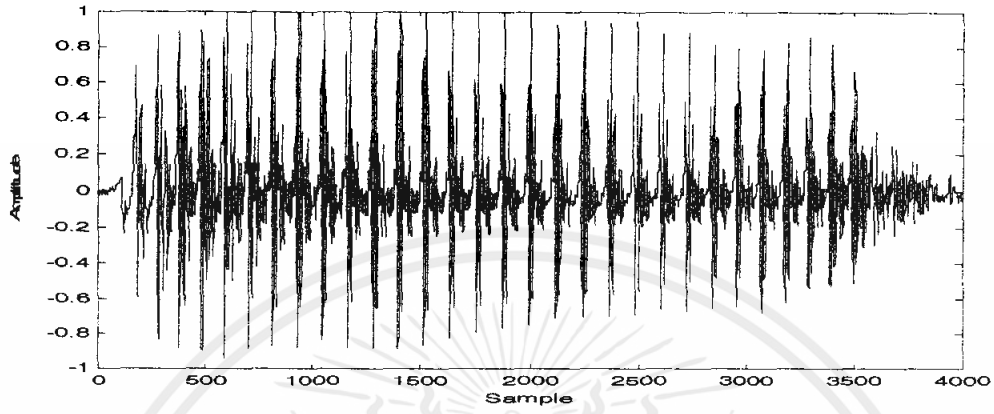
ก



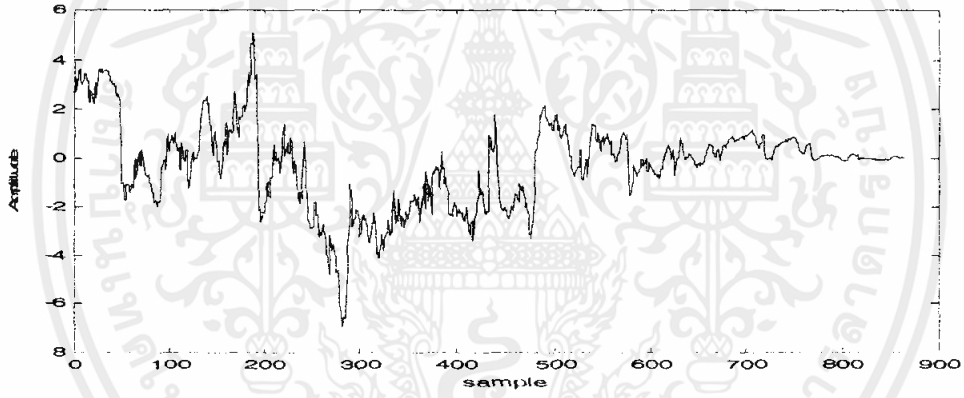
ข

รูปที่ 4.12 การเตรียมสัญญาณเสียงหนึ่ง
ก กราฟกรานอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



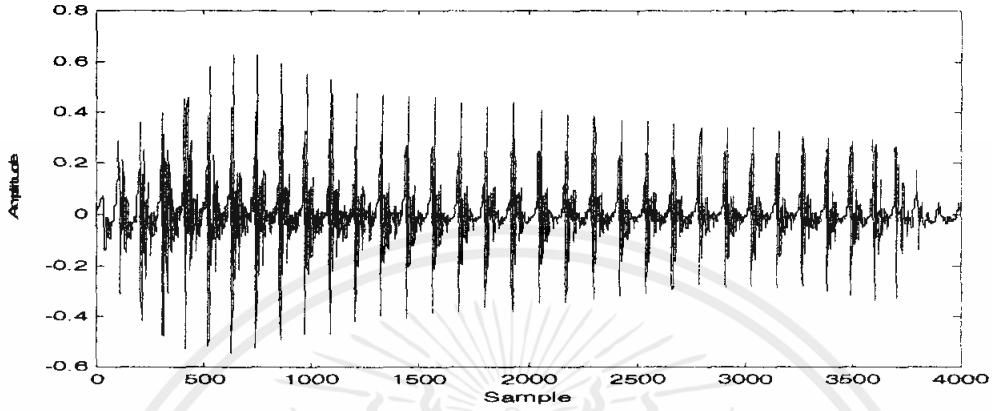
ก



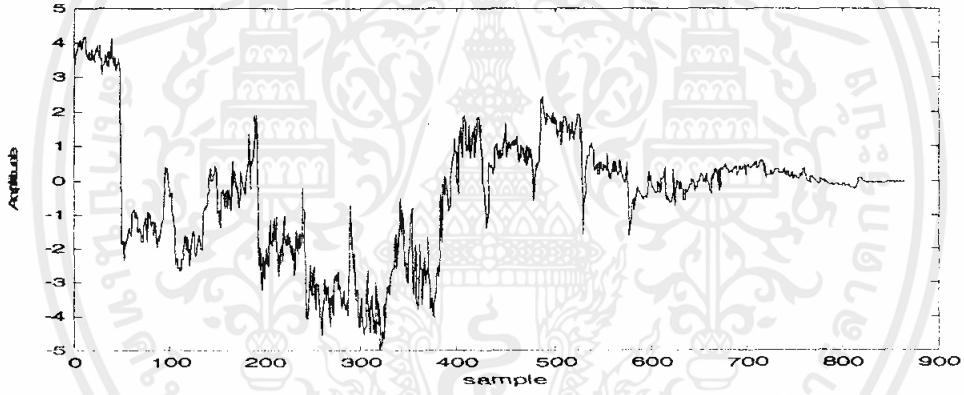
ข

รูปที่ 4.13 การเตรียมสัญญาณเสียงสอง
ก กราฟการนอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



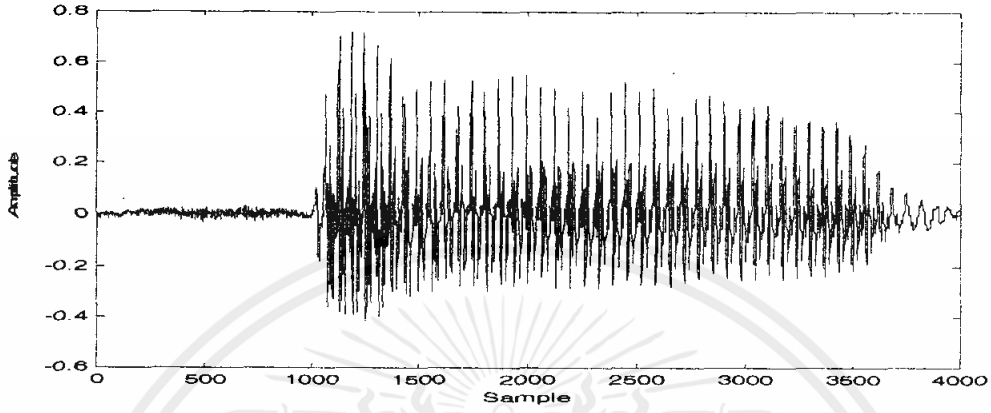
ก



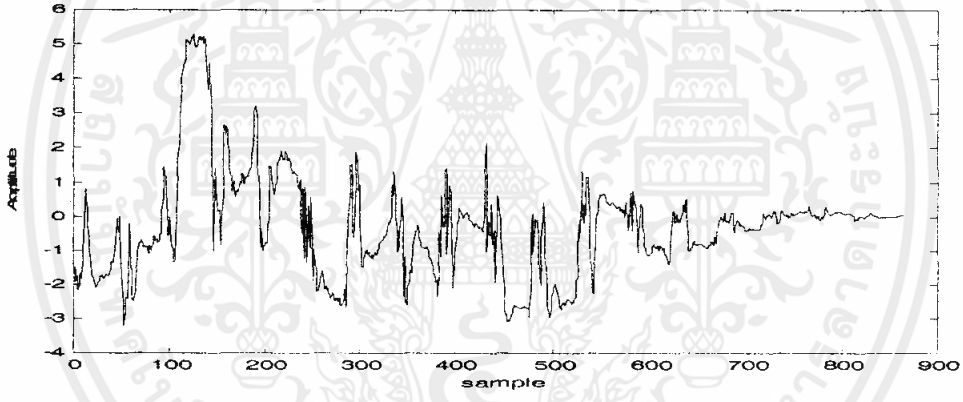
ข

รูปที่ 4.14 การเตรียมสัญญาณเสียงสาม
ก กราฟการนอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



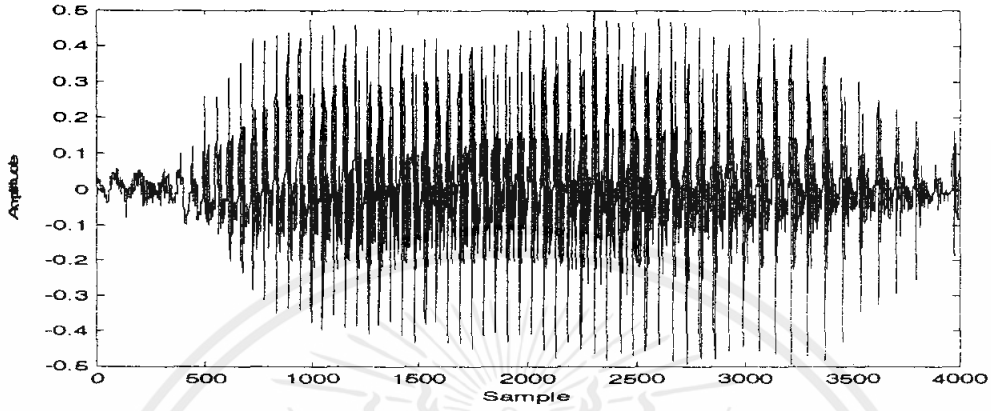
ก



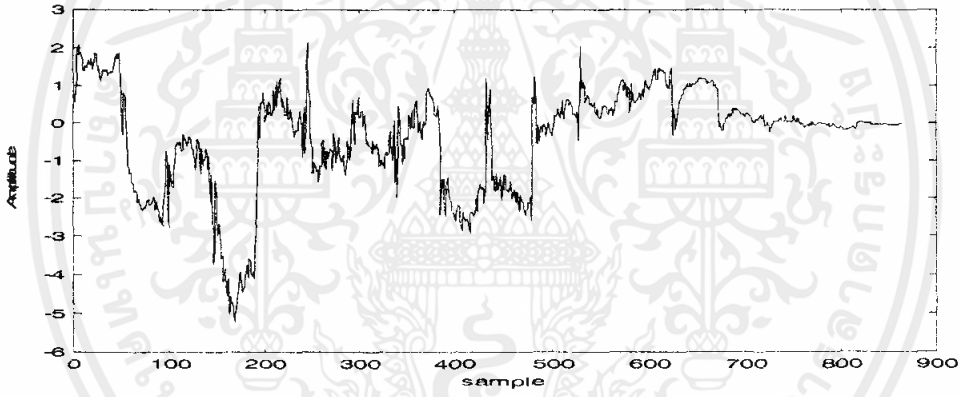
ข

รูปที่ 4.15 การเตรียมสัญญาณเสียงสี่
ก กราฟการนอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



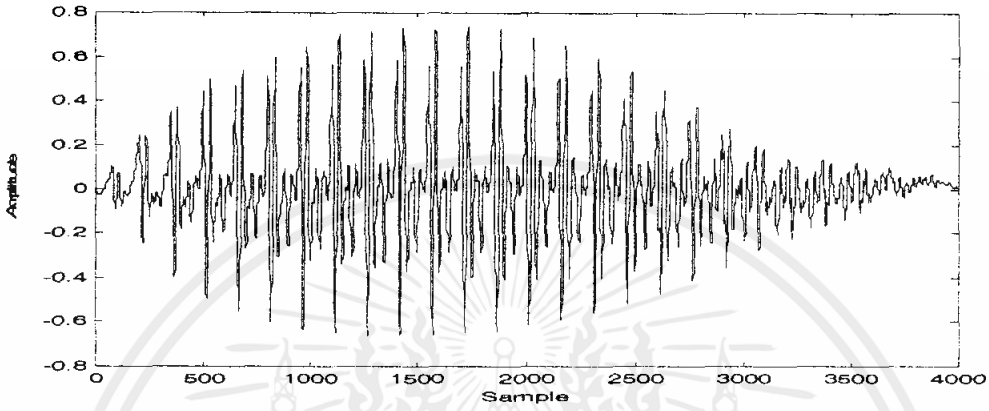
ก



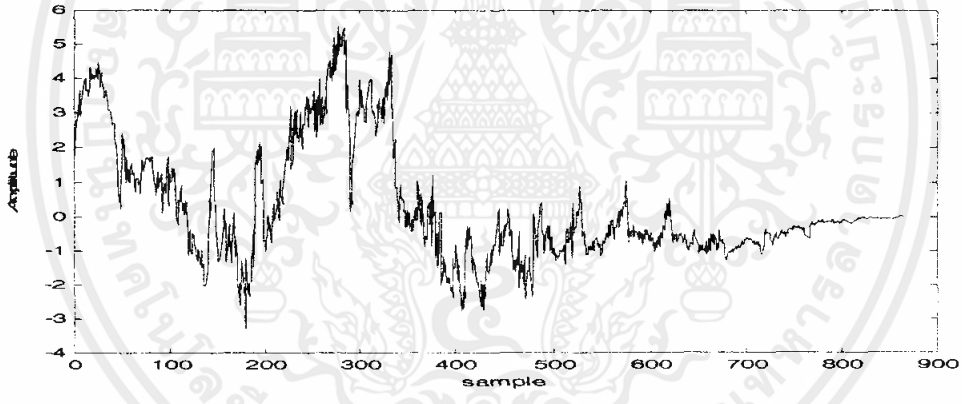
ข

รูปที่ 4.16 การเตรียมสัญญาณเสียงห้า
ก กราฟการนอร์มอลไลซ์
ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



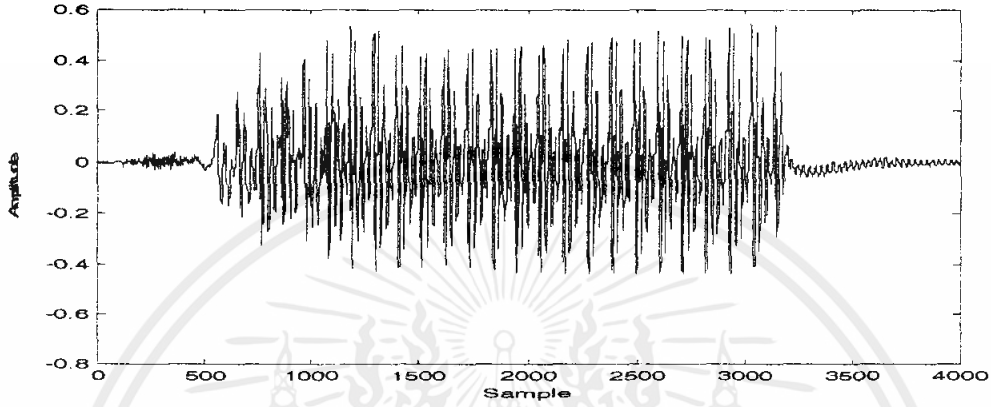
ข

รูปที่ 4.17 การเตรียมสัญญาณเสียงหก

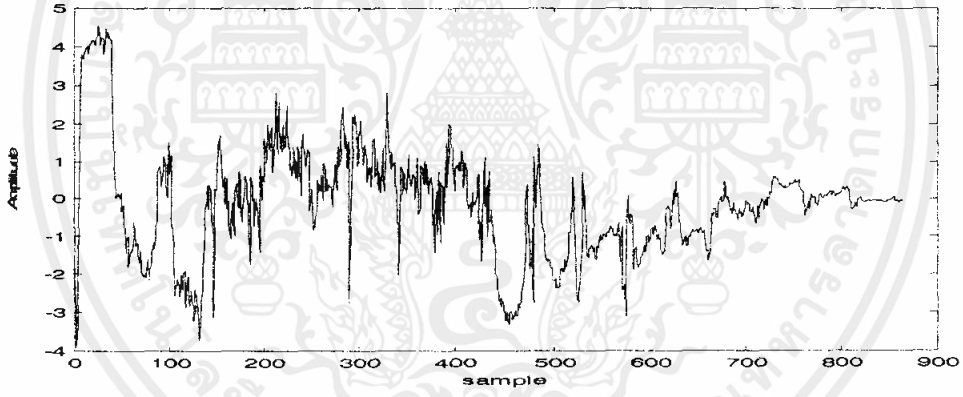
ก กราฟการนอร์มอลไลซ์

ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



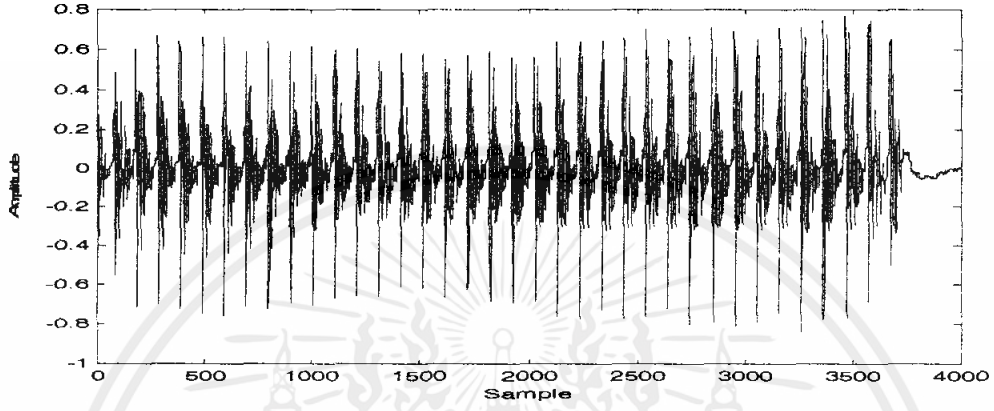
ข

รูปที่ 4.18 การเตรียมสัญญาณเสียงเจ็ด

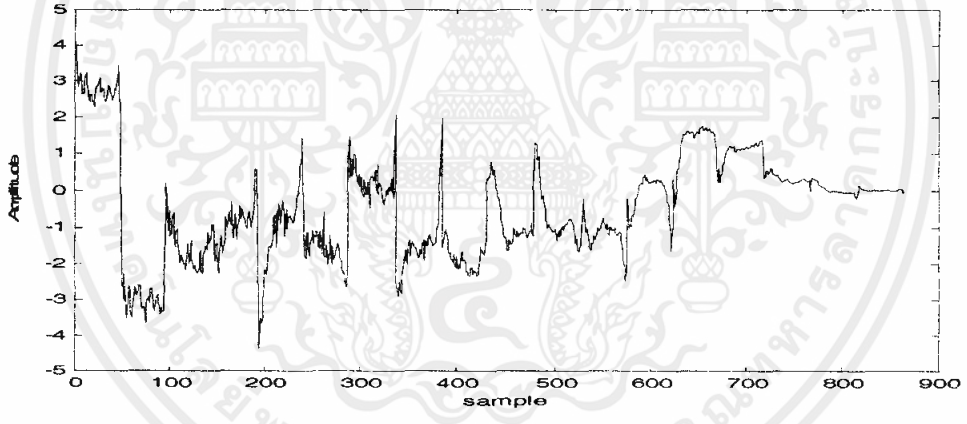
ก กราฟการนอร์มอลไลซ์

ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



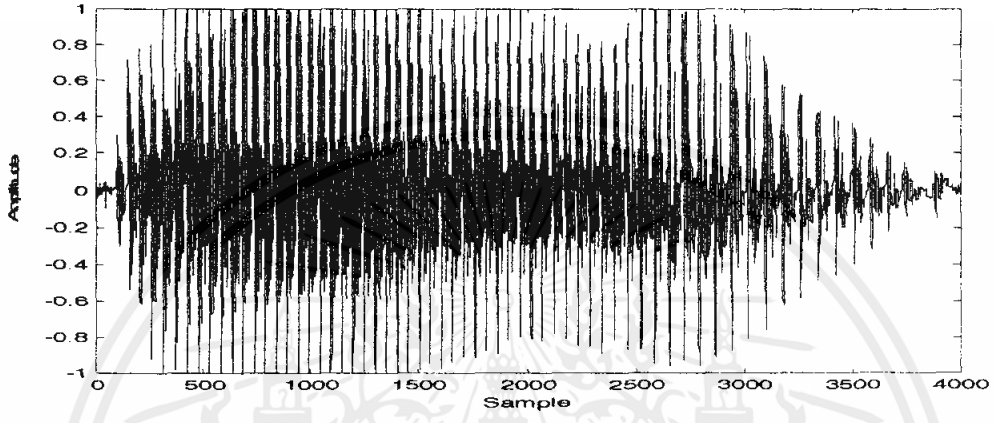
ข

รูปที่ 4.19 การเตรียมสัญญาณเสียงแปด

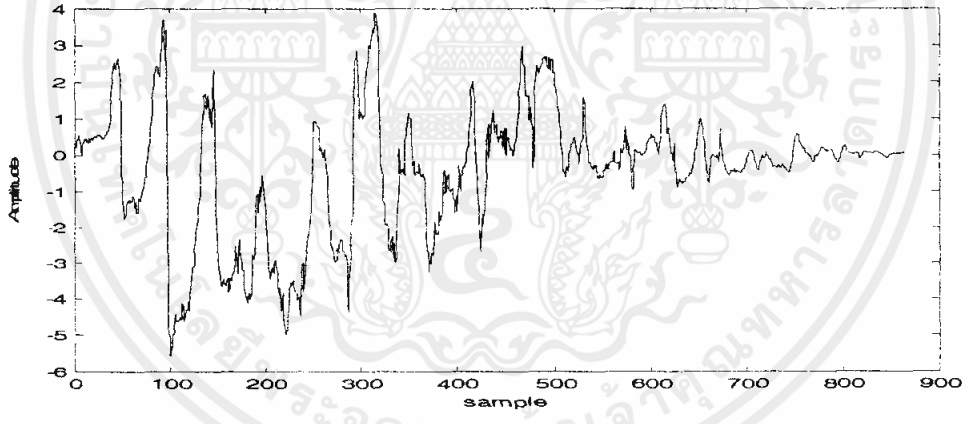
ก กราฟการนอร์มอลไลซ์

ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก



ข

รูปที่ 4.20 การเตรียมสัญญาณเสียงเก้า

ก กราฟการนอร์มอลไลซ์

ข กราฟสัมประสิทธิ์เซปตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ขั้นตอนการเรียนรู้

1. เสียงที่ใช้ในการสร้างโมเดลเป็นเสียง 0-9 มาจากการเก็บจากตัวอย่างผู้หญิง 6 คน คนละ 5 ครั้ง และผู้ชาย 6 คน คนละ 5 ครั้ง รวมตัวอย่างเสียงทั้งหมด 300 เสียง โดยเก็บเป็นนามสกุล .wav และความถี่ที่ใช้เก็บเป็น 22 kHz 8 บิท
2. ทำการหาค่าพลังงาน หาอัตราการตัดศูนย์ และการนอร์มอลไลซ์
3. ทำการวิเคราะห์หาค่าพารามิเตอร์ โดยใช้วิธีแอลพีซี
4. ทำการเปลี่ยนสัมประสิทธิ์แอลพีซีไปเป็นสัมประสิทธิ์เซปตรัม ทำการเวทค่าพารามิเตอร์เพื่อความแม่นยำในการวิเคราะห์
5. ทำเวคเตอร์ควอนไตซ์เซชั่นของสัมประสิทธิ์เซปตรัม เพื่อเป็นการลดจำนวนข้อมูลและนำไปใช้ต่อในการทำโค้ดบุคเพื่อนำมาทำการหารูปแบบของเสียงที่เป็นต้นแบบในการเปรียบเทียบ
6. สร้างแบบอ้างอิงโดยใช้แบบจำลองฮิดเดนมาร์คอฟ โดยการนำโค้ดบุคของแต่ละกลุ่มที่เก็บเอาไว้มาทำการทดลอง และนำมาเปรียบเทียบในขั้นตอนสุดท้าย

4.3 ขั้นตอนการตระหนักรู้เสียง

เป็นขั้นตอนที่จะทำให้ทราบว่า สัญญาณเสียงที่ต้องการทดสอบนั้นควรเป็นเสียงใดมากที่สุด โดยนำมาเปรียบเทียบกับแบบอ้างอิงที่ได้ทำไว้ ดังขั้นตอนต่อไปนี้

1. นำสัญญาณเสียงที่ต้องการทดสอบมาผ่านการวิเคราะห์สัญญาณ และผ่านขั้นตอนการหาสัมประสิทธิ์แอลพีซีและสัมประสิทธิ์เซปตรัมพร้อมทำการเวทค่าพารามิเตอร์เรียบร้อยแล้ว ก็นำพารามิเตอร์ที่ได้มาเปรียบเทียบกับโค้ดบุค จะได้โค้ดบุคอินเด็กซ์ของเสียงที่ต้องการทดสอบ
2. นำโค้ดบุคที่หาค่าได้มาหาค่าความน่าจะเป็นกับทุกๆ แบบอ้างอิงที่เก็บไว้ โดยในการทดลองจะใช้วิธี วิทเทอร์บี ซึ่งจะได้ความน่าจะเป็นของเสียงที่ต้องการทดสอบเทียบกับแต่ละแบบอ้างอิง และเมื่อนำค่าความน่าจะเป็นมาเปรียบเทียบกัน โดยค่าความน่าจะเป็นค่าใดมีค่ามากที่สุด ผลจะได้ว่าเสียงที่นำมาทดสอบตรงกับแบบอ้างอิงนั้น
3. แสดงผลว่าได้เป็นเสียงใด

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.1 การทดสอบการตระหนักรู้เสียงพูด

เป็นตารางผลการทดสอบ โดยจะจำแนกเป็น 2 กรณี คือ เสียงที่ใช้เป็นต้นแบบ (ที่ได้จากแบบจำลองอ้างอิง) และเสียงของบุคคลอื่น โดยใช้เสียง 0-9 ของผู้ชายและผู้หญิง เสียงละ 40 ครั้ง ทำตามขนาดของโค้ดบุค 32 64 และ 128

ตารางที่ 4.1 การตระหนักรู้เสียงต้นแบบของผู้ชายโดยใช้โค้ดบุค 32

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	32	0	3	2	2	0	0	0	0	1	80
1	0	35	2	2	0	0	0	1	0	0	87.5
2	3	0	30	3	2	0	0	1	0	1	75
3	2	0	3	29	3	0	1	1	0	1	72.5
4	1	0	2	3	34	0	0	0	0	0	85
5	0	0	1	2	0	36	0	1	0	0	90
6	1	0	1	1	1	0	36	0	0	0	90
7	1	0	1	2	0	0	0	35	0	1	87.5
8	0	1	2	2	0	0	1	0	34	0	85
9	0	1	2	1	0	1	0	0	1	36	90

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้โค้ดบุค 32

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	21	1	5	6	3	0	0	1	1	2	52.5
1	1	25	3	4	3	0	1	2	0	1	62.5
2	6	2	19	7	4	0	1	0	1	0	47.5
3	6	0	7	20	5	1	0	0	0	1	50
4	5	0	5	4	25	1	0	0	0	0	62.5
5	4	0	5	2	1	26	0	2	0	0	65
6	0	0	4	6	0	0	28	0	2	0	70
7	0	0	4	6	0	0	4	27	0	3	67.5
8	3	2	2	3	0	3	0	2	25	0	62.5
9	5	0	0	7	2	0	0	0	0	26	65

ตารางที่ 4.3 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 32

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	33	0	4	3	0	0	0	0	0	0	82.5
1	1	38	0	1	0	0	0	0	0	0	95
2	4	0	29	6	1	0	0	0	0	0	72.5
3	4	0	4	30	2	0	0	0	0	2	75
4	0	2	3	0	35	0	0	0	0	0	87.5
5	0	0	0	0	0	38	0	1	1	0	95
6	0	0	0	0	0	0	39	0	0	1	97.5
7	0	0	1	0	0	1	0	38	0	0	95
8	0	0	0	1	0	0	0	0	39	0	97.5
9	0	0	0	0	0	0	0	0	0	40	100

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 การตระหนักเสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 32

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	30	0	8	2	0	0	0	0	0	0	75
1	0	31	0	3	3	0	0	0	0	3	77.5
2	5	0	24	6	2	2	1	0	0	0	60
3	0	0	5	25	4	1	2	1	0	2	62.5
4	5	0	4	0	28	0	0	0	0	3	70
5	0	2	0	0	3	31	0	2	0	2	77.5
6	3	0	2	2	0	0	31	0	2	0	77.5
7	0	0	4	5	0	0	0	31	0	0	77.5
8	2	0	0	3	0	4	0	0	30	0	75
9	0	0	3	3	0	0	0	0	2	38	80

ตารางที่ 4.5 การตระหนักรู้เสียงต้นแบบของผู้ชายโดยใช้ โค้ดบุค 64

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	35	0	2	3	0	0	0	0	0	0	87.5
1	0	37	0	0	0	0	0	1	0	2	92.5
2	3	0	34	3	0	0	0	0	0	0	85
3	1	0	3	34	0	0	0	1	0	1	85
4	0	0	2	0	37	0	0	0	1	0	92.5
5	0	1	0	0	0	38	1	0	0	0	95
6	0	0	0	0	0	2	38	0	0	0	95
7	0	0	0	0	0	0	0	39	0	0	97.5
8	0	0	0	0	2	0	0	0	38	0	95
9	0	0	0	1	0	0	0	0	0	39	97.5

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 64

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	23	0	6	7	3	0	0	1	1	1	57.5
1	0	26	4	3	0	3	0	2	0	2	65
2	4	0	21	8	2	2	0	1	1	1	52.5
3	5	0	7	22	4	0	1	0	1	0	55
4	3	1	3	4	26	0	0	1	2	0	65
5	0	3	3	0	2	26	5	0	1	0	65
6	0	2	3	0	0	4	29	0	0	2	72.5
7	2	0	0	4	3	3	0	28	0	0	70
8	3	0	0	6	0	0	0	0	28	3	70
9	0	0	5	3	0	4	0	0	0	28	70

ตารางที่ 4.7 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 64

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	38	0	0	0	2	0	0	0	0	0	95
1	0	39	0	0	0	0	0	0	0	0	97.5
2	1	0	38	1	0	0	0	0	0	0	95
3	0	0	0	39	0	0	0	1	0	0	97.5
4	1	0	0	0	39	0	0	0	0	0	97.5
5	1	0	0	0	0	39	0	0	0	0	97.5
6	1	0	0	0	0	0	39	0	0	0	97.5
7	0	0	0	0	0	0	1	38	1	0	95
8	0	0	0	0	0	0	0	0	40	0	100
9	0	0	0	0	0	0	0	0	0	40	100

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 การตระหนักเสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 64

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	30	0	2	2	0	0	2	0	2	2	75
1	0	31	0	0	3	3	0	2	0	1	77.5
2	3	0	25	7	5	0	0	0	0	0	62.5
3	4	0	6	25	5	0	0	0	0	0	62.5
4	1	2	0	3	31	0	1	0	2	0	77.5
5	0	0	0	0	0	32	4	2	0	2	80
6	0	3	2	0	0	4	31	0	0	0	77.5
7	0	0	0	4	0	0	1	31	0	4	77.5
8	0	0	2	2	0	2	0	0	32	2	80
9	0	0	1	4	0	0	0	0	2	33	82.5

ตารางที่ 4.9 การตระหนักรู้เสียงต้นแบบของผู้ชายโดยใช้ โค้ดบุค 128

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	40	0	0	0	0	0	0	0	0	0	100
1	0	39	1	0	0	0	0	0	0	0	97.5
2	0	0	38	1	1	0	0	0	0	0	95
3	0	0	1	39	0	0	0	0	0	0	97.5
4	0	0	0	0	40	0	0	0	0	0	100
5	0	0	0	0	0	40	0	0	0	0	100
6	0	0	0	0	0	0	40	0	0	0	100
7	0	0	0	0	0	0	0	40	0	0	100
8	0	0	0	0	0	0	0	0	40	0	100
9	0	0	0	0	0	0	0	0	0	40	100

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.10 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 128

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	24	0	6	5	0	1	1	1	0	2	60
1	1	26	0	4	3	0	0	2	1	3	65
2	6	2	22	7	0	2	0	1	0	0	55
3	5	0	0	22	6	1	2	0	2	2	55
4	2	0	4	3	26	0	1	3	0	1	65
5	2	0	0	0	2	29	4	0	3	0	67.5
6	0	0	1	3	0	5	29	0	0	2	72.5
7	0	0	5	4	0	2	0	29	0	0	72.5
8	0	0	3	0	3	0	3	0	29	2	72.5
9	2	1	0	3	0	2	0	2	0	30	75

ตารางที่ 4.11 การตระหนักรู้เสียงต้นแบบของผู้หญิงโดยใช้ โค้ดบุค 128

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	40	0	0	0	0	0	0	0	0	0	100
1	0	40	0	0	0	0	0	0	0	0	100
2	0	0	39	1	0	0	0	0	0	0	97.5
3	0	0	0	40	0	0	0	0	0	0	100
4	0	0	0	0	40	0	0	0	0	0	100
5	0	0	0	0	0	40	0	0	0	0	100
6	0	0	0	0	0	0	40	0	0	0	100
7	0	0	0	0	0	0	0	40	0	0	100
8	0	0	0	0	0	0	0	0	40	0	100
9	0	0	0	0	0	0	0	0	0	40	100

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 การตระหนักรู้เสียงบุคคลอื่นของผู้ชายโดยใช้ โค้ดบุค 128

เสียงที่ต้องการ	ผลการตระหนักรู้เสียงพูด										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	30	0	4	3	1	2	0	0	0	0	75
1	0	31	1	3	3	2	0	0	0	0	77.5
2	1	0	26	5	3	0	2	2	0	1	65
3	3	0	4	26	3	0	2	0	1	1	65
4	1	1	2	3	31	0	0	2	0	0	77.5
5	1	0	0	0	0	33	4	0	2	0	82.5
6	0	0	0	3	0	3	31	1	2	0	77.5
7	1	0	2	0	2	0	1	32	0	2	80
8	2	0	3	0	0	2	1	0	32	0	80
9	0	0	2	3	1	0	0	1	0	33	82.5

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ค่าความถูกต้องการตระหนักรู้เสียง

ผู้ทดลอง	เปอร์เซ็นต์ต่ำสุด	เปอร์เซ็นต์สูงสุด	เปอร์เซ็นต์เฉลี่ย
โค้ดบุค 32	เสียงผู้ชาย		
เสียงจากต้นแบบ	72.5	90	84.5
เสียงจากบุคคลอื่น	47.5	70	60.5
ผลรวม			**72.38**
	เสียงผู้หญิง		
เสียงจากต้นแบบ	72.5	100	89.75
เสียงจากบุคคลอื่น	60	77.5	73.25
ผลรวม			**81.5**
โค้ดบุค 64	เสียงผู้ชาย		
เสียงจากต้นแบบ	85	97.5	92.5
เสียงจากบุคคลอื่น	52.5	72.5	64.25
ผลรวม			**78.25**
	เสียงผู้หญิง		
เสียงจากต้นแบบ	95	100	97.25
เสียงจากบุคคลอื่น	62.5	82.5	75.25
ผลรวม			**86.25**
โค้ดบุค 128	เสียงผู้ชาย		
เสียงจากต้นแบบ	95	100	99
เสียงจากบุคคลอื่น	55	75	66
ผลรวม			**82.5**
	เสียงผู้หญิง		
เสียงจากต้นแบบ	97.5	100	99.75
เสียงจากบุคคลอื่น	65	82.5	76.25
ผลรวม			**88**

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 การเปรียบเทียบจำนวนเสียงในแบบอ้างอิง

ในส่วนนี้ทำการทดสอบเพื่อศึกษาว่า จำนวนเสียงที่นำมาทำแบบจำลองอ้างอิงมีผลต่อการตระหนักรู้เสียงพูดอย่างไร โดยทำการทดสอบด้วยเสียงของผู้หญิง 6 คน คนละ 1 2 และ 3 ครั้งของแต่ละเสียง โดยสร้างแบบจำลองอ้างอิงจากโค้ดบุค128

ตารางที่ 4.14 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน เสียงละ 1 ครั้ง

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียง										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	13	0	6	7	8	0	2	1	1	3	32.5
1	4	14	5	7	4	0	2	1	1	2	35
2	6	2	12	7	5	0	3	0	2	3	30
3	5	1	8	12	5	2	1	3	1	2	30
4	6	0	7	8	13	0	0	3	0	3	32.5
5	4	2	4	4	5	13	7	0	0	1	32.5
6	4	1	4	4	4	8	13	0	1	1	32.5
7	4	2	3	3	4	2	4	13	3	2	32.5
8	3	2	5	5	4	0	3	0	14	4	35
9	4	1	4	5	5	2	1	1	2	14	35

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน เสียงละ 2 ครั้ง

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียง										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	20	1	5	6	3	0	1	2	0	2	50
1	4	21	3	2	2	0	2	0	2	4	52.5
2	5	1	20	8	2	0	0	0	0	4	50
3	4	1	9	19	3	2	0	0	1	1	47.5
4	2	0	6	7	21	0	1	0	2	1	52.5
5	0	0	3	3	0	21	6	0	3	4	52.5
6	1	0	2	2	0	5	22	0	3	5	55
7	1	1	1	2	3	2	0	22	2	6	55
8	3	1	3	2	2	1	2	1	20	5	50
9	4	1	5	5	0	0	2	0	11	22	55

ตารางที่ 4.16 แบบจำลองอ้างอิงเสียงผู้หญิง 6 คน เสียงละ 3 ครั้ง

เสียงที่ต้องการทดสอบ	ผลการตระหนักรู้เสียง										เปอร์เซ็นต์ความถูกต้อง
	0	1	2	3	4	5	6	7	8	9	
0	28	0	3	4	0	0	0	0	0	5	70
1	1	30	2	2	0	0	2	0	0	3	75
2	1	1	28	2	0	0	1	0	2	3	70
3	2	1	2	28	2	0	0	0	1	4	70
4	3	0	2	2	28	1	0	1	0	3	70
5	1	0	2	2	0	30	0	0	2	3	75
6	0	0	1	2	1	0	31	1	1	3	77.5
7	1	1	0	0	2	1	1	32	0	2	80
8	0	0	2	1	0	0	0	1	32	4	80
9	1	1	2	2	0	0	0	0	2	32	80

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลการเปรียบเทียบจำนวนเสียงในแบบอ้างอิง

เสียงผู้หญิง 6 คน	เปอร์เซนต์ต่ำสุด	เปอร์เซนต์สูงสุด	เปอร์เซนต์เฉลี่ย
จำนวนเสียงในแบบอ้างอิง คนละ 1 ครั้ง	30	35	32.75
จำนวนเสียงในแบบอ้างอิง คนละ 2 ครั้ง	47.5	55	52
จำนวนเสียงในแบบอ้างอิง คนละ 3 ครั้ง	70	80	74.75

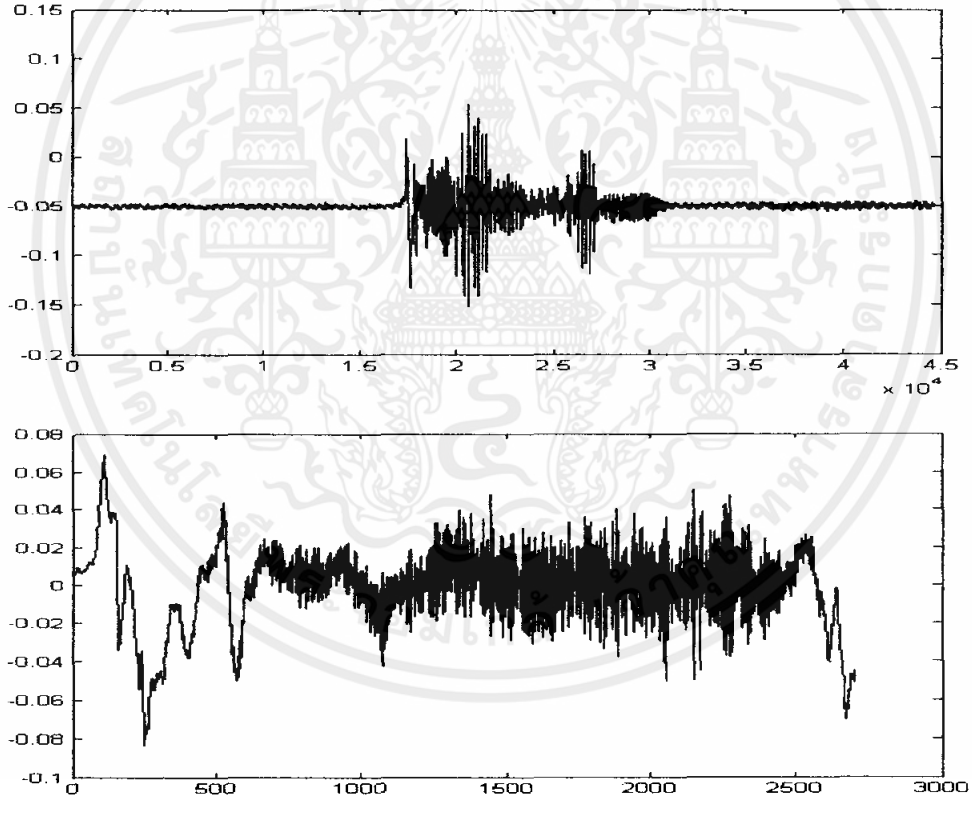
เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทวิจารณ์และบทสรุป

5.1 สรุปผลการทดลอง

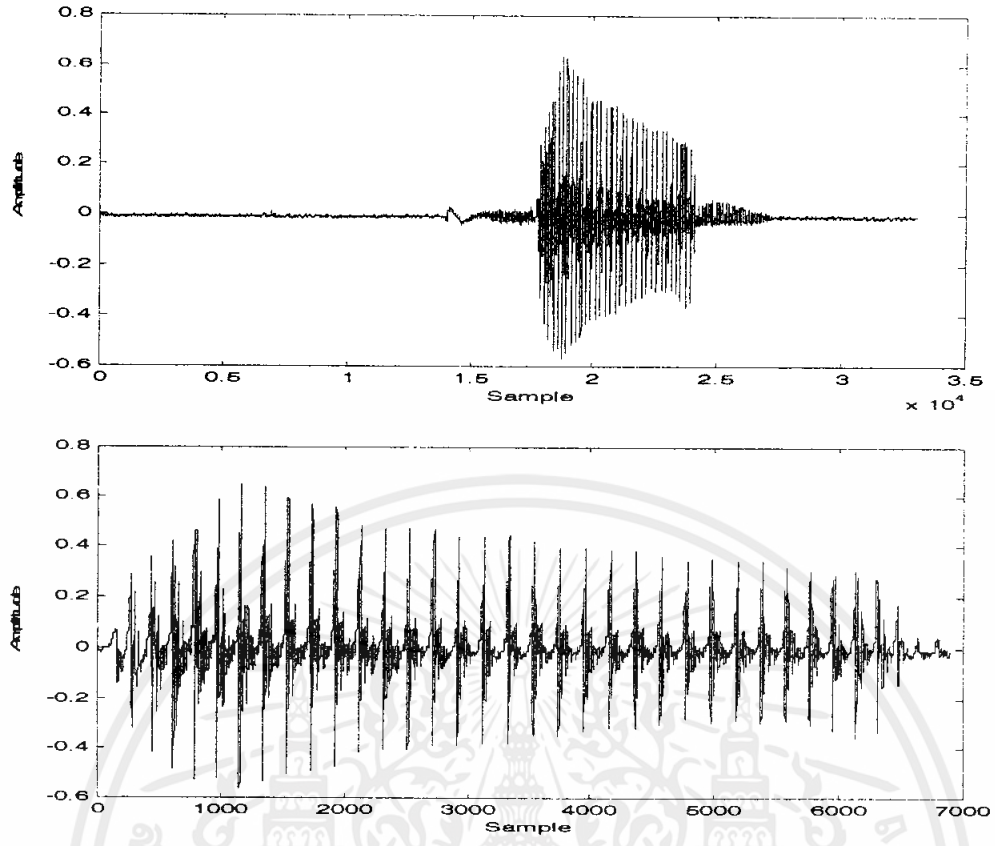
จากการทดลองระบบการตระหนักรู้เสียงพูด ของเสียง 0-9 โดยใช้สัญญาณเสียงที่สุ่มด้วยความถี่ 22 kHz

1) เสียงพูดของคนแม้จะเป็นเสียงเดียวกัน แต่พูด ณ เวลาที่ต่างกัน เมื่อนำมาวิเคราะห์แล้วค่าพารามิเตอร์ของเสียงจะไม่เหมือนกันแต่จะมีความคล้ายคลึงกัน
2) สัญญาณเสียงแต่ละสัญญาณจะมีขอบเขตของสัญญาณ ถ้าหาขอบเขตของสัญญาณได้ถูกต้องจะทำให้สามารถจดจำเสียงพูดได้ถูกต้องแม่นยำมากขึ้น



รูปที่5.1 การหาขอบเขตของสัญญาณเลขสามที่ไม่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 การหาขอบเขตของสัญญาณเลขสามที่ถูกต้อง

3) ขนาดของโค้ดบุคมีผลต่ออัตราการตระหนักรู้เสียงพูด คือ โค้ดบุค128 จะมีอัตราการตระหนักรู้เสียงพูดได้ดีกว่าโค้ดบุค32 และโค้ดบุค64 แต่โค้ดบุคขนาดใหญ่จะใช้เวลาในการคำนวณนานกว่าโค้ดบุคที่มีขนาดเล็ก
4) จำนวนสัญญาณเสียงที่นำมาสร้างเป็นแบบอ้างอิงจะมีผลต่อความถูกต้อง ในการจตระหนักรู้เสียงพูด
5) เสียงของผู้ชายมีอัตราการตระหนักรู้เสียงพูดต่ำกว่าผู้หญิงเนื่องจากผู้ชายจะมีช่องว่างความถี่ของสัญญาณหลากหลายกว่าผู้หญิง
6) ในการตระหนักรู้เสียงพูด แบบอ้างอิงฮิดเดนท์มาร์คอฟสามารถจดจำได้ดีในเสียงของบุคคลที่นำมาเป็นแบบอ้างอิง ส่วนบุคคลอื่นสามารถจดจำได้ดีในระดับหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 วิจารณ์ผลการทดลอง

1) อุปกรณ์ในการบันทึกเสียงที่แตกต่างกันจะมีผลต่ออัตราการตระหนักรู้เสียงในระดับหนึ่ง ดังนั้นควรที่จะใช้อุปกรณ์ที่มีมาตรฐานเดียวกันในการบันทึกเสียงจะทำให้มีอัตราการตระกนักรู้เสียงดีขึ้น
2) ในการบันทึกเสียงไม่ควรมีเสียงรบกวน เพราะเสียงรบกวนจะทำให้อัตตราการตระหนักรู้เสียงพูดลดลง

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] Sadaoki Furui. Digital Speech Processing, Synthesis, and Recognition. 2nd edition. New York:Marcel Dekker, Inc. 1985

[2] ประณิต อ่อนไสว, พรชัย ธรรมานุรัตพันธุ์. **"การรู้จำเสียงพูดอัตโนมัติ."** ปริญญานิพนธ์วิศวกรรมบัณฑิต สาขาวิชาโทรคมนาคม, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.2542.

[3] กิติกร พิพัฒน์นัดดา, สุพจน์ อินทร์ชัยยะ. **"การรู้จำเสียงพูด."** ปริญญานิพนธ์ วิศวกรรมบัณฑิต สาขาวิชาโทรคมนาคม, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2542.

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Detect

```
function[E,a1,b1,ref,s]=detect1(filename);
%  [E,a1,b1,ref,s]=detect1(filename);
%  this function for find endpoint dection by energy
%  when filename : file.wav
%  E:energy each frame
%  S:signal when pass detection by energy
[inpt1]=(filename);
inpt=inpt1-mean(inpt1);
%number of frame
sample = 100;
vnoloop = floor(length(inpt)/sample);
%find energy each of frame
for k=1:vnoloop;
   sum=0;
   for k1=1:sample;
      sum = sum + abs (inpt((k-1)*sample+k1));
   end
   E(k)=sum;
end
%find maximum energy
Emax=max(E);
%set value
a=1/10;
%begin process
ref=a*Emax;
for k=1:vnoloop;
  if E(k)==Emax;
     fixenergy=k;
     break;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end

for k=fixenergy:-1:1;
 if E(k)  <=ref;
    a1=k;
    break;
  end
end
for k1=fixenergy:vnoloop;
   if E(k1) <=ref;
     b1=k1;
     break;
   end
end
num=0;
if isempty(b1)==1|(b1)==vnoloop;
   b1=vnoloop-1;
   for i=(a1-1):b1;
     for i1=1:sample;
        num=num+1;
        s(num)=inpt(i*sample+i1);
      end
   end
else
   for i=(a1-1):(b1+1);
   for i1 = 1:sample;
     num=num+1;
     s(num)=inpt(i*sample+i1);
   end
   end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function zerc=sg_n(s);
for i=1:length(s);
   if s(i)>0;
   zerc=1;
   elseif s(i)==0;
      zerc=0;
   else zerc=-1;
   end
end
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Zerocrossing

```
function [S,z,a2,b2,Zmax,inpt]=zrocrs1(filename,a1,b1);
[inpt1]=(filename);
inpt=inpt1-mean(inpt1);
sample=100;
vnoloop1=floor(length(inpt)/sample);
for k=1:vnoloop1;
   sum=0;
   for k1=1:sample;
      if k1==1
         sum = sum+abs(sg_n(inpt((k-1)*sample+k1)));
      else
         sum = sum+abs(sg_n(inpt((k-1)*sample+k1))-sg_n(inpt((k-1)*sample+(k1-1))));
      end
   end
   z(k)=sum/(2*sample);
end
Zmax=max(z);
a=1/10;
for k=1:vnoloop1;
   if z(k)==Zmax;
      fixzero=k;
      break;
   end
end
refzro=a*Zmax;
for k1 = a1:-1:1;
   if z(k1)<=refzro;
      a2=k1;
      break;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
   end
end
for k2=b1:vnoloop1;
   if z(k2)<=refzro;
      b2=k2;
      break;
   end
end
n=0;
if isempty(b2)==1|b2==vnoloop1|isempty(a2)==1;
   if isempty(a2)==1;
      a2=a1-1;
      b2=vnoloop1-1;
      for k3=a2:b2;
         for k4=1:sample;
            n=n+1;
            S(n)=inpt( (k3*sample)+k4);
         end
      end
else
   b2=vnoloop1-1;
   for k3=a2:b2;
      for k4=1:sample;
         n=n+1;
         S(n)=inpt((k3*sample)+k4);
      end
   end
end
else
   for k5=a2:b2;
      for k6=1:sample;
         n=n+1;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        S(n)=inpt( (k5*sample)+k6);
      end
    end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Normalize

```
function normS=normalize(S,normsize);
normS(1) =S(1);
timestep=(length(S))/(normsize);
time=1;
j=2;
while j< normsize
   time=time+timestep;
   lowtime=floor(time);
   residue=time-lowtime;
   if (lowtime+1)>length(S)
      S(lowtime+1)=0;
   end
   normS(j)=S(lowtime)+residue*(S(lowtime+1)-S(lowtime));
   j=j+1;
end
normS(j)=S(length(S));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

LPC Process

```
function [cw,r,K]=lpcprocess(ipt,order,Q);
format long
S=ipt;
%*****************Pre-Amphasis
for i=1:length(S);
  if i==1;
    s(i)=S(i);
  else s(i)=S(i)-0.9375*S(i-1);
  end
end
%*****************Frame-Blocking
flength=240;
fshift=80;
foverlap=160;
frame=1;
lastframe=floor(length(s)/fshift)-ceil(foverlap/fshift);
%*****************window
for frame=1:lastframe;
  j=1;
  for i=(((frame-1)*fshift)+1):(((frame-1)*fshift)+flength);
    w(frame,j)=s(i)*(0.54-0.46*cos(2*pi*(j-1)/(flength-1)));
    j=j+1;
  end
  %***************auto
  for j=1:order+1;
    r(frame,j)=0;
    for i=1:(flength+1-j);
      r(frame,j)=r(frame,j)+w(frame,i)*w(frame,(i-1+j));
    end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end
%***************lpc
e(1)=r(frame,1);
K(1)=r(frame,2)/e(1);
A(1,1)=K(1);
for i=2:order
  e(i)=(1-K(i-1)^2)*e(i-1);
  j=i;
  sumAr=0;
  for k=1:i-1;
    sumAr=sumAr+A(i-1,k)*r(frame,j);
    j=j-1;
  end
  K(i)=(r(frame,i+1)-sumAr)/e(i);
  for j=i:-1:1;
    if j==i;
      A(i,j)=K(i);
    else
      A(i,j)=A(i-1,j)-K(i)*(A(i-1,i-j));
    end
  end
end
gains(frame)=sqrt(e(order));
for i=1:Q;
  if i<=order;
    a(frame,i)=A(order,i);
  else
    a(frame,i)=0;
  end
end
%%%*********************ccp
c(frame,1)=a(frame,1);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
  for i=2:Q;
    sumca=0;
    for k=1:i-1;
      sumca=sumca+(k*c(frame,k)*a(frame,i-k));
    end
    cep_sum(i)=(1/i)*sumca;
    if i<=order;
      c(frame,i)=a(frame,i)+cep_sum(i);
    else
      c(frame,i)=cep_sum(i);
    end
  end
  %**********************weight
  for i=1:Q;
    cweight(frame,i)=c(frame,i)*(1+(Q/2)*sin((pi*i)/Q));
  end
  cw=cweight;
  cw=cw(:);
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Codebook 32

```
function [cb32,CVTD32]=codebook32(TrainingSetFile);
format long e;
VectorDimension=15;
CodeBookNumber=32;
RejectValue=0.0001;
VTD=2*RejectValue;
%==========Find Number Of Frame==========%
NumberOfFrame=((length(TrainingSetFile)/VectorDimension));
%==========Random Centroid From TrainingSetFile==========%
x=floor(NumberOfFrame.*rand(1,CodeBookNumber))+1;
for nc=1:CodeBookNumber
  for nv=1:VectorDimension
    OldCentroid(((nc-1).*VectorDimension)+nv)=...
      TrainingSetFile(((x(nc)-1).*VectorDimension)+nv);
  end
end
k=0;
%==========Loop==========%
while VTD>RejectValue
%==========Find CodeBook==========%
for nf=1:NumberOfFrame
  for nc=1:CodeBookNumber
    TotalDistance=0;
    Distance=0;
    for nv=1:VectorDimension
      Distance=(TrainingSetFile(((nf-1).*VectorDimension)+nv)-...
        OldCentroid(((nc-1).*VectorDimension)+nv)).^2;
      TotalDistance=TotalDistance+Distance;
    end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    VectorDistance(nc)=TotalDistance;
  end
  [MinDistance(nf),MinIndex(nf)]=min(VectorDistance);
end
%==========Find New Centroid========%
for nc=1:CodeBookNumber
  Cnum=0;
  for nv=1:VectorDimension
    NewCentroid(((nc-1)*VectorDimension)+nv)=0;
  end
  for nf=1:NumberOfFrame
    if nc==MinIndex(nf)
      Cnum=Cnum+1;
      for nv=1:VectorDimension
        NewCentroid(((nc-1)*VectorDimension)+nv)=...
          NewCentroid(((nc-1)*VectorDimension)+nv)+...
          TrainingSetFile(((nf-1)*VectorDimension)+nv);
      end
    end
  end
  if Cnum~=0
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        NewCentroid(((nc-1)*VectorDimension)+nv)./Cnum;
    end
  else
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        OldCentroid(((nc-1)*VectorDimension)+nv);
    end
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%=============Check Distance=========%
d=NewCentroid-OldCentroid;
e=d.^2;
VTD=sum(e);%./CodeBookNumber
k=k+1;
CVTD32(k)=VTD;
%=========Copy Old Centroid To New Centroid=======%
OldCentroid=NewCentroid;
%=================================================%
end
cb32=NewCentroid;
save codebook32 cb32 CVTD32
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Codebook 64

```
function [cb64,CVTD64]=codebook64(TrainingSetFile);
format long e;
VectorDimension=15;
CodeBookNumber=64;
RejectValue=0.0001;
VTD=2*RejectValue;
%==========Find Number Of Frame=========%
NumberOfFrame=((length(TrainingSetFile)/VectorDimension));
%==========Random Centroid From TrainingSetFile========%
   x=floor(NumberOfFrame.*rand(1,CodeBookNumber))+1;
for nc=1:CodeBookNumber
   for nv=1:VectorDimension
      OldCentroid(((nc-1).*VectorDimension)+nv) =...
         TrainingSetFile(((x(nc)-1).*VectorDimension)+nv);
   end
end
   k=0;
%==============Loop=================%
while VTD>RejectValue
%==========Find CodeBook=========%
for nf=1:NumberOfFrame
  for nc=1:CodeBookNumber
     TotalDistance=0;
     Distance=0;
     for nv=1:VectorDimension
       Distance=(TrainingSetFile(((nf-1).*VectorDimension)+nv)-...
         OldCentroid(((nc-1).*VectorDimension)+nv)).^2;
       TotalDistance=TotalDistance+Distance;
     end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    VectorDistance(nc)=TotalDistance;
  end
  [MinDistance(nf),MinIndex(nf)]=min(VectorDistance);
end
%==========Find New Centroid========%
for nc=1:CodeBookNumber
  Cnum=0;
  for nv=1:VectorDimension
    NewCentroid(((nc-1)*VectorDimension)+nv)=0;
  end
  for nf=1:NumberOfFrame
    if nc==MinIndex(nf)
      Cnum=Cnum+1;
      for nv=1:VectorDimension
        NewCentroid(((nc-1)*VectorDimension)+nv)=...
          NewCentroid(((nc-1)*VectorDimension)+nv)+...
          TrainingSetFile(((nf-1)*VectorDimension)+nv);
      end
    end
  end
  if Cnum~=0
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        NewCentroid(((nc-1)*VectorDimension)+nv)./Cnum;
    end
  else
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        OldCentroid(((nc-1)*VectorDimension)+nv);
    end
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%==========Check Distance==========%
d=NewCentroid-OldCentroid;
e=d.^2;
VTD=sum(e);%./CodeBookNumber
k=k+1;
CVTD64(k)=VTD;
%==========Copy Old CentroidTo New Centroid=======%
OldCentroid=NewCentroid;
%=================================================%
end
cb64=NewCentroid;
save codebook64 cb64 CVTD64
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Codebook 128

```
function [cb128,CVTD128]=codebook128(TrainingSetFile);
format long e;
VectorDimension=15;
CodeBookNumber=128;
RejectValue=0.0001;
VTD=2*RejectValue;
%==========Find Number Of Frame==========%
NumberOfFrame=((length(TrainingSetFile)/VectorDimension));
%==========Random Centroid From TrainingSetFile========%
x=floor(NumberOfFrame.*rand(1,CodeBookNumber))+1;
for nc=1:CodeBookNumber
  for nv=1:VectorDimension
    OldCentroid(((nc-1).*VectorDimension)+nv) = ...
      TrainingSetFile(((x(nc)-1).*VectorDimension)+nv);
  end
end
k=0;
%==========Loop==========%
while VTD>RejectValue
%==========Find CodeBook==========%
for nf=1:NumberOfFrame
  for nc=1:CodeBookNumber
    TotalDistance=0;
    Distance=0;
    for nv=1:VectorDimension
      Distance=(TrainingSetFile(((nf-1).*VectorDimension)+nv)-...
        OldCentroid(((nc-1).*VectorDimension)+nv)).^2;
      TotalDistance=TotalDistance+Distance;
    end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    VectorDistance(nc)=TotalDistance;
  end
  [MinDistance(nf),MinIndex(nf)]=min(VectorDistance);
end
%==========Find New Centroid========%
for nc=1:CodeBookNumber
  Cnum=0;
  for nv=1:VectorDimension
    NewCentroid(((nc-1)*VectorDimension)+nv)=0;
  end
  for nf=1:NumberOfFrame
    if nc==MinIndex(nf)
      Cnum=Cnum+1;
      for nv=1:VectorDimension
        NewCentroid(((nc-1)*VectorDimension)+nv)=...
          NewCentroid(((nc-1)*VectorDimension)+nv)+...
          TrainingSetFile(((nf-1)*VectorDimension)+nv);
      end
    end
  end
  if Cnum~=0
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        NewCentroid(((nc-1)*VectorDimension)+nv)./Cnum;
    end
  else
    for nv=1:VectorDimension
      NewCentroid(((nc-1)*VectorDimension)+nv)=...
        OldCentroid(((nc-1)*VectorDimension)+nv);
    end
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%============Check Distance=========%
d=NewCentroid-OldCentroid;
e=d.^2;
VTD=sum(e);%./CodeBookNumber
k=k+1;
CVTD128(k)=VTD;
%========Copy Old CentroidTo New Centroid=======%
OldCentroid=NewCentroid;
%=========================================%
end
cb128=NewCentroid;
save codebook128 cb128 CVTD128
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hmm32

```
function [A,B]=hmm32(Ob);
format long e
%Initial model parameter
N=6;
M=32;
V=1:M;
[K,T]=size(Ob);
%defind TT
TT=[1 0 0 0 0 0];
%Defind Matrix A
A=[0.5 0.5 0 0 0 0
  0 0.5 0.5 0 0 0
  0 0 0.5 0.5 0 0
  0 0 0 0.5 0.5 0
  0 0 0 0 0.5 0.5
  0 0 0 0 0 1];
%Defind matrix B
for i=1:N
   for k=1:M
      B(i,k)=1/M;
   end
end
%Start HMM
%Iterative for new A and B
for round=1:50
   for n=1:K
      for t=1:T
         for i=1:N
            for k=1:M
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        if Ob(n,t)==V(k);
          b(n,i,t)=B(i,k);
        end
      end
    end
end
%The forward producc
for i=1:N
  alpha(n,1,i)=TT(i)*b(n,i,1);
end
for t=1:T-1
  for j=1:N
    sumaa=alpha(n,t,1)*A(1,j);
    for i=2:N
      sumaa=sumaa+alpha(n,t,i)*A(i,j);
    end
    alpha(n,t+1,j)=sumaa*b(n,j,t+1);
  end
end
P(n)=sum(alpha(n,T,1:N));
%The backward proceduce
for i=1:N
  beta(n,T,i)=1;
end
for t=T-1:-1:1
  for i=1:N
    sumab=A(i,1)*b(n,1,t+1)*beta(n,t+1,1);
    for j=2:N
      sumab=sumab+A(i,j)*b(n,j,t+1)*beta(n,t+1,j);
    end
    beta(n,t,i)=sumab;
  end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end
%Parameter estimation
%Find matrix A
for i=1:N
    for j=1:N
        nom=0;
        for n=1:K
            kvt=0;
            for t=1:T-1
                kvt=kvt+alpha(n,t,i)*A(i,j)*b(n,j,t+1)*...
                    beta(n,t+1,j);
            end
            nom=nom+kvt/P(n);
        end
        denom=0;
        for n=1:K
            kvt=0;
            for t=1:T-1
                for k=1:N
                    kvt=kvt+alpha(n,t,i)*A(i,k)*b(n,k,t+1)*...
                        beta(n,t+1,k);
                end
            end
            denom=denom+kvt/P(n);
        end
        A(i,j)=nom/denom;
    end
end
%Find Matrix B
for i=1:N
    for k=1:M
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        nom=0;
        for n=1:K
          kvt=0;
          for t=1:T
            if Ob(n,t)==V(k)
              ptk=1;
            else
              ptk=0;
            end
            kvt=kvt+alpha(n,t,i)*beta(n,t,i)*ptk;
          end
          nom=nom+kvt/P(n);
        end
        denom=0;
        for n=1:K
          kvt=0;
          for t=1:T
            kvt=kvt+alpha(n,t,i)*beta(n,t,i);
          end
          denom=denom+kvt/P(n);
        end
        B(i,k)=nom/denom;
      end
    end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hmm64

```
function [A,B]=hmm64(Ob);
format long e
%Initial model parameter
N=6;
M=64;
V=1:M;
[K,T]=size(Ob);
%defind TT
TT=[1 0 0 0 0 0];
%Defind Matrix A
A=[0.5 0.5 0 0 0 0
  0 0.5 0.5 0 0 0
  0 0 0.5 0.5 0 0
  0 0 0 0.5 0.5 0
  0 0 0 0 0.5 0.5
  0 0 0 0 0 1];
%Defind matrix B
for i=1:N
   for k=1:M
      B(i,k)=1/M;
   end
end
%Start HMM
%Iterative for new A and B
for round=1:50
   for n=1:K
      for t=1:T
         for i=1:N
            for k=1:M
               if Ob(n,t)==V(k);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            b(n,i,t)=B(i,k);
          end
        end
      end
    end
%The forward produce
for i=1:N
  alpha(n,1,i)=TT(i)*b(n,i,1);
end
for t=1:T-1
  for j=1:N
    sumaa=alpha(n,t,1)*A(1,j);
    for i=2:N
      sumaa=sumaa+alpha(n,t,i)*A(i,j);
    end
    alpha(n,t+1,j)=sumaa*b(n,j,t+1);
  end
end
P(n)=sum(alpha(n,T,1:N));
%The backward proceduce
for i=1:N
  beta(n,T,i)=1;
end
for t=T-1:-1:1
  for i=1:N
    sumab=A(i,1)*b(n,1,t+1)*beta(n,t+1,1);
    for j=2:N
      sumab=sumab+A(i,j)*b(n,j,t+1)*beta(n,t+1,j);
    end
    beta(n,t,i)=sumab;
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end
%Parameter estimation
%Find matrix A
for i=1:N
   for j=1:N
      nom=0;
      for n=1:K
         kvt=0;
         for t=1:T-1
            kvt=kvt+alpha(n,t,i)*A(i,j)*b(n,j,t+1)*...
               beta(n,t+1,j);
         end
         nom=nom+kvt/P(n);
      end
      denom=0;
      for n=1:K
         kvt=0;
         for t=1:T-1
            for k=1:N
               kvt=kvt+alpha(n,t,i)*A(i,k)*b(n,k,t+1)*...
                  beta(n,t+1,k);
            end
         end
         denom=denom+kvt/P(n);
      end
      A(i,j)=nom/denom;
   end
end
%Find Matrix B
for i=1:N
   for k=1:M
      nom=0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        for n=1:K
            kvt=0;
            for t=1:T
                if Ob(n,t)==V(k)
                    ptk=1;
                else
                    ptk=0;
                end
                kvt=kvt+alpha(n,t,i)*beta(n,t,i)*ptk;
            end
            nom=nom+kvt/P(n);
        end
        denom=0;
        for n=1:K
            kvt=0;
            for t=1:T
                kvt=kvt+alpha(n,t,i)*beta(n,t,i);
            end
            denom=denom+kvt/P(n);
        end
        B(i,k)=nom/denom;
    end
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hmm128

```
function [A,B]=hmm128(Ob);
format long e
%Initial model parameter
N=6;
M=128;
V=1:M;
[K,T]=size(Ob);
%defind TT
TT=[1 0 0 0 0 0];
%Defind Matrix A
A=[0.5 0.5 0 0 0 0
  0 0.5 0.5 0 0 0
  0 0 0.5 0.5 0 0
  0 0 0 0.5 0.5 0
  0 0 0 0 0.5 0.5
  0 0 0 0 0 1];
%Defind matrix B
for i=1:N
   for k=1:M
      B(i,k)=1/M;
   end
end
%Start HMM
%Iterative for new A and B
for round=1:50
   for n=1:K
      for t=1:T
         for i=1:N
            for k=1:M
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        if Ob(n,t)==V(k);
          b(n,i,t)=B(i,k);
        end
      end
    end
end
%The forward produce
for i=1:N
  alpha(n,1,i)=TT(i)*b(n,i,1);
end
for t=1:T-1
  for j=1:N
    sumaa=alpha(n,t,1)*A(1,j);
    for i=2:N
      sumaa=sumaa+alpha(n,t,i)*A(i,j);
    end
    alpha(n,t+1,j)=sumaa*b(n,j,t+1);
  end
end
P(n)=sum(alpha(n,T,1:N));
%The backward proceduce
for i=1:N
  beta(n,T,i)=1;
end
for t=T-1:-1:1
  for i=1:N
    sumab=A(i,1)*b(n,1,t+1)*beta(n,t+1,1);
    for j=2:N
      sumab=sumab+A(i,j)*b(n,j,t+1)*beta(n,t+1,j);
    end
    beta(n,t,i)=sumab;
  end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end
%Parameter estimation
%Find matrix A
for i=1:N
    for j=1:N
        nom=0;
        for n=1:K
            kvt=0;
            for t=1:T-1
                kvt=kvt+alpha(n,t,i)*A(i,j)*b(n,j,t+1)*...
                    beta(n,t+1,j);
            end
            nom=nom+kvt/P(n);
        end
        denom=0;
        for n=1:K
            kvt=0;
            for t=1:T-1
                for k=1:N
                    kvt=kvt+alpha(n,t,i)*A(i,k)*b(n,k,t+1)*...
                        beta(n,t+1,k);
                end
            end
            denom=denom+kvt/P(n);
        end
        A(i,j)=nom/denom;
    end
end
%Find Matrix B
for i=1:N
    for k=1:M
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        nom=0;
        for n=1:K
          kvt=0;
          for t=1:T
            if Ob(n,t)==V(k)
              ptk=1;
            else
              ptk=0;
            end
            kvt=kvt+alpha(n,t,i)*beta(n,t,i)*ptk;
          end
          nom=nom+kvt/P(n);
        end
        denom=0;
        for n=1:K
          kvt=0;
          for t=1:T
            kvt=kvt+alpha(n,t,i)*beta(n,t,i);
          end
          denom=denom+kvt/P(n);
        end
        B(i,k)=nom/denom;
      end
    end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Speechreg32

```
function output=speechreg32(filename);
format long e
load codebook32
[E,a1,b1,ref,s]=detect1(filename);
[S,z,a2,b2,Zmax]=zrocrs1(filename,a1,b1);
z=normalize(S,4000);
w=LPCprocess(z,12,18);
Ob=observation(w,cb32,32);
%Initial model parameter
N=6;
M=32;
V=1:M;
T=length(Ob);
%Defind TT
TT(1)=1;
for i=2:N
   TT(i)=0;
end
for num=0:9
   switch num
      case 0
         load model32_00;
      case 1
         load model32_11;
      case 2
         load model32_22;
      case 3
         load model32_33;
      case 4
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
      load model32_44;
   case 5
      load model32_55;
   case 6
      load model32_66;
   case 7
      load model32_77;
   case 8
      load model32_88;
   case 9
      load model32_99;
end
for t=1:T
   for i=1:N
      for k=1:M
         if Ob(t)==V(k)
            b(i,t)=B(i,k);
         end
      end
   end
end
% The viterbi algorithm
for i=1:N
   delta(1,i)=TT(i)*b(i,1);
   phi(1,i)=0;
end
for t=2:T
   for j=1:N
      damax=delta(t-1,1)*A(1,j)*b(j,t);
      dastatemax=1;
      for i=2:N
         if delta(t-1,i)*A(i,j)*b(j,t)>=damax
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            damax=delta(t-1,i)*A(i,j)*b(j,t);
            dastatemax=i;
          end
        end
        delta(t,j)=damax;
        phi(t,j)=dastatemax;
      end
    end
    deltamax=delta(T,1);
    deltastatemax=1;
    for i=2:N
      if delta(T,i)>=deltamax
        deltastatemax=i;
      end
    end
    qsequence(T)=deltastatemax;
    for t=T-1:-1:1
      qsequence(t)=phi(t+1,(qsequence(t+1)));
    end
    %Total probability of this path
    pathprob(num+1)=TT(qsequence(1))*b(qsequence(1),1);
    for t=2:T
      pathprob(num+1)=pathprob(num+1)*A(qsequence(t-1),...
        qsequence(t))*b(qsequence(t),t);
    end
  end
  probmax=pathprob(1);
  numprobmax=0;
  for i=1:9
    if pathprob(i+1)>=probmax
      probmax=pathprob(i+1);
      numprobmax=i;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end
output=numprobmax;
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Speechreg64

```
function output=speechreg64(filename);
format long e
load codebook64
[E,a1,b1,ref,s]=detect1(filename);
[S,z,a2,b2,Zmax]=zrocrs1(filename,a1,b1);
z=normalize(S,4000);
w=LPCprocess(z,12,18);
Ob=observation(w,cb64,64);
%Initial model parameter
N=6;
M=32;
V=1:M;
T=length(Ob);
%Defind TT
TT(1)=1;
for i=2:N
   TT(i)=0;
end
for num=0:9
   switch num
      case 0
         load model64_00;
      case 1
         load model64_11;
      case 2
         load model64_22;
      case 3
         load model64_33;
      case 4
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
      load model64_44;
    case 5
      load model64_55;
    case 6
      load model64_66;
    case 7
      load model64_77;
    case 8
      load model64_88;
    case 9
      load model64_99;
end
for t=1:T
   for i=1:N
      for k=1:M
         if Ob(t)==V(k)
            b(i,t)=B(i,k);
         end
      end
   end
end
% The viterbi algorithm
for i=1:N
   delta(1,i)=TT(i)*b(i,1);
   phi(1,i)=0;
end
for t=2:T
   for j=1:N
      damax=delta(t-1,1)*A(1,j)*b(j,t);
      dastatemax=1;
      for i=2:N
         if delta(t-1,i)*A(i,j)*b(j,t)>=damax
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            damax=delta(t-1,i)*A(i,j)*b(j,t);
            dastatemax=i;
          end
        end
        delta(t,j)=damax;
        phi(t,j)=dastatemax;
      end
    end
    deltamax=delta(T,1);
    deltastatemax=1;
    for i=2:N
      if delta(T,i)>=deltamax
        deltastatemax=i;
      end
    end
    qsequence(T)=deltastatemax;
    for t=T-1:-1:1
      qsequence(t)=phi(t+1,(qsequence(t+1)));
    end
    %Total probability of this path
    pathprob(num+1)=TT(qsequence(1))*b(qsequence(1),1);
    for t=2:T
      pathprob(num+1)=pathprob(num+1)*A(qsequence(t-1),...
        qsequence(t))*b(qsequence(t),t);
    end
  end
  probmax=pathprob(1);
  numprobmax=0;
  for i=1:9
    if pathprob(i+1)>=probmax
      probmax=pathprob(i+1);
      numprobmax=i;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end
output=numprobmax;
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Speechreg128

```
function output=speechreg128(filename);
format long e
load codebook128
[E,a1,b1,ref,s]=detect1(filename);
[S,z,a2,b2,Zmax]=zrocrs1(filename,a1,b1);
z=normalize(S,4000);
w=LPCprocess(z,12,18);
Ob=observation(w,cb128,128);
%Initial model parameter
N=6;
M=32;
V=1:M;
T=length(Ob);
%Defind TT
TT(1)=1;
for i=2:N
   TT(i)=0;
end
for num=0:9
   switch num
      case 0
         load model128_00;
      case 1
         load model128_11;
      case 2
         load model128_22;
      case 3
         load model128_33;
      case 4
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    load model128_44;
  case 5
    load model128_55;
  case 6
    load model128_66;
  case 7
    load model128_77;
  case 8
    load model128_88;
  case 9
    load model128_99;
end
for t=1:T
  for i=1:N
    for k=1:M
      if Ob(t)==V(k)
        b(i,t)=B(i,k);
      end
    end
  end
end
% The viterbi algorithm
for i=1:N
  delta(1,i)=TT(i)*b(i,1);
  phi(1,i)=0;
end
for t=2:T
  for j=1:N
    damax=delta(t-1,1)*A(1,j)*b(j,t);
    dastatemax=1;
    for i=2:N
      if delta(t-1,i)*A(i,j)*b(j,t)>=damax
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            damax=delta(t-1,i)*A(i,j)*b(j,t);
            dastatemax=i;
          end
        end
        delta(t,j)=damax;
        phi(t,j)=dastatemax;
      end
    end
    deltamax=delta(T,1);
    deltastatemax=1;
    for i=2:N
      if delta(T,i)>=deltamax
        deltastatemax=i;
      end
    end
    qsequence(T)=deltastatemax;
    for t=T-1:-1:1
      qsequence(t)=phi(t+1,(qsequence(t+1)));
    end
    %Total probability of this path
    pathprob(num+1)=TT(qsequence(1))*b(qsequence(1),1);
    for t=2:T
      pathprob(num+1)=pathprob(num+1)*A(qsequence(t-1),...
        qsequence(t))*b(qsequence(t),t);
    end
end
probmax=pathprob(1);
numprobmax=0;
for i=1:9
  if pathprob(i+1)>=probmax
    probmax=pathprob(i+1);
    numprobmax=i;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    end
end
output=numprobmax;
```


เอกสารนี้เป็นเอกสารที่สงวนไว้ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้