การรู้จำเสียงพูด

SPEECH RECOGNITION

โดย

นางสาวชวดี อิศรปรีดา

นายชัชชัย ขมสวัสดิ์

นางสาววรจิตต์ วีระจิตต์

นายวินิจ วงศ์ตั้งตน

วัน เดือน ปี 14 ส.ค. 2540

เลขทะเบียน 037217

เลขเรียกหนังสือ T 38310 ช 244 ก

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิศวกรรมโทรคมนาคม

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2538

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

037217

การรู้จำเสียงพูด
SPEECH RECOGNITION

โดย

นางสาวชวดี อิศรปรีดา 35104098

นายชัชชัย ขมสวัสดิ์ 35104099

นางสาววรจิตต์ วีระจิตต์ 35104360

นายวินิจ วงศ์ตั้งตน 35104387

อาจารย์ที่ปรึกษา

ดร. ไกรสิน ส่งวัฒนา

ปริญญานิพนธ์สำหรับปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2538

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2538

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การรู้จำเสียงพูด

SPEECH RECOGNITION

ผู้จัดทำ

1. นางสาว ชวดี อิศรปรีดา รหัสประจำตัว 35104098
2. นาย ชัชชัย ขมสวัสดิ์ รหัสประจำตัว 35104099
3. นางสาว วรจิตต์ วีระจิตต์ รหัสประจำตัว 35104360
4. นาย วินิจ วงศ์ตั้งตน รหัสประจำตัว 35104387

ไกรสิน สงวัฒนา ..... อาจารย์ที่ปรึกษา

(ดร. ไกรสิน สงวัฒนา)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การรู้จำเสียงพูด

Speech Recognition

โดย

1. นางสาว ชวดี อิศรปรีดา 35104098
2. นาย ชัชชัย ขมสวัสดิ์ 35104099
3. นางสาว วรจิตต์ วีระจิตต์ 35104360
4. นาย วินิจ วงศ์ตั้งตน 35104387

อาจารย์ที่ปรึกษา ดร. ไกรสิน ส่งวัฒนา

บทคัดย่อ

โครงงานนี้ได้ทำขึ้นโดยมีจุดประสงค์ในการรู้จำเสียงพูด สามารถแบ่งออกได้เป็น 2 ขั้นตอนใหญ่ ๆ ดังนี้

1. ขั้นตอนการเรียนรู้ (Learning) เป็นขั้นตอนที่พยายามเรียนรู้เสียงพูด โดยดึงเอาคุณลักษณะเฉพาะของเสียงพูดคำนั้น ๆ ออกมา ซึ่งทำโดยใช้วิธีการประมาณเชิงเส้น(Linear Predictive Coding : LPC) แล้วนำเอาพารามิเตอร์ที่ได้นี้ไปปรับปรุงให้เกิดความเหมาะสมแก่ขั้นตอนการรับรู้ด้วย เวคเตอร์ ควอนไตซ์เซชัน (Vector Quantization)

2. ขั้นตอนการรับรู้ (Recognition) เป็นขั้นตอนที่นำเอารูปแบบของคำที่ต้องการทดสอบว่าเป็นคำใดมาเปรียบเทียบกับรูปแบบอ้างอิงทั้งหมด แล้วเลือกเอาแบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

ABSTRACT

The purpose of this project is to recognize speech . The process is devided into 2 parts;

1. Learning : In this part the certain significant features form the speech by using Linear Predictive Coding (LPC) have been extracted . The parameters from LPC analysis are obtained and can be improved by reducing the amount of information by vector quantization technique.

2. Recognition : Compare the pattern of the unknown word and recognize to each pattern of reference words . Then select the word whose pattern is closest to the word that it can recognize. In this part , the Hidden Makov Models (HMM) for statistical recognition algorithm is used.

เอกสารนี้สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทที่ 1 บทนำ	1
บทที่ 2 ระบบการพูดและเสียงพูดของมนุษย์	4
บทที่ 3 การคำนวณค่าสัมประสิทธิ์ของ LPC และอัตราการขยาย	8
บทที่ 4 เวคเตอร์ควอนไตซ์เซชัน	17
บทที่ 5 แบบจำลองมาร์คอฟ	25
บทที่ 6 ขั้นตอนการทดลอง	41
บทที่ 7 ผลการทดลอง	46
บทที่ 8 สรุปผลการทดลอง	68
ภาคผนวก ก พารามิเตอร์ที่ใช้ในการทดลอง	72
ภาคผนวก ข โปรแกรมที่ใช้ในการคำนวณ	80
กิตติกรรมประกาศ	121
เอกสารอ้างอิง	122

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูปที่		หน้า
1.1	แสดงบล็อกไดอะแกรมการรู้จำเสียงพูด	2
2.1	ภาพตัดขวางแสดงอวัยวะในระบบการพูดของมนุษย์	4
2.2	รูปกล่องเสียง	5
2.3	แผนภาพระบบเสียงพูดมนุษย์	6
2.4	รูปสเปคตรัมของเสียง /a/ ที่ได้จากสัมประสิทธิ์ LPC	7
2.5	แสดงบล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเริ่มต้น	7
3.1	แสดงขั้นตอนการเตรียมสัญญาณในการวิเคราะห์	9
3.2	แสดงขนาดสเปคตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส	10
3.3	แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์	10
3.4	แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์	11
3.5	แสดงวินโดว์สี่เหลี่ยม	12
3.6	แสดงวินโดว์แบบแฮมมิ่ง	12
3.7	บล็อคไดอะแกรมแสดงแบบจำลองการสร้างเสียงพูดอย่างง่าย	13
4.1	แสดงการกระจายเฟรมของเสียงพูด	18
4.2	การรวมกลุ่มของเฟรมของเสียงเพื่อนำไปสร้างโค้ดบุค	18
4.3	แสดงบล็อคไดอะแกรมของเวคเตอร์ควอนไตซ์เซชัน	19
4.4	แสดงเวคเตอร์ควอนไตซ์เซอร์ที่ใช้ในระบบการรับรู้เสียงพูด	24
5.1	ตัวอย่างแบบจำลองมาร์คอฟ 5 สเตท	25
5.2	แสดงแบบจำลองต่าง ๆ ของ HMM	27
5.3	กระบวนการไปข้างหน้า	31
5.4	กระบวนการถอยหลัง	32
5.5	แสดงลำดับการคำนวณเหตุการณ์ร่วม	34
5.6	บล็อคไดอะแกรมของการรู้จำคำโดดด้วยแบบจำลองของมาร์คอฟ	40
6.1	แสดงขั้นตอนการทดลองในส่วนของเวคเตอร์ควอนไตซ์เซชั่น	42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ หน้า

7.1 กราฟเปรียบเทียบสัมประสิทธิ์เสียงของผู้ชาย 48
7.2 กราฟเปรียบเทียบสัมประสิทธิ์เสียงของผู้หญิง 49
7.3 กราฟเปรียบเทียบอัตราการขยายของเสียงเดียวกัน 50
7.4 กราฟเปรียบเทียบสัมประสิทธิ์ LPC 51
7.5 กราฟเปรียบเทียบสัมประสิทธิ์เซปสตรัม 52
7.6 กราฟเปรียบเทียบพารามิเตอร์ที่ได้เวทค่าแล้ว 53
7.7 แสดงเวคเตอร์พารามิเตอร์ของเฟรมต่าง ๆ กับโค้ดบุคที่คำนวณได้ใน 2 มิติ 54
8.1 แสดงลักษณะค่าความน่าจะเป็นตามพารามิเตอร์ค่าต่างๆ 70
ก.1 สเปคตรัมของเสียงเปรียบเทียบกับสเปคตรัมที่ได้จากสัมประสิทธิ์ LPC 73
ก.2 สัญญาณความคลาดเคลื่อน 74
ก.3 เปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกับขนาดโค้ดบุค 75
ก.4 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มต้นแบบ 76
ก.5 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มทดสอบ 77
ก.6 แสดงอัตราการผิดพลาดเมื่อใช้จำนวนสเตท (N) ค่าต่างๆ 78
ก.7 แสดงค่าความผิดพลาดในกรณีที่ใช้ $\in$ ค่าต่าง ๆ 79
ข.1 โฟลวชาร์ทแสดงการทำงานของโปรแกรม LPC 83
ข.2 โฟลวชาร์ทแสดงการทำงานของโปรแกรม CODEBOOK 92
ข.3 โฟลวชาร์ทแสดงการทำงานของโปรแกรม COMPAIR 98
ข.4 โฟลวชาร์ทแสดงอัลกอลิทึมการรีเอสติเมสพารามิเตอร์ของแบบจำลอง 103
ข.5 โฟลวชาร์ทแสดงการทำงานของโปรแกรม HMM 105

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
7.1	แสดงการเปรียบเทียบสัมประสิทธิ์ LPC จากโปรแกรมทั้งสอง	46
7.2	แสดงการเปรียบเทียบสัมประสิทธิ์ LPC จากโปรแกรมทั้งสอง	47
7.3	แสดงผลการทดสอบของเสียงจากผู้พูดคนเดิม	55
7.4	แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ	56
7.5	แสดงผลการทดสอบของเสียงจากผู้พูดเดิมทั้ง 2 คน	57
7.6	แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ	58
7.7	แสดงผลการทดสอบของเสียงจากกลุ่มผู้พูดเดิม	59
7.8	แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ	60
7.9	แสดงผลการทดสอบของเสียงจากผู้พูดเดิมทั้ง 10 คน	61
7.10	แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ	62
7.11	แสดงผลการทดสอบของเสียงจากผู้พูดคนเดิม	63
7.12	แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ	64
7.13	แสดงการเปรียบเทียบความถูกต้องในการทดสอบเสียง	66
ก.1	เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มต้นแบบ	76
ก.2	เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มทดสอบ	76

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

ในปัจจุบันนี้เทคโนโลยีทางด้านคอมพิวเตอร์ ได้เข้ามามีบทบาทในชีวิตประจำวันของมนุษย์เป็นอย่างมาก ซึ่งโดยปกติการติดต่อระหว่างคอมพิวเตอร์กับมนุษย์จะทำได้โดย เครื่องคอมพิวเตอร์จะรับคำสั่งเข้าทางแป้นพิมพ์ และจะแสดงผลออกทางจอภาพหรือเครื่องพิมพ์ เพราะฉะนั้นการติดต่อกับคอมพิวเตอร์อย่างมีประสิทธิภาพ ต้องอาศัยความชำนาญและการฝึกฝน การที่จะเพิ่มความสามารถของมนุษย์ในการติดต่อกับเครื่อง (Man-Machine Communication) จึงเป็นที่มาของการพัฒนา การติดต่อกับคอมพิวเตอร์โดยใช้เสียงพูดแทน โดยเฉพาะในยุคโลกาภิวัตน์นี้ การรู้จำเสียงพูด จึงเป็นบทบาททางเทคโนโลยีของคอมพิวเตอร์อย่างหนึ่ง ที่สามารถผนวกเข้ากับอุปกรณ์ทางโทรคมนาคม เช่น ATM โทรศัพท์มือถือ ได้อย่างมีประสิทธิภาพ

การประมาณเชิงเส้นสามารถนำไปประยุกต์ใช้กับการสื่อสารทางเสียงได้หลายด้าน เช่น การเก็บข้อมูลเสียง (voice mail) หรือการจดจำเสียงพูด (speech recognition) เนื่องจากวิธีนี้เป็นวิธีที่สามารถบีบอัด (compress) สัญญาณเสียงพูด ซึ่งมีมานานแล้ว แต่ไม่เป็นที่แพร่หลายเนื่องจากต้องอาศัยการคำนวณที่ยุ่งยากและซับซ้อน จวบจนปัจจุบันพัฒนาการทางด้านสารกึ่งตัวนำ ได้ก้าวล้ำจนสามารถสร้างเครื่องมือคำนวณที่มีประสิทธิภาพ และมีความเร็วสูง อีกทั้งยังมีการพัฒนาทางด้านการประมวลสัญญาณเชิงเลข (digital signal processing) ทำให้การประมาณสัญญาณเชิงเส้นมีความสำคัญมากยิ่งขึ้น

วัตถุประสงค์ในการทำปริญญานิพนธ์เรื่องการรู้จำเสียงพูด มีดังนี้

1. ศึกษาขั้นตอนการประมาณเชิงเส้นของสัญญาณเสียงพูด โดยวิธี ออโตคอร์รีเลชัน เพื่อนำมาใช้ในการวิเคราะห์เสียงพูดของมนุษย์

2. วิเคราะห์หาค่าพารามิเตอร์ LPC และ เซปสตรัม (Cepstrum) ของเสียง

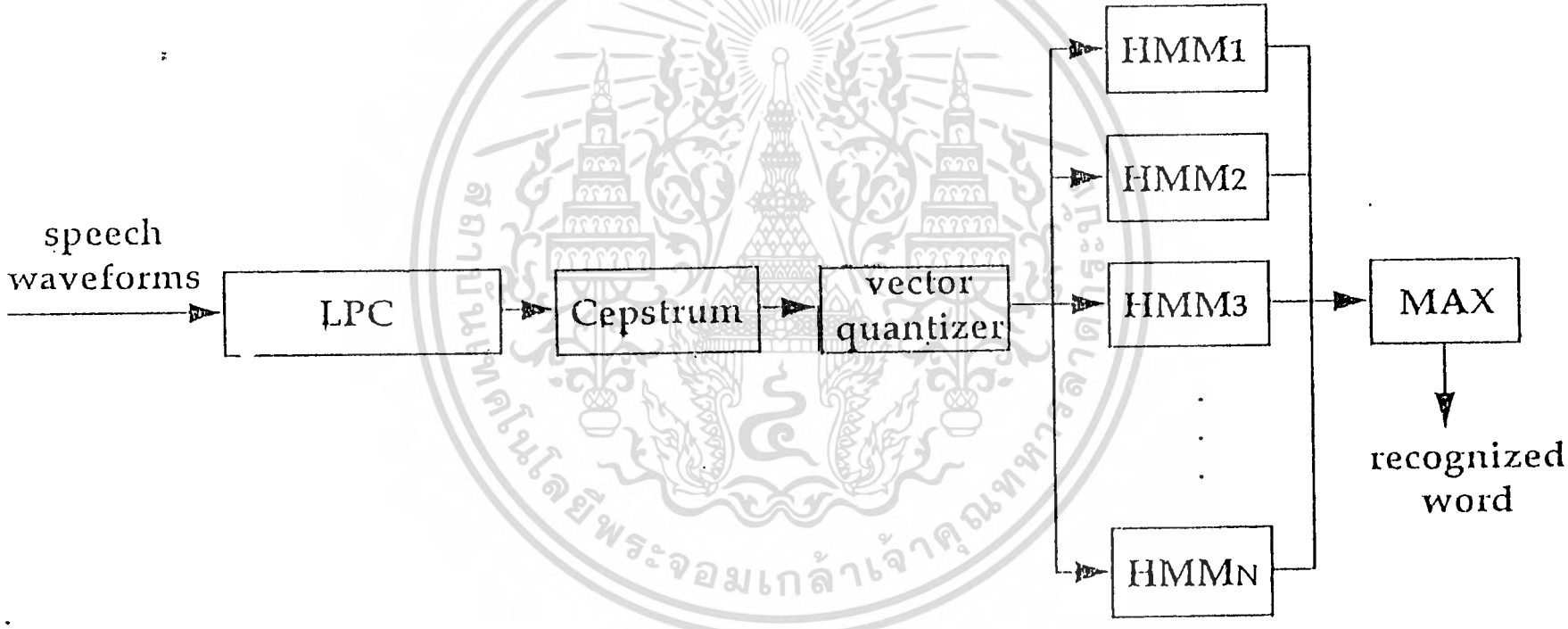
3. สร้างแบบอ้างอิง (code book) โดยใช้การจัดระดับเวคเตอร์ (Vector Quantization) ของพารามิเตอร์ที่ได้จากการประมาณเชิงเส้น

4. สร้างแบบจำลองของมาร์คอฟ (Hidden Markov Model) เพื่อใช้ในการรู้จำเสียงพูด

5. เพื่อเป็นพื้นฐานในการนำไปประยุกต์ใช้งานเพิ่มเติมต่อไป

ปริญญานิพนธ์นี้ มีจุดประสงค์หลัก คือ การรู้ จำเสียงพูด 0 - 9 ซึ่งเป็นการเริ่มต้นศึกษาหลักการประมาณเชิงเส้น ดังนั้น จึงเป็นเพียงการเขียนโปรแกรมเพื่อจำลองส่วนต่าง ๆ ในการทดลองตามบล็อกไดอะแกรม ดังรูปที่ 1.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 1.1 แสดงบล็อกไดอะแกรมการรู้จำเสียงพูด

ขั้นตอนของปริญญานิพนธ์ แบ่งออกเป็น 8 บท ดังนี้

บทที่ 2 จะกล่าวถึงระบบการพูดของมนุษย์ ในด้านองค์ประกอบสรีระ และหน้าที่ของอวัยวะต่าง ๆ ของมนุษย์ในการเปล่งเสียงพูด

บทที่ 3 จะเป็นขั้นตอนในการคำนวณค่าสัมประสิทธ์ต่างๆของเสียง และอัตราการขยาย โดยใช้ทฤษฎีพื้นฐานของการประมาณเชิงเส้น ซึ่งอาศัยหลักการ กำลังสองน้อยที่สุด (least square estimate) จนกระทั่งได้สูตรในการคำนวณค่าสัมประสิทธ์ต่าง ๆ

บทที่ 4 จะกล่าวถึงการจัดระดับเวคเตอร์ของสัมประสิทธิ์ที่ได้จากการประมาณเชิงเส้น ของเสียง ซึ่งจะได้เวคเตอร์เป็นกลุ่มๆ เพื่อจะนำไปใช้ในส่วนของแบบจำลองของมาร์คอฟต่อไป

บทที่ 5 บทนี้จะกล่าวถึงหลักการของแบบจำลองของมาร์คอฟ ซึ่งเป็นวิธีการทางสถิติ ที่ใช้ในขั้นตอนสุดท้ายการรู้จักเสียงพูด โดยจะทำการสร้างแบบจำลองของเสียง และเปรียบเทียบกับเสียงที่ต้องการรู้จัก

บทที่ 6 เป็นขั้นตอนในการทดลอง

บทที่ 7 เป็นการรวบรวมผลการทดลองที่ได้ทำการทดลองมาทั้งหมด

บทที่ 8 บทนี้จะสรุปเกี่ยวกับการทดลองทั้งหมดที่ได้ทำมา พร้อมทั้งข้อเสนอแนะสำหรับผู้ที่จะทำการวิจัย และพัฒนา การรู้จำเสียงพูดต่อไป

ในตอนท้ายของปริญญานิพนธ์นี้ ได้รวบรวมโปรแกรมต่าง ๆ ที่ใช้ ซึ่งเขียนโดยโปรแกรมที่ชื่อว่า 'MATLAB' ในการคำนวณ

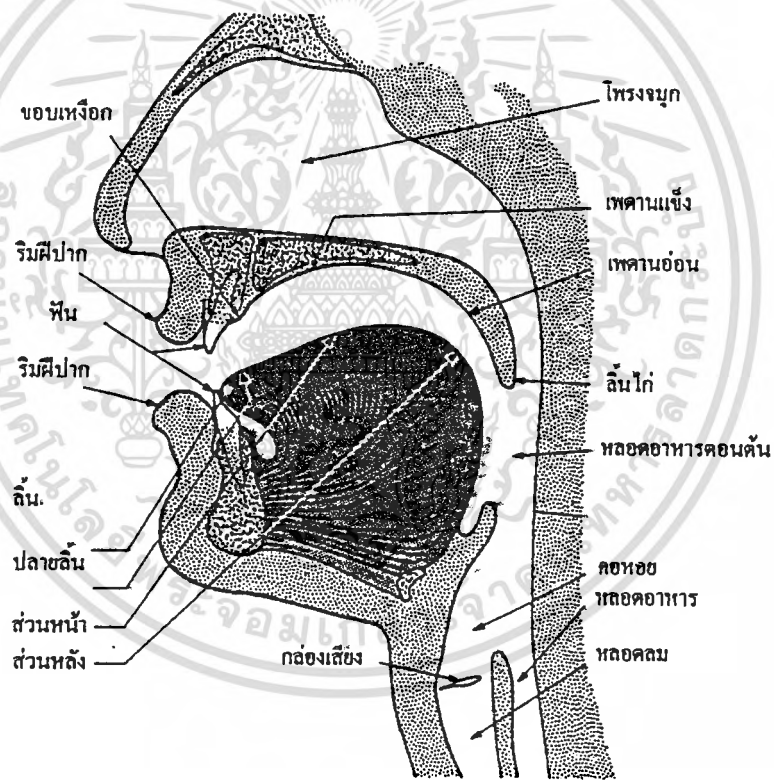
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ระบบการพูดและเสียงพูดของมนุษย์

จากการศึกษาด้าน กายวิภาคศาสตร์ของมนุษย์ (human anatomy) วิชาที่ว่าด้วย เสียงของภาษา (phonetics) และศาสตร์ทางด้านเสียง (acoutics) ช่วยให้เราเข้าใจ ขั้นตอนการทำงานร่วมกัน ของอวัยวะต่างๆ ในการเปล่งเสียงพูด ตลอดจนลักษณะทางกายภาพของเสียงพูด เพื่อนำมาวิเคราะห์ และสร้างแบบจำลองเลียนแบบเสียงพูดของมนุษย์

2.1 กายวิภาคของระบบการเปล่งเสียงของมนุษย์

การทำให้เกิดเสียงเป็นหน้าที่หนึ่งของระบบหายใจ การออกเสียงหรือการพูดของมนุษย์แต่ละครั้งจะต้องมีการทำงานร่วมกันของอวัยวะต่าง ๆ ของร่างกาย ดังรูปที่ 2.1 อันประกอบด้วย



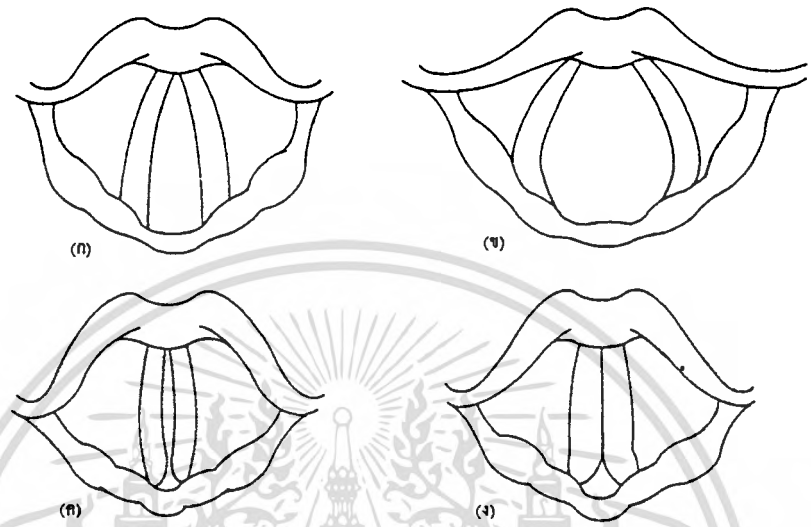
รูปที่ 2.1 ภาพตัดขวางแสดงอวัยวะในระบบการพูดของมนุษย์

2.1.1 ปอดและกระบังลม ทำหน้าที่สำคัญในการหายใจ และเป็นต้นกำเนิดการไหล ของอากาศในกระบวนการผลิตเสียง

2.1.2 หลอดลม ทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียง และเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ในการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3 กล่องเสียง เป็นอวัยวะพิเศษที่ทำหน้าที่เป็นทางเดินอากาศเวลาหายใจ และเป็นตัวผลิตพัลส์ (pulse) ของอากาศขณะเปล่งเสียง ซึ่งประกอบด้วยเส้นเสียง (vocal cords) และช่องสายเสียง (glottis) รูปร่างของกล่องเสียงและเส้นเสียงในลักษณะต่าง ๆ แสดงในรูปที่ 2.2



รูปที่ 2.2 รูปกล่องเสียงขณะ (ก) หายใจปกติ (ข) หายเข้าลึกๆ
(ค) กำลังส่งเสียง (ง) ส่งเสียงกระซิบหรือเสียงแผ่ว

2.1.4 ช่องปากและส่วนของหลอดอาหารตอนต้น อวัยวะกลุ่มนี้อยู่ต่อจากกล่องเสียงอาจเรียกว่าอวัยวะกำทอนเสียง (vocal tract) ทำหน้าที่กำทอนเสียง โดยให้กำทอนทั้งเสียงที่เกิดจากกล่องเสียงและเสียงที่เกิดภายในช่องปาก ขนาดของอวัยวะกำทอนเสียงขึ้นอยู่กับตำแหน่งของลิ้น ริมฝีปาก ขากรรไกร และเพดานอ่อน และเปลี่ยนแปลงไปตามการออกเสียง

2.1.5 โพรงจมูก เริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสอง ทำหน้าที่กำทอนเสียงร่วมกับช่องปากเมื่อมีการเปล่งเสียงที่ออกทางจมูก (nosal sounds) เช่นเสียง /ม/ ,/น/ และ /ง/ เป็นต้น

2.2 กระบวนการผลิตเสียงพูด

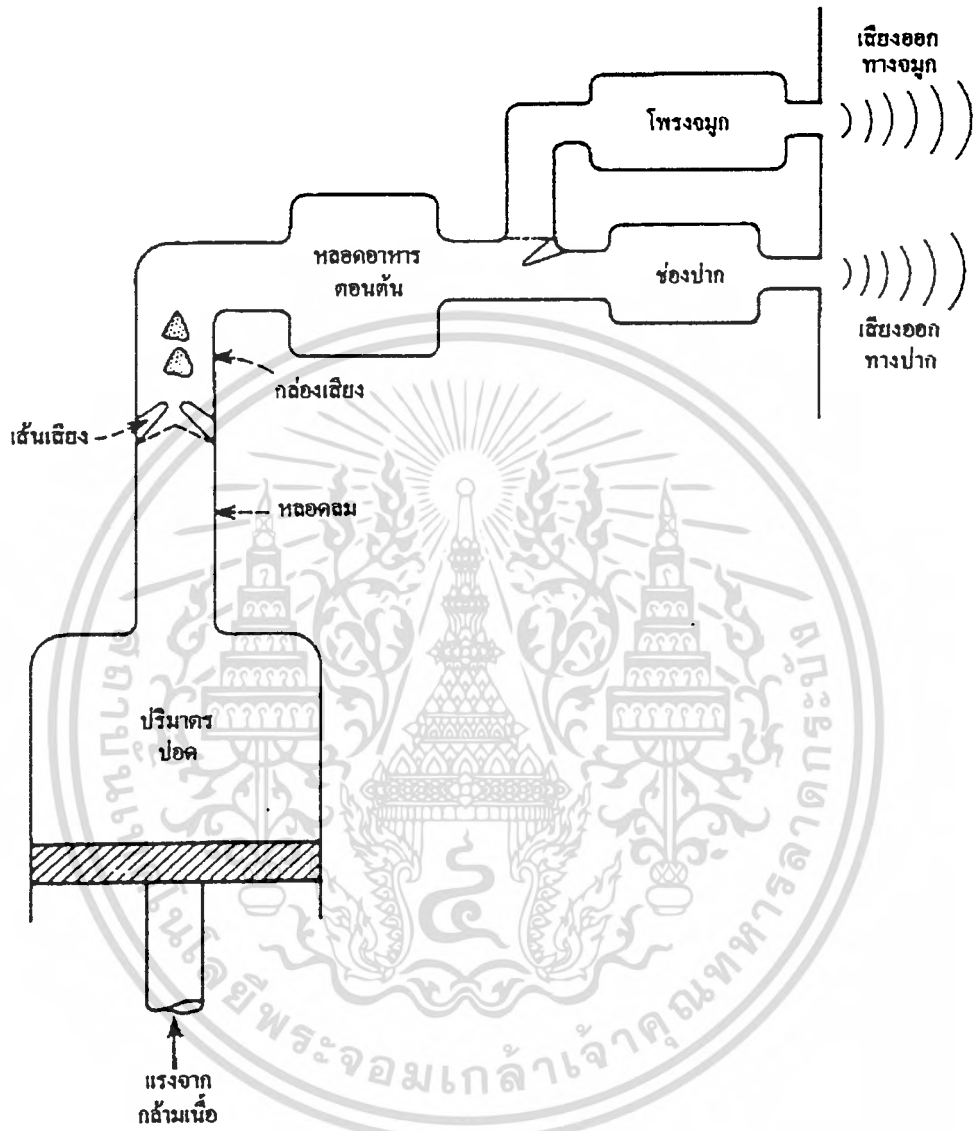
จากระบบเสียงพูด สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียง ดังรูปที่ 2.3 ซึ่งเราสามารถจำแนกกลไกการสร้างเสียงพูดของมนุษย์ได้ 3 แบบ ดังนี้

2.2.1 อากาศไหลจากปอดจะถูกมอดูเลท (modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (quasi-periodic pulse-like excitation)

2.2.2 อากาศที่ไหลจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปาก ทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (noise-like excitation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศที่มีแรงดัน พุ่งออกไปอย่างเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (transient excitation)



รูปที่ 2.3 แผนภาพระบบเสียงพูดของมนุษย์

2.3 เสียงพูดมนุษย์

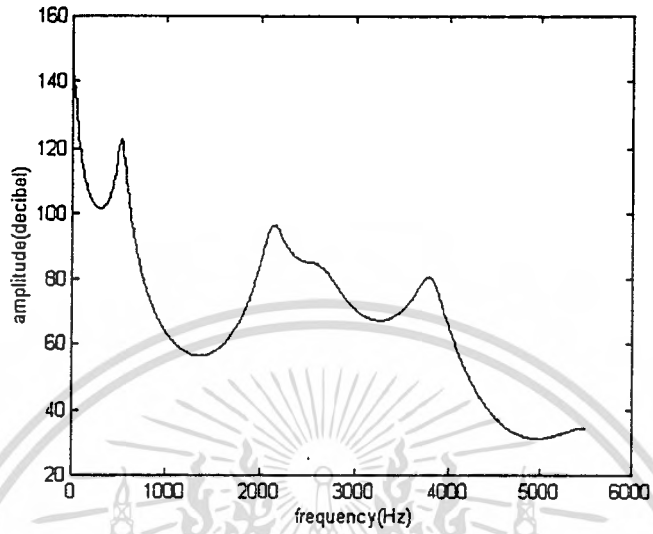
เสียงพูด เป็นคลื่นตามยาว (longitudinal wave) เกิดจากการสั่นของอนุภาคตัวกลางนั่นคืออากาศ และทิศทางการสั่นของอนุภาคจะอยู่ในทิศเดียวกันกับทิศทางของการเคลื่อนที่ คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงไปตามเวลา

2.4 ความถี่ฟอร์แมนท์ (formant frequency)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความถี่ฟอร์แมนท์ คือความถี่กำทอน (resonance frequency) ของกลุ่มอวัยวะกำทอนเสียง ความถี่ฟอร์แมนท์ของเสียงจะมีค่าเท่าใดขึ้นอยู่กับการบังคับขนาดและรูปทรงของอวัยวะส่วนนี้ อันเกิดเป็นเสียง

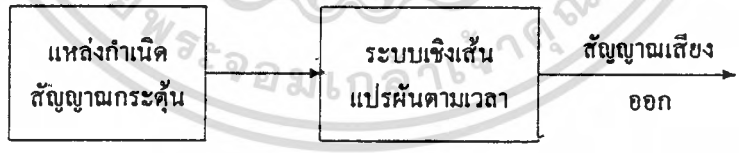
ความถี่ที่แตกต่างกันนั่นคือเสียงพูดต่าง ๆ นั่นเอง ความถี่ฟอร์แมนท์สำหรับเสียงพูดคำหนึ่งๆ อาจมีได้หลายค่า ดังรูปที่ 2.4 เป็นสเปคตรัมของเสียง /a/ ซึ่งมีความถี่ฟอร์แมนท์ที่ 1,2,3 และ 4 อยู่ ณ ตำแหน่งความถี่ ที่ 550, 2150, 2600 และ 3750 เฮิรตซ์ตามลำดับ



รูปที่ 2.4 สเปคตรัมของเสียง /a/ ที่ได้จากสัมประสิทธิ์ LPC

2.5 แบบจำลองระบบกำเนิดเสียงพูด

จากภาพรูปที่ 2.3 เราสามารถแสดงบล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเบื้องต้น ดังรูปที่ 2.5 จากรูปมีการแยกภาคแหล่งกำเนิดสัญญาณกระตุ้น ออกจากส่วนกำทอนเสียง ซึ่งแทนด้วย ระบบเชิงเส้นแปรผันตามเวลา



รูปที่ 2.5 บล็อกไดอะแกรมจำลองระบบกำเนิดเสียงเริ่มต้น

จากรูปที่ 2.5 แหล่งกำเนิดสัญญาณกระตุ้นทำหน้าที่แทนการทำงานของปอด และกล่องเสียงส่วนนี้จะผลิตขบวนพัลส์ที่มีคาบเวลาพิทช์ขณะเปล่งเสียงว้อยซ์ และให้กำเนิดเสียงซึ่งคล้ายเสียงรบกวนขณะเปล่งเสียงอันว้อยซ์

ส่วนที่สองเป็นท่อกำทอนเสียง จะแทนการทำงานของช่องปากและโพรงจมูก ทำหน้าที่เสมือนตัวกรองสัญญาณ (filter) ที่ยอมให้ความถี่ฟอร์แมนท์ผ่านได้ ซึ่งสามารถแทนด้วยระบบเชิงเส้นแปรผันตามเวลา (time-varying linear system)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การคำนวณค่าสัมประสิทธิ์ของ LPC และอัตราการขยาย

ในการหาค่าสัมประสิทธิ์ LPC (α) และอัตราการขยาย (G) เราจำเป็นต้องมีการเตรียมข้อมูลในการวิเคราะห์ก่อน ซึ่งขั้นตอนในการเตรียมสัญญาณเข้าเป็นดังรูป 3.1

การวิเคราะห์สามารถแบ่งเป็นขั้นตอนต่าง ๆ ดังนี้

3.1 การพรีเอมฟาซิส (preemphasis)

3.2 การแบ่งช่วงสัญญาณ (frame blocking)

3.3 การวินโดว์ (windowing)

3.4 การหาออโตคอร์รีเลชั่น (autocorrelation analysis)

3.5 การหาสัมประสิทธิ์ α และค่าอัตราการขยาย G

3.6 การหาสัมประสิทธิ์เซปสตรัม (cepstrum)

3.7 การเวทค่าพารามิเตอร์ (parameter weighting)

3.1 การพรีเอมฟาซิส

เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ เมื่อเทียบกับแถบความที่ถี่ปฏิบัติงาน (bandwidth) ไม่เกิน 5 กิโลเฮิรตซ์ ดังนั้น เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (signal to noise ratio : SNR) มีค่าค่อนข้างคงที่ ตลอดช่วงความถี่ที่ปฏิบัติงานนี้ เราจึงต้องมีการพรีเอมฟาซิส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้น นั่นคือ การพรีเอมฟาซิสก็คือการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (high pass filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชั่นถ่ายโอนเป็น

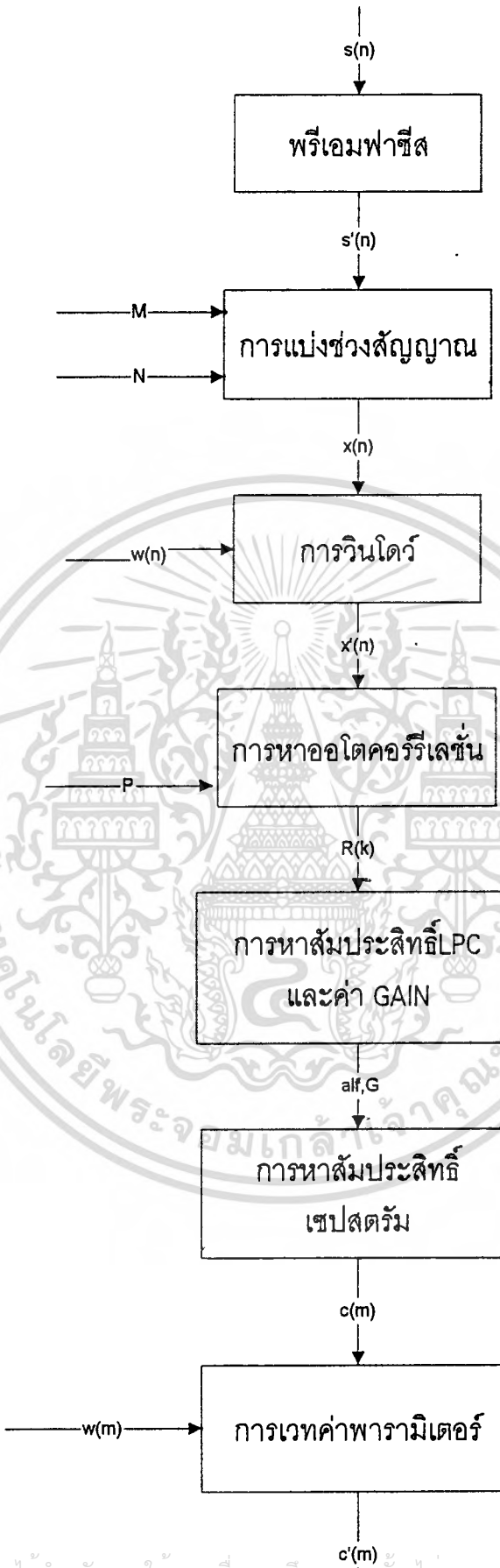
$$H(z) = 1 - a \cdot z^{-1}$$

โดยที่ 0.9<a<1.0 เมื่อเทียบกับรูป 3.1 เราจะได้ว่า

$$s'(n) = s(n) - a \cdot s(n-1)$$

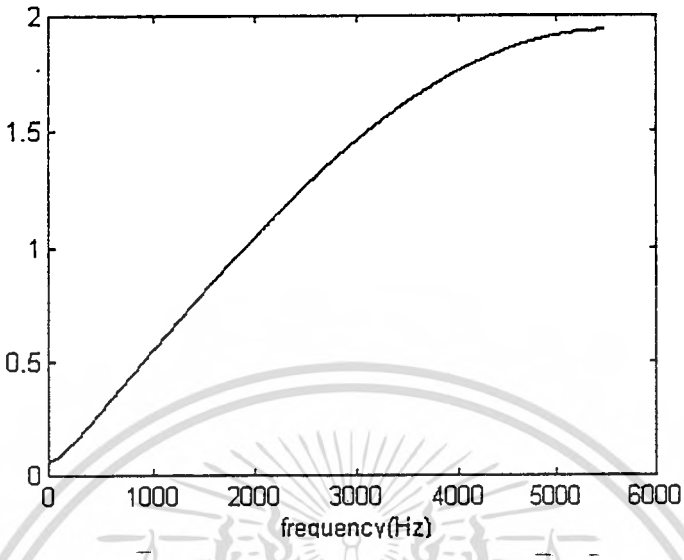
ยิ่งค่า a ใกล้ 1 เท่าใด ความถี่สูงก็จะถูกขยายมากขึ้นเท่านั้น ค่า a ที่นิยมสำหรับใช้ในการหาพารามิเตอร์ของ LPC คือ 15/16 = 0.9375 เมื่อนำฟังก์ชั่นถ่ายโอนมาพล็อตกราฟของขนาดเทียบ กับความถี่ จะได้ดังรูปที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

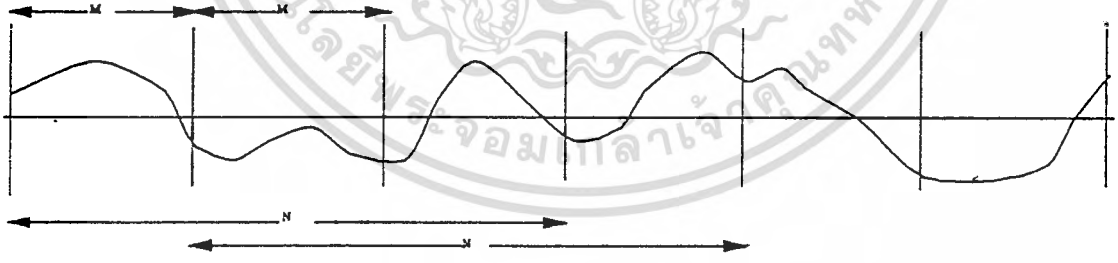
รูปที่ 3.1 แสดงขั้นตอนการเตรียมสัญญาณในการวิเคราะห์



รูปที่ 3.2 แสดงขนาดสเปคตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส

3.2 การแบ่งช่วงสัญญาณ

สัญญาณที่ผ่านการพรีเอมฟาซิสแล้ว s'(n) จะถูกตัดแบ่งออกเป็นช่วงๆ หรือ เฟรม ช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างสัญญาณ ดังรูปที่ 3.3



รูปที่ 3.3 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า M โตกว่าค่า N ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N ก็จะทำให้ตัวอย่างสัญญาณทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักการกำหนดขนาดของวินโดว์

3.2.1 วินโดว์จะต้องสั้นพอที่คุณสมบติของเสียงที่เราสนใจจะวิเคราะห์นั้น ยังไม่เปลี่ยนแปลงในวินโดว์

3.2.2 วินโดว์จะต้องยาวพอที่ จำนวนตัวอย่างสัญญาณในวินโดว์ สามารถนำมาคำนวนหาคุณสมบัติที่ต้องการได้

3.2.3 วินโดว์ที่ติดกัน ไม่ควรจะสั้นจนกระโดดข้ามข้อมูลบางส่วนไป แต่ควรเลื่อนวินโดว์ให้น้อยกว่าขนาดของเฟรม

เนื่องจาก เราใช้ความถี่ในการสุ่มสัญญาณ 11.025 กิโลเฮิรตซ์ ในการวิเคราะห์นี้ เราเลือกใช้ค่า N = 300 และค่า M = 100 นั่นคือ ช่วงในการวิเคราะห์ คือ 27.21 มิลลิวินาที และระยะในการเลื่อนเฟรมประมาณ 9 มิลลิวินาที

3.3 การวินโดว์

พิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วง ใด ๆ ที่ตัดมาวิเคราะห์ (รูปที่ 3.4) จะเห็นว่าที่ขอบของเฟรมที่ตัดมานี้มีความไม่ต่อเนื่องของสัญญาณ ถ้ามองในโดเมนความถี่สูง ก็จะมีความถี่สูงเกิดขึ้น ดังนั้นเพื่อที่จะลดองค์ประกอบทางความถี่ที่สูงเหล่านี้ เราจะคูณด้วยฟังก์ชันวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปคตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ในที่นี้จะใช้ฟังก์ชันวินโดว์แฮมมิ่ง (Hamming window function) ซึ่งนิยามโดยสมการดังนี้ คือ

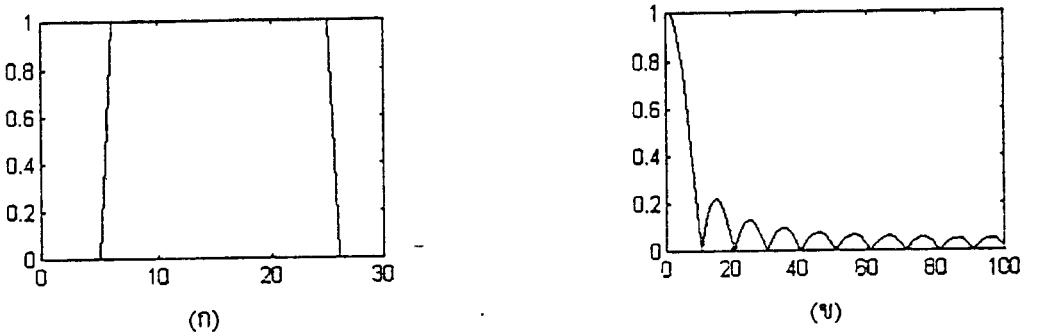
$$w(n) = 0.54 - 0.46\cos(2\pi n/(N-1)) \quad \text{เมื่อ } n = 0,1,2, \ldots, N-1 \tag{3.1}$$

รูปที่ 3.5 และ 3.6 แสดงองค์ประกอบทางเวลา และทางความถี่ของฟังก์ชันวินโดว์แบบสี่เหลี่ยม และแฮมมิ่ง ตามลำดับ จะเห็นว่าสเปคตรัมของวินโดว์แฮมมิ่งมีริพเพิล (ripple) น้อยกว่าของวินโดว์สี่เหลี่ยม ดังนั้น จากรูปที่ 3.1 เราจะได้ว่า

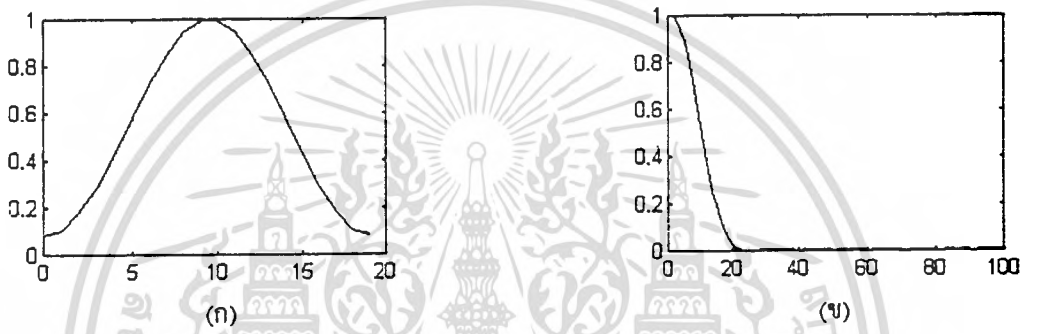
$$x'(n) = x(n)\ w(n) \tag{3.2}$$

รูปที่ 3.4 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 แสดงวินโดว์สี่เหลี่ยม (ก) ในโดเมนเวลา (ข) ในโดเมนความถี่



รูปที่ 3.6 แสดงวินโดว์แบบแฮมมิ่ง (ก) ในโดเมนเวลา (ข) ในโดเมนความถี่

3.4 การคำนวนออโตคอร์รีเลชัน

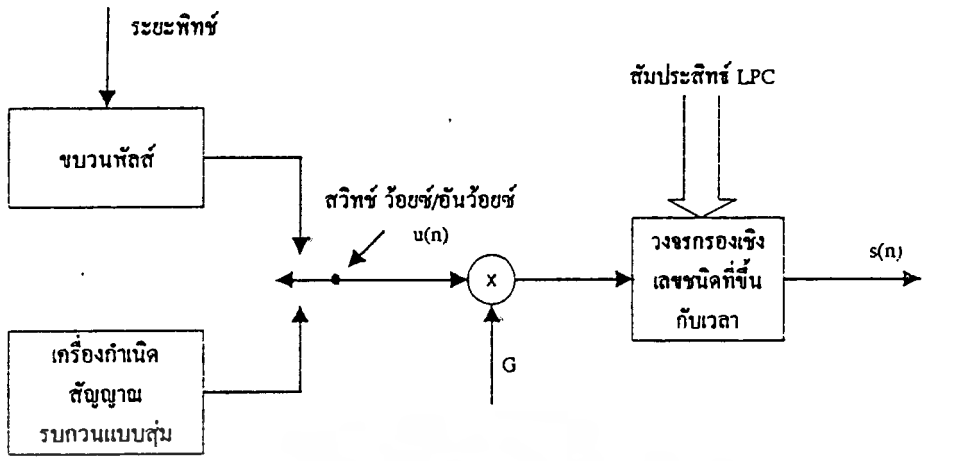
จากหลักการพื้นฐานของการประมาณเชิงเส้นคือ การประมาณค่าสัญญาณจากผลรวมเชิงเส้นของสัญญาณก่อนหน้านี้ สมมติว่าสัญญาณเดิมเป็น s(n) การประมาณค่าสัญญาณเป็น s'(n) ดังนั้นเราสามารถอธิบายการประมาณเชิงเส้นได้ด้วยสมการต่อไปนี้

$$s'(n)=\sum_{k=1}^{p} \alpha_k \, s(n-k) \qquad (3.3)$$

เมื่อ α_k เป็นค่าคงที่ เรียกวิธีการนี้ว่าการประมาณเชิงเส้นอันดับ p โดยมีเงื่อนไขว่า ค่า α_k ที่ใช้ในการประมาณจะต้องทำให้ ผลรวมของกำลังสองของความคลาดเคลื่อน $(s(n)-s'(n))^2$ มีค่าน้อยที่สุด นั่นคือ $\sum e^2(n) = \sum (s(n) - s'(n))^2$ มีค่าต่ำที่สุด ซึ่งจะใช้การประมาณเชิงเส้นวิธีออโตคอร์รีเลชั่น (Autocorreiation Method) หรือ วิธีอัตตสัมพันธ์

จากหลักการพื้นฐานของการประมาณเชิงเส้นและแบบจำลองระบบสร้างสัญญาณเสียงเราสามารถเขียน บล็อกไดอะแกรมการทำการประมาณเชิงเส้น มาสร้างสัญญาณเสียงพูดได้ดังรูปที่ 3.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 บล็อกไดอะแกรมแสดงโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย

จากรูป 3.7 สามารถเขียนสมการได้เป็น

$$s(n)=G*u(n)+\sum_{k=1}^{p}a_k\, s(n-k) \tag{3.4}$$

การประมาณเชิงเส้นโดยใช้สัมประสิทธิ์ $\{\alpha_k\}$ คือ

$$s'(n)=\sum_{k=1}^{p}\alpha_k\, s(n-k) \tag{3.5}$$

ดังนั้น ความคลาดเคลื่อน คือ

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

$$=s(n)-\sum_{k=1}^{p}\alpha_k\, s(n-k) \tag{3.6}$$

ฟังก์ชั่นถ่ายโอนระหว่าง e(n) และ s(n) คือ

$$A(z)=E(z)/S(z)$$

$$=1-\sum_{k=1}^{p}\alpha_k z^{-k} \tag{3.7}$$

จากสมการ (3.4)-(3.6) จะเห็นได้ว่า ถ้า $\{\alpha_k\} = \{a_k\}$ แล้ว

$$e(n)=G*u(n) \tag{3.8}$$

ดังนั้น ค่าผลรวมของกำลังสองของความคลาดเคลื่อน

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

$$E_n = \sum_m [s(m)-s'(m)]^2 \tag{3.9}$$

โดยที่ n คือ ช่วงที่ n ของสัญญาณที่ใช้คำนวณ เพราะฉะนั้นเพื่อให้ได้ค่า E_nต่ำที่สุดจะต้องมีเงื่อนไขว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

จากสมการ (3.9)

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

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

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

$$\sum_{k=1}^{p} \alpha_k \sum_m s_n(m-k)s_n(m-i) = \sum_m s_n(m)s_n(m-i) \qquad \text{เมื่อ i=1,2,3.....,p} \qquad (3.10)$$

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

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,j) = \phi_n(i,0) \qquad (3.11)$$

โดยสมการ (3.9)-(3.10)จะได้ว่า

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

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

$$E_n = \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k) \qquad (3.12)$$

สมมุติว่าใน 1 เฟรม ของสัญญาณ ที่ตัดมาคำนวณมี N ตัวอย่าง คือ $S_n(0), S_n(1), S_n(2),..., S_n(N-1)$ ...
ในที่นี้เราให้ $S_n(m) = 0$ เมื่อ $m < 0$ หรือ $m > N-1$ เพราะฉะนั้น

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

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

ให้ $$R_n(k) = \sum_m^{N-1-k} s_n(m)s_n(m+k) \qquad \text{เมื่อ k=0,1,2,....,p} \qquad (3.13)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น จากสมการ (3.2) และ (3.13) จะได้ว่า

$$R_n(k) = \sum_{m=0}^{N-1-k} x'(m)x'(m+k) \tag{3.14}$$

จากสมการที่ (3.11) จะได้ว่า

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \qquad \text{เมื่อ } i=1,2,3,\ldots,p \tag{3.15}$$

จากสมการ (3.15) เขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

$$\begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & \cdots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix} \tag{3.16}$$

หรือ

$$R_n \cdot \alpha = r_n \tag{3.17}$$

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

3.5 การหาสัมประสิทธิ์ α และอัตราการขาย G

เมื่อได้ค่า $R_n(0)$, $R_n(1)$, $R_n(2)$, ..., $R_n(p)$ แล้ว ก็สามารถหาค่า α ได้จากสมการ (3.17) นั่นคือ

$$\alpha = R_n^{-1} \cdot r_n \tag{3.18}$$

และจากสมการ (3.8) จะได้ว่า

e(n)=G*u(n)

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จาก สมการ (3.12) จะได้ว่า

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

$$= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \qquad (3.19)$$

และจากสมการ (3.19) เราสามารถหาค่า G โดยตรงจาก

$$G^2 = \frac{R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)}{\sum_{m=0}^{N-1} u^2(m)} \qquad (3.20)$$

3.6 การเปลี่ยนพารามิเตอร์ LPC เป็น สัมประสิทธิ์เซปสตรัม

ในการรู้จำเสียงพูดนั้น สัมประสิทธิ์เซปสตรัมนี้เป็นพารามิเตอร์ที่มีลักษณะน่าเชื่อถือได้ดีกว่าสัมประสิทธิ์ LPC ทั้งยังมีความสัมพันธ์ใกล้ชิดกับการรับรู้เสียง ตามความรู้สึกของมนุษย์โดยแท้จริง สัมประสิทธิ์เซปสตรัมสามารถหาได้โดยตรงจากสัมประสิทธิ์ LPC ดังนี้

$$C_0 = \ln G$$

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

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

$$Q \approx \frac{3}{2} p$$

3.7 การเวทค่าพารามิเตอร์ (Parameter Weighting)

เนื่องจาก สัมประสิทธิ์เซปสตรัมที่ได้นั้น ช่วงลำดับต้นๆ และลำดับท้ายๆ ของเฟรมที่นำมาวิเคราะห์จะเกิดความคลาดเคลื่อนมากกว่าบริเวณส่วนอื่น เพราะฉะนั้น จึงทำการถ่วงน้ำหนัก เพื่อลดค่าความคลาดเคลื่อนดังกล่าวนี้ ด้วยฟังก์ชั่นเวทติ่ง ดังนี้คือ

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

จะได้ $C'_m = C_m * W_m$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
เวคเตอร์ ควอนไตซ์เซชัน
(Vector Quantization)

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

ขั้นตอนในการรับรู้แบ่งเป็น 2 ขั้นตอน ดังนี้

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

ข. ขั้นรับรู้ (Recognition) จะเป็นการทดสอบการรับรู้ ระหว่างแบบอ้างอิง กับแบบทดสอบ โดยจะทำการเปรียบเทียบพารามิเตอร์ของแบบทดสอบ กับแบบอ้างอิงทั้งหมด แบบอ้างอิงที่เลือก คือ แบบอ้างอิงที่มีพารามิเตอร์ใกล้เคียงกับแบบทดสอบที่สุด

4.1 เวคเตอร์ ควอนไตซ์เซชัน

ในทฤษฎีของการสื่อสารนั้น แหล่งเข้ารหัสจะใช้เทคนิคการเปลี่ยนค่าสัญญาณ เอ้าท์พุท ไปเป็นรหัสแบบไบนารี ซึ่งเป้าหมายของเครื่องเข้ารหัสคือ การเป็นไปได้ที่จะมีการผิดเพี้ยนของข้อมูลน้อยสุด ด้วยอัตราความเร็ว(bit rate) ที่กำหนด

เวคเตอร์ ควอนไตซ์เซชัน เป็นเทคนิคเข้ารหัสเทคนิคหนึ่งที่มีคุณภาพ มันจะทำการเข้ารหัสเวคเตอร์พารามิเตอร์ของคลื่นที่แทนสเปกตรัมของเสียง ซึ่งเป็นเวคเตอร์ อินพุท ให้เป็นตัวเลขจำนวนเต็ม (index) โดยนำไปเปรียบเทียบกับโค้ดบุคที่ได้สร้างเก็บไว้ก่อนแล้ว และโค้ดบุค ที่ถูกเลือกนั้นจะเป็นตัวที่ใกล้เคียงที่สุดกับเวคเตอร์ อินพุท ผลที่ได้จาก เวคเตอร์ ควอนไตซ์เซชัน ที่แสดงออกมานั้นขึ้นกับเซตของโค้ดบุคที่สร้างเก็บไว้ก่อนแล้วนั่นเอง

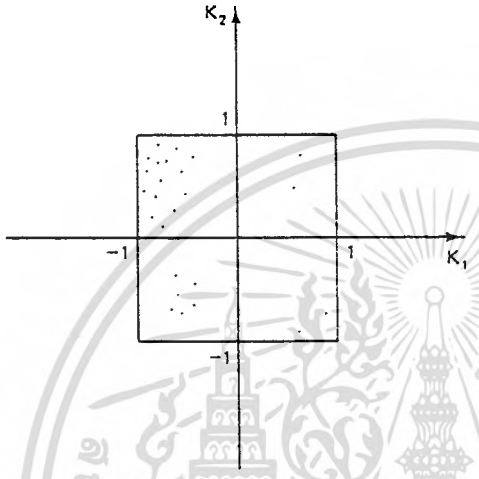
ถ้าเราเปรียบเทียบอัตราข้อมูลของการแทนเวคเตอร์สำหรับเสียงที่ยังไม่ได้เข้ารหัส เราจะเห็นว่าการวิเคราะห์เสียงนั้นต้องการการลดลงของอัตราข้อมูล พิจารณาดังนี้ เสียงพูดที่แซมเปิลด้วยความถี่ 10 kHz และมีขนาด 16 บิต จะต้องการอัตราข้อมูล 160,000 bps ในการจัดเก็บซึ่งเป็นรูปแบบที่ยังไม่ได้มีการบีบอัดข้อมูล สำหรับการวิเคราะห์เสียงพูด พิจารณา เวคเตอร์ที่มี 10 มิติ ใช้ 100 สเปกตรัมต่อวินาที ถ้าเราใช้ 16 บิตแทนแต่ละเวคเตอร์ อัตราที่ใช้ในการเก็บ คือ 100*10*16 bps หรือ 16,000 bps สามารถลดไปได้ถึง 10 เท่าของสัญญาณที่ยังไม่ได้บีบอัด ยิ่งไปกว่านั้นวิธีการเวคเตอร์ควอนไตซ์เซชันยังสามารถลดอัตราข้อมูลลงไปได้อีก โดยการสร้าง โค้ดบุค เพื่อใช้ในการชี้ว่าเวคเตอร์โค้ดบุคไหนที่เหมือนกับเวคเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

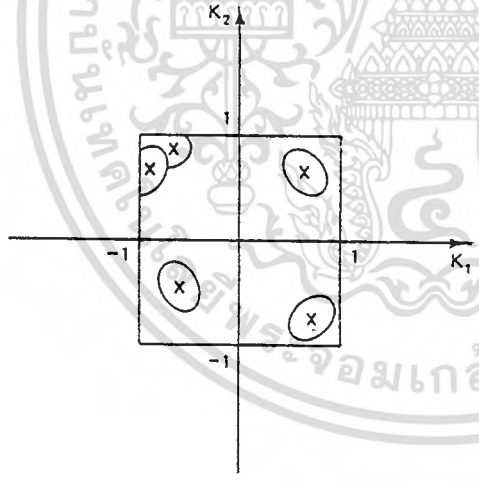
เสียงอินพุทที่เข้ามามากที่สุด สมมุติ โค้ดบุคแทนด้วยตัวเลข 10 บิต และมีอัตราข้อมูล 100 สเปกตรัมต่อวินาที เราจะได้ว่าอัตราข้อมูล 1,000 bps ซึ่งสามารถลดอัตราข้อมูลลงไปได้ถึง 16 เท่า

เวคเตอร์สเปกตรัมที่จะนำไปทำการควอนไตซ์ เป็นเวคเตอร์ที่ได้จากการประมาณเชิงเส้น และผ่านการปรับปรุงให้ดีขึ้นโดยการทำ เซปสตรัมและการเวทค่าพารามิเตอร์ ซึ่งจะแสดงโดย V_l ,l = 1 , 2 , ... , L ซึ่งแต่ละเวคเตอร์เป็นเวคเตอร์ที่มี p มิติ (p คือ order หลังจากผ่านการเวทค่าเรียบร้อยแล้ว)

เช่นในรูปตัวอย่างนี้ คือ p = 2 order = 2 ของ LPC Model



รูปที่ 4.1
แสดงการกระจายเฟรมของเสียงพูด
แต่ละจุดแทนเฟรมของเสียง



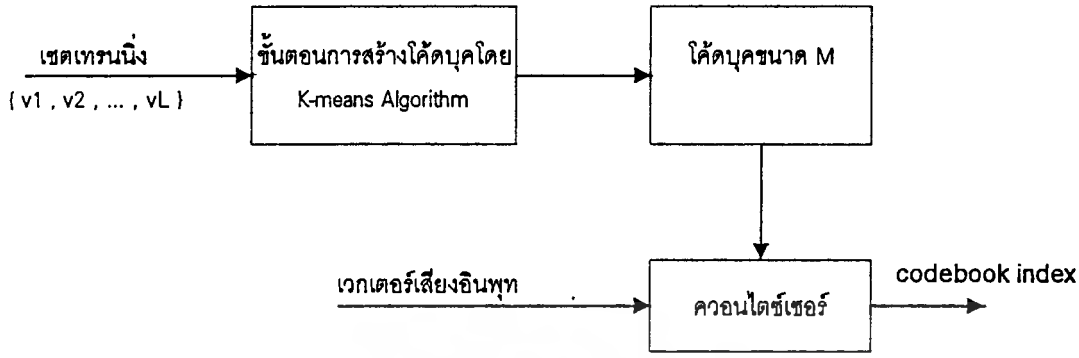
รูปที่ 4.2
การรวมกลุ่มของเฟรมของเสียงเพื่อไปสร้าง
โค้ดบุค ซึ่ง x แทนเวคเตอร์ศูนย์กลาง
เพื่อให้เวคเตอร์ทั้งหมดในกลุ่มสามารถแบ่ง
แยกได้

ข้อดีของวิธีการ เวคเตอร์ ควอนไตซ์เซชัน คือ

- ลดความจุในการเก็บข้อมูลสเปกตรัม ที่ได้จากการวิเคราะห์ลง
- ลดการคำนวณในการเปรียบเทียบความเหมือนของเวคเตอร์สเปกตรัม ที่ได้จากการวิเคราะห์ลง ซึ่งเป็นการคำนวณหลักที่ใช้ในการรับรู้เสียงพูด
- กระบวนการการเลือก โค้ดบุค ที่ดีที่สุดเพื่อแสดงแทนเวคเตอร์สเปกตรัมของเสียงพูดที่เข้ามาเป็นไปด้วยความรอบคอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ส่วนประกอบของ เวคเตอร์ควอนไตซ์เซชัน



รูปที่ 4.3 แสดงบล็อคไดอะแกรมของเวคเตอร์ ควอนไตซ์เซชัน

4.2.1 การทำงานของ เวคเตอร์ ควอนไตซ์เซชัน

1. เซตของเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ v_1 , v_2 , v_3 , ... , v_L ซึ่งเป็นเซตขนาดใหญ่ จะถูกนำมาใช้ในการสร้างเซตของ โค้ดบุค เพื่อใช้เป็นตัวแทนเวคเตอร์สเปกตรัมของเสียงที่สังเกตได้ ขนาดโค้ดบุคของเวคเตอร์ ควอนไตซ์เซชัน คือ $M = 2^B$ เวคเตอร์ (เราเรียกว่า B -bit โค้ดบุค) ซึ่งเราต้องการให้ $L >> M$ เพื่อที่จะหาเซตของ M โค้ดบุค ที่ดีที่สุด ในทางปฏิบัตินั้นเพื่อให้ เวคเตอร์ ควอนไตซ์เซชัน โค้ดบุค สามารถทำงานได้ดีนั้น ขนาด L อย่างน้อยที่สุดที่เราสามารถใช้ได้ คือ 10M

2. การวัดความเหมือน หรือ ระยะทางระหว่าง เวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ 2 เวคเตอร์ ดังนั้นสามารถจัดกลุ่มเซตของเวคเตอร์ เทรนนิ่ง เพื่อสร้าง โค้ดบุค เราแสดงระยะทางสเปกตรัม $d(v_i, v_j)$ ระหว่าง 2 เวคเตอร์ v_i และ v_j คือ d_{ij}

3. ฟังก์ชันย่อยในการคำนวณหาจุดศูนย์กลาง เพื่อแบ่งแยกเซตของ เทรนนิ่ง จำนวน L เวคเตอร์ ไปเป็น M กลุ่ม และ M โค้ดบุค ที่ได้คือจุดศูนย์กลางของแต่ละกลุ่ม

4. ฟังก์ชันย่อยในการแบ่งแยกสำหรับเวคเตอร์สเปกตรัมที่ได้จากการวิเคราะห์ ซึ่งจะเลือกโค้ดบุค ที่ใกล้เคียงที่สุดให้กับเวคเตอร์ อินพุท และผลลัพธ์ที่ได้คือ ดัชนีโค้ดบุคซึ่งเป็นตัวแทนของสเปกตรัม มักจะใช้กฎ Nearest Neighbour ในการพิจารณา ส่วนนี้เป็นหัวใจหลักของ ควอนไตซ์เซอร์ ซึ่ง อินพุท คือ เวคเตอร์สเปกตรัมของเสียง และ เอาท์พุท คือ ดัชนีโค้ดบุค ที่มีความเหมือนกับ เวคเตอร์เสียงอินพุทมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

037217

4.2.2 เซต เทรนนิ่ง (Training) ของ เวคเตอร์ ควอนไตซ์เซชัน

เซต เทรนนิ่งของเวคเตอร์ ควอนไตซ์เซชันซึ่งเป็นเสียงพูดที่จะนำมาใช้ในการสร้างโค้ดบุค ซึ่งเซต เทรนนิ่งจะต้องมีขนาดใหญ่มาก ๆ ถึงจะสามารถสร้างโค้ดบุคที่ดีได้ เซตเทรนนิ่งจะแปรตามสภาวะต่าง ๆ ดังนี้

- ผู้พูด รวมทั้งช่วงของอายุ เพศ ความเร็ว ช้าในการพูด ระดับเสียง ฯลฯ
- สภาวะแวดล้อมขณะพูด เช่น ในห้องเงียบ หรือในรถ
- ระบบการแปลงการส่ง รวมทั้ง ไมโครโฟน หูโทรศัพท์
- หน่วยของเสียงรวมถึง คำศัพท์ต่าง ๆ และข้อความที่สนทนา

4.3 หลักการของเวคเตอร์ควอนไตซ์เซชัน

เวคเตอร์ควอนไตซ์เซชัน เป็นวิธีการควอนไตซ์คลื่นเสียง ในการควอนไตซ์แบบเวคเตอร์ มีวิธีการควอนไตซ์ดังนี้ คือ จะมีเวคเตอร์โค้ด หรือ โค้ดบุคเก็บไว้ อินพุทที่เข้ามาจะถูกทำการเปรียบเทียบกับโค้ดบุคที่มีอยู่ โดยจะพิจารณาว่าอินพุทที่เข้ามานั้นห่างจากโค้ดบุคใดน้อยที่สุด อินพุทดังกล่าวจะถูกแทนด้วยเวคเตอร์โค้ดนั้น

โค้ดบุด ที่ใช้จะเหมาะสมเพียงใด ขึ้นกับความคลาดเคลื่อนรวมทั้งหมดของต้นแบบที่ใช้หา เมื่อต้นแบบที่พิจารณามีจำนวนจำกัด โค้ดบุคที่มีความคลาดเคลื่อนน้อยที่สุด จะได้จากการทำซ้ำจนกว่าจะได้ค่าที่เหมาะสม โดยเหตุนี้จึงขึ้นกับ ค่าแรนดอมที่สุมขึ้นมาจำนวนเท่ากับขนาดของโค้ดบุคเพื่อใช้เป็นค่าเริ่มต้น ในวิธีการนี้ข้อมูลต้นแบบที่ต้องการหาถูกแยกไปอยู่ในแต่ละกลุ่มของค่าเริ่มต้น จะทำการหาค่าเฉลี่ย ของแต่ละกลุ่มเพื่อเป็นค่ากลางของกลุ่ม ทำวิธีการดังกล่าวซ้ำจนกว่าความคลาดเคลื่อนรวมจะต่ำกว่าค่า ๆ หนึ่ง หรือการลดลงของความคลาดเคลื่อนรวมน้อยกว่าค่า ๆ หนึ่ง โดยค่ากลางดังกล่าวของกลุ่มจะถูกเก็บเป็นเวคเตอร์โค้ด แต่แน่นอนว่าค่าความคลาดเคลื่อนรวมจะลดลงทุกครั้งที่มีการคำนวณซ้ำใหม่ จึงขึ้นกับค่าที่กำหนดว่าต้องการให้ความคลาดเคลื่อนรวมน้อยเพียงใด

4.4 การคำนวณ เวคเตอร์ควอนไตซ์เซชัน

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

$$y = q (x) \qquad (4.1)$$

โดย q(.) เป็นโอเปอร์เรเตอร์ของควอนไตซ์ y ถูกเรียกว่าเอาท์พุทเวคเตอร์ของค่า x โดย y เป็นค่าใดค่าหนึ่งใน $Y = \{ y_i , 1 \leq i \leq M \}$ โดย $y_i = [y_{i1} \; y_{i2} \dots y_{ip}]$ Y เป็นเซ็ตของโค้ดบุด M เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขนาดของโค้ดบุค และ { y_i } เป็นเซ็ตของเวคเตอร์โค้ด y_i อาจเรียกว่าเป็นโค้ดอ้างอิง และ M อาจเรียกว่าจำนวนระดับขั้น จะทำการแบ่งเวคเตอร์ x ไปใน M เซล (C_i , $1 \leq i \leq M$) เมื่อ x อยู่ในเซล C_i

$q(x) = Y_i$ ถ้า $x \in C_i$

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

$$D = \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} d[x(n), y(n)] \qquad (4.2)$$

การคำนวณหาโค้ดบุคมีหลายวิธี แต่ที่จะนำมาใช้คือ Lloyd Algorithm (K-means Algorithm)

4.5 การวัดความคลาดเคลื่อน

การวัดความคลาดเคลื่อนเป็นส่วนจำเป็นและเป็นประโยชน์ต่อการออกแบบ โค้ดบุค ซึ่งเราวัดโดยการวัดระยะทางซึ่ง สมการพีชคณิตที่ใช้ในการหาระยะทางในปัจจุบันมีวิธีการหลายแบบ แต่ที่จะนำมาใช้ในปริญญานิพนธ์นี้ คือ square error distortion.

square error distortion :

ถ้าสัญญาณที่เข้ามามี P มิติ เราสามารถหาระยะห่างระหว่างสัญญาณเข้ากับเวคเตอร์โค้ด

$$d(v_1, v_2) = \| v_1 - v_2 \|^2 = \sum_{i=0}^{k-1} (x_i - y_i)^2 \qquad (4.3)$$

ซึ่งเป็นวิธีการคำนวณที่ง่าย และรวดเร็วในการวัดความคลาดเคลื่อน จึงนิยมใช้วิธีนี้ แต่ก็ยังไม่เหมาะสมเพียงพอ ควรใช้ Itakura distance มากกว่า แต่เป็นวิธีที่มีความซับซ้อนมาก และยุ่งยากในการคำนวณ

4.6 การจัดกลุ่มเพื่อสร้างแบบอ้างอิงโดยใช้ K-means Algorithm

การสร้างกลุ่มของแบบอ้างอิงเป็นขั้นตอนในส่วนของการเรียนรู้ โดยเริ่มต้นจากการนำข้อมูลสัญญาณเสียงที่ได้จากส่วนของการวิเคราะห์ มาทำการสร้างเป็นกลุ่มของแบบอ้างอิง ในระบบการรับรู้เสียงพูดแบบต่างบุคคล จะใช้แบบอ้างอิงของคำหนึ่ง ๆ จากผู้พูดจำนวนมาก เพื่อที่จะได้ครอบคลุมถึงความแปรปรวนต่าง ๆ ที่จะเกิดขึ้นระหว่างผู้พูดแต่ละคน แต่อย่างไรก็ตามผลที่เกิดขึ้นตามมา เมื่อจำเป็นจะต้องมีข้อมูลจำนวนมากในกลุ่มของแบบอ้างอิง คือ

ก. เวลาที่ใช้ในการตอบสนอง ในการรับรู้เสียงพูด ขั้นตอนการเปรียบเทียบแบบทดสอบกับแบบอ้างอิงทั้งหมดที่มีเป็นขั้นตอนที่สำคัญในการรับรู้ เมื่อแบบอ้างอิงมีจำนวนมาก การเปรียบเทียบจึงจำเป็นที่จะต้องอาศัยเวลาเพิ่มขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น ทั้งนี้ห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข. เนื้อที่ในหน่วยความจำสำรองที่ใช้ในการเก็บแบบอ้างอิง เมื่อมีแบบอ้างอิงจำนวนมากที่จำเป็นต้องใช้ในการเปรียบเทียบ เนื้อที่ในหน่วยจำสำรองก็จำเป็นต้องมีเพิ่มขึ้นด้วย

ค. ความถูกต้องในการรับรู้ เนื่องจากจำนวนแบบอ้างอิงในแต่ละคำไม่ได้เป็นสัดส่วนโดยตรงกับความถูกต้องในการรับรู้ กล่าวคือเมื่อเราเพิ่มจำนวนแบบอ้างอิงในแต่ละคำไปจนถึงระดับหนึ่ง ความถูกต้องในการรับรู้จะเริ่มคงที่ และในบางครั้งอาจจะมีค่าลดลงด้วย

จากเหตุผลดังกล่าว จึงทำให้มีความพยายามที่จะจัดกลุ่มของแบบอ้างอิงในแต่ละคำใหม่เพื่อให้ได้แบบอ้างอิงจำนวนที่พอเหมาะ และสามารถใช้เป็นตัวแทนของแบบอ้างอิงที่มีอยู่ทั้งหมดได้ อัลกอลิทึมที่ใช้ในการแบ่งกลุ่มมีอยู่เป็นจำนวนมาก สำหรับอัลกอลิทึมที่จะกล่าวถึงคือ K-means Algorithm ตามขั้นตอนโดยสรุป ดังนี้

4.6.1. การหาจุดศูนย์กลางของกลุ่ม (Center Cluster)

ในขั้นตอนนี้จะทำการกำหนดจำนวนกลุ่มของแบบอ้างอิงให้มีค่าเท่ากับ M และเริ่มต้นจากการแรนดอมค่ามาจำนวนเท่ากับขนาดของโค้ดบุค เพื่อเป็นค่าจุดศูนย์กลางของแต่ละกลุ่มก่อน จากนั้นจะแบ่งเวคเตอร์ต่าง ๆ เข้าไปตามกลุ่มต่าง ๆ โดยพิจารณาจากระยะทาง หรือความต่างกันน้อยที่สุดกับกลุ่มไหนก็จะให้อยู่กลุ่มนั้น จากนั้นก็จะทำการเฉลี่ยค่าของเวคเตอร์ที่อยู่ในกลุ่มเพื่อให้เป็นจุดศูนย์กลางใหม่

4.6.2. การจัดกลุ่มใหม่ (Reclassification)

แบบอ้างอิงที่เหลืออยู่ในแต่ละคำ จะถูกจัดให้เข้าไปในแต่ละกลุ่ม โดยมีเงื่อนไขคือ ระยะทางของแบบอ้างอิงนี้กับแบบอ้างอิงในกลุ่มจะมีค่าเฉลี่ยน้อยที่สุด หลังจากที่แบบอ้างอิงทุกแบบถูกจัดเข้ากลุ่มเรียบร้อยแล้ว จะทำการหาจุดศูนย์กลางของแต่ละกลุ่มใหม่ ขั้นตอนนี้จะทำซ้ำจนกว่าจุดศูนย์กลางของแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลง อย่างไรก็ตามจากขั้นตอนดังกล่าวจะเห็นว่า การกำหนดค่าจำนวนกลุ่มที่เหมาะสมจะทำได้ลำบาก ดังนั้นจึงได้มีการปรับปรุงเทคนิคนี้เป็นการกำหนดค่าระยะทางสูงสุดที่จะทำให้เกิดกลุ่มใหม่ขึ้น คือเมื่อแบบอ้างอิงใดที่ถูกเลือกให้เข้ากลุ่มด้วยระยะทางที่น้อยที่สุดแล้ว แต่ระยะทางนี้ยังคงมีค่ามากกว่าระดับที่กำหนด แบบอ้างอิงนี้จะถูกนำไปสร้างเป็นจุดศูนย์กลางของกลุ่มใหม่ทันที

4.7 การคำนวณจุดศูนย์กลางสำหรับการออกแบบ เวคเตอร์ ควอนไตซ์เซชัน โค้ดบุค

การสร้าง โค้ดบุค โดยใช้ K-means สามารถสรุปสั้น ๆ ได้ดังนี้

ก) ความเพี้ยนน้อยสุด สำหรับแต่ละเวคเตอร์ อินพุท x_t เมื่อเปรียบเทียบกับเวคเตอร์ y_i ซึ่งเป็นโค้ดบุค C ที่เข้าไป กลุ่มของเวคเตอร์ เทรนนิ่ง จะเป็นไปตาม Code words

ข) การคำนวณหาจุดศูนย์กลาง แต่ละกลุ่มของเวคเตอร์ จะคำนวณหาจุดศูนย์กลางใหม่ ซึ่งมีค่าความผิดเพี้ยนเฉลี่ยต่ำ สำหรับสมาชิกกลุ่มที่ถูกแทนโดยเวคเตอร์จุดศูนย์กลาง

ทำ 2 ขั้นตอนซ้ำจนกว่าจะเกิดการลู่เข้า (convergent) ขั้นแรกนั้นเป็นการทำโดยตรง คือนำเข้าไปเพื่อหาความผิดเพี้ยน ส่วนขั้นที่ 2 นั้นเป็นการหาจุดศูนย์กลาง ซึ่งเป็นขั้นตอนที่สำคัญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พิจารณา $(x_i)_{i=1}^{L}$ และค่าความเพี้ยน d(x,y) โดยทั่วไป เราสมมุติว่าเวคเตอร์เหล่านี้ถูกกำหนดไว้ในกลุ่มป้ายเดียวกัน (หรือ Code word) จุดศูนย์กลางของ $(x_i)_{i=1}^{L}$ ถูกจำกัดความเป็นเหมือนเวคเตอร์ $\overline{Y}$ ซึ่งเป็นค่าความเพี้ยนเฉลี่ย น้อยที่สุด

$$\overline{Y} = \underline{\underline{\Delta}} \arg \min \frac{1}{L}\sum_{i=1}^{L} d(x_i,y) \tag{4.4}$$

เมื่อ x_i และ y เป็นเวคเตอร์ โดยที่ $x_i = (x_{i1}, x_{i2}, x_{i3}, ..., x_{ik})$ และ $y = (y_1, y_2, ... y_k)$ วัดในพื้นที่ k มิติ กับระยะทางตามแบบยูคลิด จุดศูนย์กลางคือค่าเฉลี่ยของเซตของ เทรนนิ่ง

$$\overline{Y} = \frac{1}{L}\sum_{i=1}^{L} x_i \tag{4.5}$$

$\overline{Y}$ เป็นจุดศูนย์กลางซึ่งเป็นเวคเตอร์ที่อยู่ตรงกลางของ $(x_i)_{i=1}^{L}$ ซึ่งแต่ละมิติจะไม่ขึ้นแก่กัน หมายความว่า แต่ละ y_k เป็นค่ากลางของ $(x_i)_{i=1}^{L}$

4.8 กฎของการตัดสินใจ (Decision Rules)

ผลที่ได้จากการเปรียบเทียบระยะทางระหว่างแบบทดสอบกับแบบอ้างอิงแต่ละแบบ ระยะที่ได้ทั้งหมดจะนำมาผ่านขั้นตอนการตัดสินใจ เพื่อหาผลลัพธ์ของการรับรู้ กฎการตัดสินใจ (Decision Rules) ที่ใช้ในการรับรู้เสียงพูด มีดังนี้ คือ

4.8.1. กฎ Nearest Neighbor (NN)

ผลลัพธ์ของการรับรู้จากกฎการตัดสินใจนี้ ได้แก่ แบบอ้างอิงที่มีระยะทางจากแบบทดสอบน้อยที่สุด กฎนี้เหมาะสำหรับระบบการรับรู้เสียงพูดแบบบุคคลเดียว

4.8.2. กฎ K-Nearest Neighbor (KNN)

เป็นกฎที่ดัดแปลงมาจากกฎ NN อีกทีหนึ่ง เพื่อให้เหมาะสมกับการรับรู้เสียงพูดแบบต่างบุคคล ซึ่งมีแบบอ้างอิงในแต่ละคำมากกว่า 1 แบบ กฎนี้แบ่งออกเป็น 2 แบบ คือ

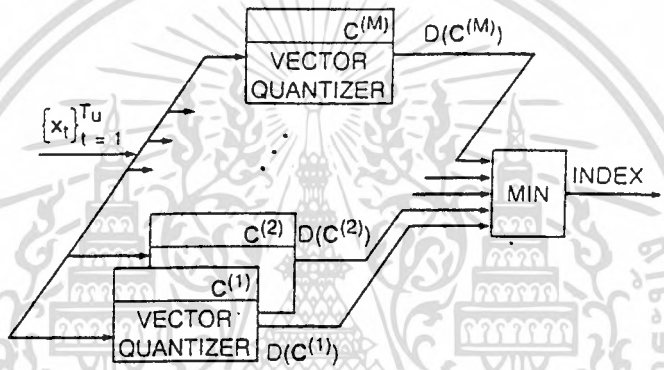
ก. แบบที่ 1 จะทำการหาแบบอ้างอิงจำนวน K แบบ ที่มีระยะทางห่างจากแบบทดสอบน้อยที่สุด ผลลัพธ์ของการรับรู้จะได้จาก แบบอ้างอิงที่มีจำนวนมากที่สุดในจำนวน K แบบ ที่ได้รับการเลือกมา

ข. แบบที่ 2 จะทำการหาแบบอ้างอิงจำนวน K แบบ ในแต่ละคำ โดยที่ทั้ง K แบบมีระยะทางห่างจากแบบทดสอบน้อยที่สุด ผลลัพธ์ของการเรียนรู้จะได้จาก แบบอ้างอิงที่ค่าเฉลี่ยของระยะทางในจำนวน K แบบ ที่ได้รับการเลือกมาน้อยที่สุด วิธีนี้จะให้ผลลัพธ์ที่ถูกต้องมากขึ้น โดยกำหนดค่า K เป็น 2 หรือ 3 เมื่อเทียบกับกฎ NN อย่างไรก็ตามวิธีนี้จะทำให้เวลาที่ใช้ในการรับรู้เพิ่มขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในระบบเสียงพูดทั่วไป จะมีการกำหนดค่าระยะทางมากที่สุดที่ยอมรับได้ (Reject Value) เพื่อเพิ่มความถูกต้องของการรับรู้ในกรณีที่ผลลัพธ์ที่ได้จากการตัดสินใจมีค่าใกล้เคียงกันมาก วิธีการกำหนดค่านี้ต้องใช้ความระมัดระวังเป็นอย่างมาก เนื่องจากถ้าเรากำหนดค่านี้สูงเกินไป หรือ ไม่ได้กำหนด โอกาสที่จะทำให้ผลลัพธ์ของการรับรู้ผิดพลาดไปจะมีมาก แต่ถ้าเรากำหนดค่านี้ต่ำเกินไปจะทำให้ผลลัพธ์ของการรับรู้สำหรับแบบอ้างอิงที่ถูกต้องเกิดขึ้นได้ยาก ดังนั้นโดยทั่วไปค่าระยะทางนี้มักจะได้มาจากการทดลอง เพื่อหาค่าที่เหมาะสมสำหรับกลุ่มของแบบอ้างอิงหนึ่ง ๆ เท่านั้น

4.9 เวคเตอร์ ควอนไตซ์เซชัน สำหรับระบบการรับรู้เสียงพูด



รูปที่ 4.4 แสดงเวคเตอร์ ควอนไตซ์เซอร์ ที่ใช้ในระบบการรับรู้เสียงพูด

เพื่อที่จะสร้าง โค้ดบุค เราต้องการเวคเตอร์ทดสอบเป็นจำนวนมาก ซึ่งจะจัดเวคเตอร์ออกเป็นกลุ่มๆ และแต่ละกลุ่ม i แทนจุดศูนย์กลางด้วย y_i ซึ่งหาจุดศูนย์กลางได้จากการเฉลี่ยค่าเวคเตอร์ทั้งหมดในกลุ่ม ซึ่ง y_i จะเป็นเวคเตอร์โค้ดที่เก็บไว้ หรือเป็นโค้ดบุคเก็บไว้สำหรับการควอนไตซ์เวคเตอร์เสียงต่อไป เวคเตอร์เสียง อินพุท ของควอนไตซ์เซอร์ $= x_i$ จะถูกเปรียบเทียบระยะทางกับ y_i $d(x_i, y_i)$ ค่า y_m ที่เราเลือกคือ ระยะทาง d น้อยที่สุด ซึ่ง y_m จะแทน x_i ที่เข้ามา จากนั้นจะส่งดัชนี m ไปยังเครื่องรับ จากเมตริกซ์ระยะทาง เราได้ว่า ถ้า x ใกล้ y_1 มากกว่า y_2 สรุปได้ว่า x เสียงเหมือน y_1 มากกว่า y_2

เวคเตอร์ ควอนไตซ์เซชัน ที่ใช้เพื่อการออกแบบการรับรู้เสียงพูดนั้น มี จำนวนควอนไตซ์เซอร์ M ตัว นั่นหมายถึง มี M ระดับของเสียงเพื่อการรับรู้ แต่ละระดับเสียงสามารถพิจารณามาจากเซตของข้อมูล เทรนนิ่ง $\{x_t^{(i)}\}$ เมื่อ i=1 , 2 , ... , M ซึ่ง M เป็นดัชนีระดับ แต่ละเซตของ เทรนนิ่ง ในแต่ละระดับจะเก็บเสียงที่อยู่ในระดับเดียวกัน เมื่อมีเสียงที่เราไม่รู้ $\{x_t\}^{Tu}$ เข้ามา จะเป็นอินพุทเข้าไปยังทุก ๆ ควอนไตซ์เซอร์ ค่าดัชนีระดับ (Index) ที่เราเลือก จะเป็นระดับที่มีความผิดเพี้ยนเฉลี่ยน้อยที่สุด $D(c^{(i)})$ เมื่อ i=1 , 2 , ... , M ซึ่งความผิดเพี้ยนเฉลี่ยนี้เราดูได้จากการวัดระยะทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใด ๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

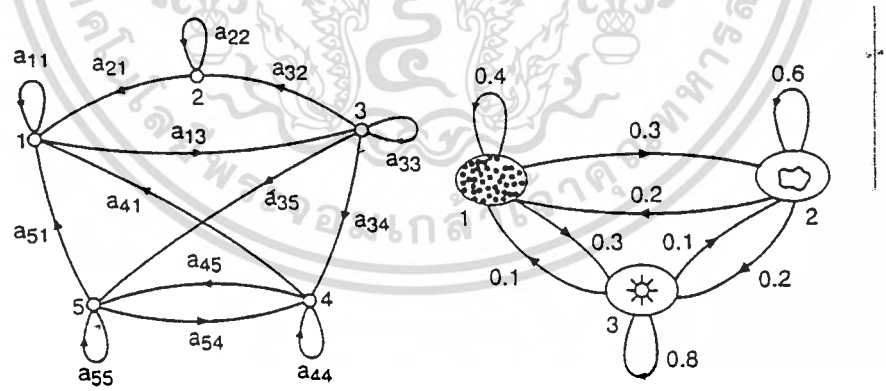
แบบจำลองมาร์คอฟ
(Hidden Markov Models)

แบบจำลองมาร์คอฟ เป็นแบบจำลอง (model) ทางสถิติซึ่งพัฒนามาเพื่อแบ่งกลุ่มของอนุกรมทางเวลา หรือสัญญาณที่ไม่คงที่ นั่นคือ ใช้สำหรับจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknown signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่ง แบบจำลองมาร์คอฟ ได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียงพูด และเป็นวิธีการที่ปริญญานิพนธ์นี้เลือกใช้

แบบจำลองมาร์คอฟ แบ่งออกเป็น 2 ประเภทคือแบบต่อเนื่อง(Continuous) และแบบไม่ต่อเนื่อง (Discrete-time) ในปริญญานิพนธ์นี้ได้เลือกใช้แบบไม่ต่อเนื่อง เพราะเป็นวิธีการที่ซับซ้อนน้อยกว่า และใช้งานได้ดีกับคำพูดสั้นๆ ที่ทำการทดลองในปริญญานิพนธ์นี้

5.1 กระบวนการ แบบจำลองมาร์คอฟ แบบไม่ต่อเนื่องทางเวลา

รูปที่ 5.1 เป็นรูปที่แสดงระบบที่มี 5 สเตท(state) การอธิบายระบบ ณ เวลาหนึ่งๆ จะอธิบายด้วยเซ็ทของดัชนีสเตท {1,2,...,N} โดยในแต่ละช่วงเวลา สามารถเปลี่ยนสเตทไป ด้วยเซ็ทของความน่าจะเป็นที่เกี่ยวข้องกับสเตท



รูปที่ 5.1 ตัวอย่างแบบจำลองมาร์คอฟ 5 สเตท

เราจะแทนค่าเวลาที่เกี่ยวข้องกับการเปลี่ยนสเตทไป ด้วย t = 1,2,...,T และแทนสเตทที่เวลา t ด้วย q_t ในห่วงโซ่ Markov (Markov Chain) นั้นจะแสดงความน่าจะเป็นในการย้ายสเตทไป โดยให้ความสนใจเพียงสเตทก่อนหน้านั้น เท่านั้น คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P[q_t=j \mid q_{t-1}=i,\ q_{t-2}=k,\ldots]=P[q_t=j \mid q_{t-1}=i] \tag{5.1}$$

และ a_{ij} จะแทนความน่าจะเป็นของการย้ายสเตท จาก i ไป j

$$a_{ij}=P[q_t=j \mid q_{t-1}=i] \quad 1 \leq i.,j \leq N \tag{5.2}$$

โดยมีคุณสมบัติว่า

$$a_{ij} \geq 0 \qquad \forall j,i \tag{5.3}$$

$$\sum_{j=1}^{N} a_{ij}=1 \qquad \forall i \tag{5.4}$$

ถ้าหากเอาท์พุท (output) ของกระบวนการเป็นเซ็ทของสเตท ที่แต่ละเวลาหนึ่ง ซึ่งแต่ละสเตทหมายถึง เหตุการณ์ที่สังเกตได้แล้วจะเรียกมาร์คอฟแบบจำลองนี้ว่า มาร์คอฟแบบจำลองที่สังเกตเห็นได้ (Observable Markov Model)

ส่วน แบบจำลองมาร์คอฟ นั้น เหตุการณ์จะเป็นฟังก์ชันของความน่าจะเป็นของสเตท คือ จะไม่ทราบเหตุการณ์ที่จะเกิดในแต่ละสเตทโดยตรง แต่จะรู้เหตุการณ์ได้โดยวิธีทางสถิติ ซึ่งจะทำให้เราได้ลำดับของเหตุการณ์ออกมา

5.2 ส่วนประกอบของ แบบจำลองมาร์คอฟ

1. N : จำนวนสเตทในแบบจำลอง โดยทั่วไปแล้วสเตทจะเกี่ยวข้องซึ่งกันและกัน ในแง่ที่ว่า **สเตทจะ**สามารถย้ายมาได้จากสเตทอื่นๆ เราจะให้เซ็ทของสเตทเป็น {1,2,...,N} และแทนสเตทที่เวลา t เป็น q_t

2. M: จำนวนของเหตุการณ์ต่างๆ ต่อหนึ่งสเตท เราจะแทนแต่ละสัญลักษณ์ เป็น $V=\{v_1,v_2,\ldots,v_M\}$

3. ความน่าจะเป็นของการย้ายสเตท $A = \{a_{ij}\}$ เมื่อ $a_{ij}=P[q_t=j \mid q_{t-1}=i]$ $1 \leq i.,j \leq N$

สำหรับกรณีพิเศษที่สเตทใดๆ จะย้ายไปยังสเตทอื่นๆ เราจะได้ $a_{ij} \geq 0$ สำหรับทุกๆ i,j สำหรับ **แบบ**จำลองมาร์คอฟ ชนิดอื่นๆ เราจะได้ $a_{ij} = 0$ สำหรับคู่ (i,j) หนึ่งคู่ หรือมากกว่า

4. การกระจายความน่าจะเป็นของสัญลักษณ์ที่สังเกตได้

$B= \{ b_j(k) \}$

โดย $b(k) = P[o_t=v_k \mid q_t=j],\ 1 \leq k \leq m$ (5.5)

เป็นนิยามการกระจายสัญลักษณ์ในสเตท j, j=1,2,...,N

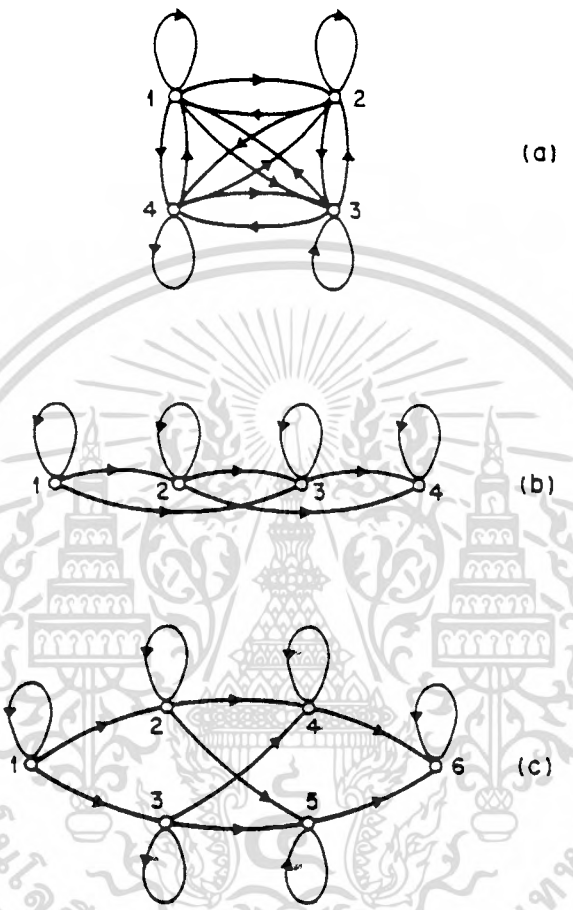
จะเห็นได้ว่า แบบจำลองมาร์คอฟ ต้องการพารามิเตอร์ของแบบจำลอง 2 ตัว คือ N และ M และกลุ่มของค่าความน่าจะเป็น A,B,π และ เพื่อความสะดวกจะใช้สัญลักษณ์เป็น (A,B,π) เพื่อแสดงเซ็ทของพารามิเตอร์ที่สมบูรณ์ของแบบจำลอง

เอกสารนี้ใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 ชนิดของ HMM

แบ่งชนิดตามลักษณะของการย้ายสเตท (Transition) ของเมตริกซ์ A

1. HMM แบบ Egordic Model หรือ Fully Connected Model : การย้ายสเตทสามารถย้ายไปยังทุก ๆ สเตทของแบบจำลอง ดังรูปที่ 5.2(a) เป็นตัวอย่างของแบบจำลองที่มี N= 4



รูปที่ 5.2 แสดงแบบจำลองต่าง ๆ ของ HMM

ซึ่งจากรูปนี้มี $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$

2. HMM แบบ Left - Right Model หรือ Bakis Model : การย้ายสเตทจะย้ายจากซ้ายไปขวา ซึ่งจะมีคุณสมบัติของสัมประสิทธิ์การย้ายสเตทดังนี้

$$w_j = 0 \quad , \quad j < i$$

คือจะไม่มีการย้ายสเตทที่ต่ำกว่าสเตทปัจจุบัน และนอกจากนี้ก็ยังมีความน่าจะเป็นของสเตทเริ่มต้นดังนี้

เอกสารนี้เป็นลิขสิทธิ์ของ ... ไม่อนุญาตให้นำไปใช้ในการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

คือลำดับของสเตทต้องเริ่มที่สเตท 1

และ Left - Right Model นี้มักจะมีกฎบังคับการย้ายสเตท เพื่อไม่ให้มีการเปลี่ยนแปลงดัชนีของสเตทมากนัก กล่าวคือ

$$a_{ij} = 0 \quad i > i + \Delta i$$

ดังรูปที่ 5.2 (b) ค่าของ $\Delta i = 2$ คือจะไม่กระโดด (ย้าย) ไปเกิน 2 สเตท จะได้เมตริกซ์การย้ายสเตทเป็น

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

จะเห็นว่าสเตทสุดท้าย สัมประสิทธิ์การย้ายสเตทจะเป็น

$$a_{NN} = 1$$

$$a_{Ni} = 0 \ , \ i < N$$

แบบจำลอง แบบนี้จะเหมาะกับสัญญาณที่มีลักษณะเปลี่ยนแปลงตามเวลาอย่างต่อเนื่อง เช่น เสียงพูด

3. HMM แบบ parallel Left - Right Model : เป็นแบบจำลอง ที่มีความยืดหยุ่นมากกว่าแบบที่ 2 Left - Right Model ดังรูปที่ 5.2 (c)

5.4 ปัญหาพื้นฐานของ แบบจำลองมาร์คอฟ

ปัญหาที่ 1

เมื่อกำหนดลำดับเหตุการณ์ $O=(o_1 o_2 \ldots o_T)$ และแบบจำลอง $\lambda = (A,B,\pi)$ จะคำนวณหา $P(O, \lambda)$ ซึ่งเป็นความน่าจะเป็นของลำดับเหตุการณ์นี้ได้อย่างไร การแก้ปัญหานี้จะทำให้เราสามารถเลือกแบบจำลองที่สอดคล้องกับเหตุการณ์ที่ดีที่สุดได้ จากแบบจำลองหลายอันที่มีอยู่

ปัญหาที่ 2

เมื่อกำหนดลำดับเหตุการณ์ $O=(o_1 o_2 \ldots o_T)$ และแบบจำลอง $\lambda = (A,B,\pi)$ เราจะเลือกลำดับสเตท

เอกสารนี้เป็น ... เพื่อการศึกษาเท่านั้น ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$q=(q_1 q_2 ....q_T)$ ที่เหมาะกับเหตุการณ์ที่สุดได้อย่างไร (อธิบายเหตุการณ์ได้ดีที่สุด) การแก้ปัญหานี้ช่วยหาลำดับของสเตทที่ถูกต้อง

ปัญหาที่ 3

เราจะปรับพารามิเตอร์ของแบบจำลอง λ =(A,B,π) เพื่อให้ได้ค่าสูงสุดของ $P(O|\lambda)$ ได้อย่างไร การแก้ปัญหานี้ช่วยให้ได้ λ ที่เหมาะที่สุด เพื่ออธิบายการได้มาซึ่งลำดับเหตุการณ์ที่กำหนดให้ อย่างดีที่สุด

ลำดับเหตุการณ์ที่ใช้ปรับค่าพารามิเตอร์ของแบบจำลอง เรียกว่าลำดับเทรนนิ่ง (training sequence) เนื่องจากเป็นลำดับเหตุการณ์ที่ถูกใช้เพื่อเทรนนิ่ง แบบจำลองมาร์คอฟ และด้วยการแก้ปัญหานี้เราจะได้แบบจำลองที่ดีที่สุดสำหรับคำนั้น

ตัวอย่าง สำหรับตัวรู้จำเสียงพูด (isolate word speech recognizer) ง่ายๆ เช่น สำหรับคำว่า 'หนึ่ง' เราต้องการออกแบบ แบบจำลองมาร์คอฟ แบบ N สเตท เราจะแทนสัญญาณคำพูดของคำด้วยลำดับเวลาของเวคเตอร์สเปคตรัม สมมติว่าข้อมูลได้ถูกเข้ารหัสมาแล้วโดยใช้ สเปคตรัม โค้ดบุค (spectral codebook) ที่มีเวคเตอร์สเปคตรัมอยู่ M แบบ นั่นคือ แต่ละเหตุการณ์แทนด้วยดัชนีของเวคเตอร์สเปคตรัมที่ใกล้เคียงกับสัญญาณเสียงเดิมที่สุด ดังนั้นแต่ละคำศัพท์จะแทนด้วยลำดับเทรนนิ่ง ซึ่งประกอบด้วย ลำดับของดัชนีโค้ดบุคของคำจำนวนมาก สิ่งแรกที่จะต้องทำคือการสร้างแบบจำลองของแต่ละคำ โดยใช้ การแก้ปัญหาที่ 3 หาพารามิเตอร์ λ ของแบบจำลอง ที่เหมาะสมที่สุด สำหรับแต่ละแบบจำลองของคำ จากนั้นใช้การแก้ปัญหาที่ 2 เพื่อแบ่งแต่ละลำดับเทรนนิ่งของคำไปยังสเตทต่างๆ และเรียนรู้ถึงคุณสมบัติของเวคเตอร์สเปคตรัม ซึ่งจะทำให้ทราบถึงเหตุการณ์ที่เกิดขึ้นในแต่ละสเตท จุดประสงค์ก็เพื่อทำให้ แบบจำลองมีความสามารถ ในการจำลองเสียงที่พูดเข้าไปนั้นเอง

เมื่อได้ แบบจำลองมาร์คอฟ ของคำว่า 'หนึ่ง' ที่เหมาะสมแล้ว ก็จะใช้การแก้ปัญหาที่ 1 ในการหาความน่าจะเป็นไปได้ของแบบจำลองในการเกิดเหตุการณ์ ตามลำดับเหตุการณ์ของเสียง ที่ไม่รู้ว่าเป็นเสียงใดเข้ามา ถ้าแบบจำลองใดของคำว่า 'หนึ่ง' ให้ค่าความน่าจะเป็นสูงกว่าแบบจำลองของคำอื่นๆ ('สอง' , 'สาม' , 'สี่',....) แสดงว่า คำที่ไม่ทราบว่าเป็นเสียงใด นี้คือคำว่า 'หนึ่ง' นั่นเอง

5.5 การแก้ปัญหาพื้นฐานของ แบบจำลองมาร์คอฟ

5.5.1 การแก้ปัญหาที่ 1 การคำนวณหาค่าความน่าจะเป็น

เมื่อเราต้องการคำนวณหาความน่าจะเป็นของลำดับเหตุการณ์ $O=(o_1 o_2 ....o_T)$ เมื่อกำหนดแบบจำลองที่มีค่าพารามิเตอร์ λ ให้เช่น $P(O,\lambda)$ วิธีหนึ่งที่จะทำได้คือระบุสเตทที่เวลาต่างๆ จนถึงเวลา T ซึ่งเป็นไปได้ถึง N^T แบบ พิจารณาลำดับที่กำหนดไว้

$$q=(q_1 q_2 ....q_T) \quad (5.6)$$

เมื่อ q_1 เป็นสเตทเริ่มต้น ความน่าจะเป็นของลำดับเหตุการณ์ O ที่กำหนดไว้ใน คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของผลงานทุกครั้งที่มีการนำไปใช้

$$P(O \mid q,\lambda) = \prod_{t=1}^{T} P(o_t \mid q_t,\lambda) \quad (5.7)$$

ความน่าจะเป็นในการเกิดเหตุการณ์คือ

$$P(O \mid q,\lambda) = b_{q1}(o_1) \cdot b_{q2}(o_2) \ldots b(o_T) \quad (5.8)$$

ความน่าจะเป็นในการย้ายสเตท q จะเป็น

$$P(q|\lambda) = \pi_{q1} a_{q1q2} a_{q2q3} a_{qT-1\ qT} \quad (5.9)$$

เมื่อเอาความน่าจะเป็นของ O และ q มารวมกัน นั่นคือ ความน่าจะเป็นที่ O และ q จะเกิดพร้อมกัน จะได้

$$P(O,q|\lambda) = P(O|q,\lambda)P(q|\lambda) \quad (5.10)$$

ความน่าจะเป็นของ O ได้มาจากการรวมกันของความน่าจะเป็นร่วม โดยคิดจากลำดับสเตท q ที่เป็นไปได้ ดังนี้

$$P(O,q|\lambda) = \sum_{allq} P(O|q,\lambda)P(q|\lambda) \quad (5.11)$$

$$= \sum_{q1,q2,\ldots q_T} \pi_{q1} b_{q1}(o_1)\, a_{q1q2}\, b_{q2}(o_2) a_{q2q3} \ldots a_{qT-1\ qT}\, b_{qT}(o_T) \quad (5.12)$$

เนื่องจากสมการดังกล่าวมีการคำนวณที่ยุ่งยาก ในรูปลำดับ $2T \cdot N^T$ ซึ่งแม้ค่า N และ T จะเป็นค่าน้อยๆ ดังนั้นจึงได้มีวิธีที่ง่ายกว่า มาช่วยแก้ปัญหาที่ 1 นี้ได้

5.5.1.1 กระบวนการไปข้างหน้า (Forward Procedure)

นิยาม $\alpha_t(i) = P(o_1 o_2 \ldots o_T, q_t = i \mid \lambda)$ (5.13)

เป็นความน่าจะเป็น ของการเกิดลำดับเหตุการณ์ $o_1 o_2 \ldots o_t$ (จนกระทั่งเวลา t) และอยู่ที่สเตท i ณ เวลา t โดยมีแบบจำลองเป็น λ เราจะสามารถหา $\alpha_t(i)$ ได้ดังนี้

1. การเริ่มต้น (Initialization)

$$\alpha_t(i) = \pi_i b_i(o_1) \quad , \quad 1 \le i \le N \quad (5.14)$$

เริ่มต้นด้วยการกำหนดความน่าจะเป็นไปข้างหน้า เป็นความน่าจะเป็นร่วมของสเตท i และเหตุการณ์เริ่มต้น o_1

2. การเหนี่ยวนำ (Induction)

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i)\, a_{ij}]\, b_j(o_{t+1}) \quad , \quad 1 \le t \le T-1 \quad (5.15)$$

$$1 \le j \le N$$

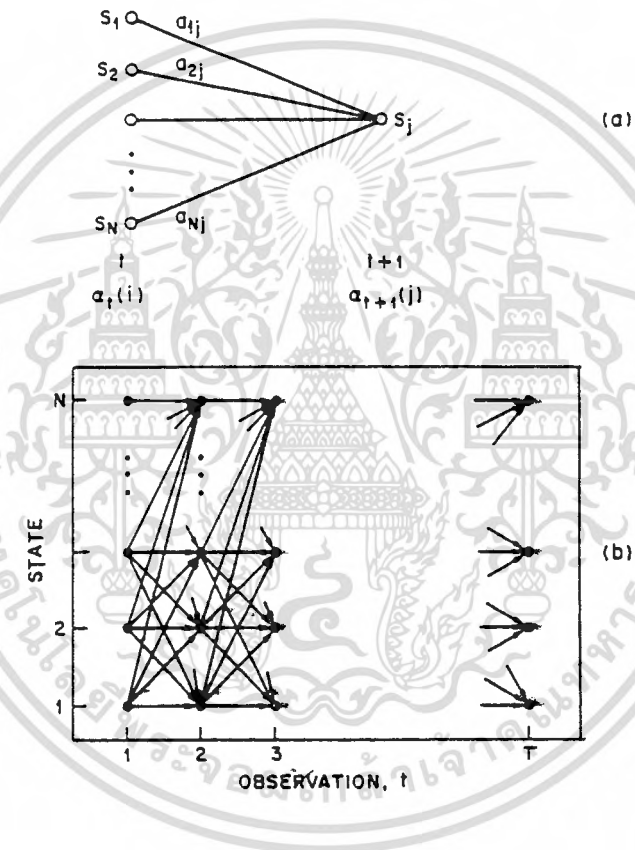
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงถึงว่า สเตท j ที่เวลา t+1 มาได้จากสเตท i ที่เป็นไปได้ถึง N สเตท , $1 \leq i \leq N$ ที่เวลา t ดังรูปที่ 5.3

3. <u>การสิ้นสุด (Termination)</u>

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \qquad (5.16)$$

หา $P(O|\lambda)$ ได้จากผลรวมของ $\alpha_t(i)$ จากทุกๆ i ($1 \leq i \leq N$)



รูปที่ 5.3 กระบวนการไปข้างหน้า

5.5.1.2 กระบวนการถอยหลัง (Backward Procedure)

นิยาม $\beta_T(i) = P(o_{t+1}o_{t+2}....o_T, q_t=i \,|\lambda)$ (5.17)

คือความน่าจะเป็นของลำดับเหตุการณ์ส่วนหลัง จากเวลา t+1 ไปจนจบ โดยกำหนดว่าต้องอยู่ที่สเตท i ที่เวลา t และมีแบบจำลองเป็น λ

เราจะคำนวณหา $\beta_T(i)$ ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

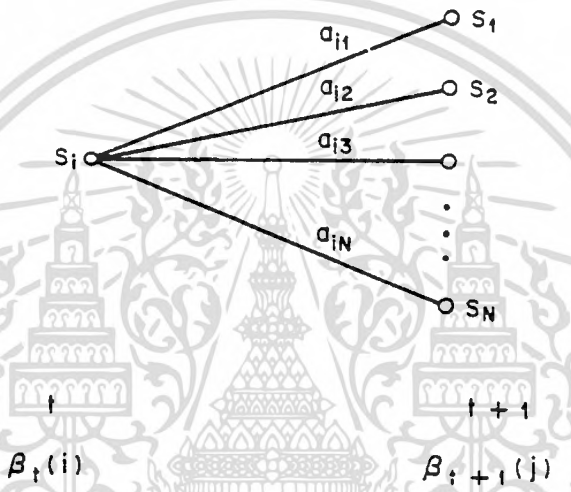
1. <u>การเริ่มต้น (Initialization)</u>

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \tag{5.18}$$

2. <u>การเหนี่ยวนำ(Induction)</u>

$$\beta_T(i) = \sum_{j=1}^{N} a_{ij}\ b_j(o_{t+1})\beta_{t+1}(i) \ ,$$

$$t=T-1,T-2,\ldots 1 \ , \quad 1 \leq i \leq N \tag{5.19}$$



รูปที่ 5.4 กระบวนการถอยหลัง

รูปที่ 5.4 แสดงว่า เพื่อที่เหตุการณ์ จะอยู่ที่สเตท i ที่เวลา t โดยคำนึงถึงลำดับเหตุการณ์จากเวลา t+1 เราต้องพิจารณาสเตท j ที่จะเป็นไปได้ทั้งหมด ณ เวลา t+1 โดยจะขึ้นอยู่กับค่า a_{ij} และ $b_j(o_{t+1})$

5.5.2 การแก้ปัญหาที่ 2 การหาลำดับสเตทที่เหมาะสม

เราจะใช้วิธี วิทเทอร์บี อัลกอริทึม (Viterbi Algoritym) เพื่อหาลำดับสเตทที่ดีที่สุด ณ เวลา t หนึ่ง ๆ เมื่อกำหนดลำดับเหตุการณ์ $O = (o_1 o_2 \ldots o_t)$ โดยนิยาม

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_t} P[q_1 q_2 \ldots q_{t-1}, \ q_t=i, \ o_1 o_2 \ldots o_t \mid \lambda] \tag{5.20}$$

$\max_i$ คือ $\delta_t(i)$ เป็นความน่าจะเป็นสูงสุดของเส้นทาง (path) ณ เวลา t ซึ่งเริ่มนับจากเหตุการณ์ที่เวลาเริ่มต้นจนถึงเวลา t ที่สเตท i และ โดยการเหนี่ยวนำ (induction) เราจะได้

$$\delta_{t+1}(i) = [\max_i \delta_t(i)\ a_{ij}] \cdot b_j(o_{t+1}) \tag{5.21}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เราจะใช้กระบวนการต่อไปนี้ ในการหาลำดับสเตทที่ดีที่สุด โดยให้ $\psi_t(i)$ เป็น อาเรย์ (array)

1. <u>การเริ่มต้น (Initialization)</u>

$$\delta_t(i) = \pi_i b_i(o_1) \quad , \quad 1 \leq i \leq N \tag{5.22}$$

$$\psi_t(i) = 0 \tag{5.23}$$

2. <u>การย้อนกลับ (Recursion)</u>

$$\delta_t(j) = [\max_{1 \leq i \leq N} \delta_{t-1}(i)\, a_{ij}] \cdot b_j(o_t), \qquad 2 \leq t \leq T \tag{5.24}$$

$$1 \leq j \leq N$$

$$\psi_1(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i)\, a_{ij}] \quad 2 \leq t \leq T \tag{5.25}$$

$$1 \leq j \leq N$$

3. <u>การสิ้นสุด (Termination)</u>

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \tag{5.26}$$

$$q_T^* = \max_{1 \leq i \leq N} \delta_T(i)] \tag{5.27}$$

4. <u>เส้นทางเดินย้อนกลับ (Path backtracking)</u>

$$q_t^* = \psi_{t+1}(q^*_{t+1}) \qquad t=T-1, T-2, \ldots, 1 \tag{5.28}$$

5.5.3 การแก้ปัญหาที่ 3

เราจะเลือกค่า $\lambda=(A,B,\pi)$ ที่ทำให้ $P(O|\lambda)$ มีค่าสูงสุด โดยใช้ กระบวนการทำซ้ำ (Iterative) ซึ่งวิธีที่เราเลือกใช้ คือวิธี บาม-เวลช์ (Baum-Welch) หรือ EM (Expectation-Maximization)

นิยาม

1. $\gamma_t(i) = P(q_t=i \mid O,\lambda)$ (5.29)

เป็นค่าความน่าจะเป็นที่จะอยู่ที่สเตท i ณ เวลา t โดยกำหนดลำดับเหตุการณ์ O และแบบจำลอง λ ให้ เราจะแสดงค่า $\gamma_t(i)$ ได้ดังนี้

$$\gamma_t(i) = P(q_t=i \mid O,\lambda) \tag{5.30}$$

$$= \frac{p(O, q_t = i|\lambda)}{p(O|\lambda)}$$

$$= \frac{p(O, q_t = i|\lambda}{\sum_{i=1}^{N} P(O, q_t = i|\lambda)}$$

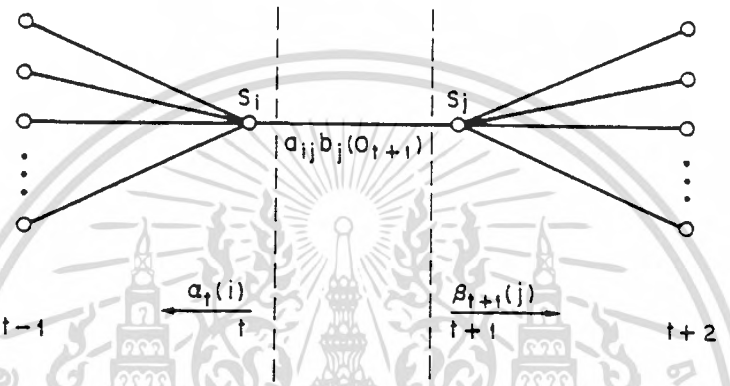
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจาก $P(O,q_t=i|\lambda)$ มีค่าเท่ากับ $\alpha_t(i)\beta_t(i)$ ดังนั้นเราจะได้ว่า

$$\gamma_t(i) = \frac{\alpha(i)\beta(i)}{\sum_{i=1}^{N}\alpha(i)\beta(i)} \tag{5.31}$$

2. $\varepsilon_t(i,j) = P(q_t=i, q_{t+1}=j \mid O,\lambda)$ (5.32)

เป็นความน่าจะเป็น ที่จะอยู่ที่สเตท i ที่เวลา t และสเตท j ที่เวลา t+1 เมื่อกำหนดแบบจำลองและลำดับเหตุการณ์ให้ ดังแสดงในรูปที่ 5.5



รูปที่ 5.5 แสดงลำดับการคำนวณเหตุการณ์ร่วมซึ่งระบบจะอยู่ในสเตท i ที่เวลา t และอยู่ที่สเตท j ที่เวลา t+1

ซึ่งจากนิยามของตัวแปรไปข้างหน้าและตัวแปลย้อนกลับ จะเขียน $\varepsilon_t(i,j)$ ได้ดังนี้

$$\varepsilon_t(i,j) = \frac{P(q_t=i, q_t+1=j, O|\lambda)}{P(O|\lambda)}$$

$$= \frac{\alpha(i)aijbj(o_t+1)\beta_t+1(j)}{P(O|\lambda)} \tag{5.33}$$

$$= \frac{\alpha(i)aijbj(o_t+1)\beta_t+1(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha(i)aijbj(o_t+1)\beta_t+1(j)}$$

จากที่เราได้นิยาม $\gamma_t(i)$ แล้ว จะนำมาสัมพันธ์กับ $\varepsilon_t(i,j)$ ได้

$$\gamma_t(i) = \sum_{j=1}^{N} \varepsilon_t(i,j) \tag{5.34}$$

และ

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{จำนวนของการย้ายสเตทจากสเตท i ในเหตุการณ์ O} \tag{5.35}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$\sum_{t=1}^{T-1} \varepsilon_t(i,j)$ = จำนวนของการย้ายสเตทจากสเตท i ไป j ในเหตุการณ์ O (5.36)

ดังนั้นเราจะสามารถหาเซ็ทของพารามิเตอร์ได้ดังนี้

π' = จำนวนครั้งในการอยู่ที่สเตท i ที่เวลา t = 1

$= \gamma_t(i)$ (5.37)

a_{ij}' = จำนวนครั้งที่คาดไว้ของการย้ายจากสเตท i ไป j / จำนวนครั้งที่คาดว่าจะย้ายจากสเตท I

$$= \frac{\sum_{t=1, o=v_k}^{T} n(j)}{\sum_{t=1}^{T-1} n(i)} \quad (5.38)$$

bj(k) = จำนวนครั้งที่คาดว่าจะอยู่ในสเตท j และเกิดเหตุการณ์ V_k / จำนวนครั้งที่คาดว่าจะอยู่ที่สเตท j

$$= \frac{\sum_{t=1, o=v_k}^{T} n(j)}{\sum_{t=1}^{T} n(j)} \quad (5.39)$$

เราจะให้ $\lambda=(A,B,\pi)$ เป็นแบบจำลองปัจจุบัน และจะใช้ λ นี้ คำนวณในด้านขวาของสมการ และให้แบบจำลองที่ได้จากการคำนวณซ้ำๆ เป็น $\lambda'=(A',B',\pi')$ เป็นแบบจำลองที่ได้จากข้างซ้ายของสมการ เราจะได้จุดวิกฤตของฟังก์ชันความน่าจะเป็นในกรณีที่ $\lambda'=\lambda$ หรือถ้า λ' มีความน่าจะเป็นมากกว่าแบบจำลอง λ ในลักษณะที่ $P(O|\lambda')>P(|\lambda)$ นั่นคือ เราก็จะได้ แบบจำลอง λ' ใหม่ ที่น่าจะทำให้เกิดลำดับเหตุการณ์ O ได้มากกว่า

จากกระบวนการข้างต้น ถ้าเราคำนวณซ้ำๆ โดยใช้ λ' แทน λ แล้ว จะทำให้ความน่าจะเป็นของการเกิดเหตุการณ์ O จะดีขึ้น จนกระทั่งถึงจุดวิกฤต ซึ่งทั้งหมดนี้ก็คือการประมาณค่าความน่าจะเป็นสูงสุด (ML estimate) ของ แบบจำลองมาร์คอฟ นั่นเอง

5.6 การปรับปรุงค่าพารามิเตอร์ของ HMM

5.6.1 การสเกลลิ่ง (Scaling)

เนื่องจาก $\alpha_t(i)$ จะประกอบด้วยผลรวมของเทอมจำนวนมาก ซึ่งแต่ละเทอม คือ

$$\left(\prod_{s=1}^{t-1} a_{q_s q_{s+1}} \prod_{s=1}^{t} b_{q_s}(O_s) \right)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น ทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ $q_t = i$ และ b เป็นความน่าจะเป็นในสมการ (5.5) เนื่องจากแต่ละเทอม a และ b มีค่าน้อยกว่า 1 แสดงว่าเมื่อ t มากขึ้น แต่ละเทอมของ $\alpha_t(i)$ จะเข้าสู่ศูนย์ ซึ่งถ้า t มากขึ้นเรื่อย ๆ Dynamic Range ของการคำนวณ $\alpha_t(i)$ เกิน Range ของคอมพิวเตอร์ ทำให้ค่าที่ได้ไม่ถูกต้อง จึงได้มีการสเกลลิ่งขึ้น

การสเกลลิ่งทำได้โดยการคูณ $\alpha_t(i)$ โดยสัมประสิทธิ์การสเกลลิ่ง ซึ่งไม่ขึ้นกับ i เพื่อทำให้ $\alpha_t(i)$ อยู่ภายใน Dynamic Range ของคอมพิวเตอร์ $1 \leq t \leq T$ การสเกลลิ่ง $B_t(i)$ ก็เช่นเดียวกัน หลังการคำนวณค่าการสเกลลิ่ง ก็จะตัดกันหมดไปเอง

พิจารณาดังนี้ จาก สมการ (5.38)

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \qquad (5.40)$$

ในการคำนวณ $\alpha_t(i)$ เราใช้ $\alpha_t(i)$ แทน α ที่ยังไม่ได้สเกลลิ่ง

$\hat{\alpha}_t(i)$ เป็น α ที่สเกลลิ่งแล้ว

และ $\hat{\hat{\alpha}}_t(i)$ แทนเวอร์ชันของ α ก่อนการสเกลลิ่ง

เมื่อ $t = 1$ จะได้ $\hat{\alpha}_t(i) = c_2 \alpha_1(i)$

เมื่อ $c_1 = \dfrac{1}{\sum_{i=1}^{N} \alpha_t(i)}$

เมื่อ $2 \leq t \leq T$ คำนวณ $\hat{\hat{\alpha}}_t(i)$ จาก สมการ 6.20 ในเทอมของ $\hat{\alpha}_t(i)$ ค่าก่อน

$$\hat{\hat{\alpha}}_t(i) = \sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t) \qquad (5.41a)$$

เมื่อ c_t สัมประสิทธิ์ การสเกลลิ่ง เป็น

$$c_t = \frac{1}{\sum_{i=1}^{N} \hat{\hat{\alpha}}(i)} \qquad (5.41b)$$

โดยให้ $\hat{\alpha}_t(i) = c_t \hat{\hat{\alpha}}_t(i)$ (5.41c)

จากสมการ (5.41) จะเขียนได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t)} \tag{5.42}$$

โดย การเหนี่ยวนำ จะได้

$$\hat{\alpha}_{t-1}(j) = (\prod_{\tau=1}^{t-1} c_\tau)\alpha_{t-1}(j) \tag{5.43a}$$

จะเขียนได้ว่า

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \alpha_{t-1}(j)(\prod_{\tau=1}^{t-1} c_\tau) a_{ji} b_i(O_t)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_{t-1}(j)(\prod_{\tau=1}^{t-1} c_\tau) a_{ji} b_i(O_t)} = \frac{\alpha_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \tag{5.43b}$$

นั่นคือจะสเกล $\alpha_t(i)$ ได้โดยหารด้วยผลรวมของ $\alpha_t(i)$ ทั้งหมด และสเกล $B_t(i)$ ด้วยค่าเดียวกันนี้ ในเทอมของการสเกลนี้สมการ (5.4) จะเป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{ij} b_j(O_{t+1}) \hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N} \hat{\alpha}_t(i) a_{ij} b_j(O_{t+1}) \hat{\beta}_{t+1}(j)} \tag{5.44}$$

โดยแต่ละ $\alpha_t(i), \hat{\beta}_{t+1}(j)$ เขียนได้เป็น

$$\hat{\alpha}_t(i) = \left[\prod_{s=1}^{t} c_s\right] \alpha_t(i) = C_t \alpha_t(i) \tag{5.45}$$

$$\hat{\beta}_{t+1}(j) = \left[\prod_{s=t+1}^{T} c_s\right] \beta_{t+1}(j) = D_{t+1} \beta_{t+1}(j) \tag{5.46}$$

ดังนั้นสมการ (5.44) จะเขียนได้เป็น

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} C_t \alpha_t(i) a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N} C_t \alpha_t(i) a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)} \tag{5.47}$$

ซึ่งเทอม $C_t D_{t+1}$ จะเขียนได้ในเทอม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$C_t D_{t+1} = \prod_{s=1}^{t} c_s \prod_{s=t+1}^{T} c_s = \prod_{s=1}^{T} c_s = C_T \tag{5.48}$$

ซึ่งไม่ขึ้นกับเวลา t ดังนั้น C_tD_{t+1} จะถูกตัดทิ้ง ทั้งเศษและส่วนของสมการ (5.47) ซึ่งทำให้ได้สูตรการคำนวณซ้ำ ๆ (reestimate) เดิมกลับคืนมา

กระบวนการ สเกลลิ่ง ดังกล่าวนี้ก็ใช้ได้กับการคำนวณซ้ำ ๆ สัมประสิทธิ์ B และ π

ในการสเกลลิ่งจะทำให้การคำนวณค่า $P(O|\lambda)$ เปลี่ยนไป เราจะไม่สามารถหาได้จากการรวมกับของเทอม $\hat{\alpha}_T(i)$ แต่จะหาจากคุณสมบัติ

$$\prod_{t=1}^{T} c_t \sum_{i=1}^{N} \alpha_T(i) = C_T \sum_{i=1}^{N} \alpha_T(i) = 1 \tag{5.49}$$

ดังนั้นจะได้

$$\prod_{t=1}^{T} c_t P(O|\lambda) = 1 \tag{5.50}$$

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^{T} c_t} \tag{5.51}$$

$$\log[P(O|\lambda)] = \sum_{t=1}^{T} \log c_t \tag{5.52}$$

ซึ่งจะได้ค่า log ของ P ไม่ใช่ค่าของ P เพราะจะเกิน dinamic range ของคอมพิวเตอร์

5.6.2 ลำดับของเหตุการณ์หลายลำดับ (Multiple Observation Sequence)

ในการใช้แบบจำลองแบบ Left-Right นั้น ไม่สามารถใช้ลำดับเหตุการณ์เพียงลำดับเดียวในการเทรนแบบจำลองได้ ต้องใช้หลาย ๆ เหตุการณ์เข้ามาเทรน เพื่อให้ได้ค่าพารามิเตอร์ที่ถูกต้องมากขึ้น

ถ้าให้เซตของ v ลำดับ เหตุการณ์ เป็น

$O=[O^{(1)},O^{(2)},.....,O^{(V)}]$

เมื่อ $O^{(V)}=O_1^{(V)}O_2^{(V)}.....O_{Tv}^{(V)}$ เป็นลำดับเหตุการณ์ลำดับที่ v เราสมมุติว่าแต่ละลำดับ เหตุการณ์เป็นอิสระต่อกัน จุดมุ่งหมายก็คือปรับค่าพารามิเตอร์ของ λ ให้ดีที่สุด

$$P(O|\lambda) = \prod_{v=1}^{V} P(O^{(V)}|\lambda) \tag{5.53}$$

$$= \prod_{v=1}^{V} P_v \tag{5.54}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากสูตรการคำนวณซ้ำ ๆ มาจากพื้นฐานของความน่าจะเป็นของการเกิดของเหตุการณ์ต่างๆ สูตรการคำนวณซ้ำของลำดับเหตุการณ์หลายๆ ลำดับ ก็จะหาได้จากการเอาความถี่ของการเกิดเหตุการณ์ต่างๆ แต่ละเหตุการณ์มารวมกัน ดังนั้นจะได้สูตรหา $a_{ij}b_j(k)$ เป็น

$$\overline{a_{ij}} = \frac{\sum_{v=1}^{V}\frac{1}{P_v}\sum_{t=1}^{T_v-1}\alpha_t^v(i)a_{ij}b_j(O_{t+1}^v)\beta_{t+1}^v(j)}{\sum_{v=1}^{V}\frac{1}{P_v}\sum_{t=1}^{T_v-1}\alpha_t^v(i)\beta_t^v(i)} \tag{5.55}$$

$$\overline{b_j}(k) = \frac{\sum_{v=1}^{V}\frac{1}{P_v}\sum_{t=1 O_t=k}^{T_v-1}\alpha_t^v(i)\beta_t^v(i)}{\sum_{v=1}^{V}\frac{1}{P_v}\sum_{t=1}^{T_v-1}\alpha_t^v(i)\beta_t^v(i)} \tag{5.56}$$

ส่วน π_iไม่ต้องคำนวณเนื่องจาก $\pi_1=1$, $\pi_i=0, i\neq 1$

การสเกลลิ่ง ที่เหมาะสมของสมการ(5.55)-(5.56)

$$\overline{a_{ij}} = \frac{\sum_{v=1}^{V}\sum_{t=1}^{T_v-1}\hat{\alpha}_t^v a_{ij}b_j(O_{t+1}^v)\hat{\beta}_{t+1}^v(j)}{\sum_{v=1}^{V}\sum_{t=1}^{T_v-1}\sum_{j=1}^{N}\hat{\alpha}_t^v(i)a_{ij}b_j(O_{t+1}^v)\hat{\beta}_{t+1}^v(j)} \tag{5.57}$$

$$\overline{b_j}(k) = \frac{\sum_{v=1}^{V}\sum_{t=1 O_t=k}^{T_v-1}\hat{\alpha}_t^v a_{ij}b_j(O_{t+1}^v)\hat{\beta}_{t+1}^v(j)}{\sum_{v=1}^{V}\sum_{t=1}^{T_v-1}\sum_{j=1}^{N}\hat{\alpha}_t^v(i)a_{ij}b_j(O_{t+1}^v)\hat{\beta}_{t+1}^v(j)} \tag{5.58}$$

5.7 ระบบแบบจำลองมาร์คอฟ

เมื่อเรามีคำศัพท์อยู่ V คำ ในการทำการรู้จำได้ เราจะต้องทำการสร้างแบบจำลองของคำ แต่ละคำด้วย แบบจำลองมาร์คอฟ ที่แตกต่างกัน คำแต่ละคำจะมี ลำดับเทรนนิ่งที่ได้มาจากคุณลักษณะเฉพาะของคำศัพท์นั้น ๆ

การที่เราจะรู้จำคำพูดได้ต้องทำได้ ดังนี้

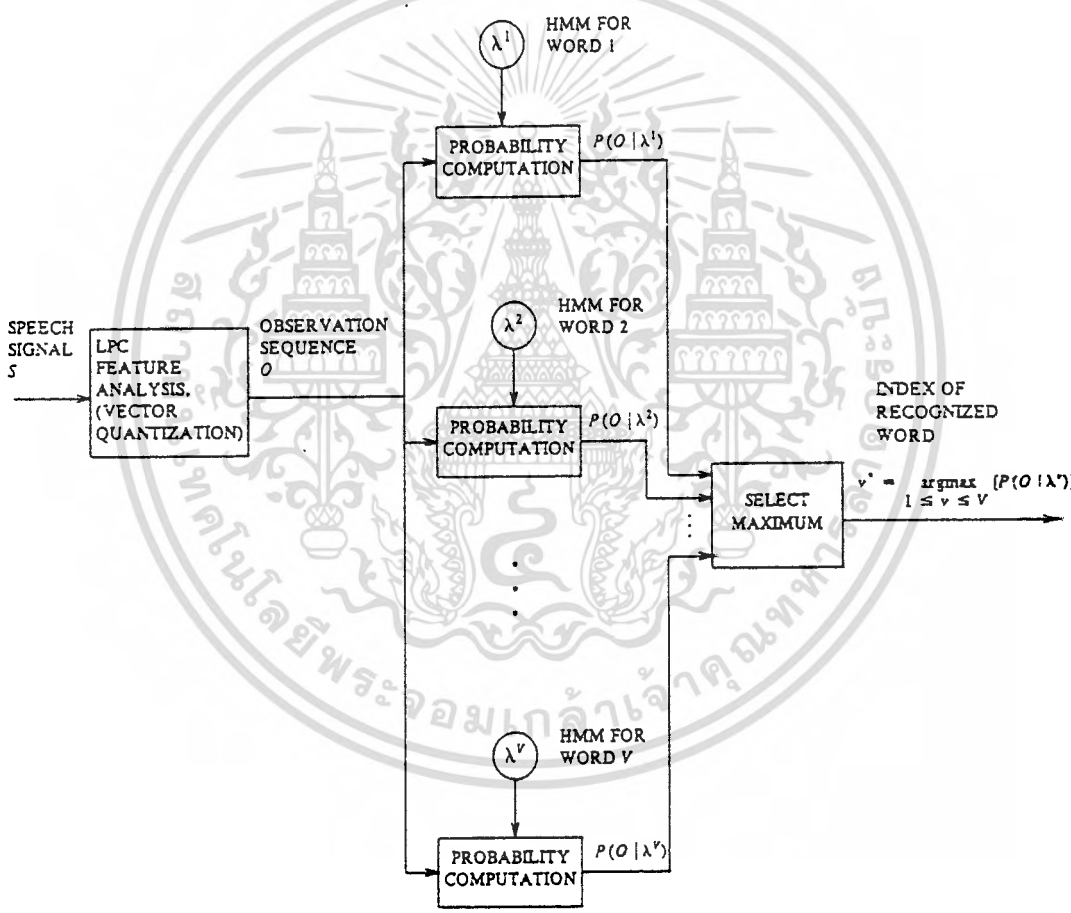
1. สำหรับแต่ละคำศัพท์ V เราต้องสร้าง แบบจำลองมาร์คอฟ :λ_V นั่นคือการหาค่า (A,B,π) ที่เหมาะสมกับลำดับเทรนนิ่งของคำ V นั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ในการจะรู้จำคำศัพท์แต่ละคำ (ดูรูป ที่ 5.6) หลังจากที่มีลำดับเหตุการณ์ของคำพูดเข้ามา (เสียงพูดที่ผ่าน LPC และ การจัดระดับเวคเตอร์ มาแล้ว) เราจะทำการหาค่า $P(O|\lambda)$ ของทุกๆ แบบจำลอง แล้วเลือกแบบจำลองที่มีค่าความน่าจะเป็นในการเกิดเหตุการณ์สูงสุด คือ

$$V^* = \arg\max P(O|\lambda_v) \qquad (5.59)$$

คำศัพท์ที่สอดคล้องกับแบบจำลองดังกล่าวนี้ จะเป็นคำเดียวกับคำศัพท์ที่เราต้องการจะรู้จำนั่นเอง โดยขั้นตอนการคำนวณหาค่าความน่าจะเป็น จะใช้วิธีวิทเทอร์บี ดังที่ได้กล่าวแล้วในตอนต้น



รูปที่ 5.6 บล็อคไดอะแกรม ของการรู้จำคำโดดด้วยแบบจำลองของมาร์คอฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
ขั้นตอนการทดลอง

ปริญญานิพนธ์นี้ได้พัฒนาโปรแกรมในส่วนต่าง ๆ ดังนี้

- ส่วนของ Linear Predictive Coding (LPC)
- ส่วนของ Vector Quantization (VQ)
- ส่วนของ Hidden Markov Model (HMM)
- ส่วนของ Viterbi Algorithm (VA)

วิธีการดำเนินการทดลอง

บันทึกเสียงพูดคำว่า 0-9 ของกลุ่มบุคคลที่ใช้เป็นเสียงต้นแบบและใช้ทดสอบเบื้องต้น เสียงละ 5 ครั้งจำนวน 40 คน เพื่อใช้ในการหาแบบจำลองของเสียง โดยข้อมูลทั้งหมดใช้ความถี่ในการสุ่มสัญญาณ 11.025 กิโลเฮิรตซ์ จากนั้นนำข้อมูลที่บันทึกไว้ จัดการปรับแต่งให้เหมาะสม โดยการตัดส่วนหัวและส่วนท้ายคำพูดที่เป็นส่วนเกินออก เนื่องจากเป็นส่วนของสัญญาณรบกวน จากนั้นนำข้อมูลไปผ่านขั้นตอนต่าง ๆ ในการหาแบบจำลองของเสียงต่อไป

6.1 การทดลองในส่วนของการประมาณเชิงเส้น

เป็นการวิเคราะห์เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับโครงสร้างระบบกำเนิดเสียง โดยแบ่งสัญญาณเสียงพูดแต่ละเสียง ที่จะทำการวิเคราะห์ ออกเป็นส่วน ๆ แต่ละส่วนใช้ระยะเวลาช่วงสั้น ๆ ประมาณ 27.21 มิลลิวินาที ซึ่งช่วงนี้สัญญาณเสียงพูด จะมีการเปลี่ยนแปลงคุณลักษณะอย่างช้า ๆ จนอาจถือว่าระบบกำเนิดเสียงมีคุณลักษณะไม่เปลี่ยนแปลง ตามขั้นตอนการวิเคราะห์ค่าสัมประสิทธิ์ในบทที่ 3 ซึ่งจะได้ค่าพารามิเตอร์ 3 ประเภท สำหรับนำไปใช้ในขั้นตอนต่อไป

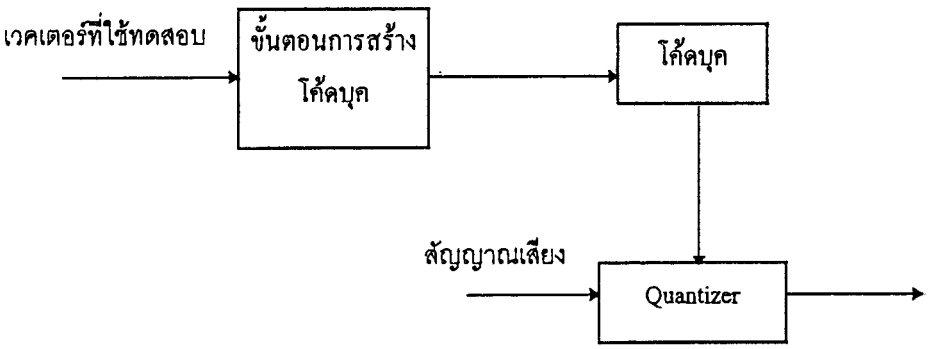
- สัมประสิทธิ์ LPC และค่าอัตราการขยาย
- สัมประสิทธิ์เซปสตรัม (Cepstrum)
- สัมประสิทธิ์ที่ได้เวทค่าแล้ว (Parameter Weighting)

6.2 การทดลองในส่วนของเวคเตอร์ควอนไตซ์เซชัน

จากรูปที่ 6.1 แสดงถึงขั้นตอนในการหาโค้ดบุค และการทดลอง โดยแบ่งเป็น 2 ขั้นตอน

ก. การหาโค้ดบุค โดยใช้ K-means Algorithm โดยมีการสุ่มค่าเริ่มต้น ซึ่งใช้ขนาดโค้ดบุคเท่ากับ 64 (ภาคผนวก) โปรแกรมที่ใช้คือ codebook.m (ภาคผนวก)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 6.1 แสดงขั้นตอนการทดลองในส่วนของเวคเตอร์ควอนไตซ์เซชัน

ข. ในการทดสอบ ใช้โปรแกรม compair.m เพื่อเปรียบเทียบเวคเตอร์เสียงอินพุทที่เข้ามากับโค้ดบุคที่มีอยู่ ให้ได้ดัชนีโค้ดบุคเพื่อนำใช้ในการสร้างแบบจำลองของ HMM ต่อไป ซึ่งใช้การหาความคลาดเคลื่อนโดยวิธี square - error

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

โดยการหาความคลาดเคลื่อนแบบ square - error ถูกใช้ทั้งตอนหาโค้ดบุคและ ตอนทดสอบ

6.2.1 การออกแบบ

ในการออกแบบ โค้ดบุค ถ้าเวคเตอร์ x มี k มิติ ถูกเปรียบเทียบกับโค้ดบุคขนาด K ค่า ขอบเขตที่เวคเตอร์ x จะอยู่ได้มี K ขอบเขต $\{C_i\}$ $(1 \leq i \leq K)$ และถูกแทนด้วย Y_i ที่เป็นเวคเตอร์โค้ดของแต่ละเซล C_i โดยเขียนแทนได้ว่า

$$q(x) = y_i$$

เมื่อ y คือค่าควอนไตซ์ของ x การหาความคลาดเคลื่อนคือ ระยะห่างระหว่างเวคเตอร์ x และ y เขียนโดย d(x,y) และสามารถหาความคลาดเคลื่อนรวมได้คือ

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เราจะบอกว่าโค้ดบุคที่ได้เหมาะสมขึ้นกับความคลาดเคลื่อนรวมต้องต่ำที่สุด การที่ความคลาดเคลื่อนรวมต่ำเนื่องจาก 2 ปัจจัย คือ

1. การใช้ nearest-neighbor

$$q(x) = y_i \; d(x, y_i) \leq d(x, y_j) \quad j \neq i \;,\; 1 \leq j \leq K$$

2. ค่าเวคเตอร์โค้ด y_i

y_i เป็นค่ากลางของ C_i หาโดยวิธีหาค่ากลาง ซึ่งขึ้นกับการวัดความคลาดเคลื่อน

6.2.2 K - means Algorithm

เป็นวิธีการหาโค้ดบุคแบบทำซ้ำ โดยพยายามแบ่งเวคเตอร์ { x(n) } ให้อยู่ในเซล { C_i } จำนวน K เซล

1. เริ่มต้น

เมื่อ m = 0 (m คือตัวเลขบอกการทำซ้ำ) เลือกค่าเริ่มต้นเวคเตอร์โค้ด { $y_i(0)$ }, ($1 \leq i \leq K$)

2. Classification

ทำการแบ่งเวคเตอร์ { x(n) } ($1 \leq n \leq L$) ไปตามเซล { C_i (m) } โดยใช้ทฤษฎี nearest- neighbor

$$x \in C_i(m)$$

$$\text{เมื่อ } d[x, y_i(m)] \leq d[x, y_j(m)] \;,\; i \neq j$$

3. ทำการปรับเวคเตอร์โค้ด

เพิ่มค่า m เป็น m +1 ปรับเวคเตอร์โค้ดทุก ๆ รอบการทำซ้ำ โดยคำนวณค่ากลางใหม่

$$y_i(m) = centroid\,(C_i(m)) \;,\; 1 \leq i \leq K$$

4.เงื่อนไข

นำค่าความคลาดเคลื่อนรวม D(m) ที่คำนวณได้เปรียบเทียบกับ D (m-1) ถ้าต่ำกว่า 0.01 (threshold) จะหยุดการทำงาน (เนื่องจากค่า 0.01 ให้ผลความคลาดเคลื่อนของโค้ดบุคกับข้อมูลที่นำมาเปรียบเทียบน้อย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.3 การทดลองในส่วนของ HMM

6.3.1 กระบวนการสร้าง HMM ของเสียงนั้นมีขั้นตอนต่าง ๆ ดังแสดงใน โฟลวชาร์ทที่ ข.4 และ ข.5 กล่าวคือ

1. สุ่มค่าเริ่มต้น a,b ส่วนค่า π นั้น จะใช้เป็น π={ 1 0 0 0 0 0 } เนื่องจากเลือกใช้แบบจำลองแบบ Left-Right

2. หาค่า α,β จากค่า a,b โดยใช้ลำดับเหตุการณ์ เข้ามาเทรนตามกระบวนการ Forward-Backward Algorithm ในปริญญานิพนธ์นี้เราใช้เหตุการณ์หลาย ๆ ลำดับ (V ลำดับ) เข้ามาเทรน เพื่อความถูกต้องมากขึ้น ตามโฟลวชาร์ทที่ ข.5

3. ทำการสเกลลิ่ง α,β ที่ได้เพื่อให้ Range ของการคำนวณอยู่ใน Range ที่คอมพิวเตอร์สามารถคำนวณได้อย่างถูกต้อง ดังได้อธิบายไว้แล้วในตอนต้น

4. หาความน่าจะเป็น (P) ที่แบบจำลอง HMM ซึ่งมีพารามิเตอร์ a ,b,π นั้นจะสร้างลำดับเหตุการณ์ที่เข้ามาเทรนโดยจุดมุ่งหมายก็เพื่อจะให้ได้ค่า a ,b,π ที่ให้ค่า P สูงสุดซึ่งหมายถึงว่าได้พารามิเตอร์ของแบบจำลองที่เหมาะสมกับคำ ๆ นั้นแล้ว

5. กระบวนการทดสอบค่าพารามิเตอร์ของแบบจำลอง เป็นการตรวจดูว่าพารามิเตอร์ของแบบจำลองที่ได้ลู่เข้าหรือยัง โดยจะดูได้จาก

(1) $P(O|\bar{\lambda}) - P(O|\lambda) \leq \varepsilon$ โดย ε เป็นค่าระดับ(threshold)ที่ตั้งไว้

(2) ดูจากการเปลี่ยนแปลงของค่าλ กับ $\bar{\lambda}$ ที่ได้จากการคำนวณซ้ำ ๆ เมื่อมีการเปลี่ยนแปลงน้อยมากจนพอใจก็จะหยุด ในปริญญานิพนธ์นี้ใช้ประมาณ 50-70 รอบ ก็จะได้ ค่าλ ของแบบจำลองที่ต้องการ

6. เก็บค่า a,b,π ที่ได้จากการคำนวนซ้ำๆ นี้ เป็นพารามิเตอร์ของแบบจำลอง

ขั้นตอนดังกล่าวมานี้ ได้ทำโดยใช้โปรแกรม hmm.m

6.3.2 กระบวนการทดสอบการรู้จำเสียง

การทดสอบการรู้จำเสียงนั้นจะใช้ Viterbi Algorithm ซึ่งได้อธิบายไว้แล้วในข้างต้น โดยหลังจากที่เราได้แบบจำลอง HMM สำหรับเสียง 0 ถึง 9 แล้ว เมื่อมีลำดับของเหตุการณ์ $O=\{O_1,O_2,....O_T\}$ ของ Unknow Word หรือ คำที่ต้องการทดสอบการรู้จำเข้ามา จะใช้ Viterbi Algorithm ในการคำนวณหา $P(O|\lambda)$ สำหรับทุก ๆ แบบจำลองของแต่ละคำศัพท์ และเลือกเอาคำศัพท์ที่มีค่าความน่าจะเป็น (P) สูงสุด

ขั้นตอนดังกล่าวมานี้ ได้ทำโดยใช้โปรแกรม recogniz.m

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.4 ขั้นตอนการทดลอง

ขั้นตอนที่1

1. เมื่อผ่านขั้นตอนการตัดหัวท้ายของเสียงแล้ว จะผ่านมาหาค่าสัมประสิทธิ์ตามหัวข้อ 6.1 ข้อมูลที่ออกมาจะเป็นข้อมูลที่แบ่งเป็นเฟรม ในแต่ละเฟรมจะประกอบด้วย สัมประสิทธิ์ 19 ค่า ที่แทนลักษณะของสัญญาณในแต่ละเฟรม

2. เมื่อผ่านการเวทค่าพารามิเตอร์เรียบร้อยแล้วจะทำการสุ่มค่าเริ่มต้น เพื่อใช้ในการหาโค้ดบุค

3. นำข้อมูลทั้งหมดมาหาโค้ดบุค ที่เหมาะสมตามหัวข้อ 6.2

4. นำข้อมูลดังกล่าวเปรียบเทียบกับโค้ดบุคว่า โค้ดบุคใดจะเป็นตัวแทนของข้อมูล

5. ทำการสุ่มค่าเริ่มต้น a_{ij} และ b_{ij}

6. ทำการหาแบบจำลองของเสียง โดยหา 1 แบบจำลองต่อ 1 เสียง

ขั้นตอนที่ 2 ใช้ในการวิเคราะห์

1. จาก เสียงที่เราต้องการทดสอบ ที่ผ่านการตัดหัวตัดท้ายแล้ว จะถูกนำมาหาค่าสัมประสิทธิ์ต่าง ๆ 19 ค่าในแต่ละเฟรม

2. นำค่าพารามิเตอร์ที่ผ่านการเวทค่าแล้ว มาทำการหาเวคเตอร์โค้ด (ดัชนีโค้ดบุค)ที่เหมาะสม

3. นำ เสียงทดสอบที่แปลงเป็นเวคเตอร์โค้ดเรียบร้อยแล้ว มาหาความน่าจะเป็นที่แบบจำลองต่าง ๆ จะให้กำเนิดค่าตามเสียงทดสอบ โดยใช้วิธีการใน 6.3.2

4. การตัดสินใจว่าเป็นแบบจำลองใด โดยเลือกค่าความน่าจะเป็นที่มากที่สุด

6.5 กรณีต่าง ๆ ที่ใช้ในการทดลอง

- กรณีที่ 1 ใช้เสียงจากผู้หญิงคนเดียว โดยที่แต่ละเสียง(0-9) พูดเสียงละ 5 ครั้ง
- กรณีที่ 2 ใช้เสียงจากผู้ชาย 1 คน ผู้หญิง 1 คน โดยที่แต่ละเสียง(0-9) พูดเสียงละ 5 ครั้ง
- กรณีที่ 3 ใช้เสียงจากบุคคล 20 คน โดยที่แต่ละเสียง(0-9) พูดเสียงละ 1 ครั้ง
- กรณีที่ 4 ใช้เสียงจากผู้ชาย 5 คน ผู้หญิง 5 คน โดยที่แต่ละเสียง(0-9) พูดเสียงละ 1 ครั้ง
- กรณีที่ 5 ใช้เสียงจากผู้ชายคนเดียว โดยที่แต่ละเสียง(0-9) พูดเสียงละ 5 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7
ผลการทดลอง

ในบทนี้ได้รวบรวมผลการทดลองในการหาค่าพารามิเตอร์ต่าง ๆ ของเสียง 0 - 9 โดยวิธีออโตคอรีรีเลชัน ในการประมาณเชิงเส้นรวมถึงพล็อตกราฟเปรียบเทียบระหว่างสัมประสิทธิ์ LPC , สัมประสิทธิ์เซปสตรัม และพารามิเตอร์ที่เวทค่าเรียบร้อยแล้ว พร้อมทั้งผลการทดสอบการรู้จำเสียงพูด

7.1 การเปรียบเทียบอัตราการขยาย และสัมประสิทธิ์ LPC ที่ได้จากโปรแกรม MATLAB ที่ได้มีการทดลองไว้แล้ว กับสัมประสิทธิ์ LPC ที่ทดลองใหม่โดยใช้ภาษา C

สัมประสิทธิ์ที่ได้จากการทดลองครั้งใหม่นี้มีค่าใกล้เคียงกับผลการทดลองเดิม ตามตารางที่ 7.1 และ 7.2

เสียง หนึ่ง	สัมประสิทธิ์ LPC	สัมประสิทธิ์ LPC	อัตราการขยาย	อัตราการขยาย
	จากภาษา C	จาก MATLAB	(Gain) จาก ภาษา C	(Gain)จาก MATLAB
1	0.270053	0.270273		
2	0.208648	0.208301		
3	0.300678	0.300861		
4	0.093551	0.093306		
5	0.361811	0.361991		
6	-0.002678	-0.002843		
7	0.279373	0.279517		
8	-0.147979	-0.148185		
9	-0.063121	-0.063221		
10	-0.125031	-0.125243		
11	-0.079102	-0.079138		
12	-0.09763	-0.097741		
			4.685026	4.660183

ตารางที่ 7.1 แสดงการเปรียบเทียบสัมประสิทธิ์ LPC จากโปรแกรมทั้งสอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

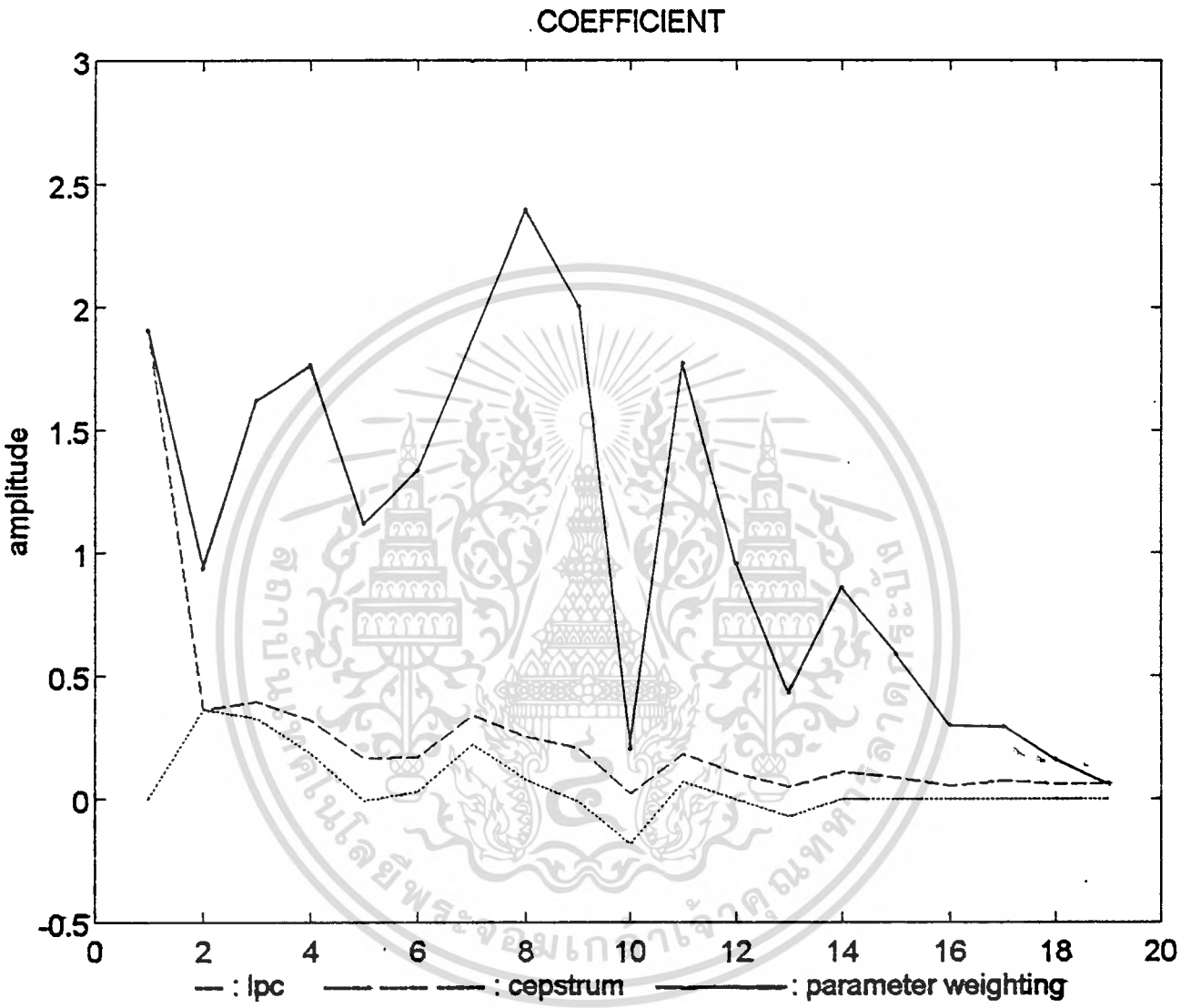
เสียง สี่	สัมประสิทธิ์ LPC	สัมประสิทธิ์ LPC	อัตราการขยาย	อัตราการขยาย
	จากภาษา C	จาก MATLAB	(Gain) จาก ภาษา C	(Gain)จาก MATLAB
1	0.232367	0.23263		
2	0.150555	0.150408		
3	0.190099	0.190196		
4	0.312022	0.312121		
5	0.217825	0.217862		
6	0.238344	0.238238		
7	0.097838	0.097814		
8	0.000625	0.000401		
9	-0.050133	-0.050026		
10	-0.047813	-0.047926		
11	-0.007011	-0.0070688		
12	-0.337265	-0.337237		
			4.583897	4.513927

ตารางที่ 7.2 แสดงการเปรียบเทียบสัมประสิทธิ์ LPC จากโปรแกรมทั้งสอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

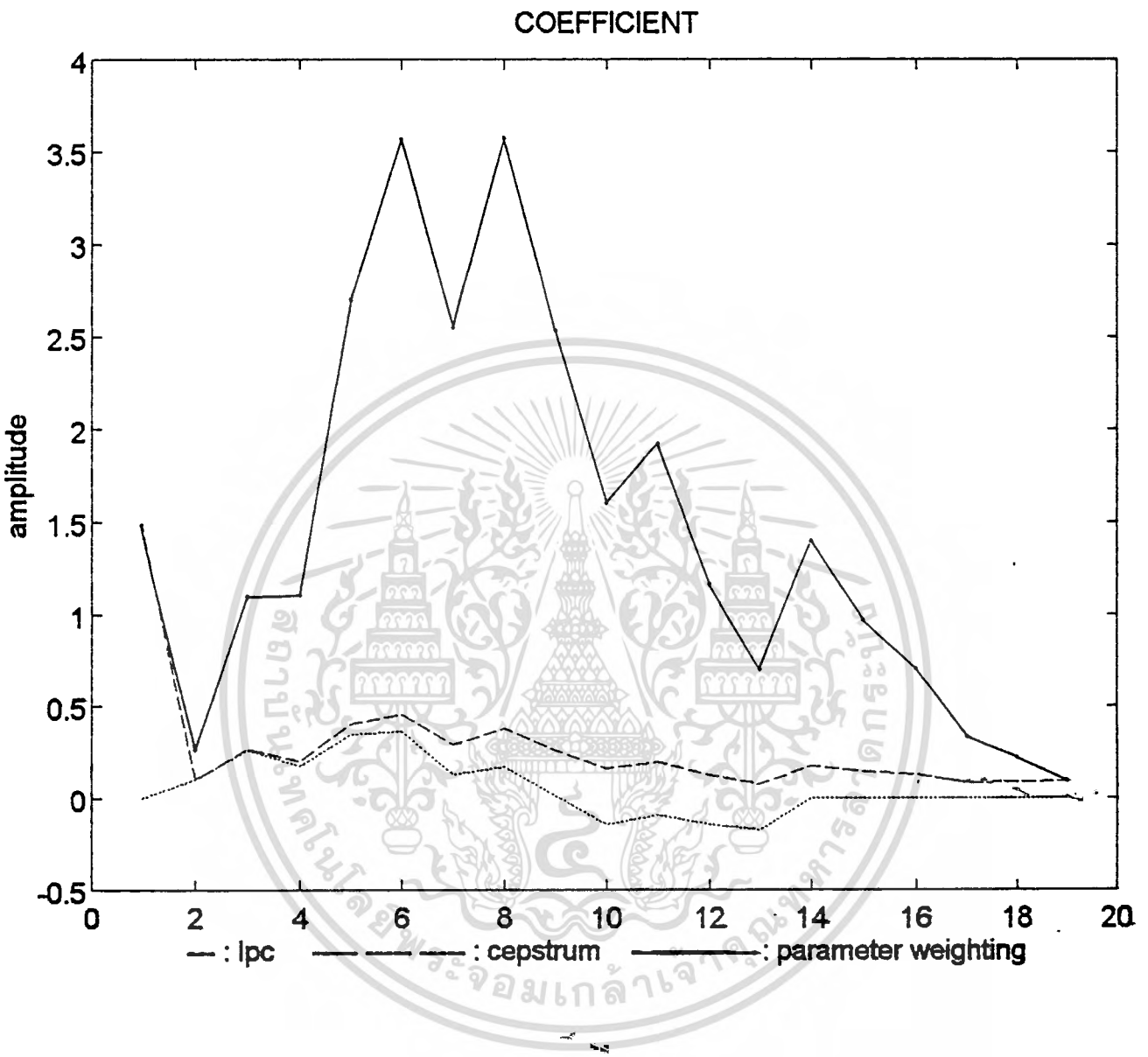
7.2 ผลการเปรียบเทียบค่าสัมประสิทธิ์ LPC , เซปสตรัม และพารามิเตอร์ที่ได้เวทค่าแล้ว

การเปรียบเทียบแสดงในรูปของกราฟ เช่น เสียง สอง ดังแสดงในรูปที่ 7.1 และ 7.2



รูปที่ 7.1 กราฟเปรียบเทียบสัมประสิทธิ์เสียงของผู้ชาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

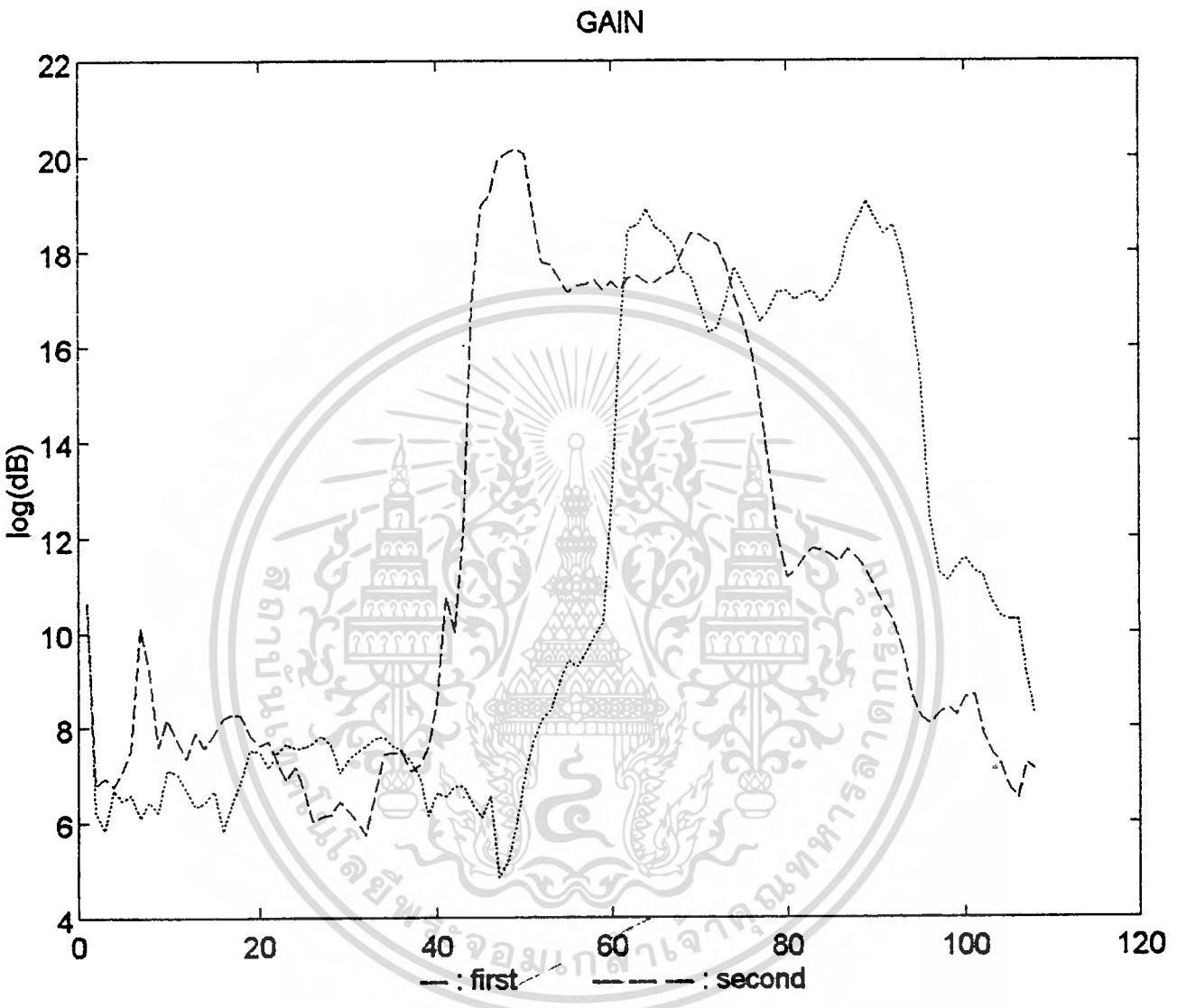


รูปที่ 7.2 กราฟเปรียบเทียบสัมประสิทธิ์เสียงของผู้หญิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.3 ผลการเปรียบเทียบค่าอัตราการขยายของเสียงเดียวกันซึ่งพูดซ้ำกัน 2 ครั้ง

ลักษณะของกราฟที่ได้มีลักษณะคล้ายคลึงกันทั้งสองครั้ง แต่จะอยู่ที่ตำแหน่งต่างกัน ดังรูปที่ 7.3



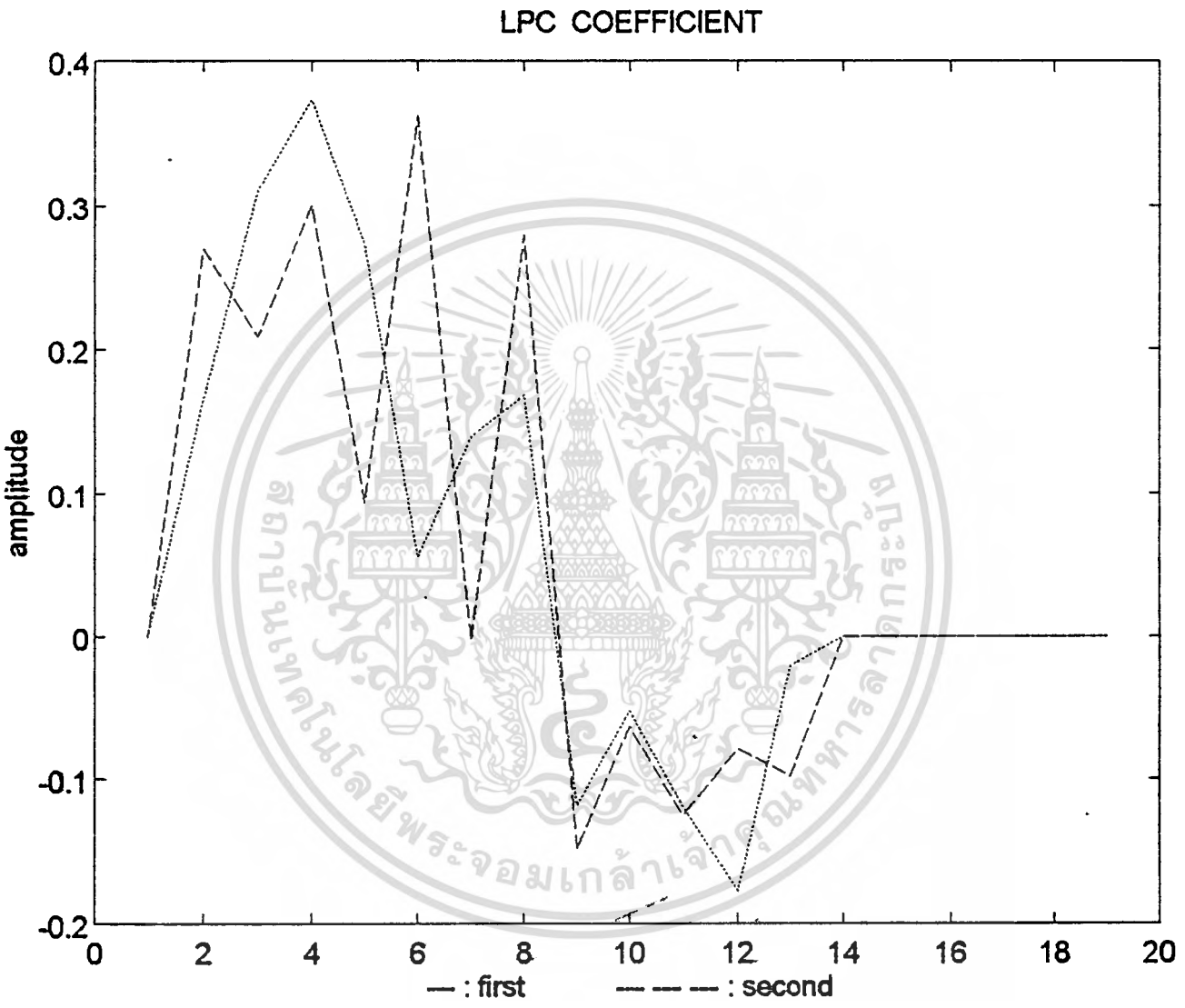
รูปที่ 7.3 กราฟเปรียบเทียบอัตราการขยายของเสียงเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.4 ผลการเปรียบเทียบสัมประสิทธิ์ ของการพูดเสียงเดียวกันซ้ำกัน 2 ครั้ง

7.4.1 ผลการเปรียบเทียบสัมประสิทธิ์ LPC

ลักษณะของสัมประสิทธิ์ LPC มีค่าแตกต่างกันอย่างชัดเจน ของการพูดทั้ง 2 ครั้ง ดังกราฟรูปที่ 7.4

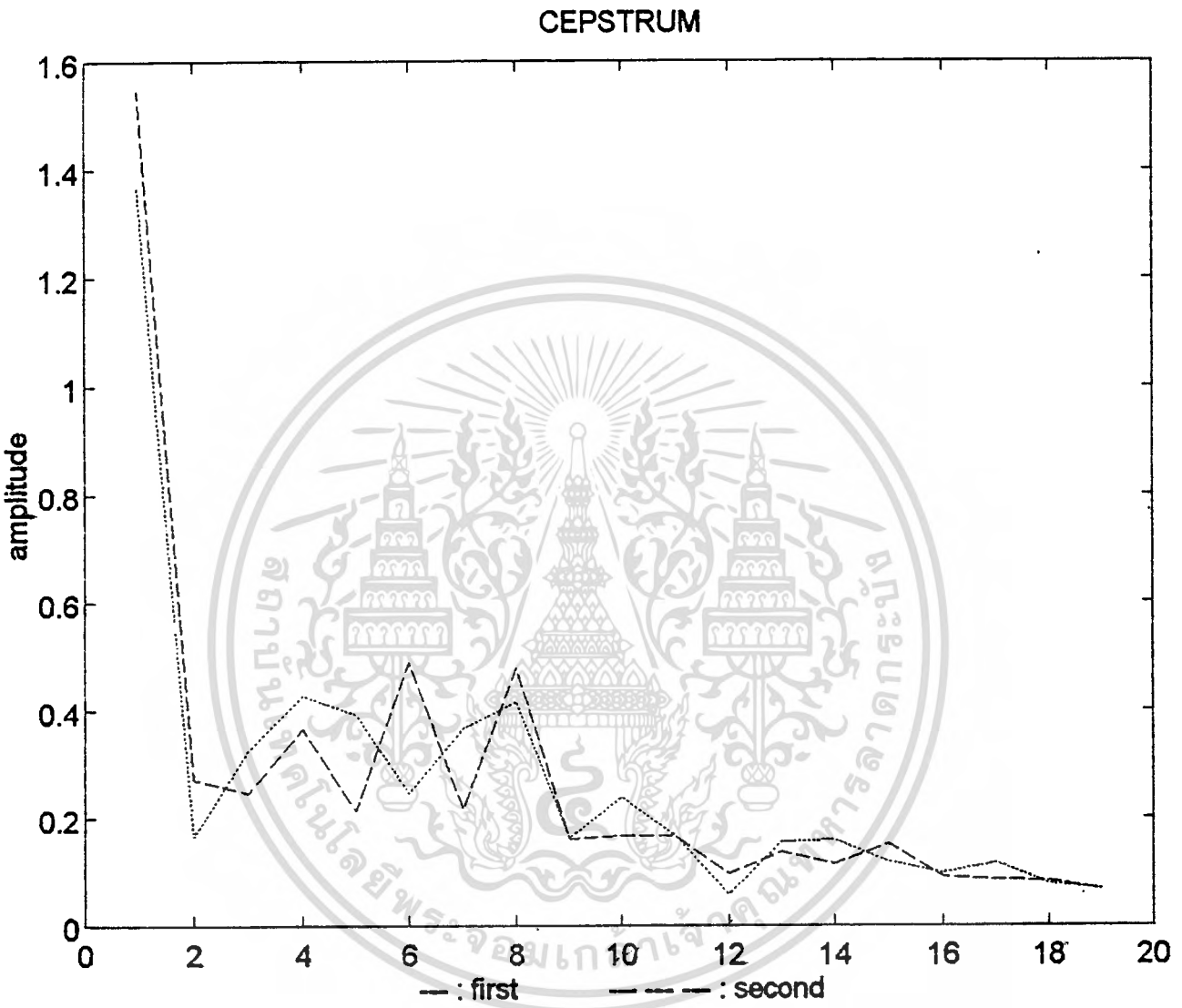


รูปที่ 7.4 กราฟเปรียบเทียบสัมประสิทธิ์ LPC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.4.2 ผลการเปรียบเทียบสัมประสิทธิ์เซปสตรัม

ลักษณะของสัมประสิทธิ์เซปสตรัม ที่ได้ ของการพูดแต่ละครั้งมีค่าใกล้เคียงกัน ดังกราฟรูปที่ 7.5

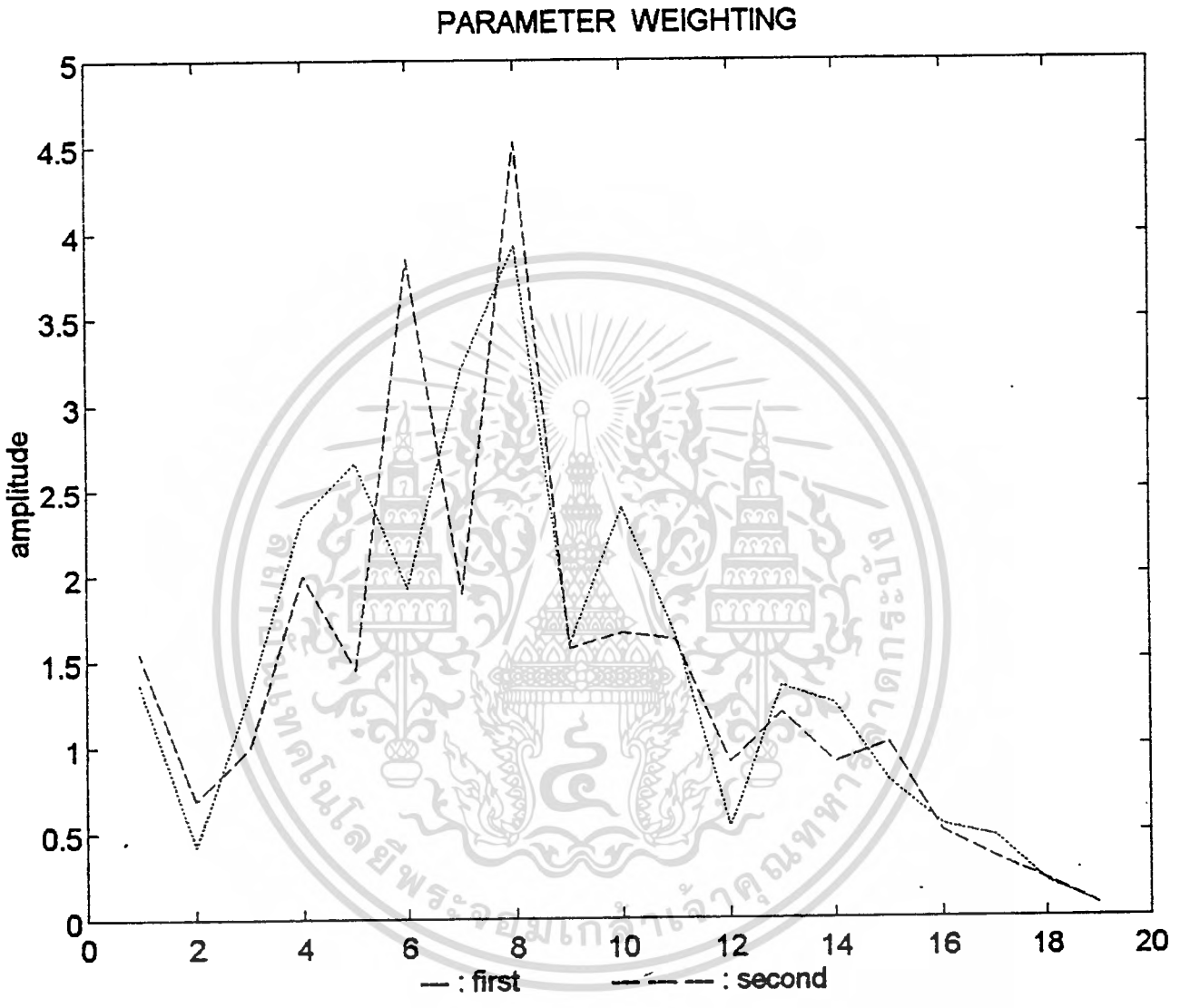


รูปที่ 7.5 กราฟเปรียบเทียบสัมประสิทธิ์เซปสตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.4.3 ผลการเปรียบเทียบพารามิเตอร์ที่ได้เวทค่าแล้ว

ลักษณะของพารามิเตอร์ที่ได้เวทค่า ของการพูดแต่ละครั้งมีค่าใกล้เคียงกัน ดังกราฟรูปที่ 7.6

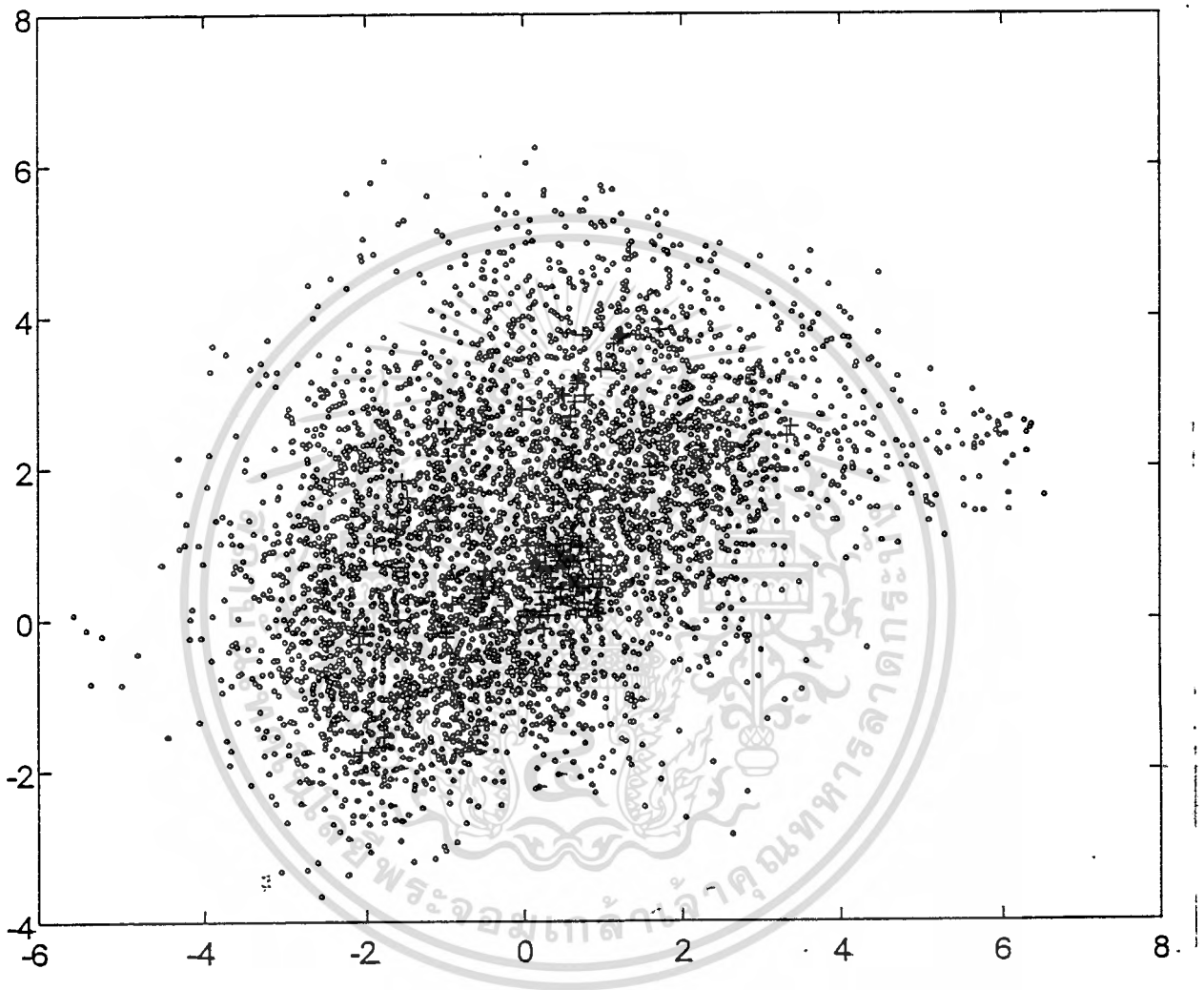


รูปที่ 7.6 กราฟเปรียบเทียบพารามิเตอร์ที่ได้เวทค่าแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.5 ผลการคำนวณโค้ดบุค

จากรูปที่ 7.7 เป็นตัวอย่างการแสดงเวคเตอร์พารามิเตอร์ของเฟรมต่าง ๆ กับโค้ดบุคที่คำนวณได้โดยแสดงใน 2 มิติ



รูปที่ 7.7 แสดงเวคเตอร์พารามิเตอร์ของเฟรมต่าง ๆ กับโค้ดบุคที่คำนวณได้ใน 2 มิติ

∘ = แทนเวคเตอร์พารามิเตอร์ใน 1 เฟรม

\+ = แทนจุดศูนย์กลางของกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการทดสอบการรู้จำเสียงพูด

7.6 **กรณีที่** 1 แบบจำลองต้นแบบจากผู้พูดผู้หญิงคนเดียว โดยที่แต่ละเสียง (0 - 9) พูดเสียงละ 5 ครั้ง นำมาใช้สร้างโค้ดบุคอ้างอิง พร้อมทั้งสร้างแบบจำลองของแต่ละเสียง (สร้างจากเสียง 50 เสียง)

7.6.1 การทดสอบการรู้จำโดยนำเสียงจากผู้พูดคนเดิม แต่ให้พูดครั้งใหม่เสียงละ 10 ครั้ง มาทดสอบ

7.6.2 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มคนทดสอบ จำนวน 10 คน เป็นผู้ชาย 3 คน และผู้หญิง 7 คน มาทดสอบ

ผลการทดสอบดังตารางที่ 7.3 และ 7.4 ดังนี้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 1	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 2	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 3	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 4	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 5	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 6	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 7	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 8	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 9	0	1	2	3	4	5	6	7	8	2
เสียงทดสอบพูดครั้งที่ 10	0	1	2	3	4	5	6	7	8	9

ตารางที่ 7.3 แสดงผลการทดสอบของเสียงจากผู้พูดคนเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	1	3	9	4	3	6	7	9	3
เสียงทดสอบผู้หญิงคนที่ 2	0	3	2	3	4	5	6	7	9	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	2	3	4	5	6	7	8	2
เสียงทดสอบผู้หญิงคนที่ 4	2	2	2	3	4	5	0	7	8	2
เสียงทดสอบผู้หญิงคนที่ 5	0	1	2	3	4	5	6	1	8	9
เสียงทดสอบผู้หญิงคนที่ 6	0	1	2	3	4	3	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 7	0	2	6	3	4	9	6	7	8	9
เสียงทดสอบผู้ชายคนที่ 1	2	3	2	3	4	5	6	3	9	2
เสียงทดสอบผู้ชายคนที่ 2	2	2	2	2	1	9	6	3	2	2
เสียงทดสอบผู้ชายคนที่ 3	0	7	6	2	7	3	0	3	9	2

ตารางที่ 7.4 แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.7 **กรณีที่** 2 แบบจำลองต้นแบบจาก ผู้ชาย 1 คน ผู้หญิง 1 คน โดยที่แต่ละเสียง (0 - 9) พูดเสียงละ 5 ครั้ง นำมาใช้สร้างโค้ดบุคอ้างอิง พร้อมทั้งสร้างแบบจำลองของแต่ละเสียง (สร้างจากเสียง 100 เสียง)

7.7.1 การทดสอบการรู้จำโดยนำเสียงจากผู้พูด 2 คนเดิม แต่ให้พูดครั้งใหม่เสียงละ 5 ครั้ง มาทดสอบ

7.7.2 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มคนทดสอบ จำนวน 10 คน เป็นผู้ชาย 5 คน และผู้หญิง 5 คน มาทดสอบ

ผลการทดสอบดังตารางที่ 7.5 และ 7.6 ดังนี้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงครั้งที่ 1	0	1	2	2	4	5	6	7	8	9
เสียงทดสอบผู้หญิงครั้งที่ 2	0	1	2	3	4	5	6	7	8	2
เสียงทดสอบผู้หญิงครั้งที่ 3	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงครั้งที่ 4	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงครั้งที่ 5	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้ชายครั้งที่ 1	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้ชายครั้งที่ 2	0	1	2	3	4	5	6	7	8	3
เสียงทดสอบผู้ชายครั้งที่ 3	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้ชายครั้งที่ 4	0	1	2	2	4	2	6	7	8	2
เสียงทดสอบผู้ชายครั้งที่ 5	0	1	2	2	4	5	6	7	8	2

ตารางที่ 7.5 แสดงผลการทดสอบของเสียงจากผู้พูดเดิมทั้ง 2 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	0	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 2	0	1	6	5	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	9	8	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 4	2	1	2	5	4	5	2	7	8	9
เสียงทดสอบผู้หญิงคนที่ 5	0	1	3	5	4	5	6	7	8	9
เสียงทดสอบผู้ชายคนที่ 1	0	8	8	9	4	5	6	7	8	9
เสียงทดสอบผู้ชายคนที่ 2	0	3	2	2	0	5	6	1	9	2
เสียงทดสอบผู้ชายคนที่ 3	6	0	2	5	4	3	2	8	8	2
เสียงทดสอบผู้ชายคนที่ 4	2	5	2	8	1	8	6	1	8	2
เสียงทดสอบผู้ชายคนที่ 5	0	0	6	2	4	3	2	1	5	6

ตารางที่ 7.6 แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.8 **กรณีที่** 3 แบบจำลองต้นแบบจากกลุ่มผู้พูดผู้หญิง 20 คน โดยที่แต่ละเสียง (0 - 9) พูดเสียงละ 1 ครั้ง นำมาใช้สร้างโค้ดบุคอ้างอิง พร้อมทั้งสร้างแบบจำลองของแต่ละเสียง (สร้างจากเสียง 200 เสียง)

7.8.1 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มผู้พูดกลุ่มเดิม จำนวน 10 คน แต่ให้พูดครั้งใหม่ มาทดสอบ

7.8.2 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มคนทดสอบ จำนวน 15 คน เป็นผู้ชาย 5 คน และ ผู้หญิง 10 คนมาทดสอบ

ผลการทดสอบดังตารางที่ 7.7 และ 7.8 ดังนี้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	1	2	3	4	9	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 2	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	2	3	4	5	6	4	8	9
เสียงทดสอบผู้หญิงคนที่ 4	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 5	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 6	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 7	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 8	0	1	2	8	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 9	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 10	0	1	2	3	4	5	6	7	8	9

ตารางที่ 7.7 แสดงผลการทดสอบของเสียงจากกลุ่มผู้พูดเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 2	0	1	0	3	4	5	6	4	8	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	6	5	4	5	0	7	8	9
เสียงทดสอบผู้หญิงคนที่ 4	0	1	2	5	4	8	6	7	8	5
เสียงทดสอบผู้หญิงคนที่ 5	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 6	0	1	2	3	1	8	0	4	8	9
เสียงทดสอบผู้หญิงคนที่ 7	0	1	2	3	4	9	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 8	0	1	2	2	4	5	6	7	8	6
เสียงทดสอบผู้หญิงคนที่ 9	0	1	6	3	4	5	0	7	8	9
เสียงทดสอบผู้หญิงคนที่ 10	0	1	2	3	4	5	6	7	8	2
เสียงทดสอบผู้ชายคนที่ 1	0	1	6	9	4	5	6	4	8	9
เสียงทดสอบผู้ชายคนที่ 2	0	1	2	2	4	5	0	1	8	2
เสียงทดสอบผู้ชายคนที่ 3	0	1	6	2	9	5	0	1	5	6
เสียงทดสอบผู้ชายคนที่ 4	0	1	6	2	1	5	0	1	2	6
เสียงทดสอบผู้ชายคนที่ 5	0	7	6	5	4	9	0	7	9	9

ตารางที่ 7.8 แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.9 กรณีที่ 4 แบบจำลองต้นแบบจาก ผู้ชาย 5 คน ผู้หญิง 5 คน โดยที่แต่ละเสียง (0 - 9) พูดเสียงละ 1 ครั้ง นำมาใช้สร้างโค้ดบุคอ้างอิง พร้อมทั้งสร้างแบบจำลองของแต่ละเสียง (สร้างจากเสียง 100 เสียง)

7.9.1 การทดสอบการรู้จำโดยนำเสียงจากผู้พูด 10 คนเดิม แต่ให้พูดครั้งใหม่เสียงละ 1 ครั้ง มาทดสอบ

7.9.2 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มคนทดสอบ จำนวน 10 คน เป็นผู้ชาย 5 คน และผู้หญิง 5 คน มาทดสอบ

ผลการทดสอบดังตารางที่ 7.9 และ 7.10 ดังนี้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 2	0	1	2	3	4	9	6	4	8	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	2	3	4	5	6	7	8	8
เสียงทดสอบผู้หญิงคนที่ 4	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 5	0	1	2	3	4	9	6	7	8	3
เสียงทดสอบผู้ชายคนที่ 1	0	1	2	3	4	3	0	7	8	9
เสียงทดสอบผู้ชายคนที่ 2	0	1	2	3	4	5	6	1	4	9
เสียงทดสอบผู้ชายคนที่ 3	0	1	6	3	4	3	6	7	3	2
เสียงทดสอบผู้ชายคนที่ 4	2	1	2	8	5	5	6	7	5	9
เสียงทดสอบผู้ชายคนที่ 5	0	1	2	3	4	3	6	4	8	9

ตารางที่ 7.9 แสดงผลการทดสอบของเสียงจากผู้พูดเดิมทั้ง 10 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 1	0	9	3	3	4	9	2	7	9	9
เสียงทดสอบผู้หญิงคนที่ 2	0	1	3	9	4	5	6	7	8	9
เสียงทดสอบผู้หญิงคนที่ 3	0	1	3	3	4	3	2	7	8	3
เสียงทดสอบผู้หญิงคนที่ 4	0	1	3	3	4	3	6	4	8	3
เสียงทดสอบผู้หญิงคนที่ 5	0	1	2	3	4	5	6	1	8	9
เสียงทดสอบผู้ชายคนที่ 1	0	7	2	3	4	5	6	7	3	2
เสียงทดสอบผู้ชายคนที่ 2	0	1	2	9	4	8	1	4	8	9
เสียงทดสอบผู้ชายคนที่ 3	2	1	2	3	7	3	6	7	8	9
เสียงทดสอบผู้ชายคนที่ 4	6	1	2	3	4	8	6	7	8	3
เสียงทดสอบผู้ชายคนที่ 5	2	1	2	3	7	5	0	1	8	8

ตารางที่ 7.10 แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.10 กรณีที่ 5 แบบจำลองต้นแบบจากผู้พูดผู้ชายคนเดียว โดยที่แต่ละเสียง (0 - 9) พูดเสียงละ 5 ครั้ง นำมาใช้สร้างโค้ดบุคอ้างอิง พร้อมทั้งสร้างแบบจำลองของแต่ละเสียง (สร้างจากเสียง 50 เสียง)

7.10.1 การทดสอบการรู้จำโดยนำเสียงจากผู้พูดคนเดิม แต่ให้พูดครั้งใหม่เสียงละ 10 ครั้ง มาทดสอบ

7.10.2 การทดสอบการรู้จำโดยนำเสียงจากกลุ่มคนทดสอบ จำนวน 10 คน เป็นผู้ชาย 7 คน และผู้หญิง 3 คน

มาทดสอบ

ผลการทดสอบดังตารางที่ 7.11 และ 7.12 ดังนี้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 1	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 2	0	1	2	3	4	5	6	4	8	9
เสียงทดสอบพูดครั้งที่ 3	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 4	0	1	2	3	4	5	6	7	8	3
เสียงทดสอบพูดครั้งที่ 5	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 6	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 7	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 8	0	1	2	9	4	5	6	7	8	9
เสียงทดสอบพูดครั้งที่ 9	0	1	2	3	4	5	6	4	8	2
เสียงทดสอบพูดครั้งที่ 10	0	1	2	3	4	5	6	7	8	9

ตารางที่ 7.11 แสดงผลการทดสอบของเสียงจากผู้พูดคนเดิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงต้นแบบ	0	1	2	3	4	5	6	7	8	9
เสียงทดสอบผู้ชายคนที่ 1	0	1	0	3	4	5	6	4	8	9
เสียงทดสอบผู้ชายคนที่ 2	0	1	6	3	4	5	6	1	8	9
เสียงทดสอบผู้ชายคนที่ 3	9	6	6	8	4	9	6	7	1	9
เสียงทดสอบผู้ชายคนที่ 4	6	1	2	3	4	5	6	4	9	2
เสียงทดสอบผู้ชายคนที่ 5	6	1	2	5	7	5	0	9	5	5
เสียงทดสอบผู้ชายคนที่ 6	0	1	2	3	4	3	6	9	8	3
เสียงทดสอบผู้ชายคนที่ 7	6	8	9	5	4	9	6	7	5	9
เสียงทดสอบผู้หญิงคนที่ 1	5	9	3	8	4	5	6	4	8	9
เสียงทดสอบผู้หญิงคนที่ 2	0	9	3	3	0	5	6	6	3	3
เสียงทดสอบผู้หญิงคนที่ 3	6	9	9	3	4	3	6	4	8	3

ตารางที่ 7.12 แสดงผลการทดสอบของเสียงจากกลุ่มคนทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.11 ผลการเปรียบเทียบความถูกต้องของการทดสอบการรู้จำเสียงพูด

ความถูกต้องของเสียงที่นำมาทดสอบคิดเป็นเปอร์เซ็นต์แล้วได้ผลดังตารางที่ 7.13 ดังนี้

การเทรน	ความถูกต้อง
กรณีที่ 1 ผู้หญิงคนเดียว	
1. การทดสอบกับคนเดิม	99%
2. การทดสอบกับกลุ่มทดสอบ	
ผู้หญิง	76%
ผู้ชาย	27%
รวม	61%
กรณีที่ 2 ผู้ชาย 1 คน ผู้หญิง 1 คน	
1. การทดสอบกับ 2 คนเดิม	92%
2. การทดสอบกับกลุ่มคนทดสอบ	
ผู้หญิง	80%
ผู้ชาย	38%
รวม	59%
กรณีที่ 3 ผู้หญิง 20 คน	
1. การทดสอบกับกลุ่มคนเดิม	97%
2. การทดสอบกับกลุ่มคนทดสอบ	
ผู้หญิง	82%
ผู้ชาย	46%
รวม	70%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเทรน	ความถูกต้อง
กรณีที่ 4 ผู้ชาย 5 คน ผู้หญิง 5 คน	
1. การทดสอบกับกลุ่มคนเดิม	81%
2. การทดสอบกับกลุ่มคนทดสอบ	
ผู้หญิง	68%
ผู้ชาย	64%
รวม	66%
กรณีที่ 5 ผู้ชาย 1 คน	
1. การทดสอบกับคนเดิม	96%
2. การทดสอบกับกลุ่มทดสอบ	
ผู้หญิง	43%
ผู้ชาย	57%
รวม	53%

ตารางที่ 7.13 แสดงการเปรียบเทียบความถูกต้องในการทดสอบเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดสอบการรู้จำเสียงพูด สามารถอธิบายผลการทดสอบได้ดังนี้

กรณีที่ 1 เนื่องจากเสียงต้นแบบเป็นเสียงผู้หญิง 1 คน ประสิทธิภาพการรู้จำของเสียงผู้หญิงคนเดิมมีสูงถึง 99 เปอร์เซ็นต์ และสามารถรู้จำเสียงของผู้หญิงคนอื่นได้ แต่มีประสิทธิภาพต่ำกว่า คือประมาณ 76 เปอร์เซ็นต์ ส่วนการรู้จำเสียงของผู้ชายนั้นมีค่าต่ำมาก

กรณีที่ 2 เมื่อเพิ่มเสียงที่ใช้ในการสร้างแบบจำลองเป็นผู้ชาย และ ผู้หญิง อย่างละ 1 คน พบว่าสามารถรู้จำเสียง

- เสียงต้นแบบได้ถูกต้อง 92 เปอร์เซ็นต์
- เสียงผู้หญิงคนอื่น ได้ 80 เปอร์เซ็นต์
- เสียงผู้ชายคนอื่นได้ 38 เปอร์เซ็นต์ ซึ่งจะเห็นได้ว่าสามารถรู้จำเสียงของผู้ชายได้มากกว่ากรณีที่ 1 เนื่องจากมีการนำเสียงของผู้ชายมาเป็นเสียงต้นแบบด้วย

กรณีที่ 3. เมื่อเพิ่มเสียงผู้หญิงเป็น 10 คน ในการสร้างแบบจำลอง สามารถรู้จำเสียงได้ดังนี้

- เสียงผู้หญิงกลุ่มต้นแบบได้ถูกต้อง 97 เปอร์เซ็นต์
- เสียงผู้หญิงคนอื่นที่ ไม่ได้อยู่ในกลุ่มต้นแบบ ถูกต้อง 82 เปอร์เซ็นต์ ซึ่งจะถูกต้องมากกว่าในทุก ๆ กรณี เพราะแบบจำลองเป็นเสียงผู้หญิงที่มีความหลากหลาย
- เสียงผู้ชายถูกต้อง 46 เปอร์เซ็นต์

กรณีที่ 4 เมื่อเพิ่มเสียงของการสร้างแบบจำลองเป็นผู้ชาย และผู้หญิง อย่างละ 5 คน พบว่าสามารถรู้จำเสียงได้ดังนี้

- เสียงต้นแบบ ถูกต้อง 81 เปอร์เซ็นต์
- เสียงผู้หญิงคนอื่นได้ 68 เปอร์เซ็นต์
- เสียงผู้ชายคนอื่นได้ 64 เปอร์เซ็นต์ ซึ่งจะเห็นได้ว่า สามารถแยกแยะเสียงระหว่างผู้ชายและผู้หญิงได้มากกว่า กรณีที่ 2

กรณีที่ 5 เสียงต้นแบบเป็นเสียงผู้ชาย 1 คน ประสิทธิภาพการรู้จำของเสียงผู้ชายคนเดิมมีค่าสูงถึง 96 เปอร์เซ็นต์ และสามารถรู้จำเสียงของผู้ชายกับผู้หญิงคนอื่นได้ แต่มีประสิทธิภาพต่ำ คือ ประมาณ 57 และ 43 เปอร์เซ็นต์ ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 8
สรุปผลการทดลอง

8.1 ขั้นตอนการเรียนรู้ (Learning)

การวิเคราะห์การเข้ารหัสโดยการประมาณเชิงเส้น (Linear Predictive Coding : LPC) เป็นวิธีที่ใช้ในการเข้ารหัสสัญญาณข้อมูลเพื่อนำไปประยุกต์ใช้งานต่าง ๆ ด้วยเหตุที่ว่า วิธีนี้เป็นการวิเคราะห์พารามิเตอร์ที่ให้ความแม่นยำสูงและสามารถย่อข้อมูลได้อย่างมีประสิทธิภาพ ในปริญญานิพนธ์นี้ จึงได้นำวิธีนี้มาใช้เพื่อวิเคราะห์พารามิเตอร์ของเสียง เพื่อนำไปใช้ในการรู้จำเสียงพูดต่อไป

หลักการของการประมาณเชิงเส้น คือการประมาณสัญญาณเสียงจากผลรวมเชิงเส้นของสัญญาณเสียงในอดีตโดยอาศัย เกณฑ์กำลังสองของความคลาดเคลื่อนที่มีค่าต่ำสุด ในการหาสัมประสิทธิ์การประมาณเชิงเส้น LPC', เซปสตรัม และพารามิเตอร์ที่ได้เวทค่า

จากการทดลอง โดยใช้สัญญาณเสียงที่สุ่มด้วยความถี่ 11.025 กิโลเฮิรตซ์ ความยาวประมาณ 0.3 ถึง 1 วินาที โดยแบ่งเป็นเฟรม เฟรมละ 300 ตัวอย่างสัญญาณเสียง และเลื่อนเฟรมทีละ 100 ตัวอย่างสัญญาณเสียง และใช้การประมาณเชิงเส้นอันดับที่ 12 เราสามารถสรุปได้ดังนี้

8.1.1 สัมประสิทธิ์ LPC ที่คำนวณได้จากโปรแกรม 'LPC' โดยใช้ภาษา C มีค่าใกล้เคียงสัมประสิทธิ์ LPC ที่ใช้โปรแกรม MATLAB จากที่เคยมีการทดลองไว้แล้ว

8.1.2 ค่าสัมประสิทธิ์ของเสียงตัวเลขใด ๆ ที่วิเคราะห์ได้ใน 1 เฟรม ประกอบไปด้วย

1. เกน	1 ค่า
2. สัมประสิทธิ์ LPC	12 ค่า
3. สัมประสิทธิ์เซปสตรัม	19 ค่า

และใน.1 เสียง จำนวนเฟรมที่วิเคราะห์จะขึ้นอยู่กับความยาวของสัญญาณเสียงที่พูด

8.1.3. สัมประสิทธิ์เซปสตรัมนั้น ใช้ในการปรับปรุงสัมประสิทธิ์ LPC ให้คงลักษณะของเสียงได้มากขึ้นและจำนวนสัมประสิทธิ์ที่ได้จากการวิเคราะห์ จะมีมากกว่า สัมประสิทธิ์ LPC

8.1.4 การเวทค่าพารามิเตอร์จะช่วยลดความผิดพลาดอันเกิดจากรอยต่อเฟรม ทำให้ได้สัมประสิทธิ์ที่ถูกต้องยิ่งขึ้น

8.1.5 การพูดเสียงเดียวกัน ในครั้งใหม่จะได้สัมประสิทธิ์ชุดใหม่ ที่มีลักษณะใกล้เคียงกับชุดเดิมแต่เปลี่ยนตำแหน่งไป เนื่องจากการเริ่มต้นวิเคราะห์เสียงในแต่ละครั้ง ไม่ได้เริ่มที่ตำแหน่งเดียวกัน

8.1.6 เวลาที่ใช้พูดของแต่ละเสียงในการทดลอง มีผลต่อสัมประสิทธิ์ที่วิเคราะห์ เพราะฉะนั้นเวลานำเสียงมาวิเคราะห์ ควรพูดให้เป็นเสียงตามธรรมชาติมากที่สุด ซึ่งจะทำให้ได้ค่าสัมประสิทธิ์ที่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8.2 ขั้นตอนการรู้จำ (Recognition)

8.2.1 โดยผลการทดสอบที่ได้แสดงไว้ในบทที่ 7 ได้แสดงให้เห็นว่าวิธีการ Hidden Markov Models (HMM) เป็นวิธีที่ใช้ในการหาต้นแบบที่เหมาะสม ในรูปของแบบจำลองของความน่าจะเป็นในการเปลี่ยน state และความน่าจะเป็นในการเกิด เหตุการณ์ จึงทำให้ผลการทดสอบกับบุคคลที่ไม่ใช่กลุ่มบุคคลต้นแบบมีความถูกต้องค่อนข้างมาก เมื่อจำนวนต้นแบบมีความหลากหลายเพียงพอ

8.2.2 ในการทดสอบความถูกต้องของการรู้จำกับจำนวนเซตต้นแบบ โดยการค่อยๆ เพิ่มจำนวนบุคคลที่ใช้เป็นต้นแบบขึ้นเรื่อย ๆ ผลความถูกต้องในการทดสอบการรู้จำจะดีขึ้นเช่นกัน ในช่วงแรกการเพิ่มจำนวนบุคคลเพียงเล็กน้อยก็มีผลทำให้ความถูกต้องในการรู้จำเพิ่มขึ้นอย่างมาก จึงสรุปได้ว่า Hidden Markov Models(HMM) ต้องการ ต้นแบบในการรู้จำจำนวนมาก ให้มีความหลากหลายเพื่อครอบคลุมความเป็นไปได้ของเสียงนั้นๆ ได้มากขึ้นเช่นกัน

8.2.3 จากผลการทดสอบการรู้จำ ความผิดพลาดจากการทดสอบอาจเกิดจาก 3 สาเหตุ คือ

1. ความผิดปกติของเสียงเอง ไม่ว่าจะเป็นการเปล่งเสียงหรือขั้นตอนในการจัดเก็บเสียง

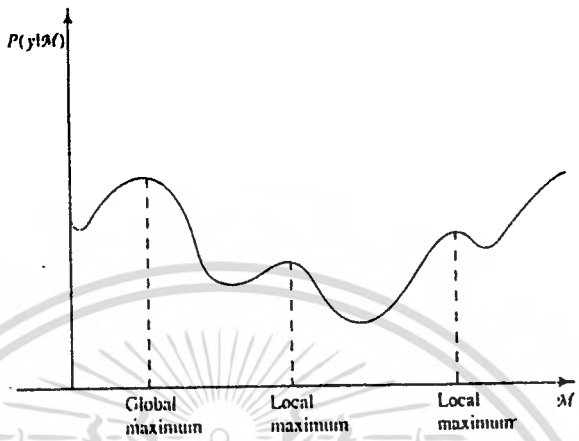
2. อาจเกิดจากขั้นตอนการควอนไตซ์ เนื่องจากโค้ดบุคไม่ได้มาตรฐานพอ ทำให้จัดระดับออกมาผิดพลาด โดยการควอนไตซ์ที่ผิดพลาดนั้นไม่จำเป็นต้องผิดพลาดทั้งหมด อาจผิดพลาดเพียงบางตัวที่มีความน่าจะเป็นมาก ก็สามารถทำให้ความน่าจะเป็นรวมเปลี่ยน ซึ่งมีผลทำให้ตัดสินใจผิดพลาดด้วย

3. อาจเกิดในขั้นตอนการสร้างแบบจำลอง เนื่องจากแบบจำลองที่ได้ไม่ดีพอ คือ พารามิเตอร์ของแบบจำลองที่คำนวณได้ไม่สอดคล้องกับเสียงนั้น ๆ ซึ่งอาจเกิดจากเซตลำดับเหตุการณ์ที่ใช้ในการเทรนแบบจำลองน้อยไป,ค่าเริ่มต้นไม่ดี,รอบของการเทรนไม่เหมาะสม เป็นต้น

8.2.4. ในขั้นตอนการเทรน HMM นั้นหากพิจารณาทางทฤษฎีนั้นจะพบว่าค่าสุ่มแต่ละครั้งในการเริ่มต้นกระบวนการจะให้ค่า Local Maximum Propability จึงต้องทำกระบวนการคำนวณซ้ำหลายๆครั้งโดยใช้ค่าสุ่มเริ่มต้นเปลี่ยนไปเรื่อยๆ แล้วเลือกเอาค่า a ,b,π ที่ให้ P สูงสุด แบบ Global Maximum เป็น λ ของแบบจำลอง ดังรูปที่ 8.1

แต่เนื่องจากเวลาที่มีจำกัดและเวลาที่ใช้ในการเทรนแต่ละครั้งใช้เวลามาก จึงเทรนด้วยค่าสุ่มเพียง 1 - 2 ครั้ง ดังนั้นค่าที่ได้นี้จึงไม่ใช่ค่า λ ที่ดีที่สุดสำหรับแบบจำลอง HMM ของคำนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 8.1 แสดงลักษณะค่าความน่าจะเป็น ตามพารามิเตอร์ค่าต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แนวทางในการพัฒนาศึกษาต่อไป

1. เสียงที่อัดจากไมโครโฟนมีสัญญาณรบกวนจากภายนอก ทำให้เสียงที่นำมาวิเคราะห์มีลักษณะที่เปลี่ยนไปบ้าง ควรมีขั้นตอนของการลดสัญญาณรบกวนเป็นขั้นตอนแรก ซึ่งเป็นส่วนที่สำคัญสำหรับการรู้จำเสียง จึงควรทำขั้นตอนนี้ให้สามารถได้เสียงที่มีความสมบูรณ์เพียงพอที่จะนำไปวิเคราะห์ต่อไป

2. เมื่อได้เสียงที่อัดมาแล้ว ควรจะมีการผ่านขั้นตอนการตัดหัวตัดท้ายของเสียง เพื่อให้ได้ค่าสัมประสิทธิ์จากเสียงอย่างแท้จริง

3. ขนาด 64 โค้ดบุค (ภาคผนวก ก.) สำหรับต้นแบบที่ใช้ทดสอบในปริญญานิพนธ์นี้ เพียงพอกับการรู้คำเสียง แต่ถ้าต้นแบบมีจำนวนมากกว่านี้ ขนาดของ โค้ดบุค ก็ควรจะมากกว่านี้

4. HMM เป็นวิธีการที่สามารถใช้กับการรู้จำคำต่อเนื่อง ดังนั้นควรพัฒนาให้รู้จำคำต่อเนื่อง ที่เป็นคำสั่งสั้น ๆ จะเป็นประโยชน์ในการใช้งานด้านอื่นๆ ต่อไป

5. ขั้นตอนการคำนวณในโปรแกรมต่าง ๆ ที่ใช้ในปริญญานิพนธ์นี้เป็นเพียงโปรแกรมที่เขียนเพื่อเป็นการจำลองการใช้งานจริงเท่านั้น ยังมิได้คำนึงถึงการใช้งานจริงแต่อย่างใด ดังนั้นการคำนวณจึงใช้เวลานานพอสมควร จึงควรพัฒนาโดยใช้โปรแกรมที่มีการประยุกต์มากขึ้น เช่น ภาษา C

6. พัฒนาขั้นตอนวิธีที่ได้ สร้างเป็นวงจร (Hardware) และนำไปใช้ในการบีบอัดสัญญาณเสียงเพื่อประโยชน์ในการติดต่อสื่อสารต่อไป

7. เพิ่มข้อมูลสัญญาณการรู้จำ จากคำเป็นประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

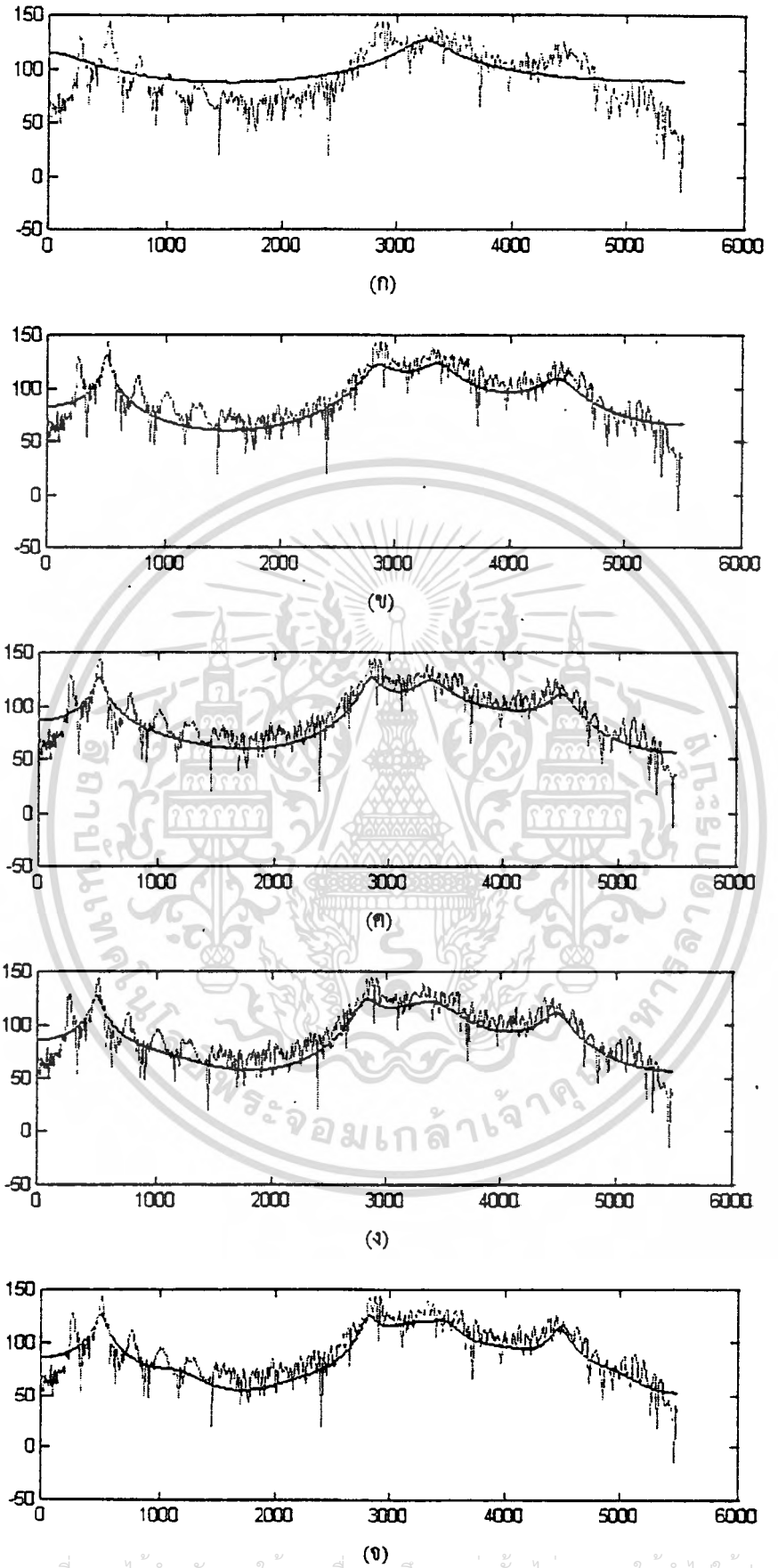
ภาคผนวก ก

การหาค่าพารามิเตอร์ที่เหมาะสมเพื่อใช้ในการทดลอง

1. จำนวนสัมประสิทธิ์ (P)

มีการทดลองการประมวลเสียงพูดโดยการประมาณเชิงเส้น(สุนทร อรอินทร์ และ อัฐ เครือพัก 2537) ซึ่งเป็นการทดลองที่มีการนำค่าสัมประสิทธิ์ LPC ที่หาได้โดยการใช้วิธีอัตสัมพันธ์ มาทำการสังเคราะห์เสียงขึ้นมาใหม่ จากการทดสอบดังกล่าว ให้ผลเป็นที่น่าพอใจ งานปริญญานิพนธ์ชิ้นนั้นมีการใช้ LPC ด้วยจำนวน P= 4 , 8 , 12 ,16 และ 20 ซึ่งใช้สำหรับเปรียบเทียบสเปคตรัม และสัญญาณความผิดพลาดระหว่างเสียงจริงและเสียงที่ได้จากการที่สังเคราะห์ขึ้นด้วยการประมาณเชิงเส้น ได้ผลสรุปว่า การประมาณสเปคตรัมจะมีความใกล้เคียงเสียงจริงมากยิ่งขึ้น เมื่อค่า P ยิ่งสูงขึ้น ดังรูปที่ ก.1 ส่วนรูปที่ ก.2 เป็นการแสดงสัญญาณความผิดพลาดที่เกิดจากการประมาณเชิงเส้น พบว่าสัญญาณความผิดพลาดจะมีขนาดลดลงเมื่อค่า P สูงขึ้น อย่างไรก็ดีถึงแม้การเพิ่มค่า P จะทำให้การประมาณเสียงมีความใกล้เคียงมากยิ่งขึ้น แต่ก็ทำให้การคำนวณมีความยุ่งยาก และใช้เวลานานขึ้น ดังนั้น เพื่อให้ได้ความเหมาะสมปริญญานิพนธ์นี้จึงเลือกใช้ค่า P=12

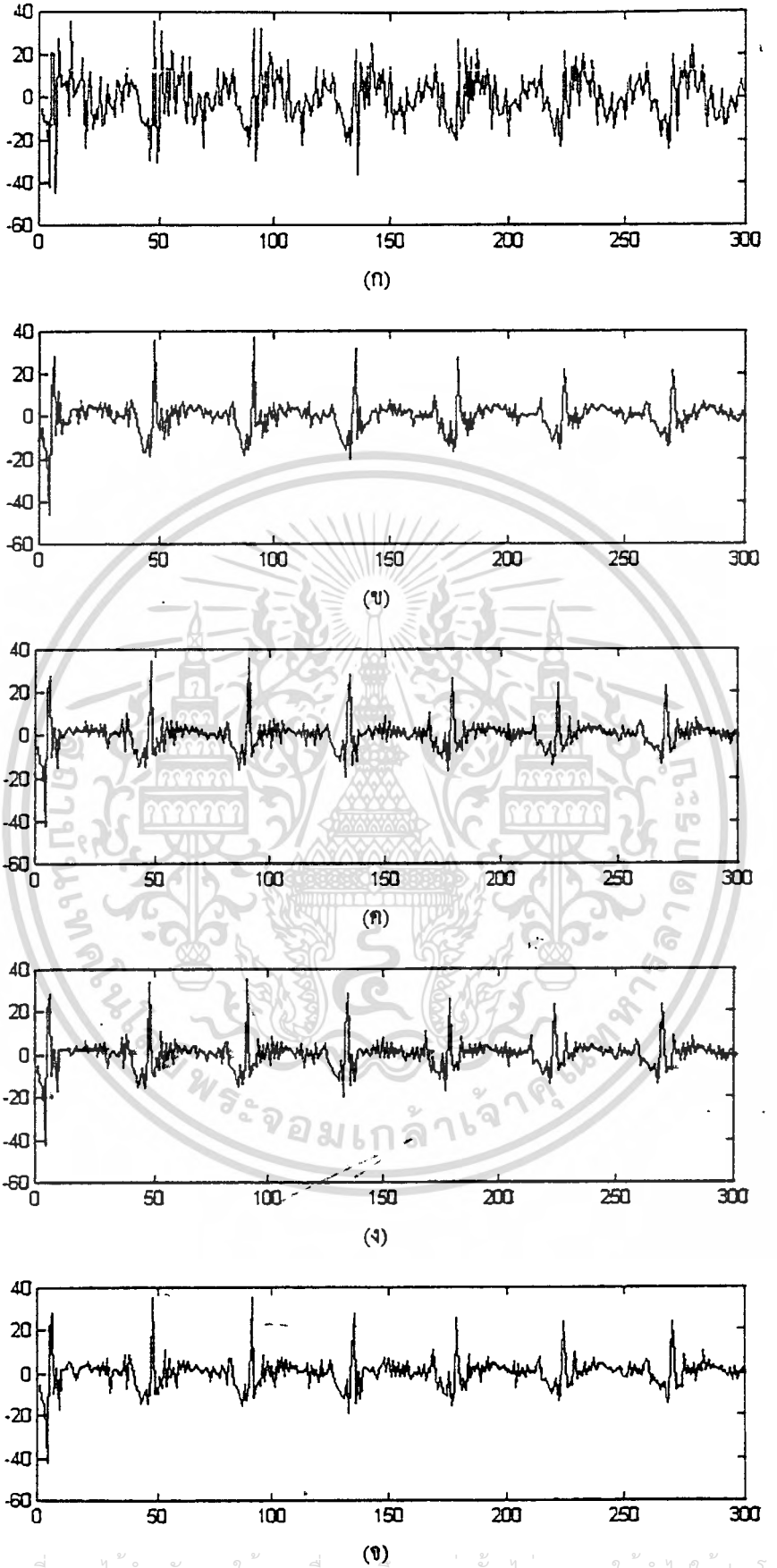
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.1 สเปคตรัมของเสียงเปรียบเทียบกับสเปคตรัมที่ได้จากสัมประสิทธิ์ LPC

(ก) p=4 (ข) p=8 (ค) p=12 (ง) p=16 (จ) p=20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.2 สัญญาณความคลาดเคลื่อน (ก) p=4 (ข) p=8 (ค) p=12 (ง) p=16 (จ) p=20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ขนาดของโค้ดบุค

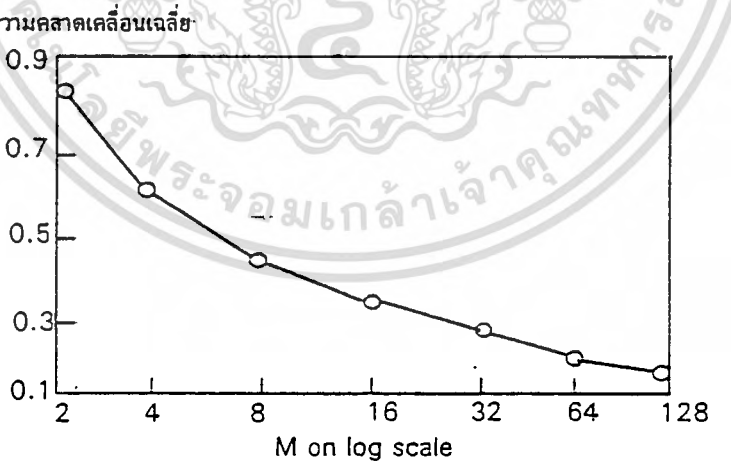
ในการทดสอบหาค่าขนาดของโค้ดบุคที่เหมาะสม (L.R. Rabiner , S.E. Levinson และ M.M. Sondhi,1982) เป็นการทดสอบโดยใช้เวคเตอร์ของ LPC จำนวน 39,708 เวคเตอร์ จากคน 100 คน (ชาย 50 คน หญิง 50 คน) พูดตัวเลขภาษาอังกฤษ 0-9 จากนั้นนำเวคเตอร์มาหาค่าควอนไตซ์ที่โค้ดบุคขนาดต่าง ๆ M = 2 ,4 , 8 , 16 , 32 , 64 , 128 ใช้การพิจารณาดังนี้

ค่าความคลาดเคลื่อนเฉลี่ย $\|D_m\|$

$$\|D_m\| = \min_{\hat{a}_m} \left\{ \frac{1}{I} \sum_{i=1}^{I} \min_{1 \le m \le M} \left[d(\hat{a}_m, a_i) \right] \right\}$$

โดย a_i เป็นเวคเตอร์ LPC I= 1,2,3,...,I $\hat{a}_m$ เป็นโค้ดบุค LPC เวคเตอร์ m = 1 , 2 , ... , M

จากผลการทดสอบจะได้ดังรูป ก.3 โดยรูป ก.3 เป็นการพลอตค่า $\|D_m\|$ กับค่า M (ขนาดโค้ดบุค โดยใช้ log scale) โดย M มีค่า 2 - 128 เราจะเห็นว่าที่ $M \geq 32$ ค่าความคลาดเคลื่อนเฉลี่ยจะต่ำกว่า 0.3 ถ้า M = 64 จะมีค่า $\|D_m\|$ ประมาณ 0.2 ซึ่งค่า $\|D_m\| < 0.3$ เป็นค่าความคลาดเคลื่อนที่น้อยมากสำหรับเวคเตอร์ ควอนไตซ์เซชัน (S.E. Levinson,L.R. Rabiner , A.E. Rosenberg และ J.G. Wilpon,1979) ดังนั้นสำหรับวิธีของ HMM ควรใช้ M = 64 เพราะจะมีค่าความคลาดเคลื่อนเฉลี่ยน้อย เมื่อ M = 64 - 128 (L.R. Rabiner , S.E. Levinson และ M.M. Sondhi,1982)



รูป ก.3 เปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกับขนาดโค้ดบุค

จากผลการทดสอบโดยใช้โปรแกรมที่พัฒนาขึ้น (วิทยานิพนธ์ เสาวลักษณ์ อารีย์พงศา 1995) นำข้อมูลจากกลุ่มต้นแบบกลุ่มที่ 1 และกลุ่มที่ 2 จำนวน 20 คน มาหาแบบจำลองต้นแบบ และทำการทดสอบความถูกต้องในการรู้จำ

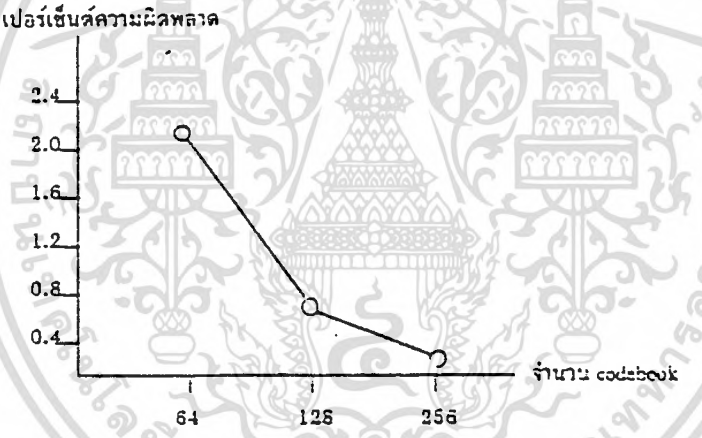
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น และห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1) ใช้กลุ่มที่ 1 และกลุ่มที่ 2 ทดสอบกับแบบจำลองต้นแบบที่คำนวณได้ เปรียบเทียบกับขนาดโค้ดบุคดังตารางที่ ก.1 และกราฟรูปที่ ก.4

ตาราง ก.1 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มต้นแบบ

จำนวนโค้ดบุค	ความผิดพลาด (%)
64	2.25
128	0.75
256	0.25

รูปที่ ก.4 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มต้นแบบ

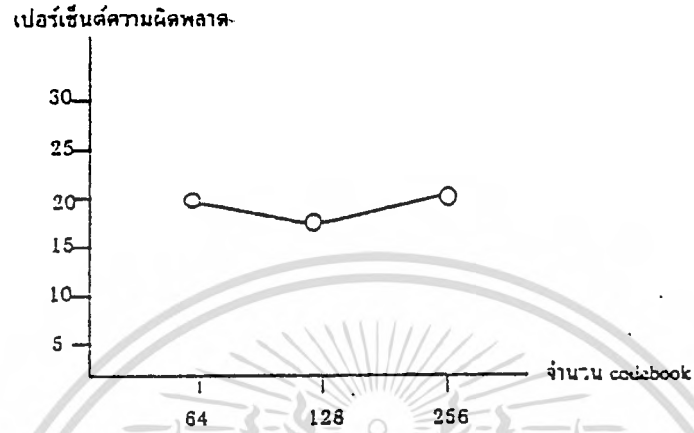


2.2) ใช้กลุ่มทดสอบมาทดสอบกับแบบจำลองต้นแบบที่คำนวณได้ ซึ่งเป็นบุคคลที่ไม่มีการฝึกฝน ดังแสดงผลความผิดพลาดในการรู้จำที่ขนาดโค้ดบุคต่าง ๆ ดังตารางที่ ก.2 และรูปที่ ก.5

ตารางที่ ก.2 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มทดสอบ

จำนวนโค้ดบุค	ความผิดพลาด (%)
64	20
128	18
256	21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



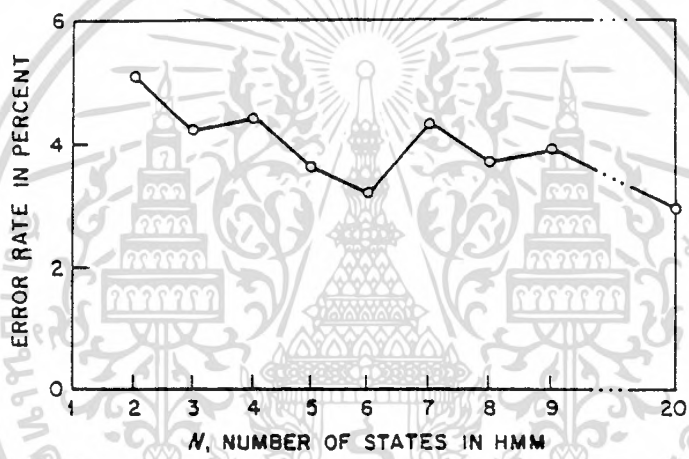
รูปที่ ก.5 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาดโค้ดบุคของกลุ่มทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. <u>การเลือกพารามิเตอร์ของแบบจำลอง (model)</u>

ในปริญญานิพนธ์นี้ได้เลือกสร้างแบบจำลอง HMM เป็นแบบ left-right เพราะเป็นแบบจำลองที่เหมาะสำหรับการจดจำรูปแบบของคำ ประเภทคำศัพท์เดี่ยว (Isolated word) เนื่องจากจะสามารถนำเวลาเข้ามาเกี่ยวข้องกับสเตท (state) ของแบบจำลองได้โดยตรง และอาจตีความหมายทางกายภาพของสเตทเป็นเสียงที่แตกต่างกันของคำได้

การเลือกจำนวนสเตท จะใช้จำนวนสเตทเท่า ๆ กันทุก ๆ คำ (0 - 9) จากรูปที่ ก.6 แสดงอัตราการผิดพลาด เมื่อใช้ จำนวนเสตท (N) ค่าต่าง ๆ สำหรับตัวเลขคำเดี่ยว ซึ่งจะเห็นว่ามีค่าต่ำสุดที่ N = 6 ในปริญญานิพนธ์นี้จึงเลือกใช้จำนวนสเตทเท่ากับ 6 สเตท

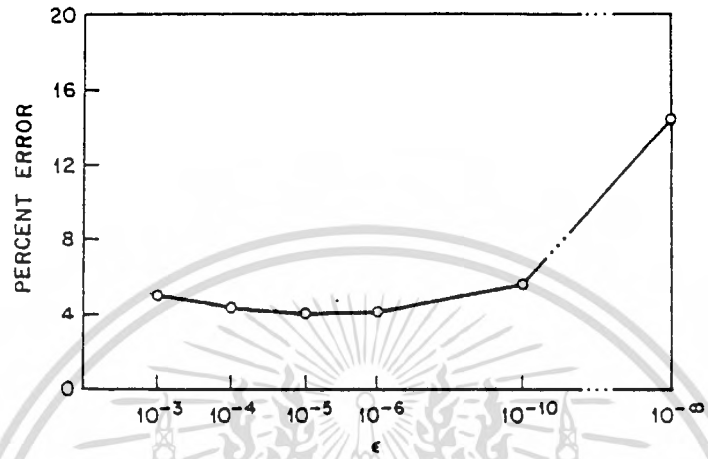


รูปที่ ก.6 แสดงอัตราการผิดพลาดเมื่อใช้จำนวนสเตท (N) ค่าต่าง ๆ

สำหรับเวคเตอร์เหตุการณ์ เนื่องจากในปริญญานิพนธ์นี้ทำแบบจำลองเป็นแบบไม่ต่อเนื่องตามเวลา เราได้ใช้วิธีการประมาณเชิงเส้นร่วมกับสัมประสิทธิ์เซปสตรัมและการเวทค่าพารามิเตอร์ ทั้งยังใช้โค้ดบุคในการสร้างสัญลักษณ์แบบไม่ต่อเนื่องขึ้น

ในการเทรน (traing) HMM นั้นต้องมีการจำกัดการคำนวณพารามิเตอร์ $b_j(k)$ เมื่อป้องกันไม่ให้คำนั้นน้อยเกินไปโดยการกำหนดค่าที่น้อยที่สุดที่จะเป็นไปได้ไว้ ($\in$) นั่นคือค่า $b_j(k)$ จะต้องมากกว่าหรือเท่ากับค่านี้ เนื่องจากอาจจะไม่เกิดเหตุการณ์ที่ k^{th} เกิดขึ้นเลยในสเตท j ในเหตุการณ์ที่เอาเข้ามาเทรน ซึ่งทำให้ได้ค่าพารามิเตอร์ที่ไม่ถูกต้อง และจำกัดค่าความน่าจะเป็นของการเกิดเหตุการณ์ลำดับของเหตุการณ์ที่เราไม่รู้ การกำหนดค่าพารามิเตอร์ $\in$ นี้จะอยู่ในช่วง $10^{-10} \leq \in \leq 10^{-3}$ ซึ่งรูปที่ ก.7 ได้แสดงค่าความผิดพลาดในกรณีที่ใช้ $\in$ ค่าต่าง ๆ ในปริญญานิพนธ์นี้เลือกใช้ค่า $\in = 10^{-4}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.7 แสดงค่าความผิดพลาดในกรณีที่ใช้ ϵ ค่าต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

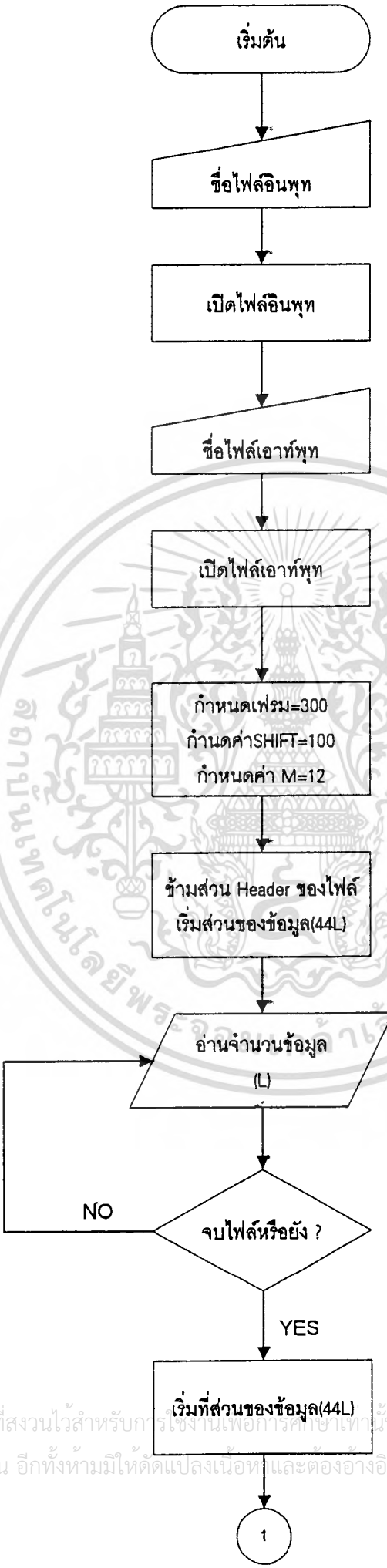
โปรแกรมที่ใช้ในการคำนวณ

ภาคผนวกนี้ได้รวบรวมโปรแกรมต่าง ๆ ที่ใช้ในการคำนวณ และวิเคราะห์โดยแบ่งออกเป็น

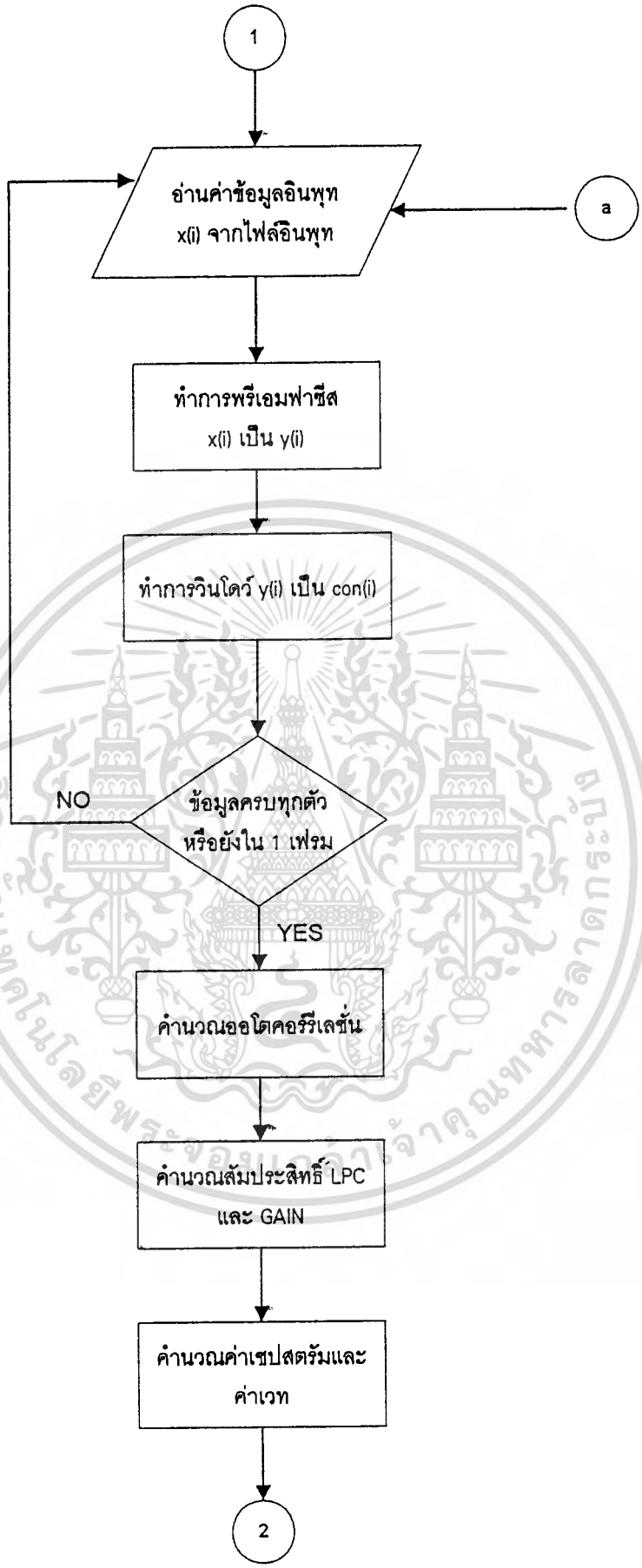
- โปรแกรมการคำนวณค่าสัมประสิทธิ์ ชื่อ LPC
- โปรแกรมการสร้างโค้ดบุค ชื่อ CODEBOOK
- โปรแกรมการเปรียบเทียบเวคเตอร์เสียง กับโค้ดบุค ชื่อ COMPAIR
- โปรแกรมการสร้างแบบจำลองเสียง ชื่อ HMM
- โปรแกรมการทดสอบเสียง ชื่อ RECOGNIZE
- โปรแกรมการทดสอบเสียงทีละเสียง ชื่อ TEST

โดยที่ โปรแกรมการคำนวณค่าสัมประสิทธิ์ ชื่อ LPC นั้นใช้โปรแกรมภาษา C ในการเขียน ส่วนโปรแกรมที่เหลือนั้นใช้โปรแกรม MATLAB ในการเขียน

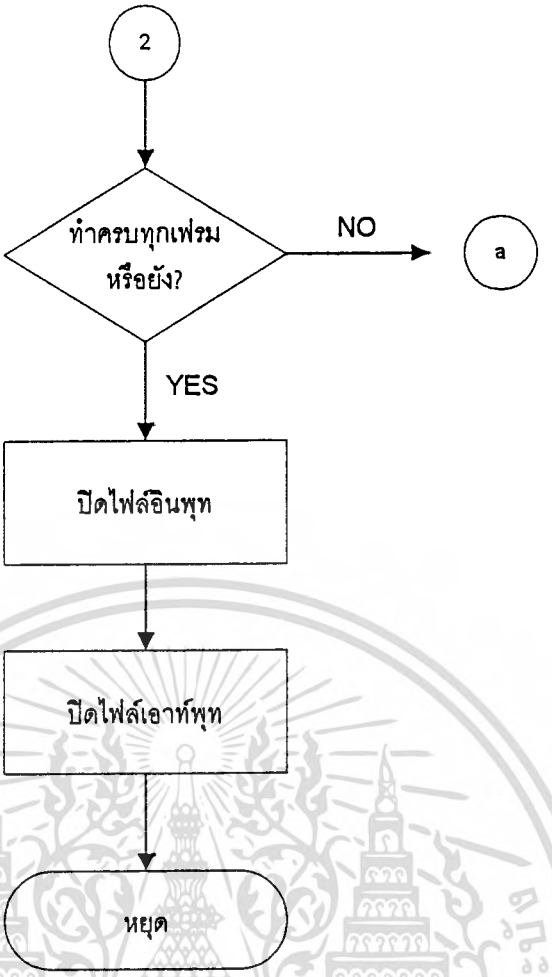
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.1 Flow Chart แสดงการทำงานของโปรแกรม LPC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม LPC

```
#include <stdio.h>
#include <stdlib.h>
#include <alloc.h>
#include <conio.h>
#include <io.h>
#include <math.h>
#define PI      3.141592654
#define FRAME   300
#define SHIFT   100
#define N       17
#define M       12
FILE *inputfile, *outputfile1, *outputfile2,*outputfile3,*outputfile4;
char  outfile1[20],outfile2[20],outfile3[20],outfile4[20],infile[20];
int   i,j,k,l,c,L,d;
void main() {
   float y[FRAME+1],con[FRAME+1];
   float a[1],R[N+1],A[N][N],X[N][N];
   float Lo[N][N],E[N],G[N],sum;
   float alf[N+2],phi[N+1],C[N+2],w[N+2];
   double g,q;
   int   x[FRAME+1],sl;
   c=0;x[0]=0;x[1]=0;a[0]=0.9375;
/* input i/o filenames to open.  */
   if(c==24){printf("Type Enter.\n");getch();c=1;};
   printf("\n Please enter input filename(and directory)...  ");c++;
   gets(infile);
   if((inputfile=fopen(infile,"rb"))==NULL) {
      if (c==24){printf("Type Enter.\n");getch();c=1;};
      printf("\n Can not open input file\007");c++;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    exit(1);
}
if (c==24){printf("Type Enter.\n");getch();c=1;};
printf("\n Please enter output filename1(and directory)... ");c++;
gets(outfile1);
if((outputfile1=fopen(outfile1,"wb"))==NULL) {
    if (c==24){printf("Type Enter.\n");getch();c=1;};
    printf("\n Can not open output file \007");c++;
    exit(1);
}
  if (c==24){printf("Type Enter.\n");getch();c=1;};
printf("\n Please enter output filename2(and directory)... ");c++;
gets(outfile2);
if((outputfile2=fopen(outfile2,"wb"))==NULL) {
    if (c==24){printf("Type Enter.\n");getch();c=1;};
    printf("\n Can not open output file \007");c++;
    exit(1);
}
  if (c==24){printf("Type Enter.\n");getch();c=1;};
printf("\n Please enter output filename3(and directory)... ");c++;
gets(outfile3);
if((outputfile3=fopen(outfile3,"wb"))==NULL) {
    if (c==24){printf("Type Enter.\n");getch();c=1;};
    printf("\n Can not open output file \007");c++;
    exit(1);
}
  if (c==24){printf("Type Enter.\n");getch();c=1;};
printf("\n Please enter output filename4(and directory)... ");c++;
gets(outfile4);
if((outputfile4=fopen(outfile4,"wb"))==NULL) {
    if (c==24){printf("Type Enter.\n");getch();c=1;};
    printf("\n Can not open output file \007");c++;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        exit(1);
    }
/* end open file */
/* Find how long of data*/
    L=0;
    fseek(inputfile,44L,SEEK_CUR);
    while (!feof(inputfile)) {
            d=fgetc(inputfile);
            L++;
    }
    if (c==24){printf("Type Enter.\n");getch();c=1;};
    printf ("File is long %d charector.\n",L);getch();c++;
    rewind(inputfile);
    fseek(inputfile,44L,SEEK_CUR);
/* end find the long of data */
/* Preemphasis and window */
    for (j=1;j<=(int)((L-FRAME+SHIFT)/SHIFT);j++) {
        if (j==1) {
            for (i=1;i<=FRAME;i++) {
                d=fgetc(inputfile);
                x[i]=(int)d;
                y[i]=0;
                y[i]=(x[i]-a[0]*(float)x[i-1]);
                con[i]=y[i]*(0.54-0.46*cos(2*PI*(i-1)/(FRAME-1)));
            }
        }/* end if (loop 1)*/
        else {
            for (i=1;i<=(FRAME-SHIFT);i++) {
                x[i]=x[i+SHIFT];
                y[i]=y[i+SHIFT];
                con[i]=y[i]*(0.54-0.46*cos(2*PI*(i-1)/(FRAME-1)));
            }
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            for (i=(FRAME-SHIFT+1);i<=FRAME;i++) {
                d=fgetc(inputfile);
                x[i]=(int)d;
                y[i]=0;
                y[i]=(x[i]-a[0]*(float)x[i-1]);
                con[i]=y[i]*(0.54-0.46*cos(2*PI*(i-1)/(FRAME-1)));
            }
        }/* end else (loop 2..n)*/
    /* end of preemphasis and windowing */
    /* find R[0..12] */
        for (i=0;i<=M;i++) {
            R[i]=0;
            for (k=0;k<=(FRAME-1-i);k++) {
                R[i]+=con[k+1]*con[k+i+1];
            }
    /*      if (c==23){printf("Type Enter.\n");getch();c=1;};
                printf(" frame %d R[%d] =%.4f \n",j,i,R[i]);c++;*/
        }
    /* define matrix */
        for (i=0;i<=M-1;i++) {
            sl=0;
            for (k=i;k<=M-1;k++) {
                A[i][k]=R[sl];sl++;
                A[k][i]=A[i][k];
            }
    /*      for (k=0;k<=N-1;k++) {
                if (c==23){printf("Type Enter.\n");getch();c=1;};
                printf(" A[%2d,%2d]= %.4f\n",i,k,A[i][k]);c++;
            } */
        }/* end define matrix */
    /* to find inverse matrix */
        for (i=0;i<=M-1;i++) {
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        for (k=0;k<=M-1;k++) {
                Lo[i][k]=0;
                if (i==k) {
                   Lo[i][k]=1;
                }
        }
}
for (i=0;i<=M-1;i++) {
         for (k=i+1;k<=M-1;k++) {
                  Lo[k][i]=A[k][i]/A[i][i];
                  for (l=0;l<=M-1;l++) {
                          A[k][l]=A[k][l]-Lo[k][i]*A[i][l];
                  }
         }
}
for (i=0;i<=M-1;i++) {
        for (k=0;k<=M-1;k++) {
                G[k]=0;
                E[k]=0;
        }
        for (k=0;k<=M-1;k++) {
                if (k==i) {
                   E[k]=1;
                }
                else {
                   E[k]=0;
                }
        }
        for (k=0;k<=M-1;k++) {
                sum=0;
                for (l=0;l<=M-1;l++) {
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                sum+=Lo[k][l]*G[l];
            }
            G[k]=E[k]-sum;
        }
        X[M-1][i]=G[M-1]/A[M-1][M-1];
        for (k=M-2;k>=0;k--) {
            sum=0;
            for (l=k+1;l<=M-1;l++) {
                sum+=A[k][l]*X[l][i];
            }
            X[k][i]=(G[k]-sum)/A[k][k];
        }
    }
/* find phi[1..N] */
    for (i=1;i<=M;i++) {
        phi[i]=R[i];
/*      if (c==23){printf("Type Enter.\n");getch();c=1;};
        printf(" phi[%d] = %.4f\n",i,phi[i]);c++;*/
    }
/* find coefficient of LPC.*/
    for (i=0;i<=M-1;i++) {
        alf[i+1]=0;
        for (k=0;k<=M-1;k++) {
            alf[i+1]+=X[i][k]*phi[k+1];
        }
    }
    alf[0]=0;  alf[13]=0;alf[14]=0;alf[15]=0;alf[16]=0;alf[17]=0; alf[18]=0;
    for (i=0;i<=N+1;i++) {
        if (c==23){printf("Type Enter.\n");getch();c=1;};
        printf("frame %d alf[%d] = %.16f\n",j,i,alf[i]);c++;
        fwrite(&alf[i],sizeof(alf[i]),1,outputfile1);
    }
```

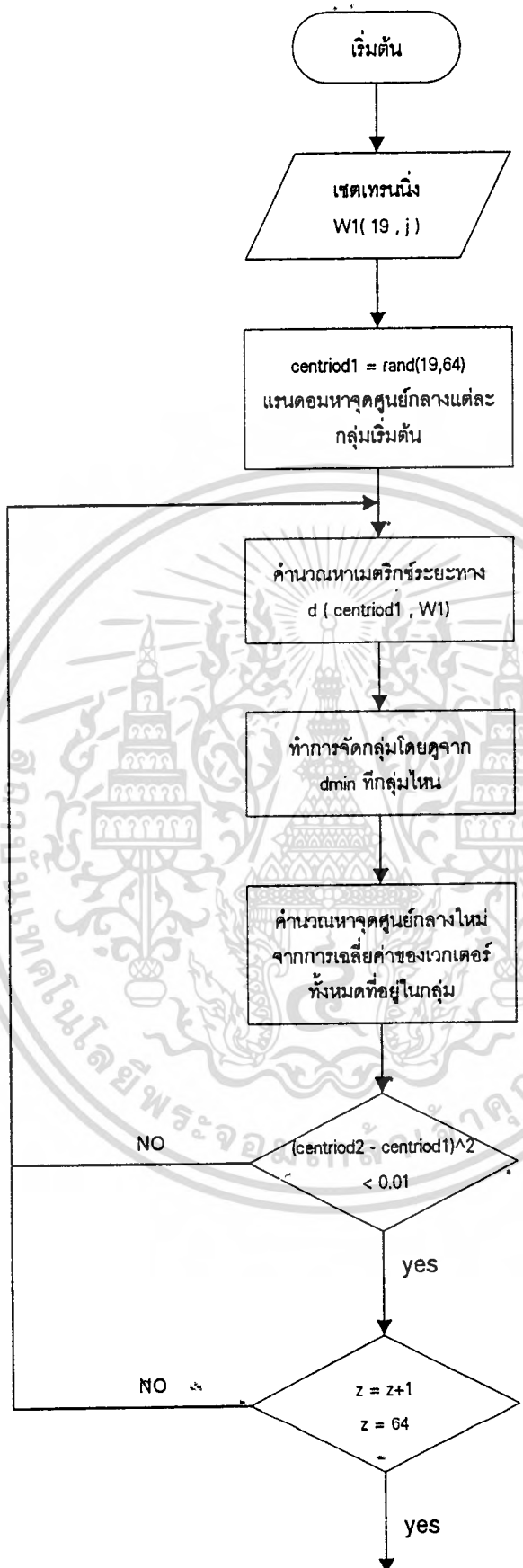
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 /* find Gain */
q=(double)R[0];
for (i=1;i<=M;i++) {
        q-=(double)alf[i]*(double)R[i];
}
 if (c==23){printf("Type Enter.\n");getch();c=1;};
  printf("\n Frame %d q= is %.16f\n\n",j,q);c++;
g=sqrt(q); fwrite(&g,sizeof(g),1,outputfile2);
if (c==23){printf("Type Enter.\n");getch();c=1;};
printf("\n Frame %d Gain is %.16f\n\n",j,g);c++;
/* find  Cepstrum */
C[0]=log(g);
for (k=1;k<=N+1;k++) {
        if (k<=12) {
           C[k]=alf[k];
        }
        else {
           C[k]=0;
        }
        for (i=1;i<=k-1;i++) {
            C[k]+=((float)i/(float)k)*C[i]*alf[k-i];
        }
}
/*find  Weight */
for (i=0;i<=N+1;i++) {
        w[i]=C[i]*(1+((N+1)/2)*sin(PI*(double)i/(N+1)));
        fwrite(&C[i],sizeof(C[i]),1,outputfile3);
        fwrite(&w[i],sizeof(w[i]),1,outputfile4);
}
for (i=0;i<=N+1;i++) {
    if (c==23){printf("Type Enter.\n");getch();c=1;};
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    printf("c[%d] = %.16f    w[%d] = %.16f\n",i,C[i],i,w[i]);c++;
        }
    }
    fclose(inputfile);
    fclose(outputfile1);
    fclose(outputfile2);
    fclose(outputfile3);
    fclose(outputfile4);
}/* main */
```


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข.2 Flow Chart แสดงการทำงานของโปรแกรม CODEBOOK

โปรแกรม CODEBOOK

```
% เวคเตอร์ควอนไตซ์เซชัน
clear;
% โหลดเทรนนิ่งเซตที่จะใช้ในการสร้างโค้ดบุค
disp(' input ');
fid=fopen('c:\matlab\bin\aliw.m','rb');
w1=fread(fid,[19,10833],'float64');
status=fclose(fid);
f=64;                % กำหนดขนาดโค้ดบุคเป็น 64
% สุ่มเวคเตอร์ขึ้นมาเพื่อให้เป็นจุดศูนย์กลางเริ่มต้นของแต่ละกลุ่ม
centroid1=rand(19,f);
disp(centroid1)
% หาเมตริกซ์ระยะทางหรือค่าความคลาดเคลื่อน
c=[];
for j=1:length(w1)
        for m=1:f
        total=0;d=0;
                for k=1:19
                    d=(w1(k,j)-centroid1(k,m))^2;
                    total=total+d;
                end
        c(m,j)=total;
        end
end
% หาระยะทางน้อยที่สุดอยู่ที่กลุ่มไหน
[s,n]=min(c);
disp('n=')
disp(n)
% หาจุดศูนย์กลางใหม่
v=[];
cent1=[];
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
centroid2=[];
for m =1:f
        count =0;
        for j=1:length(w1)
                if  n(j)== m
                    count =count+1;
                        for k =1:19
                            v(count,k)=w1(k,j);
                        end
                end
        end
        if count  ~= 0
                cent1 =mean(v);
                cent1=cent1';
                centroid2(:,m) = cent1 ;
        else
                    centroid2(:,m) = centroid1(:,m) ;
        end
end
disp('centroid2')
disp(centroid2)
% ตรวจดูว่ามีค่าความคลาดเคลื่อนน้อยกว่า 0.01 แล้วหยุดการทำงาน
true =0;
for m= 1:f
        total = 0;
        ww=0;
        for k = 1:19
            ww= (centroid2(k,m) - centroid1(k,m))^2 ;
            total = total+ww ;
        end
        disp('total1')
        disp(total)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        if total >0.01
            true = 1;
        end
    end
    while  true == 1
        e=[ ];
        for j =1: length(w1)
            for m =1:f
                total =0;
                d =0;
                for k =1:19
                    d= (w1(k,j) - centroid2(k,m))^2;
                    total =total + d;
                end
            e(m,j) = total ;
            end
        end
        [ f ,g ] = min(e);
        disp('g')
        disp(g)
        h=[ ];
        cent2 = [ ];
        centroid3=[ ];
        f=64;
        for m = 1:f
            count = 0;
            for j = 1: length(w1)
                if g(j) == m
                    count = count + 1;
                    for k =1:19
                        h(count,k) = w1(k,j);
                    end
```

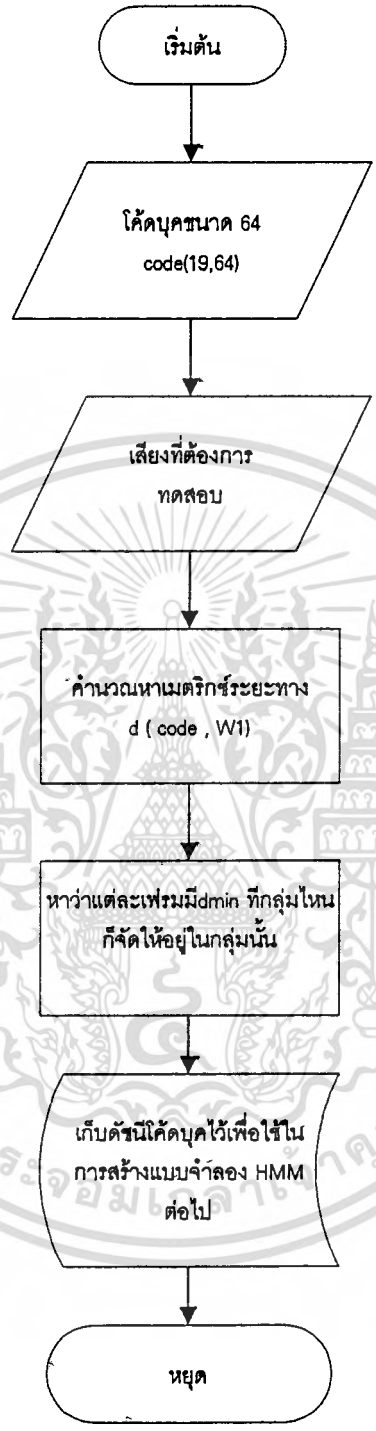
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                end
            end
            f=64;
            if count  ~= 0
                cent2 = mean(h);
                cent2 =cent2';
                centroid3(:,m) = cent2;
            else
                centroid3(:,m) = centroid2(:,m);
            end
        end
        disp('centroid3')
        disp(centroid3)
        z=0;
        for m = 1: f
            total = 0;
            for k = 1: 19
                u = (centroid3(k,m) - centroid2(k,m))^2;
                total = total + u;
            end
            disp('total2')
            disp(total)
            if total <= 0.01
                z=z+1;
            end
        end
        if z ==64 |  centroid3 == centroid2
            true =0;
        else
        centroid2 = centroid3;
        end
    end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
j=1:length(w1) ;
m = 1:64;
% พล็อตกราฟแสดงตำแหน่งของเฟรมทั้งหมดของเทรนนิ่งเซตกับเวคเตอร์โค้ดที่ได้
plot(w1(7,j),w1(8,j),'g.' ,centroid1(7,m),centroid1(8,m),'r+',centroid3(7,m),centroid3(8,m),'b+')
disp(' output ')
fid=fopen('c:\matlab\bjn\code_w20.m','wb+');
count=fwrite(fid,centroid3,'float64');
status=fclose(fid);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.3 Flow Chart แสดงการทำงานของโปรแกรม COMPAIR

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม COMPAIR

```
clear;
% โหลดไฟล์ .wav ที่จะใช้ในการวิเคราะห์ เข้ามาเก็บไว้ในตัวแปรชื่อ inpt
[inpt,fp,form]=loadwave('c:\input2\w2_zero3.wav');
% เซตค่าคงที่ต่าง ๆ ดังนี้
lpc=[];
cep=[];
p=12;           % จำนวนสัมประสิทธิ์ของวงจรกรอง
frame=300;      % ความยาวของเฟรม
shift=100;      % ระยะในการเลื่อนเฟรม
q=18;           % จำนวนสัมประสิทธิ์ของเซปสตรัม
% นำ input ชื่อ inpt มาผ่านการเอมฟาซิส แล้วเก็บไว้ในตัวแปร inlpc
inlpc=filter([1,-0.9375],[1],inpt);
% การคำนวณพารามิเตอร์ต่างๆ ในการประมาณ เก็บไว้ในตัวแปร a(j,i) และ g(i)
for i=1:(length(inpt)-frame+shift)/shift;
        format long;
        % คำนวณ input ที่จะใช้ในการคำนวณ เมื่อกำหนดขนาด(frame) และจุดเริ่มต้นของข้อมูล
        for m=1:frame;
                z(m)=inlpc(m+(shift*(i-1)));
        end
        % นำสัญญาณที่ได้มาผ่านการวินโดว์เพื่อเตรียมวิเคราะห์ต่อไป
        for m=1:frame;
                [x(m)]=z(m)*(0.54-0.46*cos(2*pi*(m-1)/(frame-1)));
        end
        % หาออโตคอร์รีเลชันและเมตริกซ์ของออโตคอร์รีเลชัน
        r=xcorr(x);
        for m=1:p,
                for j=m:p,
                        s(m,j)=r(frame+(j-m));
                        s(j,m)=s(m,j);
                end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end
for m=1:p,
        phi(m)=r(frame+m);
end
a=inv(s)*phi';
b=[1,-a'];
c=roots([1,-a']);
for m=1:p,
        if abs(c(m))>1
                b=deconv(b,[1,-c(m)]);
                b=conv(b,[1,-1/c(m)]);
        end
end
% คำนวณสัมประสิทธิ์ของวงจรกรอง(lpc)
for m=1:p+1,
        if m==1
                a(m)=0;
        else
                a(m)=-b((m-1)+1);
        end
end
a(14)=0;  a(15)=0;  a(16)=0;  a(17)=0;  a(18)=0;  a(19)=0;
lpc(:,i)=a;
% คำนวณอัตราการขยาย G
e=r(frame);
for m=1:p+2,
        e=e-a(m)*r(frame+(m-1));
end
g(i)=sqrt(e);
```

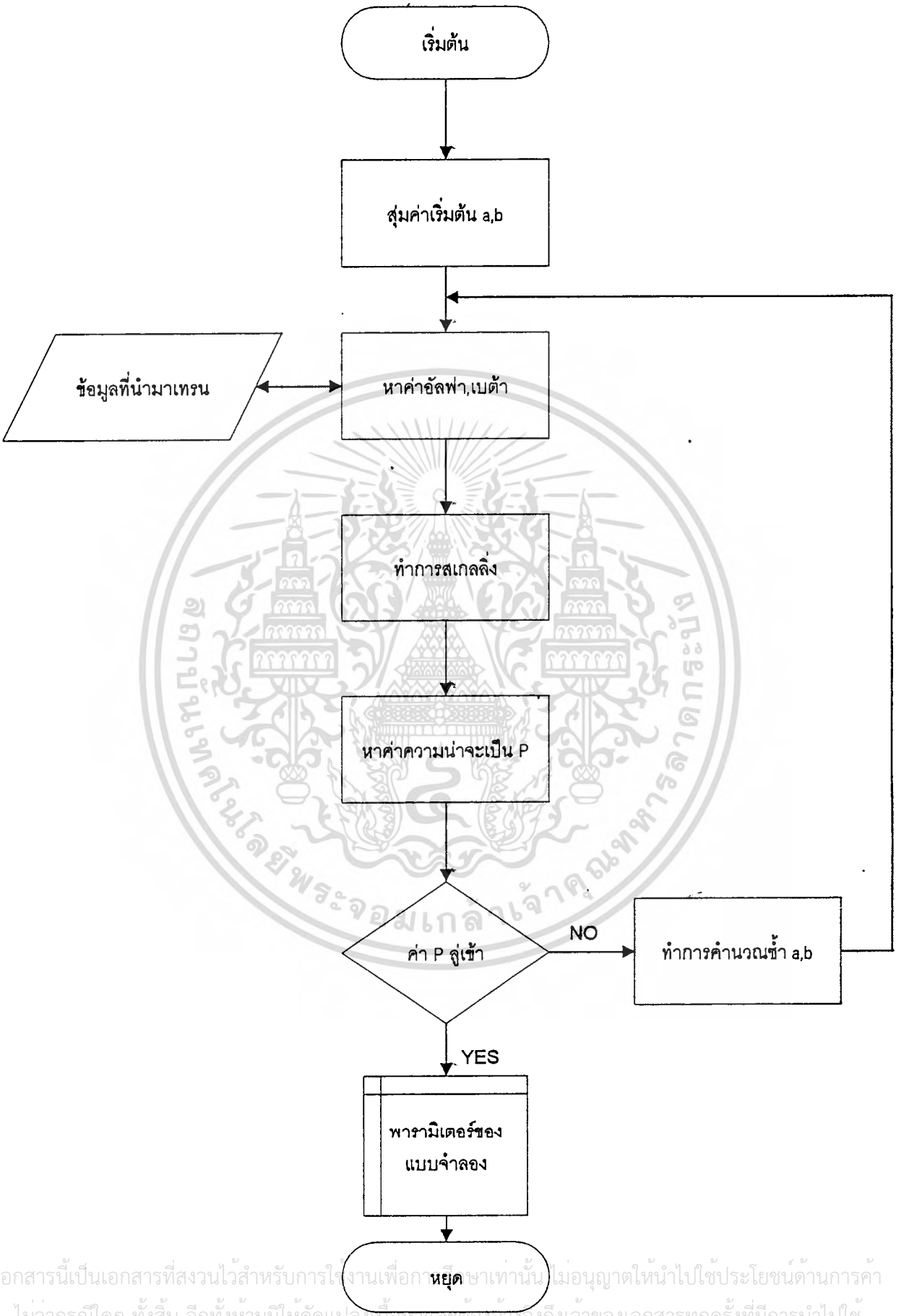
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        % คำนวณสัมประสิทธิ์เซปสตรัม(cep)
        for m=0:1,
                if m==0
                        v(m+1) = log (g(i));
                else
                        v(m+1) = lpc((m+1),i);
                end
        end
        for m=2:q,
                if m<=p
                        v(m+1)=lpc((m+1),i);
                else
                        v(m+1)=0;
                end
                for k=1:m-1,
                        v(m+1)=v(m+1)+(k/m)*v(k+1)*lpc(((m+1)-k),i);
                end
        end
        [cep(:,i)]=v';
        % การเวทค่าสัมประสิทธิ์(y)
        for m=1:q+1,
                y(m)=cep(m,i)*(1+(q/2)*sin((pi*(m-1))/q));
        end
        [w1(:,i)]=y';
end
disp(' input ');
fid = fopen('c:\matlab\bin\code_w20.m','rb');
code=fread(fid,[19,64],'float64');
disp('  output  ');
status=fclose(fid);
c=[ ];
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

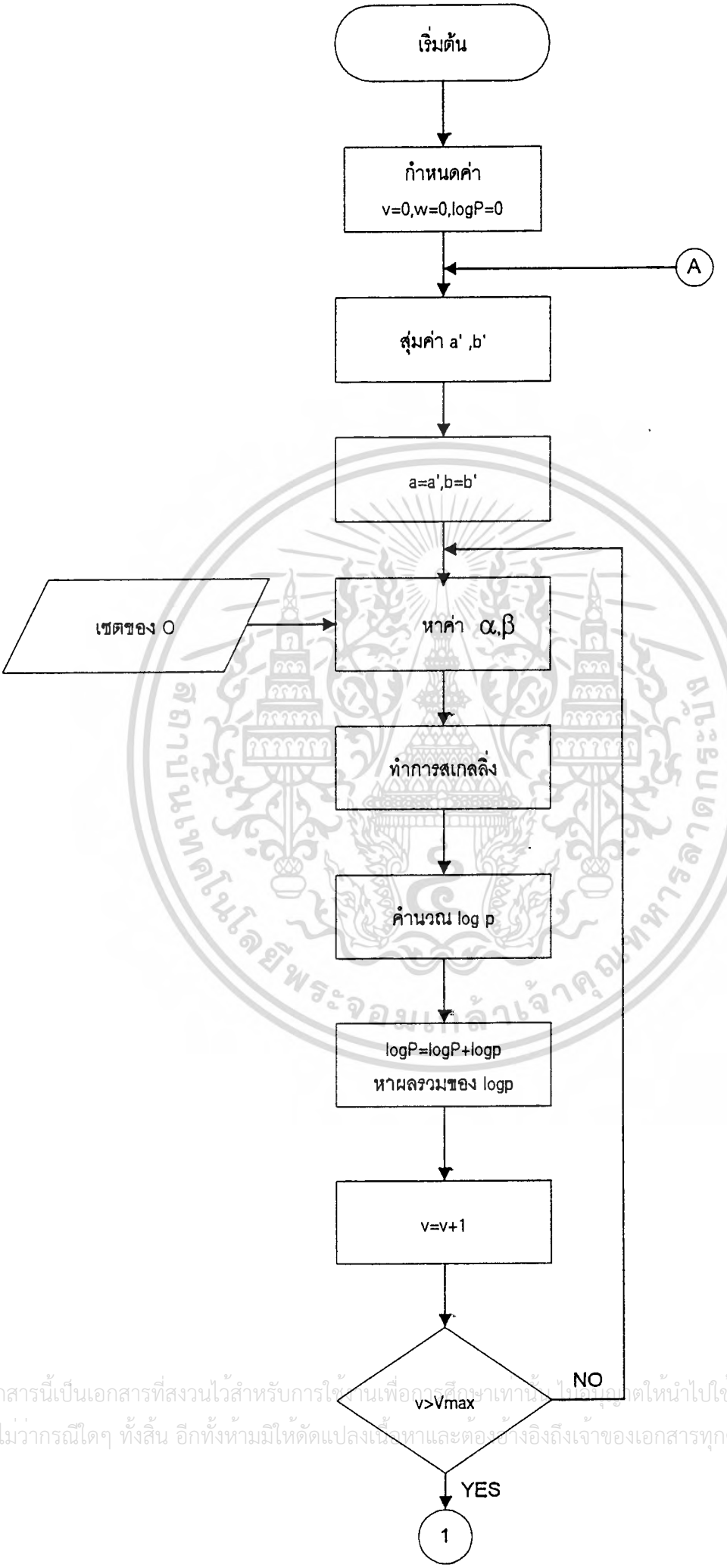
```
for j=1:length(w1)
        for m=1:64
        total=0;d=0;
                for k=1:19
                        d=( w1(k,j) - code(k,m) )^2 ;
                        total=total+d;
                end
                c(m,j)=total;
        end
end
[s,n]=min(c);
disp('n=')
disp(n)
disp(' input ');
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

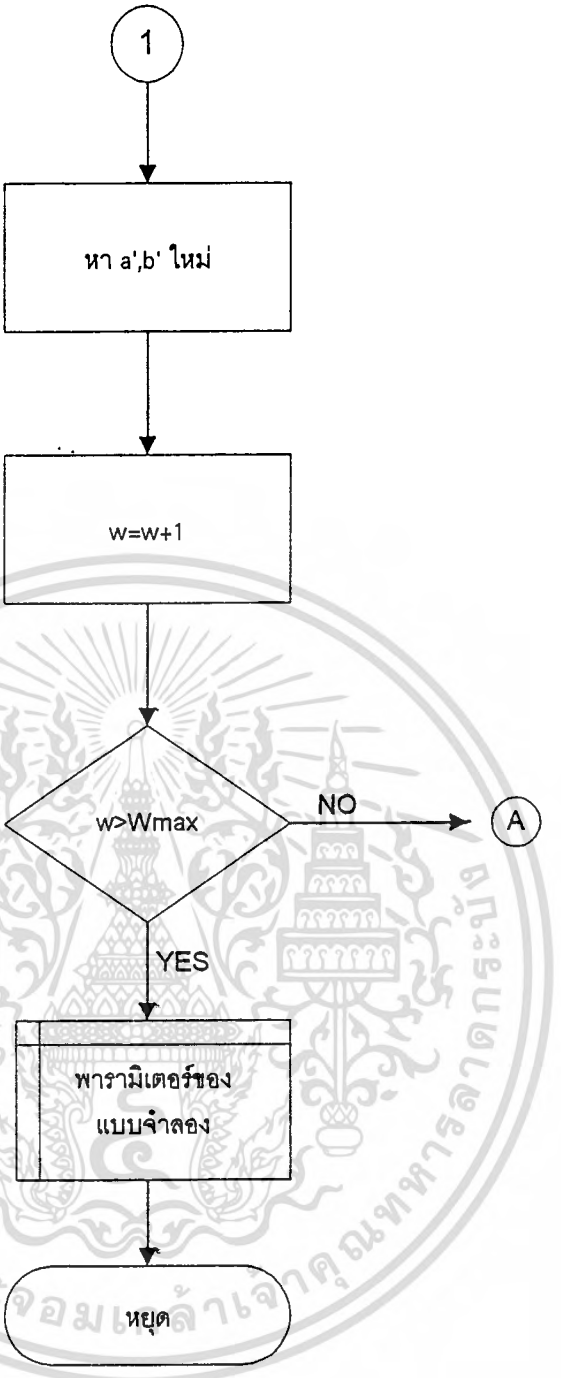


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข.4 Flow Chart แสดงอัลกอลิทึมการรีเอสติเมสพารามิเตอร์ของแบบจำลอง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.5 Flow Chart แสดงการทำงานของโปรแกรม HMM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม HMM

```
clear;
format long e;
% กำหนดค่าเริ่มต้น
load three.mat              % กำหนดลำดับ O ที่จะนำมาเทรน
N=6;                        % จำนวนสเตท
dt=[];                      % เดลต้า
K=64;                       % ค่า VQ ที่อาจปรากฏในแต่ละสเตท
x=zeros(6,6);
y=zeros(6,K);
z=zeros(1,6);
% สุ่มค่า a' , b' เริ่มต้น
a=rand(6,6);
a(1,4)=0; a(1,5)=0; a(1,6)=0; a(2,1)=0; a(2,5)=0; a(2,6)=0; a(3,1)=0; a(3,2)=0; a(3,6)=0; a(4,1)=0;
a(4,2)=0; a(4,3)=0; a(5,1)=0; a(5,2)=0; a(5,3)=0; a(5,4)=0; a(6,1)=0; a(6,2)=0; a(6,3)=0; a(6,4)=0; a(6,5)=0;
sm1=sum(a')';
for i=1:6,
        for j=1:6,
                a_prime(i,j)=a(i,j)/sm1(i);
        end
end
r=rand(6,256);
sr=sum(r')';
for j=1:6,
        for k=1:256,
                b_prime(j,k)=r(j,k)/sr(j);
        end
end
TT=[1 0 0 0 0 0];           % ค่าความน่าจะเป็นของการเริ่มที่สเตทต่างๆ
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% reestimate
for w=1:50                          % จำนวนรอบที่จะทำการ reestimate
        a=a_prime;
        b=b_prime;
        A_prime=zeros(6,6);
        AM_prime=zeros(1,6);
        B_prime=zeros(6,K);
        BM_prime=zeros(1,6);
        for v=1:10                  % จำนวนลำดับของ O ที่จะนำมาเทรน
        % หา อัลฟา,เบต้า
        % ค่าเริ่มต้นของอัลฟา,เบต้า
        lf=[];                      % อัลฟา
        Bt=[];                      % เบต้า
        lfps=zeros(T(v),6);         % อัลฟาก่อนการสเกลลิ่ง
        cB=zeros(1,T(v));
                for i=1:N
                lf(1,i)=TT(i)*b(i,O(v,1));
                Bt(T(v),i)=1;
                end
                for t=1:T(v)-1
                        for j=1:N
                        slf=0;
                                for i=1:N
                                slf=slf+lf(t,i)*a(i,j);
                                end
                        lf(t+1,j)=slf*b(j,O(v,t+1));
                        end
                end
                for t=T(v)-1:-1:1
                        for i=1:N
                                for j=1:N
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                Bt(t,i)=Bt(t,i)+a(i,j)*b(j,O(v,t+1))*Bt(t+1,j);
                end
        end
end
% ทำการสเกลลิ่ง
% หา อัลฟาสเกลลิ่ง
sc=sum(lf')';
% เมื่อ t=1
for i=1:N
lfps(1,i)=lf(1,i);
c(1)=1/sc(1);
lfs(1,i)=c(1)*lf(1,i);
end
% เมื่อ t=2-T
for t=2:T(v)
        for i=1:N
        lfps(t,i)=0;
                for j=1:N
                lfps(t,i)=lfps(t,i)+lfs(t-1,j)*a(j,i)*b(i,O(v,t));
                end
        end
c(t)=sum(lfps(t,:)')';
c(t)=1/c(t);
        for i=1:N
        fs(t,i)=c(t)*lfps(t,i);
        end
end
% หา เบต้าสเกลลิ่ง
cB(T(v))=c(T(v));
        for t=T(v)-1:-1:1
        cB(t)=cB(t+1)*c(t);
        end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        for t=1:T(v)
                for i=1:N
                Bts(t,i)=cB(t)*Bt(t,i);
                end
        end
% หา logp จาก ค่าสเกล
clog=log10(c);
plog=(-sum(clog))
% หา A_prime(ค่าเศษของ a') , AM_prime (ค่าส่วนของ a')
        for i=1:N
                for j=1:N
                x(i,j)=0;
                        for t=1:T(v)-1
                        x(i,j)=x(i,j)+lfs(t,i)*a(i,j)*b(j,O(v,t+1))*Bts(t+1,j);
                        end
                end
        end
A_prime=A_prime+x;
        for i=1:N
        z(i)=0;
                for t=1:T(v)-1
                q=0;
                        for j=1:N
                        q=q+lfs(t,i)*a(i,j)*b(j,O(v,t+1))*Bts(t+1,j);
                        end
                z(i)=z(i)+q;
                end
        end
AM_prime=AM_prime+z;
% หา B_prime (ค่าเศษของ b') , BM_prime (ค่าส่วนของ b')
        for j=1:N
                for k=1:K
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                y(j,k)=0;
                l(j)=0;
                        for t=1:T(v)-1;
                        G=0;
                                for i=1:N
                                G=G+lfs(t,j)*a(j,i)*b(i,O(v,t+1))*Bts(t+1,i);
                                end
                                if O(v,t)==k
                                y(j,k)=y(j,k)+G;
                                end
                                l(j)=l(j)+G;
                        end
                        if O(v,T(v))==k
                        y(j,k)=y(j,k)+lfs(T(v),j);
                        end
                        l(j)=l(j)+lfs(T(v),j);
                end
        end
        B_prime=B_prime+y;
        BM_prime=BM_prime+l;
        % หา  log P (ผลรวมของ logp)
        Plog=Plog+plog;
        end
Plog            % ให้แสดงค่า Plog
% หา a' , b' ใหม่
        for i=1:N
                for j=1:N
                a_prime(i,j)=A_prime(i,j)/AM_prime(i);
                end
        end
        for j=1:N
                for k=1:K
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            b_prime(j,k)=B_prime(j,k)/BM_prime(j);
            end
        end
        % ปรับปรุง b_prime
        for j=1:N
            for k=1:K
                if b_prime(j,k)<1.00e-4
                    b_prime(j,k)=1.00e-4;
                end
            end
        imb=sum(b_prime')';
        b_prime(j,:)=b_prime(j,:)/imb(j);
        end
end
% แสดงค่า a,b,TT ที่ได้จากการ reestimate
a=a_prime
b=b_prime
TT
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม RECOGNIZE

```
clear;
format long e;
N=6;                    % จำนวนสเตท
dt=[];                  % ค่าเดลต้า
K=64;                   % จำนวนค่า VQ ที่ปรากฏในแต่ละสเตท
x=zeros(6,6);
y=zeros(6,K);
z=zeros(1,6);
A_prime=zeros(6,6);
B_prime=zeros(6,K);
C_prime=zeros(1,6);
for v=1:10              % จำนวน unknown word ที่จะนำมาทดสอบ
load nine.mat           % เสียงที่จะทดสอบ
O=O(v,:);
T=T(v);
        for c=1:10      % โหลด ค่า a,b,TT จากแต่ละแบบจำลอง 0-9
        if c==1
                load hmm0.mat
        end
        if c==2
                load hmm1.mat
        end
        if c==3
                load hmm2.mat
        end
        if c==4
                load hmm3.mat
        end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
if c==5
        load hmm4.mat
end
if c==6
        load hmm5.mat
end
if c==7
        load hmm6.mat
end
if c==8
        load hmm7.mat
end
if c==9
        load hmm8.mat
end
if c==10
        load hmm9.mat
end
TT=[1 0 0 0 0 0];          % กำหนดค่าความน่าจะเป็นในการเริ่มต้นที่สเตทต่างๆ
% Viterbi Algorithm
% 1 initialization
for i=1:N,
        dt(1,i)=TT(i)*b(i,O(1));
        ar(1,i)=0;
end
% 2 recursion
for t=2:T,
        for j=1:N,
                for i=1:N
                d(i)=dt(t-1,i)*a(i,j);
                end
                d=d';
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                [dm,ar(t,j)]=max(d);
                dt(t,j)=dm*b(j,O(t));
            end
        end
        % 3 termination
        dt=dt';
        [p_st(c),q_st(T)]=max(dt(:,T));
        % 4 path
        for t=T-1:-1:1
            q_st(t)=ar(t+1,q_st(t+1));
        end
        % แสดง p,path
        q_st;
        p_st(c);
        clear a b
        end
[p,num]=max(p_st);
% แสดงตัวเลขที่ recognize ได้
num=num-1
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรม TEST

```
clear;
% โหลดไฟล์ .wav ที่จะใช้ในการวิเคราะห์ เข้ามาเก็บไว้ในตัวแปรชื่อ inpt
[inpt,fp,form]=loadwave('c:\input2\test.wav');
% เซตค่าต่าง ๆ ดังนี้
lpc=[];
cep=[];
p=12;           % จำนวนสัมประสิทธิ์ของวงจรกรอง
frame=300;      % ความยาวของเฟรม
shift=100;      % ระยะในการเลื่อนเฟรม
q=18;           % จำนวนสัมประสิทธิ์ของเซปสตรัม
% นำ input ชื่อ inpt มาผ่านการเอมฟาซิส แล้วเก็บไว้ในตัวแปร inlpc
inlpc=filter([1,-0.9375],[1],inpt);
% การคำนวนพารามิเตอร์ต่างๆ ในการประมาณ เก็บไว้ในตัวแปร a(j,i) และ g(i)
for i=1:(length(inpt)-frame+shift)/shift;
        format long;
        % คำนวน input ที่จะใช้ในการคำนวน เมื่อกำหนดขนาด(frame) และจุดเริ่มต้นของข้อมูล
        for m=1:frame;
                z(m)=inlpc(m+(shift*(i-1)));
        end
        % นำสัญญาณที่ได้มาผ่านการวินโดว์เพื่อเตรียมวิเคราะห์ต่อไป
        for m=1:frame;
                [x(m)]=z(m)*(0.54-0.46*cos(2*pi*(m-1)/(frame-1)));
        end
        % หาออโตคอร์รีเลชันและเมตริกซ์ของออโตคอร์รีเลชัน
        r=xcorr(x);
        for m=1:p,
                for j=m:p,
                        s(m,j)=r(frame+(j-m));
                        s(j,m)=s(m,j);
                end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end
for m=1:p,
        phi(m)=r(frame+m);
end
a=inv(s)*phi';
b=[1,-a'];
c=roots([1,-a']);
for m=1:p,
        if abs(c(m))>1
                b=deconv(b,[1,-c(m)]);
                b=conv(b,[1,-1/c(m)]);
        end
end
% คำนวนสัมประสิทธิ์ของวงจรกรอง(lpc)
for m=1:p+1,
        if m==1
                a(m)=0;
        else
                a(m)=-b((m-1)+1);
        end
end
a(14)=0;  a(15)=0;  a(16)=0;  a(17)=0;  a(18)=0;  a(19)=0;
lpc(:,i)=a;
% คำนวณอัตราการขยาย G
e=r(frame);
for m=1:p+2,
        e=e-a(m)*r(frame+(m-1));
end
g(i)=sqrt(e);
% คำนวนสัมประสิทธิ์เซปสตรัม(cep)
for m=0:1,
        if m==0
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                v(m+1) = log (g(i));
            else
                v(m+1) = lpc((m+1),i);
            end
        end
        for m=2:q,
            if m<=p
                v(m+1)=lpc((m+1),i);
            else
                v(m+1)=0;
            end
            for k=1:m-1,
                v(m+1)=v(m+1)+(k/m)*v(k+1)*lpc(((m+1)-k),i);
            end
        end
        [cep(:,i)]=v';
        % การเวทค่าสัมประสิทธิ์(y)
        q=18;
        for m=1:q+1,
            y(m)=cep(m,i)*(1+(q/2)*sin((pi*(m-1))/q));
        end
        [w1(:,i)]=y';
end
disp(' input ');
fid = fopen('c:\matlab\bin\code1.m','rb');
code=fread(fid,[19,64],'float64');
disp('  output  ');
status=fclose(fid);
c=[ ];
for j=1:length(w1)
        for m=1:64
        total=0;d=0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                for k=1:19
                    d=( w1(k,j) - code(k,m) )^2 ;
                    total=total+d;
                end
        c(m,j)=total;
        end
end
        [s,n]=min(c);
disp('s=')
disp(s)
disp('n=')
disp(n)
O=n;
T=length(n);
save observ.mat O T
clear;
format long e;
N=6;            % จำนวนสเตท
dt=[];          % จำนวนเดลต้า
K=64;           % จำนวนค่า VQ ที่ปรากฏในแต่ละสเตท
P=1;
Pmax=0;
x=zeros(6,6);
y=zeros(6,K);
z=zeros(1,6);
A_prime=zeros(6,6);
B_prime=zeros(6,K);
C_prime=zeros(1,6);
load observ.mat
for c=1:10%ค่า a b เริ่มต้น
        if c==1
                load arhmm0.mat
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end
if c==2
        load arhmm1.mat
end
if c==3
        load arhmm2.mat
end
if c==4
        load arhmm3.mat
end
if c==5
        load arhmm4.mat
end
if c==6
        load arhmm5.mat
end
 if c==7
        load arhmm6.mat
end
if c==8
        load arhmm7.mat
end
if c==9
        load arhmm8.mat
end
if c==10
        load arhmm9.mat
end
TT=[1 0 0 0 0 0];          %กำหนดค่าความน่าจะเป็นในการเริ่มต้นที่สเตทต่างๆ
%Viterbi  Algorithm
%1  initialization
for i=1:N,
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        dt(1,i)=TT(i)*b(i,O(1));
        ar(1,i)=0;
end
%2  recursion
for t=2:T,
        for j=1:N,
                for i=1:N
                        d(i)=dt(t-1,i)*a(i,j);
                        end
                d=d';
                [dm,ar(t,j)]=max(d);
                dt(t,j)=dm*b(j,O(t));
          end
end
%3  terminator
dt=dt';
[p_st(c),q_st(T)]=max(dt(:,T));
%4  แสดง p,path
for t=T-1:-1:1
        q_st(t)=ar(t+1,q_st(t+1));
end
%แสดง p,path
 q_st;
 p_st(c);
clear a b
end
[p,num]=max(p_st);
% แสดงตัวเลขที่ recognize ได้
num=num-1
clear O T
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้ สำเร็จลงได้ ด้วยความช่วยเหลืออย่างดียิ่งจากหลาย ๆ ฝ่าย ซึ่งคณะผู้จัดทำใคร่ขอขอบคุณทุกท่าน ที่มีส่วนร่วมสนับสนุน ช่วยเหลือ และแนะนำ ในทุก ๆ ด้าน

คณะผู้จัดทำขอขอบพระคุณ ดร. ไกรสิน ส่งวัฒนา อาจารย์ที่ปรึกษาปริญญานิพนธ์ ซึ่งได้กรุณาให้การสนับสนุนด้านอุปกรณ์ ให้คำปรึกษา แนะนำ และ ช่วยสรรหาแหล่งความรู้เพิ่มเติมช่วยให้การทำปริญญานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี

คณะผู้จัดทำได้รับความกรุณาอย่างยิ่งจาก อาจารย์ วิภา แสงพิสิทธิ์ ,อาจารย์ นภัทร สระเอี่ยม และ อาจารย์ อภิชาติ ตั้งทางธรรม อาจารย์จากมหาวิทยาลัยธรรมศาสตร์ ซึ่งได้เสียสละเวลา ช่วยให้คำปรึกษา และข้อเสนอแนะที่เป็นประโยชน์ ตลอดจนห้องทดลองและการทำงาน คณะผู้จัดทำขอขอบพระคุณอย่างสูงไว้ ณ ที่นี้

ขอขอบคุณ พี่สุนทร อรอินทร์ ที่คอยตอบคำถามที่สงสัยมาตลอด พี่สุพจน์ อินทร์ชัยยะ และพี่เสาวลักษณ์ อารีย์พงศา ที่เอื้อเฟื้อหนังสือ และเอกสาร และเพื่อน ๆ ทุกคนที่ให้ความรู้ และสนับสนุนการทำปริญญานิพนธ์เล่มนี้

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดา ของคณะผู้จัดทำ ที่ให้การสนับสนุนและคอยให้กำลังใจ จนกระทั่งปริญญานิพนธ์นี้สำเร็จลงด้วยดี

คณะผู้จัดทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] สุนทร อรอินทร์ และ อัฐ เครือฟัก ,**การประมวลเสียงพูดโดยการประมาณเชิงเส้น** ,ปริญญานิพนธ์ : สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, พ.ศ. 2537.

[2] สุพจน์ อินทร์ชัยยะ และคณะ , **การรู้จำเสียงพูด**,ปริญญานิพนธ์ : สถาบันพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , พ.ศ. 2536.

[3] เสาวลักษณ์ อารีย์พงศา ,**การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธี ฮิดเดน มาร์คอฟโมเดล และเวคเตอร์ควอนไตซ์เซชัน** , วิทยานิพนธ์ :จุฬาลงกรณ์มหาวิทยาลัย ,1995.

[4] ธันวา ศรีประโมง , **การเขียนโปรแกรมภาษาซีสำหรับวิศวกรรม** , กรุงเทพ : มหาวิทยาลัยมหานคร, พ.ศ. 2537.

[5] A.Stolx, **The Sound Blaster Book** , Michigan : Abacus,1993.

[6] A.V. Oppenheim and R.W. Schafer, **Digital Signal Processing,**New York : John Wiley & Sons,1976.

[7] D. Fowley and M. Horton , **The Student Edition of MATLAB** , New Jersey : Prentice - Hall , 1995.

[8] J.G. Proakis and D.G. Manoladis, :**Digital Signal ProcessingPrincipkes,Algorithm,and Application**,New York :Macmillam,1988.

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

[10] L.R. Rabiner and R.W. Schafer,**Diginal Processing of Speech Signals**,New Jersey : Prentice-Hall,1978.

[11] P.E. Papamichalis,**Practical Approches to Speeh Coding**, New Jersey : Prentice-Hall,1987.

[12] R.J. Schalkoff ,**Pattern Recognition** : Singapore ,1992.

[13] T.W. Parsons,**Voice and Speech Processing**,U.S.A. : McGraw-Hill ,1987

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้