สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การจดจำเสียงพูด

SPEECH RECOGNITION

โดย

นายประณิต อ่อนไสว

นายพรชัย ธรรมานุรัตพันธุ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2542

เลขหมู่.....
เลขทะเบียน 37143
วัน, เดือน, ปี - 4 ก.ย. 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจดจำเสียงพูด

SPEECH RECOGNITION

โดย

นายประณิต อ่อนไสว 39014295

นายพรชัย ธรรมานุรัตพันธุ์ 39014347

อาจารย์ที่ปรึกษา

อ. กฤษณ์ วงรุจิระ

ดร. สุทธิชัย นพนาคีพงษ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมโทรคมนาคม

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2542

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ ปีการศึกษา 2542

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง **การจดจำเสียงพูด**

SPEECH RECOGNITION

ผู้จัดทำ

1. นายประณิต อ่อนไสว **39014295**

2. นายพรชัย ธรรมานุรัตพันธุ์ **39014347**

อาจารย์ที่ปรึกษา

(อ.กฤษณ์ วงรุจิระ)

อาจารย์ที่ปรึกษา

(ดร.สุทธิชัย นพนาคีพงษ์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจดจำเสียงพูด

SPEECH RECOGNITION

โดย นายประณิต อ่อนไสว 39014295
นายพรชัย ธรรมานุรัตพันธุ์ 39014347

อาจารย์ที่ปรึกษา อ.กฤษณ์ วงรุจิระ
ดร.สุทธิชัย นพนาคีพงษ์

บทคัดย่อ

ปริญญานิพนธ์นี้เป็นการจดจำเสียงพูด โดยทำการวิเคราะห์สัญญาณเสียง ซึ่งขั้นตอนแรกต้องทำการหาขอบเขตของคำและเลือกมาเฉพาะส่วนที่พิจารณา หลังจากนั้นจะทำการหาค่าพารามิเตอร์ของสัญญาณ นำไปเปรียบเทียบความคล้ายคลึงกับแบบอ้างอิงที่ทำการเก็บพารามิเตอร์ไว้ก่อนแล้ว และผ่านขั้นตอนการตัดสินใจ เพื่อจำแนกสัญญาณเสียงเป็นขั้นตอนสุดท้าย

ABSTRACT

This thesis presents the "Speech Recognition" by analyze speech signal . First , we find bound of word , and select section that we consider . After that , finding parameters of speech signal will be used . Test pattern and reference pattern is compared at pattern similarity , and pass decision rule . Finally , selecting word is shown .

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทที่ 1 บทนำ	1
บทที่ 2 ทฤษฎีหรือหลักการ	2
2.1 ลักษณะของเสียงพูด	2
2.2 การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้นๆ	4
2.3 การหาค่าพลังงานของสัญญาณเสียง	5
2.4 การหาค่าอัตราการตัดศูนย์	6
2.5 การประมาณเชิงเส้น	6
2.6 เวคเตอร์ควอนไตเซชั่น	10
2.7 ฮิดเดนมาร์คอฟโมเดล	13
2.8 กฎการตัดสินใจ	17
บทที่ 3 การคำนวณและการสร้าง	18
3.1 การหาขอบเขตของคำ	18
3.2 การนอร์มอลไลซ์	24
3.3 การหาค่าพารามิเตอร์	25
3.4 การสร้างโค้ดบุ๊ค	31
3.5 เวคเตอร์ควอนไตเซชั่น	31
3.6 รูปแบบอ้างอิง	34
3.7 ฮิดเดนมาร์คอฟโมเดล	34
3.8 ภาคการตัดสินใจ	34
บทที่ 4 การทดลองและผลการทดลอง	37
บทที่ 5 บทวิจารณ์และบทสรุป	68
ภาคผนวก	
กิตติกรรมประกาศ	
หนังสืออ้างอิง	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

	หน้า
รูปที่ 2.1 แสดงบล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย	7
รูปที่ 2.2 แสดงขั้นตอนการทำงานของเวคเตอร์ควอนไตเซชั่น	11
รูปที่ 3.1 แสดงบล็อกไดอะแกรมการจำแนกเสียงพูดโดยทั่วไป	18
รูปที่ 3.2 แสดงบล็อกไดอะแกรมการจำแนกเสียงพูด	19
รูปที่ 3.3 แสดงผังงานการหาค่าพลังงานของสัญญาณเสียง	20
รูปที่ 3.4 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียงโดยการเปรียบเทียบพลังงาน	21
รูปที่ 3.5 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียงโดยการหาอัตราการตัดศูนย์	23
รูปที่ 3.6 แสดงขั้นตอนการนอร์มอลไลซ์สัญญาณ	24
รูปที่ 3.7 แสดงการแบ่งช่วงสัญญาณที่ใช้ในการวิเคราะห์	26
รูปที่ 3.8 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์	27
รูปที่ 3.9 แสดงผังงานขั้นตอนการเตรียมสัญญาณในการวิเคราะห์	30
รูปที่ 3.10 แสดงขั้นตอนการสร้างโค้ดบุ๊ค	32
รูปที่ 3.11 แสดงขั้นตอนการควอนไตเซชั่น	33
รูปที่ 3.12 แสดงขั้นตอนการสร้างแบบจำลองอ้างอิง	35
รูปที่ 3.13 แสดงขั้นตอนการรู้จำเสียงพูด	36
รูปที่ 4.1 แสดงขั้นตอนการหาโมเดลเสียง 0	38
รูปที่ 4.2 แสดงขั้นตอนการหาโมเดลเสียง 0	39
รูปที่ 4.3 แสดงขั้นตอนการหาโมเดลเสียง 1	40
รูปที่ 4.4 แสดงขั้นตอนการหาโมเดลเสียง 1	41
รูปที่ 4.5 แสดงขั้นตอนการหาโมเดลเสียง 2	42
รูปที่ 4.6 แสดงขั้นตอนการหาโมเดลเสียง 2	43
รูปที่ 4.7 แสดงขั้นตอนการหาโมเดลเสียง 3	44
รูปที่ 4.8 แสดงขั้นตอนการหาโมเดลเสียง 3	45
รูปที่ 4.9 แสดงขั้นตอนการหาโมเดลเสียง 4	46
รูปที่ 4.10 แสดงขั้นตอนการหาโมเดลเสียง 4	47
รูปที่ 4.11 แสดงขั้นตอนการหาโมเดลเสียง 5	48
รูปที่ 4.12 แสดงขั้นตอนการหาโมเดลเสียง 5	49
รูปที่ 4.13 แสดงขั้นตอนการหาโมเดลเสียง 6	50
รูปที่ 4.14 แสดงขั้นตอนการหาโมเดลเสียง 6	51
รูปที่ 4.15 แสดงขั้นตอนการหาโมเดลเสียง 7	52
รูปที่ 4.16 แสดงขั้นตอนการหาโมเดลเสียง 7	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ (ต่อ)

	หน้า
รูปที่ 4.17 แสดงขั้นตอนการหาโมเดลเสียง 8	54
รูปที่ 4.18 แสดงขั้นตอนการหาโมเดลเสียง 8	55
รูปที่ 4.19 แสดงขั้นตอนการหาโมเดลเสียง 9	56
รูปที่ 4.20 แสดงขั้นตอนการหาโมเดลเสียง 9	57
รูปที่ 4.21 แสดงรูปหน้าต่างหลักของโปรแกรมการจดจำเสียงพูดเมื่อใช้ในการทดสอบ	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่ 4.1 แสดงการทดสอบกับเสียงที่มีอยู่ในฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 9 คน พูดคนละ 5ครั้งเป็นฐานข้อมูล 59

ตารางที่ 4.2 แสดงการทดสอบกับเสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้งโดยใช้โมเดลของเสียงคน 9 คนพูดคนละ 5ครั้งเป็นฐานข้อมูล 60

ตารางที่ 4.3 แสดงการทดสอบกับเสียงที่มีอยู่นอกฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 9 คน พูดคนละ 5ครั้งเป็นฐานข้อมูล 61

ตารางที่ 4.4 แสดงการทดสอบกับเสียงที่มีอยู่ในฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 14 คน พูดคนละ 5ครั้งเป็นฐานข้อมูล 62

ตารางที่ 4.5 แสดงการทดสอบกับเสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้งโดยใช้โมเดลของเสียงคน 14 คนพูดคนละ 5ครั้งเป็นฐานข้อมูล 63

ตารางที่ 4.6 แสดงการทดสอบกับเสียงที่มีอยู่นอกฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 14 คน พูดคนละ 5ครั้งเป็นฐานข้อมูล 64

ตารางที่ 4.7 แสดงเปอร์เซ็นต์ถูกต้องในการตรวจสอบเสียงโดยใช้เสียงตรวจสอบที่อยู่ในฐานข้อมูล 10 ครั้งเสียงที่อยู่นอกฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 9 คนพูดคนละ 5ครั้งเป็นฐานข้อมูล 65

ตารางที่ 4.8 แสดงเปอร์เซ็นต์ถูกต้องในการตรวจสอบเสียงโดยใช้เสียงตรวจสอบที่อยู่ในฐานข้อมูล 10 ครั้งเสียงที่อยู่นอกฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงคน 14 คนพูดคนละ 5ครั้งเป็นฐานข้อมูล 65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

ในปัจจุบันการใช้คอมพิวเตอร์ได้พัฒนาไปไกลสามารถใช้งานได้สะดวกมากขึ้นไม่จำเป็นต้องจดจำคำสั่งต่างๆมากมาย โดยมีระบบการสั่งงานแบบรูปภาพแทนการใช้การสั่งงานแบบบรรทัดผู้ใช้เพียงแต่ใช้ตัวชี้ๆไปยังสัญญาณหรือรูปภาพ แทนการสั่งงานของคำสั่งต่างๆทำให้ลดความซับซ้อนและความยุ่งยากในการทำงาน

นอกจากระบบการติดต่อกับผู้ใช้โดยใช้รูปภาพแล้ว การสั่งงานด้วยเสียงจะคงเป็นอีกระบบหนึ่งที่จะเป็นมาตราฐานในการใช้งานคอมพิวเตอร์อีกระบบหนึ่งในระยะเวลาอันใกล้นี้ และเนื่องจากการ์ดเสียงในคอมพิวเตอร์แบบพีซีมีราคาถูกลงอย่างมากทำให้เหมาะที่จะใช้พัฒนาระบบการจำแนกเสียงพูดในคอมพิวเตอร์แบบพีซี เพื่อใช้เป็นพื้นฐานของการศึกษาและพัฒนาระบบสั่งงานด้วยเสียงพูดต่อไป

สำหรับปริญญานิพนธ์นี้ จะทำการศึกษาและเขียนโปรแกรมการจำแนกเสียง 0-9 โดยเสียงมนุษย์ที่พูดกันในการติดต่อสื่อสารนี้จะมีความถี่อยู่ในช่วง 300-3400 เฮิรตซ์ ซึ่งในการวิเคราะห์เสียงในแบบอนาล็อกนี้จะทำได้ยากจึงเปลี่ยนเสียงพูดให้เป็นสัญญาณดิจิตอลแทนโดยใช้ (Sound Card) เข้าช่วย สัญญาณเสียงพูดที่เป็นอนาล็อกนี้จะถูกบันทึกผ่านไมโครโฟน จากนั้นการ์ดเสียงจะเป็นตัวจัดการกับเสียงพูดให้อยู่ในรูปแบบที่เหมาะสม (ในที่นี้จะอยู่ในรูปของ ชื่อไฟล์.wav) ซึ่งรูปแบบนี้จะถูกกำหนดโดยโปรแกรมการบันทึกเสียงกับการ์ดเสียง ที่เป็นข้อมูลแบบ 8 บิต จะใช้ความถี่ในการสุ่ม 8 กิโลเฮิรตซ์ ซึ่งเป็นไปตามกฎของไนควิสต์ ที่ระบุไว้ว่าในการสุ่มจะต้องมีความถี่ในการสุ่มอย่างน้อย 2 เท่าของความถี่ของสัญญาณเพื่อสามารถเก็บข้อมูลทั้งหมดโดยไม่ผิดเพี้ยนโดยเมื่อนำสัญญาณเสียงที่เป็นสัญญาณดิจิตอลเข้ามาแล้วจะทำการตรวจสอบโดยนำไปหาค่าพารามิเตอร์แล้วเทียบกับพารามิเตอร์อ้างอิงที่หามาได้ก่อนแล้วโดยการนำไปเทียบก็จะสามารถแยกได้ว่าเสียงที่เข้าไปเป็นเสียงใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีหรือหลักการ

2.1 ลักษณะของเสียงพูด

คนเราเปล่งเสียงพูดด้วยอวัยวะที่ใช้ในการออกเสียง (organs of speech) ทำเสียงตามที่มีในระบบภาษาของตน แม้ว่าคนที่อยู่ในสังคมเดียวกันจะใช้ภาษาเดียวกันแต่ถ้าพิจารณาเสียงที่เปล่งออกมาจริง ๆ แล้ว แต่ละครั้งก็อาจจะสังเกตุลักษณะที่แตกต่างกันได้ เราจึงสามารถจำเสียง จำวิธีพูดของคนที่เราคุ้นเคยได้ เสียงพูดที่จะอธิบายด้วยหลักเกณฑ์ทางวิทยาศาสตร์แม้ว่าในภาษาหนึ่ง ๆ จะมีเสียงต่างกันมากบ้างน้อยบ้าง แต่ละเสียงก็สามารถนำมาพิจารณาและอธิบายให้รู้ลักษณะการออกเสียงและตำแหน่งที่เกิดเสียงได้คำอธิบายนี้จะทำให้เข้าใจลักษณะเสียงทุกเสียง วิชาที่ว่าด้วยเสียงพูดเรียกว่า วิชาสัทศาสตร์ (Phonetics)

ในการศึกษาเสียงพูดแบ่งได้เป็น 2 ลักษณะคือ

ก. สรีรสัทศาสตร์ (Articulatory) เป็นการศึกษาเสียงพูดจากอวัยวะและการเคลื่อนไหวอวัยวะที่ทำให้เกิดเสียงพูด การอธิบายนี้ก็จะอธิบายโดยอาศัยลักษณะและอาการเคลื่อนไหวของอวัยวะที่เกี่ยวข้องในการเปล่งเสียงพูดนั้น

ข. กลสัทศาสตร์ (Acoustic Phonetics) เป็นการศึกษาเสียงพูดจากลักษณะคลื่นเสียงที่ผู้พูดเปล่งออกมาแล้ว และผู้ฟังได้ยินว่ามีลักษณะทางกลศาสตร์อย่างไร การศึกษาตามแนวนี้ต้องอาศัยความรู้ทางฟิสิกส์และคณิตศาสตร์ช่วยอธิบายลักษณะของคลื่นเสียง

2.1.1 อวัยวะที่ใช้ในการออกเสียง

อวัยวะที่ใช้ในการออกเสียงมีอยู่หลายส่วน แต่ละส่วนสามารถทำให้เสียงพูดแตกต่างกันไปได้ อวัยวะเหล่านี้มีปากและส่วนต่าง ๆ ในปาก ช่องคอ กล่องเสียง ช่องว่างในปากและช่องว่างในจมูก อวัยวะที่ใช้ในการเปล่งเสียงพูดแบ่งออกเป็น 2 ประเภท คือ

ก. อวัยวะที่ใช้ในการทำอาการ (Articulator) คืออวัยวะที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่าง ๆ อวัยวะที่สำคัญ คือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

ข. อวัยวะซึ่งเป็นตำแหน่งที่เกิดเสียงต่าง ๆ (Point Of Articulator) คือตำแหน่งที่เกิดเสียงต่าง ๆ เช่น ริมฝีปาก ฟัน เพดาน ส่วนต่าง ๆ เป็นต้น

อวัยวะส่วนที่มีหน้าที่ในการออกเสียงโดยตรงมีดังนี้

2.1.1.1 ริมฝีปาก เป็นอวัยวะส่วนที่เคลื่อนไหวได้มากและทำให้เสียงแตกต่างกันได้มากเราอาจบังคับให้ริมฝีปากอยู่ชิดกัน ห่างกัน ยื่นออกมา เป็นต้น ก็ได้ ลักษณะริมฝีปากแบบต่าง ๆ นี้ล้วนมีอิทธิพลต่อการออกเสียงและทำให้เสียงแตกต่างกันไปทั้งสิ้น

2.1.1.2 ฟัน เป็นอวัยวะที่ทำให้เกิดเสียงหลายชนิด เช่นเมื่อฟันกดลงบนริมฝีปากล่างลมที่ผ่านออกมาโดยแรงจะลอดช่องออกมาซึ่งทำให้เกิดเสียงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1.3 ปุ่มเหงือก เป็นส่วนนูนออกมาอยู่หลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่าเป็นคลื่น ปุ่มเหงือกเป็นบริเวณที่ทำให้เกิดเสียงปุ่มเหงือก

2.1.1.4 เพดานแข็งหรือเพดานอ่อน คือส่วนที่เป็นกระดูกแข็ง

2.1.1.5 เพดานอ่อน คือส่วนเพดานที่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เวลาหายใจ เพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนลดระดับลงมา เปิดช่องว่างให้ลมออกทางจมูกเวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดหลังคอนอกจากเวลาออกเสียงนาสิกเท่านั้น

2.1.1.6 ลิ้นไก่ เป็นก้อนเนื้อเล็ก ๆ อยู่ปลายเพดานอ่อนตรงกลางปาก ลิ้นรั่วได้

2.1.1.7 ลิ้น ลิ้นเป็นส่วนที่เคลื่อนที่มากที่สุดในการออกเสียง ซึ่งสามารถแบ่งได้เป็น 3 ส่วน

ปลายลิ้น คือส่วนของปลายลิ้นซึ่งสามารถยกขึ้นไปแตะกับอวัยวะส่วนต่าง ๆ ได้

หน้าลิ้น คือส่วนที่อยู่ตรงข้ามกับเพดานแข็งถ้าวางลิ้นราบกับปากเช่นเดียวกับเวลาที่ไม่ได้พูด

หลังลิ้น ถ้าวางลิ้นราบกับปาก ลิ้นส่วนนี้จะอยู่ตรงข้ามกับเพดานอ่อน

2.1.1.8 แผ่นเนื้อปากหลอดลม เป็นก้อนเนื้อเล็ก ๆ คล้ายลิ้นไก่ อยู่ตรงโคนลิ้นลงไปในคอ มีหน้าที่ปิดช่องลมในขณะรับประทานอาหารและเปิดช่องลมเมื่อพูด

2.1.1.9 ช่องคอ อยู่ถัดจากช่องปากไปจนถึงเส้นเสียง

2.1.1.10 เส้นเสียง เป็นอวัยวะสำคัญที่ทำให้เกิดเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ แผ่นภายในกล่องเสียง ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างซึ่งเป็นช่องผ่านให้ลมไปถึงปอดและออกมาจากปอดได้ ช่องนี้เรียกว่าช่องว่างระหว่างเส้นเสียง

2.1.1.11 ช่องจมูก หมายถึงโพรงในช่องจมูกซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะป่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก

เสียงที่เกิดขึ้นนั้นไม่ว่าจะเป็นเสียงประเภทใด จะผ่านไปตามทางเดินของเสียง โดยจะเริ่มตั้งแต่ช่องว่างระหว่างเส้นเสียงถึงริมฝีปาก ในทางวิทยาศาสตร์สรุปได้ว่าทางเดินของเสียงคือ ท่อนำเสียงที่มีรูปร่างไม่แน่นอน

2.1.2 การเกิดของเสียง (Speech Production)

การเกิดของเสียงแบ่งออกเป็น 3 ขั้นตอนคือ

ขั้นตอนที่ 1 จุดเริ่มต้น เป็นขั้นตอนที่ลมเริ่มถูกขับออกจากปอด ผ่านเข้าไปสู่ขั้นตอนที่ 2

ขั้นตอนที่ 2 การดัดแปลงลมที่เส้นเสียง อวัยวะที่ใช้ในขั้นตอนนี้คือส่วนที่ต่อจากปอดขึ้นมาถึงกล่องเสียงและที่กล่องเสียงเส้นเสียงจะทำหน้าที่เป็นลิ้นปิดเปิดทำให้เกิดเสียง 2 ชนิดคือ

1. เสียงก้อง (Voiced)เกิดจากกเส้นเสียงปิดกั้นลมไว้ ลมที่ผ่านออกมาจะเพิ่มแรงดันมากขึ้นจนเส้นเสียงปิดเปิดสลับกันไป ทำให้เกิดเสียงก้องขึ้นมา ซึ่งสามารถเรียกความถี่ในการปิดเปิดของเส้นเสียงว่า ความถี่มูลฐาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เสียงไม่ก้อง (Unvoiced) เสียงชนิดนี้เส้นเสียงจะเปิดตลอดเวลาที่ลมผ่าน ลมจึงผ่านออกมาได้สะดวกทำให้เกิดเสียงไม่ก้องขึ้น

ขั้นตอนที่ 3 การเปลี่ยนแปลงลักษณะเส้นเสียง อวัยวะที่ใช้คือส่วนที่ต่อจากกล่องเสียงจนถึงริมฝีปาก โดยลมที่ผ่านออกจากกล่องเสียงจะทำให้เกิดเสียงในลักษณะต่าง ๆ ซึ่งจะเกิดจากการเปลี่ยนแปลงของอวัยวะที่กล่าวไว้ในหัวข้อที่แล้ว

2.2 การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้น ๆ

เนื่องจากว่าสัญญาณเสียงเป็นสัญญาณที่แปรตามเวลา มีการแปรเปลี่ยนที่ไม่แน่นอน เช่น ในขณะพูดช้า ๆ รูปร่างของโพรงเสียงรวมทั้งลักษณะรูปแบบของการกระตุ้นอาจจะไม่มีการเปลี่ยนแปลงในช่วงเวลาที่นานที่สุดประมาณ 200 มิลลิวินาที แต่ในขณะที่พูดอย่างรวดเร็ว อาจจะมีช่วงที่ไม่มีการเปลี่ยนแปลงสั้นมากคือประมาณ 80 มิลลิวินาที ก็ได้

ดังนั้นเทคนิคในการวิเคราะห์เสียงพูดส่วนใหญ่แล้ว จะสมมติให้สัญญาณเสียงมีคุณสมบัติที่เปลี่ยนแปลงสัมพันธ์กับเวลาอย่างเชื่องช้า นั่นก็คือเราจะต้องแบ่งทำการวิเคราะห์หาพารามิเตอร์ของสัญญาณเสียงพูดในช่วงเวลาสั้น ๆ เหมือนมองผ่านช่องแคบ ๆ ที่เรียกว่า ช็อตไทม์วินโดว์ (shot-time window) เมื่อเทียบตามเวลาที่เสียงจะมีการเปลี่ยนแปลงได้ เพื่อจะได้มองเห็นเหมือนกับว่า เราหาพารามิเตอร์นั้น ๆ ได้มาจากสัญญาณเสียงที่อยู่ภายในช่องแคบ ๆ และมีความเสถียรภายในช่วงเวลาสั้นๆ

เทคนิคส่วนใหญ่จะกำหนดให้พารามิเตอร์ได้มาจากค่าเฉลี่ยจากการวิเคราะห์ภายในช่องเวลาแคบ ๆ นั้น สำหรับกรณีที่ต้องพิจารณาพารามิเตอร์ต่าง ๆ ที่มีการเปลี่ยนแปลง ก็จะทำการแบ่งสัญญาณเสียงออกเป็นช่องหลาย ๆ ช่องหรืออาจจะเรียก กรอบการวิเคราะห์ (analysis frame) ดังนั้นพารามิเตอร์ต่าง ๆ จะสามารถหาได้ทันเพียงพอที่จะติดตามการเปลี่ยนแปลงของสัญญาณ สำหรับในช่วงที่สัญญาณมีการเปลี่ยนแปลงช้า อาจจะกำหนดให้ช่องแคบมีขนาดใหญ่ประมาณ 100 มิลลิวินาที แต่ในทางตรงข้ามถ้าสัญญาณมีการเปลี่ยนแปลงเร็ว ก็ต้องใช้ช่องแคบที่มีขนาดเล็กมาก ๆ ประมาณ 5-10 มิลลิวินาที เพื่อป้องกันการสูญหายของรายละเอียดของสัญญาณถัดไป

2.2.1 รูปแบบของช่องแคบ (Windows)

การกำหนดขนาดของช่องแคบที่ใช้ขึ้นอยู่กับ

2.2.1.1 ช่องจะต้องสั้นพอ ที่จะทำให้คุณสมบัติของเสียงที่กำลังพิจารณาไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญในช่องแคบนั้น

2.2.1.2ช่องแคบจะต้องยาวพอที่จะทำให้การจัดเตรียมตัวอย่างของเสียงเพื่อจะนำไปคำนวณหาพารามิเตอร์ให้ได้ตามต้องการอย่างเช่น ในกรณีที่มีสัญญาณรบกวนเข้ามาแทรกอยู่บางช่วงในสัญญาณเสียงด้วย ถ้าเราเลือกใช้ช่องแคบที่มีขนาดใหญ่กว่า เมื่อทำการหาค่าพารามิเตอร์โดยเฉลี่ย ก็จะทำให้ส่วนประกอบของสัญญาณรบกวนถูกตัดทิ้งหรือมองข้ามไป

2.2.1.3 ช่องแคบที่เหมาะสม ไม่ควรสั้นเกินกว่าช่วงหนึ่งคาบของสัญญาณเสียงในช่วงที่กำลังวิเคราะห์ เงื่อนไขนี้จะมีผลต่อค่าเฟรมเรท (frame rate ซึ่งก็คือจำนวนครั้งต่อวินาทีที่ทำการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณเสียง โดยการขยับช่องแคบไปเป็นคาบ ๆ ตามแกนเวลา) ตามปกติเฟรมเรทจะมีค่าประมาณ 2 เท่าของส่วนกลับของขนาดช่องแคบ นั่นก็คือช่องแคบถัด ๆ กันไปจะมีการซ้อนทับกัน 50 เปอร์เซ็นต์

การนำฟังก์ชันของช่องแคบที่มีช่วงขนาดจำกัด w(n) มาคูณเข้ากับสัญญาณ s(n) จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของช่องแคบ รูปแบบของช่องแคบที่ง่ายที่สุดก็คือ ช่องแคบสี่เหลี่ยม (rectangular window) ซึ่งมีนิยามดังนี้

$$w(n) = \begin{cases} 1, n = 0,1,...,N-1 \\ 0, n \quad other \end{cases} \tag{2.2.1}$$

ในสมการนี้ คือการกำหนดช่องของการวิเคราะห์ให้มีจำนวนตัวอย่าง N ตัวอย่าง

รูปแบบของช่องแคบในลักษณะนี้ ก็มีฟังก์ชันของช่องแคบหลายลักษณะด้วยกันตัวอย่างเช่น Blackman,Barlett,Hamming,Hanning เป็นต้น โดยฟังก์ชันของช่องแคบที่นิยมใช้กันมากในการวิเคราะห์สัญญาณเสียงก็คือ ฟังก์ชัน Hamming ซึ่งมีรูปร่างตามลักษณะของ (cosine pulse) มีนิยามดังนี้

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n)(N-1), n = 0,1,...,N-1 \\ 0, \qquad\qquad n \quad other \end{cases} \tag{2.2.2}$$

2.3 การหาค่าพลังงานของสัญญาณเสียง

พลังงานของสัญญาณเป็นตัวแทนอันหนึ่งที่เรามักจะนำมาใช้ในการวิเคราะห์ลักษณะต่างๆของสัญญาณทั่ว ๆ ไป โดยพลังงานของสัญญาณ s(n) ใด ๆ ที่แปรตามเวลาสามารถนิยามได้ว่า

$$E = \sum_{n=-\infty}^{\infty} S^2(n) \tag{2.3.1}$$

แต่สำหรับสัญญาณเสียงซึ่งเป็นสัญญาณที่แปรเปลี่ยนอยู่ตลอด ไม่มีเสถียรภาพตามเวลา เราจะต้องแบ่งสัญญาณออกมาพิจารณาเป็นช่วงเล็ก ๆ ตามแกนเวลาหรือเรียกว่าแบ่งออกเป็นเฟรม เช่น เฟรมละประมาณ 10 –30 วินาที หรือเฟรมละ 100ตัวอย่าง เป็นต้นดังนั้นก็จะสามารถหาพลังงานของเสียงในแต่ละเฟรมได้เป็น

$$E_l(m) = \sum_{n=0}^{N-1} s^2(n) \tag{2.3.2}$$

โดยที่ l แทนลำดับของเฟรมข้อมูลเสียง, l = 0,1,2,3,…,L

N แทนจำนวนข้อมูลเสียงในแต่ละเฟรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวัดค่าพลังงานดังในสมการที่ 2.3.2 นั้นมีข้อจำกัดตรงที่ว่ามันจะมีความไวต่อสัญญาณที่มีขนาดใหญ่ ๆ เนื่องจากเราใช้วิธียกกำลังสองค่าของสัญญาณอินพุท ดังนั้นการแก้ปัญหาอย่างหนึ่งก็คือวัดพลังงานของเสียงโดยใช้สมการดังนี้

$$E_1(m) = \sum_{n=0}^{N-1} |s(n)| \tag{2.3.3}$$

2.4 การหาค่าอัตราการตัดศูนย์ (Zero Crossing)

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

ในการหาค่าอัตราการตัดศูนย์สามารถทำได้โดยสมการดังนี้

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

Sign(s(n)) = 1 ; s(n)>0

0 ; at other

2.5 การประมาณเชิงเส้น (Linear Predictive Coding)

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

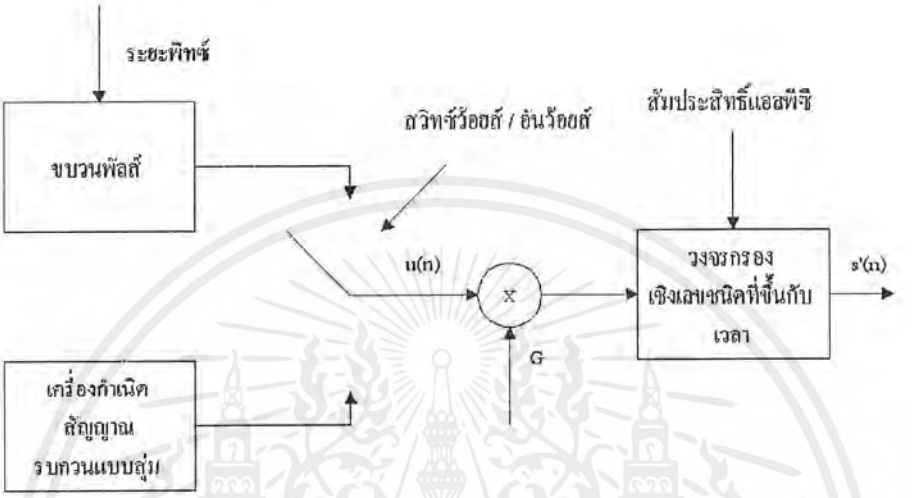
$$s'(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{2.5.1}$$

เมื่อ α_k เป็นค่าคงที่ เรียกวิธีการนี้ว่าการประมาณเชิงเส้นอันดับ p โดยมีเงื่อนไขว่า ค่า α_k ที่ใช้ในการประมาณจะต้องทำให้ ผลรวมของกำลังสองของความคลาดเคลื่อน $\{s(n)-s'(n)\}^2$ มีค่าน้อยที่สุดนั่นคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$\sum e^2(n) = \sum \{s(n) - s'(n)\}^2$ มีค่าต่ำสุด ซึ่งจะใช้การประมาณเชิงเส้นวิธีออโตคอร์รีเลชั่น (Autocorreiation Method) หรือวิธีอัตตสัมพันธ์

จากหลักการพื้นฐานของการประมาณเชิงเส้นและแบบจำลองระบบสร้างสัญญาณเสียงเราสามารถเขียน บล็อกไดอะแกรมการทำการประมาณเชิงเส้น มาสร้างสัญญาณเสียงพูดได้ดังรูปที่ 2.1



รูปที่ 2.1 แสดงบล็อกไดอะแกรมโมเดลการสร้างสัญญาณเสียงพูดอย่างง่าย

จากรูป 2.1 สามารถเขียนสมการได้เป็น

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

การประมาณเชิงเส้นโดยใช้สัมประสิทธิ์ $\{\alpha_k\}$ คือ

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

ดังนั้นความคลาดเคลื่อนคือ

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

โดยสมการ (2.5.8) – (2.5.9) จะเห็นได้ว่า

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

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

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

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

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

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

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

ดังนั้น จากสมการ (2.5.11) และ (2.5.12) จะได้ว่า

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

จากสมการ (2.5.12) จะได้ว่า

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการ 2.5.14 เขียนให้อยู่ในรูปเมตริกซ์ได้เป็น

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

หรือ $R_n.\alpha = r_n$

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

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

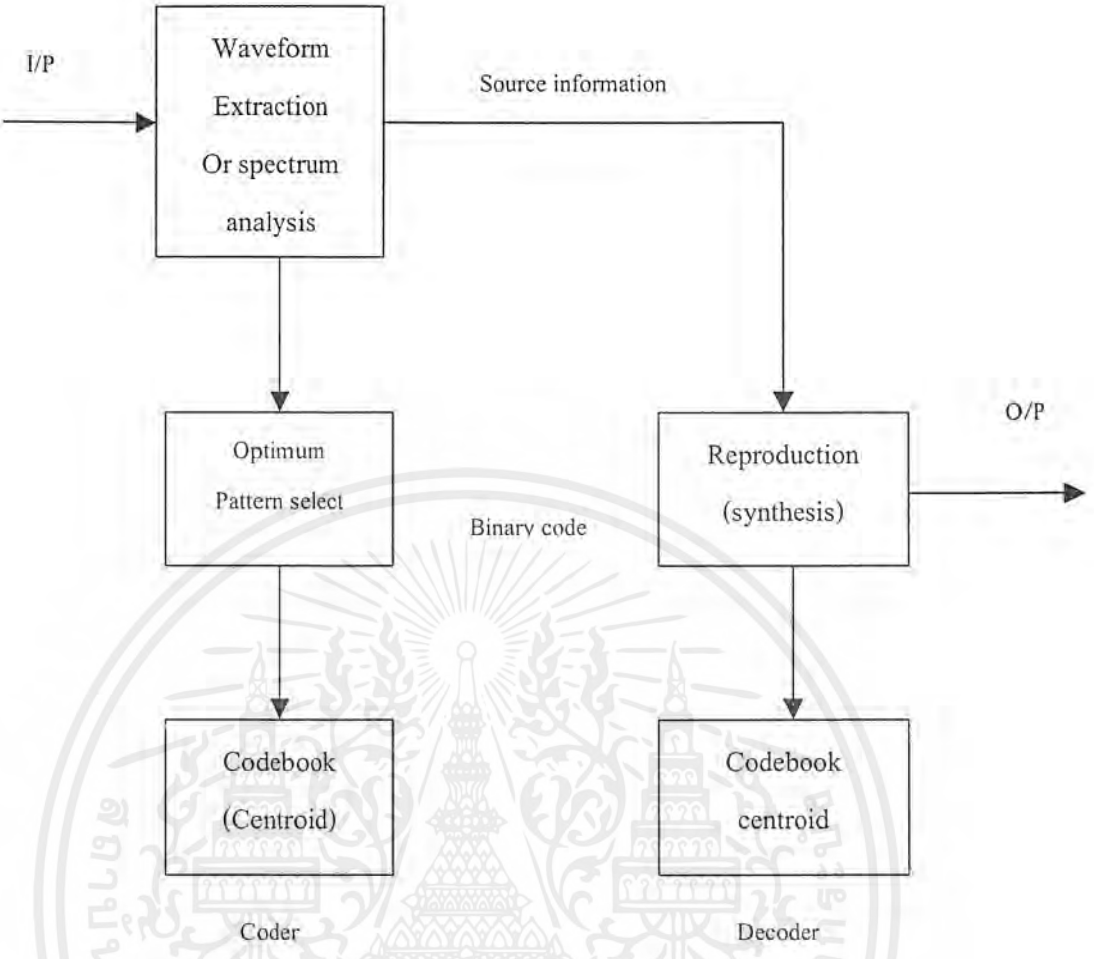
2.6 เวคเตอร์ควอนไตเซชั่น

เป็นวิธีการลดไดเมนชั่น(Dimension)ของข้อมูล โดยการหาค่าใดค่าหนึ่งซึ่งเป็นตัวแทนของข้อมูลจำนวนมากๆจำนวนหนึ่งได้ ปกติวิธีการนี้ใช้ในการบีบอัดข้อมูล แต่ในที่นี้เราใช้สำหรับเปลี่ยนข้อมูลที่ผ่านกระบวนการแอลพีซีแล้วให้เป็นข้อมูลที่ไม่ต่อเนื่อง พิจารณาได้ดังนี้

2.6.1หลักการเวคเตอร์ควอนไตเซชั่น

วิธีการเวคเตอร์ควอนไตเซชั่น เป็นวิธีการควอนไตซ์คลื่นเสียงเป็นเซต ซึ่งจะสามารถควอนไตซ์พารามิเตอร์ของแอลพีซีได้อย่างมีประสิทธภาพ มีบล็อคไดอะแกรมแสดงขั้นตอนการทำงานดังภาพต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 แสดงขั้นตอนการทำงานของเวคเตอร์ควอนไตเซชั่น

ในการทำเวคเตอร์ควอนไตเซชั่น มีวิธีการดังนี้คือ จะมีเวคเตอร์โค้ดที่เก็บไว้ อินพุทที่เข้ามาจะถูกทำการเปรียบเทียบกับโค้ดบุ้คที่มีอยู่ โดยจะพิจารณาว่าอินพุทที่เข้ามานั้นห่างจากโค้ดบุ้คใดน้อยที่สุดแล้วจะแทนอินพุทดังกล่าวด้วยเวคเตอร์โค้ดนั้น

โค้ดบุ้คที่ใช้จะเหมาะสมเพียงใด ขึ้นกับความคลาดเคลื่อนรวมทั้งหมดของต้นแบบที่ใช้หาเมื่อต้นแบบที่ใช้พิจารณามีจำนวนจำกัด โค้ดบุ้คที่มีความคลาดเคลื่อนน้อยที่สุด จะได้จากการทำซ้ำจนกว่าจะได้ค่าที่เหมาะสม โดยเหตุนี้จึงขึ้นกับแรนดอมเลินนิ่ง(random learning) และการเลือกค่าขึ้นมาจำนวนหนึ่งซึ่งเท่ากับขนาดของโค้ดบุ้คจากข้อมูลต้นแบบเพื่อใช้เป็นค่าเริ่มต้นในการคลัสเตอริ่ง(clustering) ซึ่งเป็นวิธีที่ใช้ในลอยด์อัลกอริทึม(Loyd 's algorithm) หรือเคมีนส์อัลกอริทึม (K-means algorithm) ในวิธีการนี้ข้อมูลต้นแบบที่ต้องการหาจะถูกแยกไปอยู่ในแต่ละกลุ่มของค่าเริ่มต้น แล้วจะทำการหาค่าเฉลี่ยของแต่ละกลุ่มเพื่อเป็นค่ากลางของกลุ่ม แล้วทำวิธีการดังกล่าวซ้ำจนกว่าความคลาดเคลื่อนรวมจะต่ำกว่าค่าๆหนึ่ง หรือการลดลงของความคลาดเคลื่อนรวมน้อยกว่าค่าๆหนึ่ง โดยค่ากลางดังกล่าวของกลุ่มจะถูกเก็บเป็นเวคเตอร์โค้ด ซึ่งแน่นอนว่าค่าความคลาดเคลื่อนรวมจะลดลงทุกครั้งที่มีการคำนวณซ้ำใหม่ ฉะนั้นจึงขึ้นกับค่าที่กำหนดว่าต้องการให้ค่าความคลาดเคลื่อนรวมน้อยเพียงใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2 การคำนวณเวคเตอร์ควอนไตเซชั่น

ถ้าสมมติ $x=[x_1,x_2,\ldots,x_n]^T$ มี N มิติ เราจะทำการหาควอนไตซ์ของเวคเตอร์ x กับโค้ดบุ้คซึ่งสามารถเขียนว่า y เป็นควอนไตซ์ของค่า x ได้ดังสมการ

$$y+q(x) \tag{2.6.1}$$

โดย q() เป็นโอเปอเรเตอร์ของควอนไตซ์ y เป็นเอาท์พุทเวคเตอร์ของค่า x โดย y เป็นค่าใดค่าหนึ่งใน $Y=\{y_i, 1 \le i \le L\}$ โดย $y_I=[y_{i1},y_{i2},\ldots,y_{in}]^T$ Y เป็นเซตของโค้ดบุ้ค L เป็นขนาดของโค้ดบุ้ค และ $\{y_i\}$ เป็นเซตของเวคเตอร์โค้ด y_i อาจเรียกว่าเป็นโค้ดอ้างอิง และ L อาจเรียกว่าจำนวนระดับขั้น เมื่อเทียบกับสเกลาร์ควอนไตซ์ ทำการแบ่งเวคเตอร์ x ไปใน L เซล $\{C_i,1 \le i \le L\}$ เมื่อ x อยู่ในเซล C_i

$$q(x) = y_i \quad \text{เมื่อ} \quad x \in C_i \tag{2.6.2}$$

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

$$D = \lim_{m\to\infty} \frac{1}{M} \sum_{n=1}^{M} d[x(n), y(n)] \tag{2.6.3}$$

การคำนวณหาโค้ดบุ้คมีหลายวิธี แต่ที่จะนำมาใช้คือ ลอยด์สอัลกอริทึม

2.6.3 การวัดความคลาดเคลื่อน

การวัดความคลาดเคลื่อนเป็นส่วนจำเป็นและเป็นเครื่องมือที่เป็นประโยชน์ต่อการออกแบบโค้ด ซึ่งมีอยู่หลายแบบแต่เราจะเลือกใช้วิธีสแควเออเรอร์ดิสทอร์ชั่น(Square error distortion) ซึ่ง
สำหรับสัญญาณที่เข้ามามี K มิติ เราสามารถหาระยะห่างระหว่างสัญญาณเข้ากับเวคเตอร์โค้ดได้ ดังสมการต่อไปนี้

$$d(x,y) = \|x-y\|^2 \tag{2.6.4}$$

$$= \sum_{i=0}^{k-1} (x_i - y_i)^2 \tag{2.6.5}$$

วิธีนี้เป็นวิธีแบบง่ายๆ ที่ใช้กันทั่วๆ ไป

2.6.4 การสุ่มค่าเริ่มต้น

วิธีหนึ่งที่ใช้ในการออกแบบโค้ดบุ้ค คือ การเลือกค่าเริ่มต้นของโค้ดบุ้คในการสุ่มตัวอย่างจากข้อมูลที่จะนำมาคำนวณ เราเรียกโค้ดบุ้คที่ได้จากการสุ่มค่าเริ่มต้นนี้ว่า แรนดอมโค้ดบุ้ค(random codebook) ถึงแม้ว่าการสุ่มค่าเริ่มต้น จะไม่ใช่วิธีที่ดีนักแต่โค้ดบุ้คที่ได้ก็ให้ผลเป็นที่ยอมรับได้

2.6.5 ขั้นตอนการทำงาน

ในการออกแบบโค้ดบุ้คสิ่งที่สำคัญที่สุดคือขั้นตอนการหาโค้ดบุ้คเคมีนส์อัลกอริทึม จะต้องมีการสุ่มค่าโค้ดบุ้คเริ่มต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การหาโค้ดบุ้คเคมีนส์อินิเซียลไลเซชั่น(K-means initialization) วิธีการเคมีนส์อัลกอริทึมไม่ยืนยันว่าคำตอบที่ได้จะมีค่าความคลาดเคลื่อนน้อยที่สุด จะต้องทำการสุ่มค่าโค้ดบุ้คเริ่มต้นจากหลายๆ ค่า ซึ่งโค้ดบุ้คที่มีค่าคลาดเคลื่อนน้อยที่สุดจะได้รับการเลือก
2) การเปรียบเทียบเมื่อมีโค้ดบุ้คเรียบร้อย เสียงที่จะนำไปหาโมเดล โดยวิธีฮิดเดนมาร์คอฟโมเดลจะต้องผ่านขั้นตอนนี้ เพื่อหาตัวแทนสัมประสิทธิ์การสะท้อนในรูปของควอนไตซ์

2.7 ฮิดเดนมาร์คอฟโมเดล (Hidden Markov Model)

แบบจำลองมาร์คอฟ เป็นแบบจำลอง (model) ทางสถิติซึ่งพัฒนาเพื่อแบ่งกลุ่มของอนุกรมทางเวลาหรือสัญญาณที่ไม่คงที่ นั่นคือ ใช้สำหรับการจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknow signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลองมาร์คอฟ ได้ถูกนำมาประยุกย์ใช้ในการรู้จำเสียงพูด

แบบจำลองมาร์คอฟแบ่งเป็น 2 ประเภทคือแบบต่อเนื่อง (Continuous) และแบบไม่ต่อเนื่อง (Discrete-time) ในที่นี้จะเลือกใช้แบบไม่ต่อเนื่องเพราะเป็นวิธีการที่ซับซ้อนน้อยกว่าและใช้ได้ดีกับคำพูดสั้นๆ

2.7.1 ส่วนประกอบของแบบจำลอง HMM

2.7.1.1 Nคือจำนวนสเตทในแบบจำลอง โดยสามารถย้ายจากสเตทหนึ่งไปอีกสเตทหนึ่งได้เราให้เซตของสเตทเป็น { 1,2,...,N } และสเตทที่เวลา t ใดๆเป็น q_t

2.7.1.2 M คือจำนวนของค่าปรากฏต่อ 1 สเตท แทนด้วยสัญลักษณ์

$$V = \{V_1, V_2, ..., V_M\}$$

2.7.1.3 $A = \{a_{ij}\}$ คือความน่าจะเป็นในการเปลี่ยนสเตทที่

$$a_{ij} = P\{q_t = j \mid q_{t-1} = i\} \text{ เมื่อ } 1 \le ij \le N$$

2.7.1.4 $B = \{b_j(k)\}$ คือความน่าจะเป็นของการเกิดค่าปรากฏ ที่

$$b_j(k) = P\{O_t = V_k \mid q_t = j\} \text{ เมื่อ } j = 1,2,...,N$$

2.7.1.5 π_i คือความน่าจะเป็นที่แต่ละสเตทจะเป็นสเตทเริ่มต้น เมื่อ

$$\pi_i = P(q_t ; t = 1)$$

2.7.2 โครงสร้างแบบจำลอง HMM

แบ่งตามลักษณะการเปลี่ยนสเตท (transition) ของเมตริกซ์ A

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2.1 แบบ Egordic Model หรือ Fully Connected Model แบบจำลองนี้ทุกสเตท สามารถเปลี่ยนสเตทไปสเตทอื่นๆได้ทุกสเตท

2.7.2.2 แบบ Left – Right Model หรือ Bakis Model แบบจำลองนี้การเปลี่ยนสเตทจะเปลี่ยนจากซ้ายไปขวา มีคุณสมบัติการเปลี่ยนสเตทดังนี้

2.7.2.2.1 $a_{ij} = 0, j < 1$ หมายความว่าเมื่อผ่านสเตทใดไปแล้วจะไม่มีการย้อนกลับมายังสเตทนั้นอีก

2.7.2.2.2 $\pi_i = \{0; i \neq 1 \ , 1; i = 1\}$ หมายความว่า ลำดับของสเตทต้องเริ่มต้นที่สเตทที่ 1 สเตทที่เหลือจึงมีความน่าจะเป็นที่จะเป็นสเตทเริ่มต้นเท่ากับศูนย์ I eft – Right Model นี้มีกฎข้อบังคับการเปลี่ยนสเตทดังนี้ $a_{ij} = 0$ เมื่อ $i > i + \Delta i$ โดยค่าของ $\Delta i = 2$ หมายความว่าการเปลี่ยนสเตทจะสามารถเปลี่ยนได้เกิน 2 สเตท จะได้เมตริกซ์การเปลี่ยนสเตทเป็น

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

จะเห็นว่าสเตทสุดท้ายมีสัมประสิทธิ์การเปลี่ยนสเตทเป็น

$a_{NN} = 1, a_{Ni} = 0$ เมื่อ $i < N$ แบบจำลองนี้จึงเหมาะกับสัญญาณที่มีการเปลี่ยนแปลงอย่างต่อเนื่องเช่น คำพูด

2.7.2.3แบบ Parallel Left – Right Model มีคุณสมบัติการเปลี่ยนสเตทคล้ายกับแบบที่ 2 แต่มีความยืดหยุ่นมากกว่า

ปัญหาของ HMM มี 3 ข้อซึ่งต้องใช้อัลกอริทึมวิธีต่างๆในการคำนวณเพื่อแก้ปัญหานี้

<u>ปัญหาที่ 1</u> เมื่อลำดับของค่าปรากฏ $O = \{O_1, O_2, ..., O_T\}$ และมีแบบจำลอง $\lambda = (A, B, \pi)$

เราจะคำนวณหาค่า $P(O \mid \lambda)$ ของลำดับของค่าปรากฏได้อย่างไร

<u>ปัญหาที่ 2</u> เมื่อมีลำดับของค่าปรากฏ $O = \{O_1, O_2, ..., O_T\}$ และมีแบบจำลอง $\lambda = (A, B, \pi)$

เราจะหาลำดับสเตท $q = \{q_1 q_2 ... q_T\}$ ที่เหมาะสมในการให้ค่าปรากฏนั้นได้อย่างไร

<u>ปัญหาที่ 3</u> จะหาแบบจำลอง $\lambda = (A, B, \pi)$ ที่ให้ค่า $P(O \mid \lambda)$ มากที่สุดได้อย่างไร

ลำดับของค่าปรากฏที่ใช้ปรับค่าพารามิเตอร์ A,B และ π เพื่อให้ได้แบบจำลองที่ดีที่สุดนั้นเรียกว่าลำดับเทรนนิ่ง (training sequence)

2.7.3 การคำนวณเพื่อแก้ปัญหาของ HMM

2.7.3.1 การแก้ปัญหาที่ 1 เป็นการคำนวณว่าแบบจำลองจะให้ความน่าจะเป็นที่จะได้ลำดับค่าปรากฏมากน้อยเพียงใด มีวิธีการเพื่อช่วยแก้ปัญหาโดยใช้กระบวนการต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.3.1.1 กระบวนการไปข้างหน้า (Forward Procedure)

เมื่อกำหนดให้ตัวแปรไปข้างหน้า(forward variable) $\alpha_t(i) = P(O_1, O_2, ..., O_T, q_t = i \mid \lambda)$ หมายถึงความน่าจะเป็นของการเกิดลำดับค่าปรากฏ $O_1, O_2, ..., O_T$ ที่จะอยู่ในที่สเตท i ณ เวลา t โดยมีแบบจำลองเป็น λ โดยเราจะคำนวณหา $\alpha_t(i)$ ได้ดังนี้

2.7.3.1.1.1 การเริ่มต้น (initialization)

เมื่อกำหนด $\alpha_t(i) = \pi_i b_i(O_t)$ ที่เวลาเริ่มต้น t=1 และเหตุการณ์เริ่มต้น O_1 เมื่อ $1 \le i \le N$

2.7.3.1.1.2 การเหนี่ยวนำ (induction)

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

หมายถึง ความน่าจะเป็นของสเตท j ที่เวลา t+1 ได้มาจากสเตท i ที่เป็นไปได้ถึง N สเตท ที่เวลา t

2.7.3.1.1.3 การสิ้นสุด (termination)

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$ ความน่าจะเป็นของลำดับค่าปรากฏ O ได้จากผลรวมของ $\alpha_T(i)$ จากทุกๆสเตทเมื่อ $1 \le j \le N$

2.7.3.1.2 กระบวนการย้อนกลับ (Backward Procedure)

เมื่อกำหนดให้ตัวแปรย้อนกลับ (backward variable)

$\beta_T(i) = P(o_{t+1} o_{t+2} ..... o_T, q_t = i \mid \lambda)$ หมายถึงความน่าจะเป็นของลำดับค่าปรากฏส่วนหลังจากเวลา t+1 ไปจนจบ โดยกำหนดว่าต้องอยู่ที่สเตท i ที่เวลา t และมีแบบจำลองเป็น λ เราจะคำนวณหา $\beta_T(i)$ ได้ดังนี้

2.7.3.1.2.1 การเริ่มต้น (initialization)

$$\beta_T(i) = 1; 1 \le j \le N$$

2.7.3.1.2.2 การเหนี่ยวนำ (induction)

$$\beta_T(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(i)$$; $1 \le t \le T-1$ และ $1 \le i \le N$

2.7.3.2 การแก้ปัญหาที่ 2 เพื่อหาลำดับสเตทที่เหมาะสม

เราจะใช้วิธี วีเทอร์บี อัลกอริทึม (Viterbi Algorithm) เพื่อหาลำดับสเตทที่ดีที่สุด ณ เวลา t หนึ่งๆ เมื่อกำหนดลำดับเหตุการณ์ $O = (o_1 o_2 ..... o_t)$ โดยนิยามให้

$$\delta_t(i) = \max_{q1 q2 ... qt} P[q_1 q_2 ... q_{t-1}, q_t = i, o_1 o_2 ..... o_t \mid \lambda]$$

หมายถึง ความน่าจะเป็นสูงสุดของเส้นทาง (path) ณ เวลา t ซึ่งเริ่มนับจากเหตุการณ์ที่ เวลาเริ่มต้นจนถึงเวลา t ที่สเตท i และโดยการอาศัยคุณสมบัติการเหนี่ยวนำ (induction) เราจะได้

$$\delta_{t+1}(i) = [\max \delta_t(i) a_{ij}] b_j(o_{t+1})$$

เราสามารถหาลำดับสเตทที่ดีที่สุดได้โดยใช้กระบวนการต่อไปนี้ เมื่อกำหนดให้

$\psi_t(i)$ เป็น อาเรย์ (array)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.3.2.1 การเริ่มต้น (initialization)

$$\delta_t(i) = \pi_i b_i(o_t) \quad \text{เมื่อ } 1 \le j \le N$$

$$\psi_t(i) = 0$$

2.7.3.2.2 การย้อนกลับ (recursion)

$$\delta_t(i) = [\max_{1 \le i \le N} \delta_{t-1}(i) a_{ij}] b_j(o_t) \quad \text{เมื่อ } 2 \le t \le T, 1 \le j \le N$$

$$\psi(i) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] \quad \text{เมื่อ } 2 \le t \le T, 1 \le j \le N$$

2.7.3.2.3 การสิ้นสุด (termination)

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

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

2.7.3.2.4 เส้นทางเดินย้อนกลับ (Path backtracking)

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

2.7.3.3 การแก้ปัญหาที่ 3 เพื่อหาโมเดลที่จะให้ผลตามลำดับค่าปรากฏหนึ่งๆ โดยเลือกค่าพารามิเตอร์ A, B, π ที่ดีที่สุดโดยใช้ กระบวนการทำซ้ำ (Iterative) วิธีที่เราเลือกใช้ คือ วิธีบาม-เวลช์ (Baum-Welch) หรือ EM (Expectation-Maximization)

เมื่อนิยามให้

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

หมายถึง ความน่าจะเป็นที่จะอยู่ที่สเตท i ณ เวลา t โดยกำหนดลำดับเหตุการณ์ O และแบบจำลอง ให้สามารถแสดงค่า $\gamma_t(i)$ ได้ดังนี้

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

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

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

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

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

หมายถึง ความน่าจะเป็นที่จะอยู่ที่สเตท i ที่เวลา t และสเตท j ที่เวลา t+1 เมื่อกำหนดแบบจำลองและลำดับค่าปรากฏให้ ซึ่งจากนิยามของตัวแปรไปข้างหน้าและตัวแปรย้อนกลับ สามารถนำมาสัมพันธ์กับ $\xi_t(i, j)$ ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

$$\xi_t(i,j) = \frac{P(q_t = i, q_{t+1} = j, O \mid \lambda)}{P(O \mid \lambda)}$$
$$= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O \mid \lambda)}$$
$$= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}$$

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

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

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

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

ดังนั้นสามารถหาค่าพารามิเตอร์ได้ดังนี้

$$\pi_i = \gamma_1(i) \quad ; 1 \le i \le N$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$b_j(k) = \frac{\sum_{t=1, o_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

จากกระบวนการข้างต้น ถ้าเราจะคำนวณซ้ำๆ โดยให้ $\lambda' = (A', B', \pi')$ แทน $\lambda = (A, B, \pi)$ ซึ่งเป็นแบบจำลองเริ่มต้นแล้ว จะทำให้ความน่าจะเป็นของการเกิดลำดับค่าปรากฏ O ดีขึ้นจนกระทั่งถึงจุดวิกฤตจึงหยุด ซึ่งเราจะได้จุดวิกฤตของฟังก์ชันความน่าจะเป็นในกรณีที่ $\lambda' = \lambda$ หรือถ้า λ' มีความน่าจะเป็นมากกว่าแบบจำลอง λ

ในลักษณะที่ $P(O \mid \lambda') > P(O \mid \lambda)$ นั่นก็คือเราจะได้แบบจำลอง λ' ใหม่ที่น่าจะทำให้เกิดลำดับค่าปรากฏ O ได้ดีกว่า

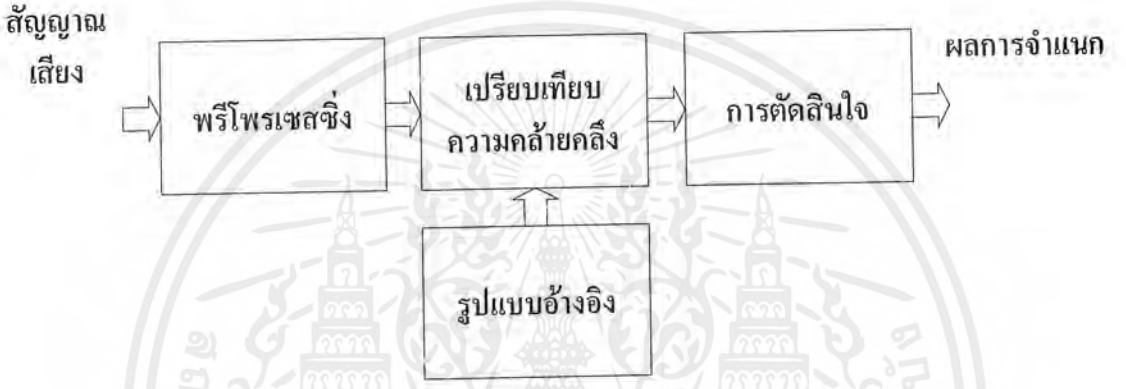
2.8 กฎการตัดสินใจ

การตัดสินใจจะดูว่าความน่าจะเป็นของคำศัพท์ที่เข้าไปเปรียบเทียบกับแบบจำลองอันไหนให้ความน่าจะเป็นมากที่สุดคำศัพท์ที่เข้าไปเปรียบเทียบก็จะเป็นคำศัพท์คำนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
37143

บทที่ 3
การคำนวณและการสร้าง

ระบบการจำแนกเสียงพูดโดยทั่วไปสามารถแบ่งออกเป็นภาคต่างๆตามบล็อกไดอะแกรมจากรูปที่ 3.1 ซึ่งประกอบไปด้วยภาคพรีโพรเซสซิ่ง ภาคการหาค่าพารามิเตอร์ ภาคแบบอ้างอิง ภาคการเปรียบเทียบความคล้ายคลึง และภาคการตัดสินใจ สำหรับในส่วนของพรีโพรเซสซิ่ง เป็นขั้นตอนการปรับข้อมูลเสียงจากอินพุทให้เหมาะสม นั่นก็คือ การหาขอบเขตที่เป็นส่วนของคำที่แท้จริง ก่อนที่จะส่งไปภาคอื่นๆต่อไป



รูปที่ 3.1 แสดงบล็อกไดอะแกรมการจำแนกเสียงพูดโดยทั่วไป

สำหรับปริญญานิพนธ์นี้ แต่ละภาคจะมีวิธีการทำดังแสดงในบล็อกไดอะแกรมในรูปที่ 3.2 ซึ่งสามารถอธิบายเป็นส่วน ๆ ได้ดังนี้

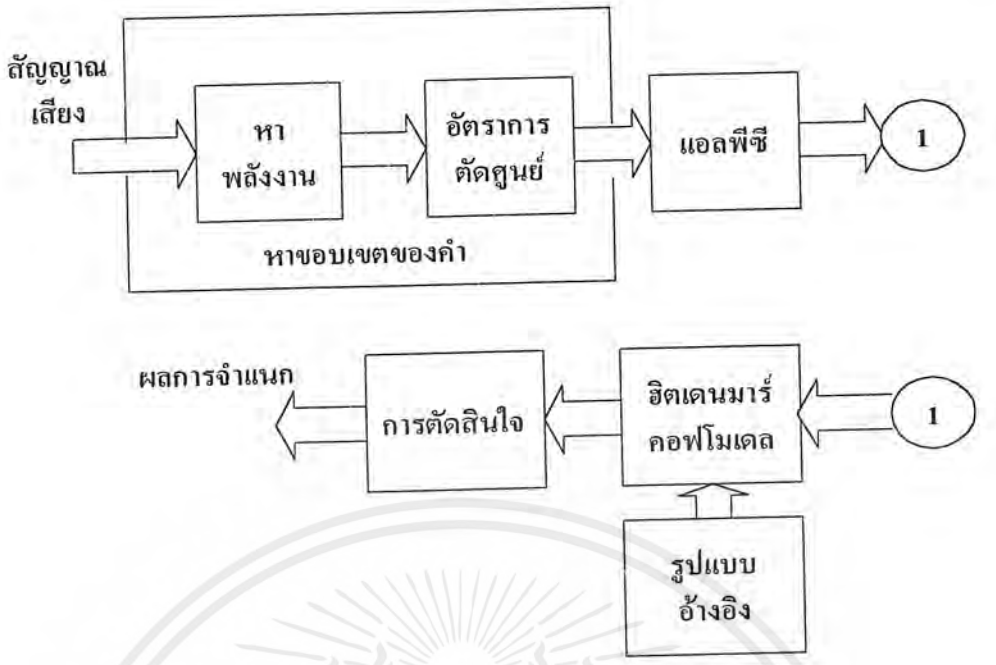
<u>ภาคพรีโพรเซสซิ่ง</u> เป็นขั้นตอนการหาค่าพารามิเตอร์ต่างๆเพื่อเตรียมการเข้าไปเปรียบเทียบในขั้นตอนการเปรียบเทียบความคล้ายคลึง จะประกอบด้วย

3.1 การหาขอบเขตของคำ คือ ขั้นตอนในภาคพรีโพรเซสซิ่ง โดยจะใช้หลักการหาค่าพลังงานของสัญญาณเสียง และการหาค่าอัตราการตัดศูนย์ในโดเมนเวลา ซึ่งมีวิธีการดังต่อไปนี้

ขั้นตอนที่ 1 ใช้การวิเคราะห์หาค่าพลังงานของสัญญาณ ในการหาขอบเขตของคำ ซึ่งมีวิธีการดังนี้

1. คำนวณหาค่าพลังงานของสัญญาณในช่วงเวลาที่ถูกบันทึกทั้งหมด โดยการคำนวณหาค่าพลังงานของสัญญาณซึ่งจะทำการคำนวณเป็นเฟรม กำหนดให้เฟรมหนึ่งมีจำนวนแซมเปิล100แซมเปิล และทำการคำนวณเป็นเฟรมต่อไปเรื่อย ๆ จนครบทุกแซมเปิลของสัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 แสดงบล็อกไดอะแกรมการจำแนกเสียงพูด

2. พิจารณาหาค่าพลังงานสูงสุดของสัญญาณ (E_max) และดูว่าที่ค่าพลังงานสูงสุดนี้อยู่ที่เฟรมใด แล้วกำหนดให้เป็นเฟรมที่ m

3. ทำการกำหนดพลังงานอ้างอิง (E_reference) โดยที่ซึ่งพลังงานอ้างอิง สามารถคำนวณได้จาก a เท่าของพลังงานสูงสุด (a*E_max) ซึ่งในที่นี้ค่า a มีค่า=0.2

พลังงานอ้างอิงนี้จะเป็นตัวบอกว่า จุดเริ่มต้นและจุดสุดท้ายของคำจะอยู่นอกช่วงนี้

4. เมื่อกำหนดพลังงานอ้างอิง เป็นที่เรียบร้อยแล้ว ทำการหาจุดเริ่มต้นของคำ พิจารณาโดยเฟรมที่มีค่าพลังงานสูงสุด จะถูกพิจารณาเป็นเฟรมแรก (เฟรมที่ m) ซึ่งนำค่าพลังงานของสัญญาณเฟรมที่ m ไปเปรียบเทียบกับพลังงานอ้างอิง แล้วเลื่อนเฟรมที่จะเปรียบเทียบพลังงานกับพลังงานอ้างอิงไปด้านหลังเรื่อยๆ (เฟรมที่ m-1,m-2,…,2,1) จนกว่าจะพบเฟรมที่มีค่าพลังงานน้อยกว่าค่าพลังงานอ้างอิงเป็นเฟรมแรก ก็จะได้จุดเริ่มต้นของสัญญาณ(a1) โดยถือว่าค่าพลังงานของสัญญาณค่าแรกที่มีค่าน้อยกว่าระดับอ้างอิง (E_ref) เป็นจุดเริ่มต้นของสัญญาณ

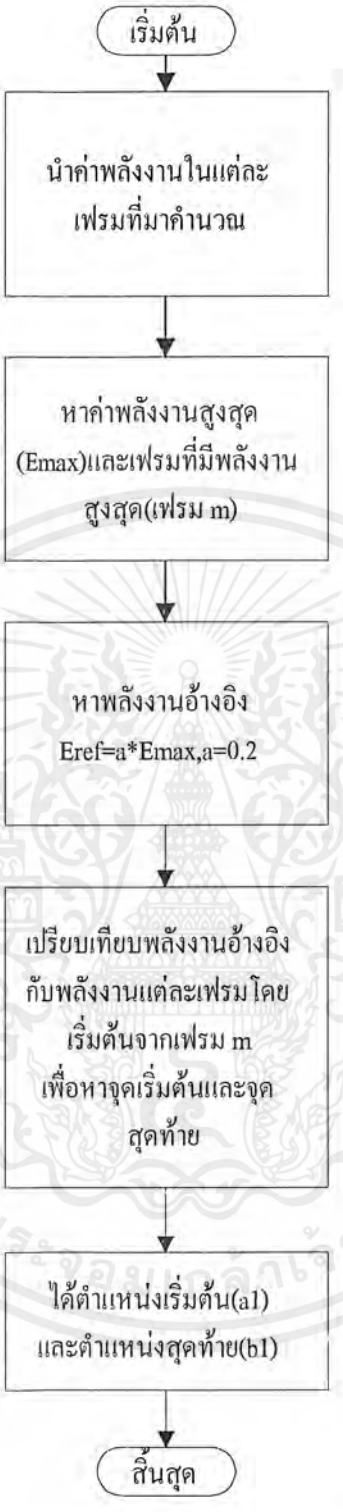
5. การหาจุดสุดท้ายของคำ จะทำการพิจารณาเช่นเดียวกับการหาจุดเริ่มต้นของคำ คือเริ่มเปรียบเทียบจากเฟรมที่มีพลังงานสูงสุด (เฟรมที่ m) เช่นกัน แต่การเลื่อนเฟรมจะเลื่อนไปทางด้านหน้าแทน (เฟรมที่ m+1,m+2,…) จนกว่าจะพบเฟรมแรกที่มีค่าพลังงานน้อยกว่าค่าพลังงานอ้างอิง ก็จะได้จุดสุดท้ายของสัญญา(b1) โดยถือว่าค่าพลังงานของสัญญาณค่าแรกที่มีค่าน้อยกว่าพลังงานอ้างอิง (E_ref) เป็นจุดสุดท้ายของคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากขั้นตอนทั้งหมดที่กล่าวมานี้สามารถพิจารณารูป 3.3 และรูป 3.4 ตามลำดับ

รูปที่ 3.3 แสดงผังงานการหาค่าพลังงานของสัญญาณเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียงโดยการเปรียบเทียบพลังงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 2 ใช้การวิเคราะห์หาค่าอัตราการตัดศูนย์ สาเหตุที่ใช้วิธีนี้ช่วยเนื่องจากการหาขอบเขตของคำโดยการเปรียบเทียบค่าพลังงานอาจมีข้อผิดพลาดได้ จึงต้องใช้วิธีหาค่าอัตราการตัดศูนย์ (Zero crossing) เข้าช่วยพิจารณา ซึ่งมีวิธีการดังนี้

1. คำนวณหาค่าอัตราการตัดศูนย์ในช่วงเวลาที่ถูกบันทึกทั้งหมด โดยการคำนวณหาค่าอัตราการตัดศูนย์จะคำนวณเป็นเฟรม กำหนดให้แต่ละเฟรมมีตัวอย่าง 100 ตัวอย่าง และทำการคำนวณจนครบช่วงเวลาของเสียงที่ถูกบันทึกทั้งหมด

2. กำหนดระดับอ้างอิง (Z_max) กำหนดให้เป็นเฟรมที่ p

3. กำหนดค่าอัตราการตัดศูนย์ (Z_ref) ซึ่งมีค่าเป็น a เท่าของอัตราการตัดศูนย์สูงสุด (a*Z_ref) โดยค่า a จะเท่ากับ 0.2 เช่นกัน

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

5. การหาจุดสุดท้าย จากจุดสุดท้าย (b2)ที่ได้จากขั้นตอนที่ 1 จะพิจารณาเหมือนข้อ 3 คือจะเป็นเฟรมแรกที่จะเปรียบเทียบอัตราการตัดศูนย์อ้างอิง ถ้ามีค่ามากกว่า จุดสุดท้ายจะถูกเปลี่ยนตำแหน่ง โดยจะเลื่อนเฟรมไปทางด้านหน้า(เฟรมที่ b1+1,b1+2,...) จนกว่าจะเจอจุดแรกที่มีค่าน้อยกว่าอัตราการตัดศูนย์อ้างอิง(b2) และจะถือว่าจุดนี้เป็นจุดสุดท้ายที่แท้จริง แต่ถ้าอัตราการตัดศูนย์ที่ b1 น้อยกว่าอัตราการตัดศูนย์อ้างอิงก็กล่าวได้ว่า b1 เป็นจุดสุดท้ายที่ถูกต้องอยู่แล้วเช่นกัน

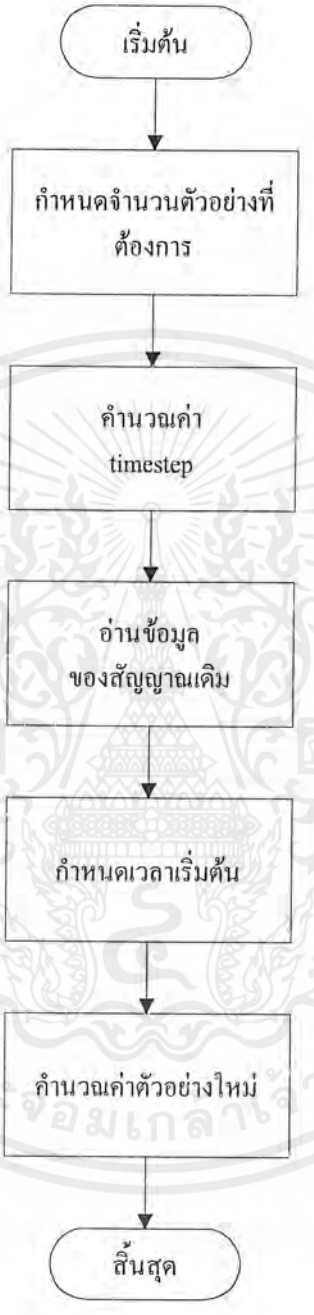
จากขั้นตอนทั้งหมดที่กล่าวมานี้สามารถพิจารณารูป 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.5 แสดงผังงานการหาจุดเริ่มต้นและจุดสุดท้ายของสัญญาณเสียง
โดยการหาอัตราการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2การนอร์มอลไลซ์(Normalization)เนื่องจากสัญญาณเสียงพูดแต่ละคำมีความยาวไม่เท่ากันจึงต้องมีการนอร์มอลไลซ์สัญญาณเสียงพูดให้ยาวเท่ากัน ขั้นตอนการนอร์มอลไลซ์สัญญาณเสียงพูดแสดงดังรูปที่3.6



รูปที่3.6แสดงขั้นตอนการนอร์มอลไลซ์สัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การหาค่าพารามิเตอร์ ซึ่งเป็นขั้นตอนการหาค่าพารามิเตอร์ สำหรับปริญญานิพนธ์นี้จะทำการหาค่าพารามิเตอร์ที่สำคัญของสัญญาณเสียง โดยจะใช้วิธีลิเนียร์พรีดิกทีฟ (Linear Predictive Coding – LPC) และหาค่าสัมประสิทธิ์เซปสตรัม ก่อนที่จะหาค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนักต่อไป

การวิเคราะห์แบบลิเนียร์พรีดิกทีฟเป็นการวิเคราะห์เสียงพูดที่ดีวิธีหนึ่ง โดยการวิเคราะห์นี้ใช้ในการหาค่าพารามิเตอร์พื้นฐานของเสียง ซึ่งพารามิเตอร์ที่หาได้จากวิธีนี้มีความถูกต้องและใช้เวลาในการคำนวณไม่มากนัก

การหาค่าพารามิเตอร์ของเสียงนี้ สำหรับลิเนียร์พรีดิกทีฟจะทำการหาค่าสัมประสิทธิ์แอลพีซี (a) โดยที่เราจะต้องเตรียมสัญญาณในการวิเคราะห์ก่อน ซึ่งขั้นตอนในการวิเคราะห์แสดงดังรูปที่ 3.9

การวิเคราะห์สามารถแบ่งเป็นขั้นตอนต่าง ๆ ดังนี้

3.3.1 การพรีเอมฟาซีส (Preemphasis)

3.3.2 การแบ่งช่วงสัญญาณ (Frame Blocking)

3.3.3 การวินโดว์ (Windowing)

3.3.4 การหาออโตคอร์รีเลชั่น (Autocorrelation Analysis)

3.3.5 การหาสัมประสิทธิ์แอลพีซี a (LPC Analysis)

3.3.6 การหาอัตราการขยาย G

3.3.7 การเปลี่ยนพารามิเตอร์แอลพีซีเป็นสัมประสิทธิ์เซปสตรัม

3.3.8 การเวทค่าพารามิเตอร์ (Parameter Weighting)

3.3.1 การพรีเอมฟาซีส

เนื่องจากสัญญาณเสียงพูด จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ ดังนั้นเพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน มีค่าค่อนข้างคงที่ เราจึงต้องมีการพรีเอมฟาซีส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้น นั่นคือ การพรีเอมฟาซีสเป็นการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (High pass filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชั่นถ่ายโอนดังนี้

$$H(z) = 1\text{-}az^{-1} \quad 0.9 \leq a \leq 1.0 \qquad (3.3.1)$$

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

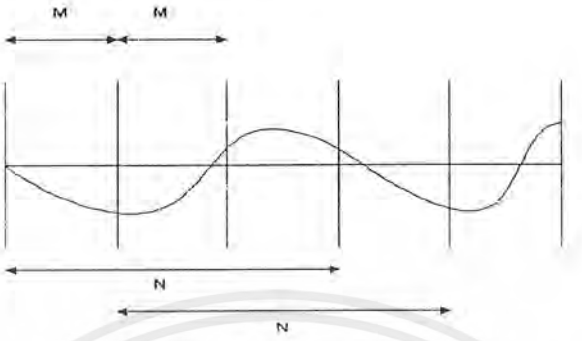
$$s'(n) = s(n)\text{-}as(n\text{-}1) \qquad (3.3.2)$$

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 การแบ่งช่วงสัญญาณ

สัญญาณที่ผ่านการพรีเอมฟจซีสแล้ว s'(n) จะถูกแบ่งออกเป็นช่วง ๆ หรือเฟรมโดยมีช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างของสัญญาณ ดังรูป3.7



รูปที่ 3.7 แสดงการแบ่งช่วงสัญญาณที่ใช้ในการวิเคราะห์

โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า M ใหญ่กว่า N ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N จำทำให้ตัวอย่างของสัญญาณทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่จะทำให้การคำนวณช้าลง

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

$$x_l(n) = s'(Ml+n) \qquad n=0,1.....,N-1. \qquad l=0,1,.....,L-1 \qquad (3.3.3)$$

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

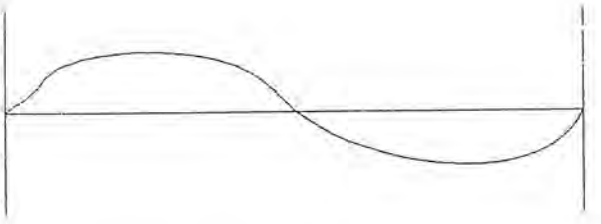
3.3.3 การวินโดว์

พิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วงใด ๆ ที่ตัดมาวิเคราะห์ (รูปที่ 3.8) จะเห็นว่าที่ขอบของเฟรมนี้มีความไม่ต่อเนื่องของสัญญาณ ถ้าพิจารณาในโดเมนความถี่สูง ก็จะมีความถี่สูงเกิดขึ้น ดังนั้นเพื่อที่จะลดองค์ประกอบทางความถี่สูงเหล่านี้ เราจะคูณด้วยฟังก์ชั่นวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปคตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

$$x_l'(n) = x_l(n)w(n), \qquad 0 \leq n \leq N-1 \tag{3.3.4}$$



รูปที่ 3.8 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์

ในที่นี้จะใช้ฟังก์ชั่นวินโดว์แฮมมิ่ง (Hamming window function) ซึ่งนิยามโดยสมการดังนี้ คือ

$$w(n) = 0.54 - 0.46\cos(2\pi n/N-1), \quad 0 \leq n \leq N-1 \tag{3.3.5}$$

3.3.4 การหาออโตคอร์รีเลชั่น

สัญญาณในแต่ละเฟรมที่ผ่านวินโดว์แล้ว จะถูกผ่านการเข้าออโตคอร์รีเลชั่น

$$R(m) = \sum_{m=0}^{N-1-m} x'(n)x'(n+m) \qquad m = 0,1,\ldots,p \tag{3.3.6}$$

เรียกวิธีการนี้ว่าการประมาณเชิงเส้นอันดับ p โดยเรียกอันดับ p ว่า อันดับของการวิเคราะห์แอลพีซี โดยทั่วไปค่า p จะมีค่าระหว่าง 8 ถึง 16

3.3.5 การหาสัมประสิทธิ์แอลพีซี a

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

$E_0 = R(0)$

เมื่อ $m = 1,2,\ldots..,p$

$$K_m = -\left[R(m) + \sum a_{m-1}(k)R(m-k)\right] / E_{m-1} \tag{3.3.7}$$

$a_m(m) = K_m$

$a_m(k) = a_{m-1}(k) + aK_{m-1}(m-k), \qquad 1 \leq k \leq m-1$

$E_m = (1-K_m^2)E_{m-1}$

โดยที่สัมประสิทธิ์ออโตคอร์รีเลชั่น $R(i) = \sum x_l'(n)x_l'(n+m), \quad m=0,1,\ldots..,p;$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

3.3.6 การหาอัตราการขยาย G

จากสมการ 2.5.6 จะได้ว่า

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

และจากสมการ 2.5.11 จะได้ว่า

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

$$= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \qquad (3.3.8)$$

และจากสมการ 3.3.8 เราสามารถหาค่า G โดยตรงจาก

$$G^2 = \frac{R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)}{\sum_{m=0}^{N-1} u^2(m)} \qquad (3.3.9)$$

3.3.7 การเปลี่ยนพารามิเตอร์แอลพีซีเป็นสัมประสิทธิ์เซปสตรัม

ในการรู้จำเสียงพูดนั้น สัมประสิทธิ์เซปสตรัมนี้เป็นพารามิเตอร์ที่น่าเชื่อถือได้ดีกว่าสัมประสิทธิ์ LPC ทั้งยังมีความใกล้ชิดกับการรับรู้เสียงตามความรู้สึกของมนุษย์โดยแท้จริง สัมประสิทธิ์เซปสตรัมสามารถหาได้โดยตรงจากสัมประสิทธิ์ LPC ดังนี้

$$C_o = \ln G$$
$$C_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) C_k a_{m-k} \quad ,1 \le m \le p$$
$$C_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) C_k a_{m-k} \quad ,m \succ p \qquad (3.3.10)$$
$$Q = \frac{3}{2} P$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.8 การเวทค่าพารามิเตอร์ (Parameter Weighting)

เนื่องจากสัมประสิทธิ์เซปสตรัมที่ได้นั้น ช่วงลำดับต้นๆ และลำดับท้ายๆของเฟรมที่นำมาวิเคราะห์จะเกิดความคลาดเคลื่อนมากกว่าบริเวณส่วนอื่นเพราะฉะนั้นจึงทำการถ่วงน้ำหนักเพื่อลดค่าความเคลื่อนดังกล่าวนี้ ด้วยฟังก์ชันเวทติ่งดังนี้คือ

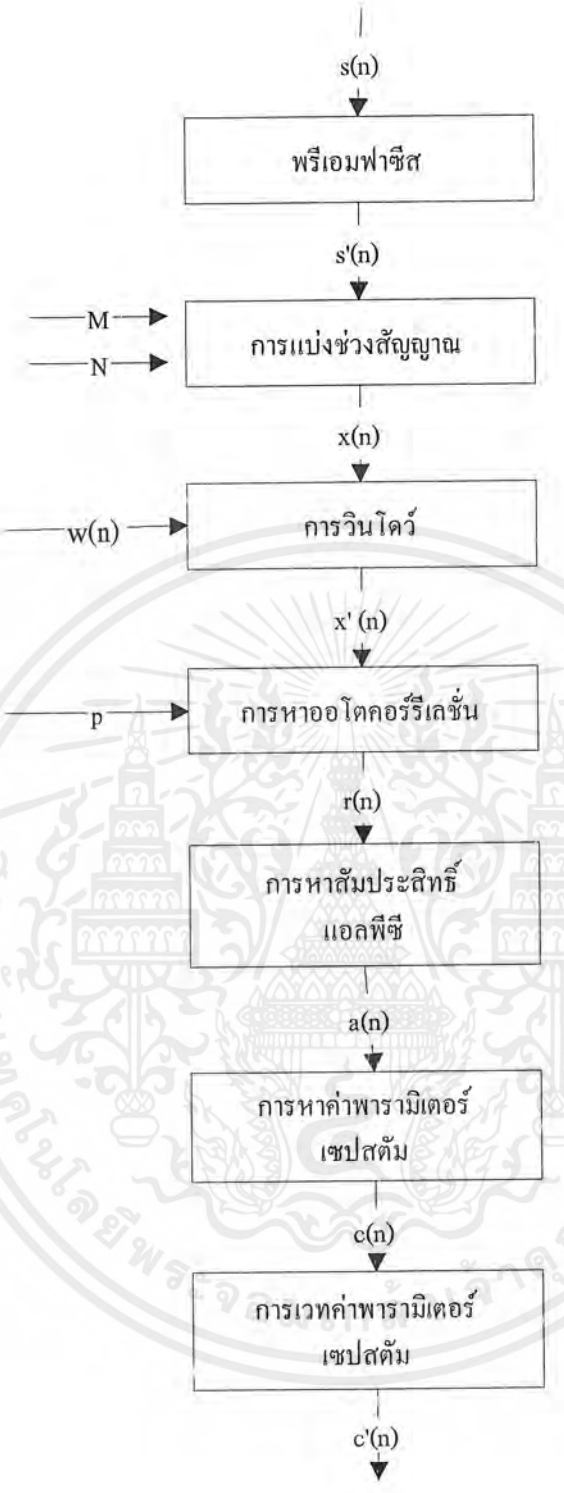
$$W_m = [1 + \frac{Q}{2}\sin(\frac{\pi m}{Q})] \quad ,1 \le m \le Q \tag{3.3.11}$$

จะได้

$$C'_m = C_m * W_m$$

จากที่กล่าวมาแล้วทั้งหมด สามารถแสดงขั้นตอนการเตรียมสัญญาณทั้งหมดได้ดังรูปที่3.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 แสดงผังงานขั้นตอนการเตรียมสัญญาณในการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การสร้างโค้ดบุ้ค

จากการทดสอบ(L.R.Rabiner.S.E.Levinson และ Sonthi.1982.) จะได้ว่าที่ขนาดโค้ดบุ้คเท่ากับ 64 จะมีค่าความคลาดเคลื่อน โดยวิธีการคำนวณแบบสแควเออเรอร์ดิสทอชั่น(Square error distortion)ในการหาระยะทางเนื่องจากเป็นวิธีที่ง่ายและรวดเร็ว

3.4.1 การสร้างโค้ดบุ้คโดยนำเทรนนิ่งเซตที่ได้จากการประมาณเชิงเส้นมาผ่านกระบวนการดังนี้

3.4.2 สุ่มค่าโค้ดบุ้คเริ่มต้นมา 64 ตัว ตัวละ 12 มิติ

3.4.3 หาระยะทางระหว่างโค้ดบุ้คกับเทรนนิ่งเซตแต่ละตัวโดยใช้ความคลาดเคลื่อนกำลังสอง

3.4.5 จัดกลุ่มของเวคเตอร์อินพุทโดยพิจารณาจากระยะทางที่น้อยที่สุด

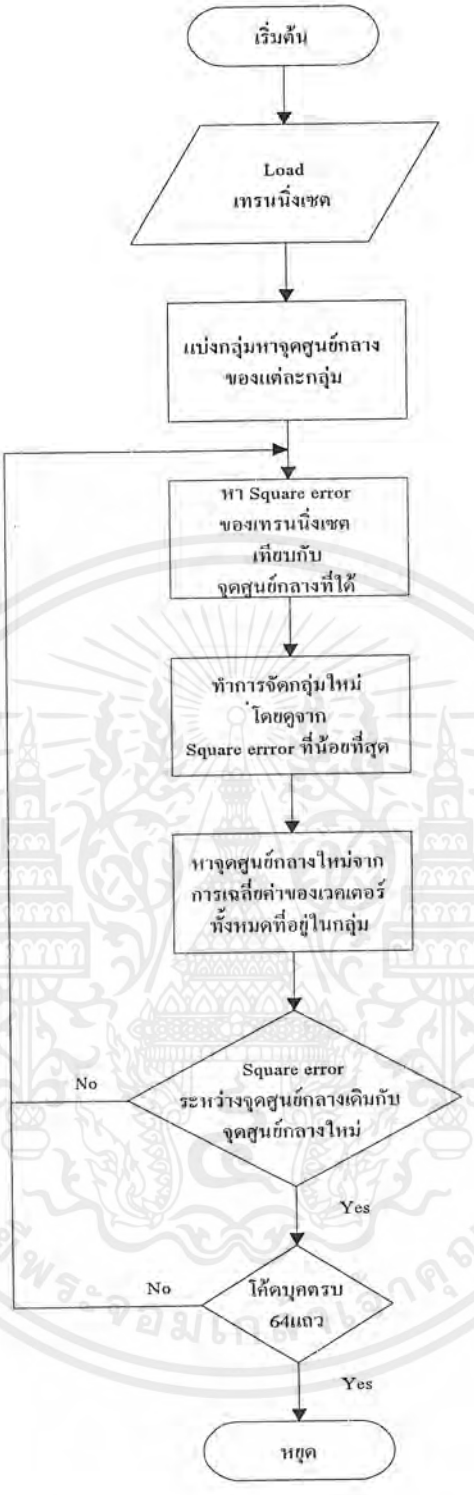
3.4.6 หาจุดศูนย์กลางของกลุ่ม

ทำขั้นตอน 3 และ 4 จนกว่าความคลาดเคลื่อนรวมจะน้อยกว่า 0.0001 ซึ่งจุดศูนย์กลางที่ได้ก็คือโค้ดบุ้คนั่นเอง จะแสดงขั้นตอนการทำงานดังรูปที่ 3.10

3.5 เวคเตอร์ควอนไตเซชั่น

จะควอนไตซ์ค่าที่ได้กับ โค้ดบุ้คที่คำนวณมาได้โดยดูความคลาดเคลื่อนที่น้อยที่สุดโดยวิธีสแควเออร์เรอร์ จะแสดงขั้นตอนการทำงานดังรูปที่ 3.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แสดงขั้นตอนการสร้างโค้ดบุค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.11 แสดงขั้นตอนการควอนไตเซชั่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาครูปแบบอ้างอิง

3.6 รูปแบบอ้างอิง คือส่วนที่เก็บค่าพารามิเตอร์ของสัญญาณ (สัมประสิทธิ์แอลพีซี) ที่ใช้เป็นสัญญาณแบบอ้างอิง ในที่นี้คือสัญญาณเลข 0-9 กล่าวคือเมื่อมีสัญญาณเสียงทดสอบเข้ามาและผ่านภาคการหาขอบเขตของคำ และหาค่าพารามิเตอร์ของสัญญาณแล้ว ก็นำมาเปรียบเทียบความคล้ายคลึงโดยวิธีฮิดเดนมาร์คอฟโมเดลกับสัญญาณอ้างอิง 0-9 ในภาคเปรียบเทียบความคล้ายคลึงต่อไป

ภาคเปรียบเทียบความคล้ายคลึง คือขั้นตอนการเปรียบเทียบความคล้ายคลึง โดยการเปรียบเทียบระหว่างสัญญาณแบบทดสอบโดยนำข้อมูลมาจากส่วน พรีโพรเซสซิ่ง กับพารามิเตอร์ของสัญญาณแบบอ้างอิง 0-9 ที่เก็บไว้ในภาคแบบอ้างอิงนั้นเอง ซึ่งในปริญญานิพนธ์นี้จะใช้วิธีของ ฮิดเดนมาร์คอฟโมเดล (Hidden Markov Model) HMM ในขั้นตอนการเปรียบเทียบความคล้ายคลึงนี้

3.7 ฮิดเดนมาร์คอฟโมเดล ในการสร้างแบบจำลอง และการรู้จำเสียงเพื่อเปรียบเทียบความคล้ายคลึงของสัญญาณเสียงแบบ HMM ของเสียงมีขั้นตอนดังต่อไปนี้

3.7.1 สุ่มค่าเริ่มต้น a,b และกำหนดให้ $\pi = [100000]$ ตามเงื่อนไขแบบจำลองแบบ Left – Right

3.7.2 หาค่า α, β จากค่า a,b เริ่มต้น และลำดับค่าปรากฏ $O = \{O_1, O_2 O_3 ... O_T\}$ ซึ่งเรียกว่าลำดับเทรนนิ่ง ตามวิธีของ Forward – Backward Procedure โดยใช้ลำดับของค่าปรากฏหลายๆ ลำดับเข้ามาเทรน เพื่อความถูกต้องมากขึ้น

3.7.3 หาค่าพารามิเตอร์ a, b, π ที่ให้ความน่าจะเป็นสูงสุด ที่จะเป็นแบบจำลอง λ ที่เหมาะสม

3.7.4 ตรวจสอบค่าพารามิเตอร์แบบจำลองที่ได้ว่าลู่เข้าหรือยังโดยใช้วิธีการคำนวณค่า a,b ซ้ำโดยในที่นี้ตั้งไว้ว่าเมื่อไม่มีการเปลี่ยนแปลงหรือค่าเท่าเดิมให้หยุดการทำซ้ำ ก็จะได้ค่าพารามิเตอร์ a, b, π ของแบบจำลองที่ต้องการ

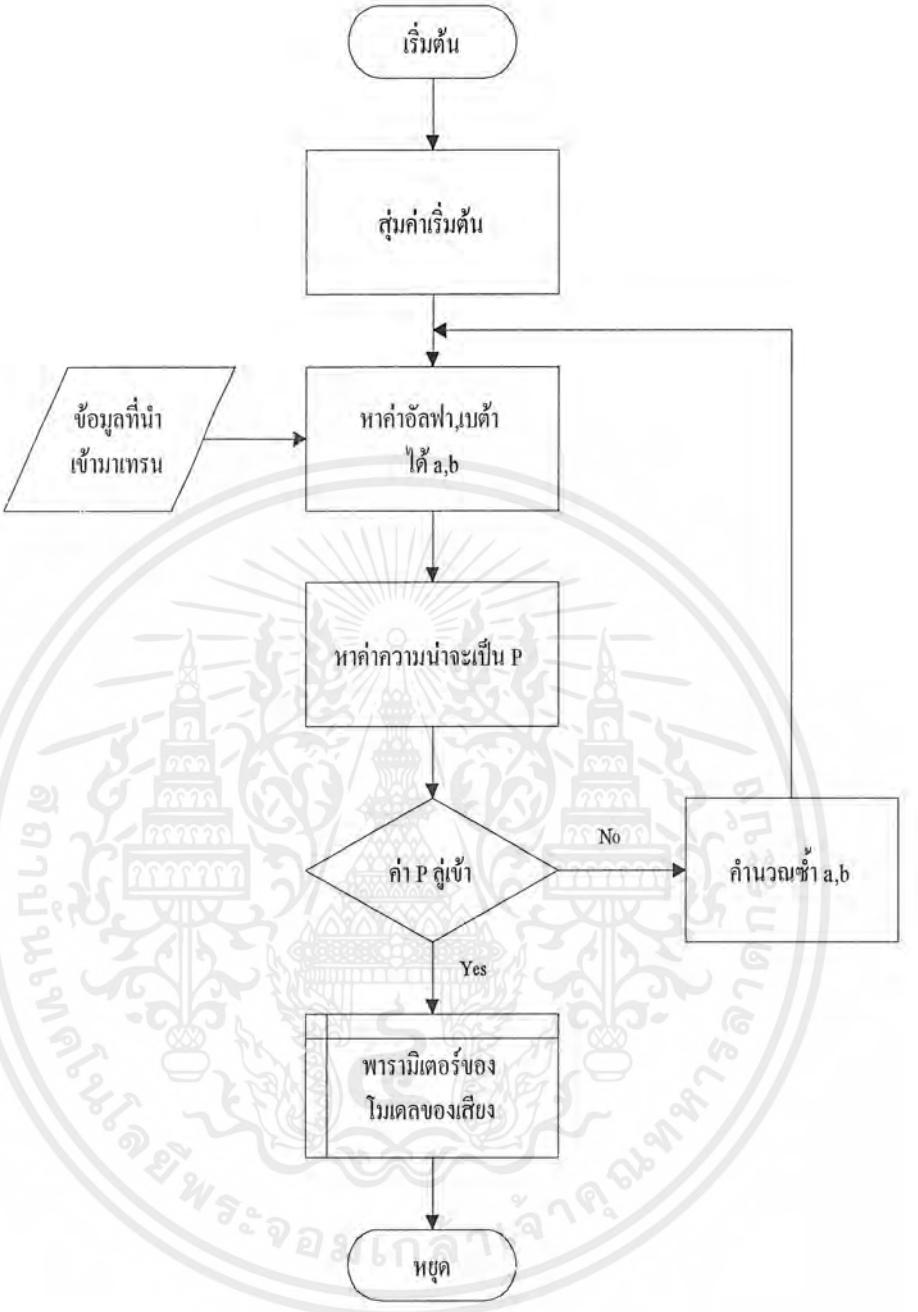
3.7.5 เก็บค่าพารามิเตอร์ a, b, π ที่ได้จากข้อ 3.5.4 เป็นพารามิเตอร์ของแบบจำลองไว้

และหลังจากได้แบบจำลอง HMM ของแต่ละคำศัพท์แล้วเมื่อมีลำดับค่าปรากฏ $O = \{O_1, O_2 O_3 ... O_T\}$ ของเสียง unknow ซึ่งเป็นเสียงที่ต้องการทดสอบเข้ามา เราจะทำการคำนวณหาความน่าจะเป็น $P(O \mid \lambda)$ ทุกแบบจำลองของแต่ละคำศัพท์โดยใช้วิธี viterbi algorithm แล้วเลือกเอาคำศัพท์ที่มีความน่าจะเป็นสูงสุด ซึ่งก็คือ คำศัพท์ที่แบบจำลองจำได้นั้นเอง ดังแสดงในรูปที่ 3.12

ภาคการตัดสินใจ เป็นขั้นตอนการตัดสินว่าเสียงนั้นเป็นเสียงอะไร โดยเปรียบเทียบกับแบบอ้างอิง

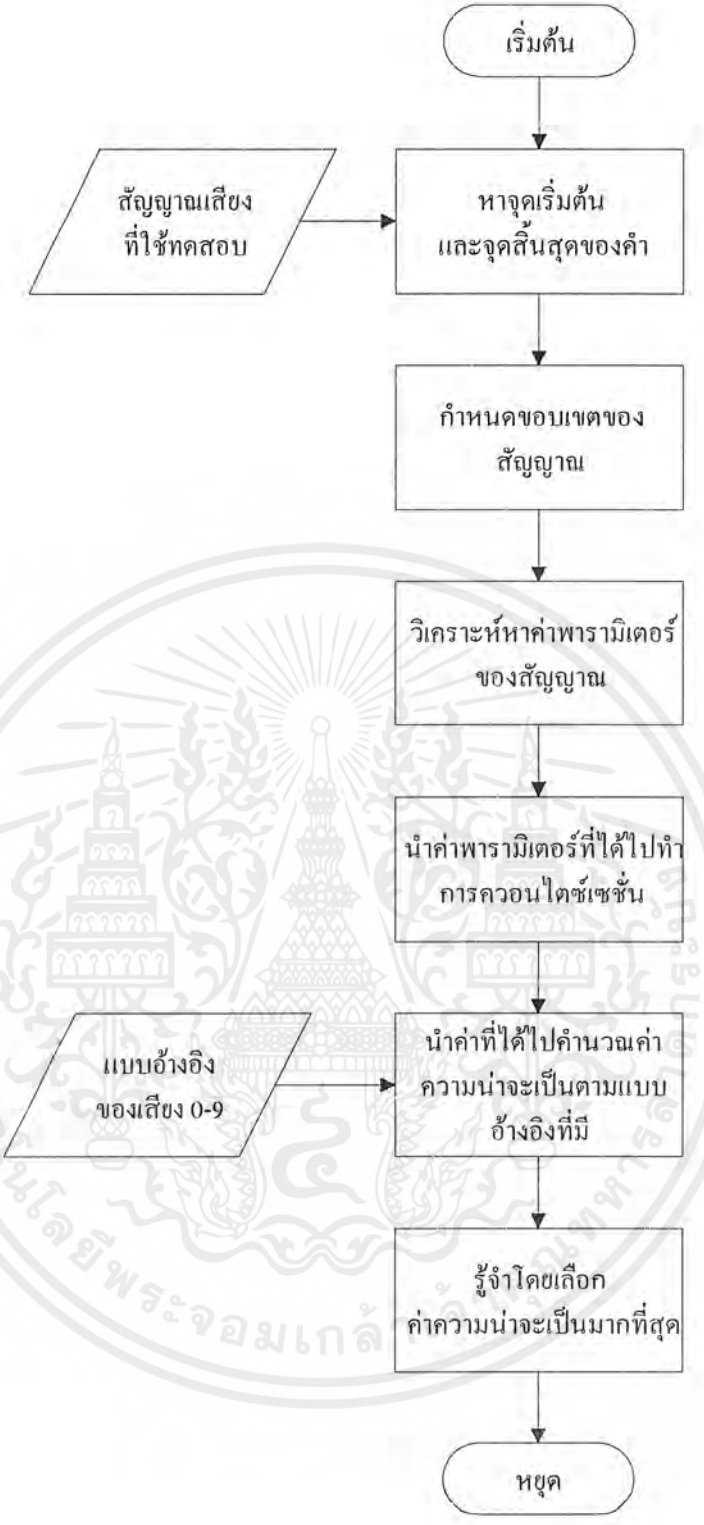
3.8 ภาคการตัดสินใจ เมื่อทำการเปรียบเทียบความคล้ายคลึงระหว่างสัญญาณแบบทดสอบกับสัญญาณแบบอ้างอิง 0-9 แล้ว หลังจากนั้นก็ทำการเลือกผลที่ได้จากการเปรียบเทียบในภาคที่แล้วที่เหมือนกันมากที่สุด ตามกฎการตัดสินใจ นั่นคือการตัดสินใจในการจำแนกเสียงพูดว่าเสียงที่นำมาทดสอบเป็นเลขอะไรนั่นเอง ดังรูปที่ 3.13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.12 แสดงขั้นตอนการสร้างแบบจำลองอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.13 แสดงขั้นตอนการรู้จำเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

จากโครงงานที่ได้ทดลอง ซึ่งมีผลการทดลองดังนี้

4.1 การหาขอบเขตของคำ แบ่งเป็นการหาค่าพลังงาน และการหาค่าอัตราการตัดศูนย์

4.2 การนำคำที่ทำการหาขอบเขตแล้วมาทำการนอร์มอลไลซ์เพื่อให้ได้ช่วงเวลาและขอบเขตที่เท่ากัน

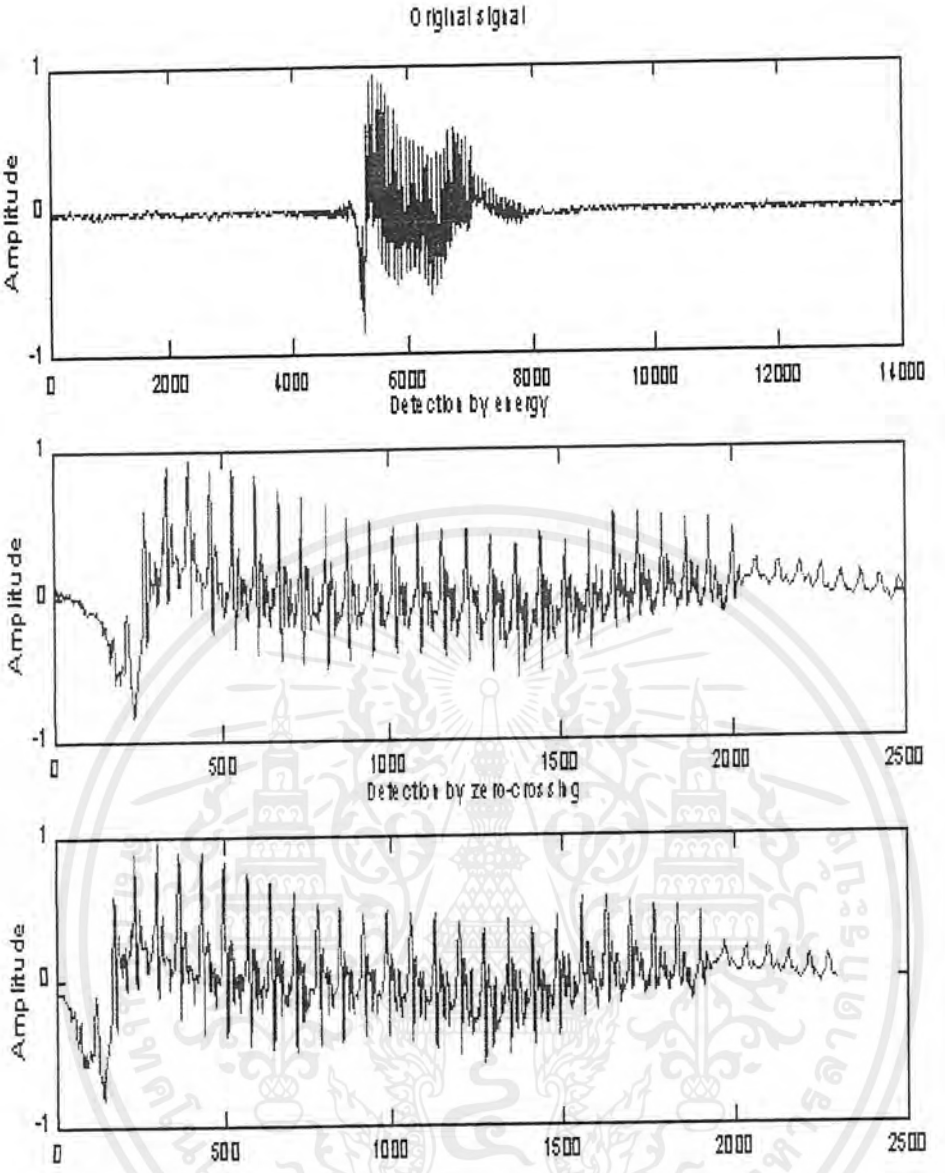
4.3 การหาค่าพารามิเตอร์ของคำที่ได้ผ่านการหาขอบเขตและนอร์มอลไลซ์แล้วซึ่งในที่นี้จะใช้ค่าพารามิเตอร์เวทเซปสตรัม

4.4 นำค่าพารามิเตอร์ที่ได้ไปทำการควอนไตซ์เซชั่นกับโค้ดบุ๊คเพื่อควอนไตซ์ค่าพำรามิเตอร์ให้มีจำนวนน้อยลง

4.5 การนำค่าพารามิเตอร์ที่ได้ควอนไตซ์แล้วมาทำการหาโมเดลโดยใช้วิธีฮิดเดนมาร์คอฟโมเดลจะได้ค่ามาเป็นโมเดลของเสียงต่างๆซึ่งจะประกอบไปด้วยค่า $\lambda = a, b, \pi$

ซึ่งแต่ละเสียงจะได้ผลการทดลองดังรูปต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



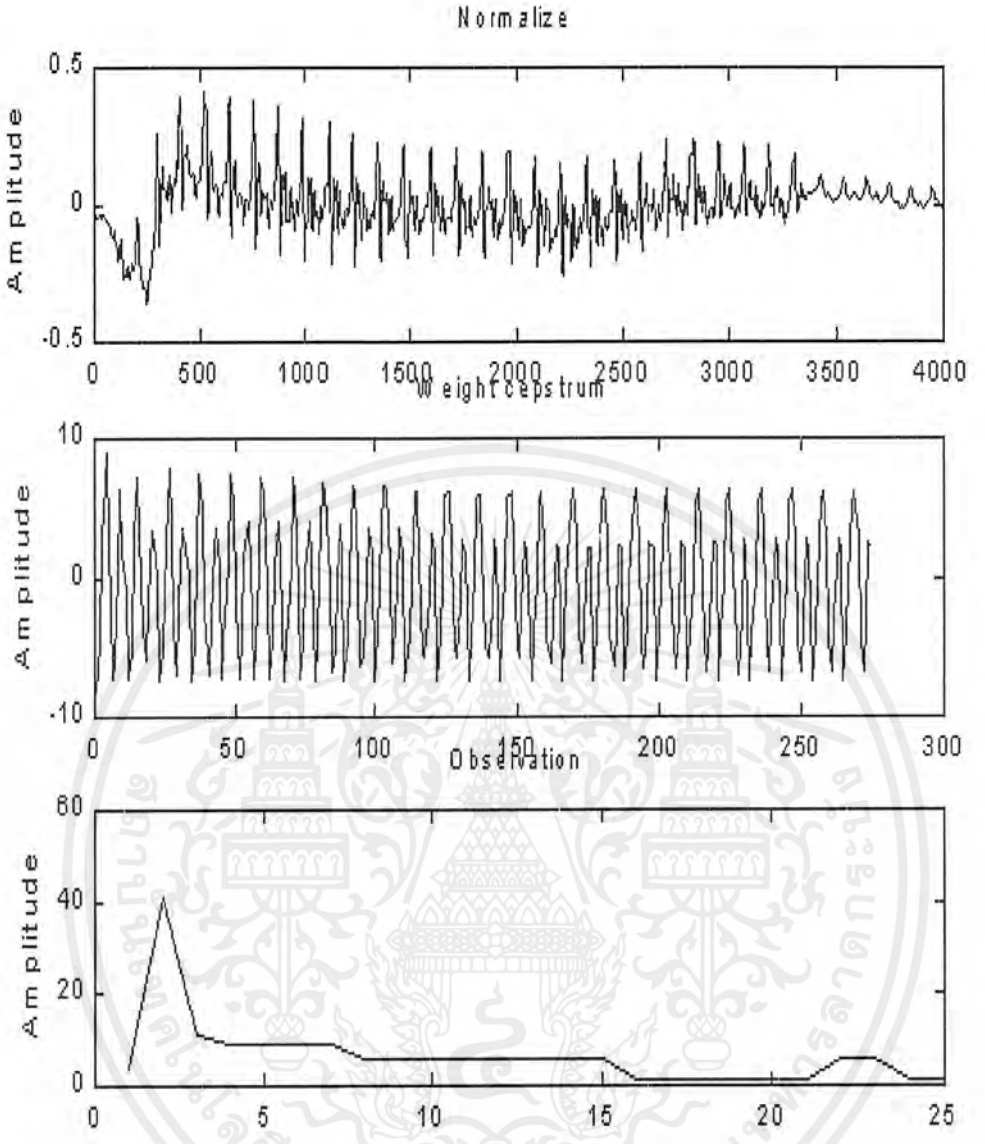
รูปที่ 4.1 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 0

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



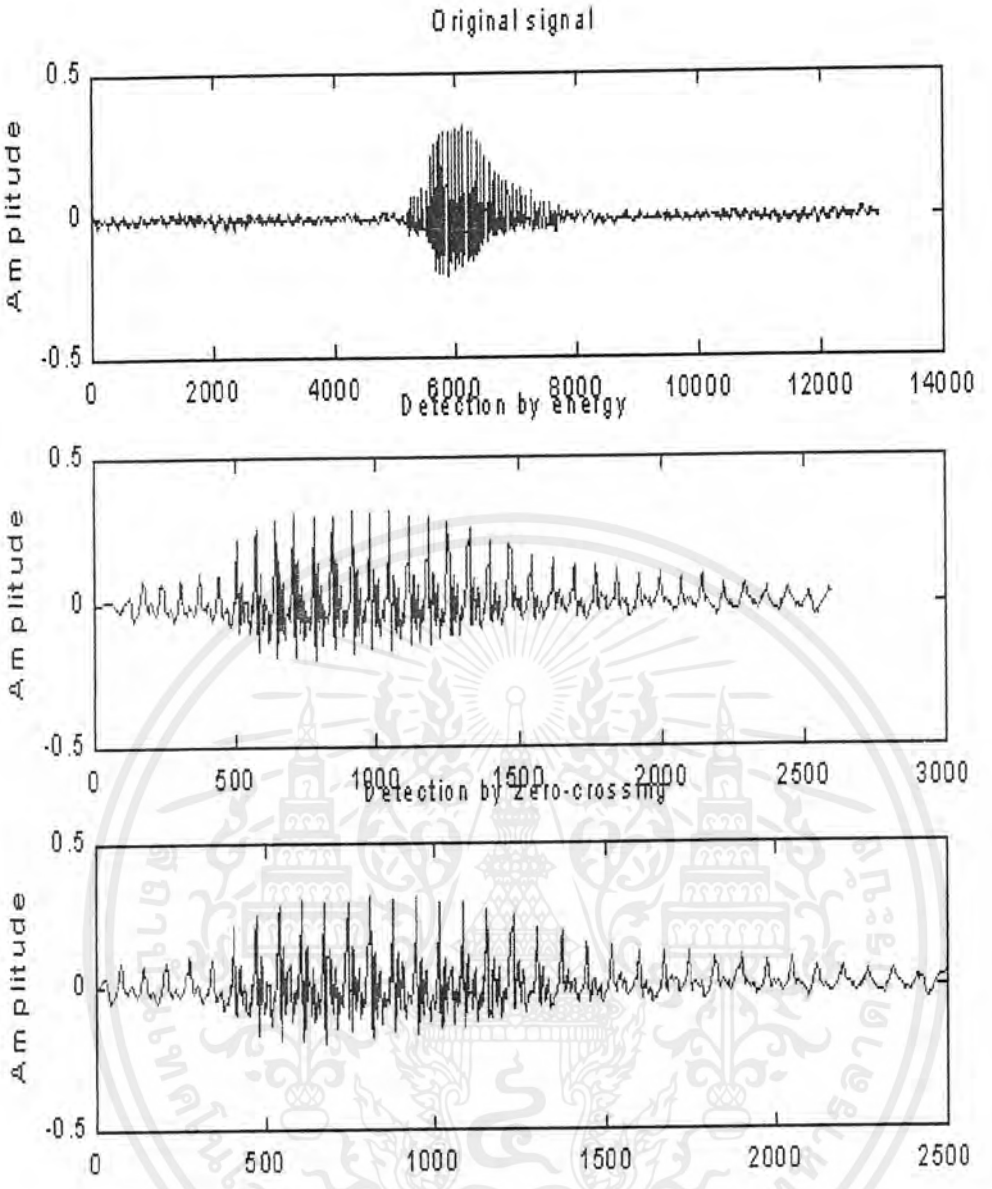
รูปที่ 4.2 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 0

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



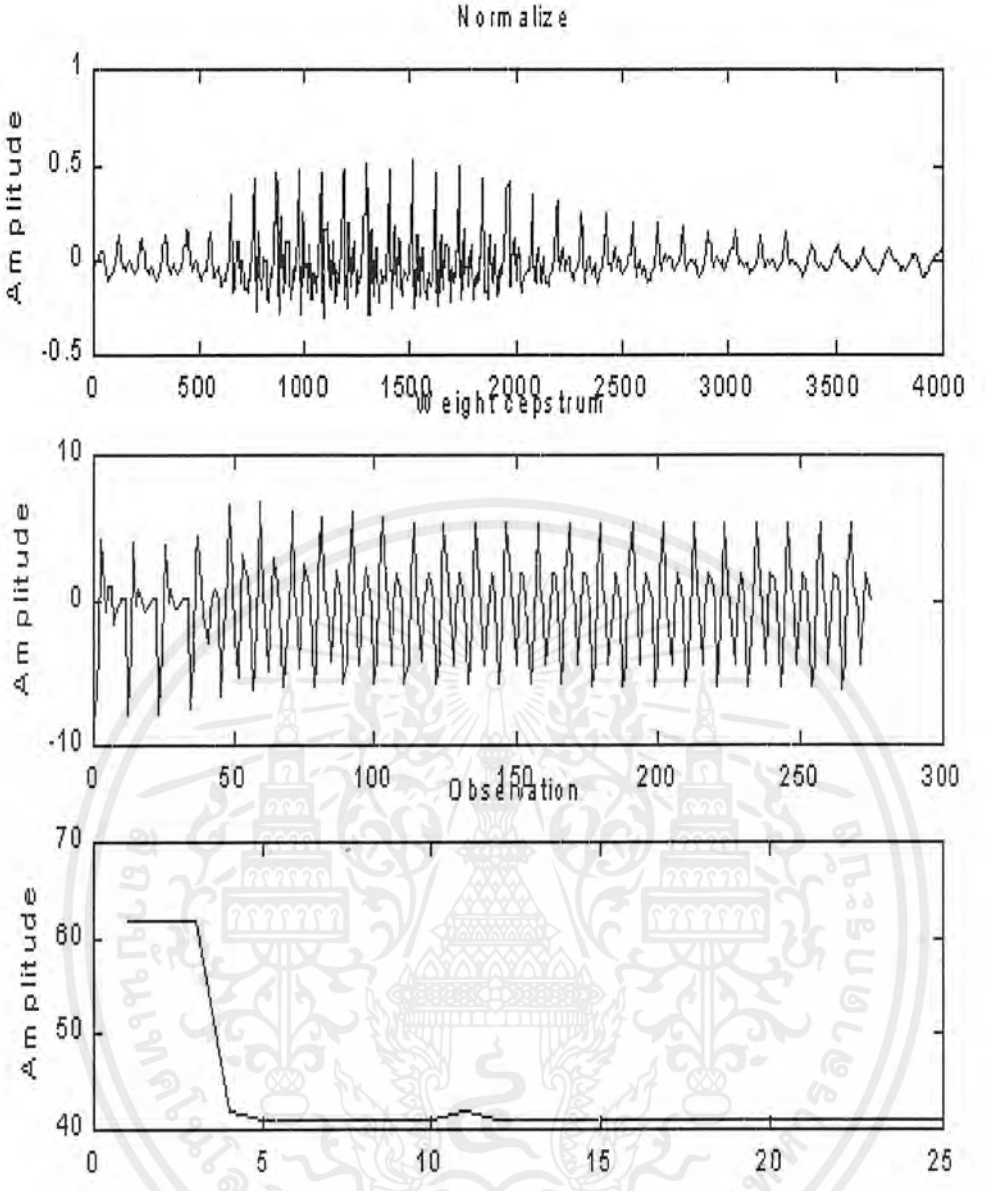
รูปที่ 4.3 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 1

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



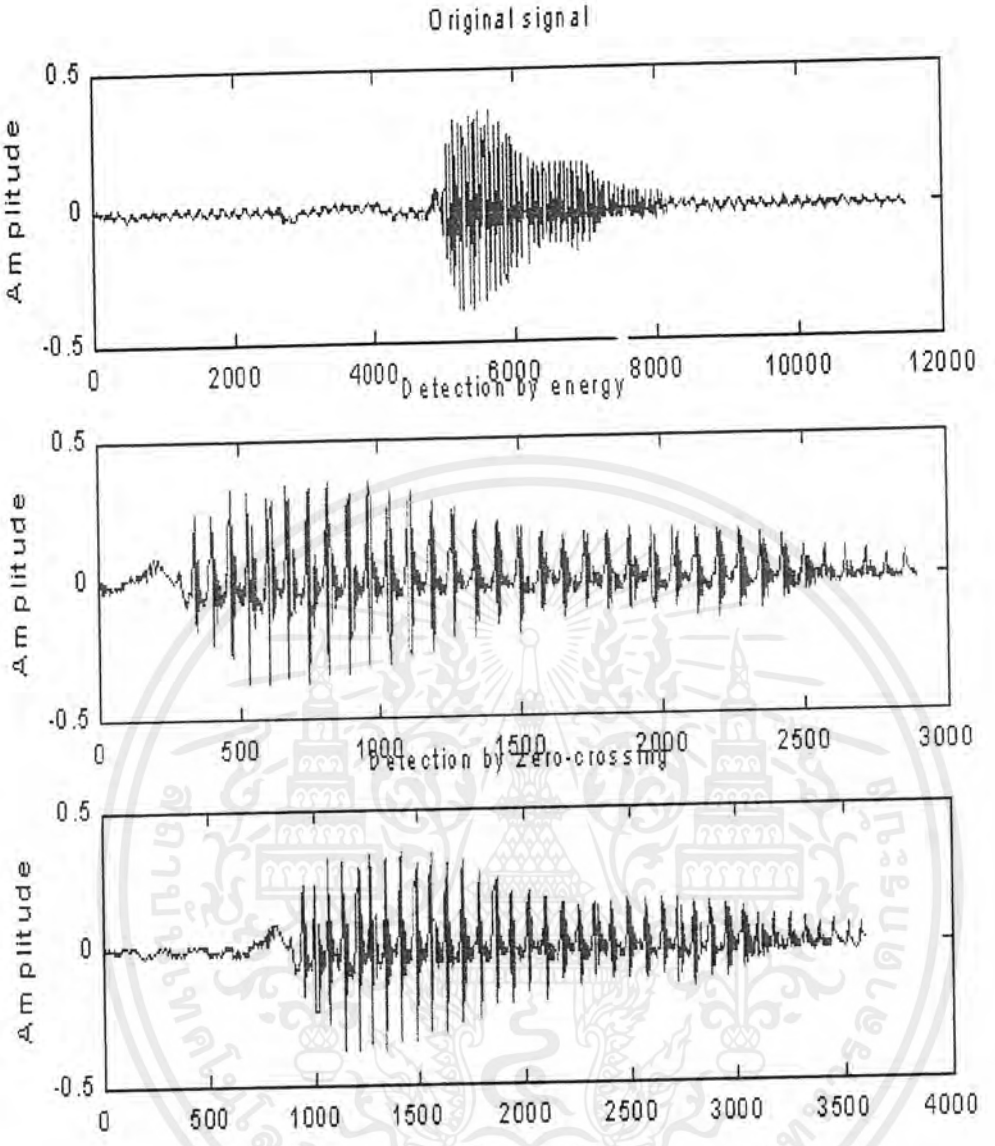
รูปที่ 4.4 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 1

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



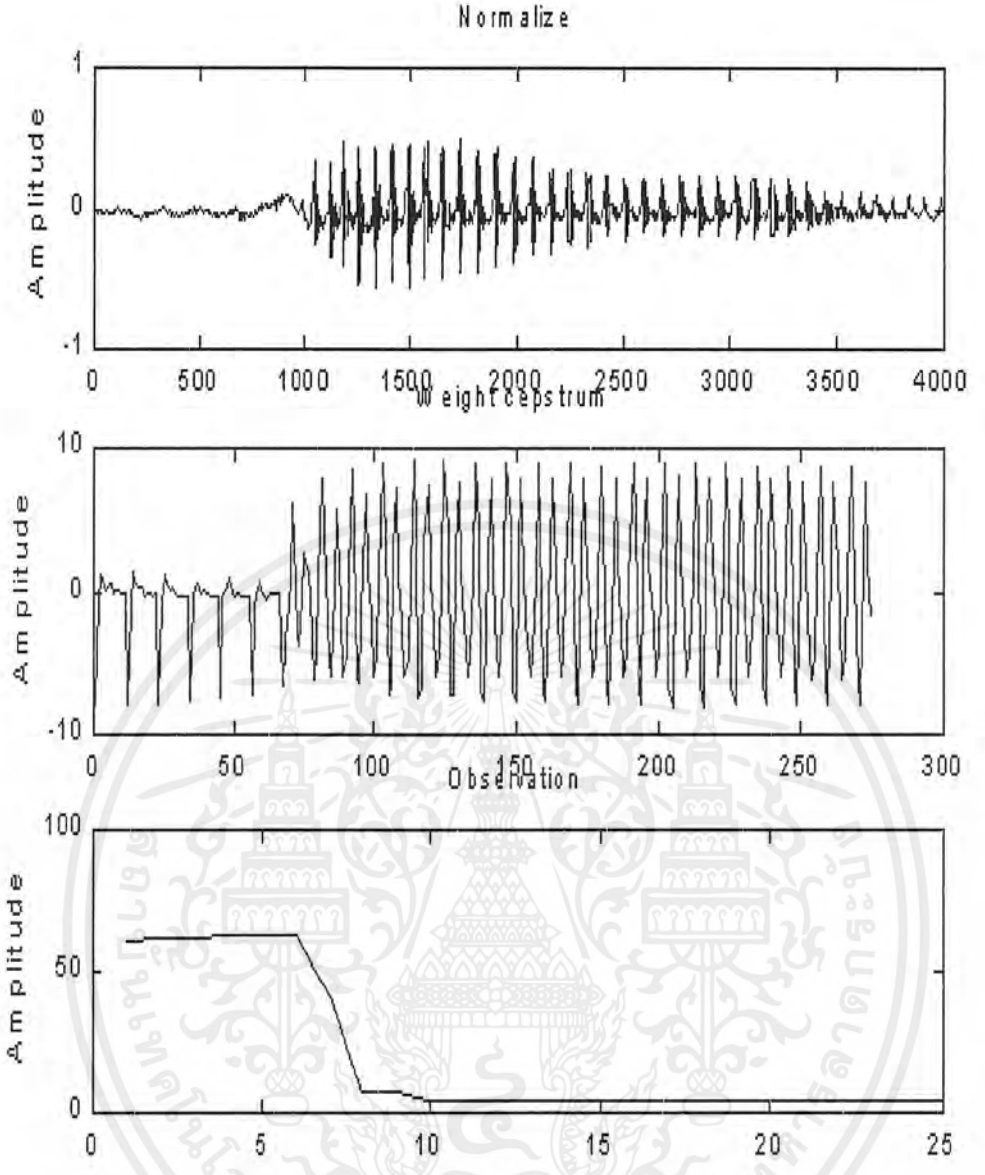
รูปที่ 4.5 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 2

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



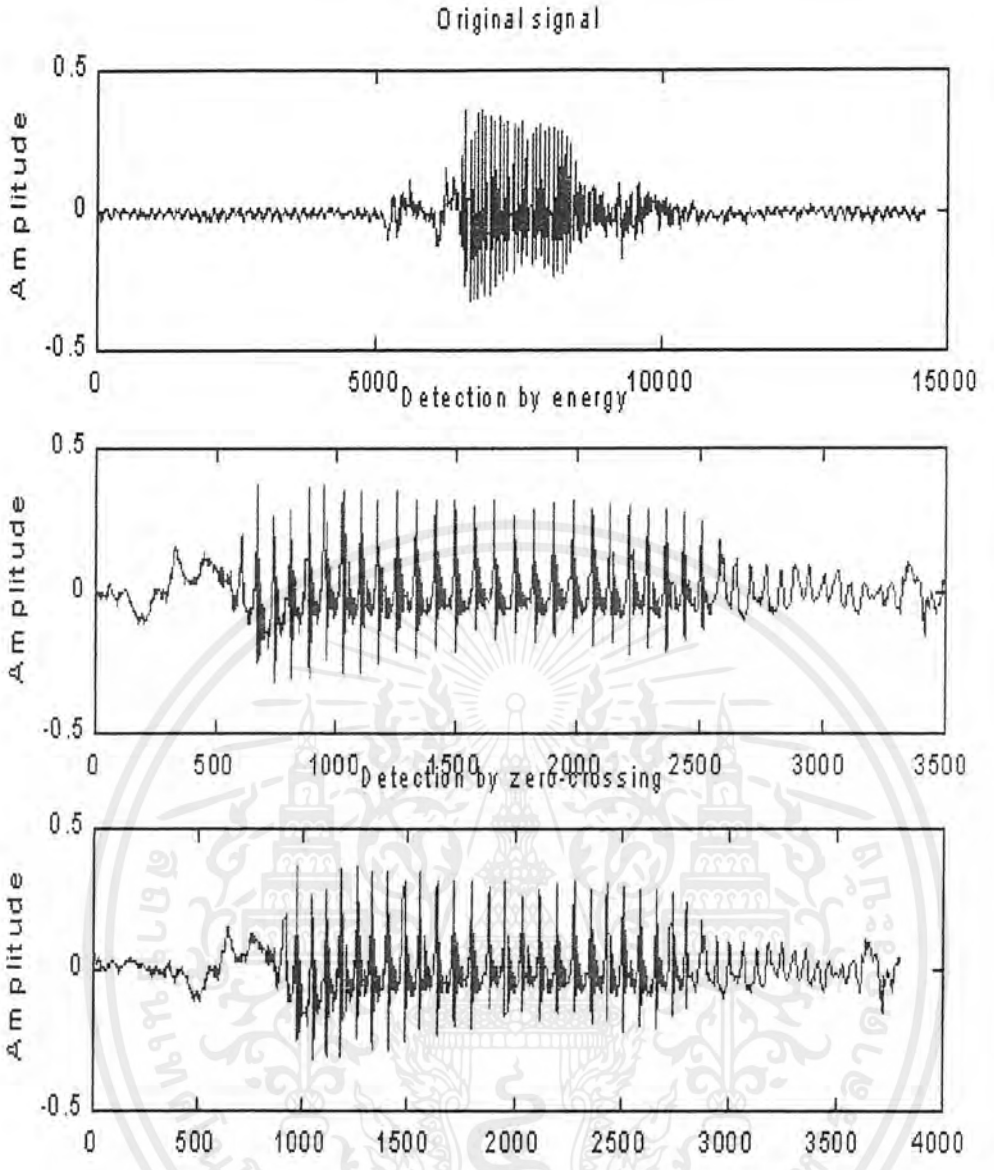
รูปที่ 4.6 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 2

รูปที่หนึ่งแสดงการนำคำมาที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



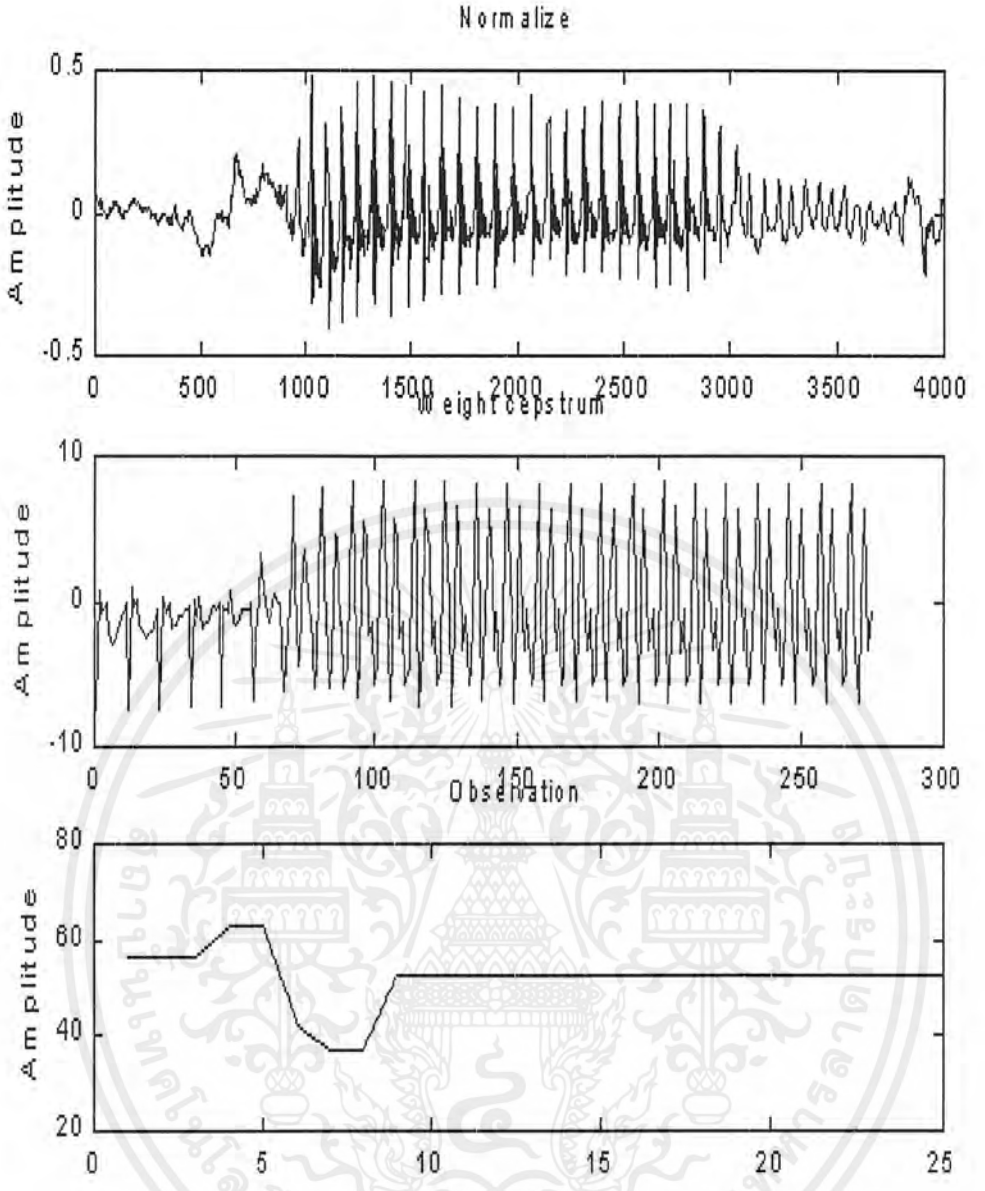
รูปที่ 4.7 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 3

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



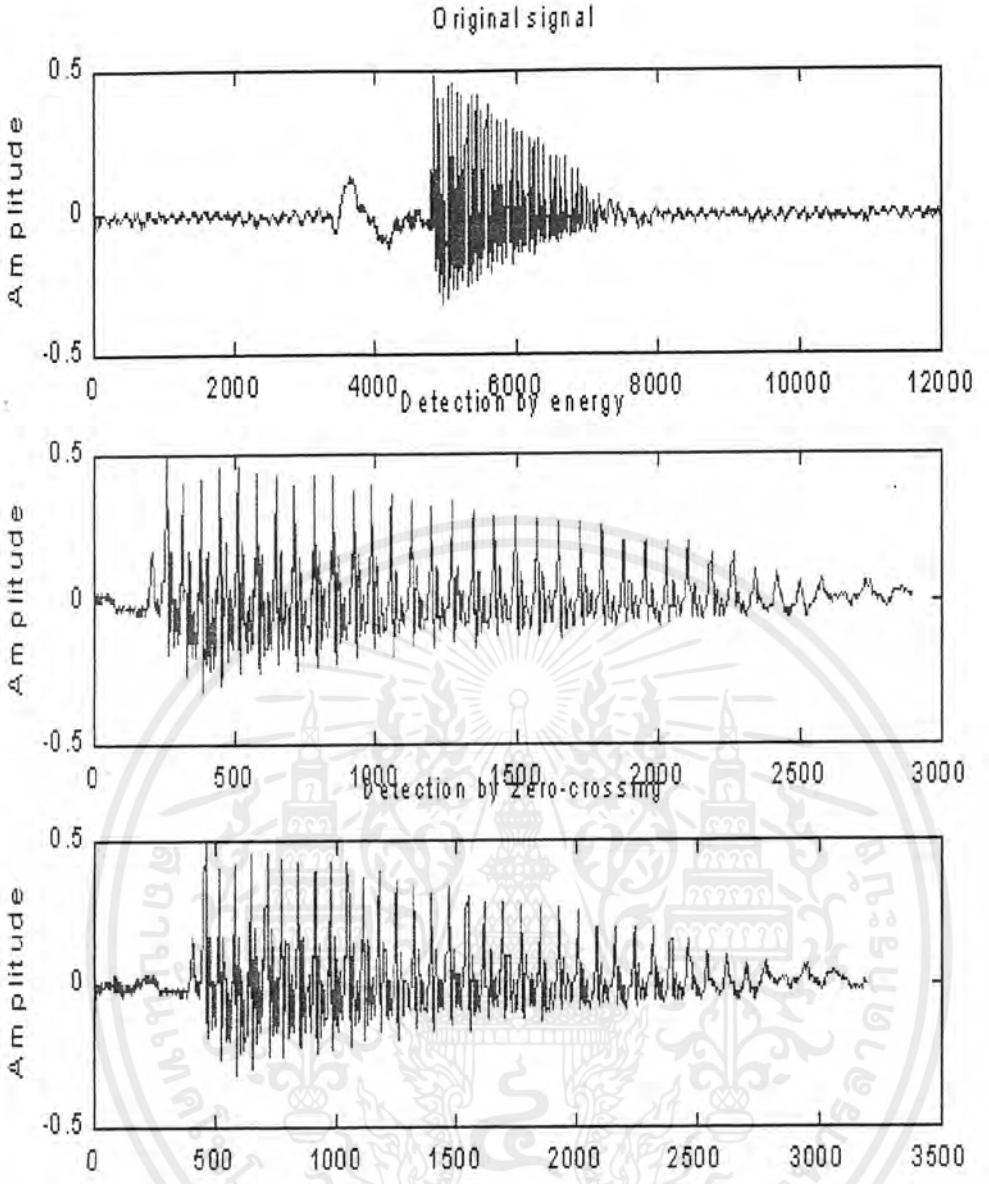
รูปที่ 4.8 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 3

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



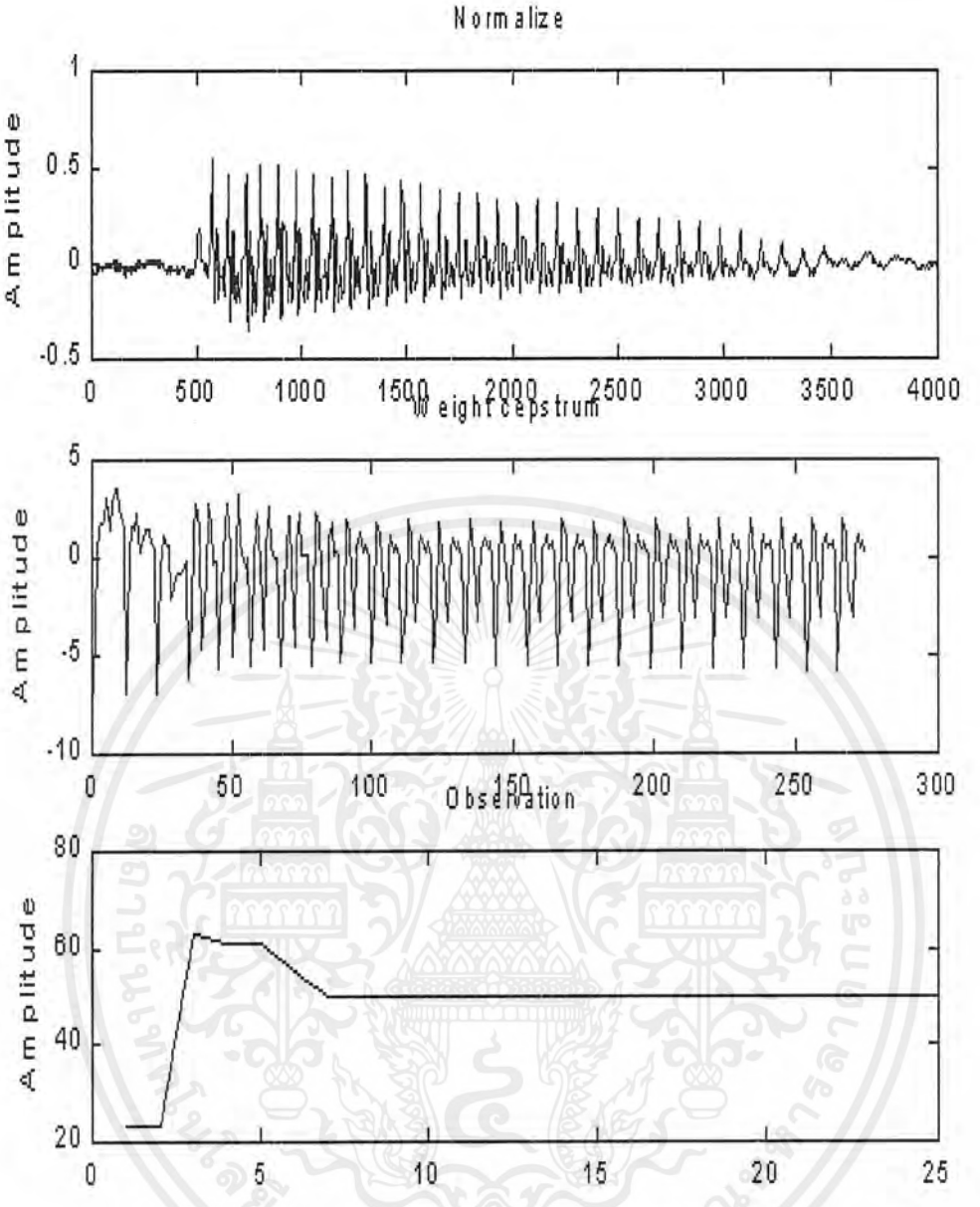
รูปที่ 4.9 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 4

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

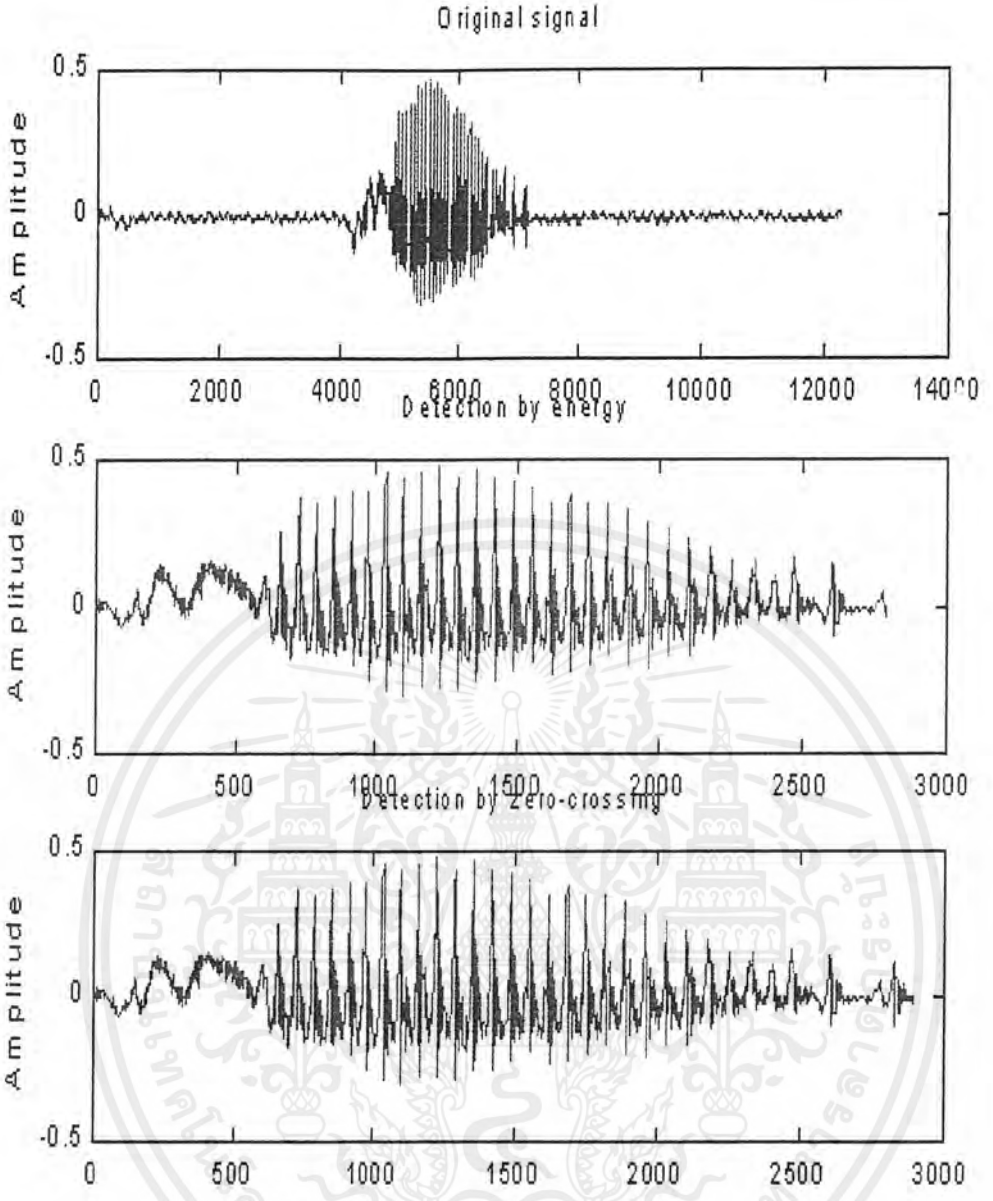
รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 4
รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์
รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม
รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



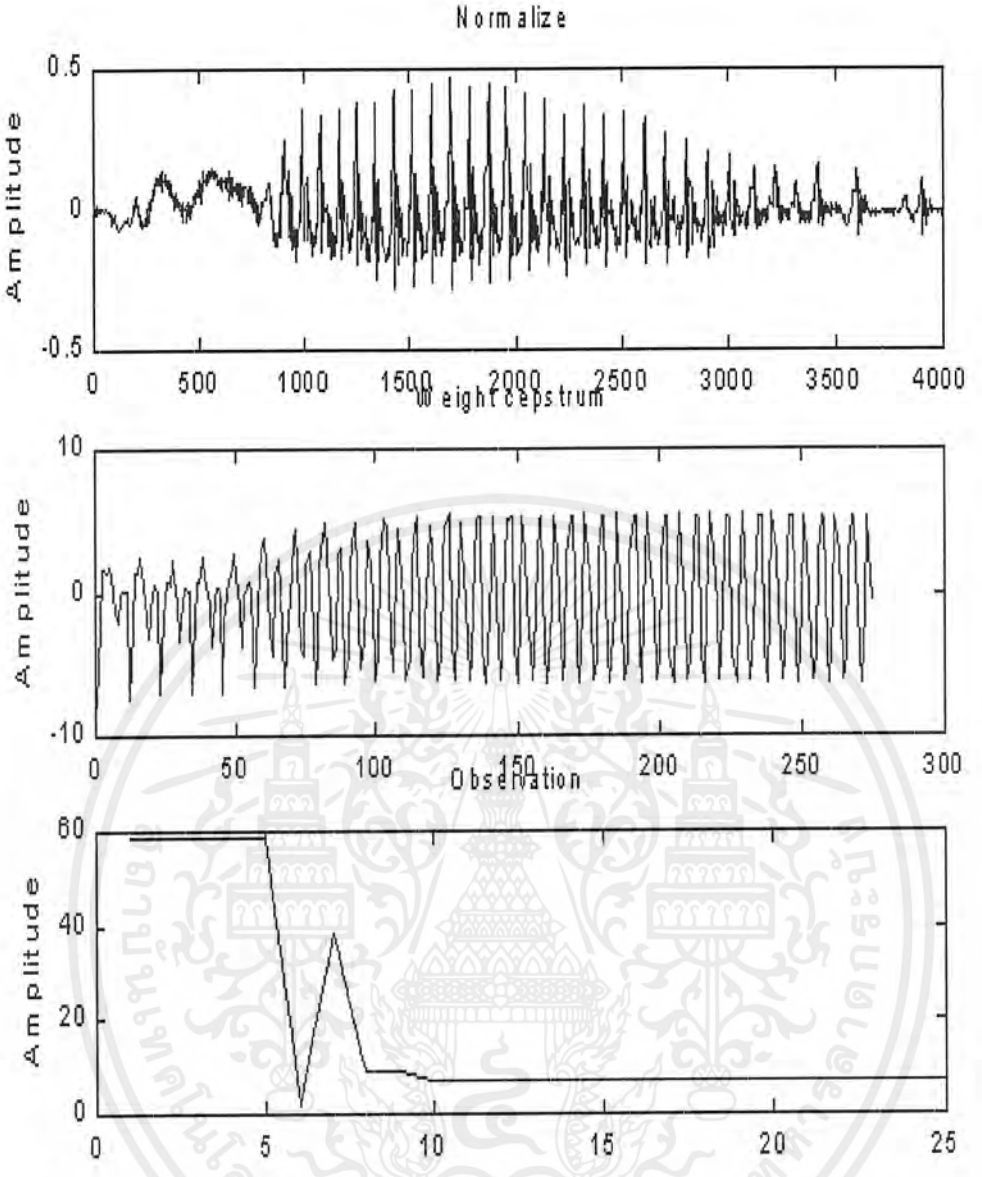
รูปที่ 4.11 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 5

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

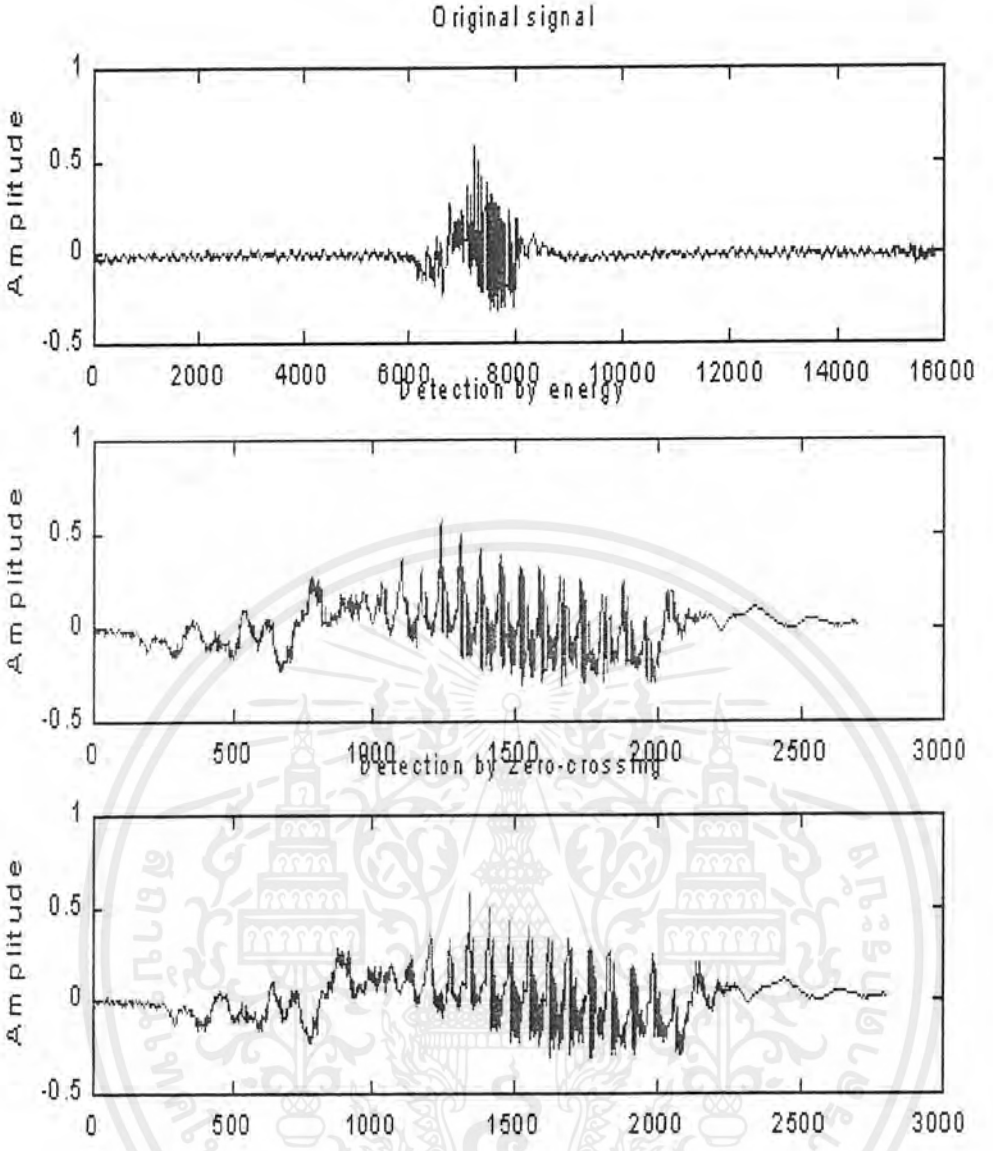
รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 5
รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์
รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม
รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



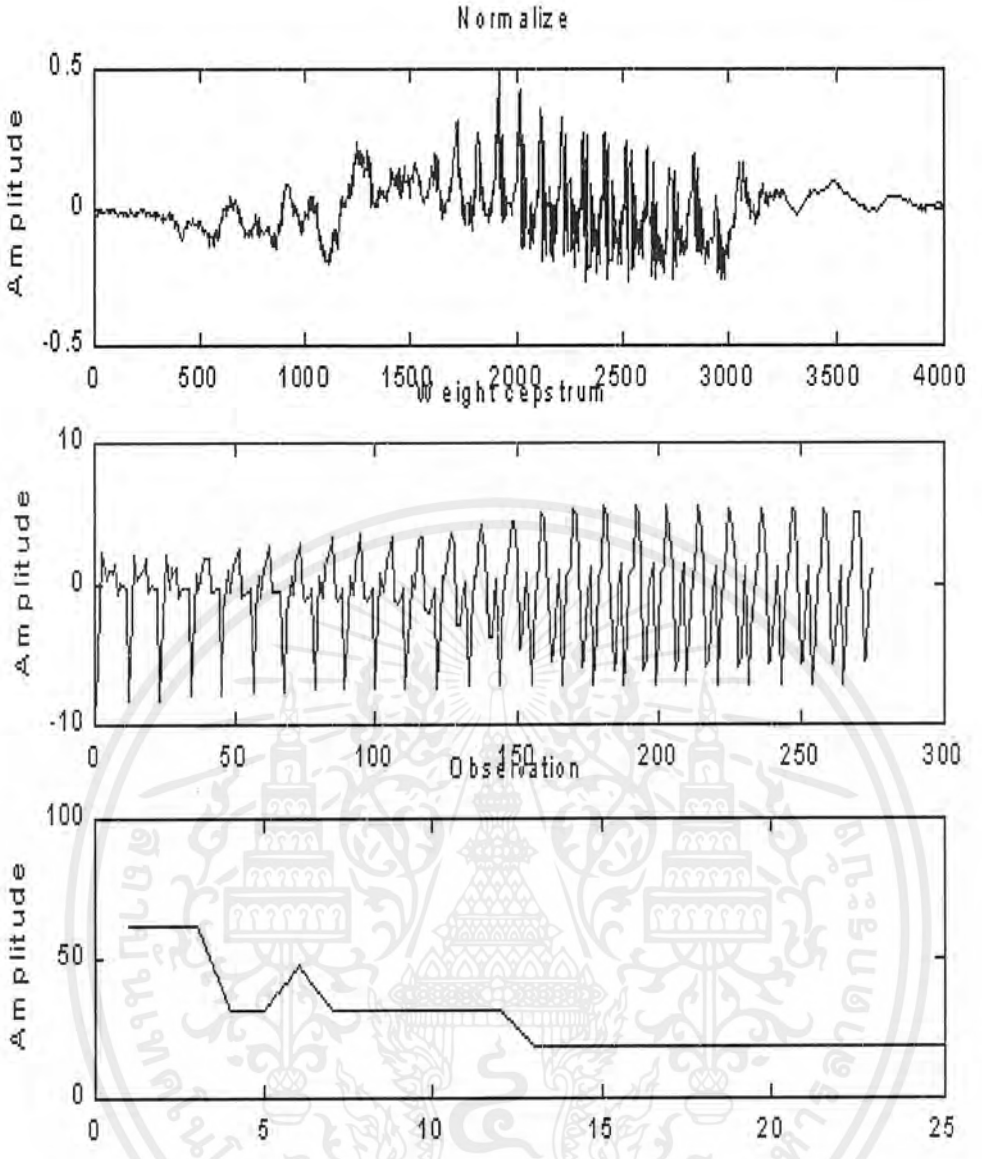
รูปที่ 4.13 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 6

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



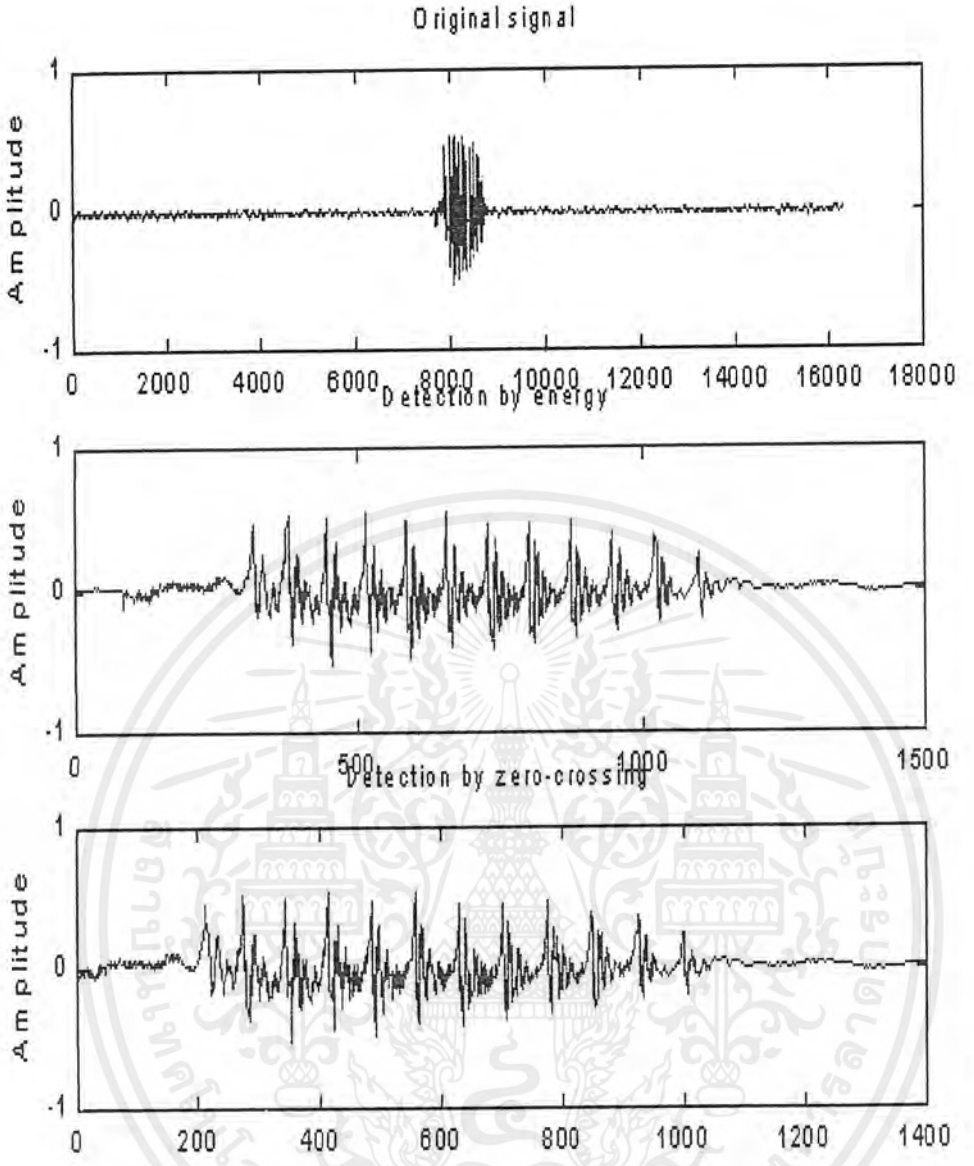
รูปที่ 4.14 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 6

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



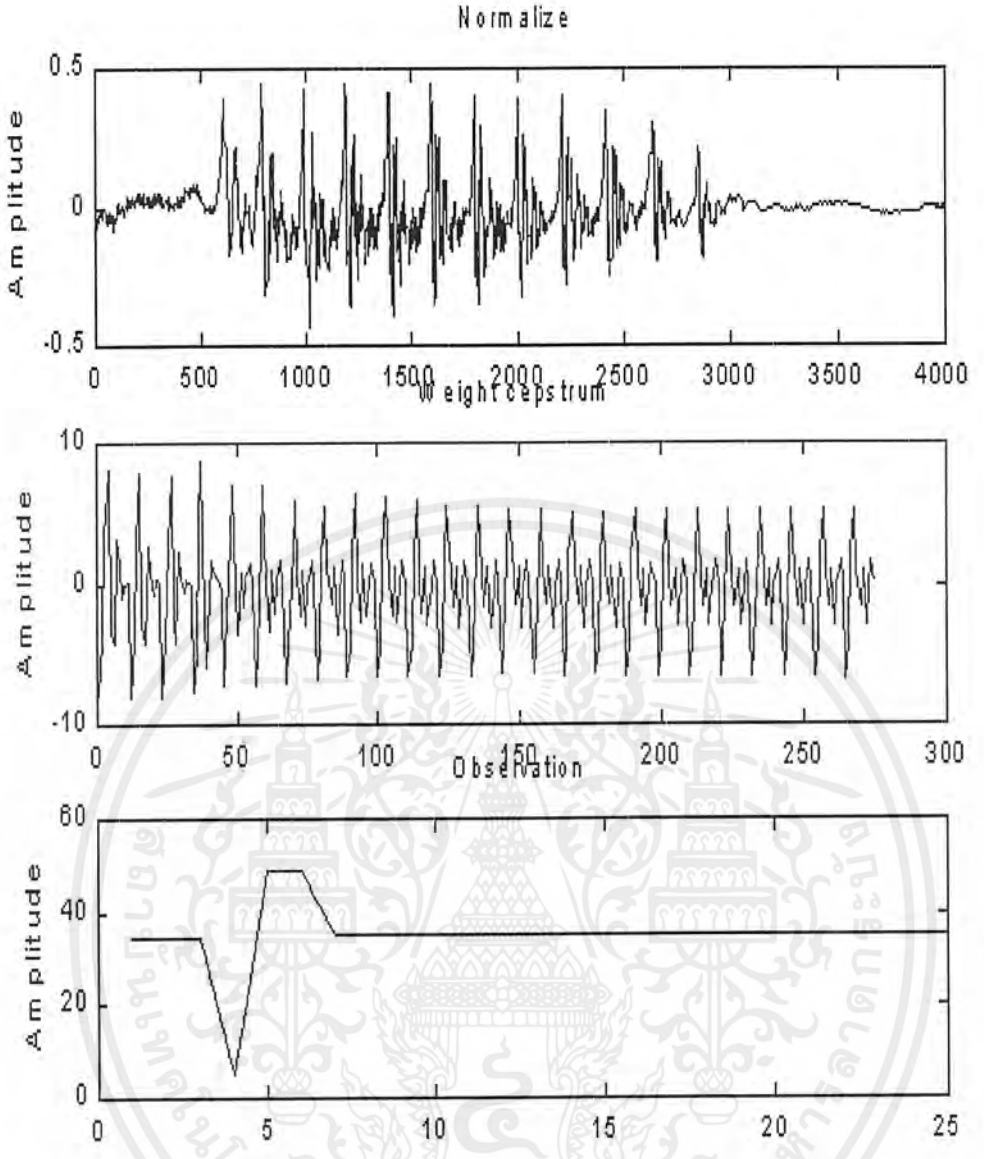
รูปที่ 4.15 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 7

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



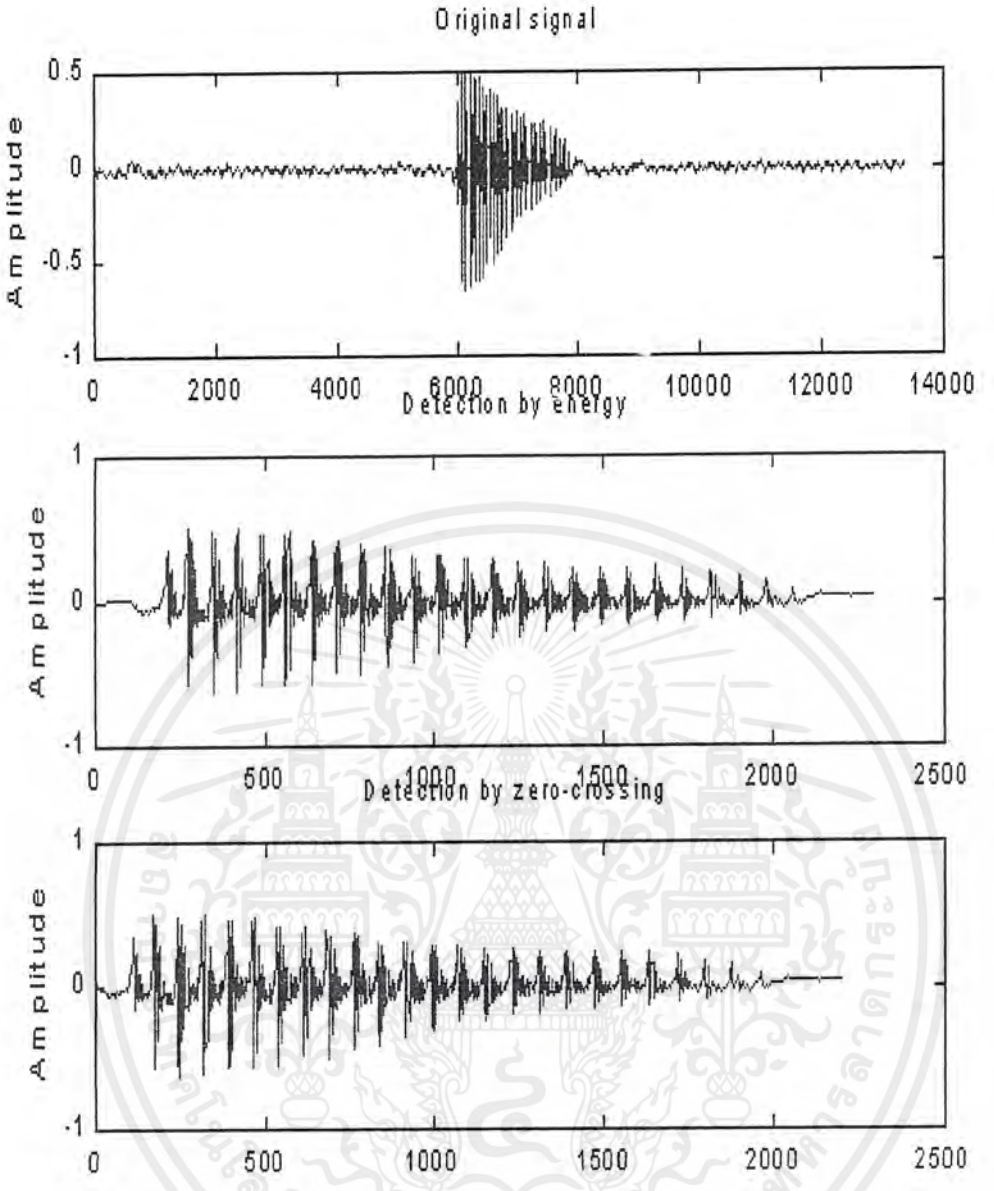
รูปที่ 4.16 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 7

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



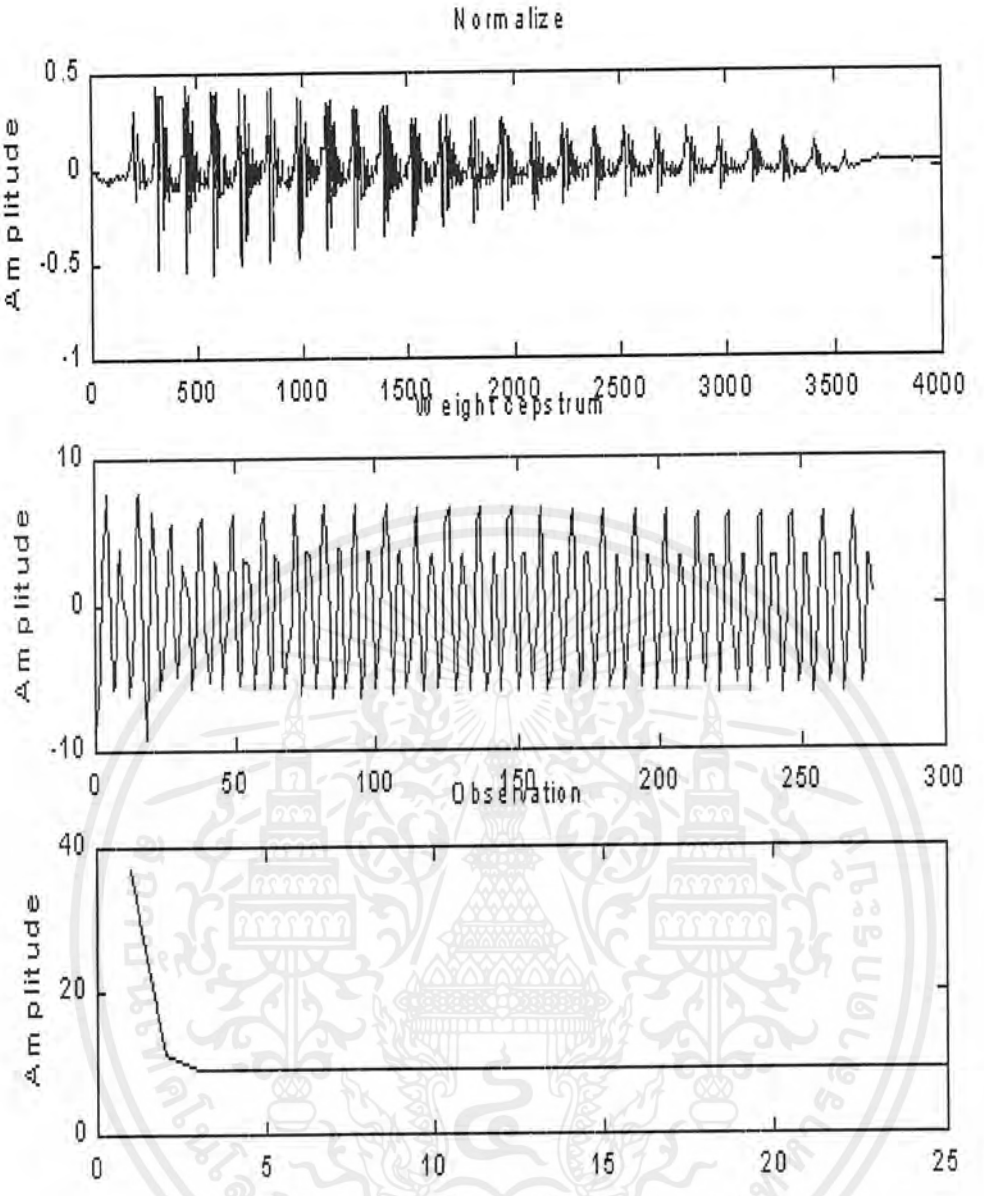
รูปที่ 4.17 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 8

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



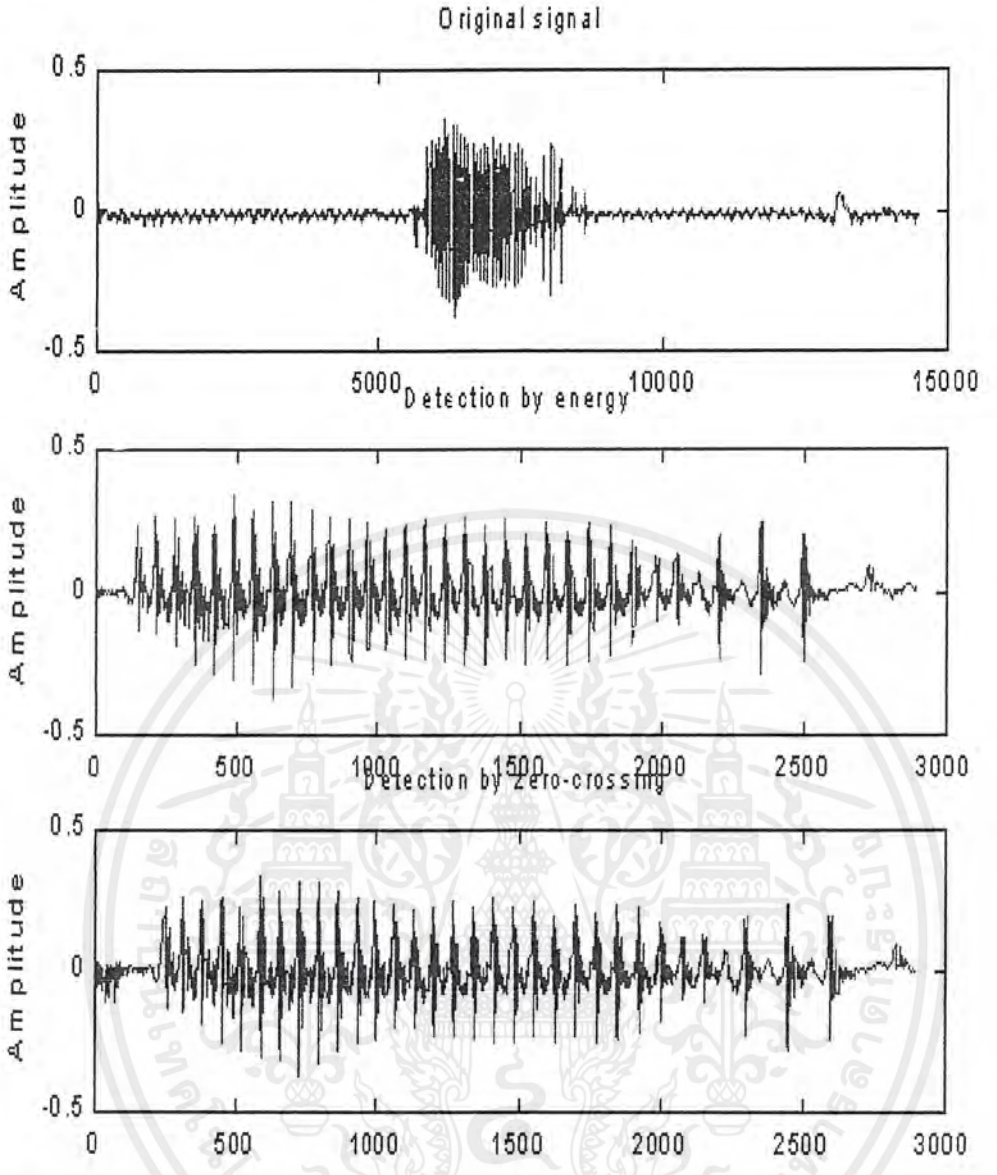
รูปที่ 4.18 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 8

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



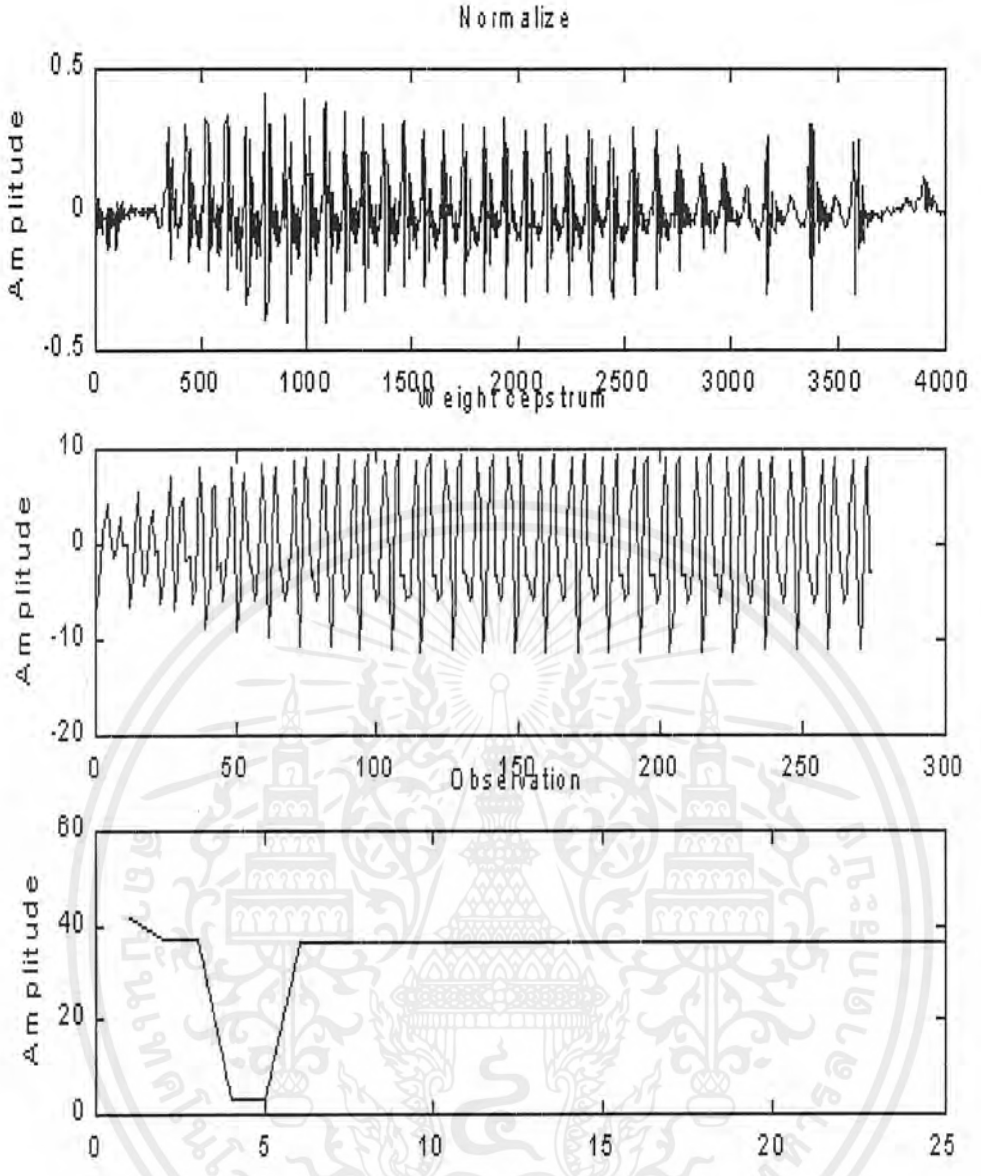
รูปที่ 4.19 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 9

รูปที่หนึ่งแสดงสัญญาณเสียง

รูปที่สองแสดงการตัดคำโดยการหาค่าพลังงาน

รูปที่สามแสดงการตัดคำโดยการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 แสดงรูปซึ่งผ่านขั้นตอนการหาโมเดลเสียง 9

รูปที่หนึ่งแสดงการนำคำที่ตัดได้มาทำการนอร์มอลไลซ์

รูปที่สองแสดงค่าพารามิเตอร์เวทเซปสตรัม

รูปที่สามแสดงค่าพารามิเตอร์ที่ควอนไตซ์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 เมื่อได้โมเดลของแต่ละเสียงแล้วเราก็สามารถนำเสียงมาทดสอบได้ซึ่งในการทดลองเราได้แบ่งโมเดลของเสียงเป็น 2 ชุด คือ

-โมเดลของเสียงที่ใช้เสียงของคน 9 คน พูดคนละ 5 ครั้งเป็นฐานข้อมูล

-โมเดลของเสียงที่ใช้เสียง ของคน 14 คน พูดคนละ 5 ครั้งเป็นฐานข้อมูล

ซึ่งจะได้ผลการทดลองดังนี้

-ส่วนที่ 1 ทดสอบโดยใช้โมเดลของเสียงของคน 9 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

โดยแบ่งการทดลองเป็น

-ทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูล 10 ครั้ง

-ทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้ง

-ทดสอบกับ เสียงที่มีอยู่นอกฐานข้อมูล 10 ครั้ง

-ส่วนที่ 2 ทดสอบโดยใช้โมเดลของเสียงของคน 14 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

โดยแบ่งการทดลองเป็น

-ทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูล 10 ครั้ง

-ทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้ง

-ทดสอบกับ เสียงที่มีอยู่นอกฐานข้อมูล 10 ครั้ง

ได้ผลดังตารางต่อ ไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 แสดงการทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงของคน 9 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

เสียงที่ทดสอบ	ตัวเลขที่รู้จำได้										คิดเป็นร้อยละ
	0	1	2	3	4	5	6	7	8	9	
0	8		1							1	80
1		8				2					80
2	2		7	1							70
3	1			3		3			1	2	30
4					10						100
5				1		7			2		70
6							9			1	90
7							1	7	1	1	70
8		1	1	1					4	3	40
9				1		1		1	1	6	60

ร้อยละความถูกต้อง	
สูงสุด	100
ต่ำสุด	30
เฉลี่ย	65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงการทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้งโดยใช้โมเดลของเสียงของคน 9 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

เสียงที่ทดสอบ	ตัวเลขที่รู้จำได้										คิดเป็นร้อยละ
	0	1	2	3	4	5	6	7	8	9	
0	4		1	1		3			1		40
1		7			2		1				70
2	1	1	3	1						4	30
3				5		2				3	50
4					9					1	90
5				4		3				3	30
6	3						5	2			50
7								6	3	1	60
8					4	1			2	3	20
9				2		2			1	5	50

ร้อยละความถูกต้อง	
สูงสุด	90
ต่ำสุด	20
เฉลี่ย	55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 แสดงการทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูล 10 ครั้งโดยใช้โมเดลของเสียงของคน 14 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

เสียงที่ทดสอบ	ตัวเลขที่รู้จำได้										คิดเป็นร้อยละ
	0	1	2	3	4	5	6	7	8	9	
0	8					1	1				80
1		8			2						80
2		1	7		1		1				70
3			1	6	1			2			60
4					10						100
5				1		7			2		70
6							9	1			90
7						1		7	1	1	70
8				4					4	3	40
9				2		1		1		6	60

ร้อยละความถูกต้อง	
สูงสุด	100
ต่ำสุด	40
เฉลี่ย	70

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 แสดงการทดสอบกับ เสียงที่มีอยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้งโดยใช้โมเดลของเสียงของคน 14 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

เสียงที่ทดสอบ	ตัวเลขที่รู้จำได้										คิดเป็นร้อยละ
	0	1	2	3	4	5	6	7	8	9	
0	4		1	1		3			1		40
1		8			2						80
2		1	3	1		1				4	30
3				5		2				3	50
4					10						100
5				1		3				6	30
6	3						5	2			50
7								9		1	90
8				4		1			2	3	20
9				2		2			1	5	50

ร้อยละความถูกต้อง	
สูงสุด	100
ต่ำสุด	20
เฉลี่ย	60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 แสดงเปอร์เซ็นต์ถูกต้องในการตรวจสอบ เสียงโดยใช้เสียงตรวจสอบเป็นเสียงที่อยู่ในฐานข้อมูล 10 ครั้ง เสียงที่อยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้ง เสียงที่อยู่นอกฐานข้อมูล 10 ครั้ง โดยใช้โมเดลของเสียงของคน 9 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

การทดสอบ	ความถูกต้อง(%)
10 คนเดิม(ไม่พูดใหม่)	65%
10 คนเดิม(พูดใหม่)	55%
10 คนใหม่	40%

ตารางที่ 4.8 แสดงเปอร์เซ็นต์ถูกต้องในการตรวจสอบ เสียงโดยใช้เสียงตรวจสอบเป็นเสียงที่อยู่ในฐานข้อมูล 10 ครั้ง เสียงที่อยู่ในฐานข้อมูลแต่พูดใหม่ 10 ครั้ง เสียงที่อยู่นอกฐานข้อมูล 10 ครั้ง โดยใช้โมเดลของเสียงของคน 14 คน พูดคนละ 5 ครั้ง เป็นฐานข้อมูล

การทดสอบ	ความถูกต้อง(%)
10 คนเดิม(ไม่พูดใหม่)	70%
10 คนเดิม(พูดใหม่)	60%
10 คนใหม่	45%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

-แสดงจำนวนตัวอย่างของสัญญาณที่ผ่านการหาขอบเขตของคำ

-แสดงจำนวนออเดอร์ของสัมประสิทธิ์แอลพีซี

-แสดงจำนวนออเดอร์ของสัมประสิทธิ์เซปสตรัม

-แสดงจำนวนพารามิเตอร์ที่สำคัญที่นำมาใช้ในการทดสอบความใกล้เคียงกับรูปแบบอ้างอิง

-ปุ่ม PLOT เป็นปุ่มที่ใช้ในการแสดงสัญญาณที่ผ่านการหาขอบเขตของคำและการนอมอลไลซ์และการควอนไตซ์

-ปุ่ม SOUND เป็นปุ่มที่ใช้ในการแสดงผลว่าเสียงที่เราเข้าไปทดสอบนั้นเป็นสัญญาณเสียงใด

-หน้าต่าง File... เป็นหน้าต่างที่ใช้แสดงชื่อไฟล์ที่เปิดสำเร็จ

-หน้าต่าง Output เป็นหน้าต่างแสดงผลที่ได้จากการจดจำเสียง

-ปุ่ม Help เป็นปุ่มที่ใช้แสดงความช่วยเหลือในการใช้โปรแกรม

-ปุ่ม Exit เป็นปุ่มที่ใช้เมื่อต้องการเลิกใช้โปรแกรมการจดจำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทวิจารณ์และบทสรุป

บทสรุป

จากการทดลองการจดจำของเสียง 0 ถึง 9 โดยการใช้สัญญาณเสียงที่สุ่มด้วยความถี่ 8 กิโลเฮริตซ์ เข้ารหัสแบบ 8 บิท ซึ่งใช้แบบอ้างอิงต่าง ๆ โดยใช้ผู้พูดจำนวน 9 คน และ 14 คน คนละ 5 เสียงในการสร้างแบบอ้างอิง

ผลการทดสอบการจดจำโดยทำการทดสอบกับกลุ่มคนที่นำมาสร้างแบบอ้างอิงแบบไม่พูดใหม่และแบบพูดใหม่ และกลุ่มที่ไม่ใช่เป็นกลุ่มที่นำมาสร้างแบบอ้างอิง

จากผลการทดลองสรุปได้ดังนี้

1) เสียงพูดของคนคนเดียวกันเมื่อให้พูดแต่ละครั้งจะพบว่าเสียงที่พูดไม่เหมือนกัน

2) การหาขอบเขตของคำมีผลต่อความถูกต้องของการจดจำเสียง การหาขอบเขตของคำที่ถูกต้องแน่นอน จะเพิ่มอัตราการจดจำเสียงให้สูงขึ้น

3) เนื่องจากสัญญาณที่ผ่านการหาขอบเขตของคำมีความยาวของเสียงที่แตกต่างกันทำให้การหาค่าเฉลี่ยของสัญญาณที่นำมาสร้างแบบอ้างอิงทำได้ไม่แน่นอน เพื่อความถูกต้องของการสร้างแบบอ้างอิง จึงทำการปรับสัญญาณก่อนที่จะมาทำการหาค่าพารามิเตอร์ (นอร์มอลไลซ์) ให้มีความยาวเท่ากันเสียก่อน

4) การหาค่าพารามิเตอร์แบบต่าง ๆ ในการทดลองนี้ใช้ค่าพารามิเตอร์เวทเซปสตรัมซึ่งจะให้ผลดีที่สุด

5) จำนวนคนที่ใช้มาทำแบบอ้างอิงจะมีผลต่อการจดจำอย่างมากโดยจำนวนคนยิ่งมาก จะให้ผลต่อการจดจำถูกต้องมากขึ้น

6) ถ้าใช้ผู้ที่มีเสียงในแบบอ้างอิงมาทดสอบจะมีความถูกต้องมากกว่าการใช้เสียงของผู้ที่ไม่มีเสียงในแบบอ้างอิง

7) การทดสอบการจดจำโดยใช้วิธีฮิดเดนมาร์คอฟจะทำการเปรียบเทียบกับค่าโมเดลซึ่งเป็นมาตรฐานเสียง 0 ถึง 9 เท่านั้นดังนั้นทำให้เวลาที่ใช้ในการคำนวณน้อยลงกว่าวิธีอื่น

บทวิจารณ์และแนวทางพัฒนา

1) การทดสอบนี้ได้ทดสอบเฉพาะเสียงผู้ชายเท่านั้น เนื่องจากช่วงความถี่ของเสียงผู้ชายและเสียงผู้หญิงจะเหลื่อมกันช่วงหนึ่ง จึงควรพัฒนาต่อไปให้สามารถจดจำได้ทั้งเสียงผู้ชายและผู้หญิง

2) เนื่องจากการทดสอบที่ใช้นี้เป็นการทดสอบให้รู้วิธีการที่สามารถจะนำมาใช้ในการจดจำเสียง จึงได้ทำการเขียนโปรแกรมบน MATLAB version 5.3 ถ้าต้องการให้เวลาคำนวณน้อยกว่านี้ควรใช้โปรแกรมอื่นๆ แทน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การหาค่าพารามิเตอร์ของสัญญาณเสียงนั้นได้ทำการหาในโดเมนของเวลาเนื่องจากวิธีการไม่ยุ่งยากเมื่อเทียบกับการหาค่าพารามิเตอร์ในโดเมนของความถี่ซึ่งมีวิธีการซับซ้อนกว่าแต่มีความถูกต้องมากกว่า ดังนั้นจึงควรพัฒนาโดยการหาค่าพารามิเตอร์ในโดเมนของความถี่เพื่อการจดจำที่ดีขึ้น

4) ปัจจุบันได้มีการพัฒนาวิธีการเปรียบเทียบความคล้ายคลึงของสัญญาณ เช่น นิวรอลเน็ตเวิร์ค ซึ่งจะเหมาะสมกับการหาพารามิเตอร์ในโดเมนความถี่ทำให้เราสามารถเลือกใช้วิธีต่างๆ เหล่านี้เพื่อใช้ในการจดจำได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการหาขอบเขตของคำ

-ฟังก์ชัน detect1 : การเปรียบเทียบโดยการหาค่าพลังงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function   [E,a1,b1,ref,s]=detect1(filename);

% [E,a1,b1,ref,s]=detect1(filename);
% this function for find endpoint detection by energy
% when filename : file.wav
% E:energy each frame
% s:signal when pass detection by energy

[inpt1,fs]=wavread(filename);
inpt=inpt1-mean(inpt1);
%number of frame
sample = 100;
vnoloop = floor(length(inpt)/sample);
%find energy each of frame
for k=1:vnoloop
    sum=0;
    for k1=1:sample
       sum = sum + abs (inpt( (k-1)*sample+k1 ) );
    end
    E(k)=sum;
end
% find maximum energy
Emax=max(E);
%set value
a=1/5;
%begin process
ref=a*Emax;
for k=1:vnoloop
    if E(k)==Emax
       fixenergy=k;
       break;
    end
end

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

-ฟังก์ชัน zrocrs1 : การเปรียบเทียบโดยการหาอัตราการตัดศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function [S,z,a2,b2,Zmax]=zrocrs1(filename,a1,b1);

%[S,z,a2,b2,Zmax]=zrocrs1(filename,a1,b1);
%when filename:file.wav
%a1,b1:initial and finish point at detect by energy
%S:signal when pass detection by zerocrossing

[inpt1,fs]=wavread(filename);
inpt=inpt1-mean(inpt1);
%find number of frame
sample=100;
vnoloop1=floor(length(inpt)/sample);
%find zero_crossing each of frame .
for k=1:vnoloop1
   sum=0;
   for k1=1:sample
      if k1==1
         sum = sum+abs( sg_n( inpt((k-1)*sample+k1) ) );
      else
         sum = sum+abs( sg_n( inpt((k-1)*sample+k1) ) -...
            sg_n( inpt((k-1)*sample+(k1-1) ) ) );
      end
   end
   z(k)= sum/(2*sample);
end
% find maximum zero_crossing
Zmax=max(z);
a=1/5;
%find position of maximum zero_crossing
for k=1:vnoloop1
   if z(k)==Zmax
      fixzero=k;
      break;
   end
end
refzro=a*Zmax;
%process find bound signal
for k1= a1:-1:1
   if z(k1)<=refzro
      a2= k1;
      break;
   end
end

for k2=b1:vnoloop1
   if z(k2) <= refzro
      b2=k2;
      break;
   end
end

%plot signal
n=0;

if isempty(b2)==1|b2==vnoloop1|isempty(a2)==1;
   if isempty(a2)==1;
      a2=a1-1;
      b2=vnoloop1-1;
       for k3=a2:b2
          for k4=1:sample
             n=n+1;
             S(n)=inpt( (k3*sample)+k4 );
          end
         end
      else
         b2=vnoloop1-1;
       for k3=a2:b2
          for k4=1:sample
             n=n+1;
             S(n)=inpt( (k3*sample)+k4 );
          end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        end
      end
else
    for k5= a2:b2
    for k6=1:sample
       n=n+1;
       S(n)= inpt( (k5*sample)+k6 );
    end
    end
end
```

```
%*****************************************************
%*************** function sign ***********************
%*****************************************************

function zerc=sg_n(s);
%zerc=sg_n(s);
%s is signal when thoung energy
for i=1:length(s)
    if s(i) >0
    zerc=1;
  elseif s(i)==0
     zerc=0;
  else zerc=-1;
    end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการนอร์มอลไลซ์สัญญาณ

-ฟังก์ชัน normalize : การนอร์มอลไลซ์สัญญาณให้ได้จำนวนตัวอย่าง 4000 ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function normS = normalize(S,normsize);

%normS = normalize(S,normsize);
%**************** input ************************
%normsize = 4000;
%when S =signal when zerocrossing;
%normsize = number of  normalize;
%************** output ************************
%     normS = signal when pass normalize;
%*********************************************
normS(1) =S(1);
timestep = (length(S))/(normsize);

time = 1;
j=2;
while j < normsize
   time = time + timestep;
   lowtime = floor(time);
   residue = time-lowtime;
   if (lowtime+1) > length(S)
      S(lowtime+1)=0;
   end

   normS(j) = S(lowtime)+residue*(S(lowtime+1)-S(lowtime));
   j=j+1;
end;
normS(j)=S(length(S));
%normS(normsize) = S(length(S));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการหาค่าพารามิเตอร์

-ฟังก์ชัน levin_Durbin : การหาค่าสัมประสิทธิ์การทำนายเชิงเส้น

-ฟังก์ชัน ceps : การหาค่าสัมประสิทธิ์เซปสตรัม

-ฟังก์ชัน cepsw : การหาค่าสัมประสิทธิ์เซปสตรัมถ่วงน้ำหนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function  [b,c,w,G]=levin_Durbin(S,N,Q);
%[b,c,w,G]=levin_Durbin(S,N,Q)'
%****************** Input ********************
% when S =   speech waveform
%      N =   Order of lpc11
%      Q =  Order of CEPSTRUM
%***************** Output ********************
% when b = lpc11 Parameter
%       C = Cepstrum Parameter
%       w = Weight cepstrum Parameter
%       G = Gain

%*********************************************
fs=8000;
sizf=length(S);
%*********************************************
% initial Pre emphasis **************************

format long;
a1=0.9375;as=0; df=0;gh=0;
for i=1:sizf
   if i==1
      s(i)=S(i);
   else s(i)=S(i)-a1*S(i-1);
   end
end

% end pre-emphasis ****************************

% initial divide frame ************************

fram=(240/fs); %0.03
t=(160/fs);     %0.02
shift=(80/fs); %0.01
z=length(s)/160;
le=length(s)-((floor(z)-1)*160);
if le==fram*fs %(fram*fs=240)
   frame=floor(z);
else
   frame=floor(z)-1;
end

% end divide frame ****************************

% initial windowing ***************************

for l=1:frame
   v=0;
         for n=(((l-1)*160)+1):(((l-1)*160)+240)
            v=v+1;
            as=as+1;
               x(l,v)=s(n)*(0.54-(0.46*cos(2*pi*v/(fram*fs))));
            h(as)=x(l,v);
         end

% end windowing *******************************************

% initial Auto correlation & lpc11 analysis ********************
          [a,g,R]=lpc11(h,N); % N : Order of lpc11
           for j=1:N+1
              df=df+1;
              b(df)=a(j);
      %        r(l,1)=R(1);
              G(l)=g;
           end

         [ce,cw]=ceps(a,g,N,Q);
           for i1=1:Q
           gh=gh+1;
              c(gh)=ce(i1);
           w(gh)=cw(i1);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            end
 end

% end Auto correlation & lpc11 analysis ************************

%************************ when final frame not 240 *************
% initial Auto correlation & lpc11 analysis ******************

 if le < 240
   v=0;
   for n=(frame*160)+1:(frame*160)+le
         v=v+1;as=as+1;
         x(frame+1,v)=s(n)*(0.54-(0.46*cos(2*pi*v/le)));
         h(as)=x(frame+1,v);
   end
   frame=frame+1;
   l=l+1;

% end windowing ******************************

% initial Auto correlation & lpc11 analysis ******************
   [a,g,R]=lpc11(h,N);
   for j=1:N+1
       %A(l,j)=a(j);
       df=df+1;
       b(df)=a(j);
       r(l,1)=R(1);
       G(l)=g;
    end
    [ce,cw]=ceps(a,g,N,Q);
    for i=1:Q
       gh=gh+1;
       c(gh)=ce(i);
       w(gh)=cw(i);
    end
 end

% ************************************************************
% ************************************************************
% ************************************************************
% ************************ lpc Function**********************
% ************************************************************
% ************************************************************
% ************************************************************

 function [a,g,R]=lpc11(h,N);
    error(nargchk(1,2,nargin))
   h =h;
    [r,c] = size(h);
    if (c>1)&(r==1)
        h = h(:);
    end
    numsigs = size(h,2);
    if nargin<2, N = size(h,1)-1; end
    if (N>size(h,1)-1),
        % disp('Warning: zero-padding short input sequence')
        h(N+1,:)=zeros(1,numsigs);
    end

    R = flipud(fftfilt(h,conj(flipud(h))))/length(h);
    % R is the autocorrelation vector,
    % Equivalent code (for the single signal case):
    %   R = xcorr(h,'biased');
    %   M = length(h);
    %   R(1:M-1) = [];

    a = levinson(R,N);
    g = real(sqrt( sum((a.').*R(1:N+1,:)) ));

% ************************************************************
% ************************************************************
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% ***************************************************************
% ******************Cepstrum Function*****************************
% ***************************************************************
% ***************************************************************
% ***************************************************************

function  [ce,cw]=ceps(a,g,N,Q);

ce(1) = 0.5*log(g);
ce(2) = a(2);
for i = N+2:Q;
   a(i) = 0;
end
for m = 3 :Q
   sum =0;
   for k = 1:m-1
      sum =sum+(k*ce(k)*a(m-k));
   end
   cep_sum(m) = (1/m)*sum;

     if m <=N+1
      ce(m) = a(m)+cep_sum(m);
     else ce(m) = cep_sum(m);
     end
   end
  cw = cepsw(Q,ce);

% ***************************************************************

% ***************************************************************
% ***************************************************************
% ***************************************************************
% ******************Weihgt Cepstrum Function*********************
% ***************************************************************
% ***************************************************************
% ***************************************************************

function cw = cepsw(Q,ce);
for m = 1:Q
   cw(m) = ce(m)*(1+( (Q/2) * sin (pi* (m/Q) ) ));
end

% ***************************************************************
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการหาโค้ดบุค

-ฟังก์ชัน codebookM : การหาโค้ดบุค 64(Main)

-ฟังก์ชัน codebookS : การหาโค้ดบุค 64(Sub)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% find codebook 64
load m
% m:file.mat is save weigth cepstrum
a=x;
i=650;
k=1;
for n=1:64
   b(k:k+10)=a(i:i+10);
   k=k+11;
   i=i+11;
end
[r]=codebookS(a,b);
z=b-r;
z=z.^2;
v=sum(z)/length(z);
round=1
while v>=0.00001
   b=r;
   [r]=codebookS(a,b);
   z=b-r;
   z=z.^2;
   v=sum(z)/length(z);
   save n r
   round=round+1
end


function [r]=codebookS(a,b);

% function is subcodebook
% r:codebook output
% a,b:value at each frame

u=1;
k=1;
for i=1:17500
   f(1:11)=a(k:k+10);
   p=1;
   for u=1:64
      g(1:11)=b(p:p+10);
      c=f-g;
      c=c.^2;
      w=sum(c);
      d(1,u)=w/length(c);
      p=p+11;
   end
   index(1,i)=find(d==min(d));
   k=k+11;
end;
for y=1:64
   w=find(index==y);
   q=length(w);
   o=[0 0 0 0 0 0 0 0 0 0 0];
   for x=1:q
      h=w(x);
      i=1+(h-1)*11;
      j(1:11)=a(i:i+10);
      o=o+j;
   end
   o=o./q;
   s=1+(y-1)*11;
   r(s:s+10)=o(1:11);
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการควอนไตซ์เซชั่น

-ฟังก์ชัน quantize : การจัดระดับค่าพารามิเตอร์ต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function [bO]=Quantize(w);

% function quantization with codebook
% bO:observation
% w:cepstrum parameter

load n
i=1;
for n=1:25
   k=1;
   j=1;
   for m=1:64
      d(1:11)=w(i:i+10)-r(j:j+10);
      d=sum(d.^2)/length(d);
      f(1,k)=d;
      k=k+1;
      j=j+11;
   end
   c=find(f==min(f));
   bO(1,n)=c;
   i=i+11;
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการหาแบบจำลองอ้างอิงโดยวิธีฮิดเดนมาร์คอฟ

-ฟังก์ชัน model : การหาแบบจำลองอ้างอิงของเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function [aN,bN,pi1,N]=model(w);

%funtion find model
%w:weigth cepstum
%clear;
%speech=input('what is your filename...','s');
%[E,a1,b1,ref,s]=detect1(speech);
%[S,z,a2,b2,Zmax]=zrocrs1(speech,a1,b1);
%normS=normalize(S,4000);
%[b,c,w,G]=levin_Durbin(normS,11,11);
% b:LPC  ,   c:cepstrum  ,  w:weight cepstrum

[bO]=Quantize(w);
%            Hidden markov model section
%find pi1 a1 a2 a3 a4 a5 a6 b1 b2 b3 b4 b5 b6
pi1 = [1 0 0 0 0 0];
a1 = [0.3 0.5 0.2 0 0 0];
a2 = [0 0.4 0.3 0.3 0 0];
a3 = [0 0 0.4 0.2 0.4 0];
a4 = [0 0 0 0.7 0.2 0.1];
a5 = [0 0 0 0 0.5 0.5];
a6 = [0 0 0 0 0 1];
b1 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
b2 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
b3 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
b4 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
b5 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
b6 = [1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64 1/64
1/64 1/64];
n = [1];
[pi1,aN,a1,a2,a3,a4,a5,a6,bN,b1,b2,b3,b4,b5,b6,Pbw] =
hmm(bO,pi1,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6);
PBW(n)=Pbw;
[pi1,aN,a1,a2,a3,a4,a5,a6,bN,b1,b2,b3,b4,b5,b6,Pbw] =
hmm(bO,pi1,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6);
PBW(n+1)=Pbw;
if (PBW(n)/PBW(n+1))>0.9999
   n = n+1;
   [pi1,aN,a1,a2,a3,a4,a5,a6,bN,b1,b2,b3,b4,b5,b6,Pbw] =
hmm(bO,pi1,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6);
   PBW(n+1)=Pbw;
end;
%find best state sequence
[N] = veterbi(bO,pi1,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%    start begin model
function [pi1,aN,a1,a2,a3,a4,a5,a6,bN,b1,b2,b3,b4,b5,b6,Pbw] = hmm
(bO,pi1,a1,a2,a3,a4,a5,a6,b1,b2,b3,b4,b5,b6);

%function find model
%part calculator value begin Alpha forward  3.4.6
A1(1) = pi1(1)*b1(bO(1));
A2(1) = pi1(2)*b2(bO(1));
A3(1) = pi1(3)*b3(bO(1));
A4(1) = pi1(4)*b4(bO(1));
A5(1) = pi1(5)*b5(bO(1));
A6(1) = pi1(6)*b6(bO(1));

for n = 1 : 1 : 24
%part calculator Alpha forward 3.4.5
   A1(n+1) = ((A1(n)*a1(1)) + (A2(n)*a2(1)) + (A3(n)*a3(1)) + (A4(n)*a4(1)) +
(A5(n)*a5(1)) + (A6(n)*a6(1)))*b1(bO(n+1));
   A2(n+1) = ((A1(n)*a1(2)) + (A2(n)*a2(2)) + (A3(n)*a3(2)) + (A4(n)*a4(2)) +
(A5(n)*a5(2)) + (A6(n)*a6(2)))*b2(bO(n+1));
   A3(n+1) = ((A1(n)*a1(3)) + (A2(n)*a2(3)) + (A3(n)*a3(3)) + (A4(n)*a4(3)) +
(A5(n)*a5(3)) + (A6(n)*a6(3)))*b3(bO(n+1));
   A4(n+1) = ((A1(n)*a1(4)) + (A2(n)*a2(4)) + (A3(n)*a3(4)) + (A4(n)*a4(4)) +
(A5(n)*a5(4)) + (A6(n)*a6(4)))*b4(bO(n+1));
   A5(n+1) = ((A1(n)*a1(5)) + (A2(n)*a2(5)) + (A3(n)*a3(5)) + (A4(n)*a4(5)) +
(A5(n)*a5(5)) + (A6(n)*a6(5)))*b5(bO(n+1));
   A6(n+1) = ((A1(n)*a1(6)) + (A2(n)*a2(6)) + (A3(n)*a3(6)) + (A4(n)*a4(6)) +
(A5(n)*a5(6)) + (A6(n)*a6(6)))*b6(bO(n+1));

end;

%part calculator  Baum-Welch forward 3.4.3
Pbw = A1(25) + A2(25) + A3(25) + A4(25) + A5(25) + A6(25);

%part calculator Backward Probability forward 3.4.13
B1(25) = 1; B2(25) = 1; B3(25) = 1; B4(25) = 1; B5(25) = 1; B6(25) = 1;
for n = 25 : -1 : 2
    B6(n-1) = (a6(1)*b1(bO(n))*B1(n)) + (a6(2)*b2(bO(n))*B2(n)) + (a6(3)*b3(bO
(n))*B3(n)) + (a6(4)*b4(bO(n))*B4(n)) + (a6(5)*b5(bO(n))*B5(n)) + (a6(6)*b6(bO
(n))*B6(n));
    B5(n-1) = (a5(1)*b1(bO(n))*B1(n)) + (a5(2)*b2(bO(n))*B2(n)) + (a5(3)*b3(bO
(n))*B3(n)) + (a5(4)*b4(bO(n))*B4(n)) + (a5(5)*b5(bO(n))*B5(n)) + (a5(6)*b6(bO
(n))*B6(n));
    B4(n-1) = (a4(1)*b1(bO(n))*B1(n)) + (a4(2)*b2(bO(n))*B2(n)) + (a4(3)*b3(bO
(n))*B3(n)) + (a4(4)*b4(bO(n))*B4(n)) + (a4(5)*b5(bO(n))*B5(n)) + (a4(6)*b6(bO
(n))*B6(n));
    B3(n-1) = (a3(1)*b1(bO(n))*B1(n)) + (a3(2)*b2(bO(n))*B2(n)) + (a3(3)*b3(bO
(n))*B3(n)) + (a3(4)*b4(bO(n))*B4(n)) + (a3(5)*b5(bO(n))*B5(n)) + (a3(6)*b6(bO
(n))*B6(n));
    B2(n-1) = (a2(1)*b1(bO(n))*B1(n)) + (a2(2)*b2(bO(n))*B2(n)) + (a2(3)*b3(bO
(n))*B3(n)) + (a2(4)*b4(bO(n))*B4(n)) + (a2(5)*b5(bO(n))*B5(n)) + (a2(6)*b6(bO
(n))*B6(n));
    B1(n-1) = (a1(1)*b1(bO(n))*B1(n)) + (a1(2)*b2(bO(n))*B2(n)) + (a1(3)*b3(bO
(n))*B3(n)) + (a1(4)*b4(bO(n))*B4(n)) + (a1(5)*b5(bO(n))*B5(n)) + (a1(6)*b6(bO
(n))*B6(n));

end;

%part calculator State Transition forward 3.4.12
a11s = 0; a12s = 0; a13s = 0; a1t = 0;
a22s = 0; a23s = 0; a24s = 0; a2t = 0;
a33s = 0; a34s = 0; a35s = 0; a3t = 0;
a44s = 0; a45s = 0; a46s = 0; a4t = 0;
a55s = 0; a56s = 0; a5t = 0;
a66s = 0; a6t = 0;

for n = 2 : 1 : 25
    a11s = a11s + (A1(n-1) * a1(1) * b1(bO(n)) * B1(n));
    a12s = a12s + (A1(n-1) * a1(2) * b2(bO(n)) * B2(n));
    a13s = a13s + (A1(n-1) * a1(3) * b3(bO(n)) * B3(n));
    a1t = a1t + (A1(n-1) * B1(n-1));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    a22s = a22s + (A2(n-1) * a2(2) * b2(bO(n)) * B2(n));
    a23s = a23s + (A2(n-1) * a2(3) * b3(bO(n)) * B3(n));
    a24s = a24s + (A2(n-1) * a2(4) * b4(bO(n)) * B4(n));
    a2t = a2t + (A2(n-1) * B2(n-1));

    a33s = a33s + (A3(n-1) * a3(3) * b3(bO(n)) * B3(n));
    a34s = a34s + (A3(n-1) * a3(4) * b4(bO(n)) * B4(n));
    a35s = a35s + (A3(n-1) * a3(5) * b5(bO(n)) * B5(n));
    a3t = a3t + (A3(n-1) * B3(n-1));

    a44s = a44s + (A4(n-1) * a4(4) * b4(bO(n)) * B4(n));
    a45s = a45s + (A4(n-1) * a4(5) * b5(bO(n)) * B5(n));
    a46s = a46s + (A4(n-1) * a4(6) * b6(bO(n)) * B6(n));
    a4t = a4t + (A4(n-1) * B4(n-1));

    a55s = a55s + (A5(n-1) * a5(5) * b5(bO(n)) * B5(n));
    a56s = a56s + (A5(n-1) * a5(6) * b6(bO(n)) * B6(n));
    a5t = a5t + (A5(n-1) * B5(n-1));

    a66s = a66s + (A6(n-1) * a6(6) * b6(bO(n)) * B6(n));
    a6t = a6t + (A6(n-1) * B6(n-1));
 end;
    a1 = [a11s/a1t  a12s/a1t  a13s/a1t 0  0  0];
    a2 = [0  a22s/a2t  a23s/a2t  a24s/a2t 0  0];
    a3 = [0  0  a33s/a3t  a34s/a3t  a35s/a3t 0];
    a4 = [0  0  0  a44s/a4t  a45s/a4t a46s/a4t];
    a5 = [0  0  0  0  a55s/a5t a56s/a5t];
    a6 = [0  0  0  0  0 a66s/a6t];
    aN = [a1; a2; a3; a4; a5; a6];


%     part find value forward 3.4.15
for n = 1 : 64
    b1N(n) = 0;  b2N(n) = 0;  b3N(n) = 0;  b4N(n) = 0;  b5N(n) = 0;  b6N(n) =
0;
end;
b1Total=0; b2Total=0; b3Total=0; b4Total=0; b5Total=0; b6Total=0;
for n = 1 : 25
    b1Total = b1Total + (A1(n) * B1(n));
    b2Total = b2Total + (A2(n) * B2(n));
    b3Total = b3Total + (A3(n) * B3(n));
    b4Total = b4Total + (A4(n) * B4(n));
    b5Total = b5Total + (A5(n) * B5(n));
    b6Total = b6Total + (A6(n) * B6(n));
    for nn = 1 : 64
            if bO(n) == nn;
                b1N(nn) = b1N(nn) + (A1(n) * B1(n));
                b2N(nn) = b2N(nn) + (A2(n) * B2(n));
                b3N(nn) = b3N(nn) + (A3(n) * B3(n));
                b4N(nn) = b4N(nn) + (A4(n) * B4(n));
                b5N(nn) = b5N(nn) + (A5(n) * B5(n));
                b6N(nn) = b6N(nn) + (A6(n) * B6(n));
            end;
    end;
end;
 b1 = b1N/b1Total;
 b2 = b2N/b2Total;
 b3 = b3N/b3Total;
 b4 = b4N/b4Total;
 b5 = b5N/b5Total;
 b6 = b6N/b6Total;
 bN = [b1;b2;b3;b4;b5;b6];
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการหาลำดับสเตทที่ดีที่สุดของการใช้วิธีฮิดเดนมาร์คอฟ

-ฟังก์ชัน veterbi : การหาลำดับสเตทที่ดีที่สุดของการใช้วิธีฮิดเดนมาร์คอฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function [N] = veterbi(bO,pi1,aN,bN);

% function find best state sequence
% N:best state sequence
%bO:observation
%pi1:probablity start
%aN:state sequence probability
%bN:observation probability

R1(1)=pi1(1)*bN(1,(bO(1)));
R2(1)=0;
R3(1)=0;
R4(1)=0;
R5(1)=0;
R6(1)=0;
R=[R1 R2 R3 R4 R5 R6];
T=max(R);
v=find(R==T);
N(1)=v;

for n=1:24
   r1=R1*aN(v,1)*bN(1,(bO(n+1)));
   r2=R1*aN(v,2)*bN(2,(bO(n+1)));
   r3=R1*aN(v,3)*bN(3,(bO(n+1)));
   r4=R1*aN(v,4)*bN(4,(bO(n+1)));
   r5=R1*aN(v,5)*bN(5,(bO(n+1)));
   r6=R1*aN(v,6)*bN(6,(bO(n+1)));
   R=[r1 r2 r3 r4 r5 r6];
   T=max(R);
   v=find(R==T);
   N(n+1)=v;
   R1=T;
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมทดสอบสัญญาณเสียงกับแบบจำลองอ้างอิง

-ฟังก์ชัน voice : การทดสอบเสียงที่เข้ามากับแบบจำลองอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function [a,S,s,w,numberlength,numberweight,g,p,q,e,normS,bO]=voice(filename)

% function test recognition
%a:value description to speech
%filename:file.wav
%g:gian
%p:order of lpc
%q:order of weigth cepstrum
%numberlength:length of speech from detect
%numberweigth:length of weigth cepstrum
%normS:speech from normalization

speech=filename;
normsize=4000;p=11;q=11;
[E,a1,b1,ref,s]=detect1(speech);
[S,z,a2,b2,Zmax]=zrocrs1(speech,a1,b1);
normS=normalize(S,normsize);
data_speech=normS;
[normal_speech]=normalizationamplijude(data_speech);
normS=normal_speech;
[b,c,w,g,e]=levin_Durbin(normS,p,q);
numberlength=length(S);
numberweight=length(w);
% b:LPC  .   c:cepstrum  .  w:weight cepstrum
[bO]=Quantize(w);

%test
load model0;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(1)=Pbw;
load model1;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(2)=Pbw;
load model2;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(3)=Pbw;
load model3;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(4)=Pbw;
load model4;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(5)=Pbw;
load model5;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(6)=Pbw;
load model6;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(7)=Pbw;
load model7;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(8)=Pbw;
load model8;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(9)=Pbw;
load model9;
[Pbw] = hmmt(bO,pi11,aN,bN);
PBW(10)=Pbw;

%[y,i]=sort(PBW);
%y
%i=i-1

if max(PBW)==PBW(1);
      a=1;
      %disp('This speech is 0 ');

     elseif max(PBW)==PBW(2);
        a=2;
        %disp('This speech is 1');

     elseif max(PBW)==PBW(3);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        a=3;
        %disp('This speech is 2');

    elseif max(PBW)==PBW(4);
        a=4;
        %disp('This speech is 3');

    elseif max(PBW)==PBW(5);
        a=5;
        %disp('This speech is 4');

    elseif max(PBW)==PBW(6);
        a=6;
        %disp('This speech is 5');

    elseif max(PBW)==PBW(7);
        a=7;
        %disp('This speech is 6');

    elseif max(PBW)==PBW(8);
        a=8;
        %disp('This speech is 7');

    elseif max(PBW)==PBW(9);
        a=9;
        %disp('This speech is 8');

    elseif max(PBW)==PBW(10);
        a=10;
        %disp('This speech is 9')
    end;
%plot(normS)
%sound(S)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%    start begin model
function [Pbw] = hmmt(bO,pi1,aN,bN,N);

% function test recognition
%bO:observation
%pi1:probability start
%aN:state sequence probability
%bN:observation probability

Pbw=pi1(N(1))*bN(N(1),bO(1));
for n=1:1:24;
      Pbw=Pbw*aN(N(n),N(n+1))*bN(N(n+1),bO(n+1));
end;
```


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมการสร้างเมนู

Graphic User Interface : GUI

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function demo(action)
% This is the machine-generated representation of a Handle Graphics object
% and its children.  Note that handle values may change when these objects
% are re-created. This may cause problems with any callbacks written to
% depend on the value of the handle at the time the object was saved.
%
% To reopen this object, just type the name of the M-file at the MATLAB
% prompt. The M-file and its associated MAT-file must be on your path.

%load demo
global recog filename readwave fs bits ce w
global numberlength numberwav numberweight p q
global minute second s S normS bO
if nargin==0
   action='initialize';
end
if strcmp(action,'initialize')
   a = figure('Color',[0.8 0.8 0.8], ...
      'MenuBar','none', ...
      'Name','SPEECH PROCESSING', ...
      'NumberTitle','off', ...
      'Position',[145 122 560 420], ...
      'Resize','off', ...
      'Tag','Fig1');
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
      'Position',[2.25 2.25 415.5 309.75], ...
      'Style','frame', ...
      'Tag','Frame1');
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[0.5 0.5 0.5], ...
      'Position',[302.25 5.25 111.75 301.5], ...
      'Style','frame', ...
      'Tag','Frame2');
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[0.752941 0.752941 0.752941], ...
      'Position',[10.5 6 284.25 123.75], ...
      'Style','frame', ...
      'Tag','Frame3');
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[0.7 0.7 0.7], ...
      'Position',[15 9 132 116.25], ...
      'Style','frame', ...
      'Tag','Frame4');
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[0.7 0.7 0.7], ...
      'Position',[156.75 9 132 116.25], ...
      'Style','frame', ...
      'Tag','Frame4');
   recog(1) = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[1 1 1], ...
      'HorizontalAlignment','left', ...
      'Max',5,...
      'Position',[18 13.5 124.5 84.75], ...
      'Style','edit', ...
      'Tag','EditText1');                                    %edit OPEN
   recog(2) = uicontrol('Parent',a, ...
      'Units','points', ...
      'BackgroundColor',[1 1 1], ...
      'HorizontalAlignment','left', ...
      'Max',10,...
      'Position',[160.5 13.5 124.5 84.75], ...
      'Style','edit', ...
      'Tag','EditText2');                                    %edit OUTPUT
   recog1 = uicontrol('Parent',a, ...
      'Units','points', ...
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    'BackgroundColor',[0.7 0.7 0.7], ...
    'FontSize',10, ...
    'Position',[16.5 105 34.5 12.75], ...
    'String','File...', ...
    'Style','text', ...
    'Tag','StaticText1');
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0.7 0.7 0.7], ...
    'FontSize',10, ...
    'Position',[162 104.25 36.75 13.5], ...
    'String','Output...', ...
    'Style','text', ...
    'Tag','StaticText2');
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'Position',[11.25 133.5 282.75 173.25],
    'Style','frame', ...
    'Tag','Frame5');
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'FontWeight','bold', ...
    'ForegroundColor',[0 0 1], ...
    'HorizontalAlignment','left', ...
    'Position',[17.25 290.25 46.5 10.5], ...
    'String','Open_wave', ...
    'Style','text', ...
   'Tag','StaticText3');
% The OPEN button
labelstr='OPEN';
recog(3) = uicontrol('Parent',a, ...
   'Callback','demo(''OPEN'')',...
    'Units','points', ...
    'FontSize',10, ...
    'FontWeight','bold', ...
    'Position',[22.5 255.75 261 26.25], ...
   'String','OPEN', ...
   'userdata',1,...
    'Tag','Pushbutton1');                          %button OPEN
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'FontWeight','bold', ...
    'ForegroundColor',[0 0 1], ...
    'HorizontalAlignment','left', ...
    'Position',[15.75 237.75 46.5 12.75], ...
    'String','Output', ...
    'Style','text', ...
   'Tag','StaticText4');
% The OUTPUT button
recog(4) = uicontrol('Parent',a, ...
   'Callback','demo(''OUTPUT'')',...
    'Units','points', ...
    'FontSize',10, ...
    'FontWeight','bold', ...
    'Position',[21.75 208.5 261 25.5], ...
   'String','OUTPUT', ...
   'userdata',1,...
    'Tag','Pushbutton2');                          %button OUTPUT
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'FontWeight','bold', ...
    'ForegroundColor',[0 0 1], ...
    'HorizontalAlignment','left', ...
    'Position',[15.75 187.5 41.25 13.5], ...
    'String','Detail', ...
    'Style','text', ...
   'Tag','StaticText5');
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% The DETAIL button
recog(5) = uicontrol('Parent',a, ...
    'Callback','demo(''DETAIL'')',...
    'Units','points', ...
    'FontSize',10, ...
    'FontWeight','bold', ...
    'Position',[21.75 162.75 125.25 22.5],
    'String','DETAIL', ...
    'Tag','Pushbutton3');                                    %button DETAIL
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'FontWeight','bold', ...
    'ForegroundColor',[0 0 1], ...
    'HorizontalAlignment','left', ...
    'Position',[157.5 188.25 39 11.25], ...
    'String','Plotwave', ...
    'Style','text', ...
    'Tag','StaticText6');
%The CLEAR button
recog(6) = uicontrol('Parent',a, ...
    'Callback','demo(''PLOT'')',...
    'Units','points', ...
    'FontSize',10, ...
    'FontWeight','bold', ...
    'Position',[157.5 162.75 123 22.5], ...
    'String','PLOT', ...
    'Tag','Pushbutton4');                                    %button PLOT
recog1 = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0 0 0], ...
    'FontSize',10, ...
    'FontWeight','bold', ...
    'ForegroundColor',[1 0 0], ...
    'Position',[22.5 138.75 255.75 15.75], ...
    'String','SPEECH RECOGNITION PROGRAM',
    'Style','text', ...
    'Tag','StaticText7');
%The Exit button
recog(7) = uicontrol('Parent',a, ...
    'Callback','close(gcbf)',...
    'Units','points', ...
    'FontSize',10, ...
    'Position',[310.5 19.5 95.25 26.25], ...
    'String','Exit', ...
    'Tag','Pushbutton5');                                    %button Exit
recog(8) = uicontrol('Parent',a, ...
    'Callback','demo(''Sound'')',...
    'Units','points', ...
    'Position',[81.75 103.5 60.75 15.75], ...
    'String','Sound', ...
    'Tag','Pushbutton6');                                    %button Sound
recog(9) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','left', ...
    'Max',10, ...
    'Position',[306 171 103.5 129.75], ...
    'Style','edit', ...
    'Tag','EditText3');                                      %edit Help
recog(10) = uicontrol('Parent',a, ...
    'Callback','demo(''Help'')',...
    'Units','points', ...
    'FontSize',8, ...
    'Position',[310.5 52.5 95.25 24.75], ...
    'String','Help', ...
    'Tag','Pushbutton7');                                    %button Help
recog(11) = uicontrol('Parent',a, ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','left', ...
    'Units','points', ...
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        'Position',[162.75 15.75 118.5 12], ...
        'Style','text', ...
        'Tag','StaticText8');                                    % count timing

    set(a,'vis','on')
    drawnow
    %action=1;
    elseif strcmp(action,'OPEN'),
        filename=[];
           [filename pathname]= uigetfile('*.wav');
           if filename ~=0
             filename1=[pathname filename];
             [readwave,fs,bits]=wavread(filename);
             numberwav=length(readwave);
           text=filename1;
        set(recog(1),...
             'string',text );
           set(recog(3),'userdata',1,'enable','off');
             msgbox('"Open wavefile complete."','Information');
        else
           msgbox('"Not a invalid input"','error');
        end
       elseif strcmp(action,'OUTPUT'),
          t1=clock;[a,S,s,w,numberlength,numberweight,g,p,q,e,normS,bO]=voice
(filename);t2=clock;
          %Time = t2-t1;
          %Tfin =((t2(:,5)*60)+t2(:,6)); Tinit =(t1(:,5)*60+t1(:,6));
          %Time1 = Tfin - Tinit; Time =Time1/60;
          %minute =num2str((Time(:,5)));second = num2str(Time(:,6));
          %minute = num2str(fix(Time));second = num2str(Time-fix(Time));
          set(recog(4),'enable','off');
          %msgbox('"Process Complete"','information');
          if a==1;
             text= str2mat('The output from recognition -',...
                'test is :.....ZERO...(0)"',...
                '',...
                'THANK YOU.....   ');
           % time =strcat(minute,':',second);
           %  text1=strcat('Time is [min sec]: ...',time);
             set(recog(2),'string',text);
           % set(recog(11),'string',text1);
          elseif a==2;
             text= str2mat('The output from recognition -',...
                'test is :.....ONE...(1)"',...
                '',...
                'THANK YOU.....   ');
           % time =strcat(minute,':',second);
            %text1=strcat('Time is [min sec]: ...',time);
            set(recog(2),'string',text);
            %set(recog(11),'string',text1);
         elseif a==3;
            text= str2mat('The output from recognition -',...
               'test is :.....TWO...(2)"',...
               '',...
               'THANK YOU.....   ');
            %time =strcat(minute,':',second);
            %text1=strcat('Time is [min sec]: ...',time);
            set(recog(2),'string',text);
            %set(recog(11),'string',text1);
       elseif a==4;
            text= str2mat('The output from recognition -',...
               'test is :.....THREE...(3)"',...
               '',...
               'THANK YOU.....   ');
            %time=strcat(minute,':',second);
            %text1=strcat('Time is [min sec]: ...',time);
            set(recog(2),'string',text);
            %set(recog(11),'string',text1);
          elseif a==5;
            text= str2mat('The output from recognition -',...
               'test is :.....FOUR...(4)"',...
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            '',...
            'THANK YOU.....   ');
        %time=strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);

    elseif a==6;
        text= str2mat('The output from recognition -',...
            'test is :....."FIVE...(5)"',...
            '',...
            'THANK YOU.....   ');
        %time=strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);

    elseif a==7;
        text= str2mat('The output from recognition -',...
            'test is :....."SIX...(6)"',...
            '',...
            'THANK YOU.....   ');
        %time =strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);

    elseif a==8;
        text= str2mat('The output from recognition -',...
            'test is :....."SEVEN...(7)"',...
            '',...
            'THANK YOU.....   ');
        %time =strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);

    elseif a==9;
        text= str2mat('The output from recognition -',...
            'test is :....."EIGHT...(8)"',...
            '',...
            'THANK YOU.....   ');
        %time =strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);

    elseif a==10;
        text= str2mat('The output from recognition -',...
            'test is :....."NINE...(9)"',...
            '',...
            'THANK YOU.....   ');
        %time =strcat(minute,':',second);
        %text1=strcat('Time is [min sec]:...  ',time);
        set(recog(2),'string',text);
        %set(recog(11),'string',text1);
    end
    msgbox('"Process Complete"','information');

elseif strcmp(action,'PLOT'),
    %plot11;
    figure;
    subplot(3,1,1),plot(readwave);
    title('Original signal');
    %xlabel('sample');
    ylabel('Amplitude');
    subplot(3,1,2),plot(s);
    title('Detection by energy');
    %xlabel('sample');
    ylabel('Amplitude');
    subplot(3,1,3),plot(S);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        title('Detection by zero-crossing');
        %xlabel('sample');
        ylabel('Amplitude');
        figure;
        subplot(3,1,1),plot(normS);
        title('Normalize');
        %xlabel('sample');
        ylabel('Amplitude');
         subplot(3,1,2),plot(w);
        title('Weight cepstrum');
        %xlabel('sample');
        ylabel('Amplitude');
        subplot(3,1,3),plot(bO);
        title('Observation');
        %xlabel('sample');
        ylabel('Amplitude');

    elseif strcmp(action,'Sound'),
        if filename ~=0
           %[readwave,fs,bits]=wavread(filename)
           if (fs==8000)&(bits==8)
              sound(readwave,fs);
           else
              msgbox('"Not sampling 8K or not 8 bits"','error');
           end
        else
           msgbox('"Please Open file .wav"','error');
        end
     elseif strcmp(action,'Help'),
        text=str2mat('Help Speech recognition',...
           '(1) Press button Open for- select .wav',...
           '(2) Run program Press- button Output',...
           '(3) Press Sound for listen- to sound',...
           '(4) Press Exit for Exit- Program',...
           '***** Thank you *****');
        set(recog(9),'string',text);
     elseif strcmp(action,'DETAIL'),
          detail(filename,fs,bits,numberlength,numberweight,numberwav,p,q);
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function detail(filename,fs,bits,numberlength,numberweight,numberwav,p,q)
% This is the machine-generated representation of a Handle Graphics object
% and its children.  Note that handle values may change when these objects
% are re-created.  This may cause problems with any callbacks written to
% depend on the value of the handle at the time the object was saved.
%
% To reopen this object, just type the name of the M-file at the MATLAB
% prompt.  The M-file and its associated MAT-file must be on your path.

%load detail
global  recog2 filename fs bits w numberlength numberweight numberwav p q
a = figure('Color',[0.8 0.8 0.8], ...
    'MenuBar','none', ...
    'Name','DETAIL SPEECH', ...
    'NumberTitle','off', ...
    'Position',[96 148 277 373], ...
    'Resize','off', ...
    'Tag','Fig1');
b = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0.5 0.5 0.5], ...
    'Position',[124.5 1.5 81.75 273.75], ...
    'Style','frame', ...
    'Tag','Frame1');
b = uicontrol('Parent',a, ...
    'Units','points', ...
    'Callback','close(gcbf)', ...
    'Position',[129 9 73.5 18.75], ...
    'String','Close', ...
    'Tag','Pushbutton1');
b = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0.752941 0.752941 0.752941], ...
    'Position',[3 0.75 120 273.75], ...
    'Style','frame', ...
    'Tag','Frame2');
recog2(1) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','left', ...
    'Max',15, ...
    'Position',[8.25 8.25 109.5 247.5], ...
    'Style','edit', ...
    'Tag','EditText1');
recog2(2) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[62.25 243 52.5 10.5], ...
    'Style','text', ...
    'Tag','StaticText3');
recog2(3) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[63 231.75 52.5 10.5], ...
    'Style','text', ...
    'Tag','StaticText3');
recog2(4) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[63 220.5 52.5 10.5], ...
    'Style','text', ...
    'Tag','StaticText3');
recog2(5) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[62.25 210 53.25 10.5], ...
    'Style','text', ...
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    'Tag','StaticText2');                                   %wave open
recog2(6) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[75.0 199.5 40.5 10.5], ...
    'Style','text', ...
    'Tag','StaticText2');                                   %Detect wave
recog2(7) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[62.25 189 53.25 10.5], ...
    'Style','text', ...
    'Tag','StaticText2');                                   %LPC order
recog2(8) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[75.75 178.5 40.5 10.5], ...
    'Style','text', ...
    'Tag','StaticText2');                                   %cepstrum order
recog2(9) = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[1 1 1], ...
    'HorizontalAlignment','right', ...
    'Position',[82.5 168.75 33.75 9.75], ...
    'Style','text', ...
    'Tag','StaticText4');                                   %comificient

b = uicontrol('Parent',a, ...
    'Units','points', ...
    'BackgroundColor',[0.752941 0.752941 0.752941], ...
    'ForegroundColor',[0.7 0 1], ...
    'HorizontalAlignment','left', ...
    'Position',[9.75 257.25 42.75 11.25], ...
    'String','Detail...', ...
    'Style','text', ...
    'Tag','StaticText1');

%clear all;
    text= str2mat('Wave name :', ...
    'Sampling rate :', ...
       'Bits is :', ...
       'Wave open :', ...
       'Detection wave :', ...
       'LPC order :', ...
       'Cepstrum order :', ...
       'Coefficient number :');
    set(recog2(1),'string',text);
    set(recog2(2),'string',filename);
    set(recog2(3),'string',fs);
    set(recog2(4),'string',bits);
    set(recog2(5),'string',numberwav);
    set(recog2(6),'string',numberlength);
    set(recog2(7),'string',p);
    set(recog2(8),'string',q);
    set(recog2(9),'string',numberweight);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงงานชิ้นนี้สำเร็จลุล่วงไปด้วยดี ก็ด้วยความช่วยเหลือจากบุคคลหลายๆ ท่านที่คอยช่วยในด้านความรู้ เครื่องมืออุปกรณ์ต่างๆ ตลอดจนให้คำแนะนำจึงต้องขอกล่าวขอบคุณบุคคลเหล่านี้คือ บิดามารดาที่คอยเป็นกำลังใจในงานชิ้นนี้ให้สำเร็จลุล่วงไปด้วยดี อาจารย์กฤษณ์ วงรุจิระ และ ดร. สุทธิชัย นพนาคีพงษ์ อาจารย์ที่ปรึกษาโครงงานที่คอยให้คำปรึกษาและช่วยเหลืออำนวยความสะดวกในการใช้เครื่องมือและอุปกรณ์ และขอขอบคุณเพื่อนและบุคคลอื่นอีกหลายท่านที่ไม่ได้กล่าวถึงไว้ ณ ที่นี้ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หนังสืออ้างอิง

1.Lawrence Rabiner and Biing-Hwang Juang, "Fundamental of Speech Revolotion",New jersy: Prentice Hall,1993.

2.John R.Deller , John G.Proakis and John H.L. Hanson, "Discrete Time Processing of Speech Signals",New York:Macmillan,1993.

3.Thomas W.Parsons, "Voice and Speech Processing",McGraw-Hill Book Company,1987

4.Cory Myers,L.R.Rabiner,Aaron E.Rosenberg, "Performance Tradeoffs in DTW Algorithms for Isolated Word Recognition",IEEE Transaction on Acoustion Acoustics Speech and Signal Processing,Vol.ASSP-28,NO.6,December 1980,P 623-635.

5.Lori F.Lamel, "An Improved Endpoint Detection for Isolated Word Recognition",IEEE Transaction on Acoustics Speech and Signal Processing, Vol.ASSP-29,No.4,August 1981, P777-785.

6.Yiu-Kei Lua and Chok-ki Chan,"Speech Recognition Based on Zero Crossing Rate and Energy ,"IEEE Transaction on Acoustics Speech and Signal Processing ,Vol.ASSP-33,No.1,February 1985,P 320-323.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้