การพัฒนาระบบรู้จำเสียงพูด/ภาพใบหน้าและแนวทางการประยุกต์ใช้งาน
DEVELOPMENT OF SPEECH/FACIAL IMAGE RECOGNITION AND THEIR APPLICATIONS

โดย

นาย ศิรเสกข์ ตราชื่นต้อง

นาย อธิษฐ์ แย้มสุข

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

การพัฒนาระบบรู้จำเสียงพูด/ภาพใบหน้าและแนวทางการประยุกต์ใช้งาน

Development of Speech/Facial Image Recognition and Their Applications

โดย

นาย ศิรเสกข์ ตราชื่นต้อง 56011212

นาย อธิษฐ์ แย้มสุข 56011398

อาจารย์ที่ปรึกษา

ผศ.ดร. ศรวัฒน์ ชิวปรีชา

ผศ. อัครพล ตรีรัตน์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2559

ผ่านการตรวจรูปเล่มแล้ว

(.....)

อาจารย์ที่ปรึกษา

...../...../.....

KMITL วิศวกรรมโทรคมนาคม Telecommunications Engineering

ผ่านการตรวจชิ้นงานแล้ว

(.....)

กรรมการผู้ตรวจชิ้นงาน

...../...../.....

KMITL วิศวกรรมโทรคมนาคม Telecommunications Engineering

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2559

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง

เรื่อง การพัฒนาระบบรู้จำเสียงพูด/ภาพใบหน้าและแนวทางการประยุกต์ใช้งาน

Development of Speech/Facial Image Recognition and Their Applications

ผู้จัดทำ 1. นาย ศิรเสกข์ ตราชื่นต้อง 56011212

2. นาย อธิษฐ์ แย้มสุข 56011398

..... อาจารย์ที่ปรึกษา

(ผศ.ดร. ศรวัฒน์ ชิวปรีชา)

..... อาจารย์ที่ปรึกษา

(ผศ. อัครพล ตรีรัตน์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์นี้ดำเนินมาได้เป็นอย่างดีด้วยความช่วยเหลือ ชี้แนะ ให้ความรู้และประสบการณ์ต่างๆ มากมายในด้านต่างๆ จาก ผศ.ดร. ศรวัฒน์ ชิวปรีชา และ ผศ. อัครพล ตรีรัตน์ ซึ่งเป็นอาจารย์ที่ปรึกษาในปริญญานิพนธ์ฉบับนี้ เหล่าข้าพเจ้ามีความซาบซึ้งใจในความช่วยเหลือเป็นอย่างมาก

ขอขอบพระคุณ บิดาและมารดาที่ให้ชีวิต คอยดูแล ให้กำลังใจในการศึกษามาโดยตลอด จนทำให้สำเร็จมาถึงจุดนี้ได้

ขอขอบคุณพี่ๆเพื่อนๆน้องๆนักศึกษาทุกคนที่ช่วยเหลือในการบันทึกเสียงเพื่อใช้ในการทดลองในปริญญานิพนธ์นี้

คุณค่าและประโยชน์อันพึงมีจากปริญญานิพนธ์นี้ เหล่าข้าพเจ้าขอมอบให้แด่ผู้มีพระคุณทุกท่าน

นายศิรเสกข์ ตราชื่นต้อง

นายอธิษฐ์ แย้มสุข

ผู้จัดทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การพัฒนาระบบรู้จำเสียงพูด/ภาพใบหน้าและแนวทางการประยุกต์ใช้งาน
DEVELOPMENT OF SPEECH/FACIAL IMAGE RECOGNITION AND THEIR APPLICATIONS

โดย นายศิรเสกข์ ตราชื่นต้อง 56011212
นายอธิษฐ์ แย้มสุข 56011398

อาจารย์ที่ปรึกษา ผศ.ดร. ศรวัฒน์ ชิวปรีชา
ผศ. อัครพล ตรีรัตน์

บทคัดย่อ

ปริญญานิพนธ์นี้นำเสนอการพัฒนาระบบโครงข่ายประสาทเทียมเพื่อการรู้จำเสียงพูดและเพื่อการรู้จำภาพใบหน้าบุคคลซึ่งถูกพัฒนาขึ้นบนโปรแกรม MATLAB PROFESSIONAL โดยในส่วนของระบบรู้จำเสียงพูดนั้น จะใช้วิธีการ MFCC มาทำการดึงคุณลักษณะเด่นของเสียง ซึ่งในที่นี้จะทำการรู้จำชื่อเพลงทั้งหมด 50 เพลง เสร็จแล้วจึงนำไปประยุกต์ใช้งานในการเปิดไฟล์เพลงบนโปรแกรม VCL MEDIA PLAYER เมื่อพูดชื่อเพลงนั้นๆ และในส่วนของระบบรู้จำภาพใบหน้าบุคคลนั้น จะใช้วิธีการไอเกนเฟสในการดึงคุณลักษณะเด่นของภาพ ซึ่งในที่นี้จะทำการรู้จำประเภทกลุ่มบุคคลทั้งหมด 4 กลุ่ม จำแนกเป็น ผู้ชาย ผู้หญิง เด็กและผู้สูงอายุ เสร็จแล้วจึงนำมาประยุกต์ใช้งานในการเปิดโฆษณาที่เหมาะสมกับบุคคลกลุ่มนั้นๆ

ABSTRACT

thesis presents speech recognition and facial image recognition which are developed by MATLAB Professional. In speech recognition we using MFCC for feature extraction and then we will introduce to application to play a music file by saying the song name. In facial image recognition we using Eigenface for feature extraction thus we will classification into 4 group that have male, female, child and old age. Then we will applied to open an video file of advertisement of each group

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
กิตติกรรมประกาศ	I
บทคัดย่อ	II
สารบัญ	III
สารบัญรูป	VII
สารบัญตาราง	XI
บทที่ 1 บทนำ	**1**
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	1
1.3 ขอบเขตของปริญญานิพนธ์	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 ทฤษฎีและหลักการที่เกี่ยวข้อง	**3**
2.1 ระบบรู้จำเสียง	3
2.2 ไอเกนเฟส	6
2.3 การตรวจจับภาพใบหน้าด้วยวิธีการ Viola-Jones	9
2.4 โครงข่ายประสาทเทียมเบื้องต้น	11
2.5 การเชื่อมต่อระหว่างระบบรู้จำเพื่อเรียกไฟล์เพลงบนโปรแกรม VCL Media Player	20
บทที่ 3 การออกแบบและจัดการปริญญานิพนธ์	**24**
3.1 การหาตัวแทนของสัญญาณเสียงเพื่อนำไปเรียนรู้ในระบบโครงข่ายประสาทเทียม	24
3.1.1 การเตรียมสัญญาณเสียงเบื้องต้น	24
3.1.2 การดึงคุณลักษณะเด่นของสัญญาณเสียง	33
3.1.3 การเลือกคุณลักษณะเด่นของสัญญาณเสียง	38
3.2 การหาตัวแทนของสัญญาณภาพเพื่อนำไปเรียนรู้ในระบบโครงข่ายประสาทเทียม	41
3.2.1 กระบวนการแปลงอินพุตข้อมูล	42
3.2.2 กระบวนการหาค่าเฉลี่ยของรูปภาพใบหน้า	44
3.2.3 กระบวนการดึงคุณลักษณะเด่น	44
3.2.4 กระบวนการหาค่า Covariance Matrix	44
3.2.5 กระบวนการหาค่า Eigenvector และ Eigenvalue	45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.2.6 กระบวนการไอเกนเฟส	46
3.2.7 กระบวนการหาค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูล (Ω)	47
3.2.8 เรียงลำดับข้อมูล	49
3.2.9 การนอมอลไลซ์ทางขนาดของข้อมูล	49
3.3 ออกแบบระบบโครงข่ายประสาทเทียมทั้งภาพและเสียง	49
3.3.1 การออกแบบโครงข่ายประสาทเทียมของการรู้จำเสียงพูด	49
3.3.2 การออกแบบโครงข่ายประสาทเทียมของการรู้จำภาพใบหน้า	53
3.3.3 แอพพลิเคชันของระบบโครงข่ายประสาทเทียมการรู้จำเสียงพูดและเสียงภาพใบหน้า	53
บทที่ 4 ผลการทดลอง	**55**
4.1 ระบบโครงข่ายประสาทเทียมสำหรับการรู้จำเสียงพูดและการประยุกต์ใช้เพื่อรู้จำชื่อเพลง	55
4.1.1 Recording	55
4.1.2 AC-Coupling	55
4.1.3 Pre-Emphasis	56
4.1.4 Amplitude Normalization	57
4.1.5 Endpoint Detection	58
4.1.6 Pre-Processing for MFCC	62
4.1.7 แฮมมิ่งวินโดว์	63
4.1.8 Power Spectrum	64
4.1.9 Mel Spectrum & Mel-Scale Filterbank	65
4.1.10 Mel Frequency Cepstral Coefficient	66
4.1.11 MFCC Arrangement	66
4.1.12 Delta & Double Delta Coefficient	67
4.1.13 Resizing	68
4.1.14 Feature Arrangement	69
4.1.15 Amplitude Normalization	71
4.1.16 Completed Feature	72
4.1.17 ระบบการรู้จำ	73

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4.1.18 ออกแบบอินเตอร์เฟสของระบบรู้จำเสียงโดยใช้ MATLAB GUI	73
4.2 ระบบโครงข่ายประสาทเทียมเพื่อรู้จำใบหน้าและการประยุกต์ใช้เพื่อรู้จำใบหน้าของกลุ่มบุคคล	74
4.2.1 กระบวนการเตรียมข้อมูล	75
4.2.2 กระบวนการหาค่าเฉลี่ยของข้อมูล	75
4.2.3 กระบวนการหาลักษณะเด่นของภาพ	76
4.2.4 กระบวนการหา Covariance Matrix	76
4.2.5 กระบวนการหาค่า Eigenvector และ Eigenvalue	76
4.2.6 กระบวนการหาค่าไอเกนเฟส	76
4.2.7 กระบวนการหาค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูล	77
4.2.8 กระบวนการจำแนกโดยใช้การหาระยะทางแบบยุคลิด	79
4.2.9 กระบวนการหาค่าสัมประสิทธิ์ใบหน้าของรูปที่จะนำมาจำแนก	80
4.2.10 กระบวนการหาระยะทางยุคลิด	80
4.2.11 Feature Arrangement & Resizing (Amplitude)	84
4.2.12 ตัวอย่างผลลัพธ์หลังจากผ่านโครงข่ายประสาทเทียม	84
4.3 ผลการทดสอบและประสิทธิภาพของระบบรู้จำ	86
4.3.1 ระบบโครงข่ายประสาทเทียมสำหรับการรู้จำชื่อเพลง	86
4.3.2 ระบบโครงข่ายประสาทเทียมสำหรับรู้จำใบหน้าของกลุ่มบุคคล	96
บทที่ 5 สรุปผลและข้อเสนอแนะ	**97**
5.1 สรุปผลผลการทดลอง	97
5.2 ข้อเสนอแนะ	97
เอกสารอ้างอิง	98
ภาคผนวก ก	99
ภาคผนวก ข	149

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 การเคลื่อนที่ของโมเลกุลของอากาศเทียบกับลักษณะของคลื่น	4
2.2 จำนวนโมเลกุลของอากาศกับแอมพลิจูดของคลื่นเสียง	4
2.3 โครงสร้างอวัยวะในการรับรู้เสียงของหู	5
2.4 รูปแบบของ Features สำหรับการตรวจจับลักษณะแบบต่างๆ	10
2.5 การทำงานของ Adaboost	11
2.6 การทำงานของ Haar Cascade Classifier	11
2.7 รูปแบบของเซลล์สมอง	12
2.8 แสดงแบบจำลองหลักๆ ของ ANN	12
2.9 ตัวอย่างของโครงข่ายประสาทเทียมที่มีโหนดซ่อนเร้น	13
2.10 บล็อกไดอะแกรมของ Neuron อย่างง่าย	14
2.11 ตัวอย่าง Activation Function แบบต่างๆ	15
2.12 Linear Separability in the Perceptron	15
2.13 บล็อกไดอะแกรมของ Multilayer Networks	17
2.14 ปัญหา Exclusive OR	17
2.15 รูปแบบการแก้ไขปัญหา Exclusive OR	18
2.16 ตัวอย่างของบล็อกไดอะแกรมปัญหา XOR	18
2.17 บล็อกไดอะแกรมของขั้นตอนการสั่งไฟล์โดย MATLAB	21
2.18 ตัวอย่าง Batch File ที่มีการชี้ไปไดร์ฟ C	22
2.19 หน้าตาไอคอนโปรแกรม VCL Media Player	23
3.1 บล็อกไดอะแกรมการทำงานของระบบรู้จำเสียง	24
3.2 บล็อกไดอะแกรมของระบบ Pre-Processing	25
3.3 บล็อกไดอะแกรมวิธีการดึงคุณลักษณะเด่นของสัญญาณเสียง	33
3.4 บล็อกไดอะแกรมของการดึงคุณลักษณะเด่น (Feature Selection)	38
3.5 บล็อกไดอะแกรมของกระบวนการไอเกนเฟส	41
3.6 บล็อกไดอะแกรมของกระบวนการรู้จำภาพวิธีไอเกนเฟส	42
3.7 วิธีการแปลงรูปภาพเป็นเวกเตอร์	43
3.8 ผลลัพธ์ของการแปลงอินพุตข้อมูล	43
3.9 การจำแนกโครงข่ายประสาทเทียม	51
3.10 ระบบโครงข่ายประสาทเทียมสำหรับรู้จำชื่อเพลง	52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.11 ระบบโครงข่ายประสาทเทียมสำหรับรู้จำภาพใบหน้า	53
3.12 หน้าต่างเลือกรูปแบบใช้งาน MATLAB GUI	54
3.13 หน้าต่างของ MATLAB GUI	54
4.1 สัญญาณเสียงคำว่า จำฝังใจ	55
4.2 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่เป็นสัญญาณที่ผ่านกระบวนการ AC-Coupling	56
4.3 ผลตอบสนองทางขนาดของวงจรกรองความถี่สูงผ่าน	56
4.4 สัญญาณที่ผ่านกระบวนการ Pre-Emphasis	57
4.5 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่เป็นสัญญาณที่ผ่านกระบวนการ Pre-Emphasis	57
4.6 สัญญาณที่ผ่านกระบวนการ Amplitude Normalization	58
4.7 พลังงานของสัญญาณที่ใช้วิธีการหาพลังงานแบบ Absolute Energy	58
4.8 กรอบสัญญาณทีทำการตัดคำ	59
4.9 กรอบสัญญาณใหม่ที่ทำการตัดคำ	59
4.10 เปรียบเทียบกรอบสัญญาณใหม่กับสัญญาณที่ผ่านกระบวนการ Amplitude Normalization	60
4.11 สัญญาณที่เป็นผลลัพธ์จากการตัดพยางค์	60
4.12 สัญญาณของเสียงพยางค์ที่ 1 หรือคำว่า “จำ”	61
4.13 สัญญาณของเสียงพยางค์ที่ 2 หรือคำว่า “ฝัง”	61
4.14 สัญญาณของเสียงพยางค์ที่ 3 หรือคำว่า “ใจ”	62
4.15 ตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”	62
4.16 กราฟแสดงแฮมมิ่งวินโดว์ (Hamming Window)	63
4.17 เปรียบเทียบแฮมมิ่งวินโดว์กับตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”	63
4.18 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่ผ่านแฮมมิ่งวินโดว์แล้ว	64
4.19 สเปกตรัมของตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”	64
4.20 กราฟแสดงฟิลเตอร์แบงค์ที่ได้จากสมการฟิลเตอร์สามเหลี่ยมชนิดแบนด์พาสฟิลเตอร์	65
4.21 แสดงสเปกตรัมของสัญญาณที่ผ่านฟิลเตอร์สามเหลี่ยมลำดับที่ 1 ถึง 12	65
4.22 ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลของเฟรมที่ 10	66
4.23 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่ 1 ถึง 12 ที่นำมาเรียงกันตามลำดับเฟรมข้อมูล	66
4.24 แสดงค่าสัมประสิทธิ์เดลต้าลำดับที่ 1 ถึง 12 ของตัวอย่างสัญญาณเสียง	67
4.25 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าลำดับที่ 1 ถึง 12 ของตัวอย่างสัญญาณเสียง	67
4.26 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 ข้อมูล	68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูป	หน้า
4.27 แสดงค่าสัมประสิทธิ์เดลต้าลำดับที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 ข้อมูล	68
4.28 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าลำดับที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 ข้อมูล	69
4.29 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกัน	69
4.30 แสดงค่าสัมประสิทธิ์เดลต้าลำดับที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกันตามละดับ	70
4.31 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าลำดับที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกัน	70
4.32 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลที่ถูกนำมาทำการนอมอไลซ์ทางขนาด	71
4.33 แสดงค่าสัมประสิทธิ์เดลต้าที่ถูกนำมาทำการนอมอไลซ์ทางขนาด	71
4.34 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าที่ถูกนำมาทำการนอมอไลซ์ทางขนาด	72
4.35 แสดงค่าสัมประสิทธิ์ที่จะถูกนำไปเรียนรู้เพื่อการรู้จำในระบบโครงข่ายประสาทเทียม	72
4.36 หน้าต่างโปรแกรมหลังจากรันโปรแกรม	73
4.37 หน้าต่างโปรแกรมหลังจากรันโปรแกรมที่สามารถตรวจสอบได้ว่าเป็นคำอะไร	73
4.38 หน้าต่างโปรแกรมหลังจากรันโปรแกรมที่ไม่สามารถตรวจสอบได้ว่าเป็นคำอะไร	74
4.39 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลผู้ชาย (10 ภาพจาก 30 ภาพ)	74
4.40 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลผู้หญิง (10 ภาพจาก 30 ภาพ)	75
4.41 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลเด็ก (10 ภาพจาก 30 ภาพ)	75
4.42 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลสูงอายุ (10 ภาพจาก 30 ภาพ)	75
4.43 ค่าเฉลี่ยของภาพทั้ง 120 ภาพ	76
4.44 ภาพในขั้นตอนการหาลักษณะเด่นของภาพ (ตัวอย่างของกลุ่มบุคคลผู้ชาย)	76
4.45 รูปไอเกนเฟส	77
4.46 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลผู้ชาย	77
4.47 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลผู้หญิง	78
4.48 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลเด็ก	78
4.49 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลสูงอายุ	79
4.50 ตัวอย่างของรูปของกลุ่มบุคคลที่นำมาใช้ทดสอบ	79
4.51 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลผู้ชาย	80
4.52 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลผู้หญิง	81
4.53 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลเด็ก	82
4.54 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลสูงอายุ	83
4.55 ตัวอย่างผลลัพธ์ของค่าที่เราจะนำไปเข้าโครงข่ายระบบประสาทเทียม	84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูป	หน้า
4.56 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลผู้ชาย	85
4.57 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลผู้หญิง	85
4.58 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลเด็ก	86
4.59 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลสูงอายุ	86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
3.1 กลุ่มรายชื่อเพลง 1 พยางค์	26
3.2 กลุ่มรายชื่อเพลง 2 พยางค์	27
3.3 กลุ่มรายชื่อเพลง 3 พยางค์	28
3.4 กลุ่มรายชื่อเพลง 4 พยางค์	29
3.5 กลุ่มรายชื่อเพลง 5 พยางค์ขึ้นไป	30
4.1 ประสิทธิภาพของกลุ่มคำ 1 พยางค์	87
4.2 ประสิทธิภาพของกลุ่มคำ 2 พยางค์	88
4.3 ประสิทธิภาพของกลุ่มคำ 3 พยางค์	89
4.4 ประสิทธิภาพของกลุ่มคำ 4 พยางค์	91
4.5 ประสิทธิภาพของกลุ่มคำ 5 พยางค์ขึ้นไป	93
4.6 สรุปผลการทดสอบระบบรู้จำเสียง	96
4.7 สรุปผลการทดสอบระบบรู้จำภาพ	96

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การพัฒนาเทคโนโลยีต่างๆ ได้มุ่งเน้นไปที่การสร้างสิ่งอำนวยความสะดวกให้กับมนุษย์ ซึ่งเทคโนโลยีที่เรียกว่า ปัญญาประดิษฐ์ (Artificial Intelligence) [1] หรือเรียกว่า เอไอ (AI) ก็เช่นเดียวกัน AI หมายถึงความฉลาดเทียมที่สร้างขึ้นให้กับสิ่งที่ไม่มีชีวิต ปัญญาประดิษฐ์เป็นสาขาหนึ่งในด้านวิทยาการคอมพิวเตอร์และวิศวกรรมเป็นหลัก และสามารถนำไปประยุกต์ใช้กับศาสตร์ในด้านอื่นๆได้อีก ซึ่ง AI นั้นเป็นการเรียนรู้เกี่ยวกับกระบวนการการคิด การกระทำ การให้เหตุผล การปรับตัว หรือการอนุมานและการทำงานของสมอง สามารถนำไปใช้งานได้หลากหลาย ซึ่งระบบรู้จำนั้น ก็อยู่ในส่วนหนึ่งของศาสตร์ปัญญาประดิษฐ์เช่นกัน

ระบบรู้จำ (Recognition System) นั้นเป็นเทคโนโลยีที่กำลังใช้กันอย่างแพร่หลาย ณ เวลานี้ ซึ่งในประเทศไทยนั้นได้มีงานวิจัยในด้านนี้มากกว่า 20 ปีแล้ว แต่ในส่วนของการใช้งานยังไม่เป็นที่แพร่หลายนัก ทางผู้จัดทำจึงได้เล็งเห็นการที่เราจะนำระบบรู้จำนี้ไปประยุกต์ใช้งานจริงได้ ทั้งส่วนของระบบรู้จำเสียงและระบบรู้จำภาพ เป็นต้น

1.2 วัตถุประสงค์

1. เพื่อศึกษาและพัฒนาระบบรู้จำพร้อมนำไปประยุกต์ใช้งานได้
2. เพื่อศึกษาและพัฒนาระบบรู้จำเสียงโดยใช้วิธีการ MFCC ได้
3. เพื่อศึกษาและพัฒนาระบบรู้จำภาพใบหน้าแบบไอเกนเฟส (Eigenface) ได้
4. เพื่อศึกษาและพัฒนาระบบโครงข่ายประสาทเทียมพร้อมนำไปประยุกต์ใช้งานได้

1.3 ขอบเขตของปริญญานิพนธ์

ในปริญญานิพนธ์นี้ เป็นการประยุกต์การใช้งานของระบบรู้จำเสียงพูดและภาพใบหน้าร่วมกับใช้โครงข่ายประสาทเทียมแบบแพร่กลับ (Back propagation) โดยส่วนของเสียงจะเป็นระบบรู้จำเสียง ซึ่งจะรู้จำชื่อเพลงจำนวนทั้งหมด 50 เพลง ซึ่งแบ่งออกเป็น 1 พยางค์ 10 เพลง 2 พยางค์ 10 เพลง 3 พยางค์ 10 เพลง 4 พยางค์ 10 เพลง และกลุ่มมากกว่า 5 พยางค์ ซึ่งประกอบไปด้วยเพลงที่มี 5 พยางค์ 6 เพลง และ 6 พยางค์ 4 เพลง เพื่อไปประยุกต์ใช้งานในการเรียกไฟล์เพลงนั้น ๆ บนโปรแกรม VCL Media Player และในส่วนของภาพจะทำระบบรู้จำกลุ่มรูปภาพใบหน้าของบุคคล ซึ่งแบ่งออกเป็น 4 กลุ่ม คือ กลุ่มบุคคลผู้ชาย (วัยกลางคน) กลุ่มบุคคลผู้หญิง (วัยกลางคน) กลุ่มบุคคลเด็ก และกลุ่มบุคคลสูงอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(อายุ 65 ปีขึ้นไป) เพื่อไปประยุกต์ใช้งานในการเรียกไฟล์วีดีโอโมษณาตามกลุ่มบุคคลนั้น ๆ บนโปรแกรม VCL Media Player

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อพัฒนาระบบรู้จำต่อจากงานวิจัยในอดีต
2. เพื่อเป็นแนวทางในการพัฒนาระบบรู้จำเสียงและภาพในแบบอื่นๆ
3. เพื่อเป็นแหล่งข้อมูลสำหรับผู้ที่จะทำวิจัยในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและหลักการที่เกี่ยวข้อง

ในบทที่ 2 นี้จะกล่าวถึงหลักการรู้จำเสียงและภาพโดยคร่าว โดยเสียงนั้นจะใช้วิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel Frequency Cepstral Coefficient: MFCC) และภาพนั้นจะใช้วิธีไอเกนเฟส หลังจากนั้นจะกล่าวถึงระบบโครงข่ายประสาทเทียมแบบต่าง ๆ ตั้งแต่หลักการเบื้องต้นจนมาถึงวิธีที่นิยมใช้ในปัจจุบัน

2.1 ระบบรู้จำเสียงพูด (Speech Recognition)

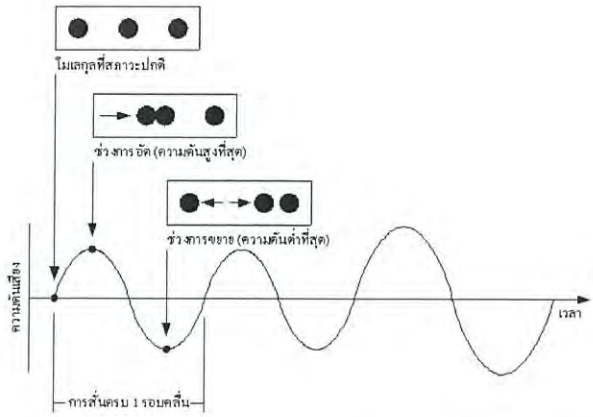
เบื้องต้นจะทำการศึกษาทฤษฎีของเสียง กระบวนการสร้างเสียงพูดและการได้ยินของมนุษย์เพื่อที่จะนำข้อมูลหรือพารามิเตอร์ในด้านเสียงพูดและการได้ยินของหูมาใช้ในการสร้างระบบรู้จำเสียงพูดที่สอดคล้องกับทฤษฎีของเสียง โดยทั่วไประบบรู้จำเสียงพูดจะมีหลักการทำงานที่สามารถแบ่งออกเป็น 4 ส่วนใหญ่ๆ [6] คือ

1. การเตรียมสัญญาณเสียงเบื้องต้น (Pre-Processing)
2. การดึงคุณลักษณะเด่นของสัญญาณเสียงพูด (Feature Extraction)
3. การเลือกคุณลักษณะเด่นของสัญญาณเสียงพูด (Feature Selection)
4. ระบบรู้จำ

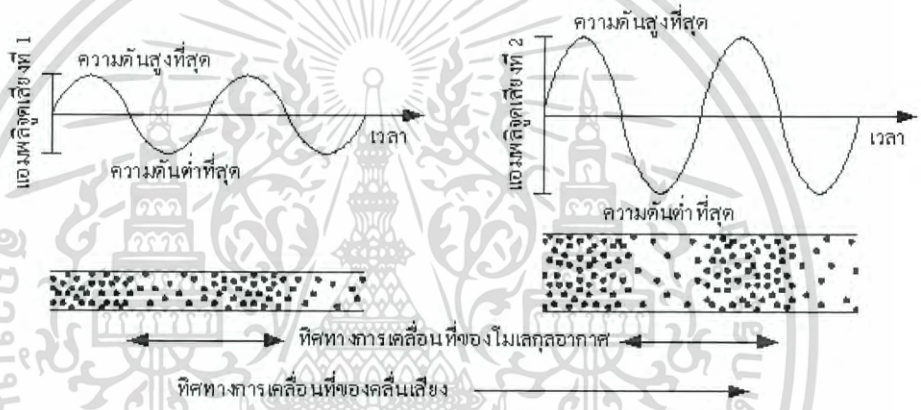
2.1.1 เสียงพูดและการได้ยินของมนุษย์

ทฤษฎีของการเกิดเสียงนั้น เสียงเกิดจากการที่วัตถุหรือแหล่งกำเนิดเสียงมีการสั่นสะเทือนโดยส่งผลต่อการเคลื่อนที่ของโมเลกุลของอากาศที่อยู่โดยรอบ ซึ่งโมเลกุลของอากาศเหล่านั้นจะเคลื่อนที่จากตำแหน่งแหล่งกำเนิดเสียงไปชนกับโมเลกุลของอากาศที่อยู่ถัดออกไป จะเกิดการถ่ายโอนโมเมนตัมจากโมเลกุลที่มีการเคลื่อนที่ไปให้กับโมเลกุลของอากาศที่อยู่ในสภาวะปกติ จากนั้นโมเลกุลที่ชนกันจะแยกออกจากกันโดยโมเลกุลของอากาศที่เคลื่อนที่มาชนจะถูกดึงกลับไปยังตำแหน่งเดิมด้วยแรงปฏิกิริยาและโมเลกุลที่ได้รับการถ่ายโอนพลังงาน ก็จะเคลื่อนที่ต่อไปและไปชนกับโมเลกุลของอากาศที่อยู่ถัดไปเป็นดังนี้ไปเรื่อย ๆ จนเคลื่อนที่ไปถึงหู เกิดการได้ยินขึ้น ปรากฏการณ์นี้จะเกิดสลับกันไปมาได้เมื่อสื่อกลางหรือตัวกลางคืออากาศซึ่งมีคุณสมบัติยืดหยุ่น การเคลื่อนที่ของโมเลกุลอากาศจะเกิดเป็นคลื่นเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 การเคลื่อนที่ของโมเลกุลของอากาศเทียบกับลักษณะของคลื่น

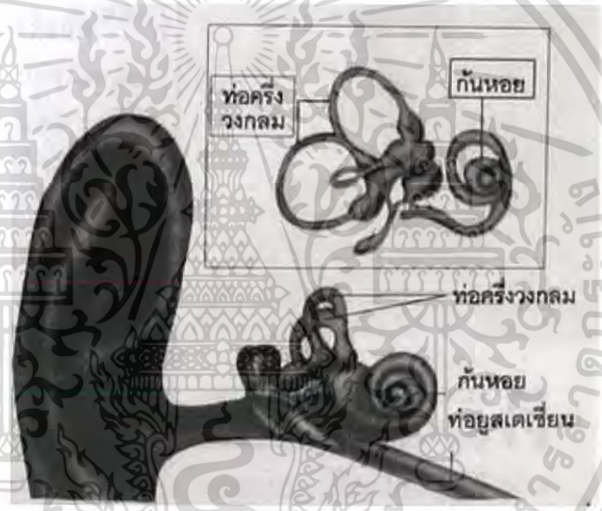


รูปที่ 2.2 จำนวนโมเลกุลของอากาศกับแอมพลิจูดของคลื่นเสียง

ต่อมาในกระบวนการสร้างเสียงพูดของมนุษย์นั้นจะเกิดจากลมที่เปล่งออกมาจากปอดผ่านอวัยวะต่าง ๆ ออกมาเป็นเสียงสูงและเสียงต่ำ โดยอวัยวะที่ทำให้เกิดเสียงก็คือกล่องเสียง เมื่อเวลาพูดเราจะเปล่งลมจากปอดผ่านหลอดคอผ่านกล่องเสียง ซึ่งภายในกล่องเสียงนี้จะมีสายเสียงซึ่งขึงอยู่ตรงกลางกล่องเสียง เมื่อลมผ่านจะทำให้สายเสียงสั้นสะเทือนเกิดเป็นเสียงสูงต่ำ ต่อมาเมื่อเสียงผ่านกล่องเสียงแล้วก็จะกระทบกับคอ เพดานปาก ฟัน และริมฝีปาก ทำให้ออกมาเป็นเสียงพูดในที่สุด โดยทั่วไปนั้นเสียงพูดของมนุษย์จะมีความถี่ของคลื่นเสียงอยู่ที่ 20-4000 เฮิรตซ์ ซึ่งต่อมามนุษย์ได้มีการสร้างภาษาขึ้นเพื่อใช้เสียงพูดนี้มาติดต่อสื่อสารกัน โดยที่เกิดจากการรวมหลาย ๆ หน่วยเสียง (phoneme) มาสร้างขึ้นมาเป็นคำ (word) จากคำหลาย ๆ คำมารวมเป็นประโยค (sentence) เป็นต้น (ซึ่งในภาษาทั่วไปหน่วยเสียงจะประกอบไปด้วย เสียงพยัญชนะและเสียงสระ สำหรับภาษาไทยจะมีหน่วยเสียงเพิ่มเติมขึ้นมา คือ เสียงวรรณยุกต์) และใช้เสียงพูดมาทำการออกเสียงที่สอดคล้องกับหลักในการสร้างภาษาต่าง ๆ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในด้านการรับรู้เสียงนั้น ๆ มนุษย์จะใช้อวัยวะที่เรียกว่า หู ในการรับรู้เสียง ทำให้สามารถแยกแยะได้ว่าเสียงที่เปล่งออกมาหรือที่ได้ยินนั้น ๆ มีองค์ประกอบของในแต่ละเสียงที่แตกต่างกันได้ กล่าวคือ หูของเราสามารถดึงคุณลักษณะเด่นของแต่ละเสียงออกมาได้ ซึ่งหูจะมีกลไกที่ค่อนข้างซับซ้อน โดยการเปลี่ยนเสียงเป็นกระแสประสาทเพื่อให้สมองแปลออกมาเป็นการได้ยิน โดยส่วนประกอบในการรับรู้นั้นเริ่มแรกที่ใบหูจะทำหน้าที่คล้ายจานรับสัญญาณเพื่อที่จะป้อนคลื่นเสียงเข้าไปในรูหู โดยที่สุดรูหูจะมีแก้วหูซึ่งเป็นแผ่นเยื่อบาง ๆ กั้นอยู่ เมื่อคลื่นเสียงมากระทบแก้วหู จะทำให้แผ่นแก้วหูนี้เกิดการสั่น โดยแผ่นแก้วหูนี้จะติดกับกระดูกชิ้นเล็ก ๆ 3 ชิ้น ที่ทำหน้าที่ในการขยายคลื่นเสียงให้แรงขึ้น ต่อมาในกระดูกชิ้นสุดท้ายจะมีรูปคล้ายโกลนของอานม้า โดยจะมีฐานของโกลนติดกับแผ่นเยื่อบาง ๆ อีกแผ่นหนึ่ง ซึ่งปิดปากท่อที่ขดเป็นรูปก้นหอยฝังอยู่ในกะโหลกศีรษะและภายในก้นหอยมีของเหลวบรรจุอยู่ เมื่อเสียงมากระทบจะทำให้เกิดคลื่นเสียงที่ผ่านมาถึงของเหลวในบริเวณนี้ ต่อมาภายในก้นหอยจะมีเซลล์ประสาทที่รับเสียงในความถี่ต่าง ๆ กัน โดยเสียงที่มากระทบจะทำให้เกิดกระแสประสาทไปที่สมอง ผ่านประสาทหูและสมองจะแปลออกมาให้เป็นการได้ยิน โดยการรับรู้เสียงของมนุษย์เราสามารถรับความถี่ของเสียงได้อยู่ที่ประมาณ 20-20000 เฮิรตซ์



รูปที่ 2.3 โครงสร้างอวัยวะในการรับรู้เสียงของหู

2.1.2 การเตรียมสัญญาณเสียงเบื้องต้น

การเตรียมสัญญาณเสียงเบื้องต้นถือเป็นขั้นตอนแรกและเป็นขั้นตอนที่มีความสำคัญขั้นตอนหนึ่ง เนื่องจากขั้นตอนนี้จะทำหน้าที่ปรับสัญญาณเสียงอินพุตที่เข้ามาให้อยู่ในรูปแบบที่เหมาะสมก่อนที่จะนำไปใช้ในขั้นตอนการดึงคุณลักษณะเด่น, การเลือกคุณลักษณะเด่นของสัญญาณและการรู้จำต่อไป ซึ่งปกติขั้นตอนการเตรียมสัญญาณเสียงเบื้องต้นจะมีวิธีการหลัก ๆ ดังนี้

1.) การกำจัดสัญญาณเสียงรบกวน (Noise) ด้วยการกรองความถี่ (Filtering)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.) การตัดคำของสัญญาณเสียง (Endpoint Detection) คือ การดึงสัญญาณเสียงที่เป็นเสียงจริง ๆ เท่านั้น โดยองค์ประกอบอื่นนอกจากสัญญาณเสียงจะทำการตัดออก

2.1.3 การดึงคุณลักษณะเด่นของสัญญาณเสียงพูด

นำสัญญาณเสียงที่ได้มาจากการเตรียมสัญญาณเสียงเบื้องต้น มาหาคุณลักษณะเด่นของสัญญาณเสียง ซึ่งเป็นตัวแทนของสัญญาณเสียงนั้น ๆ

2.1.4 การเลือกคุณลักษณะเด่นของสัญญาณเสียงพูด

คือการเลือกเก็บข้อมูลของสัญญาณเสียงเฉพาะข้อมูลสำคัญ หรือเป็นข้อมูลที่สามารถบอกหรือแยกแยะความแตกต่างของแต่ละคำพูดได้ รวมถึงการปรับพารามิเตอร์เพื่อเตรียมคุณลักษณะเด่นของสัญญาณให้เป็นคุณลักษณะเด่นของสัญญาณที่ดี ที่ส่งผลให้ระบบรู้จำมีประสิทธิภาพที่ดีขึ้น

2.1.5 การรู้จำเสียงพูด

ข้อมูลของสัญญาณเสียงที่ได้จากการเลือกคุณลักษณะเด่นของสัญญาณเสียงในหัวข้อที่ผ่านมา จะนำมาเข้าระบบรู้จำเสียงพูดซึ่งจะใช้การดึงคุณลักษณะเด่นด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้า (Delta Coefficient) และค่าสัมประสิทธิ์ดับเบิล-เดลต้า (Double Delta Coefficient) และใช้กระบวนการรู้จำด้วยโครงข่ายประสาทเทียม โดยใช้กระบวนการเรียนรู้แบบแพร่กลับ ซึ่งจะทำการประมวลผลในโปรแกรม MATLAB Professional

2.2 ไอเกนเฟส

ไอเกนเฟซ เป็นชื่อเรียกเซตของไอเกนเวกเตอร์ ซึ่งใช้ในระบบการรู้จำใบหน้า ตัวกลุ่มของเวกเตอร์นี้ พัฒนาขึ้นโดย สิโรวิชย์ และ เคอร์บี้ ในปี พ.ศ. 2530 ถูกนำมาใช้แยกแยะลักษณะในหน้ามนุษย์เป็นครั้งแรกโดย แมทธิว เตริก และ อเล็กซ์ เพนท์แลนด์ รูปแบบการใช้งานทำโดยเปรียบเทียบลักษณะของภาพกับ เวกเตอร์ในเบสิคเซต ซึ่งขั้นตอนหลักๆของไอเกนเฟสจะประกอบไปด้วย การหาค่า Covariance matrix, ไอเกนเว็กเตอร์, ไอเกนแวลูและการหาค่ายุคลิด โดยที่ค่าไอเกนเว็กเตอร์และไอเกนแวลูก็คือคุณลักษณะเด่นของสัญญาณภาพนั่นเอง

เบื้องต้น การดึงค่าไอเกนเว็กเตอร์และไอเกนแวลูนั้นจะได้มาจากสมการ Covariance matrix โดยที่สมการของ Covariance matrix จะเป็นดังสมการที่ 2.1

$$C = AA^T \tag{2.1}$$

โดยที่ A คือ เมตริกซ์ใดๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถแสดงสมการไอเกนเว็กเตอร์และไอเกนแวลูได้ดังสมการที่ 2.2

$$Xv = \lambda v \tag{2.2}$$

โดยที่ X คือ เมตริกซ์จตุรัสใดๆ

v คือ ไอเกนเว็กเตอร์

λ คือ ไอเกนแวลู

สามารถจัดสมการให้อยู่ในรูประบบสมการเอกพันธ์ *(Homogeneous System)* ได้ดังสมการที่ 2.3

$$(\lambda I - A)v = \bar{0} \tag{2.3}$$

ซึ่งระบบสมการดังกล่าวจะมีผลเฉลยที่ไม่เป็นศูนย์เมื่อดีเทอร์มินันต์ของเมตริกซ์สัมประสิทธิ์ไม่เป็น 0 ดังสมการที่ 2.4

$$|\lambda I - A| = 0 \tag{2.4}$$

โดยสามารถแสดงตัวอย่างการคำนวณไอเกนเว็กเตอร์ได้ดังนี้

ตัวอย่าง จงหาค่าไอเกนเว็กเตอร์ของ A เมื่อ

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

วิธีทำ กำหนดให้ λ เป็นไอเกนแวลูของ A

จากสมการ $Av = \lambda v$

ดังนั้น $$\begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\left(\begin{bmatrix}\lambda & 0\\ 0 & \lambda\end{bmatrix}-\begin{bmatrix}3 & 0\\ 8 & -1\end{bmatrix}\right)\begin{bmatrix}v_1\\ v_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

$$\left(\begin{bmatrix}\lambda-3 & 0\\ -8 & \lambda+1\end{bmatrix}-\begin{bmatrix}3 & 0\\ 8 & -1\end{bmatrix}\right)\begin{bmatrix}v_1\\ v_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

ผลลัพธ์จะไม่เป็น 0 เมื่อ

$$\begin{vmatrix}\lambda-3 & 0\\ -8 & \lambda+1\end{vmatrix}=0$$

$$(\lambda-3)(\lambda+1)-(-8)0=0$$

$$(\lambda-3)(\lambda+1)=0$$

$$\lambda=3,-1$$

ต่อมา ทำการคำนวณหาความสัมพันธ์ของ Covariance matrix เพื่อดึงค่าไอเกนเว็กเตอร์และไอเกนแวลู ที่เป็นคุณลักษณะเด่นของภาพได้ดังสมการที่ 2.25 ซึ่งสามารถแสดงความสัมพันธ์ได้ดังนี้ดังนี้

กำหนดให้ $C'=A^TA$

$$C=AA^T$$

จากสมการที่ 2.2 สามารถแสดงความสัมพันธ์ของไอเกนเว็กเตอร์ ไอเกนแวลูกับ Covariance matrix ได้ดังสมการที่ 2.5 ถึง 2.8

$$C'v=\lambda v \tag{2.5}$$

$$A^TAv=\lambda v \tag{2.6}$$

คูณ A ทั้ง 2 ข้างของสมการที่ 2.23 และแสดงได้ดังสมการที่ 2.24

$$AA^T(Av)=\lambda(Av) \tag{2.7}$$

$$C(Av)=\lambda(Av) \tag{2.8}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การตรวจจับใบหน้าด้วยวิธีการ Viola-Jones

ระบบวิเคราะห์ใบหน้าถือว่าเป็นหนึ่งในระบบที่ใช้ในการพิสูจน์ยืนยันตัวตนบุคคลโดยใช้คุณลักษณะจำเพาะทางสรีระ (Biometric) โดยระบบรู้จำภาพใบหน้าจะทำงานโดยการเปรียบเทียบใบหน้าจากภาพถ่ายดิจิตอลหรือภาพจากกล้องวีดีโอของบุคคลที่เราสนใจกับฐานข้อมูลใบหน้าที่มีอยู่ และเมื่อเปรียบเทียบเสร็จก็จะแสดงผลใบหน้าที่อยู่ในฐานข้อมูลที่มีใบหน้าเหมือนกับภาพที่นำมาเปรียบเทียบออกมา ระบบรู้จำใบหน้านั้นได้ถูกพัฒนามาอย่างต่อเนื่องเป็นเวลามากกว่าสิบปีมาแล้ว

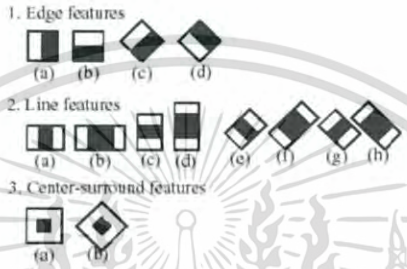
เนื่องจากเป็นระบบที่ได้รับความสนใจมากจากนักวิชาการหลายสาขาวิชาจึงทำให้ระบบรู้จำภาพใบหน้า มีผู้คนสนใจศึกษาและพัฒนากันอย่างมากมายจนทำให้มีการพัฒนาอัลกอริทึมในการทำงานของระบบออกมามากมายหลายรูปแบบแตกต่างกันไป ซึ่งการพัฒนาอัลกอริทึมก็แตกต่างกันไปตามยุคสมัยด้วย อันเนื่องมาจากปัจจัยด้านองค์ความรู้และเทคโนโลยีของอุปกรณ์ต่างๆ ที่พัฒนาขึ้นให้มีความเหมาะสมที่จะนำมาใช้ในระบบ จึงทำให้ต้องออกแบบอัลกอริทึมใหม่ให้เหมาะสมกับอุปกรณ์ใหม่ ๆ ด้วย ในปัจจุบันระบบรู้จำภาพใบหน้าได้มีการพัฒนาไปอย่างมาก ทำให้ระบบรู้จำภาพใบหน้ามีความน่าเชื่อถือมากขึ้น จนมีการนำระบบรู้จำภาพใบหน้ามาใช้ประโยชน์กันอย่างแพร่หลายในต่างประเทศ หลายๆ ประเทศได้มีการนำระบบรู้จำภาพใบหน้ามาติดตั้งในสนามบินเพื่อป้องกันคนร้ายหนีเข้าออกนอกประเทศ และมีระบบรู้จำภาพใบหน้าสำหรับการยืนยันตัวคนร้ายในคดีต่างๆด้วย

ระบบรู้จำภาพใบหน้า ถูกออกแบบมาให้ทำการเปรียบเทียบใบหน้าบุคคลที่เราสนใจกับฐานข้อมูลใบหน้าที่มีอยู่โดยอัลกอริทึมที่ใช้ในขั้นตอนการสร้างแม่แบบและขั้นตอนการเปรียบเทียบอาจแตกต่างกันไปแล้วแต่การออกแบบระบบของแต่ละระบบ แต่ไม่ว่าจะมี อัลกอริทึมในการทำงานในขั้นตอนการสร้างแม่แบบและขั้นตอนการเปรียบเทียบยังไง แต่ขั้นตอนการทำงานโดยรวมของระบบก็ยังคงเหมือนกันอยู่ โดยทั่วไประบบรู้จำภาพใบหน้าจะประกอบไปด้วย 2 ขั้นตอนหลักคือ การตรวจจับภาพใบหน้า (Face Detection) และ การรู้จำภาพใบหน้า ซึ่งการการตรวจจับภาพใบหน้า คือกระบวนการค้นหาใบหน้าของบุคคลจากภาพหรือวิดีโอ หลังจากนั้นก็จะทำการประมวลผลภาพใบหน้าที่ได้สำหรับขั้นตอนถัดไปเพื่อให้ภาพใบหน้าที่ตรวจจับได้ง่ายต่อการจำแนก และอัลกอริทึมที่นิยมใช้ในปัจจุบันนั้นมีชื่อว่า Viola-Jones

หลักการพื้นฐานของเทคนิคการตรวจจับภาพใบหนาของ Viola-Jones คือการนำภาพที่ต้องการตรวจจับมาแบงเปนภาพยอย (Sub-window) จากนั้น ภาพยอยดังกลาวจะถูกพิจารณาเปนภาพอินพุตของกระบวนการตรวจหาใบหนา เทคนิคทั่วไปในการตรวจหาใบหนาจะทำการปรับขนาดของภาพอินพุตที่แตกตางกันหลายๆขนาด และใชตัวตรวจจับ (Detector) ที่มีขนาดคงที่คนหาวัตถุ ขอเสียของการตรวจจับใบหนาแบบนี้คือ ระยะเวลาในการคำนวณไมคงที่ ดังนั้น Viola-Jones จึงเสนอเทคนิคการตรวจจับใบหนาแบบใหมโดยใชการจำลองรูปแบบ Haar เปนตัวตรวจจับทำการปรับขนาดของตัวตรวจจับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แทนการปรับขนาดของภาพอินพุต และใชตัวตรวจจับมาทำการตรวจจับใบหน่าหลายๆรอบ โดยแตละรอบจะใชขนาดของตัวตรวจจับที่แตกตางกัน ซึ่งเมื่อทำการเปรียบเทียบกับวิธีเดิม พบวาเวลาที่ใชในการคำนวณไมไดแตกตางกันมาก แต่เวลาในการคำนวณการตรวจจับภาพใบหน่าแตละรอบจะมีคาคงที่แม ขนาดของตัวตรวจจับจะแตกตางกันก็ตาม เทคนิคที่ Viola-Jones นำเสนอนี้จะทำการคำนวณการจำลองรูปแบบ Haar ดวยหลักการ Integral ภาพใบหน้า ซึ่งเป็นวิธีการตรวจจับและตีความวัตถุภายในภาพ ด้วยการสร้างคุณลักษณะเด่นที่แสดงถึงผลต่างระหว่างพื้นที่ส่วนสีขาวและส่วนที่เป็นสีดำของภาพ ซึ่งคุณลักษณะเด่นนี้นั้นสามารถเปลี่ยนแปลงขนาดและตำแหน่งได้ โดยใช้สำหรับการตรวจจับลักษณะบนภาพแบบต่าง ๆ เช่น เส้นตรง,วงกลม เป็นต้น



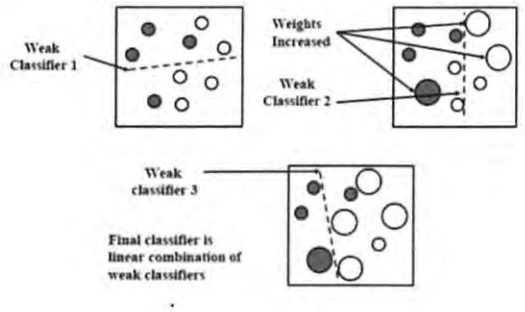
รูปที่ 2.4 รูปแบบของคุณลักษณะเด่นสำหรับการตรวจจับลักษณะในแบบต่างๆ

ในการทำ Haar นั้นจำเป็นต้องมีภาพตัวอย่างจำนวนมาก ซึ่งมีสองลักษณะคือ Positive Image คือรูปที่มี Object นั้นๆประกอบอยู่ภายในภาพ และ Negative Image คือภาพใดๆที่ไม่มี Object ที่เราต้องการอยู่ภายในภาพ ด้วยหลักการของ AdaBoost (Adaptive Boost ซึ่งสามารถแสดงได้ดังรูปที่ 2.18) โดยเป็นกระบวนการหาค่าคุณลักษณะเด่นที่มีลักษณะใกล้เคียงและแตกต่างกับภาพอินพุต สำหรับขั้นตอนการทำนั้นมีดังนี้

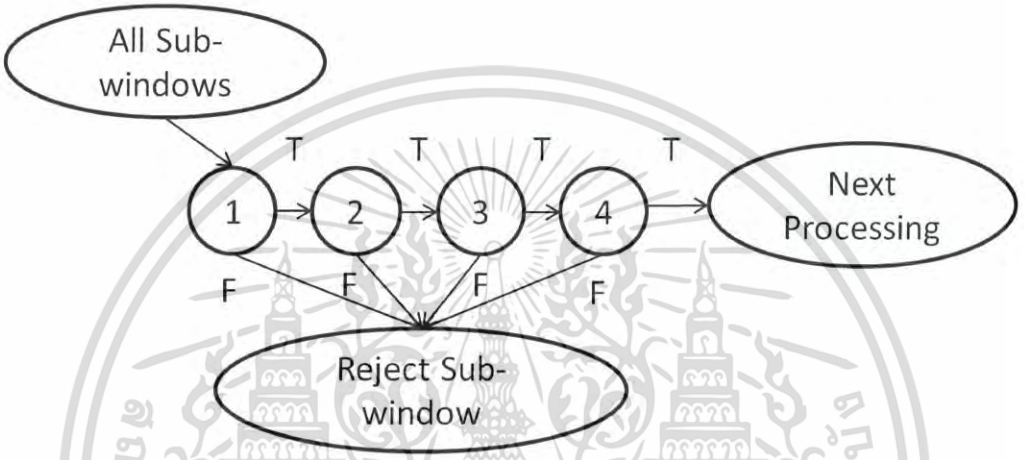
1. กำหนดค่าน้ำหนักให้กับคุณลักษณะเด่นภายในภาพตัวอย่าง
2. หาบริเวณที่ประกอบด้วยส่วนที่เราต้องการ
3. เพิ่มค่าน้ำหนักให้กับส่วนที่เหลือ เฉพาะลักษณะที่เราต้องการ โดยที่ยังไม่ได้แบ่งลักษณะไว้
4. ทำวนเช่นนี้ซ้ำไปเรื่อยๆ จนท้ายที่สุดจะนำบริเวณที่ได้ทั้งหมดมารวมกัน ซึ่งจะได้บริเวณของ Object ที่ เราต้องการหา และลักษณะในส่วนต่างๆภายใน Object นั้น

Cascade Classifiers เป็นการตีความหมายของภาพ ตามลักษณะภายในภาพ โดยตัดส่วนของ Sub window ที่เป็น Negative ออกไปก่อน แล้วใช้ส่วนที่เป็น Positive ภายในภาพ โดยเปลี่ยนลักษณะการตรวจจับไปเรื่อยๆ จนสามารถระบุได้ว่าภาพดังกล่าวเป็นภาพอะไร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 การทำงานของ Adaboost



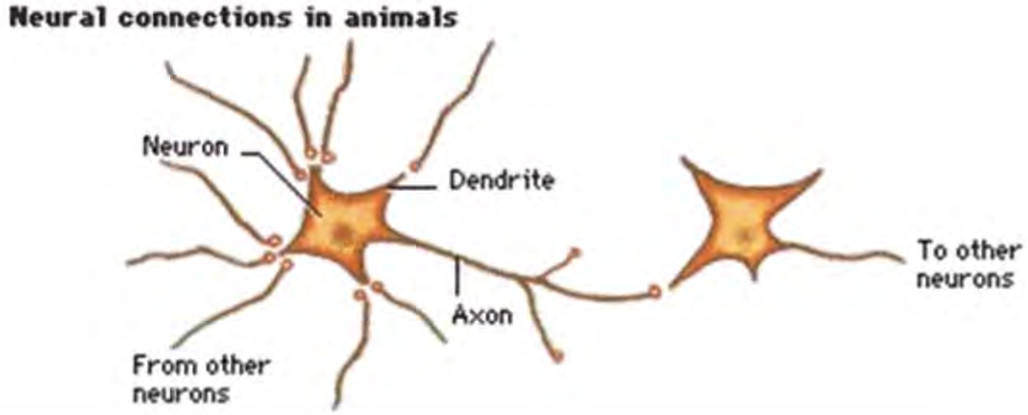
รูปที่ 2.6 การทำงานของ Haar Cascade Classifier

2.4 โครงข่ายประสาทเทียมเบื้องต้น (Artificial Neural Networks: ANN)

โครงข่ายประสาทเทียมหรือที่เรียกกันว่า Artificial Neural Networks (ANN) เป็นแบบจำลองแบบหนึ่งใน แขนงของงานวิจัยทางด้านปัญญาประดิษฐ์หรือ Artificial Intelligence (AI) โดยมีการอ้างอิงมาจากการทำงานของเซลล์ประสาทในสมองของมนุษย์ แต่ในโครงข่ายของเซลล์ประสาทในสมองของมนุษย์จริงมีขั้นตอนการทำงานที่ซับซ้อนกว่ามาก แต่กระนั้น ANN ก็ยังสามารถที่จะนำมาใช้ประโยชน์ได้จริงในงานด้านวิทยาการคอมพิวเตอร์และวิศวกรรมโดยทั่วไป

ในสมองของมนุษย์นั้น จะประกอบไปด้วยเซลล์ประสาทหรือ Neuron ประมาณ 10^{11} ตัว โดยที่เซลล์ประสาทหนึ่งๆ จะมีการเชื่อมโยงไปยังเซลล์ประสาทอื่นๆ อีกประมาณ 10^{4} และเวลาที่ใช้ในการแลกเปลี่ยนข้อมูลระหว่าง Neurons นั้นจะอยู่ในหลัก 10^{-3} วินาที โดยการทำงานของ Neurons ในสมองมนุษย์นั้นจะมีลักษณะการประมวลผลแบบขนานขั้นสูงอีกด้วย (Highly Parallel Processing) จึงทำให้มนุษย์สามารถเรียนรู้และจดจำใบหน้าพ่อแม่ของตัวเองได้ภายในเวลาเพียงไม่กี่วินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 รูปแบบของเซลล์สมอง

เป้าหมายหลักของการนำเอา ANN มาใช้ในงานด้านวิศวกรรมและวิทยาการคอมพิวเตอร์นั้น คือ การพัฒนากระบวนการเรียนรู้ของเครื่องจักร หรือที่เรียกกันว่า Machine Learning ให้มีประสิทธิภาพสูง การนำ ANN มาใช้งานแบ่งเป็น 2 ขั้นตอนหลัก คือ ขั้นตอนการฝึกหัด (Training) ขั้นตอนการเรียนรู้ (Learning) และขั้นตอนการทดสอบ (Testing)

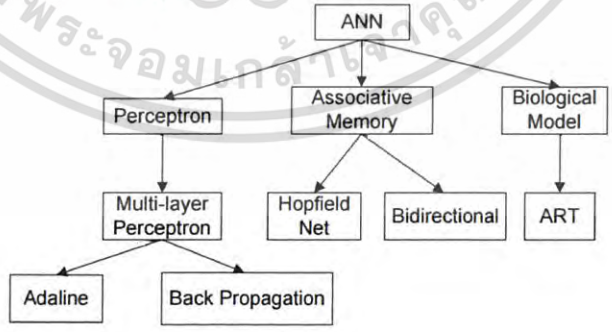
แบบจำลองของ ANN ที่ใช้กันโดยทั่วไปสามารถจำแนกได้ 3 กลุ่มใหญ่ๆ ดังนี้

1.) Perceptron : Multi-layer Perceptron

2.) Associative Memory : Hopfield Net , Bidirectional

3.) Biological Model : ART

โดยสามารถแบ่งย่อยลงไปได้อีกดังรูปที่ 2.8



รูปที่ 2.8 แสดงแบบจำลองหลักๆ ของ ANN

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณสมบัติและความสามารถที่เป็นประโยชน์ของ ANN

1.) มีความยืดหยุ่นสูงจนสามารถจำลองขบวนการของปัญหาใดๆ ได้

2.) มีความสามารถในการจำลองของคู่ input-output ที่มีความซับซ้อนมากจนไม่สามารถจำลองแบบในเชิงของความหน้าจะเป็นได้

3.) มีความสามารถในการปรับตัวเข้ากับการเปลี่ยนแปลงของสิ่งแวดล้อม

4.) มีความสามารถในการตอบสนองต่อข้อมูลที่ไม่เคยเห็น

5.) ความรู้จะกระจายอยู่ทั่วโครงสร้างของโครงข่ายประสาทเทียม

ซึ่งลักษณะของปัญหาที่เหมาะสมกับ ANN ส่วนใหญ่ๆจะมีดังนี้

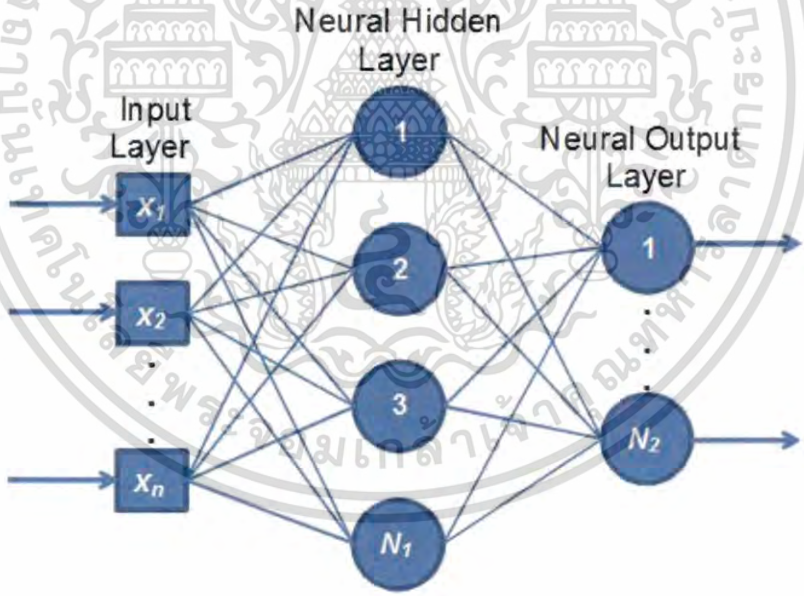
1.) ตัวอย่างข้อมูลที่ประกอบไปด้วย Attribute หลายๆ ค่า

2.) ตัวอย่างข้อมูลสามารถมีความผิดพลาดปนอยู่ได้

3.) ไม่มีปัญหากับขั้นตอนฝึกหัด เพราะอาจจะใช้เวลานานมาก

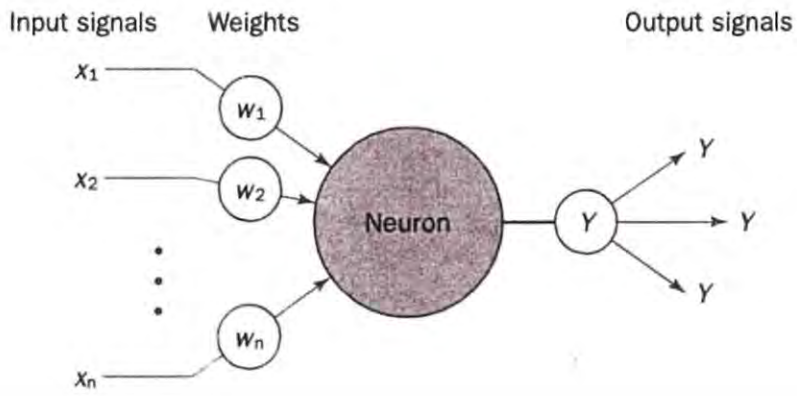
4.) ต้องการประเมินผล Output อย่างรวดเร็ว

5.) ไม่ต้องการเข้าใจวิธีการแก้ปัญหา ต้องการเพียงแค่ให้ปัญหาได้รับการแก้ไขเท่านั้น



รูปที่ 2.9 ตัวอย่างของโครงข่ายประสาทเทียมที่มีโหนดซ่อนเร้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.10 บล็อกไดอะแกรมของ Neuron อย่างง่าย

ซึ่งตามแบบจำลองที่นำเสนอโดย McCulloch and Pitts โดยถือเป็นแบบจำลอง Neuron ในยุคแรก แสดงไว้ดังแสดงในสมการที่ 2.9

$$X = \sum_{i=1}^{n} x_i w_i \tag{2.9}$$

โดยที่ X คือ net weighted input to the neuron

จากนั้น พิจารณาต่อ ดังนี้ (ซึ่งถือเป็น Transfer Characteristic หรือนิยมเรียกว่า Activation Function) ซึ่งดูได้จากสมการที่ 2.10

$$Y = \begin{cases} +1; X \geq \theta \\ -1; X < \theta \end{cases} \tag{2.10}$$

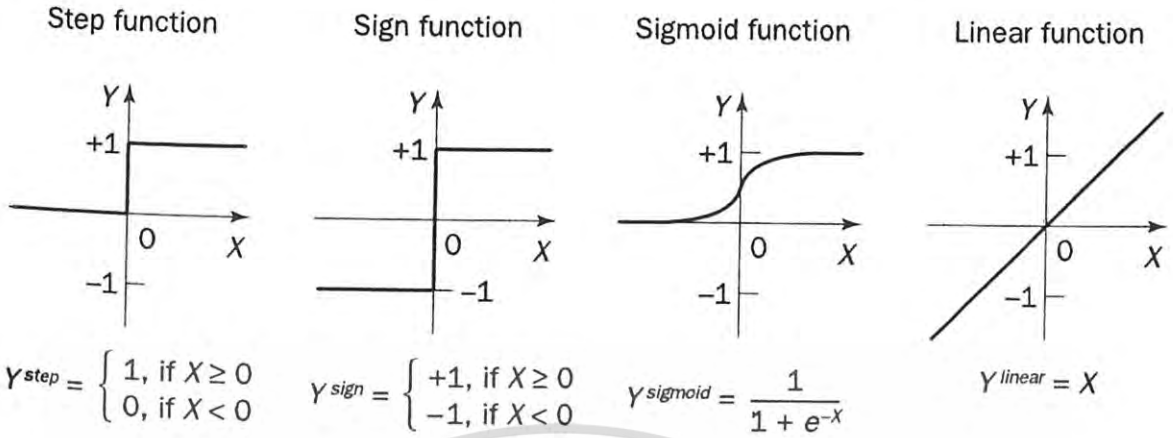
โดยที่ θ คือ ค่าเกน

ซึ่ง Activation Function ดังกล่าวนี้จะเรียกว่า Sign Function หรือ Signum Function ดังนั้น เราสามารถแสดง Actual Output ของ neuron ที่ใช้ Sign Activation Function ได้ดังสมการที่ 2.11

$$Y = sign\left[\sum_{i=1}^{n} x_i w_i - \theta\right] \tag{2.11}$$

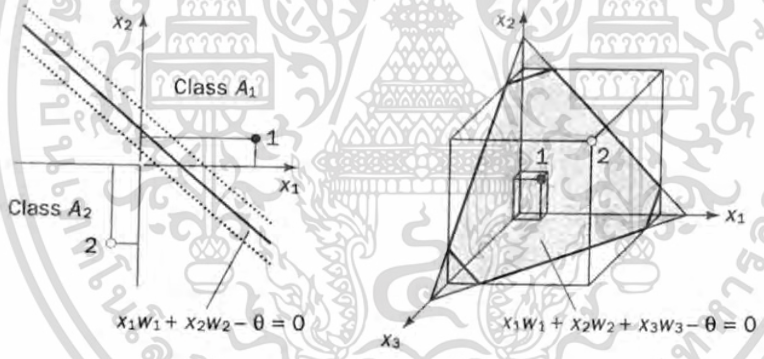
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่ง Activation Function ในแบบต่างๆ ที่นิยมใช้ในโครงข่ายประสาทเทียมแสดงได้ดังรูปที่ 2.11



รูปที่ 2.11 ตัวอย่าง Activation Function แบบต่างๆ

โดยพื้นฐานนั้น ผู้ศึกษาจำเป็นต้องเรียนรู้วิธีการใช้งาน ที่เป็นเพียง Single-Layer ก่อน โดยใช้วิธีการ Linear Perceptron ซึ่งมีทั้งหมด 4 ขั้นตอน ซึ่ง Linear Perceptron นั้น มี Hyperplane ซึ่งใช้ในการแบ่ง Decision Region จะถูกนิยามโดย Linearly Separable Function ดังรูปที่ 2.12



รูปที่ 2.12 Linear Separability in the Perceptron

ซึ่งมีขั้นตอนการทำทั้งหมด 4 ขั้นตอน ดังนี้

1. Initialization

กำหนดค่า initial weights $w_1, w_2, w_3, \ldots, w_n$ และค่า Threshold ให้มีค่าแบบสุ่มในช่วง [-0.5,0.5]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. Activation the Perceptron

โดยการ apply inputs $x_1(p), x_2(p), x_3(p), ..., x_n(p)$ และกำหนด desired output (ผลลัพธ์ที่เราต้องการในระบบรู้จำ) Y(p) จากนั้นกำหนดค่า actual output ที่ iteration p = 1 โดยใช้สมการที่ 2.12

$$Y = step\left[\sum_{i=1}^{n} x(p)w(p)\right] - x \tag{2.12}$$

step คือการ activation function ปัดเป็น 0 หรือ 1 เท่านั้น

3. Weight Training

กำหนดค่า error signal ซึ่งค่านี้จะทำให้รู้ว่าข้อมูลในจุดไหนเกิด error ได้ แทนโดย e(p) ซึ่งแสดงดังสมการที่ 2.13

$$E(p) = Y(P) - Y_{real}(p) \tag{2.13}$$

โดยที่ $Y(P)$ คือ desired output, $Y_{real}(p)$ คือ actual output (คือค่าที่ได้จากการทดลอง)

กำหนดค่า Learning Rate (α) ซึ่งจะเป็นค่าคงที่ ที่ทำให้ระบบสามารถทำการรู้จำได้ โดยถ้ายิ่ง α มีค่าเข้าใกล้ศูนย์ ความสามารถในการรู้จำก็จะแม่นยำขึ้น แต่ก็แลกมาด้วยระบบจะมีดีเลย์มากขึ้น ซึ่งจะได้ดังสมการที่ 2.14

$$\Delta w(p) = \alpha \times x(p) \times e(p) \tag{2.11}$$

โดยที่ ค่า $\Delta w(p)$ คือค่า weight correction factor และเสร็จแล้วจึงทำการ update weights ของ Perceptron ดังสมการที่ 2.12

$$w(p+1) = w(p) - \Delta w(p) \tag{2.14}$$

4. Iteration

ทำการเพิ่ม iteration p ไปอีก 1 จากนั้นกลับไปทำใน step ที่ 2 และ repeat process จนระบบเข้าสู่สภาวะเสถียร

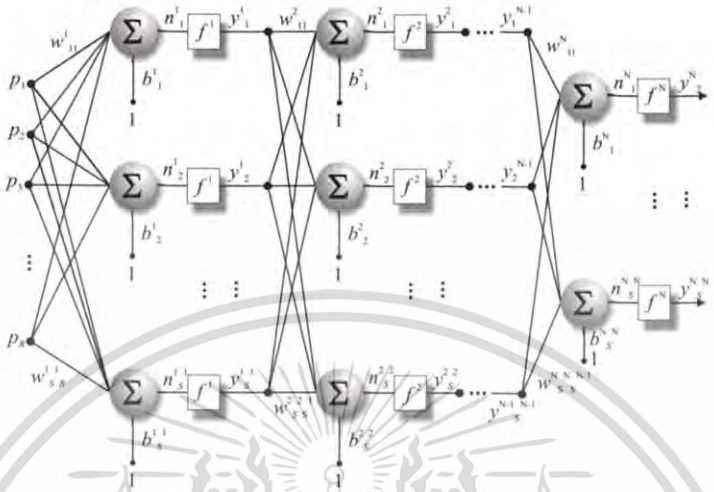
โครงข่ายประสาทเทียมโดยใช้กระบวนการเรียนรู้แบบแพร่กลับ (Back Propagation)

เนื่องจากระบบรู้จำแบบ Linear Perceptron ใช้ในการฝึกสอนเครือข่ายที่มีโครงสร้างเพียงชั้นเดียว (แต่มีหลายนิวรอนได้) เครือข่ายดังกล่าวนั้น ล้วนแต่มีข้อจำกัดที่สามารถแก้ปัญหาที่แบ่งแยกได้เชิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เส้น (Linearly Separable) เท่านั้น จึงได้มีการคิดค้นเครือข่ายแบบหลายชั้นขึ้นมา ซึ่งชนิดหนึ่งที่นิยมใช้กันอย่างแพร่หลายคือ Back Propagation ซึ่งสามารถแสดงเป็นบล็อกไดอะแกรมได้ดังรูปที่ 2.7

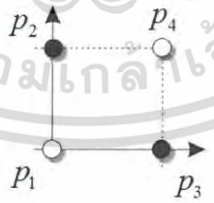
เครือข่ายหลายชั้น (Multilayer Network)



รูปที่ 2.13 บล็อกไดอะแกรมของ Multilayer Networks

จากรูปที่ 2.13 แสดงเครือข่าย N ชั้นเอาต์พุตของแต่ละชั้นจะเป็นอินพุตให้กับชั้นถัดไป แต่ละชั้นสามารถมีจำนวนนิวรอนแตกต่างกันได้ โดยปกติชั้นแรกจะเป็นชั้นอินพุต (input layer) ซึ่งรับอินพุตจากภายนอกเครือข่าย ในชั้นสุดท้ายจะเป็นชั้นเอาต์พุตสำหรับส่งค่าเอาต์พุตออกไปจากเครือข่าย ส่วนชั้นระหว่างอินพุตและเอาต์พุตเรียกว่าเป็น ชั้นซ่อนเร้น (Hidden Layer) เครือข่ายแบบหลายชั้นมีประสิทธิภาพเหนือกว่าแบบชั้นเดียวมาก ทำให้มีการประยุกต์ใช้เครือข่ายแบบหลายชั้นในการแก้ปัญหาหลายๆ อย่างได้อย่างมีประสิทธิภาพ

ซึ่งเครือข่ายหลายชั้นนี้ สามารถแก้ปัญหาที่ไม่สามารถแบ่งแยกได้แบบเชิงเส้น ยกตัวอย่างคือ Exclusive OR (XOR) ได้ดังรูปที่ 2.14

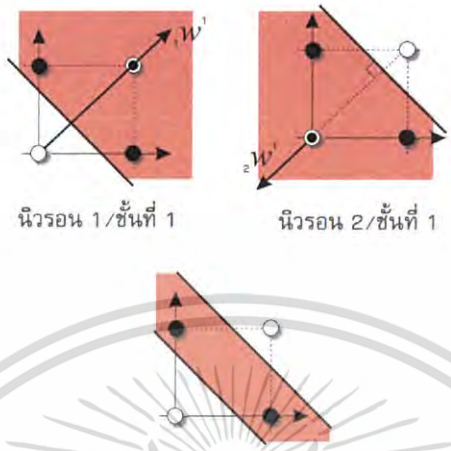


รูปที่ 2.14 ปัญหา Exclusive OR

โดยปกติแล้วเครือข่ายเพียง 2 ชั้นก็สามารถแก้ปัญหาได้ ตัวอย่างหนึ่งก็คือใช้เครือข่าย 2 ชั้นที่ชั้นแรกประกอบไปด้วย 2 นิวรอน เพื่อสร้างเส้นแบ่งพื้นที่ 2 เส้น เส้นแรกใช้แบ่งอินพุต p_1 ออกจากอินพุต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อื่นๆ และเส้นที่สองใช้แบ่งอินพุต p_1 ออก[3]. เครือข่ายชั้นที่สองใช้สำหรับรวมเส้นแบ่งพื้นที่จากชั้นแรกเข้าด้วยกันด้วยการกระทำ AND ดังนั้นเครือข่ายชั้นที่สองจึงใช้เพียงนิวรอนเดียวได้ดังรูปที่ 2.15

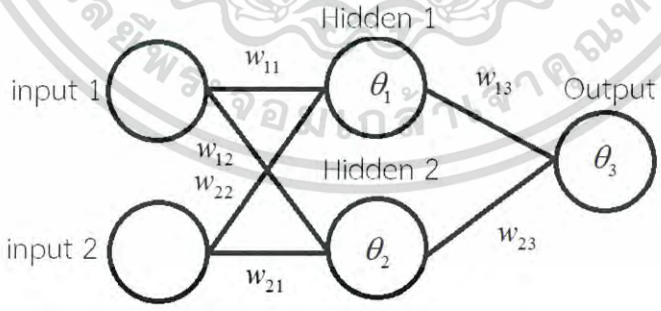


รูปที่ 2.15 รูปแบบการแก้ไขปัญหาของ Exclusive OR

ขั้นตอนการทำ Back Propagation สามารถแสดงได้ดังตัวอย่างจากรูปที่ 2.10 สามารถแสดงได้ดังนี้

Step1 : Initialization

คือกำหนดค่า Weight และค่า Threshold ทุกค่า โดยสุ่มให้มีค่าใกล้เคียงกัน



รูปที่ 2.16 ตัวอย่างของบล็อกไดอะแกรมปัญหา XOR

ยกตัวอย่างจากในรูป เราต้องกำหนด $w_{11}\ w_{12}\ w_{21}\ w_{22}\ w_{13}\ w_{23}, \theta_1$ ของ Hidden 1, θ_2 ของ Hidden 2 และ θ_3 ของ Output

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step2 : Activation

คือการหาค่าของ desired output (targets) โดยใช้หลักการเดียวกับ Linear Perceptron โดยการ apply inputs $x_1(p), x_2(p), ..., x_n(p)$ และกำหนด desired output $y_1(p), y_2(p), ..., y_n(p)$ ซึ่ง desired output ที่ได้จาก input นั้น จะกลายเป็น input ของ hidden layer ต่อไป โดยใช้สมการที่ 2.15

$$Y_j(p) = sigmoid \sum_{i=1}^{n} [x_i(p) \times w_{ij}(p) - \theta_j] \tag{2.15}$$

ซึ่งฟังก์ชัน sigmoid นั้น คือฟังก์ชันชนิดหนึ่งที่สามารถรับข้อมูลได้ช่วงกว้างๆ เป็นฟังก์ชันต่อเนื่องที่สามารถหาอนุพันธ์ได้ (ซึ่งจำเป็นกับการคำนวณหาค่าพารามิเตอร์ของโครงข่ายประสาทเทียมแบบแพร่กลับ)

โดยที่ sigmoid มีรูปแบบดังสมการที่ 2.16

$$sigmoid = \frac{1}{1 + e^{-((x_1 \times W11) + (x_1 \times W12) - \theta Hidden1)}} \tag{2.16}$$

ซึ่งหลังจากที่ได้ $Y(p)$ มาแล้ว เราใช้ $Y(p)$ เป็น input ของ hidden layer ต่อไปโดยใช้สมการที่ 2.17

$$Y_k(p) = sigmoid \sum_{j=1}^{m} [x_{ij}(p) \times w_{jk}(p) - \theta_k] \tag{2.17}$$

เสร็จแล้วทำการหาค่า error ของระบบได้ดังสมการที่ 2.18

$$e = y_{desired,output} - y_{output} \tag{2.18}$$

Step3 : Weight Training

ใน back propagation นั้น จุดเด่นของมันคือการหาค่า error ย้อนกลับเพื่อจะได้ค่า input ที่ดีกว่าเก่าโดยใช้สมการที่ 2.19

$$\delta_k(p) = y_k(p) \times [1 - y_k(p)] \times e_k(p) \tag{2.19}$$

ซึ่งค่า weight และ θ จะเปลี่ยนแปลงไปดังสมการที่ 2.20

$$\Delta w_{jk}(p) = \alpha \times y_j(p) \times \delta_k(p) \quad , \quad \theta_k = \alpha \times (-1) \times \delta_k \tag{2.20}$$

แล้วจึงทำการหาค่า error gradient ซึ่งแสดงตามสมการที่ 2.21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\delta_j(p) = y_j(p) \times [1 - y_j(p)] \times \sum_{k=1}^{l} \delta_k(p) \times w_{jk}(p) \quad (2.21)$$

เสร็จแล้วจึงหาค่า weight ที่เปลี่ยนไปของตัว input (คือการทำย้อนกลับไปเรื่อยๆ ถ้ามี Hidden Layer อยู่ N Layer ก็จำเป็นต้องทำย้อนกลับไป N ครั้ง) ดังสมการที่ 2.22

$$\Delta w_{ij}(p) = \alpha \times y_i(p) \times \delta_j(p) \quad (2.22)$$

ซึ่งค่าที่อัพเดตมาจะได้ดังสมการที่ 2.23 และ 2.24

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p) \quad (2.23)$$

$$w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p) \quad (2.24)$$

Step4 : Iteration

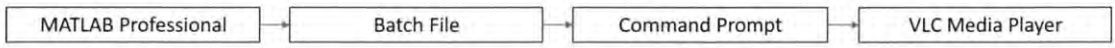
ทำการเพิ่มค่า p ขึ้นมา และย้อนกลับไปทำ step ที่ 2 ใหม่ จนกว่าจะได้ค่า Error ("Sum of Squared Errors") มีค่าน้อยกว่าที่กำหนด หลังจากนั้น ระบบจะหยุดทำงาน เราจะได้ค่าผลลัพธ์ Weight ออกมา ซึ่งเรานำ Weight นี้ มาใช้ในกระบวนการ Testing ต่อไป

2.3 การเชื่อมต่อระหว่างระบบรู้จำเพื่อเรียกไฟล์เพลงบนโปรแกรม VCL Media Player

โดยในปริญญานิพนธ์นี้จะทำการออกแบบและประมวลผลระบบรู้จำในโปรแกรม MATLAB Professional โดยผลลัพธ์จะไปทำการเปิดไฟล์เพลง (นามสกุลไฟล์ .mp3) ในโปรแกรม VCL Media Player ซึ่งถูกจำแนกเสียงในโครงข่ายประสาทเทียมตามเสียงพูดที่ได้ทำการรู้จำ

การสั่งการด้วยเสียงผ่าน MATLAB เพื่อเรียกไฟล์เสียงผ่านโปรแกรม VCL Media Player นั้น จำเป็นต้องมีการอินเตอร์เฟสระหว่าง 2 โปรแกรมเพื่อให้ระบบนั้นมีความสมบูรณ์แบบ โดยสิ่งที่นำมาอินเตอร์เฟสนั้นคือ Command Prompt โดย Command Prompt เป็นการเรียกใช้คำสั่งเพื่อการจัดการสิ่งต่าง ๆ ภายในระบบปฏิบัติการ โดยเรียกใช้ผ่าน text mode ซึ่งโดยปกติแล้วเราจะเรียกใช้โปรแกรมหรือระบบผ่านโหมดกราฟฟิก GUI ซึ่งสามารถทำงานได้ทุกอย่างได้ง่ายกว่า แต่ Command Prompt ถือเป็นสิ่งที่สำคัญมาก เพราะใช้จัดการทุกอย่างได้ ในขณะที่บางที่ GUI หรือในโหมดกราฟฟิกไม่สามารถทำการจัดการได้ นั่นคือจะทำการสั่งการด้วยเสียงใน MATLAB เพื่อไปทำการเปิด Command Prompt แล้วอ่านไฟล์ batch file ที่ทำการบรรจุที่อยู่ของไฟล์เพลงนั้น ๆ รวมถึงคำสั่งในการเปิดไฟล์เพลงผ่านโปรแกรม VCL Media Player ด้วย โดยสามารถทำการเขียนชุดคำสั่ง batch file นี้ได้ใน Notepad แล้วเซฟนามสกุลเป็น .bat

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.17 ขั้นตอนการสั่งงานโปรแกรม VLC Media Player เพื่อเรียกไฟล์เพลงด้วยโปรแกรม MATLAB Professional

2.3.1 MATLAB Professional

MATLAB เป็นภาษาคอมพิวเตอร์ระดับสูงที่ใช้สำหรับคำนวณเชิงตัวเลข (Numerical Computing) ซึ่งแสดงผลกราฟฟิก และเขียนแอพพลิเคชั่น ทำให้เราสามารถคำนวณผลลัพธ์ พัฒนาอัลกอริทึม สร้างแบบจำลอง และแอพพลิเคชั่นได้ง่ายและรวดเร็วมาก ภายในตัว MATLAB ประกอบ ด้วยภาษาคอมพิวเตอร์ ทูลบ็อกซ์[9]. (Toolbox คือกลุ่มฟังก์ชันสำเร็จรูปในแต่ละสาขาวิชา) และฟังก์ชันพื้นฐานจำนวนมาก ทำให้การวิเคราะห์ทำได้หลากหลายวิธี พร้อมกับคำตอบที่รวดเร็วกว่าโปรแกรมตารางคำนวณ (Spreadsheet) หรือภาษาคอมพิวเตอร์สมัยก่อน เช่น C, C++, Fortran, Java และอื่น ๆ

โดยสามารถนำ MATLAB ไปประยุกต์ใช้งานได้หลายสาขามาก ทั้ง การประมวลผลสัญญาณ (Signal Processing) การสื่อสาร (Communication) การประมวลผลภาพและวิดีโอ (Image and Video Processing) ระบบควบคุม (Control System) การวัดและควบคุม (Instruments and Control) การคำนวณทางเศรษฐศาสตร์ (Economic) การคำนวณทางชีววิทยา (Biology) และอื่น ๆ มีนักวิทยาศาสตร์และวิศวกรหลายล้านคนทั้งในแวดวงการศึกษาและอุตสาหกรรมที่ใช้ MATLAB ในการคำนวณเชิงตัวเลข

2.3.2 Batch File

Batch File เป็นชุดคำสั่งที่เขียนเพื่อสั่งให้คอมพิวเตอร์ทำงานตามคำสั่งผ่าน Command Prompt โดยจะใช้โปรแกรม Notepad ในการสร้าง Batch File ซึ่งสามารถทำได้โดยเขียนชุดคำสั่งที่ต้องการผ่าน Notepad แล้วทำการบันทึกไฟล์นามสกุล .bat เมื่อทำการดับเบิลคลิกไฟล์ที่บันทึกจะทำการรันไฟล์ที่ป้อนชุดคำสั่งไว้ผ่าน Command Prompt[10]. โดยอัตโนมัติ ดังรูปที่ 2.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.18 ตัวอย่าง Batch File ที่มีการชี้ไปไดร์ฟ C ที่โฟลเดอร์ Music for Project ตามด้วยโฟลเดอร์ Music2 และทำการเรียกเพลงชื่อ Jairuk ที่มีนามสกุลไฟล์เพลงเป็น .mp3

2.3.3 Command Prompt

หรืออีกชื่อที่เป็นที่รู้จักก็คือ Cmd.exe / Cmd เป็นโปรแกรมตัวหนึ่งที่เป็นส่วนหนึ่งของระบบปฏิบัติการ Windows ทำงานในรูปแบบ Command-line interpreter (ต้องสั่งงานด้วยชุดคำสั่งเป็นตัวอักษร ไม่ใช่กราฟฟิก GUI) โดยทุกวันนี้สำหรับการใช้งานคอมพิวเตอร์ในปัจจุบันจะเป็น GUI ทั้งหมด ในด้านการเรียกใช้งาน

2.3.4 โปรแกรม VLC Media Player

คือโปรแกรมเล่นไฟล์สื่อ เช่น เพลงและภาพเคลื่อนไหวได้หลายนามสกุลไฟล์ รวมทั้งไฟล์ที่เล่นบนมือถือ พัฒนาโดยโครงการ วิดีโอแลน (VideoLAN) โดยเป็นซอฟต์แวร์เสรีที่ใช้สัญญาอนุญาตแบบ GPL สำหรับเล่นไฟล์มีเดียต่าง ๆ โดยโครงการวิดีโอแลน

VLC Media Player เป็นซอฟต์แวร์เล่นไฟล์สื่อคุณภาพสูง สำหรับรับชมภาพและเสียง, บันทึกภาพและเสียง และการถ่ายทอดแบบสตรีม ซึ่งสนับสนุนไฟล์ในหลาย ๆ ประเภทที่รู้จักกันดีเช่น วีซีดี ดีวีดี และการสตรีม โพรโทคอล และยังสามารถสตรีมระหว่างเน็ตเวิร์ก รวมถึงยังสามารถแปลงไฟล์ได้ด้วยเช่นกัน วีแอลซี ย่อมาจาก วิดีโอแลน ไคลเอนต์ (VideoLan Client) และยังรองรับระบบปฏิบัติการที่หลากหลาย เช่น ไมโครซอฟท์ วินโดวส์ Mac OS X BeOS BSD Solaris

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.19 หน้าตาไอคอนโปรแกรม VCL Media Player

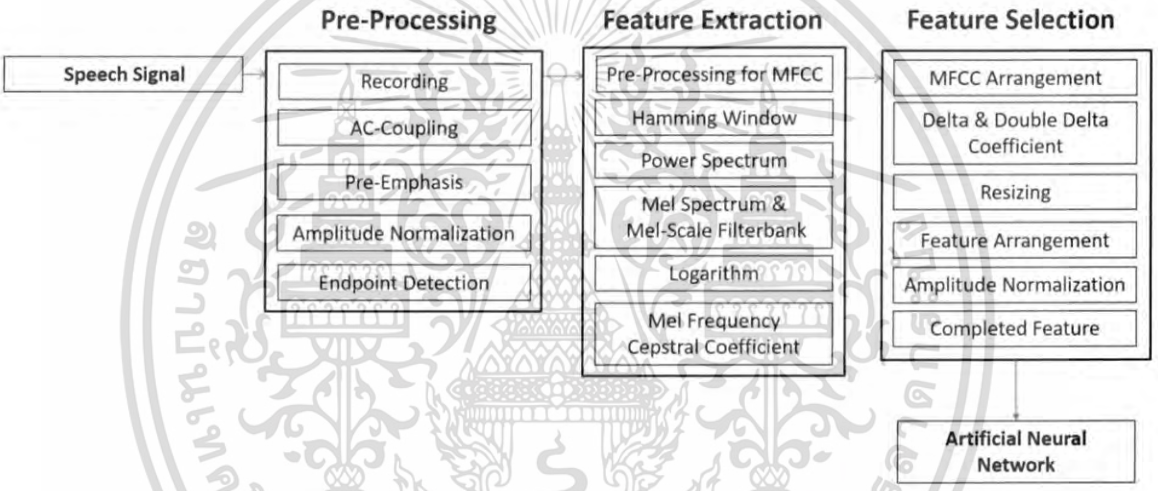
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบและจัดการปริญญานิพนธ์

3.1 การหาตัวแทนของสัญญาณเสียงเพื่อนำไปเรียนรู้ในระบบโครงข่ายประสาทเทียม

โดยการออกแบบนั้นจะทำการเตรียมสัญญาณเสียงเบื้องต้นเพื่อปรับสัญญาณเสียงอินพุตที่เข้ามาให้อยู่ในรูปแบบที่เหมาะสมก่อน จากนั้นจะทำการดึงคุณลักษณะเด่น โดยแต่ละเสียงจะมีความพิเศษที่แตกต่างกันที่สามารถทำให้หูของมนุษย์นั้นได้ยินเสียงที่แตกต่างกันเหล่านี้ได้ เพื่อเป็นตัวแทนของสัญญาณเสียงหรือความพิเศษของเสียงนั้น ๆ ต่อมาจะทำการเลือกคุณลักษณะเด่นของสัญญาณเสียงพูดโดยการปรับพารามิเตอร์เพื่อเตรียมคุณลักษณะเด่นของสัญญาณให้เป็นคุณลักษณะเด่นของสัญญาณที่ดีสำหรับการจำแนกรู้จำในโครงข่ายประสาทเทียมสามารถแสดงกระบวนการทำงานทั้งหมดได้ดังรูปที่ 3.1



รูปที่ 3.1 บล็อกไดอะแกรมการทำงานของระบบรู้จำเสียง

3.1.1 การเตรียมสัญญาณเสียงเบื้องต้น

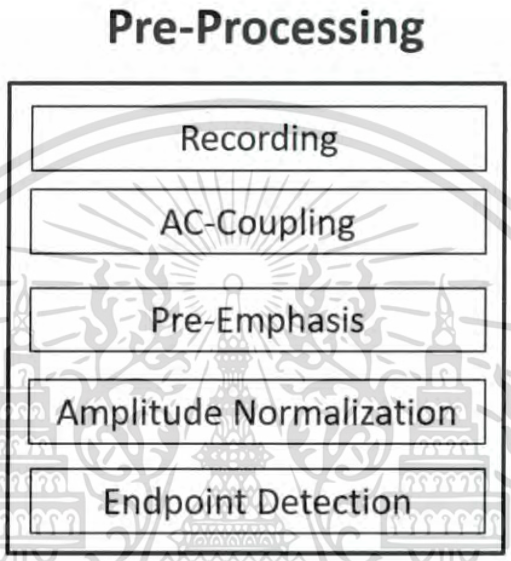
กระบวนการเตรียมสัญญาณเสียงนั้น เป็นขั้นตอนแรกที่เราจะนำสัญญาณเสียงที่เราได้ไปดึงคุณลักษณะเด่นของสัญญาณเสียงนั้น ๆ โดยใช้ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ซึ่งเรียกวิธีการนี้ว่า “การดึงคุณลักษณะเด่นของเสียง” จากนั้นจะนำไปผ่านกระบวนเลือกคุณลักษณะเด่นของเสียงและระบบการรู้จำเสียงต่อไป โดยกระบวนการเตรียมสัญญาณเสียงนั้นเป็นการทำให้สัญญาณเสียงพูดมีความเหมาะสมในด้านการใช้งาน โดยเบื้องต้นวิธีการเตรียมสัญญาณเสียงนั้น สามารถจำแนกหลักๆได้เป็น 2 ขั้นตอนดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การขจัดสัญญาณรบกวน (Noise) โดยการกรองความถี่

2. การตัดคำของสัญญาณเสียง (Endpoint Detection) คือ การดึงสัญญาณเสียงที่เป็นเสียงจริง ๆ เท่านั้น โดยองค์ประกอบอื่นนอกจากสัญญาณเสียงจะทำการตัดออก

โดยสามารถเขียนวิธีการเตรียมสัญญาณเสียงเบื้องต้นได้ในรูปบล็อกไดอะแกรมจากการออกแบบได้เป็นดังรูปที่ 3.2



รูปที่ 3.2 บล็อกไดอะแกรมของระบบกระบวนการเตรียมสัญญาณเสียงเบื้องต้น

3.1.1.1 การออกแบบระบบ

เราจะทำการเตรียมสัญญาณเสียงเบื้องต้นจากรูปที่ 1 ที่ได้กล่าวไปในบทนำ ซึ่งเราจะใช้โปรแกรม MATLAB Professional ทำการประมวลผลสัญญาณเสียงในขั้นตอนต่าง ๆ และศึกษาผลลัพธ์ที่เกิดขึ้นของสัญญาณเสียงนั้น

1.) Recording

เราจะทำการบันทึกสัญญาณเสียง ผ่านไมโครโฟนซึ่งถูก Sampling ด้วยความถี่ 44.1 kHz ขนาด 16 บิต ซึ่งไฟล์เสียงที่ได้จะถูกเก็บเป็นไฟล์นามสกุล .wav (Wave Format) และเป็นระบบเสียงแบบ Mono โดยบันทึกเสียงชื่อเพลงทั้งหมด 50 เพลง โดยแบ่งออกเป็น 1 พยางค์, 2 พยางค์, 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พยางค์, 4 พยางค์อย่างละ 10 เพลง, และกลุ่มเพลง 5 พยางค์ขึ้นไปทั้งหมด 10 เพลง โดยสามารถแสดงรายชื่อเพลงได้ดังตารางดังนี้

ตารางที่ 3.1 กลุ่มชื่อเพลง 1 พยางค์

เพลงที่	ชื่อเพลง	ศิลปิน
1	เหงา	Peacemaker
2	คอย	ฟรีเบิร์ด
3	เงา	หรั่ง Rockettra
4	ก่อน	Moderndog
5	ดาว	คริสติน
6	รอ	มาช่า
7	เคย	ออดี้
8	ดื่ม	Skalaxy
9	แช่ง	B-King
10	อาย	สิงโต นำโชค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 กลุ่มชื่อเพลง 2 พยางค์

เพลงที่	ชื่อเพลง	ศิลปิน
1	ใจรัก	สุชาติ ชวางกูล
2	สองรัก	Zeal
3	คิดถึง	พงษ์สิทธิ์ คัมภีร์
4	มีเธอ	ปุ๊ อัญชลี
5	ร่ำไร	อัสนี-วสันต์
6	อีกนาน	พงษ์พัฒน์
7	ยังจำ	สี่เต่าเธอ
8	เจ้าตาก	คาราบาว
9	แอบเจ็บ	Xyz
10	วายร้าย	UrboyTJ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 กลุ่มชื่อเพลง 3 พยางค์

เพลงที่	ชื่อเพลง	ศิลปิน
1	จำฝังใจ	ไมโคร
2	ใจปอนๆ	ไมโคร
3	เสี่ยวรำพึง	อัสนี-วสันต์
4	ตัวสำรอง	พงพัฒน์
5	เธอน่ารัก	สี่เต่าเธอ
6	ไม่ต่างกัน	พิสุทธิ์
7	ใจนักเลง	พงพัฒน์
8	แสงสุดท้าย	Bodyslam
9	ระหว่างเรา	อรอรีย์
10	ฉันยังอยู่	อิงค์ อชิตะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 กลุ่มชื่อเพลง 4 พยางค์

เพลงที่	ชื่อเพลง	ศิลปิน
1	เธอคือความฝัน	พราว
2	ได้อย่างเสียอย่าง	อัสนี-วสันต์
3	เพราะเธอหรือเปล่า	The Innocent
4	เห็นใจกันหน่อย	The Innocent
5	หัวใจสะออน	อัสนี-วสันต์
6	คนนี้ของฉัน	อิงค์-อ๋ำ
7	ไม่ตลกเลย	อ๋ำ
8	หัวใจมักง่าย	แช่ม แช่มรัมย์
9	ฟ้ายังมีฝน	แหวน ฐิติมา
10	บอกฉันบอกเธอ	Bangkok Gigolo

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.5 กลุ่มชื่อเพลง 5 พยางค์ขึ้นไป

เพลงที่	ชื่อเพลง	ศิลปิน
1	ด้วยรักและผูกพัน	เบิร์ด ธงไชย
2	รักคงยังไม่พอ	เสือ ธนพล
3	อาจจะเป็นคนนี้	แหวน ฐิติมา
4	ขอบใจที่พูดแรง	K-9
5	รักเธอให้น้อยลง	Bandwagon
6	เกินใจจะอดทน	Syam
7	ถ้าฉันเปลี่ยนใจเธอได้	Lipta
8	ครั้งหนึ่งเราเคยรักกัน	Endorphine
9	เมื่อเขามาฉันจะไป	Endorphine
10	รักที่เพิ่งผ่านพ้นไป	Sofa

2.) AC-Coupling

ในส่วนนี้จะเป็นการขจัดสัญญาณที่มีองค์ประกอบความถี่เท่ากับ 0 Hz ทิ้ง หรือ ก็คือ "สัญญาณไฟตรง" ซึ่งจะเป็นไปตามสมการที่ 3.1

$$\tilde{X}(n) = X(n) - \frac{1}{N}\sum_{n=1}^{N} X(n) \qquad (3.1)$$

3.) Pre-Emphasis

กระบวนการนี้จะนำสัญญาณที่ผ่านกระบวนการขจัดไฟตรงแล้วมาทำการผ่านวงจรกรองความถี่สูงผ่าน (High Pass Filter ชนิด First Order FIR Filter) ซึ่งมีฟังก์ชันถ่ายโอน (Transfer Function) เป็นดังสมการที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$H(z) = 1 - 0.95z^{-1} \tag{3.2}$$

4.) Amplitude Normalization

ส่วนนี้จะเป็นการทำ "นอมอลไลซ์ทางขนาดของสัญญาณ" ซึ่งเป็นการกำหนดให้แอมพลิจูดของสัญญาณเสียง $(S(n))$ ที่ผ่านฟังก์ชันถ่ายโอนข้างต้นแล้วมีค่าสูงสุดไม่เกิน ระหว่าง +1 และ -1 โดยเป็นกระบวนการที่ใช้เปรียบเทียบสัญญาณเสียงอื่นๆ[8]. ที่จุดอ้างอิงเดียวกัน ซึ่งสามารถคำนวณผลลัพธ์ได้จากสมการที่ 3.3

$$\tilde{S(n)} = \frac{S(n)}{\max(|S(n)|)} \tag{3.3}$$

5.) Endpoint Detection

โดยกระบวนการนี้จะเป็นกระบวนการที่ตัดเอาคำที่เป็นเฉพาะสัญญาณเสียงจริง ๆ เท่านั้น โดยสามารถแบ่งกระบวนการออกเป็นได้ 2 ส่วนคือ ส่วนที่ใช้สำหรับการตัดคำ และ ส่วนที่ใช้สำหรับการตรวจสอบพยางค์

<u>ส่วนที่ใช้สำหรับการตัดคำ</u>

โดยการตัดคำนั้นสำหรับเสียง 1 พยางค์และ 2 พยางค์จะใช้วิธีการตัดหัวและตัดท้ายคำของสัญญาณเสียง แต่กรณีที่เป็น 3 พยางค์ขึ้นไปจะใช้การตัดคำแบบแยกพยางค์ โดยใช้วิธีการตัดคำสัญญาณเสียงด้วยวิธีหาค่าพลังงานแบบ Absolute Energy ซึ่งการหาค่าพลังงานของสัญญาณเสียงนั้นสามารถหาได้จากสมการที่ 3.4

$$E(n) = \sum_{n=1}^{N} \left(|S_i(n)|\right)^2 \tag{3.4}$$

โดย i คือ จำนวนเฟรม

โดยจะทำการแบ่งข้อมูลเป็นเฟรมซึ่งแต่ละเฟรมนั้นมีขนาดข้อมูลเท่ากับ 200 ข้อมูลและไม่เกิดการซ้อนทับ (Overlap) กันของแต่ละเฟรม และเมื่อได้ค่าพลังงานของสัญญาณเสียงแล้ว จึงจะหาขอบเขตของแต่ละสัญญาณเสียงหรือคำที่พูด ทำได้โดยการใช้ค่าการตัดสินใจ (Thresholding Value) ซึ่งการใช้ค่าการตัดสินใจนี้ จะใช้ค่าการตัดสินใจแบบอัตโนมัติ (Automatic Thresholding) โดยจะกำหนดค่าการตัดสินใจเริ่มต้นเป็นดังสมการที่ 3.5

$$th = \frac{mean(E(n)) + median(E(n))}{4} \tag{3.5}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยค่าการตัดสินใจเริ่มต้นนี้เป็นค่าที่ได้จากการทดลองสำหรับค่าที่ต่ำที่สุดที่เป็นไปได้ในการตัดระหว่างคำเพื่อแบ่งพยางค์ของคำกรณีที่มี 2 พยางค์ขึ้นไป ซึ่งหลักการของค่าการตัดสินใจแบบอัตโนมัติ จะทำการเพิ่มค่าทีละ 0.05 จากค่าเดิม จำนวนทั้งหมด 60 ครั้ง โดยความยาวของคำจะเพิ่มขึ้นเป็นไปอย่างนัยสำคัญ ซึ่งผลสุดท้ายจะทำการเลือกค่าการตัดสินใจที่ทำให้ความยาวของคำมีค่ามากที่สุด และจะนำค่านี้ไปทำการออกแบบกรอบการตัดสัญญาณ โดยกรอบสัญญาณจะมีค่ามากสุดที่ 1 และน้อยสุดที่ 0 ซึ่งกรอบสัญญาณจะถูกออกแบบได้โดยสมการที่ 3.6

$$logic(n) = \frac{1}{2} \cdot \left(\frac{E(n) - th}{|E(n) - th|} + 1 \right) \tag{3.6}$$

แต่เนื่องจากการตัดคำยังเกิดข้อผิดพลาดขึ้นอยู่ (ตัดคำไม่สมบูรณ์) คือ เกิดการตัดคำในคำที่เป็นเสียงจริง ๆ ทำให้เกิดช่องระหว่างคำ ต้องทำการสร้างกรอบสัญญาณตัดคำใหม่ โดยนำกรอบสัญญาณที่ได้จากสมการที่ 3.6 มาทำตรวจสอบขั้นแรกคือ ถ้ากรอบสัญญาณขาขึ้นถึงกรอบสัญญาณขาลงที่มีค่าเป็น 1 (ทุก ๆ กรอบสัญญาณ) มีค่ามากกว่า 4000 จุด (ค่านี้ได้จากการทดลอง) ถือว่าคำนั้นเป็นพยางค์ของเสียงคำจริง ๆ และถ้ามีค่าน้อยกว่า 4000 จุด ให้ทุกตำแหน่งจากกรอบสัญญาณขาขึ้นถึงกรอบสัญญาณขาลง มีค่าเป็น 0 ต่อมาเมื่อทำการแก้ไขปัญหากรอบสัญญาณในขั้นแรกแล้ว ในขั้นที่สองจะเป็นการตรวจสอบระยะห่างระหว่างกรอบสัญญาณซึ่ง ถ้าระยะห่างระหว่างของกรอบสัญญาณ ทุก ๆ กรอบสัญญาณที่ติดกัน มีค่าน้อยกว่า 4000 จุด (ค่านี้ได้จากการทดลอง) ให้ระยะห่างระหว่างกรอบสัญญาณทั้งสองนั้นมีค่าเป็น 1 เป็นต้น

<u>ส่วนที่ใช้สำหรับการตรวจสอบพยางค์</u>

การตรวจสอบพยางค์จะเป็นไปตามสมการที่ 3.7

$$L_Syllable = \frac{find(logic(n+1) - logic(n))}{2} \tag{3.7}$$

โดยสมการจะเป็นการนำค่าในแต่ละตำแหน่งของสมการกรอบสัญญาณมาลบกัน ต่อมานำผลลัพธ์ที่ได้หาค่า Absolute เมื่อทำการ Absolute แล้วให้นำค่าที่ได้มาตรวจสอบว่ามีกี่ตำแหน่งที่ให้ค่าผลลัพธ์เป็น 1 ต่อมาเมื่อได้ค่าจำนวนตำแหน่งที่ได้เกิดขึ้นแล้วให้นำมาหารด้วย 2 ก็จะเป็นจำนวนพยางค์ของคำนั้น ๆ

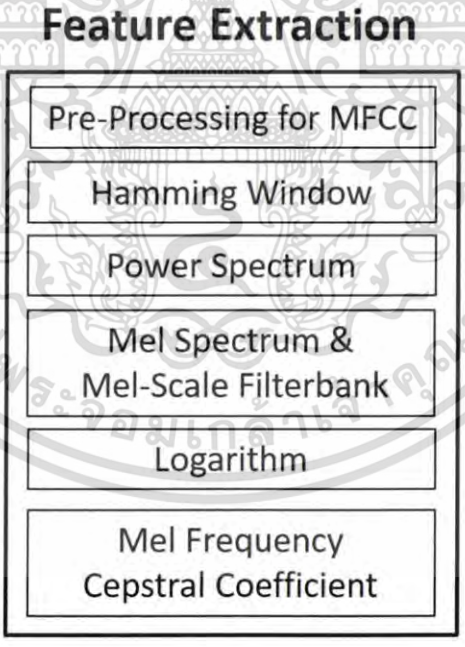
เมื่อตรวจสอบพยางค์และปรับปรุงความผิดพลาดที่เกิดจากการตัดคำแล้วจะนำแต่ละพยางค์ที่ได้ไปทำการรู้จำ (แยกพยางค์รู้จำสำหรับ 3 พยางค์ขึ้นไป) เพื่อเป็นการเพิ่มประสิทธิภาพให้ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั้นมีความถูกต้องแม่นยำมากยิ่งขึ้นโดยสามารถกล่าวรายละเอียดวิธีการตรวจสอบคำเพื่อจำแนกได้ในโครงข่ายประสาทเทียมต่อไป

3.1.2 การดึงคุณลักษณะเด่นของสัญญาณเสียง

ในส่วนของกระบวนการนี้จะเป็นกระบวนการต่อจากการเตรียมสัญญาณเสียงเบื้องต้น โดยสัญญาณเสียงพูดแต่ละเสียงนั้นจะมีความพิเศษของแต่ละสัญญาณเสียงที่พูดอยู่ ซึ่งทำให้เราสามารถบอกได้ว่าสัญญาณเสียงที่ได้นั้นมีความแตกต่างกัน เช่น เมื่อพูดคำว่า “ฉัน” หรือ “เธอ” หูเราจะรู้ได้เลยว่าสัญญาณเสียงที่พูดนั้นมีความแตกต่าง ทั้งนี้ในขั้นตอนต่อไปจำเป็นต้องหาความพิเศษของสัญญาณเสียงนั้น ๆ หรือคุณลักษณะเด่นของสัญญาณเสียงนั้น ๆ เพื่อจะจำแนกความแตกต่างของแต่ละสัญญาณเสียงทำได้โดยกระบวนการที่เรียกว่า “กระบวนการดึงคุณลักษณะเด่นของสัญญาณเสียง” ซึ่งขั้นตอนการดึงคุณลักษณะเด่นของสัญญาณเสียงที่พูดนั้น จะเป็นตัวแทนของสัญญาณเสียงพูดที่มีจำนวนข้อมูลปริมาณมาก ๆ ด้วยรูปแบบข้อมูลใหม่ ที่มีปริมาณข้อมูลน้อยลง โดยใช้ค่าที่เรียกว่า “ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล” ซึ่งค่าสัมประสิทธิ์ที่ว่านี้จะสามารถอธิบายความไม่เป็นเชิงเส้น (Non-Linearity) ของการรับรู้ของหูมนุษย์ได้ดี เนื่องจากสเปกตรัมของค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล นี้จะให้รายละเอียดที่ดีในช่วงสัญญาณความถี่ต่ำและยังมีความทนทานต่อสัญญาณรบกวน (Noise) ด้วย สามารถแสดงขั้นตอนการทำงานในส่วนของการดึงคุณลักษณะเด่นของสัญญาณได้ดังรูปที่ 3.3



รูปที่ 3.3 บล็อกไดอะแกรมวิธีการดึงคุณลักษณะเด่นของสัญญาณเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2.1 การออกแบบระบบ

เราจะทำการเตรียมสัญญาณเสียงเบื้องต้นที่ได้จากกระบวนการเตรียมสัญญาณเสียงเบื้องต้นไปใช้งาน ซึ่งเราจะใช้โปรแกรม MATLAB Professional ทำการประมวลผลสัญญาณเสียงในขั้นตอนต่าง ๆ และศึกษาผลลัพธ์ที่เกิดขึ้นของสัญญาณเสียงนั้น

1.) Pre-Processing for MFCC

ขั้นตอนนี้จะทำการแบ่งข้อมูลของสัญญาณที่ได้จากกระบวนการเตรียมสัญญาณเสียงเบื้องต้นออกเป็นเฟรม ซึ่งมีทั้งหมด N ตัวอย่าง โดยจะกำหนดให้แต่ละเฟรมมีขนาดเท่ากับ 256 แซมเปิล และทำการเลื่อนกรอบเฟรมทีละ 128 แซมเปิล (หมายความว่าเฟรมแรกจะมีแซมเปิลตั้งแต่ตัวที่ 1 ถึง 256 แซมเปิล ซึ่งในเฟรมถัดไปก็จะมีแซมเปิลเป็น 128 ถึง 383 แซมเปิล ต่อไปก็จะเป็น 257 ถึง 512 แซมเปิล , ... เป็นต้น หรือ เกิดการซ้อนทับของเฟรมทุก ๆ 128 แซมเปิล) ถ้าเฟรมสุดท้ายมีขนาดแซมเปิลน้อยกว่า 256 แซมเปิลให้ทำการตัดเฟรมนั้นทิ้งไป โดยสมการที่ทำการออกแบบการเลื่อนเฟรมนี้เป็นไปตามสมการที่ 3.8 โดยแบ่งเป็น 2 กรณี ดังนี้

<u>กรณีที่ 1</u> ถ้า $(\mathrm{N}/256) - (\lfloor N/256 \rfloor) >= 0.5$

โดยที่ $k = 1,2,3,\ldots, \lfloor N/256 \rfloor$

$f_i[k]$ คือ ฟังก์ชันของจำนวนเฟรม ซึ่งจำนวนเฟรมทั้งหมดมีค่าเท่ากับ $\lfloor N/128 \rfloor$ เฟรม โดยแต่ละเฟรมมีข้อมูลบรรจุทั้งหมด 256 แซมเปิล (หมายเหตุ: $\lfloor x \rfloor$ คือการปัดเศษเป็นจำนวนเต็มที่ต่ำกว่า)

<u>กรณีที่ 2</u> ถ้า $(\mathrm{N}/256) - (\lfloor N/256 \rfloor) < 0.5$

โดยที่ $k = 1,2,3,\ldots, \lfloor N/256 \rfloor$ -1

$f_i[k]$ คือ ฟังก์ชันของจำนวนเฟรม ซึ่งจำนวนเฟรมทั้งหมดมีค่าเท่ากับ $\lfloor N/128 \rfloor$ เฟรม โดยแต่ละเฟรมมีข้อมูลบรรจุทั้งหมด 256 แซมเปิล

$$f_i((2\times k)-1) = Output(1+(256\times(k-1)):256\times k) \qquad (3.8)$$

$$f_i(2\times k) = Output(128+(256\times(k-1)):((256\times k)+128)-1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.) แฮมมิ่งวินโดว์

เมื่อทำการแบ่งสัญญาณเสียงออกเป็นเฟรมแล้วจากนั้นจะนำสัญญาณเสียงที่ได้แต่ละเฟรมไปทำการคูณกับแฮมมิ่งวินโดว์ โดยเพื่อต้องการให้สัญญาณเสียงในแต่ละเฟรมมีลักษณะเป็นคาบและต่อเนื่อง ในตำแหน่งจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยการออกแบบแฮมมิ่งวินโดว์นั้นสามารถสร้างได้จากสมการที่ 3.9

$$w_H(m) = (1-\alpha) - [\alpha \cdot \cos(\frac{2\pi m}{Nf-1})] \quad (3.9)$$

โดยที่ $m = 0,1,2,\ldots, (Nf\text{-}1)$ และ Nf คือขนาดของวินโดว์ซึ่งมีค่าเท่ากับจำนวนข้อมูลในแต่ละเฟรม (เท่ากับ 256 แซมเปิลต่อเฟรม)

ในที่นี้จะกำหนดให้ α มีค่าเท่ากับ 0.46 ซึ่งค่าที่เลือกใช้นี้เป็นค่าที่นิยมใช้โดยทั่วไปในการวิเคราะห์สัญญาณเสียง และนำวินโดว์ที่ α เท่ากับ 0.46 ไปคูณกับทุกเฟรมข้อมูลของสัญญาณเสียงพูด จากนั้นเก็บข้อมูลไว้ในตัวแปร $x(k)$ ดังสมการที่ 3.10

$$x(k) = f_i(k) \cdot w_H(m) \quad (3.10)$$

โดยที่ $k = 1,2,3,\ldots, \lfloor N/128 \rfloor$

3.) Power Spectrum

เราจะนำเสียงที่ผ่านแฮมมิ่งวินโดว์แล้วในแต่ละเฟรมมาทำการวิเคราะห์โดยใช้สเปกตรัมเชิงกำลังของสัญญาณเสียงในโดเมนความถี่ ขั้นแรกทำการแปลงดิสครีตฟูเรียร์ทราสฟอร์ม (Discrete Fourier Transform) หรือ DFT ซึ่งเป็นไปตามสมการที่ 3.11

$$X[k] = \sum_{k=0}^{N-1} x(k) e^{\frac{-j2\pi nk}{N}} \quad (3.11)$$

โดยที่ $x(k)$ คือ เฟรมของสัญญาณเสียงที่ผ่านแฮมมิ่งวินโดว์แล้ว

$X[k]$ คือ สเปคตรัมของเฟรมสัญญาณเสียง $x(k)$

N คือ ขนาดของเฟรมสัญญาณเสียง $x(k)$

m คือ ตำแหน่งของสัญญาณเสียงอินพุตซึ่งมีค่าตั้งแต่ $m = 0$ ถึง N -1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการแปลงดิสครีตฟูเรียร์ทราสฟอร์มเสร็จแล้วขั้นตอนต่อไปจะเป็นการหาสเปกตรัมเชิงกำลังของสัญญาณเสียงในตัวแปร $P[k]$ ซึ่งเป็นไปตามสมการที่ 3.12

$$P[k] = |X[k]|^2 \tag{3.12}$$

การแปลงดิสครีตฟูเรียร์นั้นเป็นวิธีการแปลงลำดับสัญญาณทางเวลาที่มีขนาดแซมเปิล N จุดแซมเปิล ให้ได้ค่าผลลัพธ์เป็นค่าสเปกตรัมทางความถี่ที่มีจำนวนจุดแซมเปิลเท่ากับ N แซมเปิล แต่การคำนวณ DFT แต่ละครั้งจะต้องมีการคูณตัวเลขทั้งสิ้น N^2 ครั้งและ มีการบวกตัวเลขอีก $N(N\text{-}1)$ ครั้ง ดังนั้นในทางปฏิบัติจึงใช้การแปลงฟูเรียร์ด้วยวิธี Fast Fourier Transform หรือเขียนตัวย่อว่า FFT ซึ่งจะทำให้การแปลงฟูเรียร์นั้นทำได้อย่างมีประสิทธิภาพมากขึ้นและรวดเร็วขึ้น

4.) Mel Spectrum & Mel-Scale Filterbank

การจะค่าสเปคตรัมบนสเกลเมลนั้นจะต้องทำการเปลี่ยนสเกลทางความถี่ธรรมดาซึ่งเป็นสเกลเชิงเส้นให้เป็นสเกลเมลก่อน ซึ่งเป็นสเกลแบบไม่เชิงเส้น ดังสมการที่ 3.13

$$f_{mel}(f) = 1125 \cdot \ln\left(1 + \frac{f}{700}\right) \tag{3.13}$$

โดยขั้นตอนแรกนำค่าความถี่บนแกนความถี่ปกติไปแทนลงบนสมการสเกลเมลข้างต้นก่อน หลังจากนั้นทำการแปลงกลับสเกลเมล ดังสมการที่ 3.14

$$M^{-1}(m) = f(m) = 700(\exp(\frac{f_{mel}(f)}{700}) - 1) \tag{3.14}$$

ต่อมานำสัญญาณเสียงที่ผ่านการแปลงดิสครีตฟูเรียร์แล้วไปผ่านฟิลเตอร์แบงค์ที่ทำการแปลงกลับสเกลเมล (ผ่านฟิลเตอร์ก็คือการคูณกันของเฟรมสัญญาณเสียงกับฟิลเตอร์แบงค์ในโดเมนความถี่ หรือการคอนโวลูชั่นในโดเมนเวลานั่นเอง) ซึ่งเขียนได้เป็นดังสมการที่ 3.15

$$S[k] = P[k] \cdot H_m[k] \tag{3.15}$$

โดยที่ $H_m[k]$ คือ ผลตอบสนองทางขนาดของฟิลเตอร์แบงค์

m คือ จำนวนฟิลเตอร์รูปสามเหลี่ยม

ซึ่งผลลัพธ์จะเป็นการคำนวณหาค่าเฉลี่ย (Average Spectrum) รอบความถี่กลาง (Center Frequency) ของฟิลเตอร์ โดยฟิลเตอร์สามเหลี่ยมนี้จะเป็นฟิลเตอร์ชนิดแบนด์พาสฟิลเตอร์นั่นเอง ซึ่งการสร้างฟิลเตอร์แบงค์นี้จะเป็นไปตามสมการที่ 3.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$H_m[k] = \begin{cases} 0 & ,k < f(m-1) \\ \dfrac{2(k - f(m-1))}{(f(m+1) - f(m-1))(f(m) - f(m-1))} & ,f(m-1) \le k \le f(m) \\ \dfrac{2(f(m+1) - k)}{(f(m+1) - f(m-1))(f(m+1) - f(m))} & ,f(m) \le k \le f(m+1) \\ 0 & ,k > f(m+1) \end{cases} \quad (3.16)$$

จะกำหนดจำนวนฟิลเตอร์แบงค์ทั้งหมด 12 ฟิลเตอร์ (m = 1,2,3,…,12) ซึ่งเป็นค่าที่นิยมใช้ในงานทางด้านรู้จำเสียง และจะส่งผลให้มีสัมประสิทธิ์เซปสตรัมบนสเกลเมลในแต่ละเฟรมของสัญญาณเสียงเท่ากับ 12 ข้อมูลด้วยเช่นกัน

5.) Logarithm

ผลลัพธ์ที่แต่ละเฟรมสัญญาณเสียงที่ไปผ่านฟิลเตอร์แบงค์แต่ละตัวจะถูกนำไปผ่านฟังก์ชันลอการึทึมเพื่อให้สเปกตรัมเชิงกำลังของเฟรมสัญญาณเสียงอยู่บนสเกลลอการิทึม ดังสมการที่ 3.17

$$S_{mel}(k) = \log\left[\sum_{n=1}^{N} S[k]\right] \quad (3.17)$$

จากสมการอธิบายได้ว่าจะทำการบวกค่าในแต่ละ $S[k]$ ที่มีจำนวนข้อมูลอยู่ในนั้นตั้งแต่แซมเปิลที่ 1 ถึง N แซมเปิลและไปผ่านฟังก์ชันลอการิทึม

6.) กระบวนการหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล

ในขั้นตอนนี้เมื่อผ่านขั้นตอนที่ 5 มาแล้วจะนำผลลัพธ์มาหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ด้วยการแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform) ดังสมการที่ 3.18

$$C(m) = w(m) \cdot \sum_{n=1}^{M} S_{mel}(k) \cos\left(\frac{\pi}{2M}(2n-1)(m-1)\right) \ ,m = 1,2,\ldots,M \quad (3.18)$$

โดยที่ $w(m) = \begin{cases} \dfrac{1}{\sqrt{M}} & ,m = 1 \\ \sqrt{\dfrac{2}{M}} & ,2 \le m \le M \end{cases}$

C คือ จำนวนสัมประสิทธิ์เซปสตรัม

M คือ จำนวนของฟิลเตอร์แบงค์

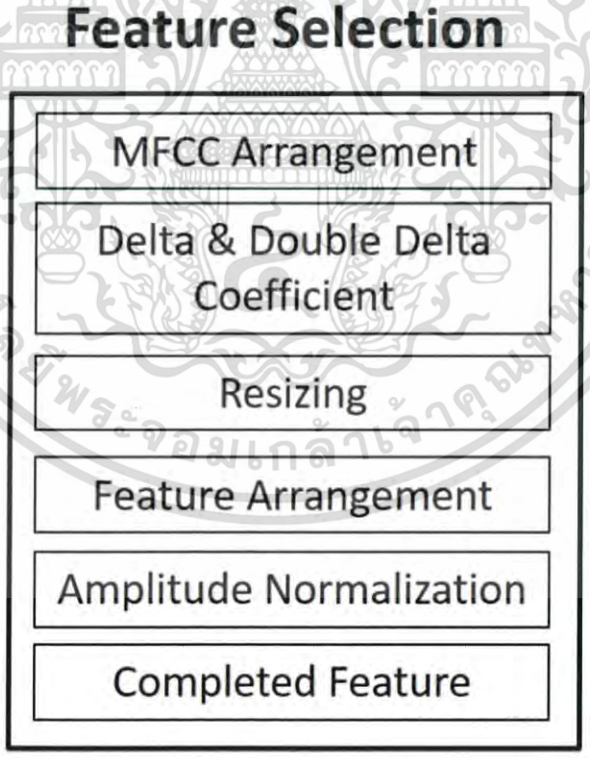
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแปลงโคซายน์แบบไม่ต่อเนื่อง (DCT) เป็นการแปลงแบบ Orthogonal ที่มีฟังก์ชันโคโซน์เป็นฐาน ซึ่งมักจะนำไปประยุกต์ใช้งานประมวลผลสัญญาณภาพ และ สัญญาณเสียงโดยเฉพาะอย่างยิ่งการเข้ารหัสข้อมูล และ การบีบอัดข้อมูลที่จะสามารถบีบอัดค่าพลังงานของสัญญาณเสียงส่วนใหญ่ไปไว้ในสัมประสิทธิ์ย่านความถี่ต่ำได้

โดยเมื่อผ่านการแปลงโคไซน์แบบไม่ต่อเนื่องแล้วจะได้ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ที่มีสัมประสิทธิ์เฟรมละ 12 ค่าตั้งแต่ค่าสัมประสิทธิ์ลำดับที่ 1 ถึงลำดับที่ 12 ในแต่ละเฟรม ซึ่งเรานำค่าที่ได้จากสัมประสิทธิ์เซปสตรัมบนสเกลเมลนี้ ไปใช้ในกระบวนการเลือกคุณลักษณะเด่น ต่อไป

3.1.3 การเลือกคุณลักษณะเด่นของสัญญาณเสียง

เมื่อผ่านส่วนของกระบวนการดึงคุณลักษณะเด่นของสัญญาณเสียงด้วย MFCC ทำให้ได้ค่าค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลทั้ง 12 ค่ามาแล้ว ต่อมาในหัวข้อนี้จะทำเลือกคุณลักษณะเด่นของสัญญาณเสียง ซึ่งจะมีการปรับปรุงสัมประสิทธิ์ที่เป็นตัวแทนของเสียงที่ส่งผลให้คุณภาพของคุณลักษณะเด่นของสัญญาณเสียงนั้นให้มีความแตกต่าง เพื่อส่งผลให้ระบบรู้จำมีประสิทธิภาพที่ดีมากยิ่งขึ้น โดยสามารถแสดงกระบวนการทำงานได้ดังรูปที่ 3.4



รูปที่ 3.4 บล็อกไดอะแกรมของการดึงคุณลักษณะเด่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3.1 การออกแบบระบบ

1.) MFCC Arrangement

นำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลในแต่ละลำดับมาทำการเรียงกันตามลำดับของเฟรม คือ ทุก ๆ เฟรมจะมีค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลอยู่ทั้งหมด 12 ค่าโดยจะนำค่าในลำดับเดียวกันของทุก ๆ เฟรมมาเรียงต่อกันตั้งแต่เฟรมแรกจนถึงเฟรมสุดท้ายทำให้เหลือค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลเพียงชุดเดียว (จากเดิมที่มีหลายชุดตามจำนวนเฟรม) เป็นต้น

2.) การหาค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า

จะทำการเพิ่มค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า เข้าไปเพื่อเพิ่มประสิทธิภาพให้ระบบ หรือสร้างความแตกต่างให้มีคุณสมบัติที่ดีของคุณลักษณะเด่นของสัญญาณเสียง ซึ่งสามารถแสดงการหาค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้าได้ดังสมการที่ 3.19 และ 3.20 ตามลำดับ

$$d(m) = \frac{c(m+1) - c(m-1)}{2}, \; m = 1, 2, ..., M \tag{3.19}$$

โดยที่ c คือ จำนวนสัมประสิทธิ์เซปสตรัม

M คือ จำนวนของฟิลเตอร์แบงค์ หรือจำนวนของค่าสัมประสิทธิ์บนสเกลเมล (ซึ่งเท่ากับ 12)

d คือ ค่าสัมประสิทธิ์เดลต้า

$$delta_d(m) = \frac{d(m+1) - d(m-1)}{2}, \; m = 1, 2, ..., M \tag{3.20}$$

โดยที่ d คือ จำนวนสัมประสิทธิ์เดลต้า

M คือ จำนวนของฟิลเตอร์แบงค์ หรือจำนวนของค่าสัมประสิทธิ์บนสเกลเมล (ซึ่งเท่ากับ 12)

$delta_d$ คือ ค่าสัมประสิทธิ์ดับเบิล-เดลต้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.) Resizing

จากนั้นจะนำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า มาทำการลดจำนวนข้อมูลด้วยวิธีการ “Decimation” ทั้งนี้ข้อมูลที่ทำการลดในแต่ละค่าสัมประสิทธิ์แต่ละชนิดนั้นจะมีค่าเท่ากัน ซึ่งเป็นสิ่งสำคัญมากเพื่อนำไปใช้ในระบบรู้จำต่อไป กล่าวคือ โครงข่ายประสาทเทียมจำเป็นต้องกำหนดข้อมูลอินพุตเท่ากัน ขึ้นอยู่กับการออกแบบโครงข่ายประสาทเทียมนั้น ๆ เมื่อทำการบันทึกหรือเรียกสัญญาณเสียงในแต่ละครั้ง ขั้นตอนของการตัดคำ จะเกิดการไม่เท่ากันของสัญญาณเสียงที่ถูกตัด จากนั้นผ่านกระบวนการต่าง ๆ ได้มาซึ่งค่าสัมประสิทธิ์ต่าง ๆ ซึ่งส่งผลให้ค่าสัมประสิทธิ์ที่ได้มีจำนวนไม่เท่ากันเช่นกัน (ตามจำนวนเฟรมของสัญญาณ) ซึ่งทำให้จำเป็นต้องมีการทำ Resizing นั่นเองอีกทั้งข้อดีของการทำ Resizing ยังทำให้ระบบทำงานเร็วขึ้นด้วย ซึ่งเกิดจากการลดจำนวนข้อมูล เป็นต้น

การทำ Decimation เป็นการ DownSampling เพื่อให้ได้จำนวนข้อมูลที่ลดน้อยลง ซึ่งสามารถใช้ฟังก์ชันใน MATLAB ในทำการประมาณค่าได้ โดยใช้คำสั่ง interp1 ชนิด spline ซึ่งเป็นการประมาณค่าด้วยเส้นโค้งทำให้ได้ค่าที่ประมาณนั้นมีความผิดพลาดน้อยลงเมื่อเทียบกับการประมาณค่าชนิดเส้นตรง ซึ่งจะกำหนดให้ทำการประมาณค่าด้วยเส้นโค้ง และให้ข้อมูลแต่ละสัมประสิทธิ์มีจำนวนเท่ากับ 120 แซมเปิล

4.) Feature Arrangement

จะนำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า ตั้งแต่ลำดับที่ 1 ถึง 12 มาเรียงต่อกันตามลำดับในแต่ละชุดของสัมประสิทธิ์

5.) การนอมอลไลซ์ทางขนาดของสัญญาณ

ขั้นตอนนี้จะเหมือนกันกับกระบวนการเตรียมสัญญาณเสียงเบื้องต้น โดยจะทำการ นอมอลไลซ์ทางขนาดของชุดสัมประสิทธิ์ทั้ง 3 ชนิด (ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า)

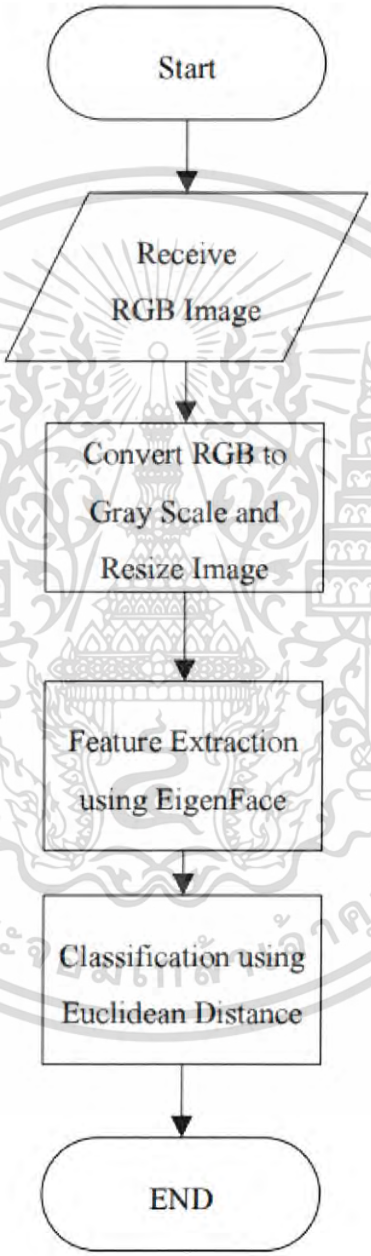
6.) Completed Feature

ในขั้นตอนสุดท้ายจะนำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า มาเรียงต่อกันตามลำดับโดยจะเป็นตัวแทนของเสียงนั้น ๆ เข้าสู่ระบบโครงข่ายประสาทเทียมเพื่อจำแนกข้อมูลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การออกแบบระบบการดึงคุณลักษณะเด่นเพื่อการรู้จำใบหน้าของกลุ่มบุคคล

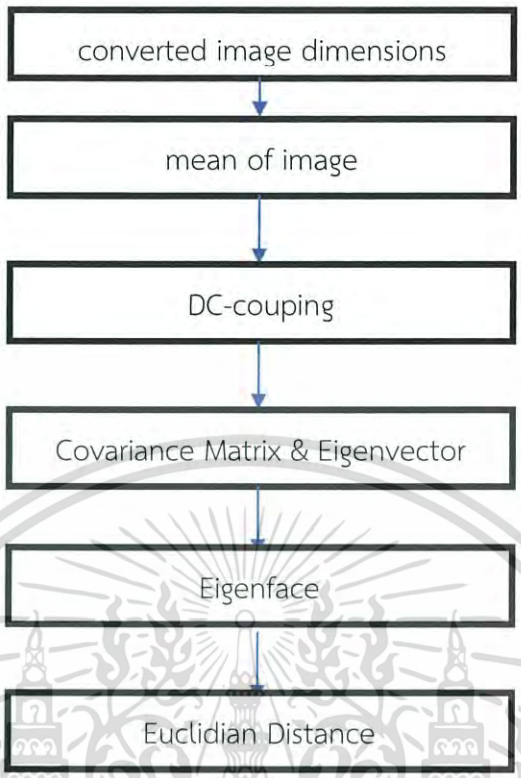
ในการออกแบบระบบรู้จำใบหน้าของกลุ่มบุคคลนั้น ผู้จัดทำได้ใช้วิธีไอเกนเฟส ซึ่งเป็นชื่อเรียกเซตของไอเกนเวกเตอร์ โดยใช้ในระบบการรู้จำภาพใบหน้า ซึ่งในที่นี้จะนำมาประยุกต์ใช้ร่วมกับโครงข่ายประสาทเทียมแบบแพร่กลับ โดยสามารถแสดงแผนภาพโฟวชาร์ตของกระบวนการไอเกนเฟสได้ดังรูปที่ 3.5



รูปที่ 3.5 โฟลวชาร์ตของกระบวนการไอเกนเฟส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถแสดงบล๊อกไดอะแกรมของระบบรวมได้ดังรูปที่ 3.6



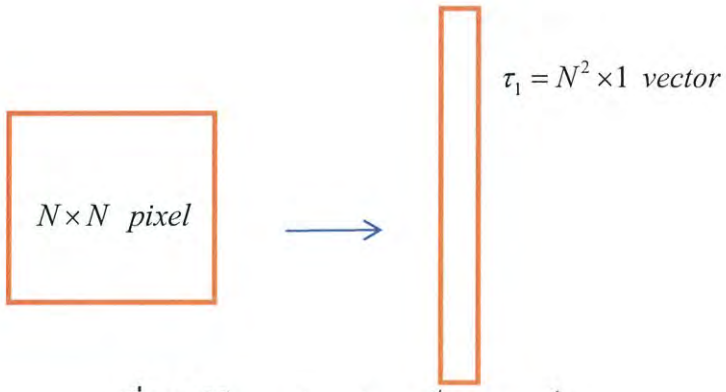
รูปที่ 3.6 บล็อกไดอะแกรมของกระบวนการรู้จำภาพวิธีไอเกนเฟส

ในขั้นตอนการเตรียมสัญญาณนั้น เราจะแบ่งข้อมูลที่ใช้เป็น 3 ส่วน ส่วนแรกคือข้อมูลที่เป็นฐานข้อมูลสำหรับไอเกนเฟส ส่วนที่ 2 คือข้อมูลส่วนที่จะนำมาเทียบค่าระยะทางยุคลิด และส่วนที่ 3 คือข้อมูลในส่วนที่นำมาเทสระบบในโครงข่ายประสาทเทียมเพื่อทำการหาค่าประสิทธิภาพของระบบ ซึ่งภาพของฐานข้อมูลในส่วนที่ 1 นั้นจะแบ่งเป็น กลุ่มบุคคลผู้ชาย (วัยกลางคน) กลุ่มบุคคลผู้หญิง (วัยกลางคน) กลุ่มบุคคลเด็กและกลุ่มบุคคลสูงอายุกลุ่มละ 30 ภาพ ซึ่งในส่วนที่ 2 และส่วนที่ 3 จะมีจำนวนข้อมูลในแต่ละกลุ่ม เท่ากับจำนวนข้อมูลในส่วนที่ 1 (ไม่มีข้อมูลที่ซ้ำกันเลย)

3.2.1 กระบวนการการแปลงอินพุตข้อมูล

กำหนดให้อินพุตนั้น เป็นรูปภาพใบหน้าขนาด NxN พิกเซล ให้ M เป็นจำนวนของรูปภาพ นำภาพขนาดนั้นไปเรียงเป็นเวกเตอร์ จะมีขนาดกลายเป็น N^2 x 1 แทน โดยให้เรียกเวกเตอร์นั้นว่า τ_M ซึ่งสามารถแสดงได้ในรูป 3.7 โดยในที่นี่ ใช้ภาพที่มีขนาดเท่ากับ 100x100 พิกเซล แล้วนำมาเรียงเป็นเว็กเตอร์ จะได้ขนาดเป็น 10,000 x 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

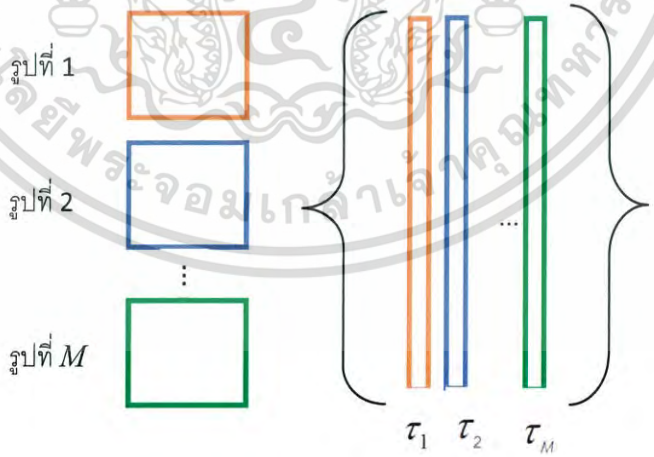


รูปที่ 3.7 วิธีการแปลงรูปภาพเป็นเวกเตอร์

ซึ่งแสดงได้ดังนี้

```
S=[];   %img matrix
figure(1);
for i=1:M
    str=strcat(int2str(i),'.jpg');    %concatenates two strings that form the name of the image
    eval('img=imread(str);');
    [irow icol]=size(img);     % get the number of rows (N1) and columns (N2)
    temp=reshape(img',irow*icol,1);     %creates a (N1*N2)x1 matrix
    S=[S temp];          %X is a (N1*N2xM matrix after finishing the sequence
                         %this is our S
end
```

เสร็จแล้วจึงเราจะทำการแปลงเป็นเวกเตอร์ทุกรูป จากนั้นนำมาเรียงกันในเมตริก A ซึ่งหลังจากนำมาเรียงกันแล้ว ขนาดของเมตริก A จะกลายเป็น $N^2 \times M$ ดังรูปที่ 3.8 ซึ่งจากการออกแบบจะได้เมตตริก A ที่มีขนาดเท่ากับ 10,000 x 120



รูปที่ 3.8 ผลลัพธ์ของการแปลงอินพุตข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 กระบวนการทำการหาค่าเฉลี่ยของรูปภาพใบหน้า

ทำการนำ $\tau_1, \tau_2, ..., \tau_n$ มาหาค่าเฉลี่ย โดยสามารถคำนวณได้จากสมการที่ 3.21

$$\psi = \frac{1}{M}\sum_{n=1}^{M} \tau_1\tau_2\tau_3, ..., \tau_n \tag{3.21}$$

โดยที่ ψ คือ ค่าเฉลี่ยของภาพทั้งหมด ซึ่งนั้นจะมีขนาดเป็น 10,000 x 1

3.2.3 กระบวนการกำจัดค่าเฉลี่ย

การกำจัดค่าเฉลี่ยออกจากภาพ ทำได้โดยหาลักษณะความแตกต่างของภาพ จากภาพอินพุตของหน้าแต่ละคนกับภาพเฉลี่ยของภาพหน้าคน สามารถทำได้โดยอาศัยสมการที่ 3.22

$$\phi_M = \tau_M - \psi \tag{3.22}$$

โดยที่ ϕ_M คือผลลัพธ์ของการกำจัดค่าเฉลี่ยของภาพที่ M โดยที่มีขนาดเท่ากับ N^2 x 1

ซึ่งสามารถออกแบบได้ดังนี้

```
%reduce the error due to lighting conditions.(Normalize)
for i=1:size(S,2)
    temp=double(S(:,i));
    m=mean(temp);
    st=std(temp);
    S(:,i)=(temp-m)*ustd/st+um;
end
```

3.2.4 กระบวนการหาค่า Covariance Matrix

นำภาพความแตกต่างที่ได้ไปหาการวิเคราะห์สหสัมพันธ์ (Correlation) ของภาพความแตกต่างโดยการสร้าง Covariance Matrix ดังสมการที่ 3.23

$$C = AA^T; A = [\phi_1, \phi_2, \phi_3, ..., \phi_n] \tag{3.23}$$

โดยที่ C คือ ผลลัพธ์หลังผ่านการคูณกันของเมตริก โดยจะมีขนาดเป็น $N^2 \times N^2$

ซึ่งสามารถออกแบบได้ดังนี้

```
%Covariance matrix C=A'A, L=AA'
A=dbx';
L=A*A';
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่เนื่องจาก Covariance Matrix นั้น มีขนาดที่กว้างมากเกินไป ทำให้ถ้าเราจะนำเมตริกนี้ไปหาค่าไอเกนแวลูและไอเกนเวกเตอร์นั้น จะทำให้เป็นการเพิ่มปริมาณของฐานข้อมูลโดยใช่เหตุ เราจึงต้องทำการแปลงสมการในหัวข้อย่อยถัดไปต่อไป

3.2.5 กระบวนการหาค่าไอเกนเวกเตอร์และไอเกนแวลู

หลังจากผ่านขั้นตอนการหาค่า Covariance Matrix แล้ว เราจะได้ค่า Covariance Matrix ซึ่งมีขนาด $N^2 \times N^2$ แต่ Covariance Matrix นี้นั้น มีขนาดที่กว้างมากเกินไปจะทำให้เป็นการเพิ่มปริมาณของฐานข้อมูลโดยใช่เหตุ ทำให้ถ้าเราต้องนำเมตริกนี้ไปแปลงขนาดก่อนโดยใช้ทฤษฎีหลักการแปลงในไอเกนเว็กเตอร์ และสามารถหาค่าไอเกนเว็กเตอร์จากความสัมพันธ์ของสมการ Covariance matrix กับไอเกนเว็กเตอร์ ไอเกนแวลู ที่สามารถอ้างอิงได้ในบทที่ 2 ในหัวข้อไอเกนเฟส ได้ดังสมการที่ 3.24

$$C(Av_M) = \lambda_M (Av_M) \tag{3.24}$$

โดยที Av_M คือ ไอเกนเว็กเตอร์

λ_M คือ ไอเกนแวลู

โดยสามารถออกแบบได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% vv are the eigenvector for L
% dd are the eigenvalue for both L=dbx'*dbx and C=dbx*dbx';
%%% [V,D] = eig(A) produces a diagonal matrix D of eigenvalues and ------%
%%%     a full matrix V whose columns are the corresponding eigenvectors ---%
%%%     so that A*V = V*D.-----------------------------------------%
[vv dd]=eig(L);
% Sort and eliminate those whose eigenvalue is zero
v=[];
d=[];
for i=1:size(vv,2)
    if(dd(i,i)>1e-4)
        v=[v vv(:,i)];
        d=[d dd(i,i)];
    end
end

%sort,  will return an ascending order
[B, index]=sort(d);
ind=zeros(size(index));
dtemp=zeros(size(index));
vtemp=zeros(size(v));
len=length(index);
for i=1:len
    dtemp(i)=B(len+1-i);
    ind(i)=len+1-index(i);
    vtemp(:,ind(i))=v(:,i);
end
d=dtemp;
v=vtemp; % d = eigenvalue  v=eigenvector for L (Covarience Metrix)
```

3.2.6 กระบวนการหาค่าไอเกนเฟส

หลังจากได้ค่าไอเกนเว็กเตอร์และไอเกนแวลูแล้วนั้นเราสามารถคำนวณหาไอเกนเฟส ซึ่งสามารถอธิบายได้ตามสมการที่ 3.25 และ 3.26

$$U = AV \tag{3.25}$$

$$V = [v_1 v_2 v_3, \ldots, v_M] \tag{3.26}$$

โดยที่ U คือผลลัพธ์ภาพไอเกนเฟส ซึ่งจะมีขนาดเป็น $1 \times M$

ซึ่งสามารถออกแบบได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%Eigenvectors of Covarance matrix
u=[];
for i=1:size(v,2)
    temp=sqrt(d(i));
    u=[u (dbx*v(:,i))./temp];
end
```

3.2.7 กระบวนการหาค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูล (Ω)

ทำการหาค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูล (Ω) ซึ่ง Ω นั้นเป็นค่าตัวแปรหนึ่งที่จะเปลี่ยนแปลงตามไอเกนเฟส โดยเราจะนำค่า Ω นี้มาเปรียบเทียบกับอินพุตข้อมูลที่เราจะนำมาหาค่าระยะทางแบบยุคลิด[2]. สามารถหาได้ตามสมการที่ 3.27

$$\Omega = U^T A \tag{3.27}$$

หรือสามารถเขียนได้อีกแบบหนึ่งตามสมการที่ 3.28

$$\Omega = [w_1, w_2, w_3, ..., w_i]; w_i = U^T[\tau_M - \psi] \tag{3.28}$$

สามารถแสดงได้ดังนี้

```
% Find the weight of each face in the training set.
omega = [];
for h=1:size(dbx,2)
    WW=[];
    for i=1:size(u,2)
        t = u(:,i)';
        WeightOfImage = dot(t,dbx(:,h)');
        WW = [WW; WeightOfImage];
    end
    omega = [omega WW];
end
```

หลังจากนั้นจึงเข้าสู่กระบวนการหาค่าระยะทางแบบยุคลิด ในส่วนขั้นตอนนี้นั้น เราจะนำรูปที่เราจะใช้หาค่ายุคลิด ในที่นี้คือใบหน้าบุคคล ซึ่งเราจะนำมาหาค่าสัมประสิทธิ์ใบหน้าของภาพนั้น โดยสามารถทำได้ตามสมการที่ 3.29

$$\Omega_{Test} = U[\tau_{Test} - \psi_{Test}] \tag{3.29}$$

โดยที่ τ_{Test} คือ เวกเตอร์ของรูปทดสอบที่จะนำมาหาค่ายุคลิด

ψ_{Test} คือ ค่าเฉลี่ยของรูปทดสอบที่จะนำมาหาค่ายุคลิด

U คือ ผลลัพธ์ภาพไอเกนเฟสซึ่งจะมีขนาดเป็น $N^2 \times M$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Ω_{Test} คือ ผลลัพธ์ของค่าสัมประสิทธิ์ใบหน้าของรูปทดสอบ

ซึ่งสามารถออกแบบได้ดังนี้

```
p = [];
 %size of data input
aa=size(u,2);
for i = 1:aa
    pare = dot(NormImage,u(:,i));
    p = [p; pare];
end
ReshapedImage = m + u(:,1:aa)*p;    %m is the mean image, u is the eigenvector
ReshapedImage = reshape(ReshapedImage,icol,irow);
ReshapedImage = ReshapedImage';

InImWeight = [];
for i=1:size(u,2)
    t = u(:,i)';
    WeightOfInputImage = dot(t,Difference');
    InImWeight = [InImWeight; WeightOfInputImage];
end
```

หลังจากนั้นเรานำค่าสัมประสิทธิ์ใบหน้าของภาพนี้ และสัมประสิทธิ์ใบหน้าของฐานข้อมูลมา เสร็จแล้วจึงคำนวณค่าระยะทางแบบยุคลิด โดยสามารถคำนวณได้ตามสมการที่ 3.30

$$e_d = \min || \Omega_{Test} - \Omega || \tag{3.30}$$

ซึ่งสามารถออกแบบได้ดังนี้

```
% Find Euclidean distance
e=[];
for i=1:size(omega,2)
    q = omega(:,i);
    DiffWeight = InImWeight-q;
    mag = norm(DiffWeight);
    e = [e mag];
end
e=e/max(e);
```

เมื่อได้ค่า e_d มาแล้วนั้น ค่า e_d ของเราจะมีขนาดเท่ากับ $1 \times M$ โดยที่ M คือจำนวนรูปที่เราใช้ในฐานข้อมูล d คือจำนวนชุดของฐานข้อมูล เสร็จแล้วจึงนำไปเข้ากระบวนการเลือกคุณลักษณะเด่นของสัญญาณเพื่อนำไปใช้ในโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.8 เรียงลำดับข้อมูล

เนื่องจากการสังเกตค่ายุคลิดของระบบนั้น เมื่อนำภาพของกลุ่มบุคคลเดียวกันมาทดสอบค่าผลลัพธ์ของยุคลิดของกลุ่มบุคคลเดียวกัน บางค่าจะสูง บางค่าจะต่ำ ซึ่งไม่มีความสัมพันธ์กันเลย จึงได้ทำการออกแบบระบบโดยทำการ "เรียงลำดับ" โดยจะใช้คำสั่ง sort เพื่อทำการเรียงลำดับของค่ายุคลิดของกลุ่มบุคคลเดียวกันจากน้อยไปมาก

3.2.9 การนอมอลไลซ์ทางขนาดของข้อมูล

เนื่องจากค่าระยะทางแบบยุคลิดของแต่ละอันนั้น มีค่าที่มีความใกล้เคียงกันมากเกินไป จึงทำการนอมอลไลซ์ทางขนาดของข้อมูลเพื่อให้มีความเหมาะสมก่อนนำไปเข้าโครงข่ายประสาทเทียม โดยนำ 1 ไปลบกับค่า e_d แล้วนำ 2.5 ไปคูณกับ e_d ใหม่ เป็นอันเสร็จสิ้นกระบวนการเลือกคุณลักษณะเด่น โดยสามารถออกแบบได้ดังนี้และจึงนำไปเข้าระบบโครงข่ายประสาทเทียมต่อไป

```
%optimize feature,sort,1-,*2
for j=1 : 1
    fak(j,1:30)=sort(e(j,1:30));
    fak(j,31:60)=sort(e(j,31:60));
    fak(j,61:90)=sort(e(j,61:90));
    fak(j,91:120)=sort(e(j,91:120));
end
fak=(1-fak);
fak=fak*2;
```

3.3 ออกแบบระบบโครงข่ายประสาทเทียมของการรู้จำภาพและเสียง

3.3.1 การออกแบบโครงข่ายประสาทเทียมส่วนของการรู้จำเสียง

การออกแบบจะมีการกำหนดพารามิเตอร์ต่าง ๆ แต่ละระบบโครงข่ายประสาทเทียมเป็นดังนี้ กำหนดอินพุตโหนดเท่ากับ 4320 โหนด จำนวนโหนดในชั้นซ้อนเร้น เท่ากับ 30 โหนด และเอาต์พุตโหนดจำนวน 10 โหนด สำหรับกลุ่ม 5 พยางค์ขึ้นไป กรณีที่เป็น 5 พยางค์จะมีจำนวนเอาต์พุตโหนดเท่ากับ 6 โหนดและ 6 พยางค์จะมีเอาต์พุตโหนด 4 โหนด ซึ่งแสดงพารามิเตอร์ของโครงข่ายประสาทเทียมสำหรับการรู้จำเสียงได้โดยง่ายดังรูปที่ 3.11 โดยการทำงานของระบบโครงข่ายประสาทเทียมเบื้องต้นจะทำตรวจสอบจำนวนพยางค์ของสัญญาณเสียงก่อน (จากกระบวนการเตรียมสัญญาณเบื้องต้น) เมื่อทราบแล้วว่าจำนวนพยางค์ที่เกิดขึ้นเป็นเท่าใดจะส่งข้อมูลไปประมวลผลในระบบโครงข่ายประสาทเทียมตามจำนวนของพยางค์นั้น ๆ และทำการประมวลผลแยกพยางค์ในโครงข่ายประสาทเทียม (โดยจะทำการรู้จำแบบแยกพยางค์กรณีคำที่มี 3 พยางค์ขึ้นไป) ผลลัพธ์ที่ได้ของเอาต์พุตโหนดของแต่ละพยางค์ที่

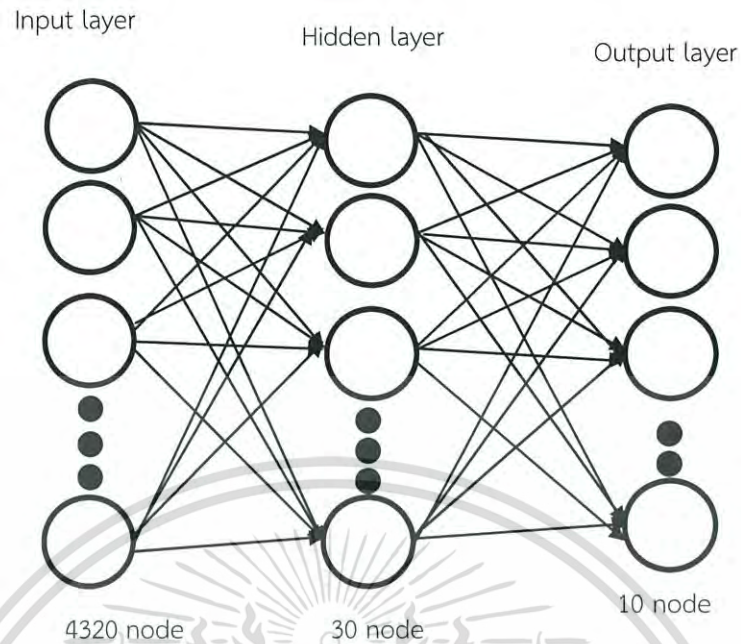
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการประมวลผลในโครงข่ายประสาทเทียมจะมีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งในโครงข่ายประสาทเทียมจะมีการแบ่งออกเป็น 2 ส่วนด้วยกันคือ ส่วนสำหรับการรู้จำหรือเรียนรู้ และส่วนสำหรับการจำแนก

ซึ่งส่วนสำหรับการรู้จำหรือเรียนรู้นั้นจะอ้างอิงหลักการทำงานได้ตามบทที่ 2 แต่จะมีการเพิ่มเติมของระบบเพื่อเพิ่มประสิทธิภาพในการจำแนกให้ดียิ่งขึ้น กล่าวคือจะทำการรู้จำแยกพยางค์สำหรับกลุ่มคำตั้งแต่ 3 พยางค์ขึ้นไปจากที่กล่าวมาข้างต้น โดยจะทำการหาค่าถ่วงน้ำหนัก (weight) และค่าไบอัส แต่ละพยางค์ของแต่ละคำเพื่อมาจำแนกในส่วนสำหรับการจำแนกต่อไป

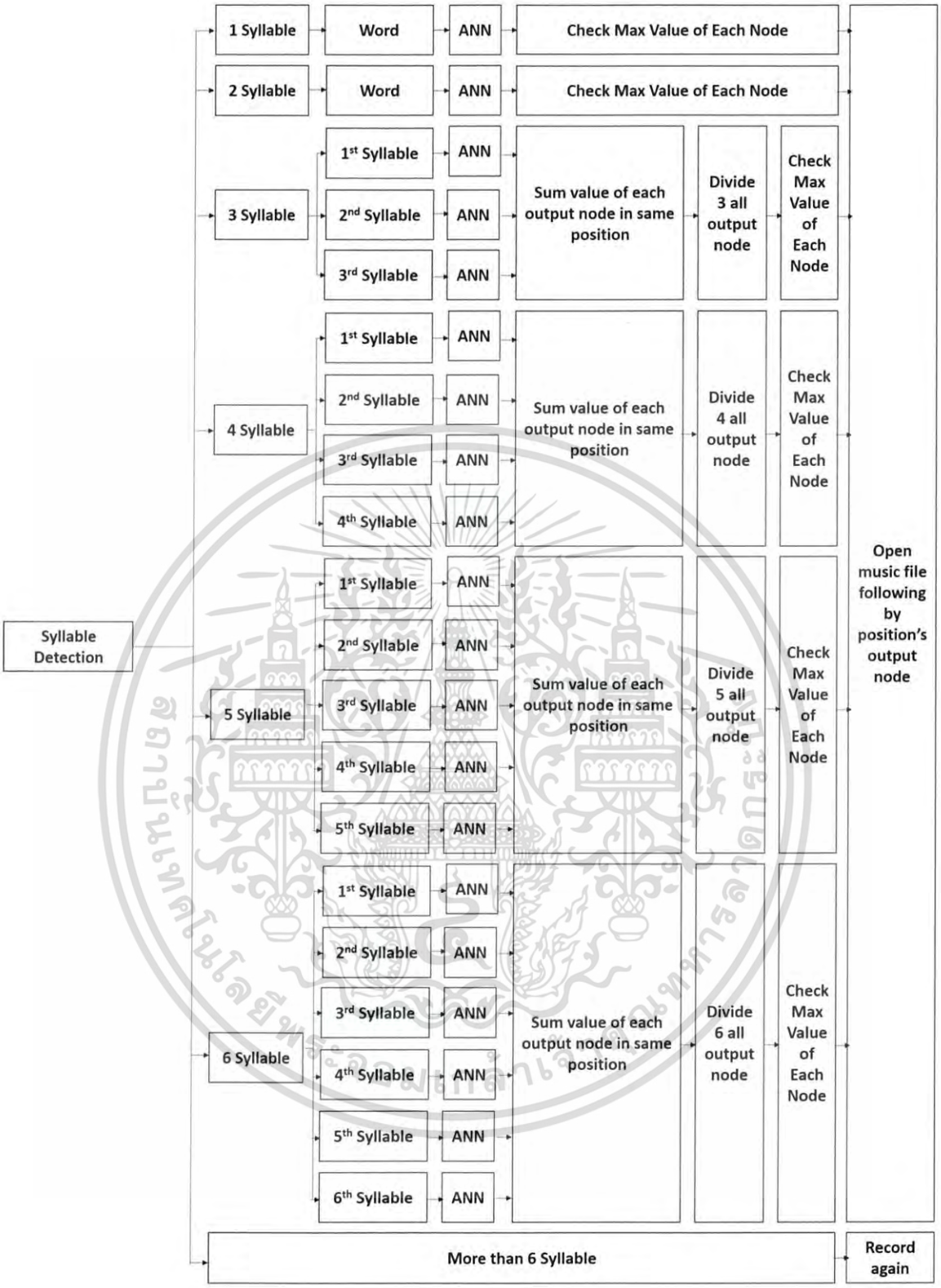
ต่อมาในส่วนสำหรับการจำแนกสามารถอธิบายได้ดังรูปที่ 3.12 โดยการทำงานเมื่อผ่านกระบวนการต่าง ๆ ที่ได้กล่าวมาข้างต้นแล้วได้มาซึ่งค่าเอาต์พุตโหนดในแต่ละโครงข่ายประสาทเทียมตามแต่ละพยางค์ จะนำค่าเอาต์พุตโหนดที่ได้ที่ตำแหน่งเดียวกันทุกระบบโครงข่ายประสาทเทียมนั้น ๆ มารวมกันแล้วหารด้วยจำนวนพยางค์ที่ได้ทำการตรวจสอบพยางค์ก่อนหน้านี้ จากนั้นจะได้ผลลัพธ์ของเอาต์พุตโหนดรวมที่มีค่าอยู่ระหว่าง 0 ถึง 1 ต่อมาทำการตรวจสอบว่าที่ตำแหน่งใดของเอาต์พุตโหนดของโครงข่ายประสาทเทียมนั้น ๆ มีค่ามากที่สุด ให้ที่ตำแหน่งของเอาต์พุตโหนดนั้นเปิดไฟล์เพลงที่ต้องการเป็นต้น เช่น กรณีเมื่อมีการพูดสัญญาณเสียง 3 พยางค์จะทำการแยกพยางค์ของเสียงนั้นออกเป็น 3 ส่วน (จากกระบวนการเตรียมสัญญาณเสียงเบื้องต้น) คือส่วนที่ 1 คือ พยางค์ที่ 1 ของคำ ส่วนที่ 2 คือ พยางค์ที่ 2 ของคำ และ ส่วนสุดท้าย คือ พยางค์ที่ 3 ของคำ ผ่านกระบวนการต่าง ๆ ได้มาซึ่งค่าสัมประสิทธิ์สำหรับนำมาประมวลผลในโครงข่ายประสาทเทียมของแต่ละส่วน ต่อมาเมื่อโครงข่ายประสาทเทียมทราบแล้วว่าเสียงที่ทำการพูดนั้นเป็น 3 พยางค์จะนำผลลัพธ์แต่ละส่วนมาประมวลผลในระบบโครงข่ายประสาทเทียมสำหรับ 3 พยางค์ที่ได้มีการออกแบบไว้ โดยส่วนที่ 1 จะทำการประมวลผลในโครงข่ายประสาทเทียมสำหรับพยางค์ที่ 1 ของสัญญาณเสียง 3 พยางค์, ส่วนที่ 2 จะทำการประมวลในโครงข่ายประสาทเทียมสำหรับพยางค์ที่ 2 ของสัญญาณเสียง 3 พยางค์ และส่วนสุดท้ายจะทำการประมวลในโครงข่ายประสาทเทียมสำหรับพยางค์ที่ 3 ของสัญญาณเสียง 3 พยางค์เป็นต้น ผลลัพธ์ที่ได้จะให้ค่าเอาต์พุตโหนดทั้งหมด 10 โหนดโดยที่มีค่าอยู่ระหว่าง 0 ถึง 1 ทั้ง 3 ระบบโครงข่ายประสาทเทียม ต่อมานำค่าของเอาต์พุตโหนดทั้ง 3 ระบบโครงข่ายประสาทเทียมมาบวกกันตามตำแหน่งของเอาต์พุตโหนด (เช่น เอาต์พุตโหนดที่ 1 ทุก 3 ระบบโครงข่ายประสาทเทียมมาบวกกันเป็นต้น) จากนั้นหารด้วย 3 (จำนวนพยางค์ที่ทำการตรวจสอบได้) แล้วทำการตรวจสอบว่าที่ตำแหน่งใดของเอาต์พุตโหนดให้ค่าสูงที่สุดให้ตำแหน่งนั้นเปิดไฟล์เพลงหรือจำแนกข้อมูลดังที่ได้ทำการออกแบบไว้เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 ระบบโครงข่ายประสาทเทียมสำหรับการรู้จำชื่อเพลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

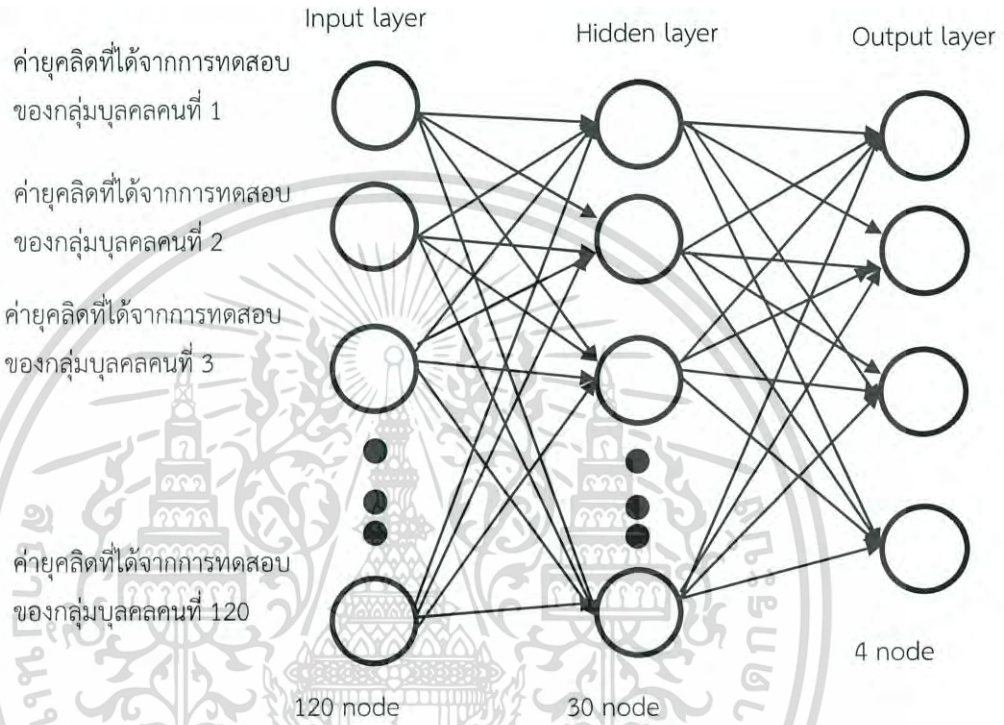


รูปที่ 3.10 กระบวนการจำแนกของระบบโครงข่ายประสาทเทียมของระบบรู้จำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.2 การออกแบบโครงข่ายประสาทเทียมในส่วนของการรู้จำภาพใบหน้ากลุ่มบุคคล

ในส่วนของภาพนั้น จะทำการดึงผลลัพธ์ค่า e_d ที่ได้จากขั้นตอนการหาค่าระยะทางแบบยุคลิด โดยทำการรู้จำกลุ่มบุลคลทั้งหมด 4 กลุ่ม คือ ผู้ชาย ผู้หญิง เด็กและผู้สูงอายุ เพื่อจะนำไปใช้ในการจำแนกกลุ่มบุคคล โดยสามารถแสดงเป็นบล็อกไดอะแกรมของโครงข่ายประสาทเทียมได้ดังรูปที่ 3.11



รูปที่ 3.11.ระบบโครงข่ายประสาทเทียมสำหรับการรู้จำภาพใบหน้า

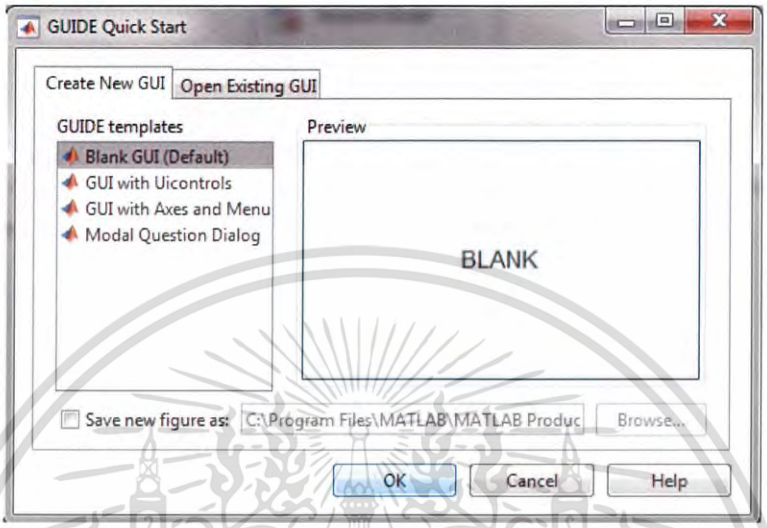
3.3.3 แอพพลิเคชันของระบบโครงข่ายประสาทเทียมของการรู้จำภาพและเสียง

หลังจากที่ออกแบบระบบรู้จำทั้ง 2 ส่วนเสร็จแล้ว จึงออกแบบรูปแบบของการแสดงผลออกมาโดยผ่านทางโปรแกรม MATLAB โดยระบบรู้จำเสียงพูดนั้นจะออกแบบโดยใช้ MATLAB GUI และระบบรู้จำภาพใบหน้าของกลุ่มบุคคลนั้นจะแสดงผลผ่าน .m file โดยที่ GUI นั้นสามารถออกแบบระบบได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

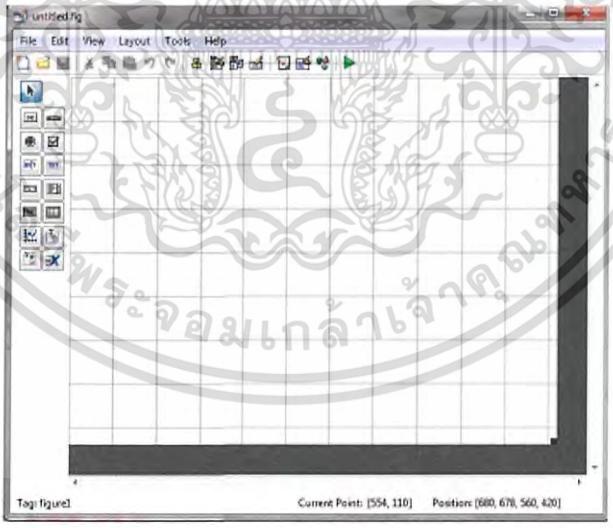
1.) เปิดโปรแกรม Matlab Professional ขึ้นมาและไปที่ Main Menu ที่มีชื่อว่า Home ->New->Graphical User Interface ดังรูป

2.) จะขึ้นหน้าต่าง GUIDE Quick Start ขึ้นมาต่อจากนั้นให้ไปที่ Main Menu ที่มีชื่อว่า Create New GUI->GUI Templates->Blank GUI (Default)->แล้วกด OK ดังรูปที่ 3.12



รูปที่ 3.12 หน้าต่างเลือกรูปแบบการใช้งาน MATLAB GUI

3.) โดยจะแสดงหน้าต่างที่ใช้ในการออกแบบ Graphical User Interface ขึ้นมาดังรูปที่ 3.13



รูปที่ 3.13 หน้าต่าง MATLAB GUI

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลอง

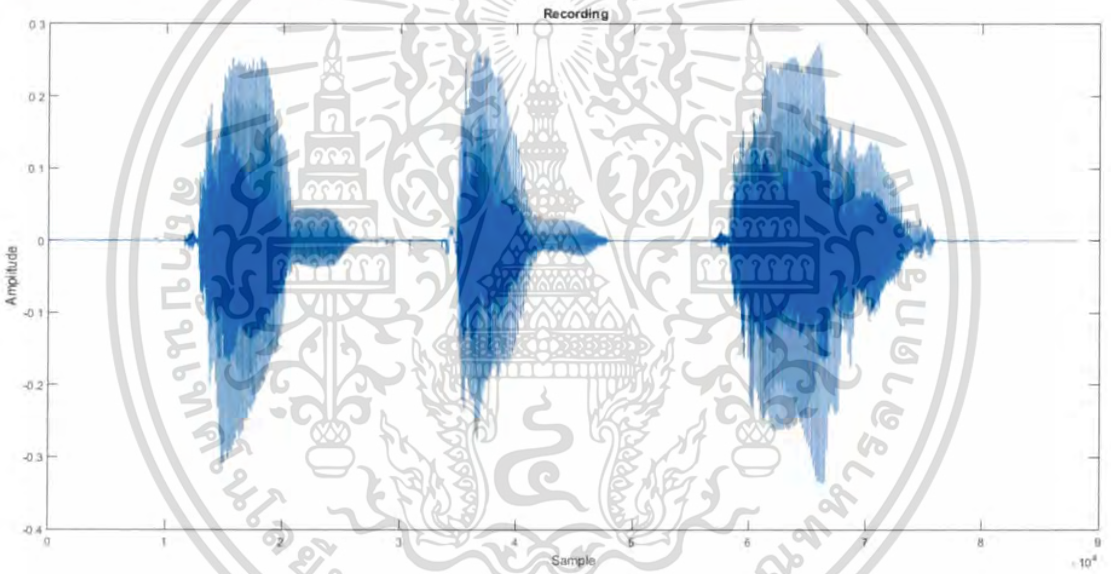
การทดลองนั้นจะมี 2 ส่วน คือผลการทดลองของระบบโครงข่ายประสาทเทียมของเสียงและภาพ

4.1 ระบบโครงข่ายประสาทเทียมสำหรับการรู้จำชื่อเพลง

โดยจะทำการแสดงผลลัพธ์ในแต่ละขั้นตอนของการทำงานซึ่งจะใช้เสียงผู้หญิงคำว่า “จำฝังใจ” มาประมวลผล

4.1.1 Recording

ทำการอัดเสียงคำว่า “จำฝังใจ” ซึ่ง Sampling ด้วยความถี่ 44.1 kHz ขนาด 16 บิตโดยใช้เวลาบันทึก 2 วินาที เป็นดังรูปที่ 4.1

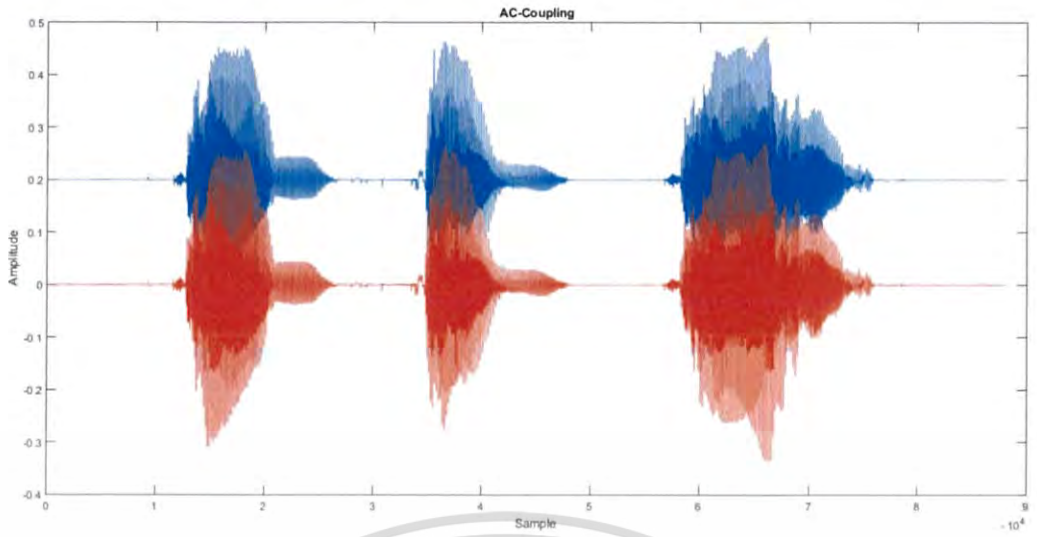


รูปที่ 4.1 สัญญาณเสียงคำว่า จำฝังใจ

4.1.2 AC-Coupling

จากนั้นจะนำผลลัพธ์ที่ได้ไปขจัดสัญญาณไฟตรง โดยทำการเปรียบเทียบสัญญาณเสียงคำว่า “จำฝังใจ” ที่ผ่านกระบวนการ AC-Coupling กับไม่ผ่านกระบวนการได้ดังนี้รูปที่ 4.2

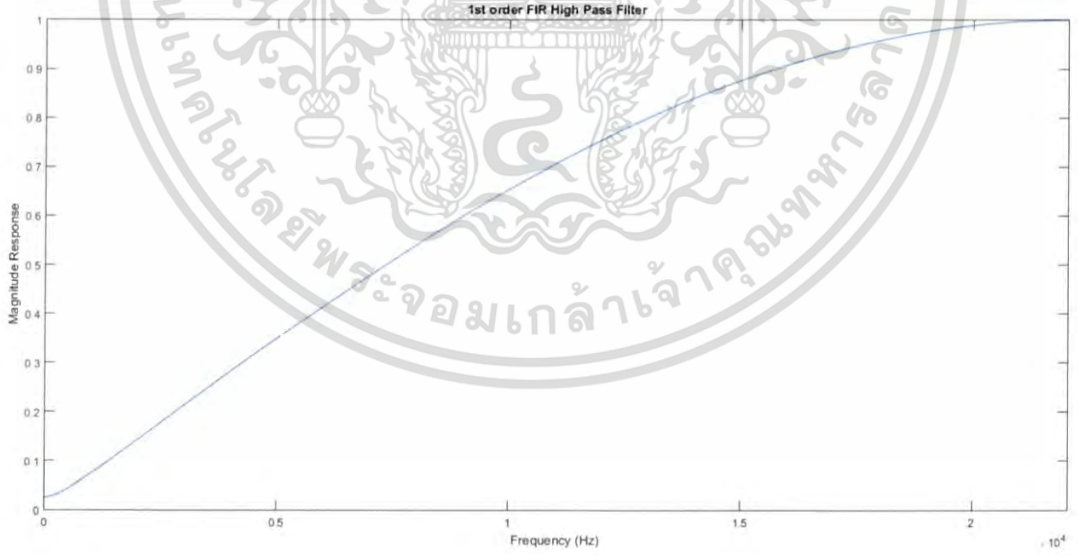
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่เป็นสัญญาณที่ผ่านกระบวนการ AC-Coupling กับสัญญาณสีน้ำเงินที่เป็นสัญญาณอินพุตที่มีไฟตรงเท่ากับ 0.2 โวลต์

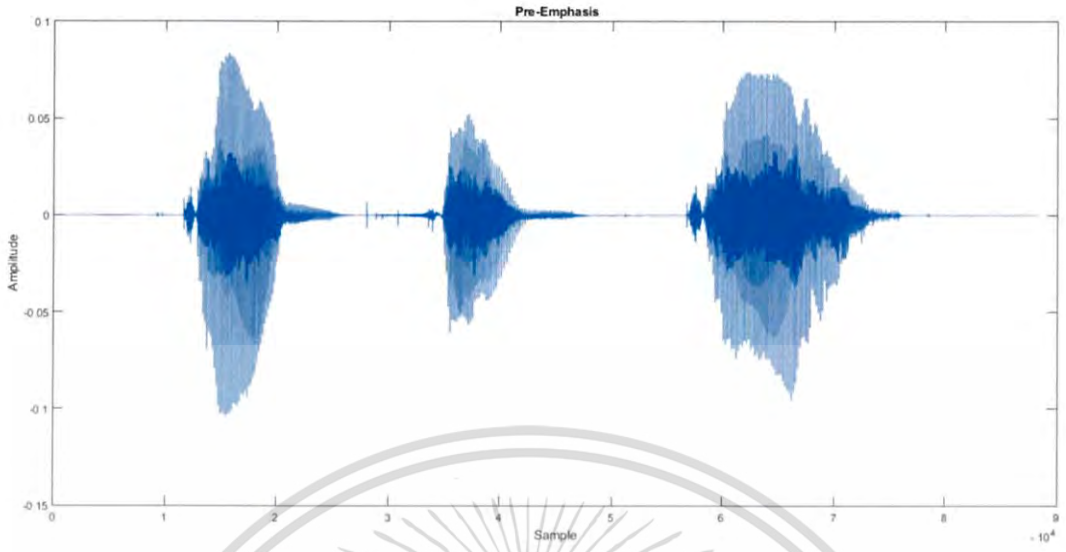
4.1.3 Pre-Emphasis

จากนั้นจะนำผลลัพธ์ที่ผ่านกระบวนการ AC-Coupling แล้ว ไปผ่านวงจรกรองความถี่สูงผ่านที่แสดงผลตอบสนองทางขนาดเป็นดังรูปที่ 4.3 โดยสามารถแสดงผลลัพธ์ได้ดังรูปที่ 4.4 และ รูปที่ 4.5



รูปที่ 4.3 ผลตอบสนองทางขนาดของวงจรกรองความถี่สูงผ่าน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



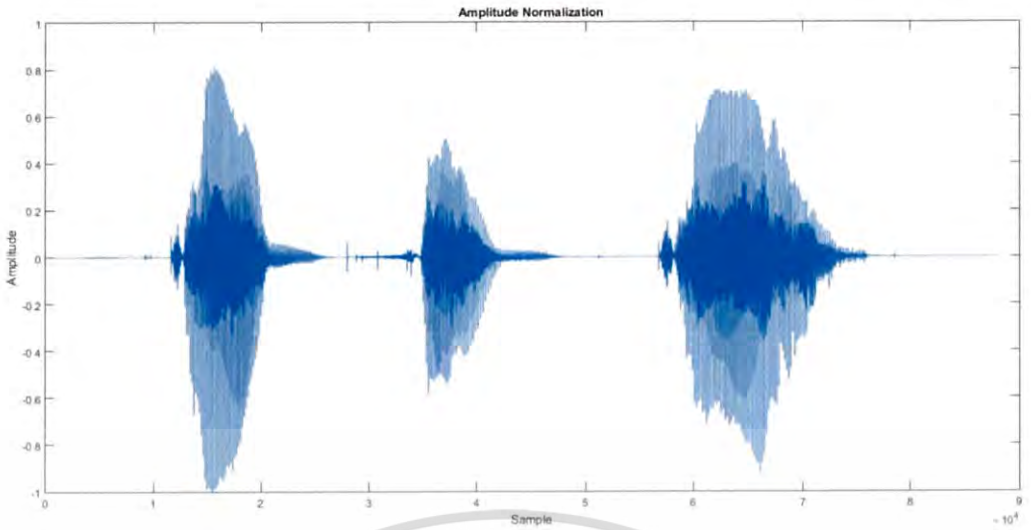
รูปที่ 4.4 สัญญาณที่ผ่านกระบวนการ Pre Emphasis

รูปที่ 4.5 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่เป็นสัญญาณที่ผ่านกระบวนการ Pre-Emphasis กับสัญญาณสีน้ำเงินที่เป็นสัญญาณที่ผ่านกระบวนการ AC-Coupling

4.1.4 กระบวนการทำการนอมอไลซ์ทางขนาด

ต่อมาจะนำสัญญาณมาทำการนอมอไลซ์ทางขนาดโดยสัญญาณที่ได้นั้นจะมีค่าสูงสุดซึ่งอยู่ระหว่าง 1 ถึง -1 และสามารถแสดงผลลัพธ์ได้ดังรูปที่ 4.6

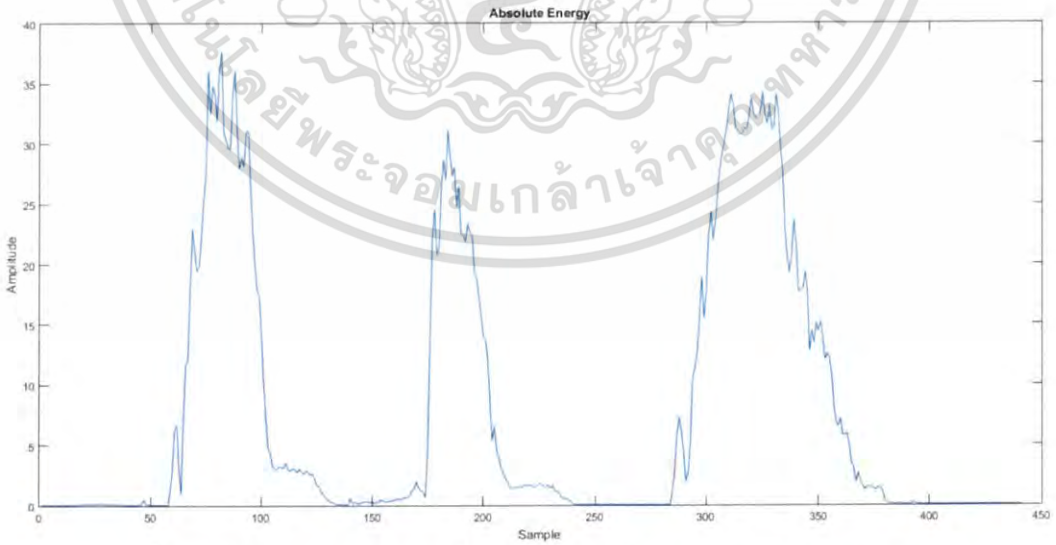
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 สัญญาณที่ผ่านกระบวนการนอมอไลซ์ทางขนาด

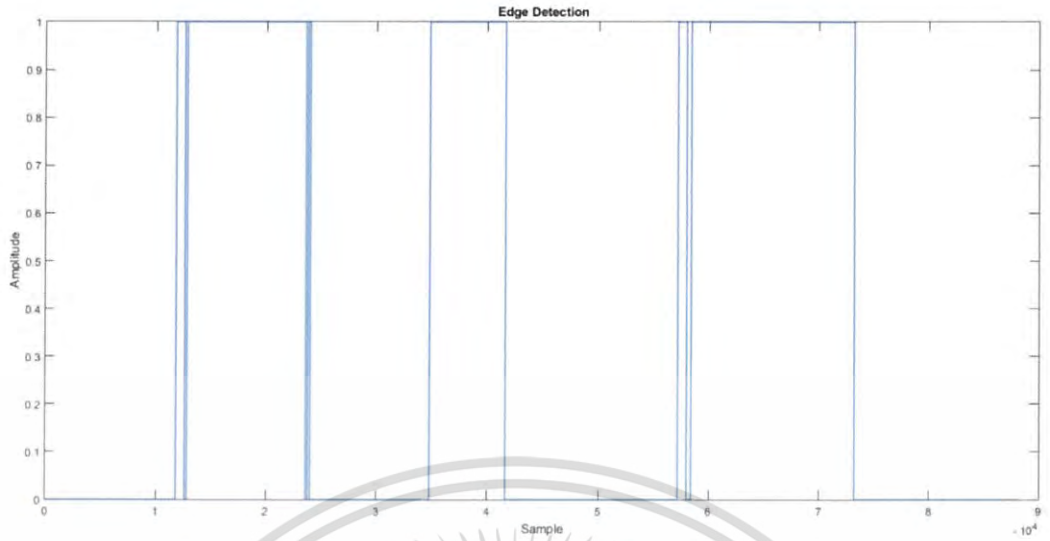
4.1.5 Endpoint Detection

นำผลลัพธ์ที่ได้จากกระบวนการนอมอไลซ์ทางขนาด ไปทำการหาพลังงานของสัญญาณ โดยแสดงผลลัพธ์ออกมาดังรูปที่ 4.7 จากนั้นนำผลลัพธ์ที่ได้ไปทำการหากรอบสัญญาณในการตัดคำได้ดังรูปที่ 4.8 ต่อมารูปที่ 4.9 แสดงผลลัพธ์ของกรอบสัญญาณใหม่ที่เกิดจากการตัดคำไม่สมบูรณ์และแสดงการเปรียบเทียบกรอบสัญญาณใหม่กับสัญญาณที่ผ่านกระบวนการนอมอไลซ์ทางขนาด ดังรูปที่ 4.10 และแสดงผลลัพธ์ของการตัดคำได้ดังรูปที่ 4.11 โดยแสดงผลลัพธ์ของแต่ละพยางค์ในการตัดคำได้ดังรูปที่ 4.12 รูปที่ 4.13 และรูปที่ 4.14

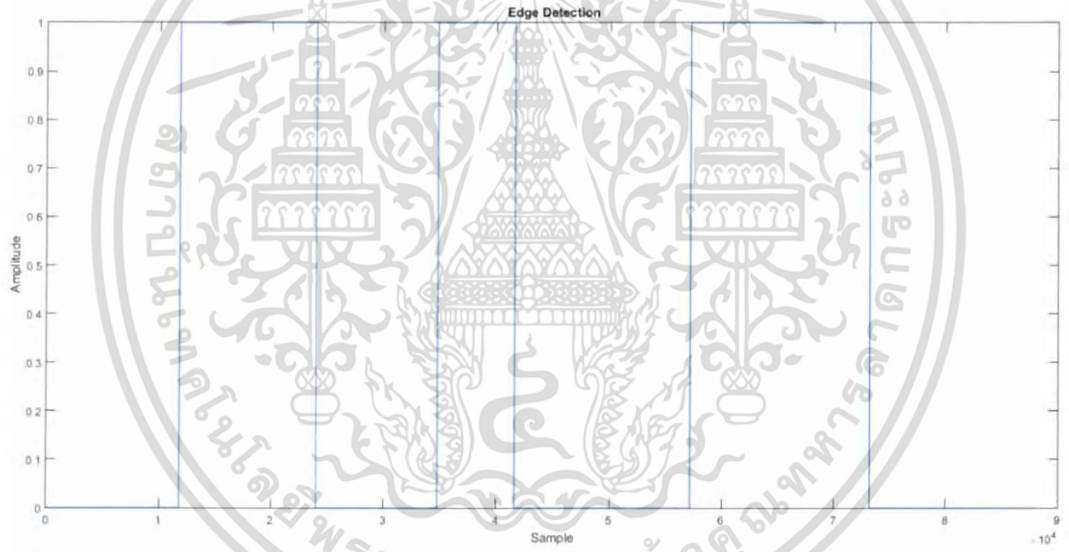


รูปที่ 4.7 พลังงานของสัญญาณที่ใช้วิธีการหาพลังงานแบบ Absolute Energy

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

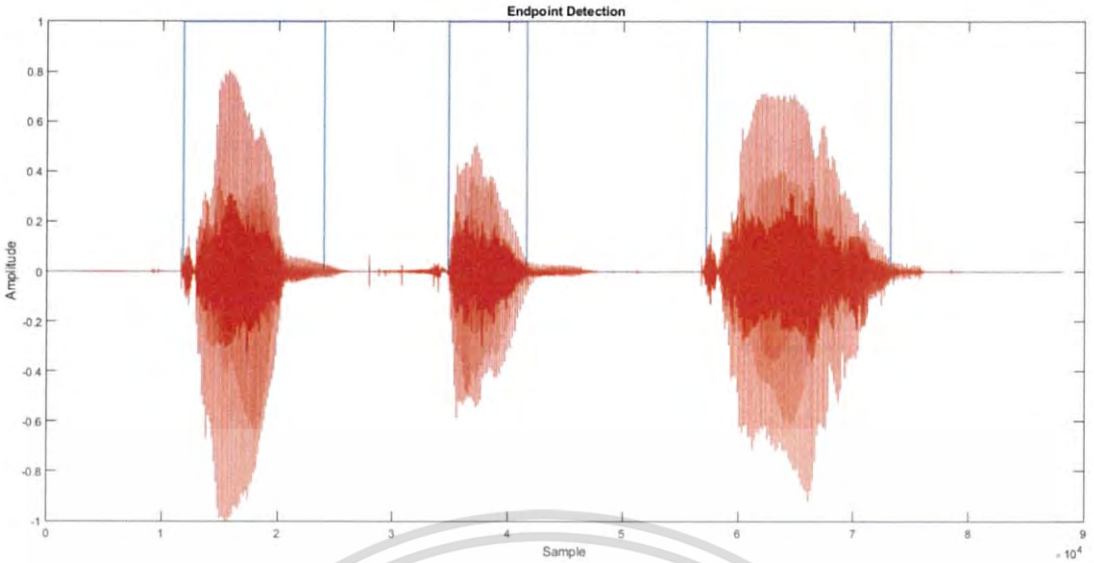


รูปที่ 4.8 กรอบสัญญาณที่ทำการตัดคำ



รูปที่ 4.9 กรอบสัญญาณใหม่ที่ทำการตัดคำ

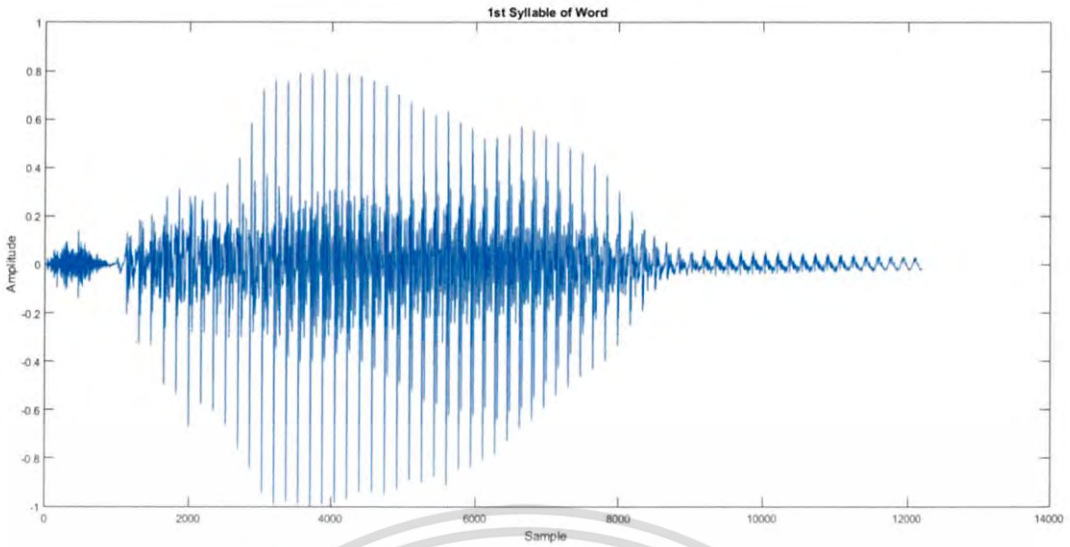
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



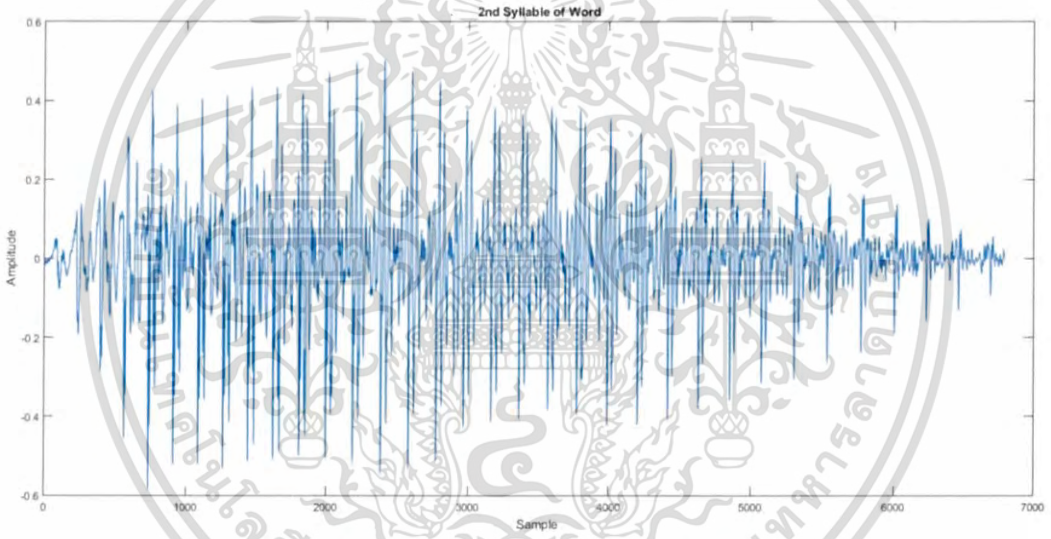
รูปที่ 4.10 เปรียบเทียบกรอบสัญญาณใหม่กับสัญญาณที่ผ่านกระบวนการนอมอไลซ์ทางขนาด

รูปที่ 4.11 สัญญาณที่เป็นผลลัพธ์จากการตัดพยางค์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

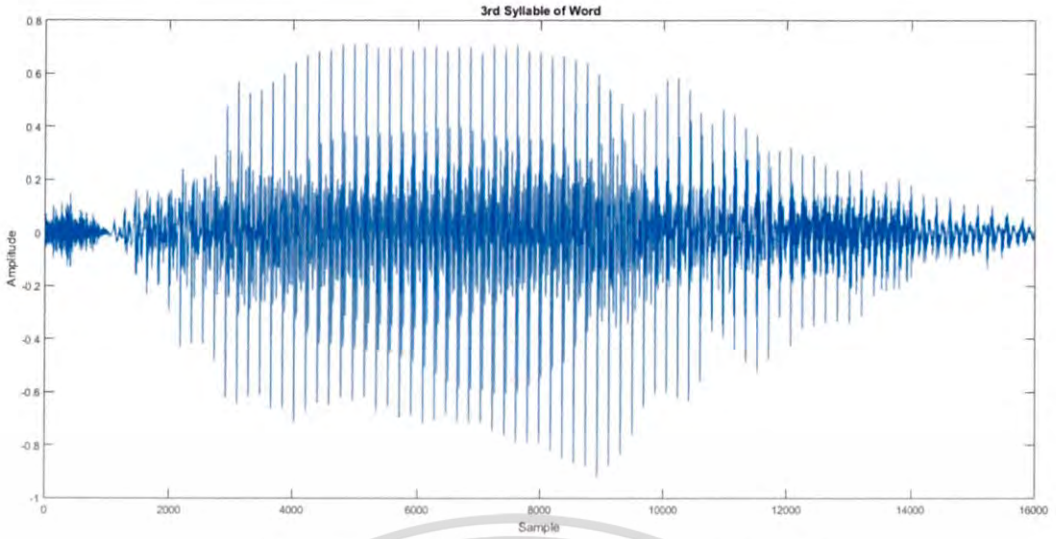


รูปที่ 4.12 สัญญาณของเสียงพยางค์ที่ 1 หรือคำว่า “จำ”



รูปที่ 4.13 สัญญาณของเสียงพยางค์ที่ 2 หรือคำว่า “ฝัง”

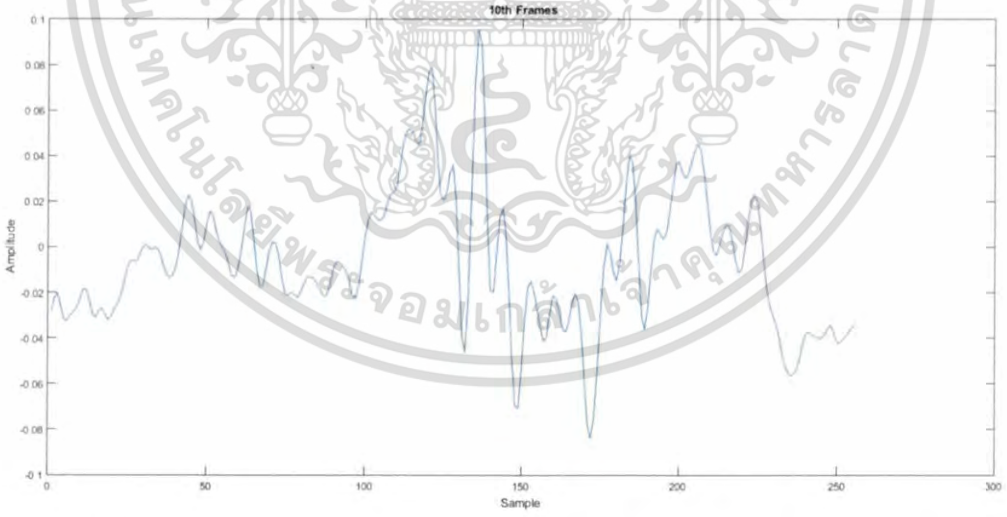
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.14 สัญญาณของเสียงพยางค์ที่ 3 หรือคำว่า “ใจ”

4.1.6 Pre-Processing for MFCC

โดยในขั้นตอนการดึงคุณลักษณะเด่นของสัญญาณจะนำสัญญาณของเสียงพยางค์ที่ 3 หรือคำว่า “ใจ” มาแสดงผลลัพธ์ในการขั้นตอนต่าง ๆ โดยในกระบวนการ Pre-Processing for MFCC จะนำสัญญาณที่ได้จากการตัดคำในกระบวนการเตรียมสัญญาณเสียงเบื้องต้นมาแบ่งเป็นเฟรมข้อมูลซึ่งสามารถแสดงตัวอย่างของเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 หรือคำว่า “ใจ” ได้ดังรูปที่ 4.15

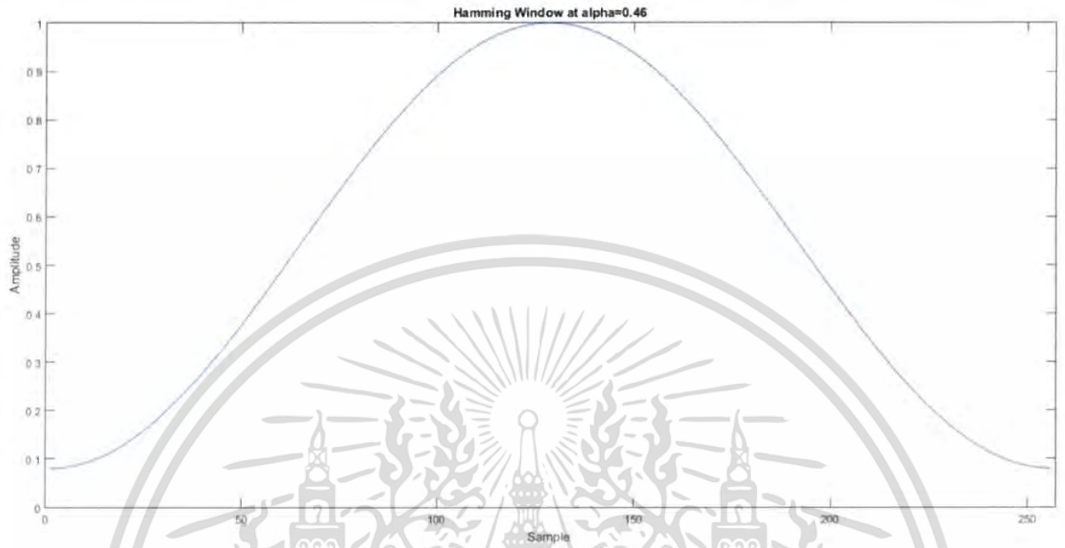


รูปที่ 4.15 ตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

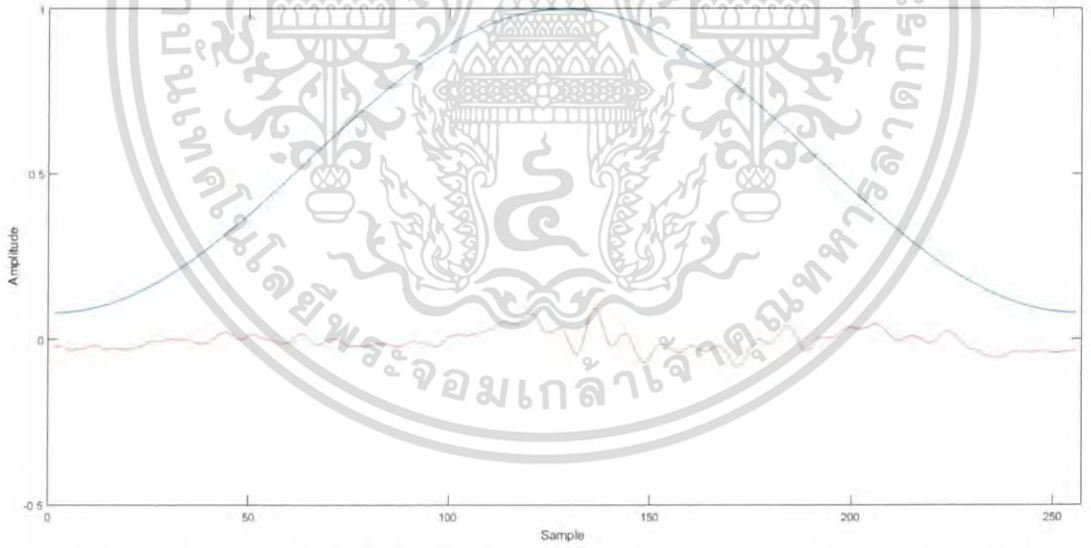
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.7 แฮมมิ่งวินโดว์

จากนั้นจะนำผลลัพธ์ที่ได้ไปผ่านแฮมมิ่งวินโดว์ ซึ่งหน้าตาของแฮมมิ่งวินโดว์แสดงได้ดังรูปที่ 4.16 เปรียบเทียบเฟรมสัญญาณกับแฮมมิ่งวินโดว์ได้ดังรูปที่ 4.17 และแสดงผลลัพธ์ของสัญญาณที่ผ่านแฮมมิ่งวินโดว์เปรียบเทียบกับสัญญาณที่ยังไม่ผ่านแฮมมิ่งวินโดว์ได้ดังรูปที่ 4.18

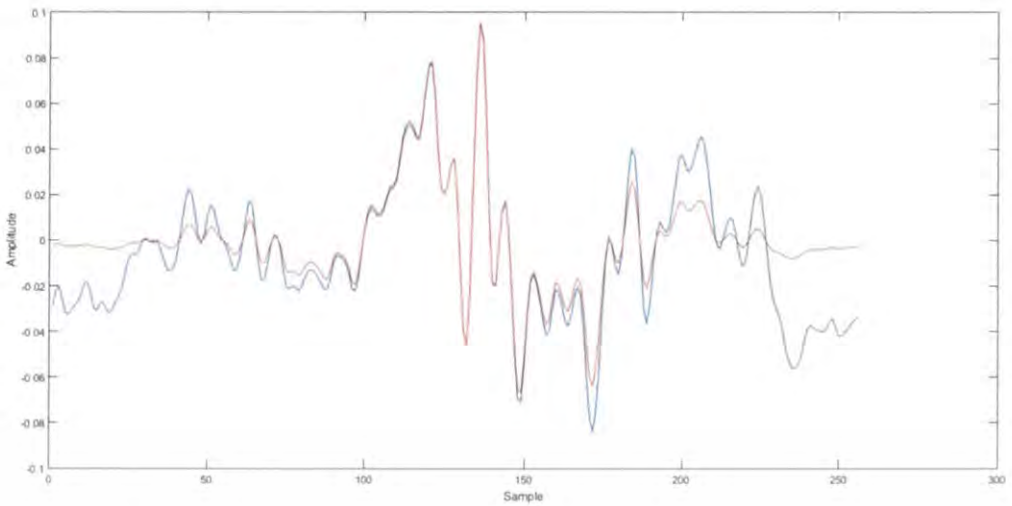


รูปที่ 4.16 กราฟแสดงแฮมมิ่งวินโดว์



รูปที่ 4.17 เปรียบเทียบแฮมมิ่งวินโดว์กับตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

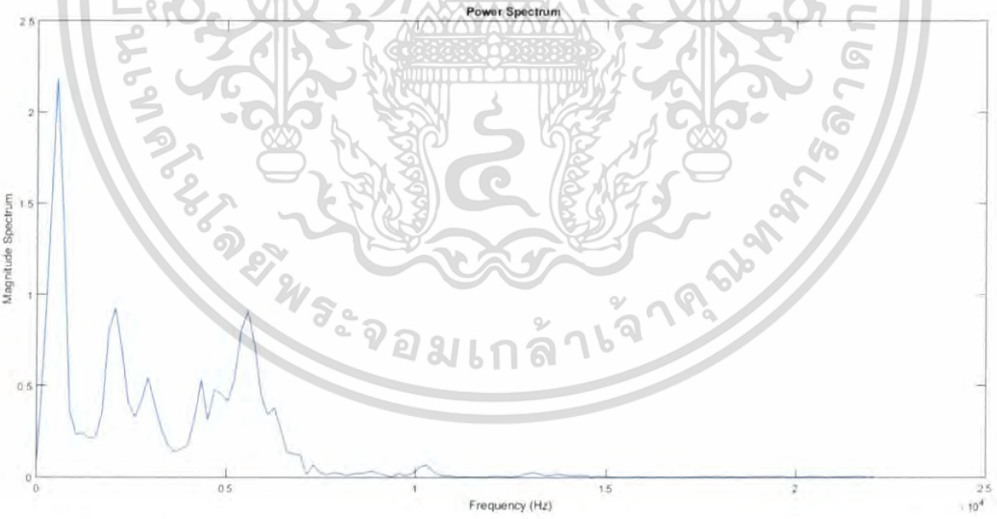
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 เปรียบเทียบสัญญาณระหว่างสัญญาณสีแดงที่เป็นตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” ที่ผ่านแฮมมิ่งวินโดว์กับสัญญาณสีน้ำเงินที่เป็นตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” ที่ยังไม่ผ่านแฮมมิ่งวินโดว์

4.1.8 Power Spectrum

จากนั้นนำเฟรมของสัญญาณที่ผ่านแฮมมิ่งวินโดว์แล้วไปทำการหาสเปกตรัมเชิงกำลังได้ดังรูปที่ 4.19

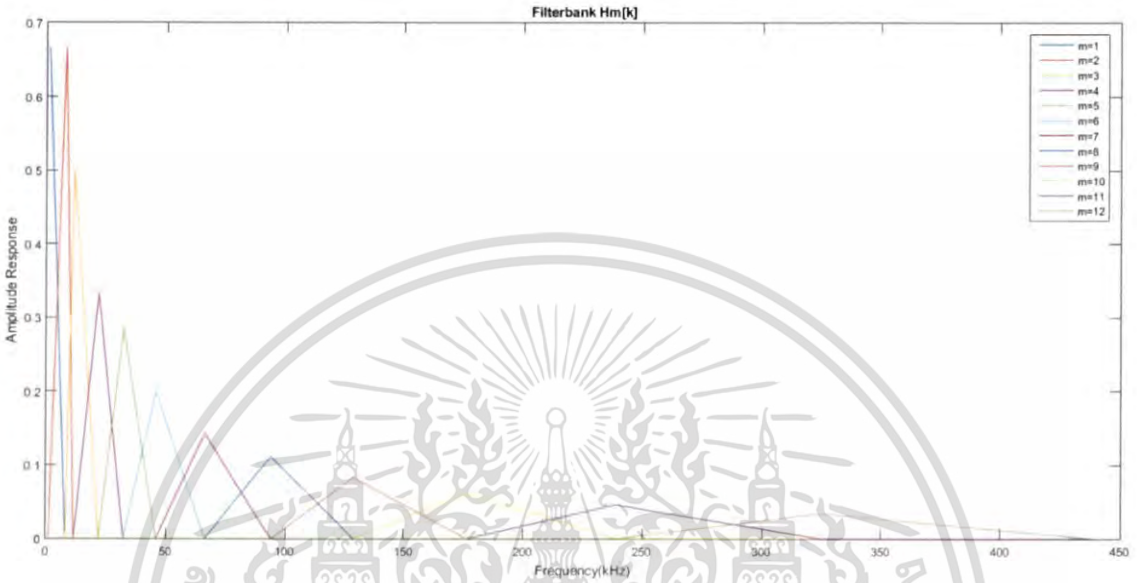


รูปที่ 4.19 สเปกตรัมของตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

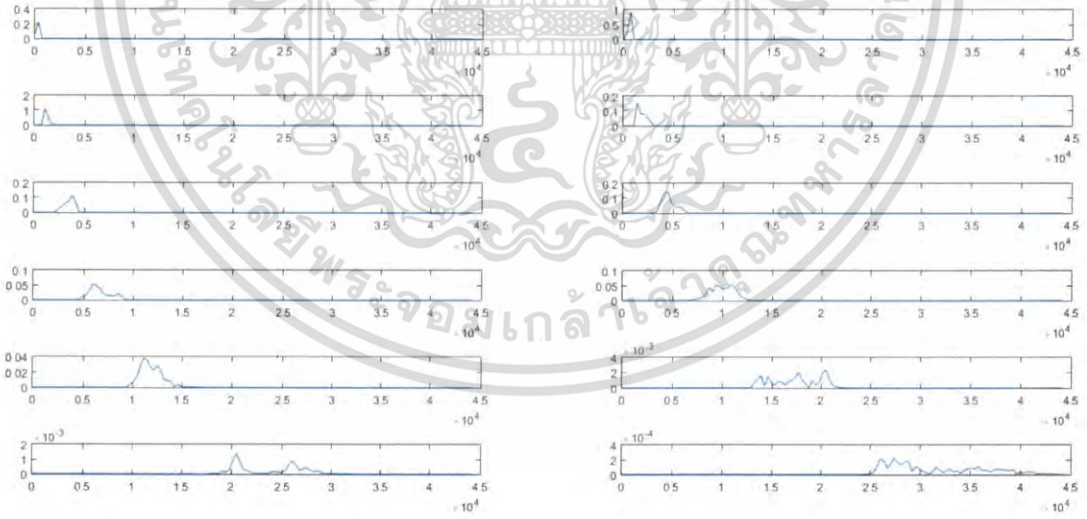
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.9 Mel Spectrum & Mel-Scale Filterbank

เมื่อได้สเปกตรัมเชิงกำลังของสัญญาณแล้วจะนำผลลัพธ์ที่ได้ไปผ่านฟิลเตอร์แบงค์ โดยที่ผลตอบสนองทางขนาดของฟิลเตอร์แบงค์แสดงได้ดังรูปที่ 4.20 และแสดงผลลัพธ์ของสเปกตรัมเชิงขนาดที่ผ่านฟิลเตอร์แบงค์แล้วดังรูปที่ 4.21



รูปที่ 4.20 กราฟแสดงผลตอบสนองทางขนาดของฟิลเตอร์แบงค์

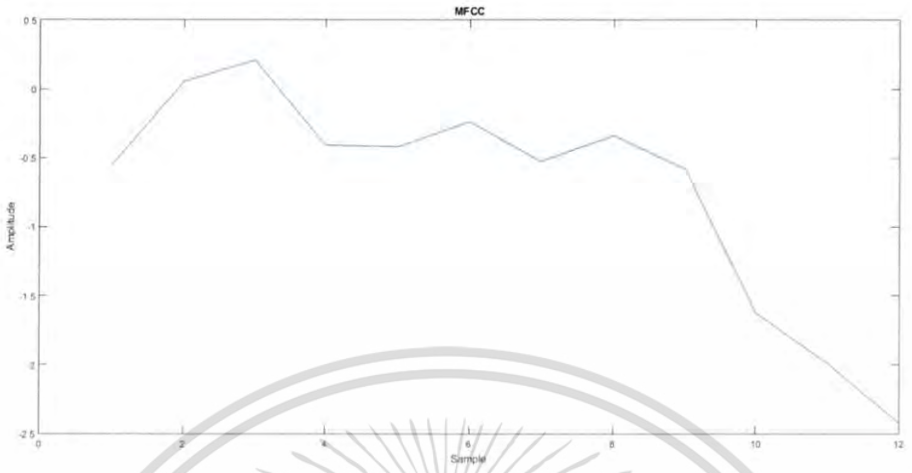


รูปที่ 4.21 แสดงสเปกตรัมของตัวอย่างเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” ที่ผ่านฟิลเตอร์แบงค์ที่ 1 ถึง 12 โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.10 กระบวนการหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล

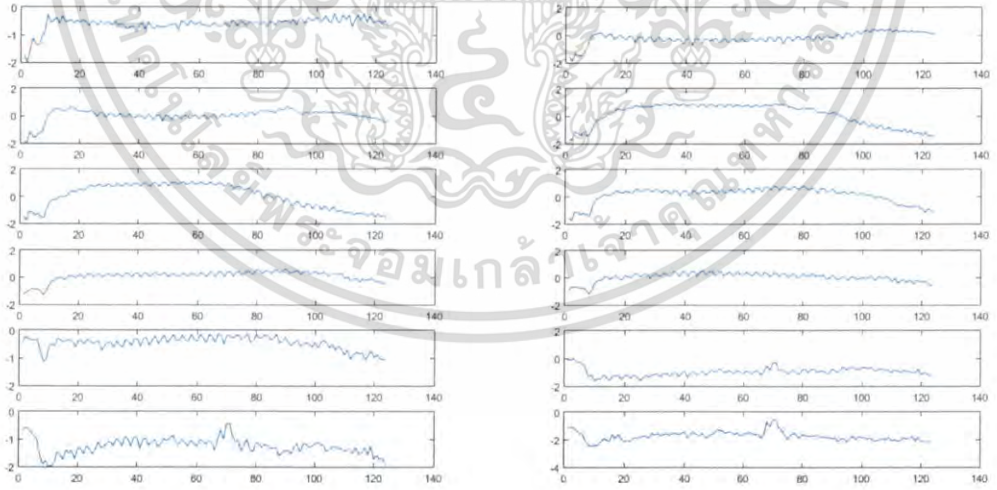
ทำการหาค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ได้ดังรูปที่ 4.22



รูปที่ 4.22 ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลของเฟรมที่ 10 ของสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

4.1.11 MFCC Arrangement

นำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลที่ลำดับเดียวกันในแต่ละเฟรมมาเรียงต่อกัน โดยเรียงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลตามลำดับของเฟรมข้อมูลตั้งแต่เฟรมที่ 1 ถึงเฟรมสุดท้ายได้ดังรูปที่ 4.23

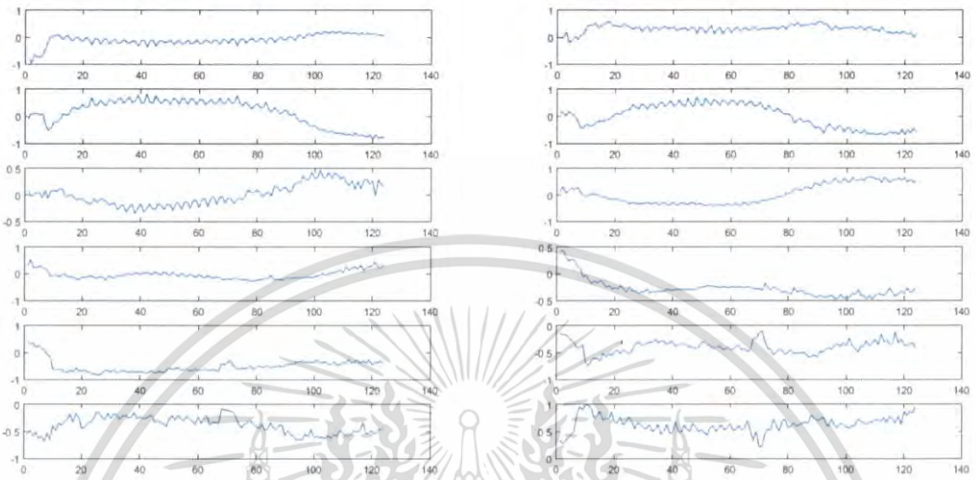


รูปที่ 4.23 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลลำดับที่ 1 ถึง 12 ที่นำมาเรียงกันตามลำดับเฟรมข้อมูลของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ

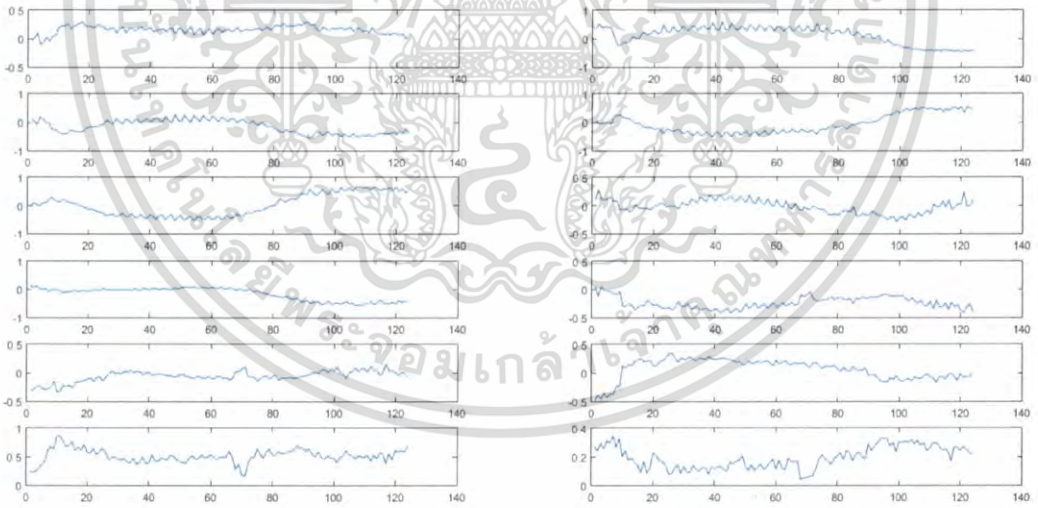
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.12 การหาค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า

เมื่อทำการเรียงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลในกระบวนการ MFCC Arrangement แล้ว จะทำการหาค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้าและแสดงผลลัพธ์ที่ได้ดังรูปที่ 4.24 และรูปที่ 4.25 ตามลำดับ



รูปที่ 4.24 แสดงค่าสัมประสิทธิ์เดลต้าที่ 1 ถึง 12 ของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ

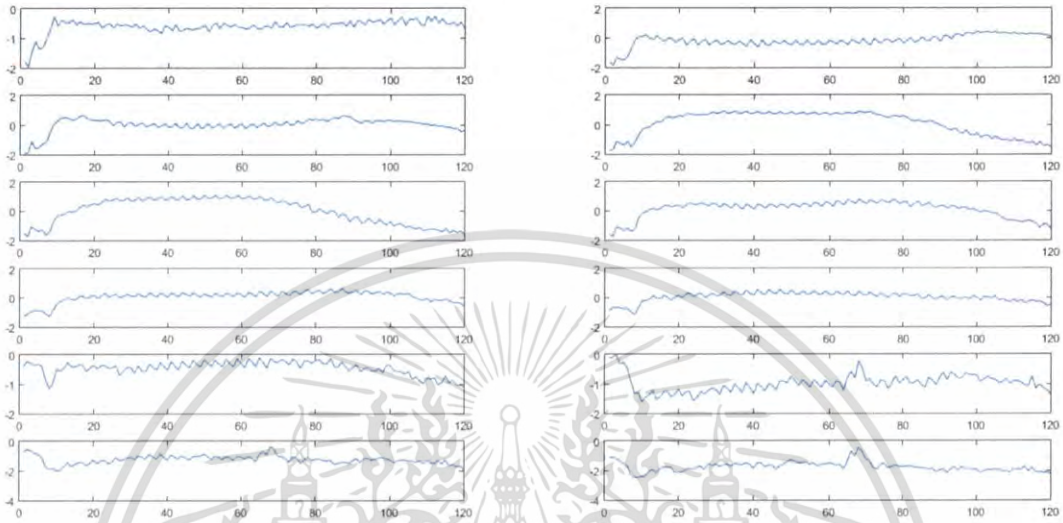


รูปที่ 4.25 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าที่ 1 ถึง 12 ของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ

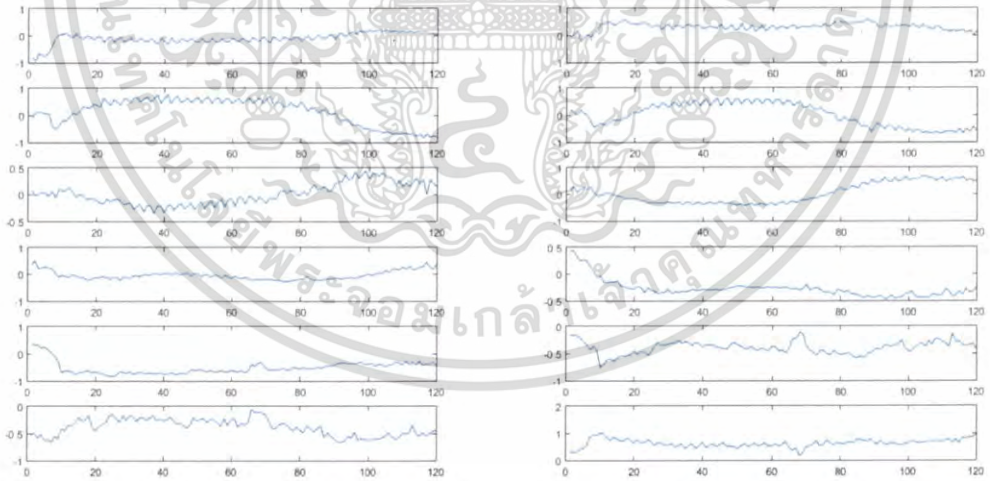
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.13 Resizing

จากนั้นจะนำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้ามาลดขนาดข้อมูลลงให้เหลือเพียง 120 แซมเปิล แสดงได้ดังรูปที่ 4.26 รูปที่ 4.27 และ รูปที่ 4.28 ตามลำดับ

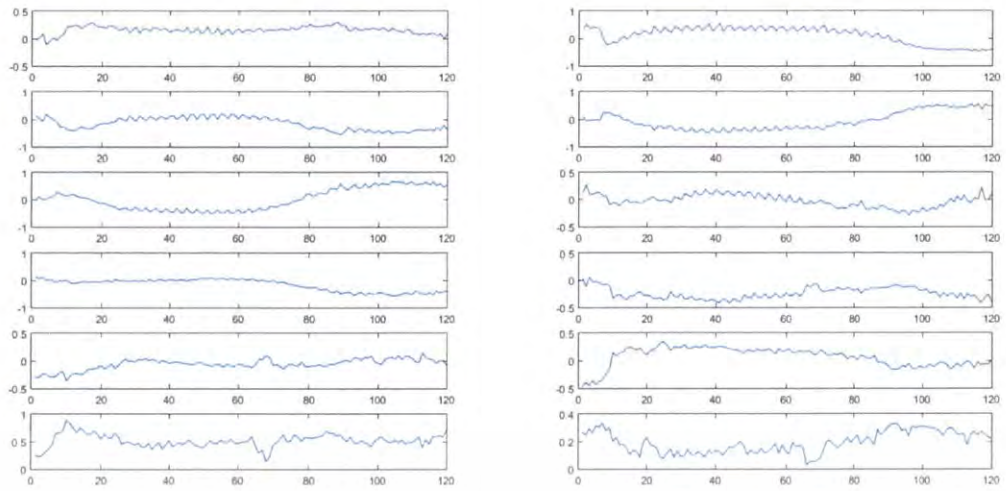


รูปที่ 4.26 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 แซมเปิลของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ



รูปที่ 4.27 แสดงค่าสัมประสิทธิ์เดลต้าที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 แซมเปิลของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ” โดยเรียงกราฟจากซ้ายไปขวาตามลำดับ

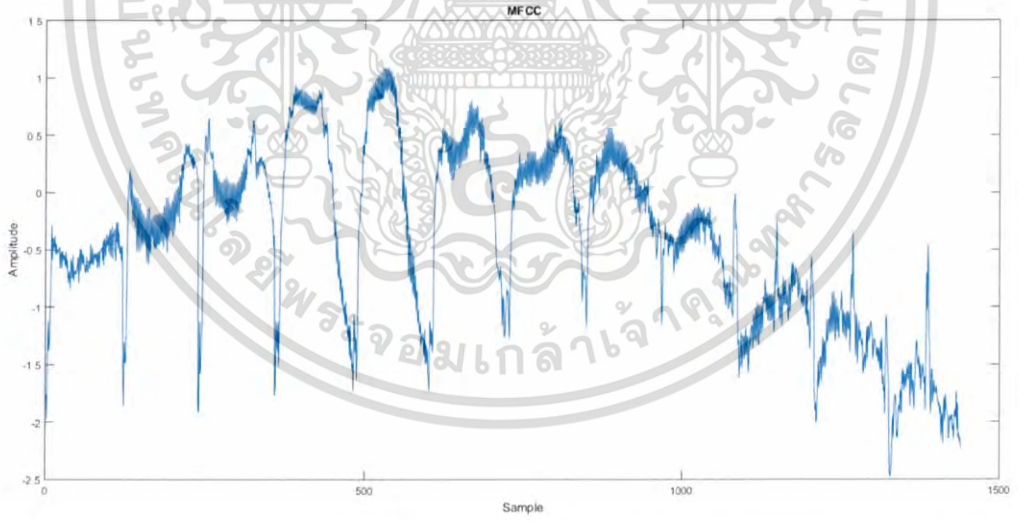
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.28 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าที่ 1 ถึง 12 ที่ถูกลดขนาดข้อมูลลงเหลือ 120 แซมเปิลของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

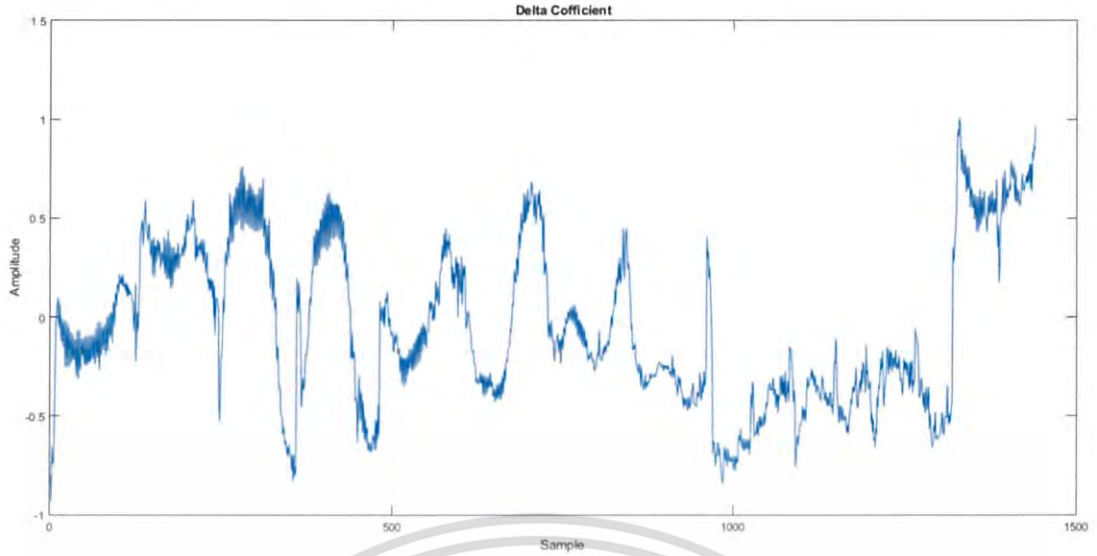
4.1.14 Feature Arrangement

จากนั้นจะนำผลลัพธ์ที่ได้ในขั้นตอน Resizing มาทำการเรียงสัมประสิทธิ์ตั้งแต่ลำดับที่ 1 ถึง 12 ทั้ง ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ดับเบิล-เดลต้า โดยแสดงผลลัพธ์ได้ดังรูปที่ 4.29 รูปที่ 4.30 และ รูปที่ 4.31 ตามลำดับ



รูปที่ 4.29 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกันตามลำดับของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



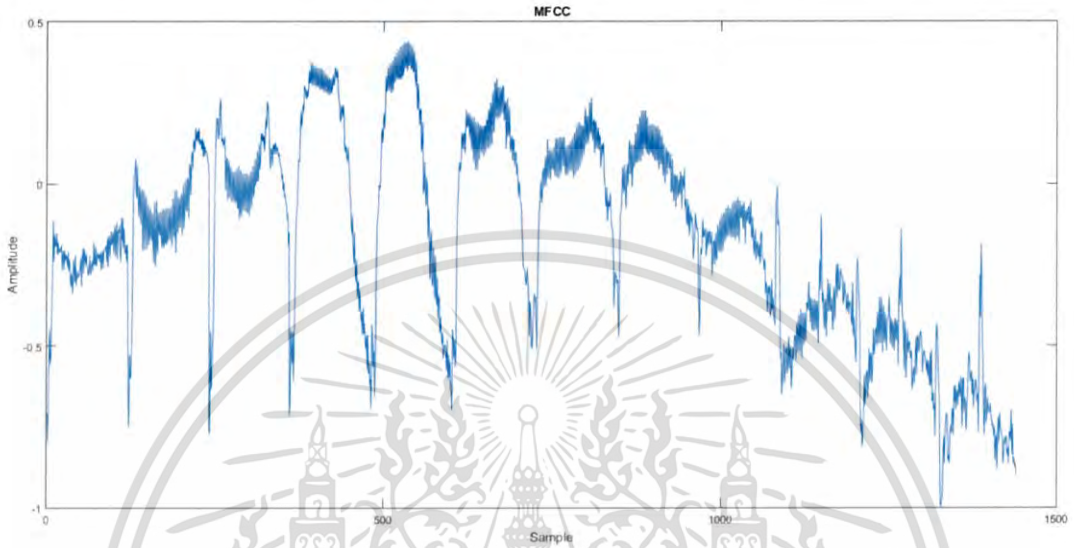
รูปที่ 4.30 แสดงค่าสัมประสิทธิ์เดลต้าที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกันตามลำดับของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

รูปที่ 4.31 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าที่ 1 ถึง 12 ที่ถูกนำมาเรียงต่อกันตามลำดับของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

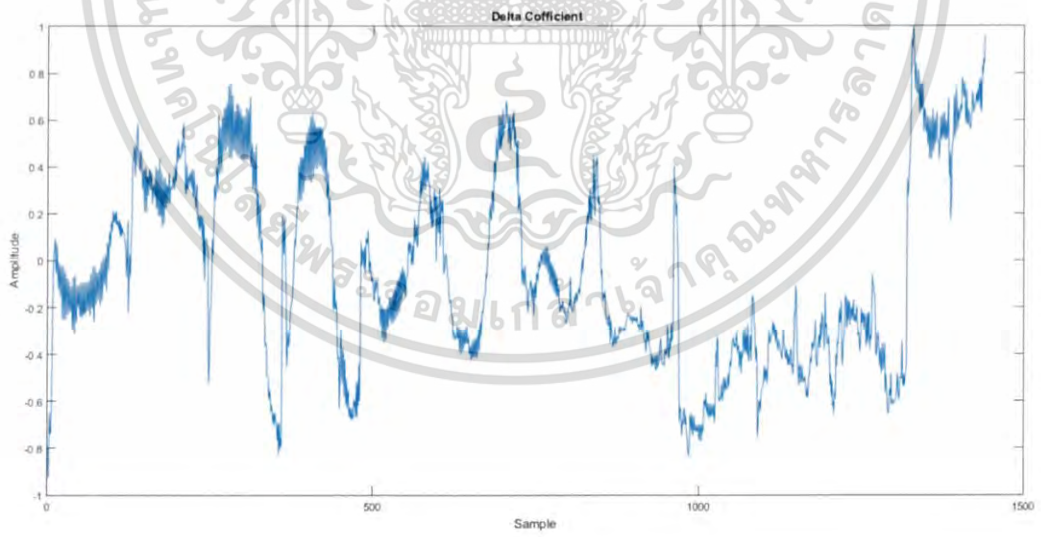
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.15 กระบวนการทำการนอมอไลซ์ทางขนาด

เมื่อผ่านกระบวนการ Feature Arrangement แล้วจะนำค่าสัมประสิทธิ์แต่ละชนิดมาทำการนอมอไลซ์ทางขนาดซึ่งจะคล้ายกับในกระบวนการเตรียมสัญญาณเสียงเบื้องต้น โดยแสดงผลลัพธ์ที่ทำการนอมอไลซ์ทางขนาดแล้วได้ดังรูปที่ 4.32 รูปที่ 4.33 และรูปที่ 4.34

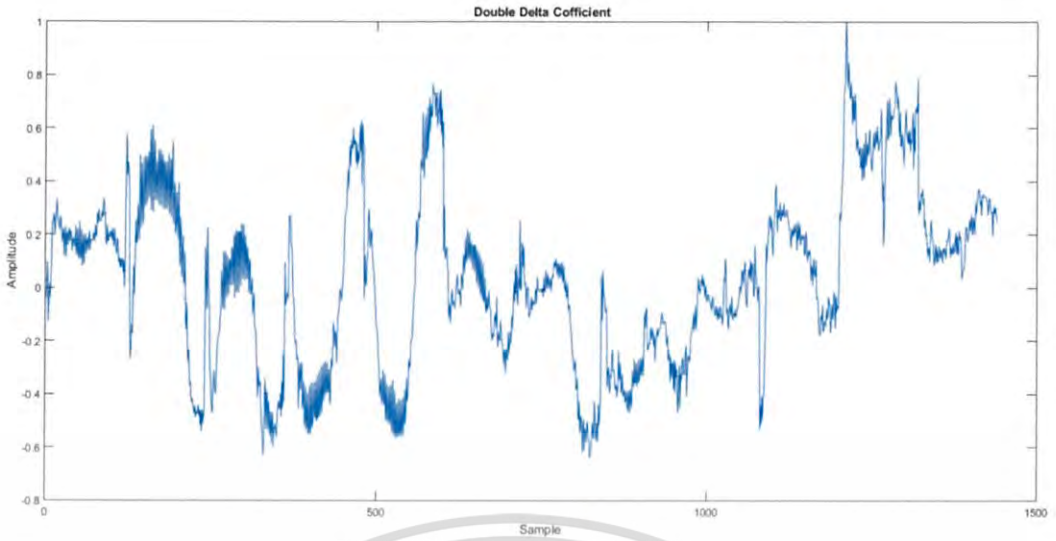


รูปที่ 4.32 แสดงค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมลที่ถูกนำมาทำการนอมอไลซ์ทางขนาดของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”



รูปที่ 4.33 แสดงค่าสัมประสิทธิ์เดลต้าที่ถูกนำมาทำการนอมอไลซ์ทางขนาดของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

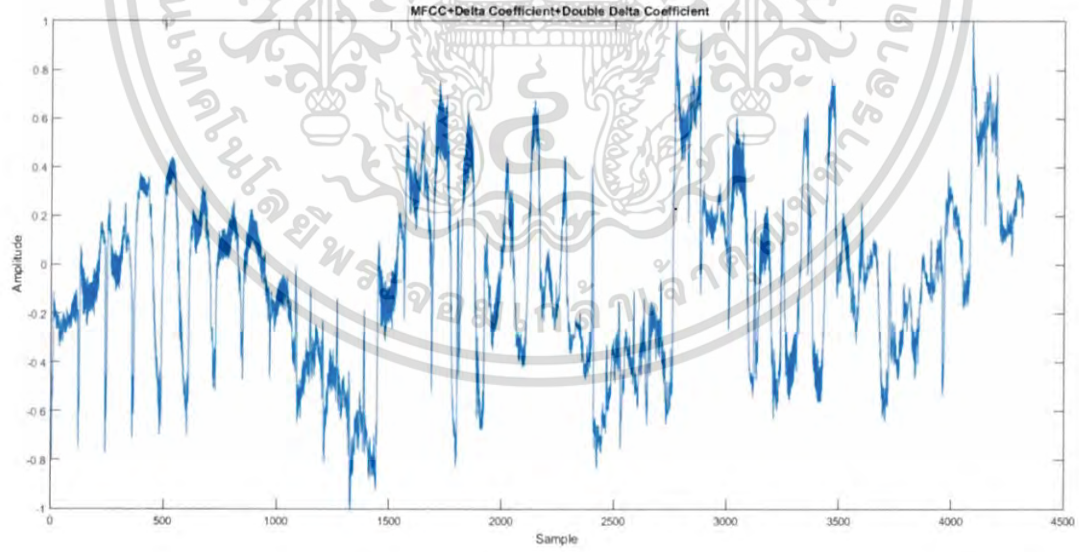
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.34 แสดงค่าสัมประสิทธิ์ดับเบิล-เดลต้าที่ถูกนำมาทำการนอมอไลซ์ทางขนาดของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

4.1.16 Completed Feature

สุดท้ายแล้วจะนำค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ค่าสัมประสิทธิ์เดลต้า และค่าสัมประสิทธิ์ดับเบิล-เดลต้ามาเรียงต่อกันเพื่อที่จะนำไปใช้เรียนรู้ในระบบโครงข่ายประสาทเทียมต่อไป ซึ่งสามารถแสดงได้ดังรูปที่ 4.35



รูปที่ 4.35 แสดงค่าสัมประสิทธิ์ที่จะถูกนำไปเรียนรู้เพื่อการรู้จำในระบบโครงข่ายประสาทเทียมของตัวอย่างสัญญาณเสียงพยางค์ที่ 3 คำว่า “ใจ”

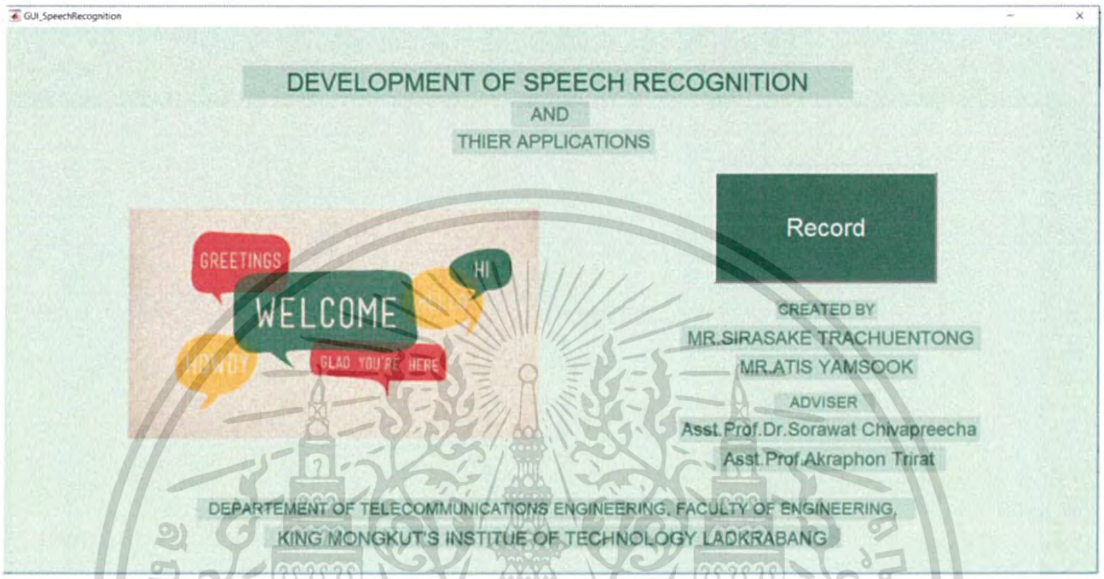
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.17 ระบบการรู้จำ

ในส่วนนี้ได้ออกแบบอัลกอริทึม ซึ่งดูเพิ่มเติมได้ที่ภาคผนวก ก

4.1.18 ออกแบบอินเตอร์เฟสของระบบรู้จำเสียงโดยใช้ MATLAB GUI

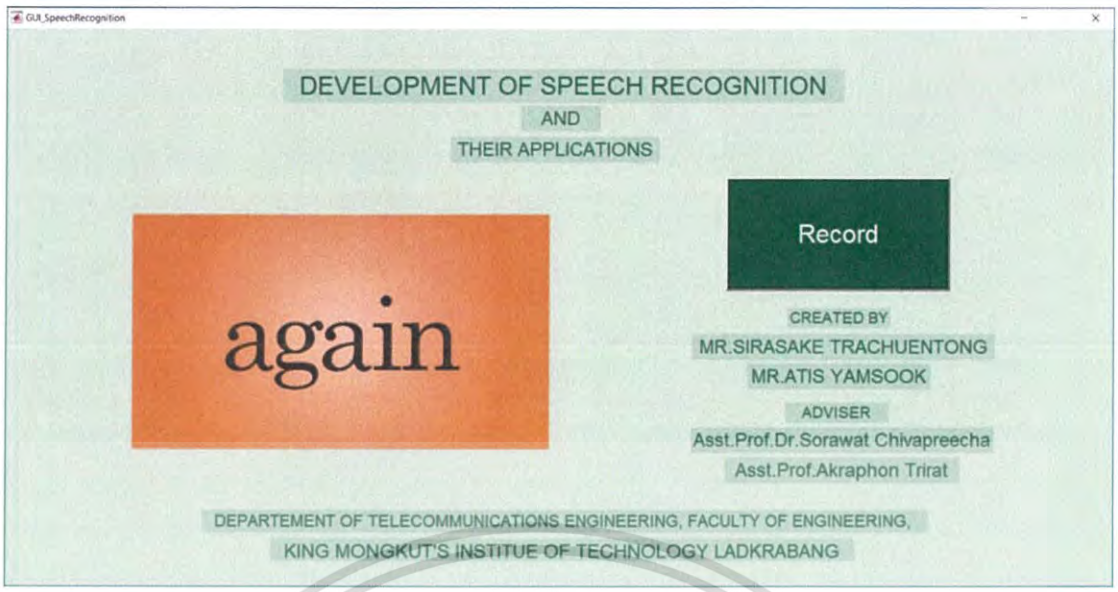
จะได้หน้าต่างผลลัพธ์ดังรูปที่ 4.36



รูปที่ 4.36 หน้าต่างโปรแกรมหลังจากรันโปรแกรม

รูปที่ 4.37 หน้าต่างโปรแกรมหลังจากรันโปรแกรมโดยสามารถตรวจสอบได้ว่าเป็นคำว่า “จำฝังใจ” โดยจะทำการเรียกไฟล์เพลง “จำฝังใจ” ผ่านโปรแกรม VLC Media Player

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.38 หน้าต่างโปรแกรมหลังจากรันโปรแกรมที่ไม่สามารถตรวจสอบได้ว่าเป็นคำอะไร หรือปัญหาที่เกิดจากการตัดพยางค์โดยให้ทำการพูดอีกครั้ง

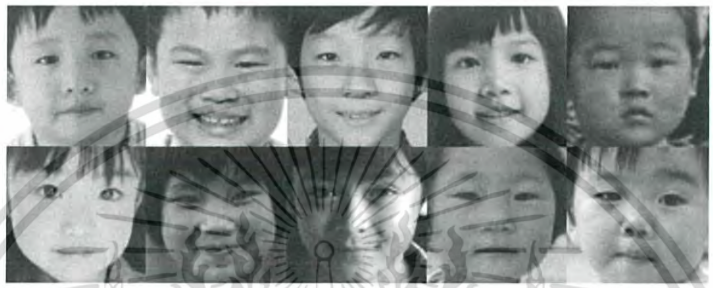
4.2 ระบบโครงข่ายประสาทเทียมสำหรับรู้จำภาพใบหน้าของกลุ่มบุคคล

การทดลองนี้ ได้ใช้ฐานข้อมูลเป็นภาพใบหน้าของกลุ่มบุคคลทั้งหมด 120 ภาพ โดยจะจำแนกออกเป็น 4 กลุ่ม คือ ผู้ชาย ผู้หญิง เด็กและผู้สูงอายุตามลำดับ แบ่งเป็นกลุ่มละ 30 ภาพเท่าๆกัน ซึ่งเราจะแบ่งข้อมูลที่ใช้เป็น 3 ส่วน ส่วนแรกคือข้อมูลที่เป็นฐานข้อมูลสำหรับไอเกนเฟส ส่วนที่ 2 คือข้อมูลส่วนที่จะนำมาเทียบค่าระยะทางยุคลิด และส่วนที่ 3 คือข้อมูลในส่วนที่นำมาทดสอบระบบในโครงข่ายประสาทเทียมเพื่อทำการหาค่าประสิทธิภาพของระบบ โดยขนาดของรูปที่ใช้ในการทดลองนั้นมีขนาดเท่ากับ 100x100 พิกเซล ซึ่งสามารถแสดงตัวอย่างของภาพฐานข้อมูลได้ดังรูปที่ 4.39 ถึง 4.42

รูปที่ 4.39 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลผู้ชาย (10 ภาพจาก 30 ภาพ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.40 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลผู้หญิง (10 ภาพจาก 30 ภาพ)



รูปที่ 4.41 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลเด็ก (10 ภาพจาก 30 ภาพ)

รูปที่ 4.42 ภาพตัวอย่างฐานข้อมูลของกลุ่มบุคคลผู้สูงอายุ (10 ภาพจาก 30 ภาพ)

4.2.1 กระบวนการเตรียมข้อมูล

ในขั้นตอนนี้ทำการแปลงรูปทั้งหมด 120 ภาพที่มีขนาด 100x100 พิกเซล ให้เป็นเวกเตอร์ จะได้เอาพุตของข้อมูลที่มีขนาดเท่ากับ 10,000 x120

4.2.2 กระบวนการหาค่าเฉลี่ยของข้อมูล

ค่าเฉลี่ยของรูปทั้ง 120 รูปสามารถแสดงได้ดังรูปที่ 4.43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.43 ค่าเฉลี่ยของภาพทั้ง 120 ภาพ

4.2.3 กระบวนการกำจัดค่าเฉลี่ย

สามารถแสดงเป็นรูปได้ดังภาพที่ 4.44

รูปที่ 4.44 ภาพในขั้นตอนการกำจัดค่าเฉลี่ย (ตัวอย่างของกลุ่มบุคคลผู้ชาย)

4.2.4 กระบวนการการหา Covariance Matrix

ในขั้นตอนนี้ ค่า Covariance Matrix ที่ได้จะมีขนาดเท่ากับ 10,000 x 10,000 ซึ่งมีขนาดใหญ่มาก ทำให้ผู้จัดทำได้แปลงสมการในสมการที่ 3.24

4.2.5 กระบวนการการหาค่าไอเกนเว็กเตอร์และไอเกนแวลู

ในขั้นตอนนี้จะได้ค่าเมตริกซ์ไอเกนที่มีขนาดเท่ากับ 120x120

4.2.6 กระบวนการหาค่าไอเกนเฟส

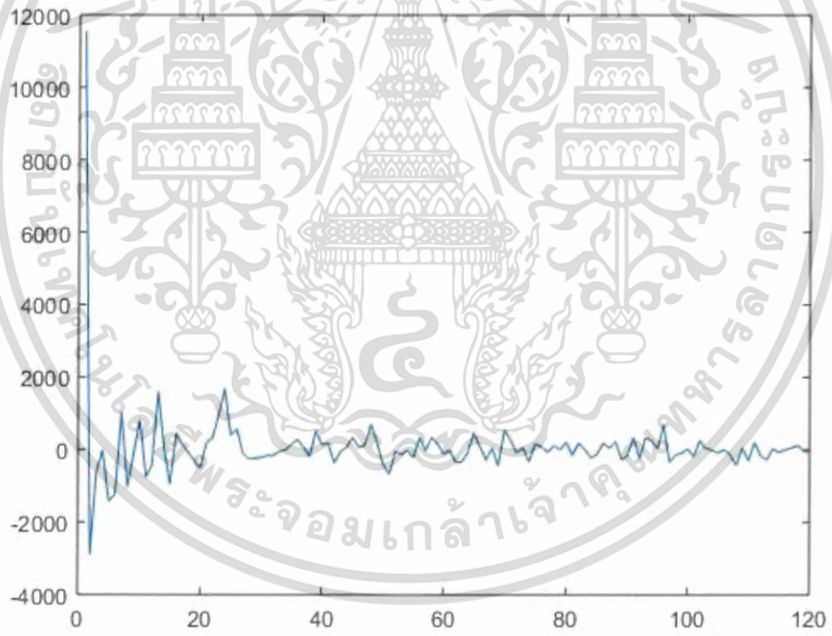
สามารถแสดงผลลัพธ์ได้ตามรูปที่ 4.45 โดยจะแสดงตัวอย่าง 10 รูปแรกของไอเกนเฟส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.45 รูปไอเกนเฟส

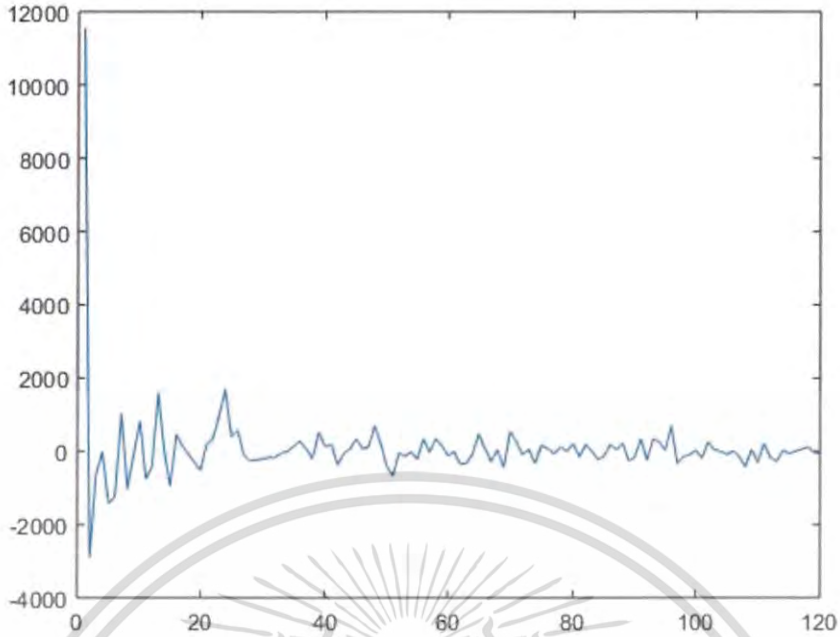
4.2.7 กระบวนการการหาค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูล

ในขั้นตอนนี้ค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลนั้นจะมีขนาดเท่ากับ 120x120 ซึ่งสามารถแสดงได้ดังรูปที่ 4.46 ถึง 4.49

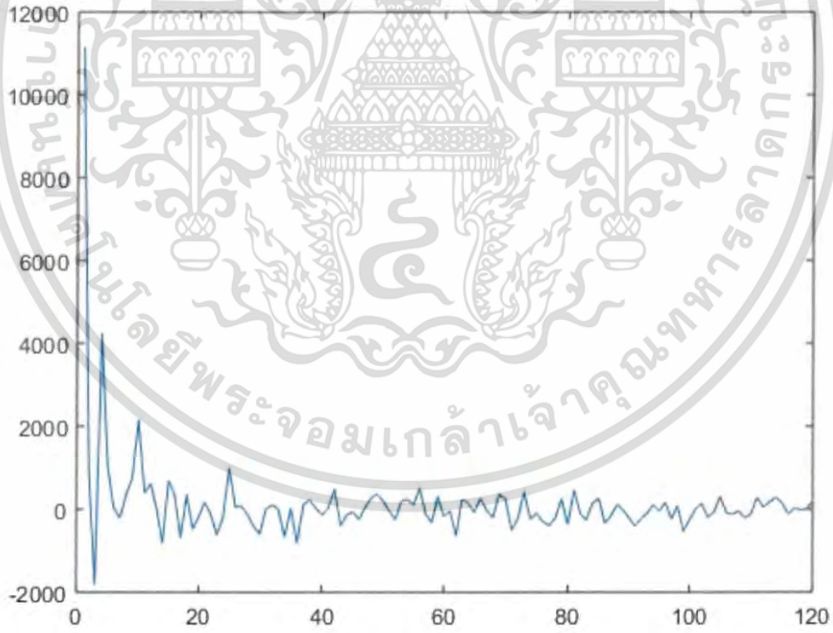


รูปที่ 4.46 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลผู้ชาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

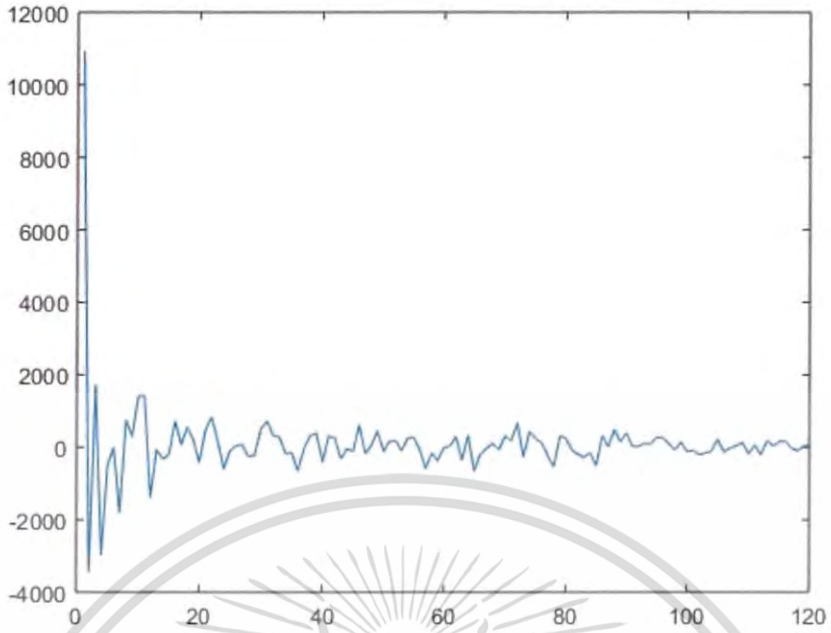


รูปที่ 4.47 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลผู้หญิง



รูปที่ 4.48 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลเด็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.49 ตัวอย่างค่าสัมประสิทธิ์ใบหน้าของฐานข้อมูลของกลุ่มบุคคลผู้สูงอายุ

4.2.8 กระบวนการจำแนกโดยใช้การหาระยะทางแบบยุคลิด

ในกระบวนการนี้ ได้นำรูปภาพของกลุ่มบุคคลทั้ง 4 แบบ แบบละ 30 ภาพมาทดสอบเช่นกัน โดยรูปที่นำมาใช้นั้นไม่ได้อยู่ในฐานข้อมูล โดยจะมีทั้งหมด 120 รูป ซึ่งในที่นี้จะขอนำเสนอตัวอย่างของกลุ่มบุคคลกลุ่มละ 2 ภาพ ดังรูปที่ 4.50

รูปที่ 4.50 ตัวอย่างของรูปของกลุ่มบุคคลที่นำมาใช้ทดสอบ โดยเรียงเป็นผู้ชาย 2 คน ผู้หญิง 2 คน เด็ก 2 คนและผู้สูงอายุ 2 คน

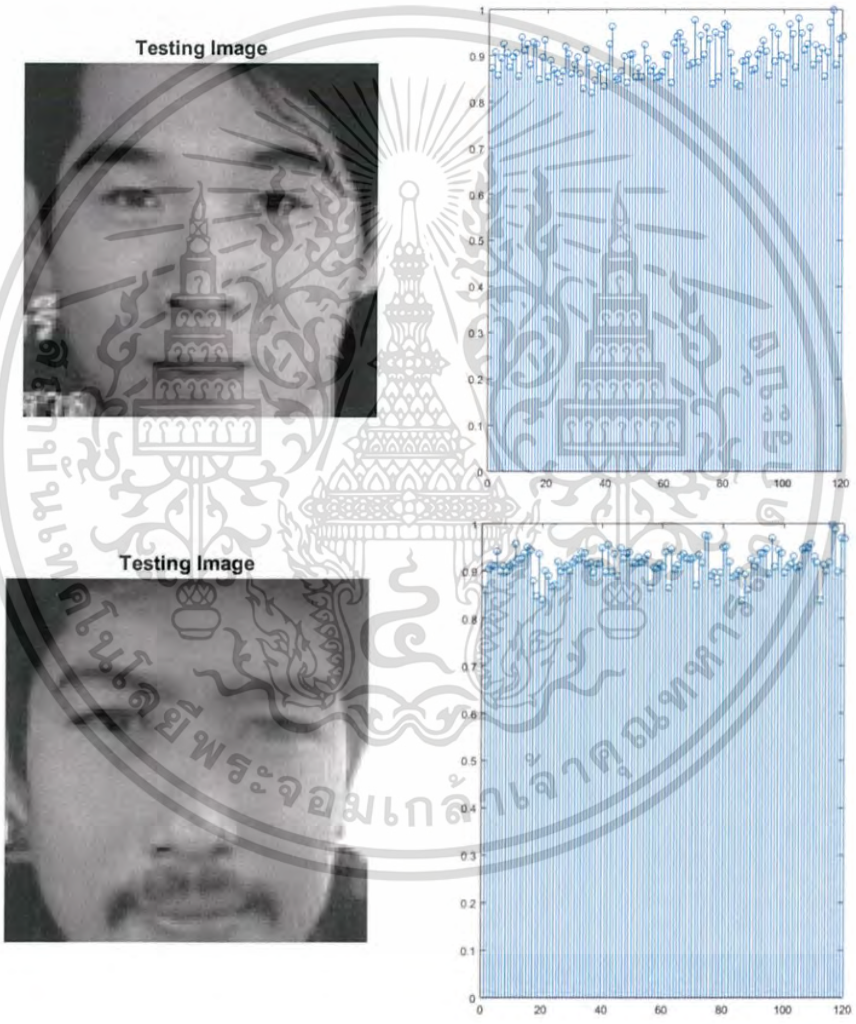
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.9 กระบวนการหาค่าสัมประสิทธิ์ใบหน้าของรูปที่จะนำมาจำแนก

จะมีขนาดของข้อมูลเท่ากับ 120x1

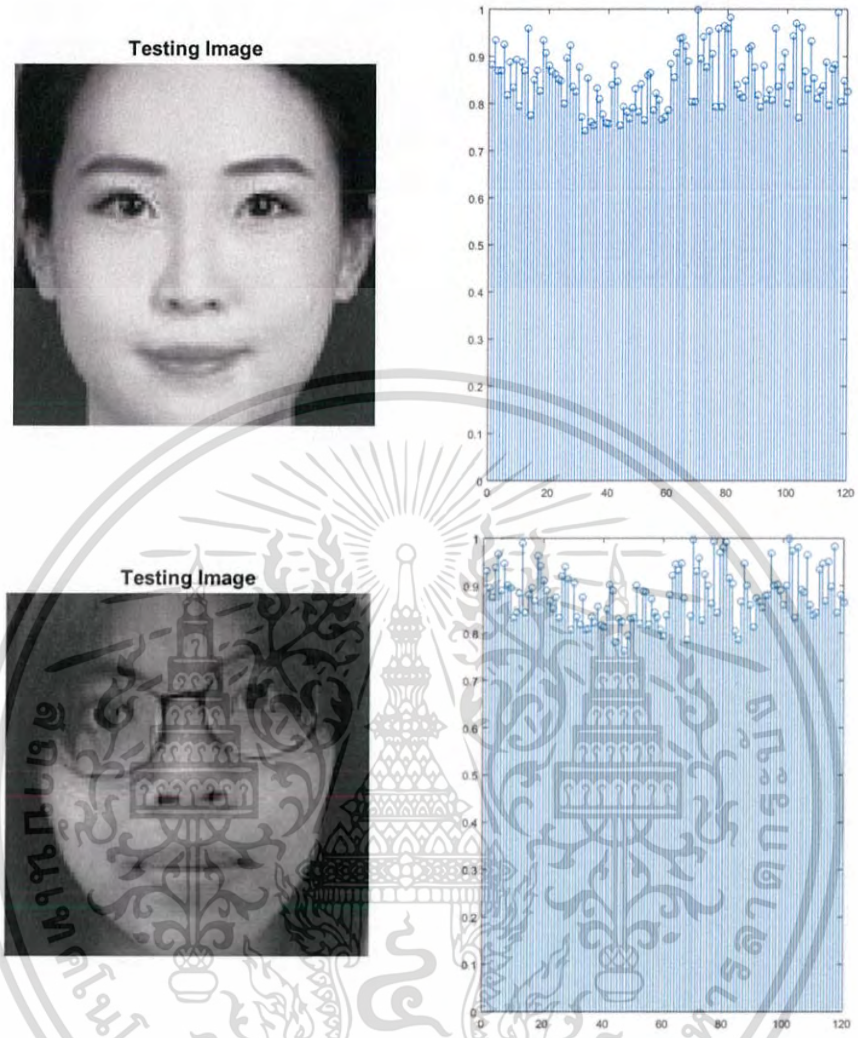
4.2.10 กระบวนการหาระยะทางยุคลิด

ในขั้นตอนนี้จะเป็นการหาระยะทางยุคลิด โดยจะแสดงผลการทดสอบได้ดังรูปที่ 4.51 ถึง 4.54 (จากกราฟระยะทางยุคลิดนั้นในแกนนอน ตั้งแต่ 1 ถึง 30 คือผลลัพธ์ของค่ายุคลิดของกลุ่มบุคคลผู้ชายที่ถูกนอมอลไลซ์แล้ว, 31 ถึง 60 คือผลลัพธ์ของค่ายุคลิดของกลุ่มบุคคลผู้หญิงที่ถูกนอมอลไลซ์แล้ว, 61 ถึง 90 คือผลลัพธ์ของค่ายุคลิดของกลุ่มบุคคลเด็กที่ถูกนอมอลไลซ์แล้ว, 91 ถึง 120 คือผลลัพธ์ของค่ายุคลิดของกลุ่มบุคคลผู้สูงอายุที่ถูกนอมอลไลซ์แล้ว)



รูปที่ 4.51 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลเพศชาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

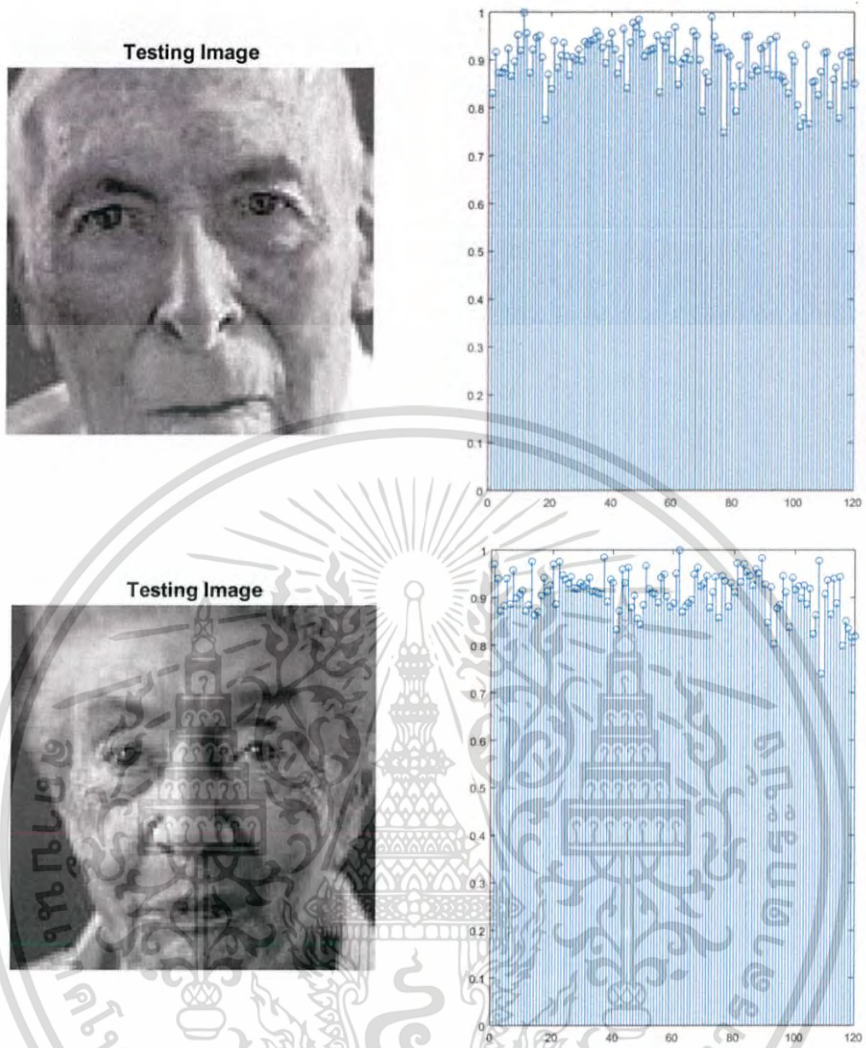


รูปที่ 4.52 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลผู้หญิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่4.53 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลเด็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

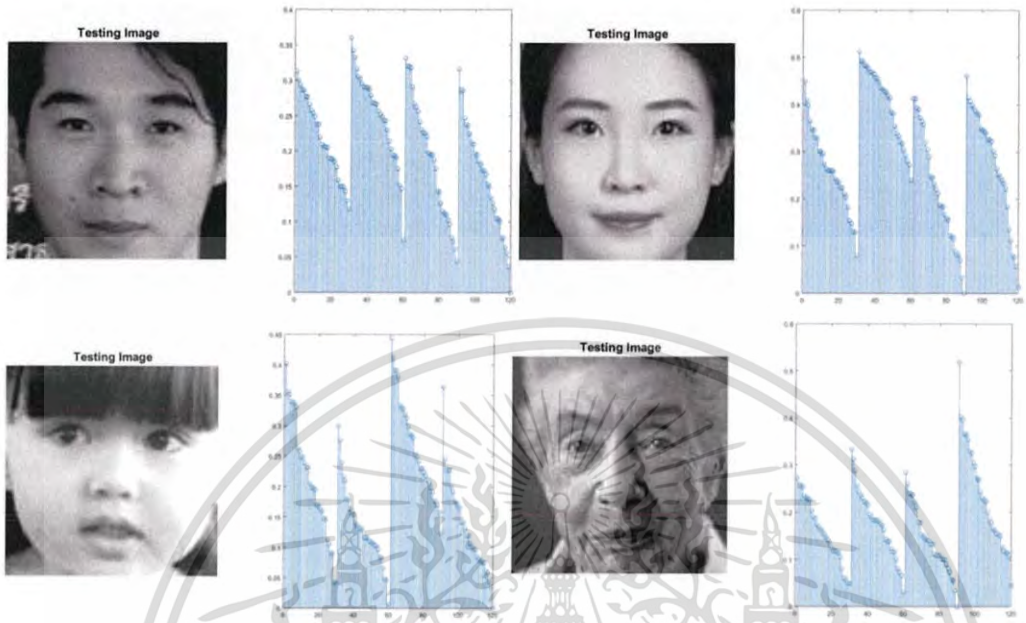


รูปที่ 4.54 รูปตัวอย่างค่าระยะทางยุคลิดของภาพตัวอย่างของกลุ่มบุคคลผู้สูงอายุ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.11 Feature Arrangement & Resizing (Amplitude)

ทำการปรับข้อมูลก่อนนำไปเข้าโครงข่ายประสาทเทียม ซึ่งผลลัพธ์จะแสดงได้ดังรูปที่ 4.55

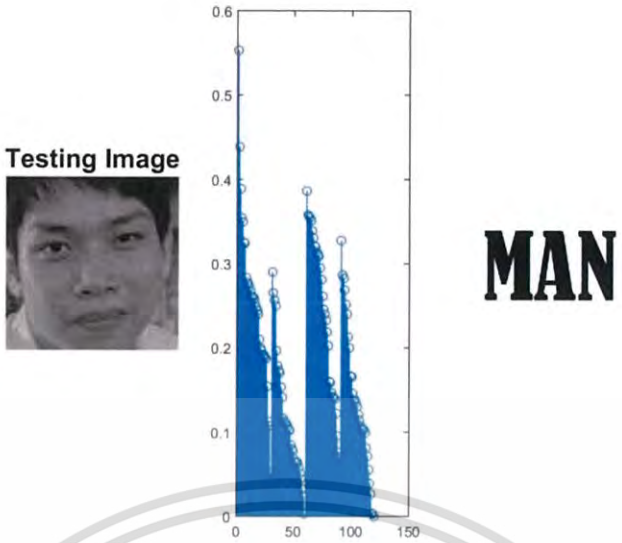


รูปที่ 4.55 ตัวอย่างผลลัพธ์ของข้อมูลที่เราจะนำไปเข้าระบบโครงข่ายประสาทเทียม

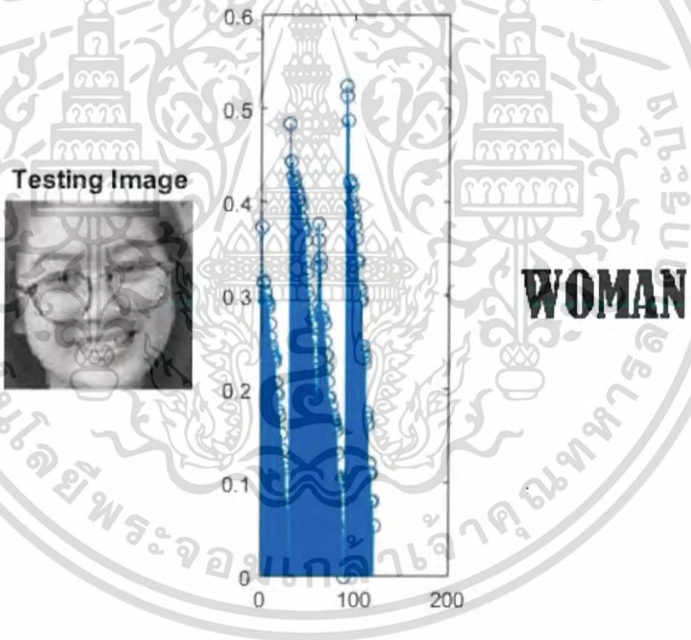
4.2.12 ตัวอย่างผลลัพธ์หลังจากผ่านระบบโครงข่ายประสาทเทียม

สามารถแสดงตัวอย่างได้ดังรูปที่ 4.56 ถึง 4.59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

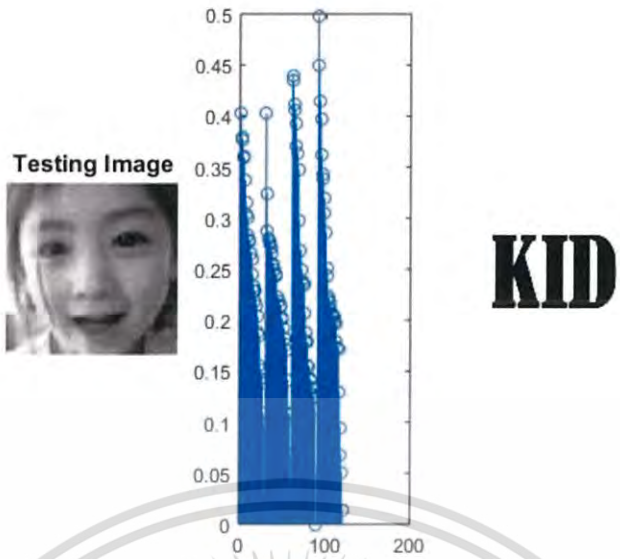


รูปที่ 4.56 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลผู้ชาย

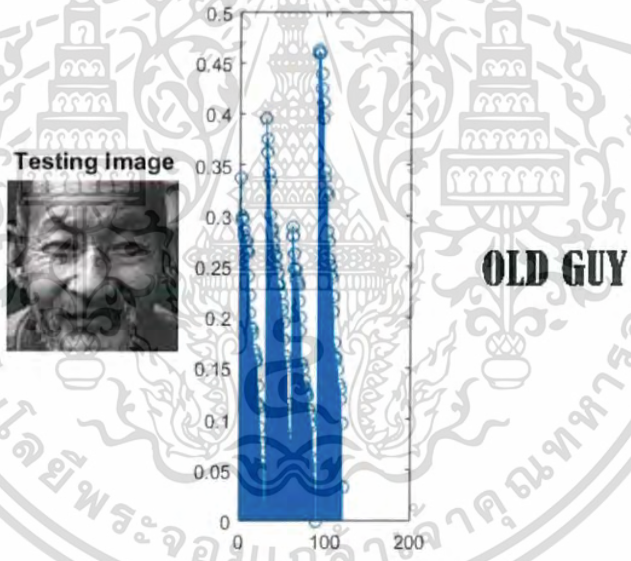


รูปที่ 4.57 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลผู้หญิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.58 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลเด็ก



รูปที่ 4.59 ตัวอย่างผลลัพธ์ของการทดสอบบุคคลผู้สูงอายุ

4.3 ผลการทดสอบและประสิทธิภาพของระบบรู้จำ

4.3.1 ระบบโครงข่ายประสาทเทียมของการรู้จำชื่อเพลง

ในที่นี้จะทำการทดสอบโดยใช้เสียงผู้ชายและผู้หญิง 2 คน โดยพูดคนละ 2 ครั้งต่อ 1 เพลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 1 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
เหงา	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
คอย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
เงา	✗	เหงา	✗	เหงา	✗	เหงา	✗	เหงา	0	✗	เหงา	✗	เหงา	✗	เหงา	✗	เหงา	0
ก่อน	✗	รอ	✗	รอ	✗	รอ	✗	รอ	0	✓	-	✗	รอ	✗	รอ	✗	รอ	25
ดาว	✗	เหงา	✗	เหงา	✓	-	✗	เหงา	25	✗	เหงา	✗	เหงา	✓	-	✗	เหงา	25
รอ	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
เคย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
ดื่ม	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
แช่ง	✓	-	✓	-	✓	-	✓	-	100	✓	-	✗	เคย	✓	-	✗	เคย	50
อาย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
% ความถูกต้องรวม	72.5									65								

ตารางที่ 4.2 ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 2 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
ใจรัก	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
สองรัก	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
คิดถึง	✗	อีกนาน	✓	-	✗	อีกนาน	✓	-	50	✓	-	✓	-	✓	-	✓	-	100
มีเธอ	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
ร่ำไร	✗	สองรัก	✓	-	✗	สองรัก	✓	-	50	✗	วายร้าย	✗	วายร้าย	✓	-	✗	วายร้าย	25
อีกนาน	✓	-	✓	-	✓	-	✗	มีเธอ	75	✓	-	✓	-	✓	-	✓	-	100
ยังจำ	✗	ร่ำไร	✗	ใจรัก	✗	ว้ายร้าย	✗	วายร้าย	0	✗	ร่ำไร	✗	ใจรัก	✗	วายร้าย	✗	วายร้าย	0
เจ้าตาก	✓	-	✓	-	✗	ใจรัก	✗	สองรัก	50	✓	-	✓	-	✓	-	✗	ใรจัก	75
แอบเจ็บ	✗	สองรัก	✗	สองรัก	✗	สองรัก	✗	สองรัก	0	✗	ใจรัก	✗	ใจรัก	✗	ใจรัก	✗	คิดถึง	0
วายร้าย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
% ความถูกต้องรวม	62.5									70								

ตารางที่ 4.3 ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 3 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
จำฝังใจ	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
ใจปอนปอน	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
เสี่ยวรำพึง	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✗	ใจนักเลง	✓	-	75
ตัวสำรอง	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✗	แสงสุดท้าย	✓	-	75
เธอน่ารัก	✗	จำฝังใจ	✗	ปัญหาจากการตัดพยางค์	✗	ปัญหาจากการตัดพยางค์	✗	แสงสุดท้าย	0	✓	-	✓	-	✓	-	✓	-	100
ไม่ต่างกัน	✓	-	✓	-	✗	ระหว่างเรา	✗	จำฝังใจ	50	✓	-	✗	จำฝังใจ	✗	แสงสุดท้าย	✗	จำฝังใจ	25
ใจนักเลง	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
แสงสุดท้าย	✓	-	✗	จำฝังใจ	✓	-	✗	ตัวสำรอง	50	✓	-	✓	-	✗	จำฝังใจ	✓	-	75

ตารางที่ 4.3 (ต่อ) ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 3 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
ระหว่างเรา	✓	-	✓	-	✗	ตัวสำรอง	✓	-	75	✓	-	✓	-	✓	-	✓	-	100
ฉันยังอยู่	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
% ความถูกต้องรวม	77.5									85								

ตารางที่ 4.4 ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 4 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
เธอคือความฝัน	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✗	บอกฉันบอกเธอ	75
ได้อย่างเสียอย่าง	✓	-	✗	ปัญหาจากการตัดพยางค์	✗	ปัญหาจากการตัดพยางค์	✓	-	50	✓	-	✓	-	✓	-	✓	-	100
เพราะเธอหรือเปล่า	✗	บอกฉันบอกเธอ	✓	-	✓	-	✗	เห็นใจกันหน่อย	50	✗	ปัญหาจากการตัดคำ	✓	-	✗	บอกฉันบอกเธอ	✓		50
เห็นใจกันหน่อย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100

ตารางที่ 4.4 (ต่อ) ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำ 4 พยางค์

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
หัวใจสะออน	✓	-	✓	-	✓	-	✗	ปัญหาจากการตัดพยางค์	75	✓	-	✓	-	✓	-	✓	-	100
คนนี้ของฉัน	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✗	บอกฉันบอกเธอ	75
ไม่ตลกเลย	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
หัวใจมักง่าย	✓	-	✓	-	✓	-	✓	-	50	✓	-	✓	-	✓	-	✓	-	100
บอกฉันบอกเธอ	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
% ความถูกต้องรวม	80									90								

ตารางที่ 4.5 ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำตั้งแต่ 5 พยางค์ขึ้นไป

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
ด้วยรักและผูกพัน	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
รักคงยังไม่พอ	✓	-	✓		✓		✓	-	100	✓	-	✓		✓	-	✓	-	100
อาจจะเป็นคนนี้	✗	ด้วยรักและผูกพัน	✓	-	✗	ด้วยรักและผูกพัน	✓	-	50	✗	ปัญหาจากการตัดพยางค์	✗	ปัญหาจากการตัดพยางค์	✓	-	✓	-	50
ขอบใจที่พูดแรง	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
รักเธอให้หน่อยลง	✓	-	✓	-	✓	-	✗	ปัญหาจากการตัดพยางค์	75	✓	-	✓	-	✗	ปัญหาจากการตัดพยางค์	✗	ปัญหาจากการตัดพยางค์	50

ตารางที่ 4.5 (ต่อ) ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำตั้งแต่ 5 พยางค์ขึ้นไป

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
เกินใจจะอดทน	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✗	ปัญหาจากการตัดพยางค์	✓	-	75
ถ้าฉันเปลี่ยนใจเธอได้	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✓	-	✓	-	100
ครั้งหนึ่งเราเคยรักกัน	✗	ถ้าฉันเปลี่ยนใจเธอได้	✓	-	✓	-	✗	ปัญหาที่เกิดจากการตัดคำ	50	✓	-	✓	-	✓	-	✓	-	50

ตารางที่ 4.5 (ต่อ) ผลการทดลองและทดสอบประสิทธิภาพของกลุ่มคำตั้งแต่ 5 พยางค์ขึ้นไป

	ผู้ชายคนที่ 1				ผู้ชายคนที่ 2					ผู้หญิงคนที่ 1				ผู้หญิงคนที่ 2				
เพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	ครั้งที่ 1	คำที่ได้	ครั้งที่ 2	คำที่ได้	% ความถูกต้องรายเพลง
เมื่อเขามาฉันจะไป	✓	-	✓	-	✓	-	✗	ปัญหาที่เกิดจากการตัดคำ	75	✓	-	✓	-	✓	-	✓	-	75
รักที่เพิ่งผ่านพ้นไป	✓	-	✓	-	✓	-	✓	-	100	✓	-	✓	-	✗	ถ้าฉันเปลี่ยนใจเธอได้	✗	ถ้าฉันเปลี่ยนใจเธอได้	50
% ความถูกต้องรวม	80.21									83.34								

ตารางที่ 4.6 สรุปผลการทดลองและทดสอบประสิทธิภาพของระบบรู้จำเสียง

% ความถูกต้อง / กลุ่มพยางค์	% ความถูกต้องของเสียงผู้ชาย	% ความถูกต้องของเสียงผู้หญิง	% ความถูกต้องตามกลุ่มพยางค์
กลุ่มคำ 1 พยางค์	72.5	65	68.75
กลุ่มคำ 2 พยางค์	62.5	70	66.25
กลุ่มคำ 3 พยางค์	77.5	85	81.25
กลุ่มคำ 4 พยางค์	80	90	85
กลุ่มคำมากกว่า 5 พยางค์	80.21	84.34	81.78
% ความถูกต้องรวมของระบบรู้จำเสียง	75.49	79.45	77.98

4.3.2 ระบบโครงข่ายประสาทเทียมของการรู้จำภาพใบหน้าของกลุ่มบุคคล

ในที่นี้จะใช้ภาพใบหน้าจำนวน 120 ภาพ ที่นำมาจากที่ต่างๆในอินเทอร์เน็ตซึ่งแบ่งเป็นกลุ่มละ 30 ภาพในการทดสอบเช่นกัน โดยจะเป็นภาพที่ไม่ได้อยู่ในฐานข้อมูลและไม่ได้เป็นภาพที่ใช้ในการหาค่าระยะทางยุคลิด

ตารางที่ 4.7 สรุปผลการทดลองและทดสอบประสิทธิภาพของระบบรู้จำภาพใบหน้าของบุคคล

ภาพที่นำมาทดสอบ	กลุ่มผู้ชาย	กลุ่มผู้หญิง	กลุ่มเด็ก	กลุ่มผู้สูงอายุ	เปอร์เซ็นต์ความถูกต้อง
ผู้ชาย 30 ภาพ	**25**	3	1	1	83.33%
ผู้หญิง 30 ภาพ	1	**28**	0	1	93.33%
เด็ก 30 ภาพ	10	1	**18**	1	60.00%
ผู้สูงอายุ 30 ภาพ	6	4	1	**19**	63.33%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการทดลอง

ในปริญญานิพนธ์นี้นำเสนอกระบวนการของระบบการรู้จำชื่อเพลงและภาพใบหน้ากลุ่มบุคคล โดยใช้โครงข่ายประสาทเทียมแบบแพร่กลับ ในส่วนของระบบรู้จำชื่อเพลงนั้นใช้วิธีการหาค่าสัมประสิทธ์เซปสตรัมบนสเกลเมล ซึ่งเมื่อผู้ทดลองพูดชื่อเพลงที่อยู่ในตาราง ผลลัพธ์จะปรากฏในหน้าต่างโปรแกรมจะแสดงผลเช่นเดียวกับคำที่พูดไป และในกรณีที่พูดผิด หน้าต่างโปรแกรมจะแสดงว่าให้พูดใหม่อีกครั้ง ซึ่งผลการทดลองปรากฏว่าประสิทธิภาพของระบบรู้จำเสียงอยู่ที่ 77.98% และในส่วนของระบบรู้จำภาพใบหน้ากลุ่มบุคคลนั้นใช้วิธีไอเกนเฟส เมื่อระบบได้ภาพใบหน้าของบุคคลที่ทดสอบแล้ว ระบบจะประมวลผลและแสดงโฆษณาที่เหมาะสมกับบุคคลนั้นๆขึ้นมา ซึ่งผลการทดลองปรากฏว่าประสิทธิภาพของระบบรู้จำภาพอยู่ที่ 75.00%

5.2 ข้อเสนอแนะ

เนื่องจากระบบโครงข่ายประสาทเทียมนั้นอัตราความถูกต้องจะขึ้นอยู่กับจำนวนของฐานข้อมูลและความคล้ายคลึงกันของข้อมูลชนิดเดียวกัน จึงจำเป็นต้องเตรียมข้อมูลที่จะนำไปเข้าโครงข่ายประสาทเทียมให้ดี โดยจะต้องเลือกใช้ข้อมูลที่มีมาตรฐาน เนื่องจากระบบค่อนข้างมีความอ่อนไหวสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] Artificial neural network." Wikipedia. https://en.wikipedia.org/wiki/Artificial_neural_network

[2] J. M. Kim and M. Kang. "A Study of Face Recognition using the PCA and Error BackPropagation" **Proc. International Conference on Intelligent HumanMachine Systems and Cybernetics (IHMSC)**, August 2010. pp. 241 – 244

[3] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R. Howard,W. Hubbard,and L. Jackel. "Handwritten digit recognition with a back-propagation network" **Proc. Advances in Neural Information Processing Systems**, 1990. pp. 399

[4] L. Rabiner and B. H. Jung, "Fundamentals of Speech Recognition", **Prentice Hall, New Jersey**, (1993).

[5] Yoav Freund and Robert E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt"'95, pages 23–37. Springer-Verlag, 1995.

[6] B. Petek and J. Tebelskis, "Context- Dependent Hidden Control Neural Network Architecture for Continuous Speech Recognition", **Proceeding IEEE International Conference on Acoustics**, Speech and Signal Processing, (1992)

[7] C. Papageorgiou, M. Oren, and T. Poggio. **A general framework for object detection. In International Conference on Computer Vision**, 1998.

[8] C. Ittichaichareon, S. Suksri and T. Yingthawornsuk, speech Recognition using MFCC, **International Conference on Computer Graphics Simulation and Modeling**, 2012.

[9] Jamal Price, sophomore student, **Design an automatic speech recognition system using maltab** , University of Maryland Estern Shore Princess Anne.

[10] Allen, J.B.: **How do humans process and recognize speech IEEE Trans. on Speech and Audio Processing**, vol. 2, no. 4, pp.567--577 (1994.)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

Source Code Speech Recognition

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function varargout =GUI_SpeechRecognition(varargin)
%GUI_SPEECHRECOGNITION MATLAB code for GUI_SpeechRecognition.fig
%  GUI_SPEECHRECOGNITION, by itself, creates a new GUI_SPEECHRECOGNITION
or raises the existing
%  singleton*.
%
%  H =GUI_SPEECHRECOGNITION returns the handle to a new
GUI_SPEECHRECOGNITION or the handle to
%  the existing singleton*.
%
%  GUI_SPEECHRECOGNITION('CALLBACK',hObject,eventData,handles,...)calls the
local
%  function named CALLBACK in GUI_SPEECHRECOGNITION.M with the given
input arguments.
%
%  GUI_SPEECHRECOGNITION('Property','Value',...)creates a new
GUI_SPEECHRECOGNITION or raises the
%  existing singleton*. Starting from the left, property value pairs are
%  applied to the GUI before GUI_SpeechRecognition_OpeningFcn gets
called. An
%  unrecognized property name or invalid value makes property
application
%  stop. All inputs are passed to GUI_SpeechRecognition_OpeningFcn via
varargin.
%
%  *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only one
%  instance to run (singleton)".
%
%See also:GUIDE, GUIDATA, GUIHANDLES

%Edit the above text to modify the response to help
GUI_SpeechRecognition

%Last Modified by GUIDE v2.5 27-Mar-2017 16:55:18

%Begin initialization code -DO NOT EDIT
gui_Singleton =1;
gui_State =struct('gui_Name',       mfilename, ...
     'gui_Singleton',  gui_Singleton, ...
     'gui_OpeningFcn', @GUI_SpeechRecognition_OpeningFcn, ...
     'gui_OutputFcn',  @GUI_SpeechRecognition_OutputFcn, ...
     'gui_LayoutFcn',  [], ...
     'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback =str2func(varargin{1});
end

if nargout
 [varargout{1:nargout}]=gui_mainfcn(gui_State, varargin{:});
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT
end


% --- Executes just before GUI_SpeechRecognition is made visible.
function GUI_SpeechRecognition_OpeningFcn(hObject, eventdata, handles,
varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to GUI_SpeechRecognition (see
VARARGIN)

% Choose default command line output for GUI_SpeechRecognition
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes GUI_SpeechRecognition wait for user response (see UIRESUME)
% uiwait(handles.figure1);
imshow(imread('welcome.JPG'));
end


% --- Outputs from this function are returned to the command line.
function varargout = GUI_SpeechRecognition_OutputFcn(hObject, eventdata,
handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;
clear all;
clc;
end

% --- Executes on button press in Record.
function Record_Callback(hObject, eventdata, handles)
% hObject    handle to Record (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%Fs=44100;
%record=audiorecorder(Fs,16,1);
%msgbox('Start speaking ......','Record')
%recordblocking(record,8);
%close;
%msgbox('End of Recording ......','Record');
%close;
%data =(getaudiodata(record));
%data(1:5000)=0;
data =wavread('3.1.0F11.wav');
N =200;
%AC Coupling,Offset
offset =data -((1/length(data))*(sum(data)));
beta =-0.95;

%Pre-emphasis
pre_emp =filter([1 beta],1,offset);

%absolute normalization
abs_nor =(pre_emp/max(abs(pre_emp)));

%%%
```

```
f3=find(f2>0.3);
f4=length(f3)+1;

%%%
```

```
    logic_n =(((energy{k}-opt_th)/(abs((energy{k}-opt_th)))+1)*(1/2));
    logic_n1{k} =logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame/200values to 24000/1values
    logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%
```

```
   for k=1:(floor(length(outp_new1)/256))
            b1(:,1)=outp_new1(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new1(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);

   end
  elseif((length(outp_new1)/256)-(floor(length(outp_new1)/256)))<0.5
   for k=1:((floor(length(outp_new1)/256))-1)
            b1(:,1)=outp_new1(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new1(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%      fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
  end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %(1,cor)
%-----------------------------------
%plot(hw);
%title('Hamming Window at alpha =0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------
for k=1:N %invert 1,cor -> row,1
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
        sss{k}=sig_pass_hw;
 end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %S[k]
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 =12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3)/Fs);
for m=2:13
 for k=1:128
  if k<f(m-1);
            H(m,k)=0;
  elseif f(m-1)<=k && k<=f(m)
            H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
  elseif f(m)<=k && k<=f(m+1)
            H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
  elseif k>f(m+1)
            H(m,k)=0;
  end
 end
end
%for m=2:13
% for k=1:128
%  if k<f(m-1);
%   H(m,k)=0;
%  elseif f(m-1)<=k && k<=f(m)
%   H(m,k)=((k-f(m-1)))/((f(m)-f(m-1)));
%  elseif f(m)<=k && k<=f(m+1)
%   H(m,k)=((f(m+1)-k))/((f(m+1)-f(m)));
%  elseif k>f(m+1)
%   H(m,k)=0;
%  end
% end
%end
%-----------------------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
 for m=2:13
  for k=1:N/2
            A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
  end
 end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%Log
for i=1:length(spt_sig)
 for m=2:13
        S_mel(m-1,i)=log10(sum(A{m-1,i}));
 end
end
%DCT
for i=1:length(spt_sig)
 for m=2:13
    C_n(m-1,i)=dct(S_mel(m-1,i));
 end
end
%Feature Selection
for m=2:13
 for i=1:length(spt_sig)
    C_nn{m-1}(i,1)=C_n(m-1,i);
 end
end
%delta
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation --decimation
for m=2:13
    x=1:length(C_nn{m-1});
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    x1=x';
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
    x2=1:length(d{mm});
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
        C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_1payang');
  %----After training section So this section is working section----
 %x1=input('please input =' );
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %  4.1554  36.2801];
 [row col]=size(wij);
     N_hidden=col;
 %wjk=[-1.0315
 %  1.8269];
 [row1 col1]=size(wjk);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    N_output=col1;
%zeta_j=[5.1079 -0.0603];
%zeta_k=[1.6810];
for j=1:N_hidden
        yj(j)=sigmoid(xx*wij(:,j)-zeta_j(j));
end
for k=1:N_output
        yk(k)=sigmoid(yj*wjk(:,k)-zeta_k(k));
end
    yk
if max(yk)==yk(1);
        dos('start1.bat &')
elseif max(yk)==yk(2);
        dos('start2.bat &')
elseif max(yk)==yk(3);
        dos('start3.bat &')
elseif max(yk)==yk(4);
        dos('start4.bat &')
elseif max(yk)==yk(5);
        dos('start5.bat &')
elseif max(yk)==yk(6);
        dos('start6.bat &')
elseif max(yk)==yk(7);
        dos('start7.bat &')
elseif max(yk)==yk(8);
        dos('start8.bat &')
elseif max(yk)==yk(9);
        dos('start9.bat &')
elseif max(yk)==yk(10);
        dos('start10.bat &')
end
else
    imshow('again.JPG');
end
%%%
```

```
    energy{k}=a3;
    energy_new(k,1)=energy{k};
end
%Automatic Thresholding
ini=(mean(energy_new)+median(energy_new))/4;
th(1)=ini;
for j=1:60
    clear x_new
    logic =((energy_new-th(j))./(abs(energy_new-th(j)))+1).*(1/2);
    logic1 =zeros(1,length(abs_nor));
 for i=1:length(abs_nor)./(N)
        logic1(((i*N)-N)+1:N*i)=logic(i,1);
 end
    I =find(logic1);
    cnt=1;
 for i=I(1):I(length(I))
        x_new(cnt)=abs_nor(i);
        cnt=cnt+1;
 end
    th(j+1)=th(j)+0.05;
    l(j)=length(x_new);
end
dL=abs(diff(l));
I=find(dL==max(dL));
I=I+1;
opt_th=th(I);
if length(opt_th)>1
    opt_th=opt_th(1);
end
opt_th;
%-----------------------------------------------------
for k=1:(length(abs_nor)/N)%logic(n)
    logic_n =(((energy{k}-opt_th)./(abs((energy{k}-opt_th)))+1)*(1/2));
    logic_n1{k} =logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame 200values to 24000 1values
    logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%
```

```
 end
end

f_op=find(logic_n_new==0);
f_op1=diff(f_op);
f_op2=find(f_op1>1);
for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=0;
 end
end
logic_n_new(1:f_op(f_op2(1)),1)=0;
logic_n_new(f_op(f_op2(length(f_op2))+1):end,1)=0;

f1=diff(logic_n_new);
f4_new=length(find(abs(f1)==1));

for k=1:(length(abs_nor)-200)%Output
    outp{k}=abs_nor(k,1)*logic_n_new(k,1);
    outp_new(k,1)=outp{k};
end
j1 =find(abs(outp_new)>0);
outp_new1 =outp_new(j1);

if(f4_new/2)==2;

%--------Mel Frequency Cepstral Coefficient (MFCC)------

%Hamming Window
%Windowing
 if((length(outp_new1)/256)-(floor(length(outp_new1)/256)))>=0.5
  for k=1:(floor(length(outp_new1)/256))
            b1(:,1)=outp_new1(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new1(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%    fc((2*k)-1,1)=b2(k,1);
%    fc(2*k,1)=b4(k,1);

  end
 elseif((length(outp_new1)/256)-(floor(length(outp_new1)/256)))<0.5
  for k=1:((floor(length(outp_new1)/256))-1)
            b1(:,1)=outp_new1(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new1(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
  end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %1,cor)
%-----------------------------------------------
%plot(hw);
%title('Hamming Window at alpha =0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------------------
for k=1:N %invert 1,cor -> row,1
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
        sss{k}=sig_pass_hw;
 end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %S(k)
end

%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 =12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3)/Fs);
for m=2:13
 for k=1:128
  if k<f(m-1);
            H(m,k)=0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    elseif f(m-1)<=k && k<=f(m)
                H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
    elseif f(m)<=k && k<=f(m+1)
                H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
    elseif k>f(m+1)
                H(m,k)=0;
    end
   end
  end
%for m=2:13
% for k=1:128
%  if k<f(m-1);
%    H(m,k)=0;
%  elseif f(m-1)<=k && k<=f(m)
%    H(m,k)=((k-f(m-1)))/((f(m)-f(m-1)));
%  elseif f(m)<=k && k<=f(m+1)
%    H(m,k)=((f(m+1)-k))/((f(m+1)-f(m)));
%  elseif k>f(m+1)
%    H(m,k)=0;
%  end
% end
%end
%-----------------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
 for m=2:13
  for k=1:N/2
                A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
  end
 end
end
%Log
for i=1:length(spt_sig)
 for m=2:13
        S_mel(m-1,i)=log10(sum(A{m-1,i}));
 end
end
%DCT
for i=1:length(spt_sig)
 for m=2:13
    C_n(m-1,i)=dct(S_mel(m-1,i));
 end
end
%Feature Selection
for m=2:13
 for i=1:length(spt_sig)
    C_nn{m-1}(i,1)=C_n(m-1,i);
 end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%delta
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation - decimation
for m=2:13
    x=1:length(C_nn{m-1});
    x1=x';
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
    x2=1:length(d{mm});
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
        C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_1payang');
  %----After training section So this section is working section---
 %x1=input('please input ');
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %  4.1554  36.2801];
 [row col]=size(wij);
     N_hidden=col;
 %wjk=[-1.0315
  % 1.8269];
 [row1 col1]=size(wjk);
    N_output=col1;
 %zeta_j=[5.1079  -0.0603];
 %zeta_k=[1.6810];
 for j=1:N_hidden
        yj(j)=sigmoid(xx*wij(:,j)-zeta_j(j));
 end
 for k=1:N_output
        yk(k)=sigmoid(yj*wjk(:,k)-zeta_k(k));
 end
    yk
 if max(yk)==yk(1);
        dos('start11.bat &')
 elseif max(yk)==yk(2);
        dos('start12.bat &')
 elseif max(yk)==yk(3);
        dos('start13.bat &')
 elseif max(yk)==yk(4);
        dos('start14.bat &')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 elseif max(yk)==yk(5);
        dos('start15.bat &')
 elseif max(yk)==yk(6);
        dos('start16.bat &')
 elseif max(yk)==yk(7);
        dos('start17.bat &')
 elseif max(yk)==yk(8);
        dos('start18.bat &')
 elseif max(yk)==yk(9);
        dos('start19.bat &')
 elseif max(yk)==yk(10);
        dos('start20.bat &')
 end
else
    imshow('again.JPG');
end
%%%
```

```
%Endpoint detection
for k=1:(length(abs_nor)/N)%Absolute Energy 1 Frame=N=200
   a1(:,1)=abs_nor(1+(N*(k-1)):N*k);
     a2 =abs(a1(:,1));
     a3 =sum(a2);
     energy{k}=a3;
     energy_new(k,1)=energy{k};
end
%Automatic Thresholding
ini=(mean(energy_new)+median(energy_new))/4;
th(1)=ini;
for j=1:60
     clear x_new
     logic =((energy_new-th(j))./(abs(energy_new-th(j)))+1).*(1/2);
     logic1 =zeros(1,length(abs_nor));
 for i=1:length(abs_nor)./(N)
         logic1(((i*N)-N)+1:N*i)=logic(i,1);
 end
     I =find(logic1);
     cnt=1;
 for i=I(1):I(length(I))
         x_new(cnt)=abs_nor(i);
         cnt=cnt+1;
 end
     th(j+1)=th(j)+0.05;
     l(j)=length(x_new);
end
dL=abs(diff(l));
I=find(dL==max(dL));
I=I+1;
opt_th=th(I);
if length(opt_th)>1
     opt_th=opt_th(1);
end
opt_th;
%-----------------------------------------------------
for k=1:(length(abs_nor)/N)%logic(n)
     logic_n =(((energy{k}-opt_th)./(abs((energy{k}-opt_th)))+1)*(1/2));
     logic_n1{k} =logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame/200values to 24000/1values
     logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%
```

```
f_op2=find(f_op1>1);

for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=1;
 end
end

f_op=find(logic_n_new==0);
f_op1=diff(f_op);
f_op2=find(f_op1>1);
for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=0;
 end
end
logic_n_new(1:f_op(f_op2(1)),1)=0;
logic_n_new(f_op(f_op2(length(f_op2))+1):end,1)=0;

f1=diff(logic_n_new);
f4_new=length(find(abs(f1)==1));

if(f4_new/2)==f4;
%%%
```

```
%Hamming Window
%Windowing
 if((length(outp_new_payang3{round})/256)-
(floor(length(outp_new_payang3{round})/256)))>=0.5
   for k=1:(floor(length(outp_new_payang3{round})/256))
            b1(:,1)=outp_new_payang3{round}(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new_payang3{round}(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);

   end
 elseif((length(outp_new_payang3{round})/256)-
(floor(length(outp_new_payang3{round})/256)))<0.5
   for k=1:((floor(length(outp_new_payang3{round})/256))-1)
            b1(:,1)=outp_new_payang3{round}(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new_payang3{round}(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
 end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %(1,cor)
%-----------------------------------------
%plot(hw);
%title('Hamming Window at alpha =0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------------
for k=1:N %invert 1,cor -> row,1
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
        sss{k}=sig_pass_hw;
  end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %s[k]
end

%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 -12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3)/Fs);
for m=2:13
  for k=1:128
   if k<f(m-1);
            H(m,k)=0;
   elseif f(m-1)<=k && k<=f(m)
            H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
   elseif f(m)<=k && k<=f(m+1)
            H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
   elseif k>f(m+1)
            H(m,k)=0;
   end
  end
end

%-----------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
  for m=2:13
   for k=1:N/2
            A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
   end
  end
end
%Log
for i=1:length(spt_sig)
  for m=2:13
        S_mel(m-1,i)=log10(sum(A{m-1,i}));
  end
end
%DCT
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
for i=1:length(spt_sig)
 for m=2:13
    C_n(m-1,i)=dct(S_mel(m-1,i));
 end
end
%Feature Selection
for m=2:13
 for i=1:length(spt_sig)
    C_nn{m-1}(i,1)=C_n(m-1,i);
 end
end
%delta
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation -decimation
for m=2:13
    x=1:length(C_nn{m-1});
    x1=x';
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    x2=1:length(d{mm});
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
        C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_1payang');
  %---After training section So this section is working section---
 %x1=input('please input -' );
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %  4.1554  36.2801];
 [row col]=size(wijj{round});
     N_hidden=col;
 %wjk=[-1.0315
  %  1.8269];
 [row1 col1]=size(wjkk{round});
    N_output=col1;
 %zeta_j=[5.1079 -0.0603];
 %zeta_k=[1.6810];
 for j=1:N_hidden
        yj(j)=sigmoid(xx*wijj{round}(:,j)-zetaj{round}(j));
 end
 for k=1:N_output
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        yk(k)=sigmoid(yj*wjkk{round}(:,k)-zetak{round}(k));
 end
    yk_all(round,:)=yk;
end
yk_n(1,1)=find(yk_all(1,:)==max(yk_all(1,:)));
yk_n(1,2)=find(yk_all(2,:)==max(yk_all(2,:)));
yk_n(1,3)=find(yk_all(3,:)==max(yk_all(3,:)));

if yk_n(1,1)==2;
    yk_all(1,7)=yk_all(1,2);
elseif yk_n(1,1)==7;
    yk_all(1,2)=yk_all(1,7);
end

yk_test=sum(yk_all)/3

 if max(yk_test)==yk_test(1);
        dos('start21.bat &')
 elseif max(yk_test)==yk_test(2);
        dos('start22.bat &')
 elseif max(yk_test)==yk_test(3);
        dos('start23.bat &')
 elseif max(yk_test)==yk_test(4);
        dos('start24.bat &')
 elseif max(yk_test)==yk_test(5);
     dos('start25.bat &')
 elseif max(yk_test)==yk_test(6);
        dos('start26.bat &')
 elseif max(yk_test)==yk_test(7);
        dos('start27.bat &')
 elseif max(yk_test)==yk_test(8);
        dos('start28.bat &')
 elseif max(yk_test)==yk_test(9);
        dos('start29.bat &')
 elseif max(yk_test)==yk_test(10);
        dos('start30.bat &')
 end

else
    imshow('again.JPG');
end

%%%
```

```
load('zeta_k_4payang_1st.mat')
load('wij_4payang_1st.mat')
load('wjk_4payang_1st.mat')
zetaj{1}=zeta_j;
zetak{1}=zeta_k;
wijj{1}=wij;
wjkk{1}=wjk;

load('zeta_j_4payang_2nd.mat')
load('zeta_k_4payang_2nd.mat')
load('wij_4payang_2nd.mat')
load('wjk_4payang_2nd.mat')
zetaj{2}=zeta_j;
zetak{2}=zeta_k;
wijj{2}=wij;
wjkk{2}=wjk;

load('zeta_j_4payang_3rd.mat')
load('zeta_k_4payang_3rd.mat')
load('wij_4payang_3rd.mat')
load('wjk_4payang_3rd.mat')
zetaj{3}=zeta_j;
zetak{3}=zeta_k;
wijj{3}=wij;
wjkk{3}=wjk;

load('zeta_j_4payang_4th.mat')
load('zeta_k_4payang_4th.mat')
load('wij_4payang_4th.mat')
load('wjk_4payang_4th.mat')
zetaj{4}=zeta_j;
zetak{4}=zeta_k;
wijj{4}=wij;
wjkk{4}=wjk;

%Endpoint detection
for k=1:(length(abs_nor)/N)%Absolute Energy 1 Frame N=200
    a1(:,1)=abs_nor(1+(N*(k-1)):N*k);
    a2 =abs(a1(:,1));
    a3 =sum(a2);
    energy{k}=a3;
    energy_new(k,1)=energy{k};
end
%Automatic Thresholding
ini=(mean(energy_new)+median(energy_new))/4;
th(1)=ini;
for j=1:60
    clear x_new
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    logic =((energy_new-th(j))./(abs(energy_new-th(j)))+1).*(1/2);
    logic1 =zeros(1,length(abs_nor));
 for i=1:length(abs_nor)./(N)
        logic1(((i*N)-N)+1:N*i)=logic(i,1);
 end
    I =find(logic1);
    cnt=1;
 for i=I(1):I(length(I))
        x_new(cnt)=abs_nor(i);
        cnt=cnt+1;
 end
    th(j+1)=th(j)+0.05;
    l(j)=length(x_new);
end
dL=abs(diff(l));
I=find(dL==max(dL));
I=I+1;
opt_th=th(I);
if length(opt_th)>1
    opt_th=opt_th(1);
end
opt_th;
%----------------------------------------------
for k=1:(length(abs_nor)/N)%logic(n)
    logic_n =(((energy{k}-opt_th)./(abs((energy{k}-opt_th)))+1)*(1/2));
    logic_n1{k} =logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame 200values to 24000 1values
    logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%optimization logic
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
f_op=find(logic_n_new==1);
f_op1=diff(f_op);
f_op2=find(f_op1>1);

for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=1;
 end
end

f_op=find(logic_n_new==0);
f_op1=diff(f_op);
f_op2=find(f_op1>1);
for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=0;
 end
end
logic_n_new(1:f_op(f_op2(1)),1)=0;
logic_n_new(f_op(f_op2(length(f_op2))+1):end,1)=0;

f1=diff(logic_n_new);
f4_new=length(find(abs(f1)==1));

if(f4_new/2)==f4;
%%%
```

```
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);

   end
  elseif((length(outp_new_payang4{round})/256)-
(floor(length(outp_new_payang4{round})/256)))<0.5
   for k=1:((floor(length(outp_new_payang4{round})/256))-1)
            b1(:,1)=outp_new_payang4{round}(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new_payang4{round}(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
  end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %1,cor)
%-----------------------------------------
%plot(hw);
%title('Hamming Window at alpha = 0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------------
for k=1:N %invert 1,cor -> row,1
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
        sss{k}=sig_pass_hw;
 end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %S[k]
end

%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 =12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3/Fs);
for m=2:13
 for k=1:128
  if k<f(m-1);
            H(m,k)=0;
  elseif f(m-1)<=k && k<=f(m)
            H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
  elseif f(m)<=k && k<=f(m+1)
            H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
  elseif k>f(m+1)
            H(m,k)=0;
  end
 end
end

%-----------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
 for m=2:13
  for k=1:N/2
            A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
  end
 end
end
%Log
for i=1:length(spt_sig)
 for m=2:13
       S_mel(m-1,i)=log10(sum(A{m-1,i}));
 end
end
%DCT
for i=1:length(spt_sig)
 for m=2:13
    C_n(m-1,i)=dct(S_mel(m-1,i));
 end
end
%Feature Selection
for m=2:13
 for i=1:length(spt_sig)
    C_nn{m-1}(i,1)=C_n(m-1,i);
 end
end
%delta
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation -decimation
for m=2:13
    x=1:length(C_nn{m-1});
    x1=x';
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
    x2=1:length(d{mm});
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
    C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_1payang');
  %----After training section So this section is working section----
 %x1=input('please input = ');
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %  4.1554  36.2801];
 [row col]=size(wijj{round});
     N_hidden=col;
 %wjk=[-1.0315
 %  1.8269];
 [row1 col1]=size(wjkk{round});
    N_output=col1;
 %zeta_j=[5.1079  -0.0603];
 %zeta_k=[1.6810];
 for j=1:N_hidden
        yj(j)=sigmoid(xx*wijj{round}(:,j)-zetaj{round}(j));
 end
 for k=1:N_output
        yk(k)=sigmoid(yj*wjkk{round}(:,k)-zetak{round}(k));
 end
    yk_all(round,:)=yk;
end
yk_n(1,1)=find(yk_all(1,:)==max(yk_all(1,:)));
yk_n(1,2)=find(yk_all(2,:)==max(yk_all(2,:)));
yk_n(1,3)=find(yk_all(3,:)==max(yk_all(3,:)));
yk_n(1,4)=find(yk_all(4,:)==max(yk_all(4,:)));

if yk_n(1,1)==5;
    yk_all(1,8)=yk_all(1,5);
elseif yk_n(1,1)==8;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    yk_all(1,5)=yk_all(1,8);
end

if yk_n(1,2)==4;
    yk_all(2,5)=yk_all(2,4);
    yk_all(2,8)=yk_all(2,4);
elseif yk_n(1,2)==5;
    yk_all(2,4)=yk_all(2,5);
    yk_all(2,8)=yk_all(2,5);
elseif yk_n(1,2)==8;
  yk_all(2,4)=yk_all(2,8);
    yk_all(2,5)=yk_all(2,8);
end

yk_test=sum(yk_all)/4

 if max(yk_test)==yk_test(1);
        dos('start31.bat &')
 elseif max(yk_test)==yk_test(2);
        dos('start32.bat &')
 elseif max(yk_test)==yk_test(3);
    dos('start33.bat &')
 elseif max(yk_test)==yk_test(4);
        dos('start34.bat &')
 elseif max(yk_test)==yk_test(5);
        dos('start35.bat &')
 elseif max(yk_test)==yk_test(6);
        dos('start36.bat &')
 elseif max(yk_test)==yk_test(7);
        dos('start37.bat &')
 elseif max(yk_test)==yk_test(8);
        dos('start38.bat &')
 elseif max(yk_test)==yk_test(9);
        dos('start39.bat &')
 elseif max(yk_test)==yk_test(10);
        dos('start40.bat &')
 end

else
    imshow('again.JPG');
end
%%%
```

```
load('wij_5payang_1st.mat')
load('wjk_5payang_1st.mat')
zetaj{1}=zeta_j;
zetak{1}=zeta_k;
wijj{1}=wij;
wjkk{1}=wjk;

load('zeta_j_5payang_2nd.mat')
load('zeta_k_5payang_2nd.mat')
load('wij_5payang_2nd.mat')
load('wjk_5payang_2nd.mat')
zetaj{2}=zeta_j;
zetak{2}=zeta_k;
wijj{2}=wij;
wjkk{2}=wjk;

load('zeta_j_5payang_3rd.mat')
load('zeta_k_5payang_3rd.mat')
load('wij_5payang_3rd.mat')
load('wjk_5payang_3rd.mat')
zetaj{3}=zeta_j;
zetak{3}=zeta_k;
wijj{3}=wij;
wjkk{3}=wjk;

load('zeta_j_5payang_4th.mat')
load('zeta_k_5payang_4th.mat')
load('wij_5payang_4th.mat')
load('wjk_5payang_4th.mat')
zetaj{4}=zeta_j;
zetak{4}=zeta_k;
wijj{4}=wij;
wjkk{4}=wjk;

load('zeta_j_5payang_5th.mat')
load('zeta_k_5payang_5th.mat')
load('wij_5payang_5th.mat')
load('wjk_5payang_5th.mat')
zetaj{5}=zeta_j;
zetak{5}=zeta_k;
wijj{5}=wij;
wjkk{5}=wjk;


%Endpoint detection
for k=1:(length(abs_nor)/N)%Absolute Energy 1 Frame/N-200
    a1(:,1)=abs_nor(1+(N*(k-1)):N*k);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    a2 =abs(a1(:,1));
    a3 =sum(a2);
    energy{k}=a3;
    energy_new(k,1)=energy{k};
end
%Automatic Thresholding
ini=(mean(energy_new)+median(energy_new))/4;
th(1)=ini;
for j=1:60
    clear x_new
    logic =((energy_new-th(j))./(abs(energy_new-th(j)))+1).*(1/2);
    logic1 =zeros(1,length(abs_nor));
 for i=1:length(abs_nor)./(N)
        logic1(((i*N)-N)+1:N*i)=logic(i,1);
 end
    I =find(logic1);
    cnt=1;
 for i=I(1):I(length(I))
        x_new(cnt)=abs_nor(i);
        cnt=cnt+1;
 end
    th(j+1)=th(j)+0.05;
    l(j)=length(x_new);
end
dL=abs(diff(l));
I=find(dL==max(dL));
I=I+1;
opt_th=th(I);
if length(opt_th)>1
    opt_th=opt_th(1);
end
opt_th;
%-----------------------------------------
for k=1:(length(abs_nor)/N)%logic(n)
    logic_n =(((energy{k}-opt_th)./(abs((energy{k}-opt_th)))+1)*(1/2));
    logic_n1{k}=logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame 200values to 24000 1values
    logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%
```

```
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=1;
 end
end

f_op=find(logic_n_new==0);
f_op1=diff(f_op);
f_op2=find(f_op1>1);
for pp=1:length(f_op2);
 if(f_op(f_op2(pp)+1)-f_op(f_op2(pp)))<4000;
        logic_n_new(f_op(f_op2(pp)):f_op(f_op2(pp)+1),1)=0;
 end
end
logic_n_new(1:f_op(f_op2(1)),1)=0;
logic_n_new(f_op(f_op2(length(f_op2))+1):end,1)=0;

f1=diff(logic_n_new);
f4_new=length(find(abs(f1)==1));

if(f4_new/2)==f4;
%%%
```

```
%Hamming Window
%Windowing
 if((length(outp_new_payang5{round})/256)-
(floor(length(outp_new_payang5{round})/256)))>=0.5
   for k=1:(floor(length(outp_new_payang5{round})/256))
            b1(:,1)=outp_new_payang5{round}(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new_payang5{round}(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);

   end
  elseif((length(outp_new_payang5{round})/256)-
(floor(length(outp_new_payang5{round})/256)))<0.5
   for k=1:((floor(length(outp_new_payang5{round})/256))-1)
            b1(:,1)=outp_new_payang5{round}(1+(256*(k-1)):256*k);
            b2(k,1)=b1((256/2),1);
            b3(:,1)=outp_new_payang5{round}(128+(256*(k-1)):((256*k)+128)-1);
            b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
            b5{(2*k)-1}=b1(:,1);
            b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
  end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %(1,cor)
%-----------------------------------------------
%plot(hw);
%title('Hamming Window at alpha = 0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------------------
for k=1:N %invert 1,cor -> row,1
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        sss{k}=sig_pass_hw;
  end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %s[k]
end

%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 =12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3)/Fs);
for m=2:13
  for k=1:128
    if k<f(m-1);
              H(m,k)=0;
    elseif f(m-1)<=k && k<=f(m)
              H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
    elseif f(m)<=k && k<=f(m+1)
              H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
    elseif k>f(m+1)
              H(m,k)=0;
    end
  end
end

%-----------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
  for m=2:13
    for k=1:N/2
            A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
    end
  end
end
%Log
for i=1:length(spt_sig)
  for m=2:13
        S_mel(m-1,i)=log10(sum(A{m-1,i}));
  end
end
%DCT
for i=1:length(spt_sig)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
  for m=2:13
      C_n(m-1,i)=dct(S_mel(m-1,i));
  end
 end
%Feature Selection
for m=2:13
  for i=1:length(spt_sig)
      C_nn{m-1}(i,1)=C_n(m-1,i);
  end
 end
%delta
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation -decimation
for m=2:13
    x=1:length(C_nn{m-1});
    x1=x';
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
    x2=1:length(d{mm});
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
        C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_lpayang');
  %----After training section So this section is working section----
 %x1=input('please input =' );
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %    4.1554  36.2801];
 [row col]=size(wijj{round});
     N_hidden=col;
 %wjk=[-1.0315
 %    1.8269];
 [row1 col1]=size(wjkk{round});
    N_output=col1;
 %zeta_j=[5.1079  -0.0603];
 %zeta_k=[1.6810];
 for j=1:N_hidden
        yj(j)=sigmoid(xx*wijj{round}(:,j)-zetaj{round}(j));
 end
 for k=1:N_output
        yk(k)=sigmoid(yj*wjkk{round}(:,k)-zetak{round}(k));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 end
    yk_all(round,:)=yk;
end
yk_n(1,1)=find(yk_all(1,:)==max(yk_all(1,:)));
yk_n(1,2)=find(yk_all(2,:)==max(yk_all(2,:)));
yk_n(1,3)=find(yk_all(3,:)==max(yk_all(3,:)));
yk_n(1,4)=find(yk_all(4,:)==max(yk_all(4,:)));
yk_n(1,5)=find(yk_all(5,:)==max(yk_all(5,:)));

if yk_n(1,1)==5;
    yk_all(1,2)=yk_all(1,5);
elseif yk_n(1,1)==2;
    yk_all(1,5)=yk_all(1,2);
end

if yk_n(1,2)==6;
    yk_all(2,4)=yk_all(2,6);
elseif yk_n(1,2)==4;
    yk_all(2,6)=yk_all(2,4);
end

yk_test=sum(yk_all)/5

 if max(yk_test)==yk_test(1);
        dos('start41.bat &')
 elseif max(yk_test)==yk_test(2);
        dos('start42.bat &')
 elseif max(yk_test)==yk_test(3);
        dos('start43.bat &')
 elseif max(yk_test)==yk_test(4);
        dos('start44.bat &')
 elseif max(yk_test)==yk_test(5);
        dos('start45.bat &')
 elseif max(yk_test)==yk_test(6);
        dos('start46.bat &')
 end
else
    imshow('again.JPG');
end

%%%
```

```
load('wjk_6payang_1st.mat')
zetaj{1}=zeta_j;
zetak{1}=zeta_k;
wijj{1}=wij;
wjkk{1}=wjk;

load('zeta_j_6payang_2nd.mat')
load('zeta_k_6payang_2nd.mat')
load('wij_6payang_2nd.mat')
load('wjk_6payang_2nd.mat')
zetaj{2}=zeta_j;
zetak{2}=zeta_k;
wijj{2}=wij;
wjkk{2}=wjk;

load('zeta_j_6payang_3rd.mat')
load('zeta_k_6payang_3rd.mat')
load('wij_6payang_3rd.mat')
load('wjk_6payang_3rd.mat')
zetaj{3}=zeta_j;
zetak{3}=zeta_k;
wijj{3}=wij;
wjkk{3}=wjk;

load('zeta_j_6payang_4th.mat')
load('zeta_k_6payang_4th.mat')
load('wij_6payang_4th.mat')
load('wjk_6payang_4th.mat')
zetaj{4}=zeta_j;
zetak{4}=zeta_k;
wijj{4}=wij;
wjkk{4}=wjk;

load('zeta_j_6payang_5th.mat')
load('zeta_k_6payang_5th.mat')
load('wij_6payang_5th.mat')
load('wjk_6payang_5th.mat')
zetaj{5}=zeta_j;
zetak{5}=zeta_k;
wijj{5}=wij;
wjkk{5}=wjk;

load('zeta_j_6payang_6th.mat')
load('zeta_k_6payang_6th.mat')
load('wij_6payang_6th.mat')
load('wjk_6payang_6th.mat')
zetaj{6}=zeta_j;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
zetak{6}=zeta_k;
wijj{6}=wij;
wjkk{6}=wjk;

%Endpoint detection
for k=1:(length(abs_nor)/N)%Absolute Energy 1 Frame=N=200
    a1(:,1)=abs_nor(1+(N*(k-1)):N*k);
    a2 =abs(a1(:,1));
    a3 =sum(a2);
    energy{k}=a3;
    energy_new(k,1)=energy{k};
end
%Automatic Thresholding
ini=(mean(energy_new)+median(energy_new))/4;
th(1)=ini;
for j=1:60
    clear x_new
    logic =((energy_new-th(j))./(abs(energy_new-th(j)))+1).*(1/2);
    logic1 =zeros(1,length(abs_nor));
 for i=1:length(abs_nor)./(N)
        logic1(((i*N)-N)+1:N*i)=logic(i,1);
 end
    I =find(logic1);
    cnt=1;
 for i=I(1):I(length(I))
        x_new(cnt)=abs_nor(i);
        cnt=cnt+1;
 end
    th(j+1)=th(j)+0.05;
    l(j)=length(x_new);
end
dL=abs(diff(l));
I=find(dL==max(dL));
I=I+1;
opt_th=th(I);
if length(opt_th)>1
    opt_th=opt_th(1);
end
opt_th;
%-----------------------------------------------------------------
for k=1:(length(abs_nor)/N)%logic(n)
    logic_n =(((energy{k}-opt_th)./(abs((energy{k}-opt_th)))+1)*(1/2));
    logic_n1{k} =logic_n;
end
for k=1:(length(abs_nor)/N)%amplify 120 frame*200values to 24000*1values
    logic_n_new(1+(N*(k-1)):N*k,1)=logic_n1{k};
end

%%%
```

```
%optimization logic
%%%
```

```
outp_new_payang6{6}=outp_new(f_op(f_op2(5)+1):f_op(end),1);
%j1 =find(abs(outp_new)>0);
%outp_new1 =outp_new(j1);
%sound(outp_new1,Fs);
for round=1:f4;
%--------Mel Frequency Cepstral Coefficient (MFCC)--------

%Hamming Window
%Windowing
 if((length(outp_new_payang6{round})/256)-
(floor(length(outp_new_payang6{round})/256)))>=0.5
   for k=1:(floor(length(outp_new_payang6{round})/256))
              b1(:,1)=outp_new_payang6{round}(1+(256*(k-1)):256*k);
              b2(k,1)=b1((256/2),1);
              b3(:,1)=outp_new_payang6{round}(128+(256*(k-1)):((256*k)+128)-1);
              b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
              b5{(2*k)-1}=b1(:,1);
              b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);

   end
 elseif((length(outp_new_payang6{round})/256)-
(floor(length(outp_new_payang6{round})/256)))<0.5
   for k=1:((floor(length(outp_new_payang6{round})/256))-1)
              b1(:,1)=outp_new_payang6{round}(1+(256*(k-1)):256*k);
              b2(k,1)=b1((256/2),1);
              b3(:,1)=outp_new_payang6{round}(128+(256*(k-1)):((256*k)+128)-1);
              b4(k,1)=(b3((256/2),1)+b3((256/2)+1,1))/2;
              b5{(2*k)-1}=b1(:,1);
              b5{2*k}=b3(:,1);
%     fc((2*k)-1,1)=b2(k,1);
%     fc(2*k,1)=b4(k,1);
   end
 end

%Generailzed Haming Window
alpha=0.46;
N=256; %size of window
y=0:(N-1);
hw =((1-alpha)-(alpha*cos((2*pi*y)/(N-1)))); %(1,cor)
%-----------------------------------------------
%plot(hw);
%title('Hamming Window at alpha =0.47');
%xlabel('Sample');ylabel('Amplitude');
%axis([1 256 0 1]);
%-----------------------------------------------
for k=1:N %invert 1,cor -> row,1
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    hw_new(k,1)=hw(1,k);
end

%Multiplier between Output with Hamming window
for k=1:length(b5)
 for i=1:N
        sig_pass_hw(i,1)=b5{k}(i,1)*hw_new(i,1);
        sss{k}=sig_pass_hw;
 end
%|FFT|
spt_sig{k}=abs(fftshift(fft(sig_pass_hw))); %s[k]
end

%Mel Scale Filterbank
Fs=44100;
fmin =300;
fmax =44100;
m1 =1125*log10(1+fmin/700)/log10(exp(1));
m2 =1125*log10(1+fmax/700)/log10(exp(1));
m  =linspace(m1,m2,14); %14 - 12 filter bank
h1 =700*(exp(m1/1125)-1);
h2 =700*(exp(m2/1125)-1);
h  =linspace(h1,h2,14);
h3 =700*(exp(m/1125)-1);
f =floor((128+1)*(h3)/Fs);
for m=2:13
 for k=1:128
  if k<f(m-1);
            H(m,k)=0;
  elseif f(m-1)<=k && k<=f(m)
            H(m,k)=(2*(k-f(m-1)))/((f(m+1)-f(m-1))*(f(m)-f(m-1)));
  elseif f(m)<=k && k<=f(m+1)
            H(m,k)=(2*(f(m+1)-k))/((f(m+1)-f(m-1))*(f(m+1)-f(m)));
  elseif k>f(m+1)
            H(m,k)=0;
  end
 end
end

%-------------------------------------------------------------
%Spectrum(b9)Through Filterbank
for i=1:length(spt_sig)
 for m=2:13
  for k=1:N/2
            A{m-1,i}(1,k)=spt_sig{i}(k+N/2,1)*H(m,k);
  end
 end
end
%Log
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
for i=1:length(spt_sig)
 for m=2:13
        S_mel(m-1,i)=log10(sum(A{m-1,i}));
 end
end
%DCT
for i=1:length(spt_sig)
 for m=2:13
    C_n(m-1,i)=dct(S_mel(m-1,i));
 end
end
%Feature Selection
for m=2:13
 for i=1:length(spt_sig)
    C_nn{m-1}(i,1)=C_n(m-1,i);
 end
end
%delta
d{1}=(C_nn{2})/2;
d{2}=((C_nn{3}-C_nn{1})/2);
d{3}=((C_nn{4}-C_nn{2})/2);
d{4}=((C_nn{5}-C_nn{3})/2);
d{5}=((C_nn{6}-C_nn{4})/2);
d{6}=((C_nn{7}-C_nn{5})/2);
d{7}=((C_nn{8}-C_nn{6})/2);
d{8}=((C_nn{9}-C_nn{7})/2);
d{9}=((C_nn{10}-C_nn{8})/2);
d{10}=((C_nn{11}-C_nn{9})/2);
d{11}=((C_nn{12}-C_nn{10})/2);
d{12}=(-(C_nn{11}))/2;
%delta_delta
dd{1}=(d{2})/2;
dd{2}=((d{3}-d{1})/2);
dd{3}=((d{4}-d{2})/2);
dd{4}=((d{5}-d{3})/2);
dd{5}=((d{6}-d{4})/2);
dd{6}=((d{7}-d{5})/2);
dd{7}=((d{8}-d{6})/2);
dd{8}=((d{9}-d{7})/2);
dd{9}=((d{10}-d{8})/2);
dd{10}=((d{11}-d{9})/2);
dd{11}=((d{12}-d{10})/2);
dd{12}=(-(d{11}))/2;
%C_nn_delta=[d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12];
%interpolation --decimation
for m=2:13
    x=1:length(C_nn{m-1});
    x1=x';
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 for k=1:120;
    C_nnn{m-1}(k,1)=interp1(x1,(C_nn{m-1}),(length(C_nn{m-1})/120)*k,'spline');
 end
end

for mm=1:12
    x2=1:length(d{mm});
    x3=x2';
 for k2=1:120;
    d_new{mm}(k2,1)=interp1(x3,(d{mm}),(length(d{mm})/120)*k2,'spline');
 end
end

for mmm=1:12
    xdd=1:length(dd{mmm});
    x3=xdd';
 for k3=1:120;
    dd_new{mmm}(k3,1)=interp1(x3,(dd{mmm}),(length(dd{mmm})/120)*k3,'spline');
 end
end
        C_total=[C_nnn{1}' C_nnn{2}' C_nnn{3}' C_nnn{4}' C_nnn{5}'
C_nnn{6}' C_nnn{7}' C_nnn{8}' C_nnn{9}' C_nnn{10}' C_nnn{11}'
C_nnn{12}'];
  %C(:,iii)=C_total/max(abs(C_total));
        C_nn_delta=[d_new{1}' d_new{2}' d_new{3}' d_new{4}' d_new{5}'
d_new{6}' d_new{7}' d_new{8}' d_new{9}' d_new{10}' d_new{11}'
d_new{12}'];

        C_nn_delta_delta=[dd_new{1}' dd_new{2}' dd_new{3}' dd_new{4}'
dd_new{5}' dd_new{6}' dd_new{7}' dd_new{8}' dd_new{9}' dd_new{10}'
dd_new{11}' dd_new{12}'];

        C1(1,1:(12*120))=C_total/max(abs(C_total));
        C1(1,((12*120)+1):((12*120)*2))=C_nn_delta/max(abs(C_nn_delta));

C1(1,(((12*120)*2)+1):((12*120)*3))=C_nn_delta_delta/max(abs(C_nn_delta_delta));
  %save('total_unnor_c14_interp1_20_1payang');
  %---After training section So this section is working section---
 %x1=input('please input  ' );
    x_1=C1;
    xx =x_1;
 %wij=[4.3554  36.1801;
 %  4.1554  36.2801];
 [row col]=size(wijj{round});
     N_hidden=col;
 %wjk=[-1.0315
  %  1.8269];
 [row1 col1]=size(wjkk{round});
    N_output=col1;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 %zeta_j=[5.1079 -0.0603];
 %zeta_k=[1.6810];
 for j=1:N_hidden
        yj(j)=sigmoid(xx*wijj{round}(:,j)-zetaj{round}(j));
 end
 for k=1:N_output
     yk(k)=sigmoid(yj*wjkk{round}(:,k)-zetak{round}(k));
 end
    yk_all(round,:)=yk;
end
yk_n(1,1)=find(yk_all(1,:)==max(yk_all(1,:)));
yk_n(1,2)=find(yk_all(2,:)==max(yk_all(2,:)));
yk_n(1,3)=find(yk_all(3,:)==max(yk_all(3,:)));
yk_n(1,4)=find(yk_all(4,:)==max(yk_all(4,:)));
yk_n(1,5)=find(yk_all(5,:)==max(yk_all(5,:)));
yk_n(1,6)=find(yk_all(6,:)==max(yk_all(6,:)));

if yk_n(1,6)==3;
    yk_all(6,4)=yk_all(6,3);
elseif yk_n(1,6)==4;
    yk_all(6,3)=yk_all(6,4);
end

yk_test=sum(yk_all)/6

 if max(yk_test)==yk_test(1);
        dos('start47.bat &')
 elseif max(yk_test)==yk_test(2);
        dos('start48.bat &')
 elseif max(yk_test)==yk_test(3);
        dos('start49.bat &')
 elseif max(yk_test)==yk_test(4);
        dos('start50.bat &')
 end
else
    imshow('again.JPG');
end
%%%
```

```
%%%
```

ภาคผนวก ข

Source Code Facial Image Recognition

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
function varargout = untitled(varargin)
% UNTITLED MATLAB code for untitled.fig
%      UNTITLED, by itself, creates a new UNTITLED or raises the existing
%      singleton*.
%
%      H = UNTITLED returns the handle to a new UNTITLED or the handle to
%      the existing singleton*.
%
%      UNTITLED('CALLBACK',hObject,eventData,handles,...) calls the local
%      function named CALLBACK in UNTITLED.M with the given input arguments.
%
%      UNTITLED('Property','Value',...) creates a new UNTITLED or raises the
%      existing singleton*.  Starting from the left, property value pairs are
%      applied to the GUI before untitled_OpeningFcn gets called.  An
%      unrecognized property name or invalid value makes property application
%      stop.  All inputs are passed to untitled_OpeningFcn via varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help untitled

% Last Modified by GUIDE v2.5 18-Apr-2017 15:20:00

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @untitled_OpeningFcn, ...
                   'gui_OutputFcn',  @untitled_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT



end
% --- Executes just before untitled is made visible.
function untitled_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to untitled (see VARARGIN)

% Choose default command line output for untitled

handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

end
% UIWAIT makes untitled wait for user response (see UIRESUME)
% uiwait(handles.figure1);


% --- Outputs from this function are returned to the command line.
function varargout = untitled_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;
end


% --- Executes on button press in pushbutton1.
function pushbutton1_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton1 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

 zeta_k=[];
 zeta_j=[];
 load('weig.mat')
clc;
camlist=webcamlist;
% Create the point tracker object.
pointTracker = vision.PointTracker('MaxBidirectionalError', 2);
cam=webcam(2);
preview(cam)

while (1)
    img = snapshot(cam);
     pause(2)
  %    image(img);
  %    title('input form camera')
    % Create the face detector object.
%To detect Face
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
FDetect = vision.CascadeObjectDetector;
detector.MergeThreshold = 2;
%Read the input image
I = img;

%Returns Bounding Box values based on number of objects
BB = step(FDetect,I);
%------------------For single face--------------------
% a = insertObjectAnnotation(I,'rectangle',BB,'I');
%     face = imcrop(I,BB);
%     figure,imshow(face);
%     clear BB

%-------------------For multi face------------------
for i = 1:size(BB,1)
   a=  rectangle('Position',BB(i,:),'LineWidth',5,'LineStyle','-','EdgeColor','r');

   face = imcrop(I,BB(i,:));

%-----------------classifily crop image true or false---------
if (length(face)>150 & length(face)<600)
     %-----------------Change img for pre-processing--------
     FaceGray=rgb2gray(face);
     FacePreprocessing=imresize(FaceGray,[100,100]);

M=120;
%Chosen std and mean.
%It can be any number that it is close to the std and mean of most of the images.
um=100;

ustd=80;

%read and show images(jpg);
S=[];   %img matrix
figure(1);
for i=1:M
     str=strcat(int2str(i),'.jpg');    %concatenates two strings that form the name of the image
     eval('img=imread(str);');
     [irow icol]=size(img);     % get the number of rows (N1) and columns (N2)
     temp=reshape(img',irow*icol,1);     %creates a (N1*N2)x1 matrix
     S=[S temp];              %X is a N1*N2xM matrix after finishing the sequence
                               %this is our S
end


%reduce the error due to lighting conditions.(Normalize)
for i=1:size(S,2)
     temp=double(S(:,i));
     m=mean(temp);
     st=std(temp);
     S(:,i)=(temp-m)*ustd/st+um;
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%mean image;
m=mean(S,2);    %Find mean of img.
tmimg=uint8(m);
img=reshape(tmimg,icol,irow);    %creates a N2xN1 matrix
img=img';        %creates a N1xN2 matrix by transposing the image.
m2=m;

% Change image for manipulation
dbx=[];    % A matrix
for i=1:M
      temp=double(S(:,i));
      dbx=[dbx temp];
end

%Covariance matrix C=A'A, L=AA'
A=dbx';
L=A*A';
% vv are the eigenvector for L
% dd are the eigenvalue for both L=dbx'*dbx and C=dbx*dbx';
%%% [V,D] = eig(A) produces a diagonal matrix D of eigenvalues and ------%
%%%      a full matrix V whose columns are the corresponding eigenvectors ---%
%%%      so that A*V = V*D.-----------------------------------------------%
[vv dd]=eig(L);
% Sort and eliminate those whose eigenvalue is zero
v=[];
d=[];
for i=1:size(vv,2)
     if(dd(i,i)>1e-4)
          v=[v vv(:,i)];
          d=[d dd(i,i)];
     end
 end

 %sort,   will return an ascending order
 [B, index]=sort(d);
 ind=zeros(size(index));
 dtemp=zeros(size(index));
 vtemp=zeros(size(v));
 len=length(index);
 for i=1:len
     dtemp(i)=B(len+1-i);
     ind(i)=len+1-index(i);
     vtemp(:,ind(i))=v(:,i);
 end
 d=dtemp;
 v=vtemp;
% d = eigenvalue   v=eigenvector for L (Covarience Metrix)

%Eigenvectors of Covarance matrix
u=[];
for i=1:size(v,2)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        temp=sqrt(d(i));
        u=[u (dbx*v(:,i))./temp];
end


% Find the weight of each face in the training set.
omega = [];
for h=1:size(dbx,2)
        WW=[];
        for i=1:size(u,2)
                t = u(:,i)';
                WeightOfImage = dot(t,dbx(:,h)');
                WW = [WW; WeightOfImage];
        end
        omega = [omega WW];
end

%obtain image form webcam(2)
InputImage = FacePreprocessing;
InImage=reshape(double(InputImage)',irow*icol,1);
temp=InImage;
me=mean(temp);
st=std(temp);
temp=(temp-me)*ustd/st+um;
NormImage = temp;
Difference = temp-m;

p = [];
 %size of data input
aa=size(u,2);
for i = 1:aa
        pare = dot(NormImage,u(:,i));
        p = [p; pare];
end
ReshapedImage = m + u(:,1:aa)*p;    %m is the mean image, u is the eigenvector
ReshapedImage = reshape(ReshapedImage,icol,irow);
ReshapedImage = ReshapedImage';

InImWeight = [];
for i=1:size(u,2)
        t = u(:,i)';
        WeightOfInputImage = dot(t,Difference');
        InImWeight = [InImWeight; WeightOfInputImage];
end

% Find Euclidean distance
e=[];
for i=1:size(omega,2)
        q = omega(:,i);
        DiffWeight = InImWeight-q;
        mag = norm(DiffWeight);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        e = [e mag];
    end
    e=e/max(e);
    %optimize feature,sort,1-,*2 and gg
    for j=1 : 1

        fak(j,1:30)=sort(e(j,1:30));

        fak(j,31:60)=sort(e(j,31:60));

        fak(j,61:90)=sort(e(j,61:90));

        fak(j,91:120)=sort(e(j,91:120));

    end

    fak=(1-fak);
    fak=fak*2;

            %----After training section So this section is working section----
        %x1=input('please input =' );

        xx = fak;
        %wij=[4.3554  36.1801;
       %       4.1554  36.2801];
        [row  col] =size(wij);
         N_hidden=col;
       % wjk=[-1.0315
         %    1.8269];
        [row1  col1]=size(wjk);
        N_output=col1;
       % zeta_j=[5.1079  -0.0603];
        %zeta_k=[1.6810];
        for j=1:N_hidden
            yj(j)=sigmoid(xx*wij(:,j)-zeta_j(j));
        end
        for k=1:N_output
            yk(k)=sigmoid(yj*wjk(:,k)-zeta_k(k));
        end
    yk
            if max(yk)==yk(:,1)
                disp('manz')
                axes(handles.axes1);
                imshow(face)
               axes(handles.axes3);
                imshow('MANN.jpg')
                dos('man.bat &')
      %           implay('manz.mp4');
                pause(20)
            elseif max(yk) == yk(:,2)
                disp('womanz')
                axes(handles.axes1);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                imshow(face)
                 axes(handles.axes3);
                imshow('WOMAN.jpg')
                dos('woman.bat &')
%                 implay('whoman.avi');
                pause(20)
            elseif max(yk) == yk(:,3)
                 disp('kidz')
                  axes(handles.axes1);
                imshow(face)
                 axes(handles.axes3);
                 imshow('KID.jpg')
                 dos('kid.bat &')
%                   implay('kid.avi');
                 pause(20)
            else max(yk) == yk(:,4)
                disp('Oldguy')
                 axes(handles.axes1);
                imshow(face)
                 axes(handles.axes3);
                imshow('OLD.jpg')
%                   implay('oldguy.avi');
                dos('oldguy.bat &')
                pause(20)
            end

else
 disp('Waitingforface')

end
end
end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้