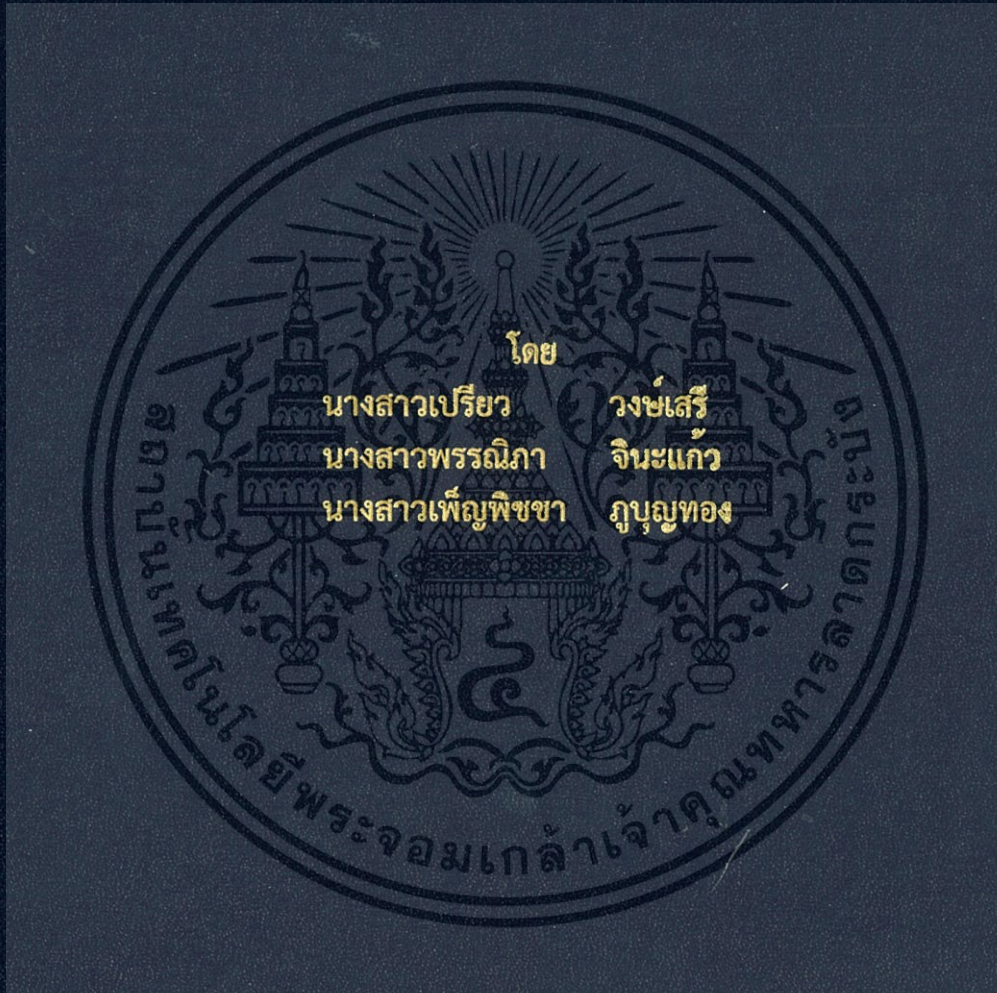


ระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง
ACCESS CONTROLLING SYSTEM USING SPEECH RECOGNITION



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมโทรคมนาคม
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

ระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง
ACCESS CONTROLLING SYSTEM USING SPEECH RECOGNITION



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมโทรคมนาคม
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง
ACCESS CONTROLLING SYSTEM USING SPEECH RECOGNITION

โดย

นางสาวเปรี๊ยะ	วงศ์เสรี	56010771
นางสาวพรรณิภา	จินะแก้ว	56010810
นางสาวเพ็ญพิชชา	ภุญญทอง	56010893

อาจารย์ที่ปรึกษา
รศ.ดร.ปราโมทย์ วาดเขียน
รศ.ดร.จิรสุดา โกษิยาภรณ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมโทรคมนาคม
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2559

ผ่านการตรวจรูปเล่มแล้ว

อาจารย์ที่ปรึกษา
24/5/60
วิศวกรรมโทรคมนาคม

ผ่านการตรวจชิ้นงานแล้ว

กรรมการผู้ตรวจชิ้นงาน
24/5/60
วิศวกรรมโทรคมนาคม

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งที่ 24/5/60 ของเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำ 24/5/60

ปริญญาานิพนธ์ปีการศึกษา 2559

ภาควิชาวิศวกรรมโทรคมนาคม

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง

ACCESS CONTROLLING SYSTEM USING SPEECH RECOGNITION

ผู้จัดทำ

นางสาวเปรี๊ยะ วงษ์เสรี 56010771

นางสาวพรรณนิภา จินะแก้ว 56010810

นางสาวเพ็ญพิชชา ภูบุญทอง 56010893


.....
(รศ.ดร.ปราโมทย์ วาดเขียน)

อาจารย์ที่ปรึกษา


.....
(รศ.ดร.จีรสุดา โกษียากรณ์)

อาจารย์ที่ปรึกษาร่วม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการฉบับนี้สำเร็จลุล่วงได้ด้วยดีจากคำปรึกษา แนะนำ ร่วมถึงความเมตตากรุณา จากท่านอาจารย์ รศ.ดร.ปราโมทย์ วาดเขียน และ รศ.ดร.จิรัสตา โกเชียภรณ์ ผู้ให้โอกาส แนวคิด และความรู้ทางวิชาการโดยมิได้หวังผลสิ่งใดตอบแทน คณะผู้จัดทำรู้สึกซาบซึ้งในความอนุเคราะห์ จากท่านอาจารย์ทั้งสองท่าน และใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณคณาจารย์ภาควิชาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกๆ ท่าน ที่ท่านได้ประสิทธิ์ประสาทวิชา ความรู้ต่างๆ ให้กับคณะผู้จัดทำนำมาประยุกต์ใช้กับโครงการฉบับนี้

ขอขอบคุณ พี่ศักรินทร์ สินไชย ที่คอยให้คำปรึกษาในด้านวิชาการความรู้ที่เกี่ยวข้อง กับปริญญาานิพนธ์ฉบับนี้ ตลอดจนคอยให้กำลังใจในการทำงาน เพื่อนๆ พี่ๆ น้องๆ ในภาควิชา วิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง ที่คอยให้กำลังใจและคอยเป็นแรงผลักดันให้โครงการฉบับนี้เกิดขึ้น

นางสาวเปรี๊ยะ วงษ์เสรี
นางสาวพรรณนิภา จินะแก้ว
นางสาวเพ็ญพิชชา ภูบุญทอง
ผู้จัดทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง
ACCESS CONTROLLING SYSTEM USING SPEECH
RECOGNITION

โดย นางสาวเปรี๊ยะ วงษ์เสรี 56010771
นางสาวพรธนิภา จินะแก้ว 56010810
นางสาวเพ็ญพิชชา ภูบุญทอง 56010893

อาจารย์ที่ปรึกษา รศ.ดร.ปราโมทย์ วาดเขียน
รศ.ดร.จิรสุดา โกษิยาภรณ์

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้นำเสนอระบบควบคุมการเข้าถึงโดยให้หลักการรู้จำเสียง โดยแบ่งการควบคุมการเป็นสองส่วนด้วยกันได้แก่ ควบคุมโดยใช้ระบบรู้จำเสียง และระบบระบุลักษณะของวัตถุด้วยคลื่นความถี่วิทยุ (RFID) ในส่วนแรกการควบคุมการเข้าถึงโดยใช้ระบบรู้จำเสียงสามารถทำได้โดยการวิเคราะห์คลื่นเสียงที่ได้รับมาเพื่อสกัดหาคุณลักษณะเด่น และเมื่อได้คุณลักษณะเด่นมาแล้วจะถูกป้อนเข้าสู่โครงข่ายประสาทเทียม (NEURAL NETWORK) เพื่อใช้ในการจำแนกเสียงผู้พูด เมื่อมีเสียงที่ตรงกับเสียงที่อยู่ในระบบจะอนุญาตให้สามารถใช้งานได้ ในอีกส่วนสามารถควบคุมการการเข้าถึงโดยการใช้อาร์เอฟไอดีในการระบุตัวตนในการควบคุมประตูเป็นอีกทางเลือก ซึ่งวิธีการที่นำเสนอมาเพื่ออำนวยความสะดวกต่อการเปิดปิดประตู ให้ใช้งานได้อย่างเหมาะสม และมีทางเลือกที่หลากหลายมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ABSTRACT

This project presents an access controlling system using speech recognition. The controlling technique can be accomplished by using speech recognition and FRID technique. For a speech recognition system, sample voice is analyzed to extracting voice feature by using Mel-frequency cepstral coefficients (MFCCs) technique and then coefficients are used for train in neural network to distinguishing user. When trained voice is input of controlling system, user is allowed to using access controlling system. Alternative method is RFID technique. This project provides convenient choice for users to select as desired.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
กิตติกรรมประกาศ	I
บทคัดย่อ	II
สารบัญ	IV
สารบัญรูป	VIII
สารบัญตาราง	XI
บทที่ 1	
บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	1
1.3 ขอบเขตของปริญญานิพนธ์	1
บทที่ 2	
ทฤษฎีและหลักการที่เกี่ยวข้อง	
2.1 เสียง (Sound)	3
2.1.1 การกำเนิดเสียง	3
2.1.2 แบบจำลองการกำเนิดเสียง	3
2.1.3 การรู้จำเสียง (Speech Recognition)	5
2.2 การสกัดคุณลักษณะของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์ เชปสตรัมบน เมลสเกล (Mel Frequency Cepstral Coefficient: MFCC)	6
2.2.1 พรีเอมฟาซิส (Pre-emphasis)	6
2.2.2 การแบ่งเฟรม (Framing)	6
2.2.3 แฮมมิงวินโดว์ (Hamming window)	6
2.2.4 การแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform: DFT)	7
2.2.5 เมลฟิลเตอร์แบงก์ (Mel filter bank)	8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.2.6 การแปลงโคไซน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform: DCT)	9
2.2.7 การหาสัมประสิทธิ์พลังงานและเดลด้าเซปสตรัม	10
2.4 โครงข่ายประสาทเทียม	12
2.4.1 เซลล์ประสาท (neuron)	12
2.4.2 โครงข่ายประสาทเทียมกับเซลล์ประสาทของมนุษย์	14
2.4.3 การศึกษาและการสร้างโครงข่ายประสาทเทียม	14
2.4.4 การทำงานของโครงข่ายประสาทเทียม	16
2.5 ความไวและความจำเพาะ	18
2.5.1 ความไว (Sensitivity)	18
2.5.2 ความจำเพาะ (Specificity)	18
2.5.3 ค่าทำนายผลบวกหรือค่าพยากรณ์ผลบวก (Positive predictive value)	18
2.5.4 ค่าทำนายผลลบหรือค่าพยากรณ์ผลลบ (Negative predictive value)	18
2.6 อาร์เอฟไอดี	19
2.6.1 ส่วนประกอบของระบบอาร์เอฟไอดี	19
2.6.2 ระบบบัส SPI	21
บทที่ 3 การออกแบบและการจัดทำปริญญาานิพนธ์	
3.1 การออกแบบ	26
3.1.1 การเตรียมสัญญาณสำหรับใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)	25

สารบัญ (ต่อ)

	หน้า
3.1.2 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)	25
3.1.4 โครงข่ายประสาทเทียม	31
3.1.5 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้การรู้จำเสียง	36
3.1.6 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี	37
3.1.7 ส่วนควบคุมประตูแม่เหล็ก	40
3.2 เครื่องมือที่ใช้ในการทดลอง	42
3.2.1 โปรแกรม Matlab	42
3.2.2 ราสเบอร์รี่พาย (Raspberry Pi)	42
3.2.3 เครื่องอ่านอาร์เอฟไอดี MFRC522	43
3.2.4 แท๊ก	43
3.2.5 ออสซิลโลสโคป (Oscilloscope)	44
3.3 การจัดเก็บผลการทดลอง	44
3.3.1 การสกัดคุณลักษณะของสัญญาณเสียงพูด	44
3.3.2 โครงข่ายประสาทเทียม	44
3.3.3 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี	45
3.3.4 ระบบรวม	46
บทที่ 4 ผลการทดลอง	
4.1 กล่าวนำ	47
4.2 การระบุตัวตนโดยใช้การรู้จำเสียง	47
4.2.1 ระบบการจำแนกคำพูด	47
4.2.2 ระบบการจำแนกบุคคล	56
4.3 การควบคุมการเปิด-ปิดประตู โดยระบบอาร์เอฟไอดี	62

สารบัญ (ต่อ)

	หน้า
4.3.1 การทดลองอ่านข้อมูลของตัวแท็ก โดยเครื่องอ่านอาร์เอฟไอดี	62
4.3.2 การติดต่อสื่อสารกันระหว่างเครื่องอ่านอาร์เอฟไอดี กับราสเบอร์รี่พาย	66
4.3.3 การสร้างฐานข้อมูลบนราสเบอร์รี่พาย	68
4.4 ระบบรวม (ระบบรู้จำเสียง และระบบ RFID)	69
บทที่ 5	
สรุปผลและข้อเสนอแนะ	
5.1 สรุปผล	76
5.2 ปัญหาและข้อเสนอแนะ	76
บรรณานุกรม	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1	4
2.2	5
2.3	7
2.4	11
2.5	13
2.6	14
2.7	15
2.8	17
2.9	22
3.1	24
3.2	28
3.3	28
3.4	29
3.5	32
3.6	33
3.7	34
3.8	34
3.9	35
3.10	36
1 และ 2	
3.11	37
3.12	38
3.13	39
3.14	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า	
3.15	กลไกควบคุมประตูแม่เหล็ก	41
3.16	โปรแกรม MATLAB	42
3.17	ราสเบอร์รี่พาย	43
3.18	เครื่องอ่านอาร์เอฟไอดี MFRC522	43
3.19	แท็ก	43
3.20	ออสซิลโลสโคป	44
3.21	การออกแบบการจัดเก็บผลการทดลองของการสกัดคุณลักษณะของสัญญาณเสียงพูด และโครงข่ายประสาทเทียม	45
3.22	การออกแบบการจัดเก็บผลการทดลองของการอ่านค่าแท็ก	45
3.23	บล็อกไดอะแกรมการจัดเก็บผลการทดลอง	46
4.1	การตัดส่วนของสัญญาณที่ไม่ใช่เสียงพูด	48
4.2	แถบความถี่ของเสียงก่อนและหลังทำกระบวนการพรีเอมฟาซิส	49
4.3	เฟรมย่อยของสัญญาณเสียง	50
4.4	สัญญาณที่ถูกทำแฮมมิงวินโดว์	50
4.5	เฟรมย่อยของสัญญาณเสียงในโดเมนเวลา (ซ้าย) และโดเมนความถี่ (ขวา)	51
4.6	เมลฟิลเตอร์แบงก์และสเปคตรัมในแต่ละตัวกรอง	52
4.7	เมลฟิลเตอร์แบงก์และสเปคตรัมในแต่ละตัวกรองทั้งหมด	52
4.8	ค่าสัมประสิทธิ์เซปสตรัมทั้งหมด 26 ค่า	53
4.9	ค่าสัมประสิทธิ์เซปสตรัม 12 ค่า จากตำแหน่งที่ 2 - 13	54
4.10	สัมประสิทธิ์ 26 ค่าจากอัลกอริทึม MFCC	55
4.11	ผลการทดสอบระบบแยกคำพูด เมื่ออินพุตเป็นสัมประสิทธิ์คำว่า 'เปิด' (ด้านซ้าย) และเมื่ออินพุตเป็นสัมประสิทธิ์ของคำอื่นๆ (ด้านขวา)	55
4.12	การแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณเอาร์ทพุตของ SAGC	57
4.13	กรอบสัญญาณของระดับเสียง	58
4.14	การแปลงฟูเรียร์แบบไม่ต่อเนื่องของพิทช์	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.15 การแปลงฟูเรียร์แบบไม่ต่อเนื่องของกรอบสัญญาณของพิทช์	59
4.16 ผลการทดสอบระบบจำแนกบุคคล	60
4.17 การเชื่อมต่อระหว่างคอมพิวเตอร์พกพา และ เครื่องอ่านอาร์เอฟไอดี	63
4.18 สัญญาณข้อมูลเมื่อแตะแท็กที่เครื่องอ่านอาร์เอฟไอดี	63
4.19 รูปแบบการส่งสัญญาณ TTL	64
4.20 การอ่านบล็อกข้อมูล UID ที่ 1 และ 2	64
4.21 การอ่านบล็อกข้อมูล UID ที่ 3 และ 4	65
4.22 ฐานข้อมูลที่มีการบันทึกค่า UID	66
4.23 การเชื่อมต่อระหว่างเครื่องอ่านอาร์เอฟไอดีกับรอสเบอร์รี่พาย	66
4.24 คำสั่งในการกำหนดหมายเลขไอดี ให้สามารถเปิดหลอดแอลอีดีได้	67
4.25 ผลการทดลองการเปิดหลอดแอลอีดีโดยใช้การ์ดทั้งสามใบ	67
4.26 ฐานข้อมูลที่สร้างขึ้น	68
4.27 ตารางสำหรับบันทึกรายชื่อผู้เข้าใช้งานประตู	69
4.28 การติดตั้งแม่เหล็กที่ประตูเพื่อเป็นตัวล็อก	70
4.29 วิธีการใช้งานระบบรู้จำเสียง	70
4.30 การทำงานของวงจรควบคุมการทำงานของแม่เหล็กโดยการสั่งการผ่านระบบรู้จำเสียง	71
4.31 จอ LCD แสดงชื่อผู้สั่งการที่อยู่ในระบบ	71
4.32 ไม่มีไฟเลี้ยงแม่เหล็ก ประตูสามารถเปิดออกได้	72
4.33 จอ LCD ขึ้นข้อความให้สั่งการใหม่	72
4.34 วิธีการใช้งานระบบ RFID	73
4.35 การทำงานของวงจรควบคุมการทำงานของแม่เหล็กโดยการสั่งการผ่านระบบ RFID	74
4.36 จอแสดงชื่อเจ้าของแท็ก RFID	74
4.37 จอแสดงผลสำหรับ RFID ที่ไม่ได้ถูกบันทึกในฐานข้อมูล	75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
4.1 ผลการทดสอบระบบโดยใช้คำ 100 คำ	56
4.2 แสดงประสิทธิภาพของระบบจำแนกคำพูด	56
4.3 ผลการทดสอบระบบโดยใช้เสียง 100 เสียง	61
4.4 แสดงประสิทธิภาพของระบบจำแนกบุคคล	61
4.5 แสดงประสิทธิภาพของการรวมระบบจำแนกคำพูด และระบบจำแนกบุคคล	62



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันวิวัฒนาการของเทคโนโลยี ได้พัฒนาขึ้นอย่างรวดเร็ว ซึ่งเข้ามามีบทบาทกับรูปแบบในการดำเนินชีวิตเป็นอย่างมาก โดยเฉพาะในด้านการอำนวยความสะดวก ด้านความคล่องตัว นอกจากความสะดวกสบายและคล่องตัวแล้ว ยังสามารถช่วยรักษาความปลอดภัยในการเข้าใช้งานอีกด้วย อาทิเช่น การนำลายนิ้วมือ, ใบหน้า, ดวงตา มาใช้ในการสั่งการควบคุมอุปกรณ์ต่างๆ ซึ่งสิ่งที่กล่าวมาข้างต้นนั้น ใช้ประโยชน์ของความเป็นอัตลักษณ์ของแต่ละบุคคล

ในปฏิญานิพนธ์นี้ ได้นำเสนอ การนำเสียงมาใช้ในการระบุตัวตน สำหรับการเข้าถึงระบบเปิด-ปิดประตู ซึ่งเสียงของแต่ละบุคคล มีลักษณะเด่นที่แตกต่างกัน โดยหลักการรู้จำเสียงจะถูกนำมาใช้ในการจดจำลักษณะเด่นของเสียง และเพื่อรองรับการทำงานจากระบบ การเข้าถึงระบบเปิด - ปิดประตูโดยใช้บัตรอาร์เอฟไอดี จะถูกนำมาใช้เป็นอีกทางเลือกหนึ่งของระบบที่นำเสนอด้วย

1.2 วัตถุประสงค์

1.2.1 เพื่ออำนวยความสะดวกต่อการเปิด-ปิดประตู โดยสามารถเลือกวิธีการเข้าถึงด้วยการใช้เสียงหรือใช้บัตรอาร์เอฟไอดี

1.2.2 เพื่อกำหนดตัวบุคคลที่อนุญาตให้เข้า/ออกผ่านระบบได้

1.2.3 เพื่อศึกษาหลักการรู้จำเสียง ในการระบุตัวตน

1.2.4 เพื่อสร้างระบบควบคุมการเปิดประตูโดยใช้การรู้จำเสียง

1.2.5 เพื่อศึกษาการใช้บัตรอาร์เอฟไอดี เป็นทางเลือกในการควบคุมการเปิด-ปิดประตู

1.2.6 เพื่อเก็บบันทึกข้อมูลการใช้งานระบบในฐานะข้อมูล

1.3 ขอบเขตของปฏิญานิพนธ์

ในปฏิญานิพนธ์นี้ได้นำเสนอระบบเปิด-ปิดประตูอัตโนมัติแบบบานพับ แบ่งออกได้เป็น 2 วิธีใหญ่ๆ ได้แก่

1.3.1 การควบคุมการเปิดประตูโดยใช้การรู้จำเสียงพูด (Speech Recognition) มีหลักการการทำงานคือเริ่มจากการวิเคราะห์คลื่นเสียงที่ได้รับมาเพื่อสกัดหาคุณลักษณะเด่น และเมื่อได้คุณลักษณะเด่นมาแล้วจะถูกป้อนเข้าสู่โครงข่ายประสาทเทียม (Neural Networks) เพื่อใช้ในการจำแนกเสียงผู้พูดนั้น ซึ่งในการทำงานแบบเจาะจงผู้พูด ผู้พูดจะต้องลงทะเบียนเสียงของตนกับระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก่อน ระบบจึงจะสามารถรู้จำเสียงของผู้พูดได้ ซึ่งในการใช้งานระบบ หากระบบตรวจสอบได้ว่าเป็นเสียงที่อยู่ในระบบ ระบบจะสั่งให้ประตูเปิดออก

1.3.2 การควบคุมการเปิดประตูโดยใช้อาร์เอฟไอดี เป็นทางเลือก ซึ่งเครื่องอ่านอาร์เอฟไอดี จะปล่อยคลื่นแม่เหล็กไฟฟ้าออกมาตลอดเวลา และคอยตรวจจับว่ามีแท็กในบริเวณสนามแม่เหล็กไฟฟ้าหรือไม่ เมื่อมีแท็กเข้ามาอยู่ในบริเวณสนามแม่เหล็กไฟฟ้า แท็กจะได้รับพลังงานไฟฟ้าที่เกิดจากการเหนี่ยวนำของคลื่นแม่เหล็กไฟฟ้าเพื่อให้แท็กเริ่มทำงาน และจะส่งข้อมูลในหน่วยความจำที่ผ่านการมอดูเลตกับคลื่นพาห์แล้วออกมาทางสายอากาศที่อยู่ภายในแท็ก คลื่นพาห์ที่ถูกส่งออกมาจากแท็ก จะเกิดการเปลี่ยนแปลงแอมพลิจูด เครื่องอ่านจะตรวจจับความเปลี่ยนแปลงของคลื่นพาห์แปลงออกมาเป็นข้อมูลแล้วทำการถอดรหัสเพื่อนำข้อมูลไปตรวจสอบ หากรหัสตรงกันประตูจะถูกสั่งให้เปิดออก โดยระบบนี้สามารถรองรับผู้ใช้งานได้ทั้งหมด 5 คน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1 เสียง (Sound)

เสียงเป็นคลื่นเชิงกลที่เกิดจากการสั่นสะเทือนของวัตถุ เมื่อวัตถุสั่นสะเทือน ก็จะทำให้เกิดการอัดตัวและขยายตัวของคลื่นเสียง และถูกส่งผ่านตัวกลาง เช่น อากาศไปยังหู แต่เสียงสามารถเดินทางผ่านสสารในสถานะก๊าซ ของเหลว และของแข็งได้ แต่ไม่สามารถเดินทางผ่านสุญญากาศได้ เมื่อการสั่นสะเทือนนั้นมาถึงหู มันจะถูกแปลงเป็นพัลส์ประสาท ซึ่งจะถูกส่งไปยังสมองทำให้เรารับรู้และจำแนกเสียงต่างๆ ได้ [1]

2.1.1 การกำเนิดเสียง

เสียงเริ่มเกิดขึ้นเมื่อวัตถุหรือแหล่งกำเนิดเสียงมีการสั่นสะเทือนส่งผลต่อการเคลื่อนที่ของโมเลกุลของอากาศที่อยู่โดยรอบ กล่าวคือโมเลกุลของอากาศเหล่านี้จะเคลื่อนที่จากตำแหน่งเดิมไปชนกับโมเลกุลที่อยู่ถัดไป ก่อให้เกิดการถ่ายโอนโมเมนตัมจากโมเลกุลที่มีการเคลื่อนที่ให้กับโมเลกุลที่อยู่ในสภาวะปกติ จากนั้นโมเลกุลที่ชนกันนี้จะแยกออกจากกันโดยโมเลกุลที่เคลื่อนที่มาจะถูกดึงกลับไปยังตำแหน่งเดิมด้วยแรงปฏิกิริยาและโมเลกุลที่ได้รับการถ่ายโอนพลังงานจะเคลื่อนที่ไปชนกับโมเลกุลที่อยู่ถัดไป ปรากฏการณ์นี้จะเกิดขึ้นสลับกันไปมาได้เมื่อสื่อกลาง (ในที่นี้คืออากาศ) มีคุณสมบัติของความยืดหยุ่น การเคลื่อนที่ของโมเลกุลอากาศนี้จึงเกิดเป็นคลื่นเสียง

2.1.2 แบบจำลองการกำเนิดเสียง

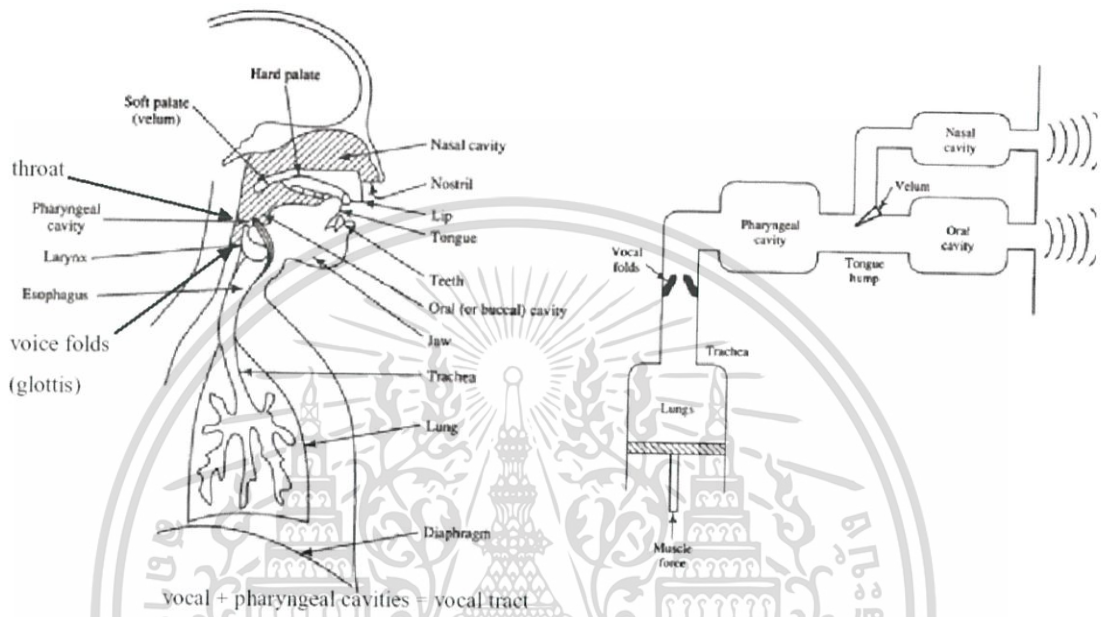
กระบวนการการสร้างเสียงจากมนุษย์เพื่อใช้ในการรู้จำเสียง เกี่ยวข้องกับขั้นตอนดังต่อไปนี้

1. การเปิดและปิดอย่างรวดเร็วของสายเสียง (glottis) เพื่อสร้างการสั่นสะเทือนในอากาศ
2. การกำทอนของช่องคอหอย (pharyngeal cavity) โพรงจมูก (nasal cavity) และช่องปาก (oral cavity)
3. การสั่นสะเทือนของอากาศ
4. การสั่นสะเทือนของกระบอกหู (ear drum) หรือเยื่อแก้วหู (tympanum)
5. การรับรู้ของหูชั้นใน (inner ear)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. การรู้จำโดยสมอง

จากขั้นตอนการกำเนิดเสียง สามารถแสดงบล็อกไดอะแกรมของกลไกการสร้างเสียงของมนุษย์ได้ดังรูปที่ 2.1 [15]



รูปที่ 2.1 บล็อกไดอะแกรมของกลไกการสร้างเสียงของมนุษย์

เนื่องจากความดันของอากาศที่ผลักจากปอด เส้นเสียงสามารถเปิดและปิดได้อย่างรวดเร็วซึ่งสร้างการสั่นในอากาศ การสั่นถูกมอดูเลตโดยการสะท้อนของคอหอย (pharyngeal cavities) โพรงจมูก (nasal cavities) และช่องปาก (oral cavities) ทำให้เกิดคุณภาพของเสียงที่แตกต่างกัน นั่นคือความถี่ของการสั่นสะท้อนของเส้นเสียง กำหนดพิทช์ของเสียง (Pitch) โดยรูปของพิทช์แสดงได้ดังรูปที่ 2.2 ตำแหน่งหรือรูปทรงของริมฝีปาก ลิ้น และจมูก เป็นตัวกำหนดคุณภาพของเสียง และการบีบอัดอากาศจากปอดจะช่วยกำหนดความดังของเสียง [16]

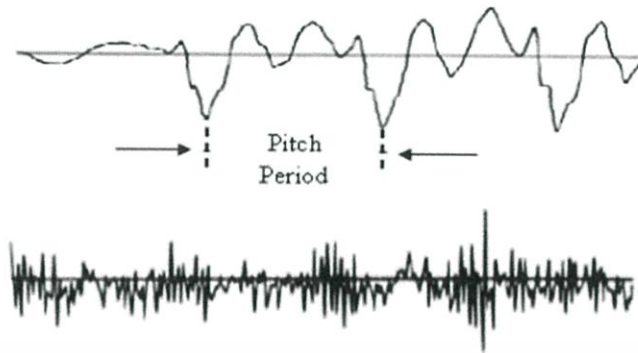
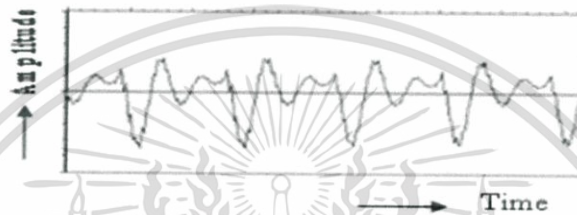


Fig. 5 Examples of Voiced (Top) and Unvoiced (Bottom) Speech Signal



รูปที่ 2.2 ระดับของเสียง (Pitch)

2.1.3 การรู้จำเสียง (Speech Recognition)

การรู้จำเสียง (Speech Recognition) คือการประมวลผลโดยแปลงสัญญาณเสียงพูดให้อยู่ในรูปแบบของคำที่มีการเรียงลำดับด้วยอัลกอริทึมที่ใช้ในโปรแกรมคอมพิวเตอร์เช่น การป้อนเสียงที่เป็นตัวเลขของชั้นบนลิพด์โดยสารอัจฉริยะผ่านไมโครโฟน เมื่อระบบรู้จำเสียงได้รับเสียงเข้าก็จะทำการประมวลผลโดยให้ลิพด์โดยสารชั้นลงตามจำนวนตัวเลขชั้นที่ได้รับมา โดยตัวเลขชั้นต่างๆ จะต้องมีการสอน (Training) ให้ระบบรู้จำก่อน เพื่อให้คอมพิวเตอร์สามารถเรียนรู้เสียงพูดจากคน ซึ่งจะแตกต่างกันไปตามลักษณะของคนเช่น เพศ อายุ แล้วจดจำไว้ใช้ในอนาคต เนื่องจากข้อจำกัดดังกล่าว จึงทำให้ระบบรู้จำเสียงในยุคแรก มีความสามารถในการแยกแยะคำได้อย่างถูกต้องก็ต่อเมื่อผู้พูดมีการเว้นจังหวะคำพูด หรือพูดทีละคำ (Isolated Word) ต่อมาได้มีการพัฒนาระบบรู้จำเสียงให้สามารถรู้จำเสียงพูดอย่างต่อเนื่อง (Continuous speech) ได้ [2]

2.2 การสกัดคุณลักษณะของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนเมลสเกล (Mel Frequency Cepstral Coefficient: MFCC)

การดึงลักษณะเด่นของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนเมลสเกลเป็นวิธีการสกัดคุณลักษณะที่ใช้มากที่สุดในการรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition : ASR) เพื่อดึงเวกเตอร์คุณลักษณะที่มีข้อมูลทั้งหมดที่เกี่ยวกับข้อความทางภาษา MFCC จำลองการรับรู้แบบลอกการิทึมของเสียงดังและระดับเสียงของระบบการฟังของมนุษย์และพยายามจะกำจัดคุณลักษณะที่เป็นผลมาจากผู้พูดยกเว้นความถี่มูลฐานและฮาร์โมนิก นอกจากนี้แสดงลักษณะธรรมชาติของคำพูดแล้ว MFCC ยังรวมถึงการเปลี่ยนแปลงเมื่อเวลาผ่านไปของคุณลักษณะของเสียงอีกด้วย โดยขั้นตอนของการทำ MFCC นั้นประกอบด้วยหลักการต่างๆ ดังนี้

2.2.1 프리เอมฟาซิส (Pre-emphasis)

ฟรีเอมฟาซิสคือการเพิ่มปริมาณของพลังงานที่ความถี่สูงของสัญญาณเสียง ซึ่งโดยทั่วไปสัญญาณเสียงมีปริมาณพลังงานที่ความถี่ต่ำมากกว่าความถี่สูง จึงใช้ตัวกรองความถี่สูงผ่านอันดับที่หนึ่ง แสดงความสัมพันธ์ระหว่างข้อมูลอินพุต $x[n]$ และข้อมูลเอาออก $y[n]$ ได้ดังสมการที่ (2.1) [3]

$$y[n] = x[n] - 0.95x[n-1] \quad (2.1)$$

2.2.2 การแบ่งเฟรม (Framing)

การแบ่งเฟรมเป็นการเตรียมสัญญาณก่อนทำการแอมมิงวินโดว์โดยแบ่งสัญญาณเสียงให้มีขนาดสั้นลง สัญญาณเสียงจะถูกแบ่งออกเป็นเฟรมขนาด N จุดข้อมูล โดยเฟรมถัดไปจะมีข้อมูลเหลื่อมทับกันจำนวน M จุดข้อมูลกับเฟรมก่อนหน้า [4]

2.2.3 แอมมิงวินโดว์ (Hamming window)

แอมมิงวินโดว์ถูกนำไปคูณกับสัญญาณเสียงที่ผ่านการแบ่งเฟรมแล้ว เพื่อให้แต่ละเฟรมย่อยนั้นมีลักษณะเป็นสัญญาณที่ต่อเนื่องในจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยเมื่อนำเฟรมสัญญาณเสียงนี้ไปแปลงฟูเรียร์จะทำให้สเปกตรัมความถี่มูลฐานและความถี่ฮาร์โมนิกของสัญญาณเสียงแสดงออกมาในลักษณะสัญญาณยอดแหลมจำนวนมากที่มีระยะห่างใกล้เคียงกัน สมการของแอมมิงวินโดว์แสดงได้ดังสมการที่ (2.2) และสัญญาณที่ได้จากการทำแอมมิงวินโดว์แสดงได้ดังสมการที่ (2.3)

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2.2)$$

$$y(n) = x(n) \cdot w(n) \quad (2.3)$$

โดยที่ $w(n)$ คือค่าสัมประสิทธิ์ของแฮมมิงวินโดว์
 $x(n)$ คือสัญญาณอินพุตก่อนการทำวินโดว์
 $y(n)$ คือสัญญาณเอาต์พุตที่ได้จากการทำวินโดว์
 N คือขนาดของเฟรมสัญญาณเสียง
 และ n มีค่าตั้งแต่ 0 ถึง $N-1$

โดยตัวอย่างของแฮมมิงวินโดว์สามารถแสดงได้ดังรูปที่ 2.3



รูปที่ 2.3 แฮมมิงวินโดว์

2.2.4 การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform: DFT)

การแปลงฟูรีเยร์แบบไม่ต่อเนื่องได้ถูกนำมาใช้ในการประมาณค่าสเปกตรัมของสัญญาณในรูปของลำดับของตัวเลข โดยมีวิธีการแปลงตามสมการที่ (2.4)

$$X(m) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi nm}{N}} \quad ; \quad m = 0, 1, 2, \dots, N-1 \quad (2.4)$$

โดยที่ $x(n)$ คือสัญญาณในโดเมนเวลา
 $X(m)$ คือค่าการแปลงฟูรีเยร์ของสัญญาณ $x(n)$
 N คือขนาดของเฟรมเสียงที่ผ่านการทำวินโดว์
 และ m มีค่าตั้งแต่ 0 ถึง $N-1$

ซึ่งอาจเรียก $X(m)$ ว่าเป็น DFT ขนาด N จุด (N -point DFT) ของ $x(n)$ เพราะจำนวนจุดที่ใช้ในการคำนวณ DFT นั้นมีผลกับการใช้งาน DFT แต่ปัญหาของการใช้การแปลงฟูเรียร์แบบไม่ต่อเนื่องจากสมการที่ (2.4) จะเห็นได้ว่าต้องคูณจำนวนเชิงซ้อนเป็นสัดส่วนโดยตรงกับ N^2 ซึ่งถ้าจำนวนเฟรมข้อมูล N มีค่ามากจะต้องใช้เวลาในการคำนวณมากขึ้นไปด้วย ดังนั้นในทางปฏิบัติจึงใช้การแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform : FFT) ซึ่งจะช่วยให้การแปลงฟูเรียร์มีประสิทธิภาพและรวดเร็วมากยิ่งขึ้น

2.2.5 เมลฟิลเตอร์แบงก์ (Mel filter bank)

การได้ยินของมนุษย์มีความไวในการได้ยินที่ไม่เท่ากันในแต่ละช่วงความถี่ โดยมีความไวในการได้ยินต่ำในช่วงความถี่สูงที่เหนือกว่า 1000 เฮิรตซ์ ซึ่งในการสร้างแบบจำลองการได้ยินของมนุษย์ระหว่างการสกัดคุณลักษณะนั้นช่วยเพิ่มประสิทธิภาพในการรู้จำเสียง [5]

ในที่นี้ได้อาศัยหลักการของ MFCC โดยนำความถี่ที่ได้จากการทำ DFT เปลี่ยนให้เป็นความถี่บนเมลสเกล การทำเมลสเกลจะเป็นหน่วยการวัดที่เหมาะสมกับระดับการได้ยินของเสียง การแปลงระหว่างความถี่ในหน่วยเฮิรตซ์เป็นความถี่บนเมลสเกลจะให้ค่าความเป็นเชิงเส้นที่ความถี่ต่ำกว่า 1000 เฮิรตซ์และเป็นลอการิทึมที่ความถี่สูงกว่า 1000 เฮิรตซ์ขึ้นไป ความถี่เมลสามารถคำนวณได้จากสมการที่ (2.5) และสามารถคำนวณความถี่ในหน่วยเฮิรตซ์จากความถี่บนเมลสเกลได้จากสมการที่ (2.6)

$$mel(f) = 1125 \ln \left(1 + \frac{f}{700} \right) \quad (2.5)$$

$$f(m) = M^{-1}(m) = 700(\exp(m/1125) - 1) \quad (2.6)$$

โดยที่ $mel(f)$	คือค่าความถี่บนเมลสเกล
f	คือความถี่ในหน่วยเฮิรตซ์
$f(m)$	คือความถี่ในหน่วยเฮิรตซ์
m	คือค่าความถี่บนเมลสเกล

สำหรับการคำนวณ MFCC นั้น เมลฟิลเตอร์แบงก์จะถูกสร้างจากค่าความถี่เมลที่คำนวณได้ (การคำนวณฟิลเตอร์แบงก์จะกล่าวถึงในบทที่ 3) ความกว้างของฟิลเตอร์แบงก์มีความ

เป็นเชิงเส้นที่ความถี่ต่ำกว่า 1000 เฮิรตซ์และเป็นลอการิทึมที่ความถี่สูงกว่า 1000 เฮิรตซ์ ซึ่งสมการในการคำนวณฟิลเตอร์แบงค์แสดงได้ดังสมการที่ (2.7)

$$H_m(f) = \begin{cases} 0 & ; f < f(m-1) \\ \frac{f}{f(m)-f(m-1)} - \frac{f(m-1)}{f(m)-f(m-1)} & ; f(m-1) \leq f \leq f(m) \\ -\frac{f}{f(m+1)-f(m)} + \frac{f(m+1)}{f(m+1)-f(m)} & ; f(m) \leq f \leq f(m+1) \\ 0 & ; f > f(m+1) \end{cases} \quad (2.7)$$

โดยที่ m คือจำนวนของตัวกรองที่เราต้องการ
 f คือความถี่ใด ๆ

2.2.6 การแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform: DCT)

การแปลงโคซายน์แบบไม่ต่อเนื่องใช้สำหรับการแปลงจากค่าพลังงานลอการิทึมเมลสเปคตรัมที่อยู่ในโดเมนความถี่มาเป็นโดเมนเวลา [3,4] และเนื่องจากฟิลเตอร์แบงค์ที่มีการซ้อนทับกันค่าพลังงานที่ได้นั้นก็มีความสัมพันธ์กันอยู่ด้วย จึงใช้การแปลงโคซายน์แบบไม่ต่อเนื่องในการลดความสัมพันธ์และความซับซ้อนของพลังงานลงเพื่อการจำลองรูปแบบของคุณลักษณะได้ [5] ซึ่งค่าสัมประสิทธิ์ที่ได้คือค่าสัมประสิทธิ์เซปสตรัม การแปลงโคซายน์แบบไม่ต่อเนื่องแสดงได้ดังสมการที่ (2.8)

$$C(k) = w(k) \sum_{n=1}^N e(n) \cos\left(\frac{\pi}{2N} (2n-1)(k-1)\right), \quad k=1,2,\dots,N \quad (2.8)$$

$$w(k) = \begin{cases} \frac{1}{\sqrt{2}}, & k=1 \\ \sqrt{\frac{2}{N}}, & 2 \leq k \leq N \end{cases}$$

โดยที่ $C(k)$ คือค่าสัมประสิทธิ์เซปสตรัม

$e(n)$ คือค่าพลังงานลอการิทึมเมลสเปคตรัม

N คือจำนวนของค่าพลังงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.7 การหาสัมประสิทธิ์พลังงานและเดลด้าเซปสตรีม

เนื่องจากพลังงานมีความสัมพันธ์กับอัตลักษณ์ของเสียงและเพื่อเป็นประโยชน์ในการตรวจสอบอัตลักษณ์ พลังงานในแต่ละเฟรมซึ่งเป็นผลรวมของค่ากำลังงานของจุดข้อมูลของสัญญาณเสียงในแต่ละเฟรมที่ผ่านกระบวนการทำวินโดว์จะถูกนำมาใช้เป็นหนึ่งในลักษณะเด่นของเสียง โดยพลังงานในแต่ละเฟรมสามารถหาได้ดังสมการที่ (2.9)

$$Energy = \sum_{n=n_1}^{n_2} x^2[n] \quad (2.9)$$

โดยที่ $x[n]$ คือสัญญาณเสียงที่ผ่านการทำวินโดว์

การทำเดลด้าเซปสตรีมเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรีมสำหรับค่าการเปลี่ยนแปลงระหว่างเฟรม ซึ่งค่าสัมประสิทธิ์เซปสตรีมของแต่ละเฟรมเป็นค่าของสัมประสิทธิ์หนึ่งชุด โดยการคำนวณหาค่าสัมประสิทธิ์เดลด้าเซปสตรีม สามารถคำนวณได้โดยใช้ความสัมพันธ์ดังแสดงในสมการที่ (2.10)

$$D_{f_n} = \frac{C_{f_{n+1}} - C_{f_{n-1}}}{2} \quad (2.10)$$

$$C_{f_n} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}$$

โดยที่ D_{f_n} คือชุดสัมประสิทธิ์เดลด้าเซปสตรีมของเฟรมที่ n
 C_{f_n} คือชุดสัมประสิทธิ์เซปสตรีมของเฟรมที่ n
 N คือจำนวนของสัมประสิทธิ์เซปสตรีม

การหาค่าสัมประสิทธิ์เดลด้าพลังงานเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์พลังงานระหว่างเฟรม โดยการคำนวณหาค่าสัมประสิทธิ์เดลด้าพลังงานสามารถคำนวณได้โดยใช้ความสัมพันธ์ดังแสดงในสมการที่ (2.11)

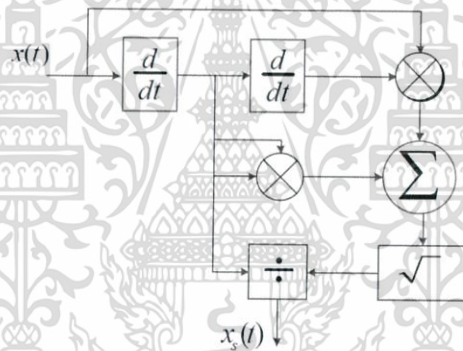
$$dE_{f_n} = \frac{E_{f_{n+1}} - E_{f_{n-1}}}{2} \quad (2.11)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ dE_{f_n} คือค่าสัมประสิทธิ์เดลต้าพลังงานของเฟรมที่ n
 E_{f_n} คือค่าสัมประสิทธิ์พลังงานของเฟรมที่ n

2.3 การสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้ในการระบุบุคคลโดยอาศัยวิธีการควบคุมอัตราขยายสัญญาณคลื่นรูปไซน์อัตโนมัติ (Sinusoidal Automatic Gain Control: SAGC)

วิธีการควบคุมอัตราขยายสัญญาณคลื่นรูปไซน์อัตโนมัติเป็นการปรับแอมพลิจูดสัญญาณอินพุตรูปไซน์ให้มีแอมพลิจูดสัญญาณเอาต์พุตที่มีความเหมาะสม การประยุกต์ใช้งานนิยมใช้ในการปรับสัญญาณของเครื่องรับวิทยุ อันเนื่องมาจากความแตกต่างของความแรงของสัญญาณที่ได้รับโดยปรับขนาดของสัญญาณที่ส่งมาจากต่างสถานีให้มีความเหมาะสม จะเห็นได้ว่าการปรับขนาดของสัญญาณเสียงเป็นประโยชน์ในการปรับความเหมาะสมของสัญญาณได้อย่างมีประสิทธิภาพ โดยบล็อกไดอะแกรมของ SAGC แสดงได้ดังรูปที่ 2.4



รูปที่ 2.4 บล็อกไดอะแกรมของ SAGC

โดยที่ $x(t)$ คือสัญญาณอินพุตรูปไซน์และเอาต์พุตสุดท้ายของระบบคือ $x_s(t)$ ซึ่งในการเปรียบเทียบกับสัญญาณอินพุตรูปไซน์ สัญญาณเอาต์พุตมีขนาดหนึ่งหน่วยและเฟสเลื่อน 90 องศาสำหรับอินพุตของระบบ ถ้าอินพุตเป็นสัญญาณรูปไซน์ สัญญาณที่ได้จากขั้นตอนการหาค่ารากที่สองนั้นคือกรอบสัญญาณของพิทช์ (Pitch envelope) ของเสียงพูด ซึ่งในปริภูมิตวินนอร์มนี้ได้นำค่าเอาต์พุตของระบบ ($x_s(t)$) และค่าเอาต์พุตจากขั้นตอนการหาค่ารากที่สอง ไปใช้ในการสกัดลักษณะเด่นของเสียงต่อไป สมการของขั้นตอนต่างๆ แสดงได้ดังนี้

สัญญาณอินพุตรูปไซน์

$$x(t) = KA \sin[\omega t + \phi] \quad (2.12)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลของการทำอนุพันธ์อันดับหนึ่งและอนุพันธ์อันดับสองของ $x(t)$ ถูกแสดงตามลำดับ

$$x'(t) = \frac{dx(t)}{dt} = KA\omega \cos(\omega t + \varphi) \quad (2.13)$$

$$x''(t) = \frac{d^2x(t)}{dt^2} = -KA\omega^2 \sin(\omega t + \varphi) \quad (2.14)$$

โดยยกกำลังสองทั้งสองข้างของสมการ (2.13) ผลลัพธ์คือ

$$(x'(t))^2 = (KA\omega)^2 \cos^2(\omega t + \varphi) \quad (2.15)$$

นำสมการ (2.14) คูณด้วยสมการ (2.12) ผลลัพธ์คือ

$$x(t)x''(t) = -(KA\omega)^2 \sin^2(\omega t + \varphi) \quad (2.16)$$

นำสมการ (2.15) ลบด้วยสมการ (2.16) ได้ผลลัพธ์คือ

$$(x'(t))^2 - x(t)x''(t) = (KA\omega)^2 \quad (2.17)$$

ทำการถอดรากที่สองของสมการ (2.17) ได้เป็น

$$\sqrt{(x'(t))^2 - x(t)x''(t)} = |KA\omega| \quad (2.18)$$

หลังจากนั้นหารสมการ (2.13) ด้วยสมการ (2.18) ผลลัพธ์สุดท้ายคือ

$$x_s(t) = \cos(\omega_0 t + \varphi) \quad (2.19)$$

2.4 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ โครงข่ายประสาทเทียมมีวัตถุประสงค์ที่จะสร้างเครื่องมือที่มีความสามารถในการเรียนรู้การจดจำแบบรูป และการทำนายอนาคต เช่นเดียวกับความสามารถที่มีในสมองมนุษย์ [6]

2.4.1 เซลล์ประสาท (neuron)

เซลล์ประสาทเป็นส่วนที่เล็กที่สุดของระบบประสาท เซลล์ประสาทหนึ่งเซลล์มีส่วนประกอบที่สำคัญอยู่ 4 ส่วน ดังแสดงในรูปที่ 2.5 [7] ดังนี้

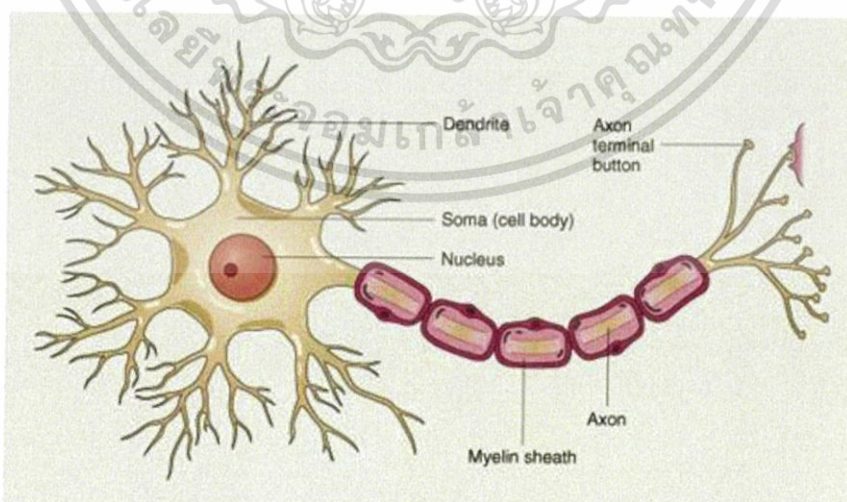
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.1.1 ตัวเซลล์ (cell body) เป็นจุดศูนย์กลางของเซลล์ประสาท ประกอบด้วยนิวเคลียส (nucleus) อยู่ตรงกลางเซลล์ ล้อมรอบด้วยของเหลวที่เรียกว่าไซโตพลาส (cytoplasm) มีผนังเซลล์ทำหน้าที่เป็นผนังห่อหุ้มเซลล์

2.4.1.2 เดนไดรต์ (dendrite) เป็นเส้นใยที่ยื่นออกจากตัวเซลล์มีหน้าที่รับความรู้สึกมีกิ่งก้านสาขาเป็นแขนงสั้น ๆ มีลักษณะคล้ายรากแขนงของต้นไม้

2.4.1.3 แอกซอน (axon) เป็นเส้นใยเดี่ยว ๆ ที่ยื่นออกจากตัวเซลล์ ทำหน้าที่ส่งความรู้สึกของเซลล์ไปยังเซลล์ประสาทตัวอื่น ๆ แอกซอนมีเปลือกหุ้มเรียกว่า ไมอีลินชีท (myelin sheath) ปลายสุดของแอกซอนเป็นพุ่มต่อกับอวัยวะเรียกเอนด์บรัส (end brust) ใยแอกซอนจะมีความยาวมากเป็นพิเศษ แต่ละเซลล์จะมีเพียงเส้นเดียวเท่านั้น ปลายแขนงย่อยของแอกซอน ทุกแขนงจะมีตุ่มเล็ก ๆ เรียกว่าตุ่มปลายประสาท (terminal buttons) การทำงานของแอกซอนจะเกิดขึ้น เมื่อตัวเซลล์ได้รับกระแสประสาทความรู้สึกจากเดนไดรต์จากนั้นจะส่งกระแสความรู้สึกนั้นไปยังแอกซอน แล้วแอกซอนจะส่งกระแสประสาทความรู้สึกนั้น ต่อไปยังเซลล์ประสาทตัวอื่น ๆ หรือส่งไปยังอวัยวะต่าง ๆ ที่ต้องการให้เกิดความรู้สึก

2.4.1.4 ซิแนปส์ (synaps) เป็นจุดต่อระหว่างใยแอกซอนของเซลล์ประสาทตัวหนึ่งกับเดนไดรต์ของเซลล์ประสาทอีกตัวหนึ่ง โดยที่เมื่อเซลล์ประสาทตัวหนึ่งส่งกระแสประสาทความรู้สึกเข้าสู่แอกซอนจนถึงปลายตุ่มประสาทแล้ว กระแสความรู้สึกนั้นจะถูกส่งเข้าสู่บริเวณซิแนปส์ จากนั้นซิแนปส์จะรับกระแสประสาทและส่งต่อไปยังเดนไดรต์ เพื่อเข้าสู่เซลล์ประสาทอีกตัวหนึ่งทันที ซิแนปส์จึงทำหน้าที่เป็นตัวเชื่อมสัญญาณกระแสประสาทระหว่างเซลล์ประสาทตัวหนึ่งกับเซลล์ประสาทอีกตัวหนึ่ง

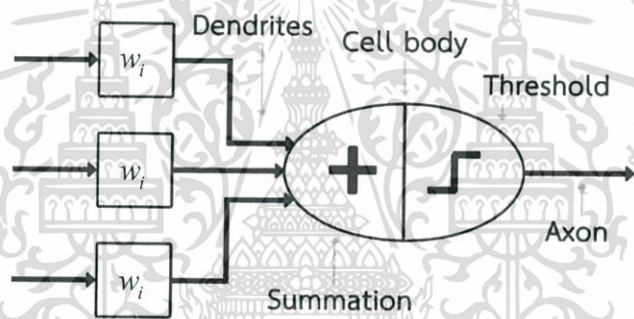


รูปที่ 2.5 ส่วนต่างๆ ของเซลล์ประสาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.2 โครงข่ายประสาทเทียมกับเซลล์ประสาทของมนุษย์

จากการทำงานของเซลล์ประสาทของมนุษย์ ทำให้เกิดโครงข่ายประสาทเทียมที่จำลองการทำงานและหน้าที่ในส่วนต่างๆของเซลล์จนเกิดเป็นโครงข่าย ซึ่งประกอบด้วยปลายการรับกระแสประสาท เดนไดรต์ เปรียบเสมือนเป็นอินพุทของระบบ และปลายการส่งกระแสประสาท เรียกว่า แอกซอน ซึ่งเปรียบเสมือนเป็นเอาต์พุทของระบบ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้า กระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสจะกระตุ้นเซลล์อื่นๆต่อไปผ่านทางแอกซอน ตามโมเดลนี้ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทจนเป็นเครือข่ายที่ทำงานร่วมกัน ซึ่งรูปที่ 2.6 จะเป็นการอธิบายโครงข่ายประสาทเทียมเทียบกับเซลล์ประสาทของมนุษย์ โดยที่ w_i คือค่าถ่วงน้ำหนัก



รูปที่ 2.6 โครงข่ายประสาทเทียมเทียบกับเซลล์ประสาทมนุษย์

2.4.3 การศึกษาและการสร้างโครงข่ายประสาทเทียม

ในการสร้างโครงข่ายประสาทเทียมพื้นฐานมี 3 ส่วนที่จะต้องพิจารณา ได้แก่

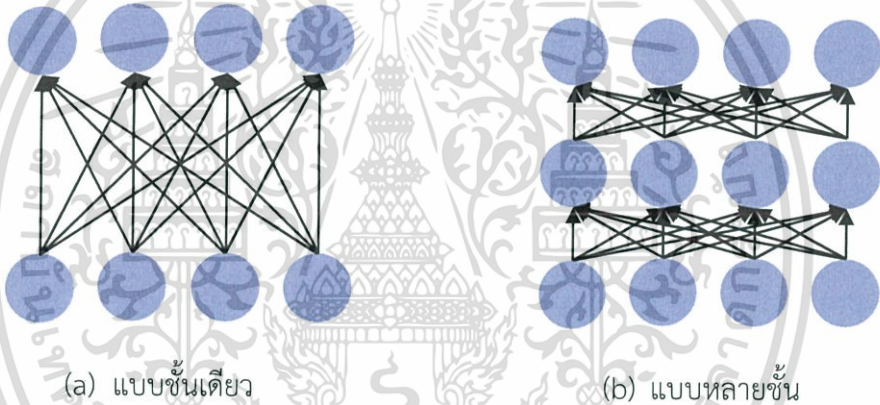
2.4.3.1 สถาปัตยกรรมของโครงข่ายประสาทเทียม (Network Architecture)

การเชื่อมโยงเซลล์ประสาทเทียมจำนวนหนึ่งเข้าด้วยกันเป็นโครงข่ายประสาทเทียมนั้นสามารถเชื่อมโยงแบบใดก็ได้โดยไม่ต้องมีขอบเขตจำกัด อย่างไรก็ตาม ในทางปฏิบัติแล้ว เทคนิคการเรียนรู้ของโครงข่ายประสาทเทียมมักจะถูกออกแบบมาให้ใช้งานได้กับสถาปัตยกรรมโครงข่ายประสาทเทียมที่มีลักษณะเฉพาะเท่านั้น สถาปัตยกรรมโครงข่ายประสาทเทียมที่พบทั่วไปจะมีลักษณะหลักๆ คือ มีการจัดเซลล์ประสาทเทียมเป็นชั้นๆ (layer) ชั้นที่รับข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชั้นอินพุท (input layer) ชั้นที่ผลิตผลตอบของโครงข่ายเรียกว่า ชั้นเอาต์พุท (output layer) ส่วนชั้นอื่นๆ ที่มีส่วนในการช่วยทำการประมวลผลอยู่ภายในเรียกว่า ชั้นซ่อน (hidden layer) ในโครงข่ายประสาทเทียมอาจมีชั้นซ่อนได้หลายชั้นโครงสร้างพื้นฐานจะมีลักษณะเป็นการประกอบกันของรูปแบบ ดังต่อไปนี้

1) แบบป้อนไปข้างหน้า (Feed forward) อาจจัดได้เป็นสองแบบย่อยคือ แบบมีชั้นของเซลล์ประสาทชั้นเดียว (Single feed forward) แสดงดังรูปที่ 2.7(a) และแบบมีชั้นของเซลล์ประสาทหลายชั้น (Multi-Feed forward) แสดงดังรูปที่ 2.7(b) โดยปกติแล้วการเชื่อมโยงจะถูกกำหนดขึ้นระหว่างชั้นที่ติดกัน โดยจะมีการเชื่อมโยงระหว่างเซลล์ประสาทเทียมทุกตัว จากชั้นหนึ่งๆ ไปยังเซลล์ประสาทเทียมทุกตัวในชั้นต่อไป ในบางสถาปัตยกรรมอาจมีการเชื่อมโยงข้ามชั้นก็ได้



รูปที่ 2.7 สถาปัตยกรรมแบบป้อนไปข้างหน้า

2) แบบมีการป้อนไปเวียนกลับ (Recurrent) ในสถาปัตยกรรมบางแบบ โครงข่ายประสาทเทียมอาจมีการเชื่อมโยงที่ถูกกำหนดขึ้นระหว่างเซลล์ประสาทเทียมในชั้นหนึ่งๆ ย้อนกลับไปยังชั้นอื่นๆ ก่อนหน้านั้น หรือแม้แต่ภายในชั้นเดียวกันเอง

ในปริยฐานิพนธ์นี้ได้เลือกใช้สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าหลายชั้น เนื่องจากมีความง่ายในการสร้าง

2.4.3.2 การกำหนดค่าถ่วงน้ำหนัก (Setting weights)

วิธีในการกำหนดค่าถ่วงน้ำหนักให้กับโครงข่ายประสาทเทียมเพื่อทำให้ผลลัพธ์ทางเอาต์พุทตรงกับข้อมูลอินพุทที่ต้องการ ซึ่งกระบวนการในการกำหนดค่าถ่วงน้ำหนักนั้นคือ การเรียนรู้ (Learning) หรือ การสอน (Training) ซึ่งมีอยู่สองประเภทหลักๆ คือ

1) การเรียนรู้แบบมีผู้สอน (Supervised Training) เป็นการสอนที่ต้องมีคู่ข้อมูลของการปรับสอน นั่นคือ ข้อมูลอินพุตและข้อมูลเป้าหมาย โดยค่าเอาต์พุตที่คำนวณได้จากข้อมูลอินพุตจะถูกเปรียบเทียบกับข้อมูลเป้าหมาย โดยกระบวนการปรับสอนจะต้องกำหนดเงื่อนไขการหยุดการปรับสอนว่าจะใช้วิธีใด เช่น การนับจำนวนรอบการปรับสอน หรืออีกวิธีหนึ่งคือการพิจารณาจากความผิดพลาดที่ได้จากการเปรียบเทียบระหว่างค่าเอาต์พุตที่คำนวณได้จากอินพุตกับข้อมูลเป้าหมาย

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Training) เป็นการสอนที่ต้องการข้อมูลอินพุตเพียงอย่างเดียว โดยไม่จำเป็นต้องกำหนดข้อมูลเป้าหมายขึ้นมาก่อน ซึ่งเครือข่ายการเรียนรู้ชนิดนี้จะมีกลไกการทำงานที่เหมือนสมองมนุษย์ กล่าวคือ เครือข่ายมีการเรียนรู้และทำงานไปพร้อมๆ กัน ในเวลาเดียวกัน

ในปฏิญานิพนธ์นี้ได้เลือกใช้กระบวนการสอนแบบมีผู้สอน เนื่องจากมีความง่ายและมีความซับซ้อนน้อยกว่าการสอนระบบแบบไม่มีผู้สอน

2.4.3.3 ฟังก์ชันกระตุ้น (Activation function)

ฟังก์ชันกระตุ้นใช้ในการคำนวณผลตอบสนองทางเอาต์พุต (Output response) ของแต่ละเซลล์ประสาท โดยฟังก์ชันนี้มีทั้งที่ให้ผลตอบสนองเป็นแบบเชิงเส้น และแบบไม่เป็นเชิงเส้น ซึ่งการเลือกใช้งานจะแล้วแต่ความเหมาะสม

2.4.4 การทำงานของโครงข่ายประสาทเทียม

การประมวลผลต่าง ๆ เกิดขึ้นในหน่วยประมวลผลย่อย เรียกว่า โหนด ซึ่งโหนดเป็นการจำลองลักษณะการทำงานมาจากเซลล์การส่งสัญญาณ ระหว่างโหนดที่เชื่อมต่อกัน จำลองมาจากการเชื่อมต่อของใยประสาท และแกนประสาทในระบบประสาทของสมองมนุษย์ ภายในโหนดจะมีฟังก์ชันกำหนดสัญญาณส่งออกที่เรียกว่า ฟังก์ชันผลรวม และฟังก์ชันกระตุ้น ซึ่งทำหน้าที่เปรียบเสมือนกระบวนการทำงานในเซลล์ โครงข่ายประสาทเทียมประกอบด้วย 5 องค์ประกอบดังนี้

- 1) ข้อมูลอินพุต (Input) เป็นข้อมูลที่เป็นตัวเลข หากเป็นข้อมูลเชิงคุณภาพ ต้องแปลงให้อยู่ในรูปเชิงปริมาณที่โครงข่ายประสาทเทียมยอมรับได้
- 2) ข้อมูลเอาต์พุต (Output) คือ ผลลัพธ์ที่เกิดขึ้นจริง จากกระบวนการเรียนรู้ของโครงข่ายประสาทเทียม

3) ค่าน้ำหนัก (Weights) และค่าไบอัส (Biases) คือ สิ่งที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียม หรือเรียกอีกอย่างหนึ่งว่า ค่าความรู้ ค่านี้จะถูกเก็บเป็นทักษะเพื่อใช้ในการจดจำข้อมูลอื่น ๆ ที่อยู่ในรูปแบบเดียวกัน

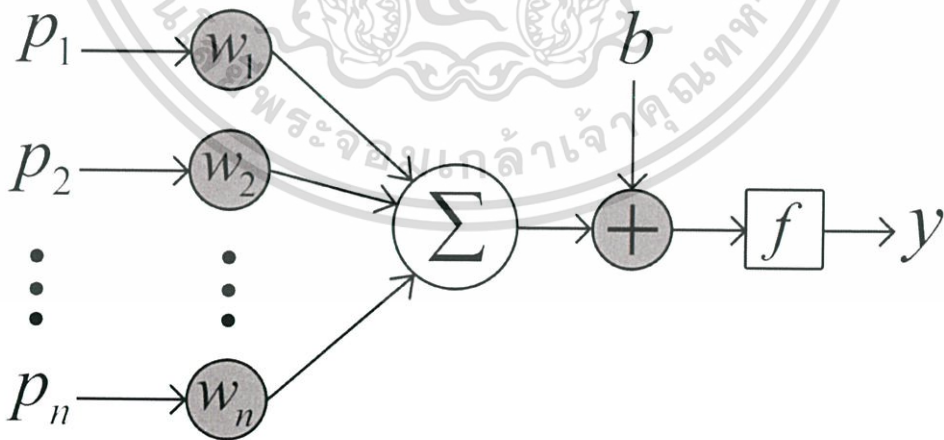
4) ฟังก์ชันผลรวม (Summation function: S) เป็นผลรวมของข้อมูลป้อนเข้า (p_i) ค่าน้ำหนัก (w_i) และค่าไบอัส (b) โดยที่ฟังก์ชันผลรวมสามารถคำนวณได้ตามสมการที่ (2.20)

$$S = \left(\sum_{i=1}^n p_i w_i \right) + b \quad (2.20)$$

5) ฟังก์ชันกระตุ้น (Activation function) ฟังก์ชันกระตุ้นเป็นส่วนที่ทำหน้าที่รวมค่าเชิงตัวเลขจากเอาต์พุตของโครงข่ายประสาทเทียม แล้วทำการตัดสินใจว่าจะส่งสัญญาณเอาต์พุตออกไปในรูปใด ฟังก์ชันการแปลงสามารถเป็นได้ทั้งแบบเชิงเส้นหรือไม่เป็นเชิงเส้น การเลือกใช้ฟังก์ชันการแปลงจะขึ้นอยู่กับลักษณะของระบบ ในปฏิญานิพนธ์นี้ได้เลือกใช้ฟังก์ชันซิกมอยด์แบบเส้นสัมผัสไฮเพอร์โบลาร์ (Hyperbolic tangent sigmoid function) ซึ่งมีสมการตามสมการที่ (2.21) โดยที่ S เป็นค่าที่ได้จากฟังก์ชันผลรวม

$$y = \frac{e^S - e^{-S}}{e^S + e^{-S}} \quad (2.21)$$

โดยภาพรวมการทำงานและการคำนวณของโครงข่ายประสาทเทียมสามารถแสดงได้ดังรูปที่ 2.8



รูปที่ 2.8 การทำงานของระบบโครงข่ายประสาทเทียม

2.5 ความไวและความจำเพาะ

ความไวและความจำเพาะเป็นค่าวัดทางสถิติที่ใช้ประเมินประสิทธิภาพของการทดสอบที่ให้ผลเป็นสองส่วน (เช่น เป็นบวกและลบ) [8]

2.5.1 ความไว (Sensitivity)

ความไวคือสัดส่วนของผลบวกที่เป็นจริงสำหรับภาวะนั้น ๆ (เช่น สัดส่วนของการตรวจพบโรคในผู้ที่ป่วยจริง) มีไวพจน์เป็นคำอื่น ๆ รวมทั้ง อัตราผลบวกจริง (true positive rate) สามารถคำนวณได้ตามสมการที่ (2.22)

2.5.2 ความจำเพาะ (Specificity)

ความจำเพาะคือสัดส่วนของผลลบที่เป็นจริงสำหรับภาวะนั้น ๆ (เช่น สัดส่วนของการตรวจไม่พบโรคในผู้ที่ไม่ป่วย) มีไวพจน์เป็นคำอื่น ๆ รวมทั้ง อัตราผลลบจริง (true negative rate) สามารถคำนวณได้ตามสมการที่ (2.23)

2.5.3 ค่าทำนายผลบวกหรือค่าพยากรณ์ผลบวก (Positive predictive value)

ค่าทำนายผลบวกหรือค่าพยากรณ์ผลบวก คือสัดส่วนของจำนวนผลการตรวจที่เป็นผลบวกแท้ ("การวินิจฉัยถูกต้อง") ต่อจำนวนผลการตรวจที่เป็นผลบวกทั้งหมด เป็นค่าที่มีความสำคัญมากในการประเมินประสิทธิภาพของวิธีหรือเครื่องมือในการวินิจฉัยโรค เนื่องจากเป็นค่าที่บ่งบอกว่าผลบวกที่ได้จากการตรวจนั้นแสดงว่าเป็นโรคจริงๆ ได้ดีเพียงใด ค่าทำนายผลบวกหรือค่าพยากรณ์ผลบวกสามารถคำนวณได้ตามสมการที่ (2.24)

2.5.4 ค่าทำนายผลลบหรือค่าพยากรณ์ผลลบ (Negative predictive value)

ค่าทำนายผลลบหรือค่าพยากรณ์ผลลบ คือสัดส่วนของจำนวนผลการตรวจที่เป็นผลลบ ต่อจำนวนผลการตรวจที่เป็นผลลบทั้งหมด เป็นค่าที่บ่งบอกว่าผลลบที่ได้จากการตรวจนั้นแสดงว่าไม่ได้เป็นโรคจริงๆ ได้ดีเพียงใด ดังนั้นความไวจึงมีประโยชน์ในการวินิจฉัยแยกกับผลลบปลอม (false negative) เพราะว่าการทดสอบยิ่งไวเท่าไร โอกาสการได้ผลลบ (เช่น การพบว่าไม่มีโรค) ที่ไม่เป็นจริง (เช่น บุคคลจริง ๆ มีโรค) ก็น้อยลงเท่านั้น สามารถคำนวณได้ตามสมการที่ (2.25) ดังนั้น ถ้าความไวอยู่ที่ 100% โอกาสได้ผลลบปลอมก็อยู่ที่ 0% และความจำเพาะจึงมีประโยชน์ในการยืนยันภาวะที่มี โดยกันผลบวกปลอม (false positive) เพราะว่าการทดสอบยิ่งจำเพาะเท่าไร โอกาสการได้ผลบวก (เช่น การพบว่ามโรค) ที่ไม่เป็นจริง (เช่น บุคคลจริง ๆ ไม่มีโรค) ก็น้อยลงเท่านั้น และดังนั้น ถ้าความจำเพาะอยู่ที่ 100% โอกาสได้ผลบวกปลอมก็อยู่ที่ 0% [9]

$$\text{Sensitivity} = TP / (TP + FN) \quad (2.22)$$

$$\text{Specificity} = TN / (FP + TN) \quad (2.23)$$

$$\text{Positive predictive} = TP / (TP + FP) \quad (2.24)$$

$$\text{Negative predictive} = TN / (FN + TN) \quad (2.25)$$

โดยที่

TP (True Positive) คือ ผลบวกจริง (พูดเปิดแล้วเปิด)

FP (False Positive) คือ ผลบวกหลวง (พูดคำอื่นแต่เปิด)

TN (True Negative) คือ ผลลบจริง (พูดเปิดแต่ไม่เปิด)

FN (False Negative) คือ ผลลบหลวง (พูดคำอื่นแล้วไม่เปิด)

2.6 อาร์เอฟไอดี

อาร์เอฟไอดี (Radio Frequency Identification) หรือเรียกย่อๆว่า RFID เป็นเทคโนโลยีการระบุข้อมูลที่แสดงเอกลักษณ์ของวัตถุหรือบุคคลด้วยคลื่นความถี่วิทยุที่ได้ถูกพัฒนาขึ้นมาในยุค ค.ศ. 1970 โดยที่อุปกรณ์อาร์เอฟไอดีที่มีการประดิษฐ์ขึ้นใช้งานเป็นครั้งแรกนั้น เป็นผลงานของ Leon Theremin ซึ่งสร้างให้กับรัฐบาลของประเทศรัสเซียในปี ค.ศ. 1945 ซึ่งอุปกรณ์ที่สร้างขึ้นมาในเวลานั้นทำหน้าที่เป็นเครื่องมือดักจับสัญญาณ ไม่ได้ทำหน้าที่เป็นตัวระบุเอกลักษณ์อย่างที่ใช้กันอยู่ในปัจจุบัน เทคโนโลยีอาร์เอฟไอดีมีวัตถุประสงค์เพื่อนำไปใช้ในการบ่งชี้วัตถุในระยะไกลได้ โดยมีจุดเด่นคือสามารถอ่านข้อมูลจากป้าย (Tag) แบบไร้สัมผัส และสามารถอ่านค่าได้แม้ในสภาพที่ทัศนวิสัยไม่ดี ทนต่อความเปียกชื้น แรงสั่นสะเทือน การกระทบกระแทก และสามารถจะอ่านข้อมูลได้ด้วยความเร็วสูง โดยข้อมูลจะถูกเก็บไว้ในไมโครชิปที่อยู่ในป้าย ในปัจจุบันได้มีการนำอาร์เอฟไอดีไปประยุกต์ใช้งานในด้านอื่นๆ ได้แก่ การใช้งานในบัตรชนิดต่างๆ เช่น บัตรประจำตัวพนักงาน (ID card) บัตรโดยสาร บัตรสำหรับผ่านเข้าออกห้องพัก เป็นต้น [10]

2.6.1 ส่วนประกอบของระบบอาร์เอฟไอดี

ในระบบอาร์เอฟไอดี จะมีองค์ประกอบหลักสองส่วน ส่วนแรก คือ ทรานสพอนเดอร์ หรือ แท็ก ที่ใช้ติดกับวัตถุต่างๆที่เราต้องการ โดยป้ายนั้นจะประกอบด้วยสายอากาศและไมโครชิปที่มีการบันทึกหมายเลข (ID) หรือข้อมูลที่เกี่ยวข้องกับวัตถุชิ้นนั้นๆ ส่วนที่สอง คือ เครื่องอ่าน/เขียนข้อมูลภายในแท็ก เรียกว่า Reader ทำหน้าที่เชื่อมต่อเพื่ออ่านหรือเขียนข้อมูลลงในป้ายด้วยสัญญาณความถี่วิทยุ

2.6.1.1 ทรานสปอนเดอร์ (แท็ก)

โครงสร้างภายในของแท็กประกอบด้วย 2 ส่วนหลัก ได้แก่ ส่วนของไมโครชิป ที่ทำหน้าที่เก็บข้อมูลของวัตถุ (ID) และสายอากาศ ทำหน้าที่รับ-ส่ง สัญญาณคลื่นความถี่วิทยุ และสร้างพลังงานป้อนให้ส่วนของไมโครชิป โดยทั่วไปตัวแท็ก จะมีรูปร่างที่แตกต่างกันไป ทั้งนี้ขึ้นอยู่กับวัสดุที่จะนำไปติด และมีหลายรูปแบบ เช่น บัตร เหยี่ยงู กระดุม ฉลากสินค้า หรือ แคลปซูล เป็นต้น ทั้งนี้เราสามารถแบ่งแท็กที่มีใช้งานกันอยู่ได้เป็น 3 ชนิด ได้แก่ แท็กแบบแพสซีฟ แบบกึ่งแพสซีฟ และแบบแอคทีฟ

1) แท็กแบบแพสซีฟ แท็กชนิดนี้ทำงานได้ไม่ต้องอาศัยแหล่งจ่ายไฟภายนอกใด ๆ โดยทั่วไปการทำงานของแท็ก RFID แบบแพสซีฟ ในย่านความถี่ต่ำและสูง (LF และ HF) จะใช้หลักการคู่ควบแบบเหนี่ยวนำ (Inductive coupling) ซึ่งเกิดจากการอยู่ใกล้กับขดลวดจากเครื่องอ่านที่กำลังทำงานและสายอากาศของแท็ก ทำให้เกิดการถ่ายเทพลังงานจากเครื่องอ่านไปยังไมโครชิปในแท็กผ่านสนามแม่เหล็กไฟฟ้าที่เกิดขึ้น เมื่อไมโครชิปได้รับพลังงานก็จะทำงานตามลักษณะเฉพาะของข้อมูลรหัสประจำตัว ปฏิกริยาของไมโครชิปดังกล่าวเครื่องอ่านจะรับรู้ได้ผ่านสนามแม่เหล็กและจะทำการตีความความเป็นข้อมูลดิจิทัลแสดงถึงรหัสประจำตัวที่ส่งมาจากแท็กได้ ลักษณะเงื่อนไขในการทำการเหนี่ยวนำแบบขั้วพาทําให้การอ่านข้อมูลทำได้ไม่ไกลมากนัก โดยทั่วไประยะอ่านสูงสุดจะประมาณ 1 เมตรขึ้นอยู่กับกำลังงานของเครื่องส่งและคลื่นความถี่วิทยุที่ใช้

โดยปกติแท็กชนิดนี้มักมีหน่วยความจำขนาดเล็ก โดยทั่วไปประมาณ 16 ถึง 1,024 ไบต์ มีขนาดเล็กและน้ำหนักเบา ราคาต่อหน่วยต่ำ ไมโครชิปหรือไอซีของป้ายชนิดแพสซีฟที่มีการผลิตออกมาจะมีทั้งขนาดและรูปร่างเป็นไปได้อย่างตั้งแต่แบบแท่ง หรือ แผ่นขนาดเล็กจนแทบไม่สามารถมองเห็นได้ไปจนถึงขนาดใหญ่สะดุดตา ซึ่งต่างก็มีความเหมาะสมกับชนิดการใช้งานที่แตกต่างกัน

ส่วนในระบบความถี่สูงยิ่ง (UHF) แทนที่จะใช้การสร้างสนามแม่เหล็กไฟฟ้าจะใช้คู่ควบแบบแผ่กระจาย (Propagation coupling) โดยที่สายอากาศของเครื่องจ่ายจะทำการส่งพลังงานแม่เหล็กไฟฟ้าในรูปคลื่นวิทยุออกมา ซึ่งเมื่อแท็กได้รับสัญญาณผ่านสายอากาศ แท็กก็จะทำงานโดยการสะท้อนคลื่นที่ได้รับ ซึ่งมีการระบุนอติของแท็กไปยังเครื่องอ่านด้วย (backscattering) ทั้งนี้การทำงานในย่านความถี่ต่างกันจะส่งผลให้ประสิทธิภาพโดยรวมขึ้นกับเงื่อนไขอื่น ๆ เช่น ขนาดของสายอากาศหรือสัญญาณรบกวน เป็นต้น

2) แท็กแบบกึ่งแฟลชชิพ แท็กชนิดนี้จะต้องอาศัยแหล่งจ่ายไฟจากแบตเตอรี่ภายนอก ทำให้สามารถส่งข้อมูลได้ระยะไกลกว่าแท็กแบบแฟลชชิพ โดยตัวแท็กจะรอรับสัญญาณกระตุ้นให้ทำงานจากเครื่องอ่านแล้วจึงจะส่งข้อมูลกลับไปยังเครื่องอ่าน

3) แท็กแบบแอคทีฟ แท็กชนิดนี้จะต้องอาศัยแหล่งจ่ายไฟจากแบตเตอรี่ภายนอก เพื่อจ่ายพลังงานให้กับวงจรภายในทำงาน โดยหลักใหญ่อาจสามารถแบ่งตามหลักการทำงานได้เป็นทรานสพอนเดอร์แบบแอคทีฟ ซึ่งจะทำการส่งข้อมูลออกก็ต่อเมื่อได้รับสัญญาณจากเครื่องอ่านและเครื่องบอกตำแหน่ง หรือบีคอน (beacon) ซึ่งสัญญาณจะถูกปล่อยออกมาเป็นระยะ ๆ ตลอดเวลาการใช้งานของแท็กหรือทรานสพอนเดอร์แบบแอคทีฟนั้นอาจพบได้ในระบบ เช่น ระบบจ่ายเงินในทางด่วน หรือ ด้านตรวจ ขณะที่บีคอนอาจพบได้ในระบบที่ต้องการการบ่งชี้พิกัดแบบเวลาจริง (Real – time locating system, RTLS) เช่น การจัดการการขนส่งสินค้า เป็นต้น โดยแท็กแบบนี้สามารถมีหน่วยความจำภายในขนาดใหญ่ได้ถึง 1 เมกะไบต์และสามารถอ่านได้ในระยะสูงสุดประมาณ 100 เมตร ข้อเสียของแท็กแบบนี้คือ มีราคาต่อหน่วยสูง มีขนาดค่อนข้างใหญ่ และมีอายุการใช้งานที่จำกัดตามอายุของแบตเตอรี่ซึ่งจะมีอายุใช้งานประมาณ 3-7 ปี [10]

2.6.1.2 เครื่องอ่าน/เขียนข้อมูล

หน้าที่ของเครื่องอ่านก็คือ การเชื่อมต่อเพื่ออ่านหรือเขียนข้อมูลลงในป้ายด้วยสัญญาณความถี่วิทยุ ภายในเครื่องอ่านจะประกอบด้วยเสาอากาศที่ทำจากขดลวดทองแดงเพื่อใช้รับส่งสัญญาณ ภาคส่งสัญญาณวิทยุ และวงจรควบคุมการอ่าน-เขียนข้อมูลซึ่งมักจะเป็นวงจรจำพวกไมโครคอนโทรลเลอร์ และส่วนของการติดต่อกับคอมพิวเตอร์ โดยลักษณะ ขนาด และรูปร่างของเครื่องอ่านจะอ่านแตกต่างกันไปตามประเภทของการใช้งาน

2.6.2 ระบบบัส SPI

ระบบบัส SPI ย่อมาจาก Serial Peripheral Interface Bus เป็นรูปแบบหนึ่งของการสื่อสารข้อมูลระหว่างอุปกรณ์แบบดิจิทัลที่พบเห็นได้บ่อย และใช้กับอุปกรณ์ได้มากกว่าสองอุปกรณ์ขึ้นไปและนำมาต่อกันเป็นบัส (Bus) บัส SPI ส่งและรับข้อมูลที่ละบิต (Bit Serial) และใช้สัญญาณนาฬิกาเป็นตัวกำหนดจังหวะการทำงาน ดังนั้นจึงเรียกว่า Synchronous, Bit-Serial Data Communication มีการกำหนดบทบาทในการทำงานของอุปกรณ์ในระบบบัส แบ่งเป็น SPI Master และ SPI Slave โดยที่ SPI Master เป็นฝ่ายเริ่มการสื่อสารข้อมูล และสร้างสัญญาณนาฬิกาที่กำหนดจังหวะการส่งและรับข้อมูล และด้าน SPI Slave จะเป็นฝ่ายคอยตอบสนอง และ

มากำหนดจังหวะการส่งและรับข้อมูล และด้าน SPI Slave จะเป็นฝ่ายคอยตอบสนอง และในระบบ บัส SPI อาจมีอุปกรณ์ที่เป็น SPI Slave ได้มากกว่าหนึ่งอุปกรณ์ (Single-Master, Multi-Slave)

2.6.2.1 ส่วนประกอบของระบบบัส SPI

ระบบ SPI ใช้สัญญาณ 4 เส้น (ใช้งานในแบบที่เรียกว่า 4-Wire SPI) ได้แก่

- 1) SCK (Serial Clock) เป็นสัญญาณนาฬิกาที่ถูกสร้างโดย อุปกรณ์ที่เป็น SPI Master
- 2) MOSI (Master-Out Slave-In) เป็นสัญญาณสำหรับส่งข้อมูล บิตออกจาก SPI Master ไปยัง SPI Slave
- 3) MISO (Master-In Slave-Out) เป็นสัญญาณสำหรับส่งข้อมูล บิตออกจาก SPI Slave ไปยัง SPI Master
- 4) SS (Slave Select, Active-Low) เป็นสัญญาณที่สร้างโดย SPI Master เพื่อใช้ระบุว่า ต้องการสื่อสารกับ SPI Slave หรือไม่ ในกรณีที่มียุกรณ์ SPI Slave มากกว่าหนึ่งชุด จะต้องมียุกรณ์ Slave Select มากกว่าหนึ่งเส้น และแยกสำหรับแต่ละอุปกรณ์ โดยรูปแบบการติดต่อสื่อสารด้วย SPI สามารถแสดงได้ดังรูปที่ 2.9



รูปที่ 2.9 รูปแบบการติดต่อสื่อสารด้วย SPI

2.6.2.2 การทำงานของระบบบัส SPI

เมื่อทำการส่ง-รับข้อมูลผ่านบัส SPI (เรียกว่า SPI Data Transfer) สัญญาณ Slave Select (SS) จะต้องเปลี่ยนจาก HIGH เป็น LOW จากนั้นข้อมูลหนึ่งไบต์จะถูกเลื่อนบิตและส่งออกทีละบิตจาก SPI Master ตามจังหวะของ SCK และเลือกได้ว่าจะให้บิต MSB (Most-Significant Bit) หรือ LSB (Least-Significant Bit) ถูกส่งออกมาก่อน และในขณะเดียวกันก็จะรับข้อมูลที่ทีละบิตจาก SPI Slave จนได้ครบหนึ่งไบต์ (หรือกล่าวได้ว่า Data Frame = 8 บิต) ดังนั้นเมื่อ SPI Master ส่งข้อมูลจำนวนหนึ่งไบต์ไปยัง SPI Slave ก็จะได้ข้อมูลหนึ่งไบต์จาก SPI Slave เช่นกัน

ในช่วงเวลาที่สัญญาณ SS เป็น LOW อาจมีการส่ง-รับข้อมูลได้มากกว่าหนึ่งไบต์ (Multi-byte SPI transfer)

การทำงานของ SPI มี 4 โหมด จำแนกตามพารามิเตอร์สองตัวที่เรียกว่า CPOL (Clock Polarity) และ CPHA (Clock Phase) ซึ่งจะเป็นตัวกำหนดลักษณะการทำงาน อย่างเช่น จะส่ง-รับบิตที่ขอบขาขึ้นหรือลงของสัญญาณ CLK และสัญญาณ CLK จะอยู่ที่ลอจิก HIGH หรือ LOW เมื่อไม่อยู่ในช่วงของการส่งข้อมูลใดๆในบัส SPI (ช่วงที่เรียกว่า Idle) แต่โดยทั่วไป จะใช้ SPI Mode 0 [12]



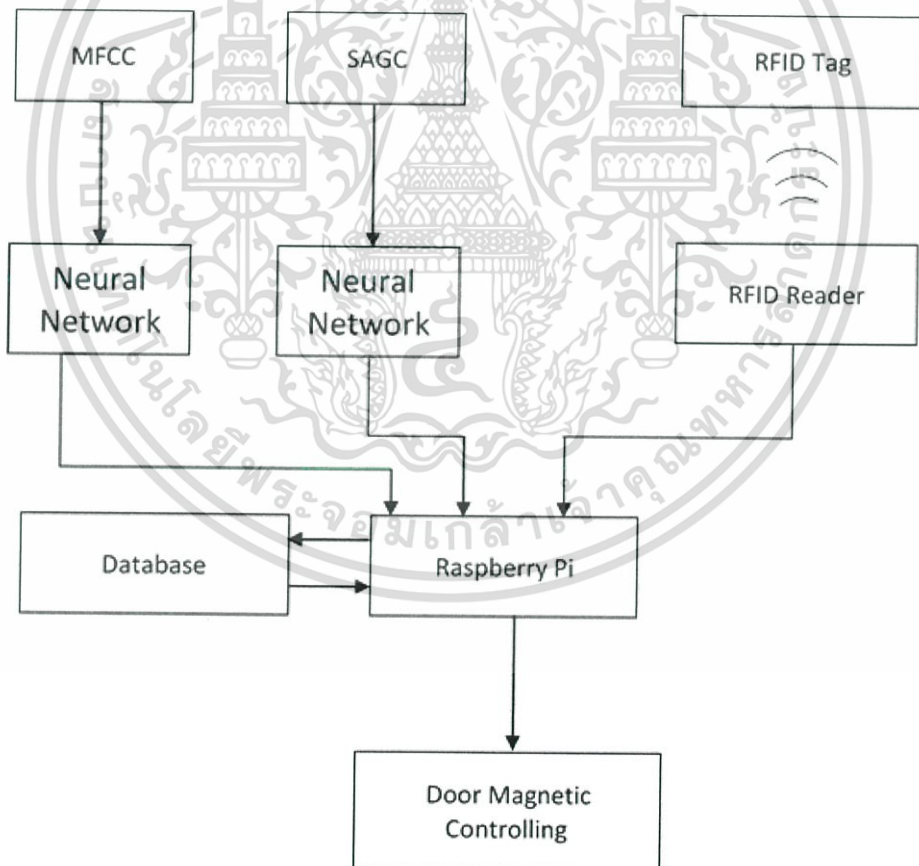
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบและการจัดทำปฏิญานิพนธ์

3.1 การออกแบบ

ปฏิญานิพนธ์นี้ได้นำเสนอระบบรู้จำเสียงเพื่อใช้ในการเข้าถึงการควบคุม ซึ่งในบทนี้จะนำเสนอในส่วนของการออกแบบระบบ ภาพรวมการทำงานของระบบสามารถแสดงได้ดังรูปที่ 3.1 โดยการทำงานของระบบรวมจะแบ่งออกเป็นสองส่วน ส่วนแรกเป็นการรู้จำเสียง (Speech Recognition) ซึ่งประกอบด้วยส่วนของอัลกอริทึม MFCC สำหรับการแยกคำพูด ส่วนของการจำแนกบุคคลจากเสียงพูดโดยอาศัยขั้นตอนของ SAGC และโครงข่ายประสาทเทียม ในส่วนที่สองจะเป็นการทำงานของระบบ RFID ซึ่งสามารถเลือกใช้เป็นทางเลือกในกรณีที่ระบบรู้จำเสียงไม่สามารถใช้งานได้



รูปที่ 3.1 บล็อกไดอะแกรมของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของระบบในส่วนแรกจะเริ่มจากการนำเสียงมาผ่านระบบ Mel Frequency Cepstral Coefficient (MFCC) เพื่อสกัดหาค่าคุณลักษณะเด่นออกมาเป็นค่าสัมประสิทธิ์ เมื่อจบขั้นตอนจะนำค่าสัมประสิทธิ์ที่ได้ไปใช้เป็นอินพุตของระบบโครงข่ายประสาทเทียมระบบที่ 1 และเสียงจะถูกทำ SAGC เพื่อหาค่าสัมประสิทธิ์และนำไปใช้เป็นอินพุตของระบบโครงข่ายประสาทเทียมระบบที่ 2 เช่นกัน เอาท์พุทที่ได้จากโครงข่ายประสาทเทียมจะนำไปเปรียบเทียบกับฐานข้อมูล หากเป็นค่าที่ตรงกันจะส่งเอาท์พุทไปยังระบบควบคุมแม่เหล็กประตู่ให้ทำการเปิด ในส่วนการทำงานของระบบที่สอง ตัวอ่านอาร์เอฟไอดีจะปล่อยสัญญาณออกมาเพื่อตรวจสอบบริเวณนั้นว่ามีแท็กเข้ามาหรือไม่ หากพบแท็ก จะนำค่าข้อมูลในแท็กเข้าไปตรวจสอบกับฐานข้อมูล หากเป็นค่าที่ตรงกันจะส่งเอาท์พุทไปที่ระบบควบคุมแม่เหล็กประตู่ให้ทำการเปิดเช่นเดียวกัน

3.1.1 การเตรียมสัญญาณสำหรับใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)

การบันทึกเสียงพูดเป็นการบันทึกเสียงในโปรแกรม Matlab โดยพูดผ่านไมโครโฟน และทำการบันทึกเสียงพูดยาว 2 วินาที ความถี่สุ่ม 8000 เฮิร์ตซ์ในโปรแกรม Matlab โดยใช้คำสั่ง audiorecorder ซึ่งในความยาว 2 วินาทีนี้ประกอบไปด้วยส่วนที่เป็นเสียงพูดและส่วนที่ไม่ใช่เสียงพูด ดังนั้นจึงทำการตัดส่วนที่ไม่ใช่เสียงพูดออกด้วยวิธีการกำหนดขอบเขตที่รับได้ (Thresholding)

3.1.2 การสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้แยกคำพูดโดยใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)

3.1.2.1 프리เอมฟาซิส (pre-emphasis) ของสัญญาณเสียงพูด

นำสัญญาณเสียงที่ตัดส่วนที่ไม่ใช่เสียงพูดออกมาทำฟรีเอมฟาซิส เพื่อเพิ่มระดับพลังงานที่ความถี่สูงด้วยการใช้ตัวกรองความถี่สูงผ่านอันดับที่หนึ่ง แสดงความสัมพันธ์ดังสมการที่ 2.1 โดยใช้คำสั่ง filter ในโปรแกรม Matlab สามารถแสดงโปรแกรมที่เขียนได้ดังนี้

```
b = [1 -0.95]; a = 1;
```

% a คือ zero b คือ pole

```
preem = filter(b,a,y);
```

โดยที่ y คือสัญญาณเสียง

และ preem คือสัญญาณที่ได้จากการกรอง

3.1.2.2 การแบ่งเฟรมเสียงพูด

ขั้นตอนการแบ่งเฟรมจะนำเสียงที่ผ่านการทำพรีเอมฟาซีสมาแบ่งเป็นเฟรมย่อย โดยเสียงพูดที่ถูกตัดมานั้นจะมีความยาวที่ไม่เท่ากัน ซึ่งก่อนทำการแบ่งเฟรมสัญญาณเสียงจะผ่านการทำ zero-pad เพื่อปรับให้สัญญาณเสียงแต่ละสัญญาณมีความยาวเท่ากันที่ความยาว 5600 จุดข้อมูลหรือ 0.7 วินาที เนื่องจากสัญญาณเสียงแต่ละสัญญาณที่ถูกตัดมานั้นจะมีความยาวที่ไม่เกิน 0.7 วินาที เมื่อทำการ zero-pad สัญญาณเสียงเรียบร้อยแล้ว เสียงที่ได้จะถูกนำมาแบ่งเฟรมให้เป็นเฟรมย่อยที่มีขนาด 256 จุดข้อมูลและข้อมูลของเฟรมต่อไปเลื่อนจากเฟรมก่อนหน้า 100 จุดข้อมูล ซึ่งสัญญาณเสียงพูดหนึ่งสัญญาณจะถูกแบ่งเป็นเฟรมย่อยได้ 54 เฟรม โดยขั้นตอนนี้ทำการแบ่งเฟรมที่ 256 จุดข้อมูลเพื่อให้เหมาะสมกับขั้นตอนการแปลงฟูรีเยร์แบบไม่ต่อเนื่อง [3,4]

3.1.2.3 การทำแฮมมิงวินโดว์

ขั้นตอนการทำแฮมมิงวินโดว์เป็นการนำแต่ละเฟรมย่อยของสัญญาณเสียงพูดที่มากคูณกับสัมประสิทธิ์แฮมมิงวินโดว์ดังสมการใน 2.2 โดยใช้คำสั่ง hamming ในโปรแกรม Matlab สามารถแสดงโปรแกรมที่เขียนได้ดังนี้

```
w = hamming(L);
```

```
ham = k(1,:).*w';
```

โดยที่ L คือขนาดของวินโดว์

W คือแฮมมิงวินโดว์

และ ham คือสัญญาณได้จากการทำแฮมมิงวินโดว์

3.1.2.4 การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง

ขั้นตอนการแปลงฟูรีเยร์แบบไม่ต่อเนื่องจะนำเฟรมสัญญาณที่ผ่านการทำวินโดว์มาแปลงจากโดเมนเวลาเป็นโดเมนความถี่ โดยใช้คำสั่ง fft ในโปรแกรม Matlab โดยมีความยาวขนาด 256 จุด เพื่อทำการประมาณหาค่าสเปกตรัมของสัญญาณเสียงแต่ละเฟรมเพื่อนำไปเข้ากระบวนการขั้นถัดไปคือการทำเมลฟิเตอร์แบงค์ซึ่งเป็นการคำนวณสัญญาณในแกนความถี่ซึ่งสามารถแสดงโปรแกรมที่เขียนได้ดังนี้

```
y = fft(x,n);
```

โดยที่ n คือขนาดของการแปลงฟูรีเยร์

x คือสัญญาณอินพุตในโดเมนเวลา

และ y คือสัญญาณเอาต์พุตจากการแปลงฟูรีเยร์

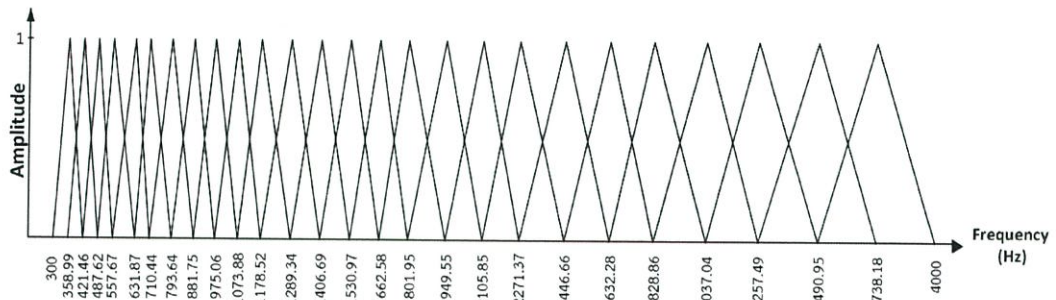
3.1.2.5 การทำเมลฟิลเตอร์แบงค์

การทำเมลฟิลเตอร์แบงค์จะเริ่มจากการออกแบบตัวกรองรูปสามเหลี่ยมซึ่งเป็นตัวกรองแบบความถี่ผ่าน ในที่นี้ใช้ตัวกรองสามเหลี่ยม 26 ตัวกรอง ซึ่งในการออกแบบความกว้างย่านความถี่ของตัวกรองแต่ละตัวกรองนั้นจะถูกคำนวณในโดเมนความถี่บนเมลสเกล โดยการคำนวณเมลฟิลเตอร์แบงค์สามารถทำได้ตามขั้นตอนต่อไปนี้ [5]

1) เลือกความถี่ล่างและความถี่บนซึ่งโดยปกติค่าความถี่ล่างจะใช้ค่าความถี่ 300 เฮิรตซ์และจากการบันทึกเสียงโดยใช้ความถี่สุ่ม 8000 เฮิรตซ์ ทำให้ความถี่บนถูกจำกัดที่ความถี่ 4000 เฮิรตซ์ เมื่อได้ค่าความถี่ล่างและความถี่บนแล้วก็ทำการแปลงความถี่นี้เป็นความถี่บนเมลสเกล โดยการแปลงความถี่ในหน่วยเฮิรตซ์เป็นความถี่บนเมลสเกลทำได้ดังสมการที่ 2.5 ได้เป็นค่าความถี่ล่างและความถี่บนบนเมลสเกลคือ 401.25 และ 2142.26 ตามลำดับ ซึ่งในปริภูมิตวินนอร์มัลใช้ตัวกรอง 26 ตัวกรอง จึงต้องคำนวณให้ได้ค่าความถี่รวมทั้งหมด 28 ค่า ในช่วงความถี่ 401.25 – 2142.26 โดยนำช่วงความถี่สุทธิตหารด้วยจำนวนแบนด์ที่ต้องการ แล้วนำค่าที่หารได้มาบวกกับค่าความถี่เริ่มต้น เป็นค่าความถี่ 401.25, 465.74, 530.22, 594.70, 659.18, 723.66, 788.14, 852.63, 917.11, 981.59, 1046.07, 1110.55, 1175.04, 1239.52, 1304.00, 1368.48, 1432.96, 1497.44, 1561.93, 1626.41, 1690.89, 1755.37, 1819.85, 1884.34, 1948.82, 2013.30, 2077.78 และ 2142.26

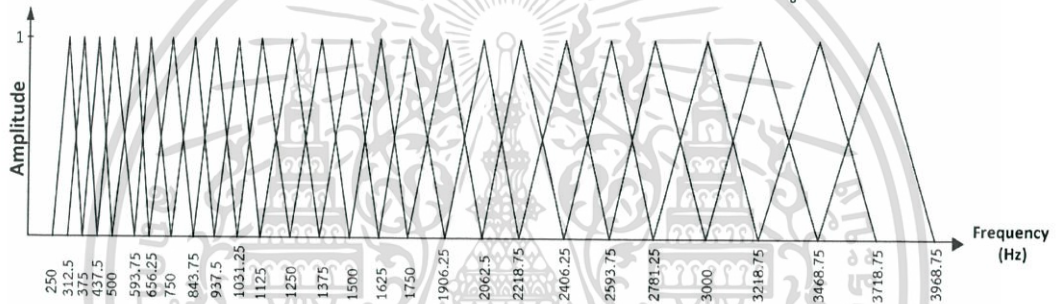
2) แปลงค่าความถี่บนเมลสเกลทั้ง 28 ค่ากลับไปเป็นความถี่ในหน่วยเฮิรตซ์โดยสามารถคำนวณได้จากสมการที่ 2.6 ได้เป็นความถี่ในหน่วยเฮิรตซ์คือ 300, 358.99, 421.46, 487.62, 557.67, 631.87, 710.44, 793.64, 881.75, 975.06, 1073.88, 1178.52, 1289.34, 1406.69, 1530.97, 1662.58, 1801.95, 1949.55, 2105.85, 2271.37, 2446.66, 2632.28, 2828.86, 3037.04, 3257.49, 3490.95, 3738.18 และ 4000 เฮิรตซ์ตามลำดับ

3) ขั้นตอนของการสร้างตัวกรองสามเหลี่ยม จากค่าความถี่ที่คำนวณได้จากข้อที่ 2) จะสร้างตัวกรองสามเหลี่ยมตัวแรกโดยเริ่มจากค่าความถี่ต่ำสุดโดยมียอดสูงสุด (ความถี่กึ่งกลางของแบนด์) ที่จุดความถี่ลำดับที่สองจากขั้นตอนที่ 2) และลดลงมาเป็นศูนย์ที่ความถี่ลำดับที่สาม ตัวกรองตัวที่สองจะเริ่มจากค่าความถี่ลำดับที่สอง มีค่าขั้นสูงสุดที่ความถี่ลำดับที่สามและมีค่าเป็นศูนย์ที่ความถี่ลำดับที่สี่ และตัวกรองลำดับถัดไปจะมีรูปแบบความสัมพันธ์แบบนี้ไปเรื่อยๆ และสามารถแสดงความสัมพันธ์ของตัวกรองทั้งหมดได้ด้วยสมการที่ 2.7 โดยตัวกรองสามเหลี่ยมทั้งหมด 26 ตัวกรอง แสดงได้ดังรูปที่ 3.2



รูปที่ 3.2 ตัวกรองสามเหลี่ยมในความถี่แอนะล็อก

แต่เนื่องจากนำมาประมวลผลบนคอมพิวเตอร์โดยใช้โปรแกรม Matlab ซึ่งความถี่ที่ใช้เป็นความถี่ดิจิทัล ความถี่ที่คำนวณได้จากสมการ 2.7 จะถูกนำไปประมาณค่าตำแหน่งที่มีค่าความถี่ใกล้เคียงกับความถี่ที่คำนวณได้ และนำค่าความถี่นั้นมาสร้างฟิลเตอร์แบงค์แสดงได้ดังรูปที่ 3.3



รูปที่ 3.3 ตัวกรองสามเหลี่ยมในความถี่ดิจิทัล

ซึ่งชุดของตัวกรองสามเหลี่ยมที่ได้นี้คือฟิลเตอร์แบงค์ ฟิลเตอร์แบงค์นี้จะถูกนำไปคูณกับค่าสเปกตรัมของแต่ละเฟรมเสียงที่ได้จากขั้นตอนการแปลงฟูเรียร์แบบไม่ต่อเนื่อง ภายในช่วงความถี่ของแต่ละตัวกรองนั้นจะได้ค่าสเปกตรัมที่ความถี่ต่างๆ ซึ่งสเปกตรัมความถี่นั้นๆ จะถูกรวมเป็นสเปกตรัมความถี่เพียงค่าเดียวและทำการหาค่าพลังงานที่มีอยู่ในแต่ละตัวกรอง จากขั้นตอนนี้จะทำให้ได้ค่าพลังงานออกมาทั้งหมด 26 ค่า ค่าพลังงานนี้จะถูกทำลอการิทึมและนำไปใช้ในขั้นตอนการแปลงโคซายน์แบบไม่ต่อเนื่องต่อไป [14]

3.1.2.6 การแปลงโคซายน์แบบไม่ต่อเนื่อง

ค่าการแปลงโคซายน์แบบไม่ต่อเนื่องจะถูกนำมาใช้กับค่าพลังงานที่ได้จากขั้นตอนการใช้เมลฟิลเตอร์แบงค์ใช้สำหรับการแปลงจากค่าที่อยู่ของลอการิทึมที่อยู่ในโดเมนความถี่มาเป็นโดเมนเวลา โดยใช้คำสั่ง `dct` ในโปรแกรม Matlab แสดงได้ดังนี้ $mcep = dct(cep, 26)$ โดย `cep` คือค่าลอการิทึมพลังงาน และ `mcep` คือค่าสัมประสิทธิ์เซปสตรัม

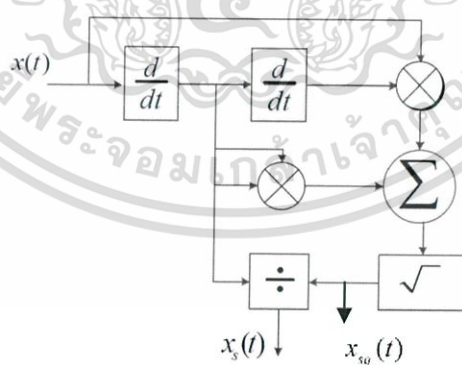
โดยค่าสัมประสิทธิ์เซปสตรีมที่ได้มีทั้งหมด 26 ค่า แต่ค่าสัมประสิทธิ์ลำดับที่ 2-13 เท่านั้นที่ถูกนำมาใช้ จากขั้นตอนนี้ทำให้ได้ค่าสัมประสิทธิ์ทั้งหมด 12 ค่า

3.1.2.7 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรีม

การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรีมเป็นการคำนวณเพื่อเพิ่มค่าสัมประสิทธิ์ โดยค่าสัมประสิทธิ์พลังงานนั้นได้จากการหาค่าพลังงานจากเฟรมเสียงย่อยที่ผ่านกระบวนการแสมมิงวินโดว์ ได้เป็นค่าสัมประสิทธิ์อีกหนึ่งค่าออกมา ส่วนการหาสัมประสิทธิ์เดลต้าเซปสตรีมหาจากอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรีมระหว่างเฟรม จากการคำนวณนี้จะได้ค่าสัมประสิทธิ์อีก 12 ค่า และค่าสัมประสิทธิ์ตัวสุดท้าย สามารถหาได้จากการคำนวณอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์พลังงานระหว่างเฟรมซึ่งเรียกว่าเดลต้าพลังงาน จากขั้นตอนทั้งหมดจะได้ค่าสัมประสิทธิ์จากการสกัดลักษณะของเฟรมเสียงย่อยทั้งหมด 26 ค่า (26 MFCCs) คือ ค่าสัมประสิทธิ์เซปสตรีม 12 ค่า ค่าสัมประสิทธิ์พลังงาน 1 ค่า ค่าสัมประสิทธิ์เดลต้าเซปสตรีม 12 ค่าและค่าเดลต้าพลังงานอีก 1 ค่า

3.1.3 การสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้ในการจำแนกบุคคลโดยอาศัยขั้นตอนของ SAGC

วิธีการควบคุมอัตราขยายสัญญาณคลื่นรูปไซน์อัตโนมัติเป็นการปรับแอมพลิจูดสัญญาณอินพุตรูปไซน์ให้มีสัญญาณเอาต์พุตแอมพลิจูดหนึ่งหน่วย ซึ่งขั้นตอนการทำงานของ SAGC สามารถแสดงบล็อกไดอะแกรมได้ดังรูปที่ 3.4



รูปที่ 3.4 บล็อกไดอะแกรมของ SAGC

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับการสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้ในการจำแนกบุคคลได้อาศัยขั้นตอนของ SAGC เพื่อทำการวิเคราะห์สัญญาณสำหรับการนำไปใช้เพื่อหาคุณลักษณะของสัญญาณเสียงพูดโดยใช้การแปลงฟูเรียร์แบบไม่ต่อเนื่อง

จากระบบ ค่าเอาต์พุต $x_s(t)$ และ $x_{sq}(t)$ ถูกนำมาประมวลผลหาค่าสัมประสิทธิ์โดยนำสัญญาณเอาต์พุตของ SAGC ($x_s(t)$) ทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องเพื่อหาสัมประสิทธิ์ที่มีขนาดสูงที่สุดซึ่งเป็นตัวกำหนดระดับเสียงสูงต่ำของสัญญาณเสียง ส่วนกรอบสัญญาณของพิทช์ ($x_{sq}(t)$) จะถูกทำการกรองด้วยตัวกรองความถี่ต่ำผ่าน แล้วนำสัญญาณที่ได้ไปทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องเพื่อหาค่าสัมประสิทธิ์ที่มีขนาดสูงที่สุดซึ่งจะเป็นตัวแทนของความถี่ของพิทช์ (Pitch) โดยสัมประสิทธิ์อีกชุดได้จากการทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณ $x_{sq}(t)$ เพื่อเก็บค่าสเปกตรัมของสัญญาณ จากขั้นตอนนี้จะได้ค่าสัมประสิทธิ์สำหรับเป็นอินพุตให้กับระบบโครงข่ายประสาทเทียมสำหรับการจำแนกบุคคล สามารถแสดงโปรแกรมที่เขียนบน Matlab ได้ดังนี้

```
testvoic = xcut;
p_cutoff = 500;
[bl,a1] = butter(4, lp_cutoff/(Fs/2), 'low');
pre = filter(bl, a1, testvoic);
tc = linspace(0,n,n*8000);
fdsig=diff(pre)./diff(tc); % first derivative
fdsig=[fdsig 0]; % pad the last bit of the first derivative
sdsig=diff(pre,2)./(diff(tc(1:length(tc)-1))).^2; % second derivative
sdsig=[sdsig 0 0]; % pad the last bit of the second derivative
xsqfd=fdsig.^2; % squaring the first derivative
xmul=pre.*sdsig; % multiplying between signal and second derivative
y=xsqfd-xmul; % the difference
sq=sqrt(y); %square root
sagc=fdsig./sq; % divided
rsq = real(sq);
rsagc = real(sagc);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

fftrsagc = fft(rsagc,nn);
fftrsqr = fft(rsq,nn);
fftenve = fft(enve,nn);

```

คำสั่งในกรอบสี่เหลี่ยมแสดงการทำ SAGC โดยที่อินพุตเป็น xcut และเอาต์พุตเป็น rsagc
fftrsagc คือ สัมประสิทธิ์การแปลงฟูเรียร์แบบไม่ต่อเนื่องของ rsagc
fftrsqr คือ สัมประสิทธิ์การแปลงฟูเรียร์แบบไม่ต่อเนื่องของ rsq
และ fftenve คือ สัมประสิทธิ์การแปลงฟูเรียร์แบบไม่ต่อเนื่องของ enve

3.1.4 โครงข่ายประสาทเทียม

ปฏิยานิพนธ์นี้จะต้องใช้โครงข่ายประสาทเทียมสองประเภท ประเภทแรกคือ ออกแบบเพื่อใช้ในการแยกคำว่า 'เปิด' ออกจากเสียงอื่นๆ เพื่อใช้ในการสั่งการแก่ระบบ และ ประเภทที่สอง คือ ออกแบบเพื่อใช้จำแนกบุคคลที่มีสิทธิ์และไม่มีสิทธิ์ในการใช้งานระบบออกจากกัน เพื่อจำกัดการเข้าถึงการควบคุมระบบ ในการออกแบบระบบโครงข่ายประสาทเทียม จะต้องพิจารณาถึงองค์ประกอบ ดังนี้

3.1.4.1 ระบบแยกคำพูด

ในการออกแบบสถาปัตยกรรมโครงข่ายประสาทเทียมของปฏิยานิพนธ์นี้ ได้เลือกใช้สถาปัตยกรรมแบบแบบป้อนไปข้างหน้าแบบหลายชั้น ซึ่งในการออกแบบส่วนประกอบของสถาปัตยกรรมแบบป้อนไปข้างหน้าแบบหลายชั้น รวมทั้งสิ้น 3 ชั้นมีดังนี้

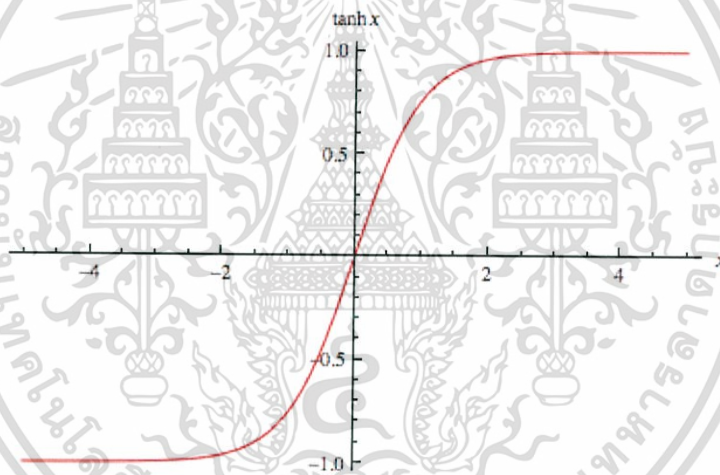
1) ชั้นอินพุต เป็นชั้นที่รองรับข้อมูลเข้าสู่โครงข่ายประสาทเทียม ข้อมูลอินพุตที่เข้ามาในระบบการรู้จำเสียง เป็นค่าสัมประสิทธิ์ที่ได้จากการทำ MFCC ในขั้นตอนข้างต้น ค่าสัมประสิทธิ์ที่ได้จากทุกเฟรมย่อยมีทั้งหมด 1404 ค่า ซึ่งเป็นจำนวนโหนดของอินพุตของระบบ

2) ชั้นซ่อน เป็นชั้นที่เพิ่มประสิทธิภาพในการจัดกลุ่มและการตัดสินใจข้อมูล เพื่อลดความผิดพลาดของเอาต์พุต ซึ่งในปฏิยานิพนธ์นี้ได้ออกแบบให้ชั้นซ่อนมี 1 ชั้น 5 โหนด

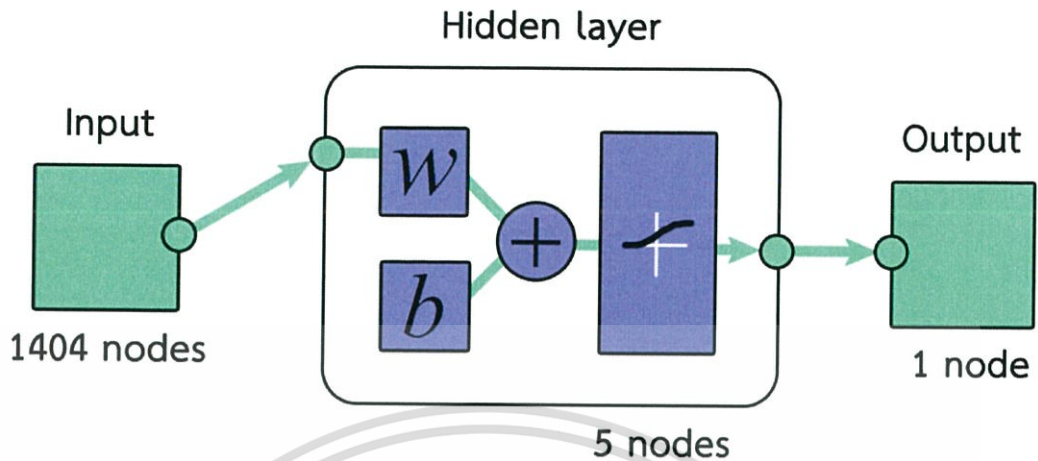
3) ชั้นเอาต์พุต เป็นชั้นที่คำนวณหาผลลัพธ์ของโครงข่ายประสาทเทียม จำนวนโหนดในชั้นเอาต์พุตจะขึ้นอยู่กับจำนวนกลุ่มข้อมูล ซึ่งในปฏิยานิพนธ์นี้ได้ออกแบบให้ระบบมีเอาต์พุต 1 โหนด นั่นคือได้ค่าออกมาเป็นผลตัดสินใจเป็นคำว่าเปิดหรือไม่

กระบวนการในการฝึกฝนระบบ ได้เลือกใช้วิธีการฝึกฝนแบบ Bayesian Regularization Backpropagation วิธีการนี้จะใช้เวลาในการฝึกฝนนาน แต่ให้ผลการฝึกฝนที่ดีสำหรับชุดข้อมูลที่มีค่าใกล้เคียงกัน ซึ่งยากต่อการแยกแยะ การฝึกฝนจะสิ้นสุดลงเมื่อค่าถ่วงโครงข่ายมีการเปลี่ยนแปลงเล็กน้อย

ฟังก์ชันกระตุ้นใช้ในการคำนวณผลตอบสนองทางเอาต์พุตของแต่ละเซลล์ประสาท โดยการออกแบบได้เลือกใช้ฟังก์ชันซิกมอยด์แบบเส้นสัมผัสไฮเพอร์โบลาร์ (Hyperbolic tangent sigmoid function) มาเป็นฟังก์ชันกระตุ้น กราฟของฟังก์ชันแทนเจนต์สามารถแสดงได้ดังรูปที่ 3.5 ซึ่งค่าที่ออกมาจากฟังก์ชันแทนเจนต์จะมีค่าอยู่ระหว่าง -1 ถึง 1 โดยที่ภาพรวมของการออกแบบระบบสามารถแสดงได้ดังรูปที่ 3.6 เมื่อ w คือค่าถ่วงน้ำหนักที่ได้จากการฝึกสอนโครงข่ายประสาทเทียม และ b คือค่าไบอัส ที่นำมาบวกเพิ่มเพื่อให้ค่าที่ถูกน้ำหนักแต่ละตัวมีความแตกต่างกันมากยิ่งขึ้น



รูปที่ 3.5 กราฟของฟังก์ชันซิกมอยด์แบบเส้นสัมผัสไฮเพอร์โบลาร์

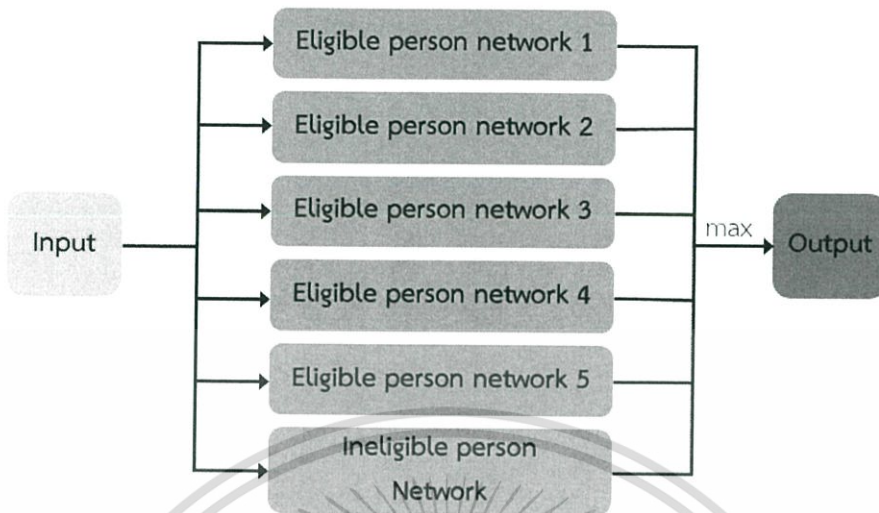


รูปที่ 3.6 ภาพรวมของการออกแบบระบบที่ใช้ในการจำแนกคำพูด

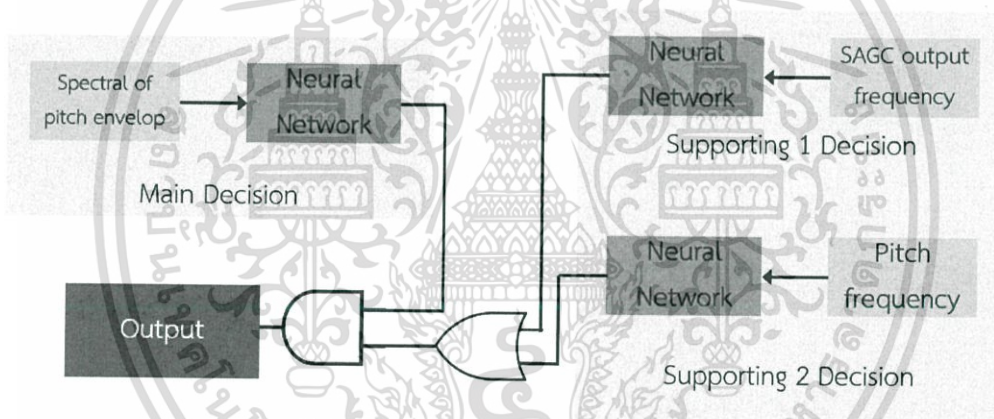
3.1.4.2 ระบบจำแนกบุคคล

ระบบจำแนกบุคคลถูกออกแบบให้เป็นไปตามบล็อกไดอะแกรมดังรูปที่ 3.7 โดยที่ในระบบจำแนกบุคคลจะประกอบไปด้วยโครงข่ายประสาทเทียมย่อยอีก 6 โครงข่าย โดย 5 โครงข่ายแรกนั้นเป็นโครงข่ายในการแยกบุคคลจำนวน 5 คน ที่สามารถใช้งานระบบได้ และอีก 1 โครงข่ายเป็นโครงข่ายที่ใช้แยกบุคคลที่ไม่สามารถเข้าใช้ระบบได้ ซึ่งในแต่ละโครงข่ายจะนำค่าสัมประสิทธิ์ที่ได้จากการทำ SAGC มาใช้เป็นอินพุต โดยภายในแต่ละโครงข่ายสามารถแสดงได้ดังรูปที่ 3.8 โดยจะแบ่งการตัดสินใจออกเป็น 2 ส่วนใหญ่ๆ ส่วนแรกจะเป็นการตัดสินใจหลัก (Main decision) และอีกส่วนจะเป็นการตัดสินใจรอง (Supporting decision)

การตัดสินใจหลักจะใช้สัมประสิทธิ์ที่เป็นสเปกตรัมของกรอบสัญญาณของพิทช์ทั้งหมด 512 ค่าเป็นอินพุตให้แก่โครงข่ายประสาทเทียม เพื่อใช้ในการจำแนกบุคคล นอกจากนี้เพื่อความแม่นยำ ยังได้มีการใช้ตัวตัดสินใจรองเพื่อมาช่วยเสริม โดยตัวตัดสินใจรองประกอบไปด้วยค่าสัมประสิทธิ์ 2 ชุด ชุดแรกจะเป็นสัมประสิทธิ์ที่ได้จากค่าตำแหน่งที่สเปกตรัมของสัญญาณเอาร์ทพูด SAGC มีค่าสูงที่สุด 1 ค่า และชุดที่สองเป็นสัมประสิทธิ์ที่ได้จากค่าความถี่ของพิทช์ 1 ค่า สัมประสิทธิ์ทั้งสองตัวนี้จะถูกนำไปฝึกฝนในโครงข่ายประสาทเทียมเพื่อใช้ในการตัดสินใจรอง โดยเงื่อนไขการตัดสินใจจะเลือกจากโครงข่ายตัดสินใจหลักเป็นหลัก แล้วเลือกโครงข่ายการตัดสินใจรองอย่างน้อย 1 ตัวเพื่อมาเสริมการตัดสินใจหลัก เงื่อนไขการตัดสินใจสามารถแสดงลोजิกดังรูปที่ 3.8



รูปที่ 3.7 บล็อกไดอะแกรมแสดงการออกแบบระบบจำแนกบุคคล



รูปที่ 3.8 บล็อกไดอะแกรมแสดงวิธีการตัดสินใจหลัก การตัดสินใจรอง และลอจิกสำหรับการเลือกตัดสินใจ

ซึ่งในการออกแบบของแต่ละโครงข่ายได้ออกแบบโครงข่ายประสาทเทียมเหมือนกัน เพียงแต่เปลี่ยนอินพุตให้เป็นสัมประสิทธิ์ตามที่ได้ระบุไว้ด้านบน เป็นของผู้ที่สามารถเข้าใช้ระบบได้ทั้งหมด 5 คน และสัมประสิทธิ์ของผู้ที่ไม่สามารถเข้าใช้ระบบได้ ตามลำดับ โดยมีโครงสร้างดังนี้

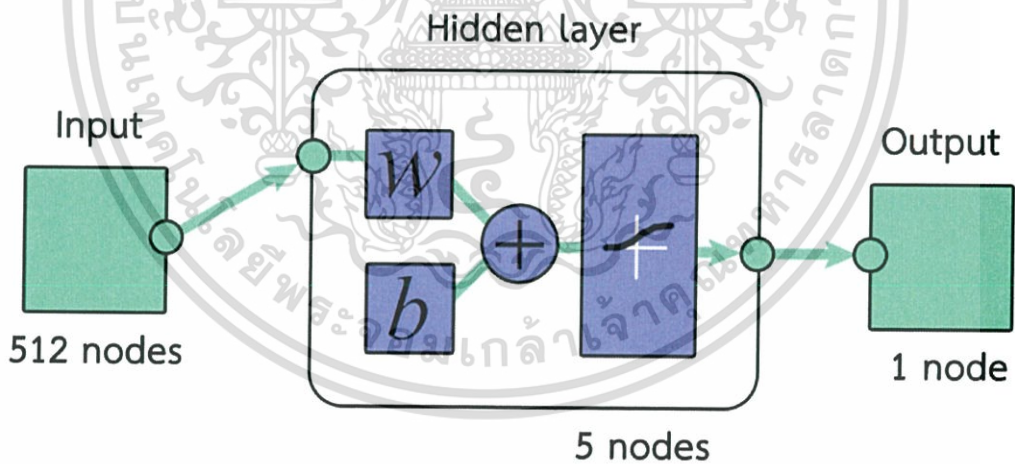
- 1) ชั้นอินพุต เป็นชั้นที่รองรับข้อมูลเข้าสู่โครงข่ายประสาทเทียม ข้อมูลอินพุตที่เข้ามาในระบบการรู้จำเสียง เป็นค่าสัมประสิทธิ์ที่ได้จากสเปกตรัมของกรอบของพิทช์ ค่าสัมประสิทธิ์ที่ได้มีทั้งหมด 512 ค่า ซึ่งเป็นจำนวนโหนดของอินพุตของระบบการตัดสินใจหลัก

ส่วนสัมประสิทธิ์ที่ได้จากค่าตำแหน่งที่ค่าสเปกตรัมของสัญญาณสูงที่สุดซึ่งเป็นตัวกำหนดระดับเสียงสูงต่ำของสัญญาณเสียง 1 ค่า และสัมประสิทธิ์ที่ได้จากค่าความถี่ของพิทช์ 1 ค่าเป็นจำนวนโหนดของอินพุตของระบบการตัดสินใจครั้งที่ 1 และ 2 ตามลำดับ

2) ชั้นซ่อน เป็นชั้นที่เพิ่มประสิทธิภาพในการจัดกลุ่มและการตัดสินใจข้อมูล เพื่อลดความผิดพลาดของเอาต์พุต ในระบบการตัดสินใจหลักได้ออกแบบให้ชั้นซ่อนมี 1 ชั้น 5 โหนด การตัดสินใจครั้งที่ 1 และ 2 ใช้ชั้นซ่อน 1 ชั้น 30 โหนด เนื่องจากโหนดอินพุตมีจำนวนน้อย จะต้องใช้ชั้นซ่อนหลายๆ ชั้น เพื่อการฝึกฝนที่แม่นยำ

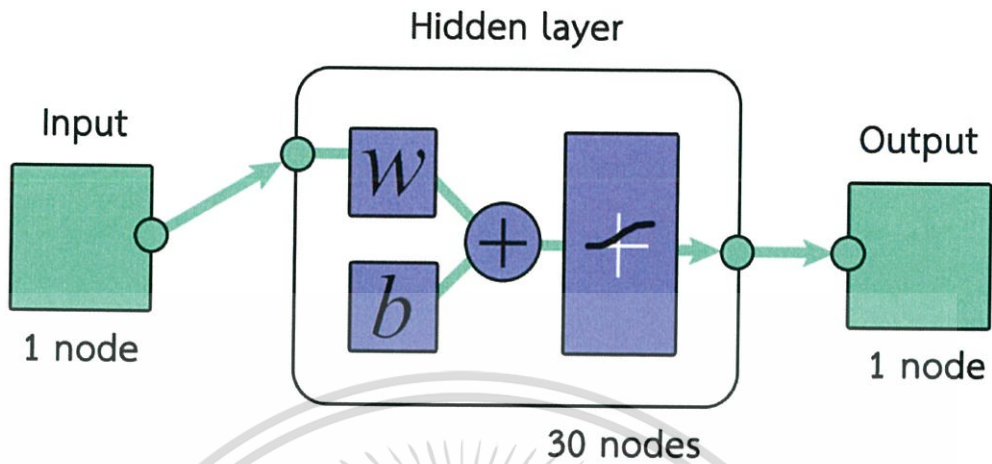
3) ชั้นเอาต์พุต เป็นชั้นที่คำนวณหาผลลัพธ์ของโครงข่ายประสาทเทียม จำนวนโหนดในชั้นเอาต์พุตจะขึ้นอยู่กับจำนวนกลุ่มข้อมูล ซึ่งในแต่ละโครงข่ายประสาทเทียมได้ออกแบบให้ระบบมีเอาต์พุต 1 โหนด ได้ค่าออกมาตัดสินใจว่าเป็น 1 ใน 5 บุคคลในฐานข้อมูลหรือไม่

การออกแบบโครงข่ายประสาทเทียมของแต่ละระบบการตัดสินใจหลัก และการออกแบบโครงข่ายประสาทเทียมสำหรับแต่ละระบบการตัดสินใจครั้งที่ 1 และ 2 สามารถแสดงได้ดังรูปที่ 3.9 และรูปที่ 3.10 ตามลำดับ



รูปที่ 3.9 การออกแบบโครงข่ายประสาทเทียมของแต่ละระบบการตัดสินใจหลัก

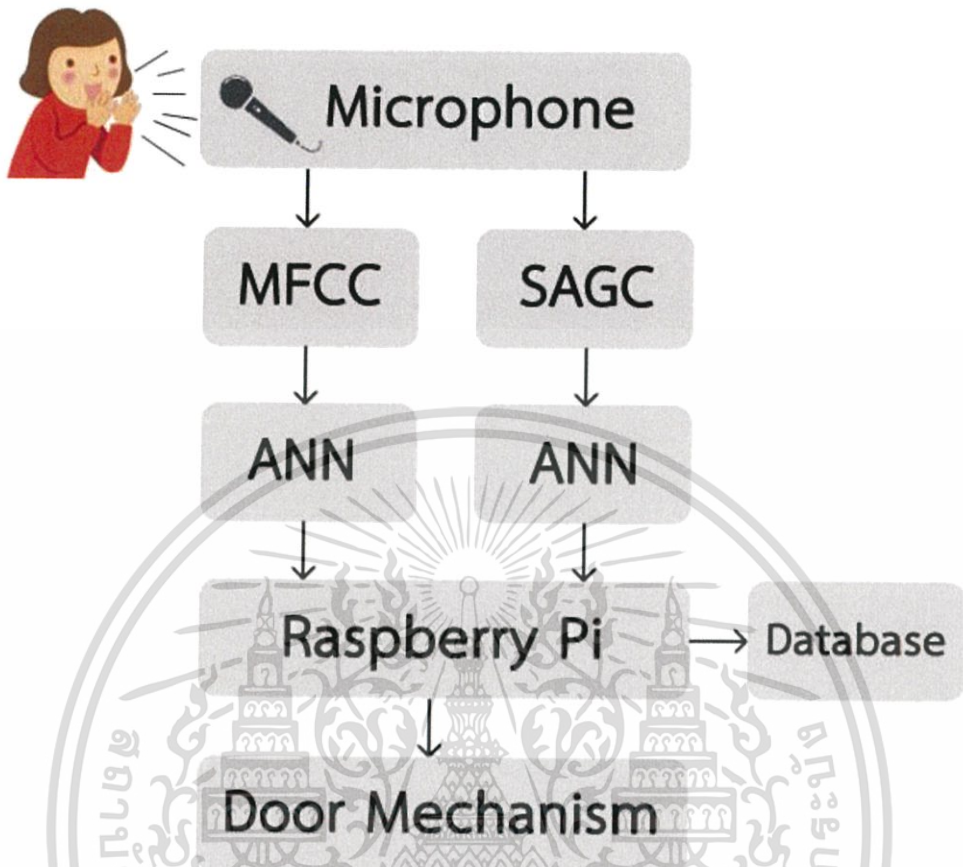
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 การออกแบบโครงข่ายประสาทเทียมสำหรับแต่ละระบบการตัดสินใจครั้งที่ 1 และ 2

3.1.5 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้การรู้จำเสียง

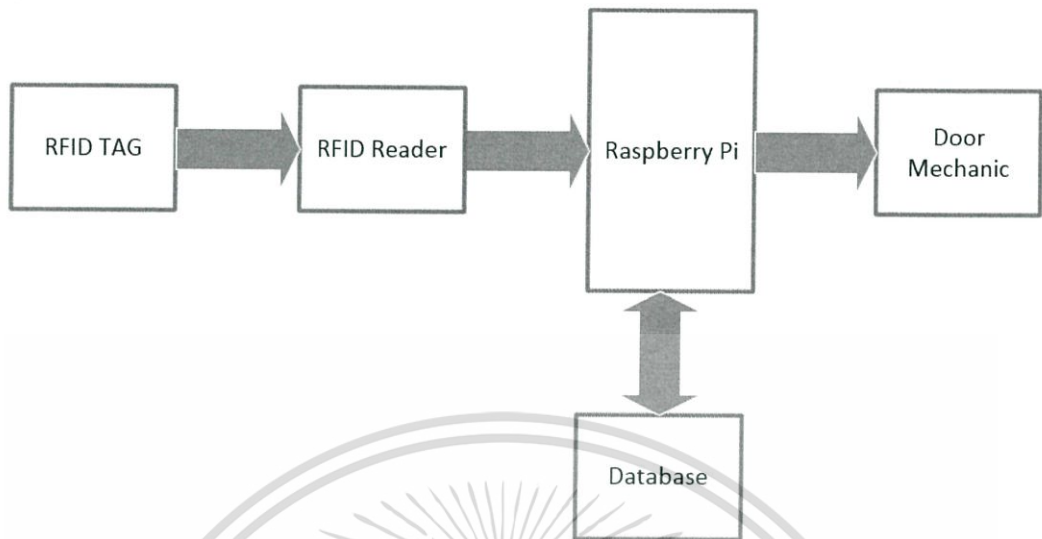
ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้การรู้จำเสียง สามารถแสดงบล็อกไดอะแกรมได้ดังรูปที่ 3.11 โดยเริ่มจากผู้ใช้งานพูดผ่านไมโครโฟน ระบบจะทำการประมวลผลสัญญาณเสียงผ่านอัลกอริทึม MFCC และ SAGC แล้วตัดสินใจโดยระบบโครงข่ายประสาทเทียมว่าผู้พูดจะสามารถเข้าถึงการควบคุมได้หรือไม่ ซึ่งถ้าสามารถเข้าถึงระบบได้รหัสเบอร์รีพายจะบันทึกข้อมูลการเข้าใช้งานในฐานข้อมูล และรหัสเบอร์รีพายก็จะสั่งให้ประตูเปิดออก



รูปที่ 3.11 บล็อกไดอะแกรมส่วนการควบคุมการเปิด-ปิดประตูโดยใช้การรู้จำเสียง

3.1.6 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี

ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี จะแสดงได้ดังบล็อกไดอะแกรมของระบบอาร์เอฟไอดีในรูปที่ 3.12 จะเห็นได้ว่าเครื่องอ่านอาร์เอฟไอดีจะทำการอ่านค่าไอดีของแท็ก จากนั้นเครื่องอ่านอาร์เอฟไอดีจะส่งค่าไอดีไปยังรหัสเบอร์รีพาย ซึ่งเป็นตัวควบคุมระบบการทำงาน จากนั้นรหัสเบอร์รีพายจะทำการเช็คข้อมูลในฐานข้อมูล ถ้าข้อมูลที่เข้ามาตรงกับฐานข้อมูล รหัสเบอร์รีพายจะสั่งการให้ประตูเปิดได้



รูปที่ 3.12 บล็อกไดอะแกรมระบบอาร์เอฟไอดี

การทำงานของระบบอาร์เอฟไอดี ประกอบไปด้วยอุปกรณ์ราสเบอร์รี่พายโมเดลB+ เครื่องอ่านอาร์เอฟไอดี MFRC522 และแท็ก ในการสั่งการเปิด-ปิดประตูแม่เหล็ก ซึ่งหลักการทำงานของระบบอาร์เอฟไอดีสามารถเขียนเป็นโฟลว์ชาร์ตได้ดังรูปที่ 3.13 จะเห็นได้ว่าการทำงานของระบบเริ่มจากการนำแท็กแตะกับเครื่องอ่านอาร์เอฟไอดี เครื่องอ่านอาร์เอฟไอดีจะทำการอ่านค่าไอดีแล้วส่งไปยังราสเบอร์รี่พาย จากนั้นราสเบอร์รี่พายจะตรวจสอบความถูกต้องของไอดีในฐานข้อมูล ถ้าไอดีตรงกัน ในหน้าต่างคอมพิวเตอร์ของราสเบอร์รี่พาย จะแสดงข้อมูลที่เกี่ยวข้อง พร้อมกับสั่งการให้ประตูเปิด ถ้าข้อมูลไม่ตรงกับฐานข้อมูล ระบบจะกลับไปเริ่มต้นใหม่ในขั้นตอนรอรับการแตะแท็ก



รูปที่ 3.13 โฟลว์ชาร์ตระบบอาร์เอฟไอดี

3.1.6.1 การออกแบบให้แท็กติดต่อสื่อสารกับเครื่องอ่านอาร์เอฟไอดี

แท็กที่นำมาใช้เป็นแท็กชนิดแพสซีฟ สามารถทำงานได้ไม่ต้องอาศัยแหล่งจ่ายไฟภายนอก มีขนาดเล็ก น้ำหนักเบา และสามารถอ่านข้อมูลได้ในระยะไม่เกิน 1 เมตร ส่วนเครื่องอ่านอาร์เอฟไอดีที่นำมาใช้เป็นรุ่น MFRC522 ใช้งานที่ความถี่ 13.56 MHz มีระยะในการอ่านสูงสุดประมาณ 10 เซนติเมตร ใช้สำหรับการอ่านและการจัดเก็บข้อมูลภายในแท็กที่ต้องการความปลอดภัยสูง ในส่วนการทดสอบการติดต่อกันระหว่างตัวแท็กและอาร์เอฟไอดีจะใช้โปรแกรม

Mifare ในการทดลองให้เครื่องอ่านอาร์เอฟไอดีอ่านข้อมูลของตัวแท็กจำนวน 10 ใบ พบว่าข้อมูลที่
ได้แตกต่างกัน

3.1.6.2 การออกแบบให้เครื่องอ่านอาร์เอฟไอดีติดต่อสื่อสารกับรหัสเบอร์รี่
พาย

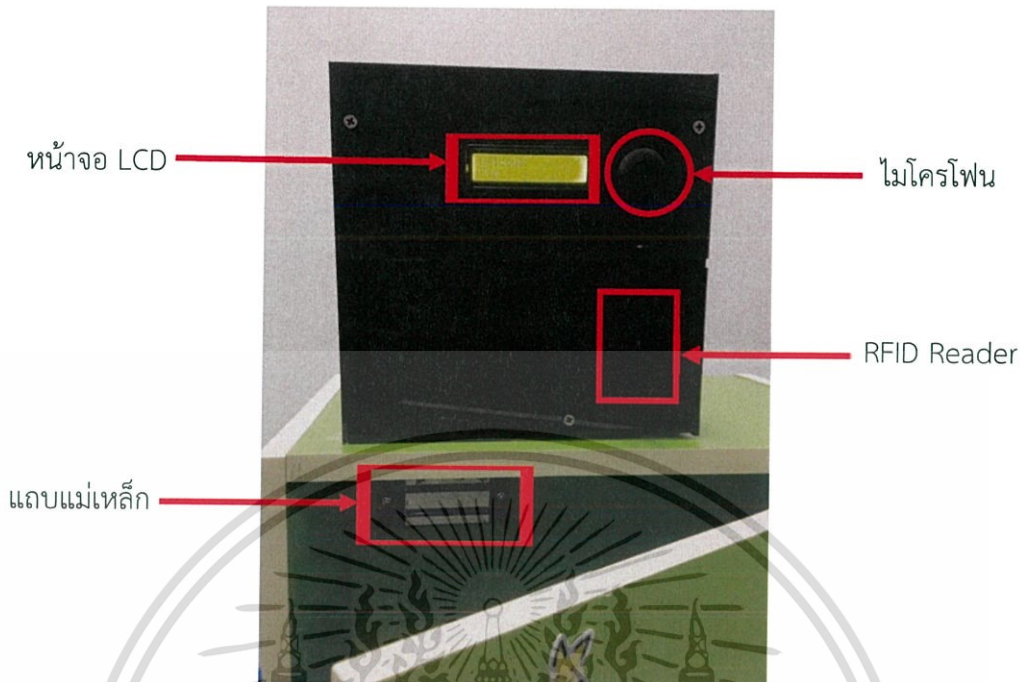
ในการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี ได้เลือกใช้
อุปกรณ์รหัสเบอร์รี่พายโมเดล B+ ในการทำงาน โดยออกแบบชุดคำสั่งสำหรับการติดต่อด้วย
โปรแกรม putty เพื่อติดต่อกับรหัสเบอร์รี่พายผ่านทางคอมพิวเตอร์ ในเครือข่ายท้องถิ่น ซึ่งใช้ภาษา
ไพธอนในการสร้างชุดคำสั่งต่างๆ จากนั้นนำเครื่องอ่านอาร์เอฟไอดีต่อเข้ากับพอร์ตจีพีไอโอของ
รหัสเบอร์รี่พาย แล้วเขียนโปรแกรมเก็บหมายเลขไอดีของแท็กจำนวน 10 ค่า เพื่อใช้ในการสั่งการ
เปิด-ปิดประตูต่อไป

3.1.6.3 การออกแบบฐานข้อมูลบนรหัสเบอร์รี่พาย

การสร้างฐานข้อมูลบนรหัสเบอร์รี่พาย มีระบบปฏิบัติการบนพื้นฐานของ
ลินุกซ์ซึ่งทำงานเช่นเดียวกับเว็บเซิร์ฟเวอร์ ในปริญญาานิพนธ์นี้จะสร้างฐานข้อมูลโดยใช้
phpMyAdmin ซึ่งเป็นโปรแกรมที่ใช้ในการจัดการฐานข้อมูล MySQL ผ่านทางเบราว์เซอร์ในรหัส
เบอร์รี่พาย

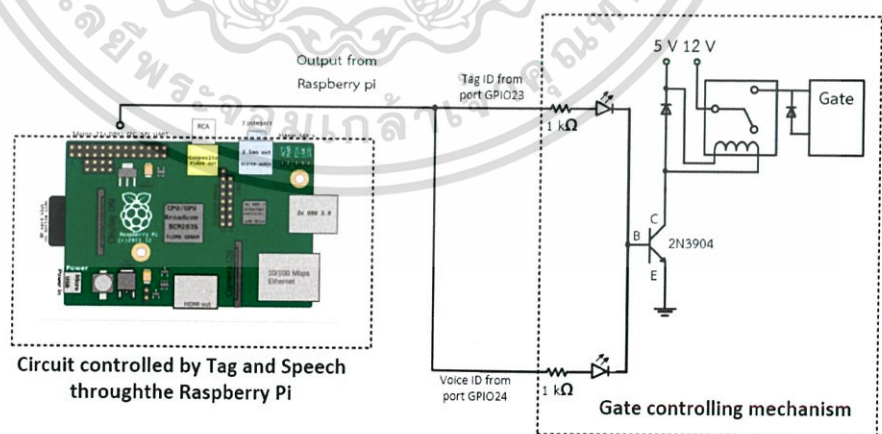
3.1.7 ส่วนควบคุมประตูแม่เหล็ก

ในส่วนของการควบคุมการเปิด-ปิดประตู ได้จัดทำต้นแบบประตูที่เปิด-ปิด
ล็อกด้วยแถบแม่เหล็ก ดังแสดงรูปที่ 3.14 โดยมีการติดตั้งไมโครโฟนเพื่อให้ผู้ใช้งานสามารถสั่งการ
ด้วยเสียง และส่วนของ RFID Reader เพื่อการใช้งานด้วย RFID Tag พร้อมทั้งมีการแสดงผลการสั่ง
การประตูบนหน้าจอ LCD อีกด้วย



รูปที่ 3.14 ต้นแบบประตูที่เปิด-ปิดด้วยแถบแม่เหล็ก

กลไกควบคุมประตูด้วยทรานซิสเตอร์ มีอินพุตเป็นสัญญาณที่ได้จากส่วนควบคุมทั้ง 2 ส่วนได้แก่อินพุตจาก voice id และ tag id ซึ่งเป็นสัญญาณความต่างศักย์ไฟฟ้าขนาด 3.3 โวลต์ที่ขาเบสของทรานซิสเตอร์ชนิดเอ็นพีเอ็น ทำให้กระแสไฟฟ้าไหลจากขาเบสสู่ขาอีมิเตอร์ลงกราวนด์ โดยต่อรีเลย์อนุกรมระหว่างไฟเลี้ยงและขาคอลเลคเตอร์ ทำให้สามารถควบคุมการเปิดปิดประตูได้จากการตัดไฟเลี้ยงแม่เหล็ก ซึ่งแสดงกลไกควบคุมประตูด้วยทรานซิสเตอร์ดังรูปที่ 3.15



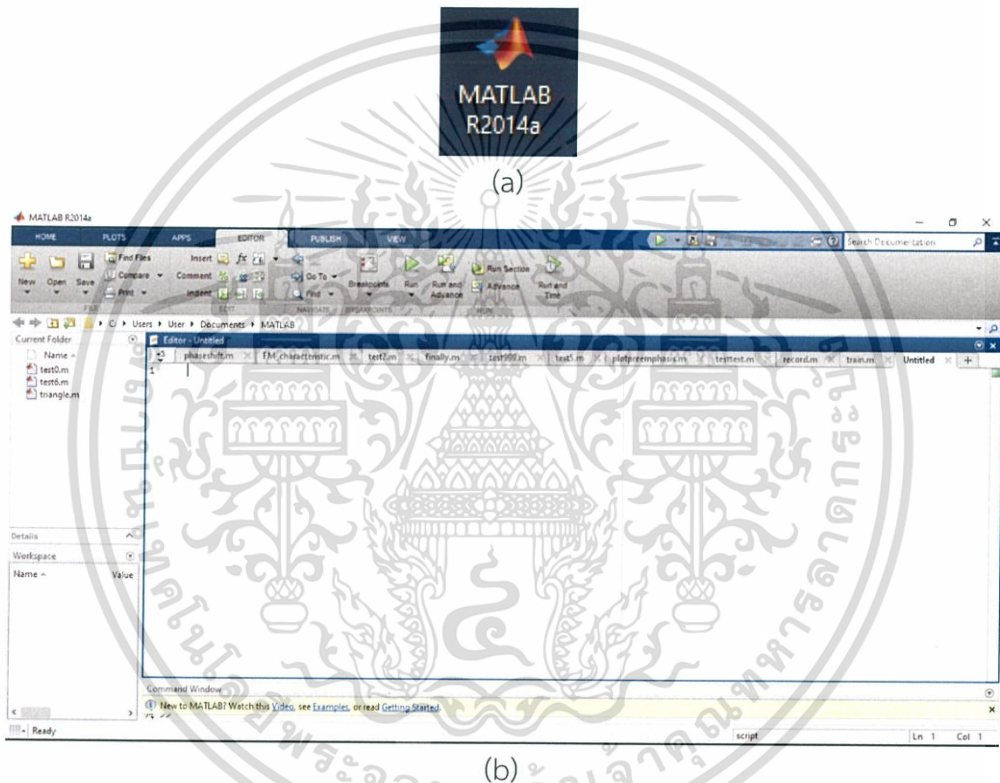
รูปที่ 3.15 กลไกควบคุมประตูแม่เหล็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 เครื่องมือที่ใช้ในการทดลอง

3.2.1 โปรแกรม Matlab

Matlab เป็นภาษาคอมพิวเตอร์ระดับสูงที่มาพร้อมด้วยสภาพแวดล้อมการทำงานเชิงโต้ตอบ (คล้ายเครื่องคิดเลข) ซึ่งสามารถคำนวณคณิตศาสตร์ที่ซับซ้อนได้อย่างรวดเร็วมากกว่าภาษาคอมพิวเตอร์สมัยก่อน เช่น ภาษา C, C++ หรือ Fortran โดยไอคอนและหน้าต่างการใช้งานโปรแกรม Matlab แสดงได้ดังรูปที่ 3.16(a) และ 3.16(b) ตามลำดับ



รูปที่ 3.16 ไอคอนและหน้าต่างการใช้งานโปรแกรม Matlab

(a) ไอคอนโปรแกรม Matlab

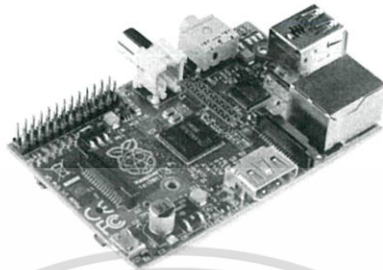
(b) หน้าต่างการใช้งานโปรแกรม Matlab

3.2.2 ราสเบอร์รี่พาย (Raspberry Pi)

ราสเบอร์รี่พายเป็นบอร์ดเอ็มเบดเต็ด ลินุกซ์ (Embedded Linux) โดยคำว่า พาย (pi) นั้นมาจาก ไพธอน (python) ซึ่งเป็นภาษาหลักที่ใช้ในการพัฒนาดังแสดงในรูปที่ 3.17 มีขนาดเท่ากับบัตรเครดิตประชาชน ความสามารถคือสามารถติดตั้งระบบปฏิบัติการลงไปได้และระบบปฏิบัติการที่ถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

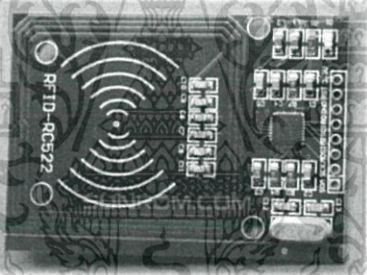
ติดตั้งลงไปในที่ถูกออกแบบพัฒนามาเพื่อราสเบอร์รี่นี้โดยเฉพาะ คือ ราสเบียน (Raspbian) ราสเบอร์รี่พายมีความสามารถหลายอย่างเหมือนคอมพิวเตอร์พีซีที่ใช้กันทั่วไป เพียงแต่ความเร็วในการประมวลผลอาจจะด้อยกว่า



รูปที่ 3.17 ราสเบอร์รี่พาย

3.2.3 เครื่องอ่านอาร์เอฟไอดี MFRC522

เครื่องอ่านอาร์เอฟไอดี MFRC522 เป็นเครื่องอ่านที่สามารถประมวลผลได้โดยไร้สัมผัส รองรับมาตรฐาน ISO14443A. รูปที่ 3.18 แสดงเครื่องอ่านอาร์เอฟไอดี MFRC522



รูปที่ 3.18 เครื่องอ่านอาร์เอฟไอดี MFRC522

3.2.4 แท็ก

แท็กที่ใช้คือรุ่น MFICS50 ซึ่งเป็นชนิดพาสซีฟ แสดงได้ดังรูปที่ 3.19

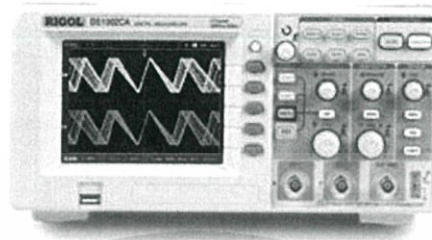


รูปที่ 3.19 แท็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.5 ออสซิลโลสโคป (Oscilloscope)

ออสซิลโลสโคปเป็นเครื่องมือที่สร้างขึ้นมาเพื่อใช้แสดงรูปสัญญาณทางไฟฟ้า สามารถแสดงรูปสัญญาณเดี่ยว หรือหลายๆ รูปสัญญาณบนจอเดียวกัน แสดงได้ดังรูปที่ 3.20



รูปที่ 3.20 ออสซิลโลสโคป

3.3 การจัดเก็บผลการทดลอง

3.3.1 การสกัดคุณลักษณะของสัญญาณเสียงพูด

การสกัดคุณลักษณะของสัญญาณเสียงพูด จากขั้นตอนต่างๆ สัญญาณเสียงจะผ่านการประมวลผลสัญญาณและจัดเก็บผลการทดลองโดยใช้โปรแกรม Matlab สามารถแสดงได้ดังรูปที่ 3.21

3.3.2 โครงข่ายประสาทเทียม

การจัดเก็บผลการทดลองในส่วนของโครงข่ายประสาทเทียมนั้น ใช้โปรแกรม Matlab ในการฝึกฝนและทดสอบระบบ โดยการจัดเก็บผลการทดลองแบ่งออกเป็น 2 ส่วน

3.3.2.1 การจำแนกคำพูด ในการจำแนกคำพูดจะเก็บผลการทดลองโดยใช้สัมประสิทธิ์ของคำว่า 'เปิด' 50 คำ และคำอื่นๆ อีก 50 คำ เพื่อทดสอบความสามารถในการแยกคำของระบบ

3.3.2.2 การจำแนกบุคคล ในการจำแนกบุคคลได้เก็บผลการทดลองโดยใช้สัมประสิทธิ์ของบุคคลที่สามารถเข้าใช้งานระบบได้ทั้งหมด 5 คน และบุคคลที่ไม่สามารถเข้าใช้งานระบบได้ 5 คน ในการทดสอบระบบจำแนกบุคคล

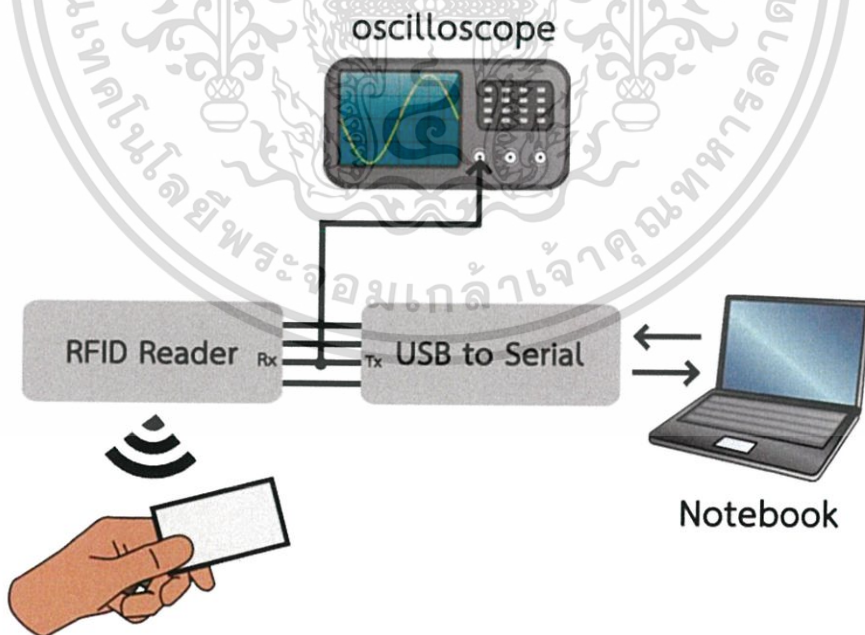
โดยการที่ระบบจะอนุญาตแต่ละบุคคลนั้นจะต้องผ่านทั้งสองระบบ นั่นคือ ระบบจำแนกคำพูด และระบบจำแนกบุคคล การจัดเก็บผลการทดลองสามารถแสดงได้ดังรูปที่ 3.21



รูปที่ 3.21 การออกแบบการจัดเก็บผลการทดลองของการสกัดคุณลักษณะของสัญญาณเสียงพูด และโครงข่ายประสาทเทียม

3.3.3 ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี

ส่วนการควบคุมการเปิด-ปิดประตูโดยใช้ระบบอาร์เอฟไอดี จะส่งการผ่านรหัสเบอร์รี้พาย โดยได้ทำการเก็บหมายเลขไอดีการ์ดไว้ในฐานข้อมูล และเขียนชุดคำสั่งเพื่อควบคุมประตู เมื่อนำแท็กแต่ละที่เครื่องอ่านอาร์เอฟไอดี แท็กที่ค่าไอดีตรงกับฐานข้อมูลจะสามารถเปิดประตูได้ ส่วนแท็กที่ค่าไอดีไม่ตรงกับฐานข้อมูลจะไม่สามารถเปิดประตูได้ การออกแบบการจัดเก็บผลการทดลองของการอ่านค่าแท็กสามารถแสดงได้ดังรูปที่ 3.22



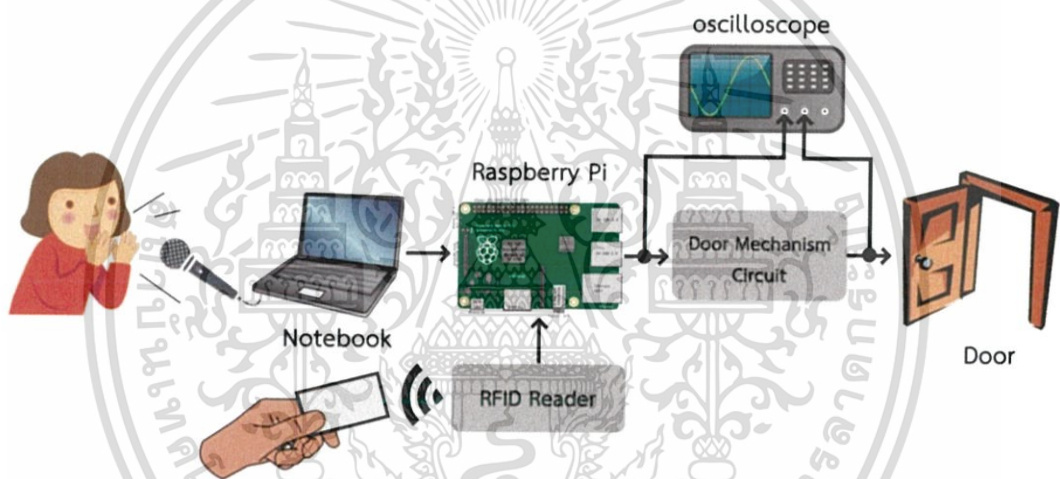
รูปที่ 3.22 การออกแบบการจัดเก็บผลการทดลองของการอ่านค่าแท็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.4 ระบบรวม

ทำการทดสอบระบบรวม ตั้งแต่การใช้งานระบบรู้จำเสียงในการสั่งการและการใช้งานระบบด้วย RFID โดยวัดสัญญาณความต่างศักย์ไฟฟ้าที่ใช้การควบคุมประตูที่ได้จากราสเบอร์รี่ พาย และความต่างศักย์ไฟฟ้าที่แอมป์หลักด้วยออสซิลโลสโคป

เมื่อผู้ใช้พูดผ่านไมโครโฟน สัญญาณเสียงที่ได้จะถูกบันทึกและประมวลผลโดยโปรแกรม Matlab ผลการตัดสินใจที่ได้จากโครงข่ายประสาทเทียมจะถูกส่งไปยังราสเบอร์รี่พาย เพื่อให้ราสเบอร์รี่พายจ่ายความต่างศักย์ไฟฟ้าไปควบคุมกลไกควบคุมประตู จึงใช้ออสซิลโลสโคปวัดสัญญาณที่ได้จากราสเบอร์รี่พายและวัดสัญญาณไฟเลี้ยงของแอมป์หลัก ซึ่งสามารถแสดงบล็อกไดอะแกรมได้ดังรูปที่ 3.23



รูปที่ 3.23 บล็อกไดอะแกรมการจัดเก็บผลการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลอง

4.1 กล่าวนำ

ในบทนี้เป็นการนำเสนอผลจากการทดลองเพื่อยืนยันผลการออกแบบในบทที่ 3 โดยการทดลองแบ่งหัวข้อเพื่อนำเสนอเป็นสองส่วนหลัก โดยส่วนแรก คือการระบุตัวตนโดยใช้การรู้จำเสียง และส่วนที่สอง คือผลการทดลองการเข้าถึงด้วยอาร์เอฟไอดี ก่อนจะทดสอบระบบรวมเป็นลำดับสุดท้าย

4.2 การระบุตัวตนโดยใช้การรู้จำเสียง

การระบุตัวตนโดยใช้การรู้จำเสียงสามารถแยกระบบได้ออกเป็นสองระบบ คือระบบการจำแนกคำพูด และระบบจำแนกผู้พูด โดยที่ระบบการจำแนกคำพูดจะทำการแยกคำสั่งที่ใช้ในการสั่งการควบคุม ในที่นี้ใช้คำว่า 'เปิด' ในการสั่งการ ส่วนระบบจำแนกผู้พูดจะทำการระบุตัวตนของผู้ใช้งาน ซึ่งจะจำแนกว่าบุคคลนี้เป็นใคร สามารถใช้งานได้หรือไม่

4.2.1 ระบบการจำแนกคำพูด

ระบบจำแนกคำพูดเป็นระบบที่ใช้ในการแยกคำว่า 'เปิด' ออกจากคำอื่นๆ เพื่อเป็นการสั่งการให้แก่ระบบ โดยที่ระบบจะอนุญาตให้เข้าใช้งานได้ก็ต่อเมื่อผู้พูดพูดคำว่า 'เปิด' โดยที่ระบบจำแนกคำพูดจะแบ่งออกเป็นสองส่วนใหญ่ๆ นั่นคือการสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้อัลกอริทึม MFCC และระบบโครงข่ายประสาทเทียม

4.2.1.1 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึม

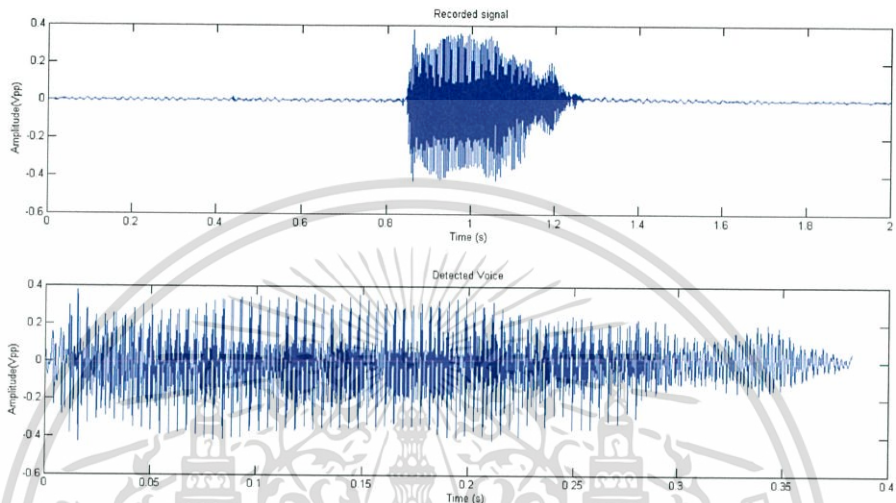
MFCC

การสกัดคุณลักษณะของสัญญาณเสียงพูดซึ่งเลือกใช้อัลกอริทึม MFCC มีขั้นตอนของการเตรียมสัญญาณเสียงก่อนการทำการสกัดคุณลักษณะเด่นของเสียงทั้งหมด 7 ขั้นตอน ซึ่งประกอบด้วย 1) การทำพรีเอมฟาซิสกับสัญญาณเสียงพูด 2) การแบ่งเฟรมเสียงพูด 3) การทำแฮมมิงวินโดว์ 4) การแปลงฟูเรียร์แบบไม่ต่อเนื่อง 5) การทำเมลฟิลเตอร์แบงค์ 6) การแปลงโคซายน์แบบไม่ต่อเนื่อง และ 7) การหาสัมประสิทธิ์พลังงานและเดลด้าเซปสตรัม โดยแสดงผลการทดลองในแต่ละขั้นตอนตามลำดับดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การเตรียมสัญญาณเสียง

สัญญาณเสียงที่ถูกบันทึกขนาด 2 วินาทีนั้นประกอบไปด้วยส่วนที่เป็นเสียงพูดและส่วนที่ไม่ใช่เสียงพูด ดังนั้นจึงทำการตัดส่วนที่ไม่ใช่เสียงพูดออกด้วยวิธีการกำหนดขอบเขตที่รับได้ (Thresholding) แสดงได้ในรูปที่ 4.1

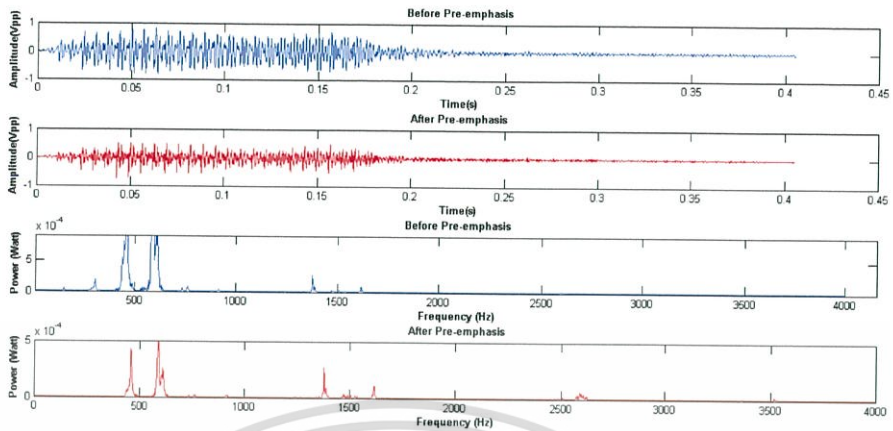


รูปที่ 4.1 การตัดส่วนของสัญญาณที่ไม่ใช่เสียงพูด

จากรูปที่ 4.1 แลกรูปด้านบนแสดงสัญญาณเสียงที่ถูกบันทึก และแลกรูปด้านล่างแสดงสัญญาณเสียงที่ถูกตัดส่วนสัญญาณที่ไม่ใช่เสียงพูดออก

2) ฟรีเอมฟาซิสของสัญญาณเสียงพูด

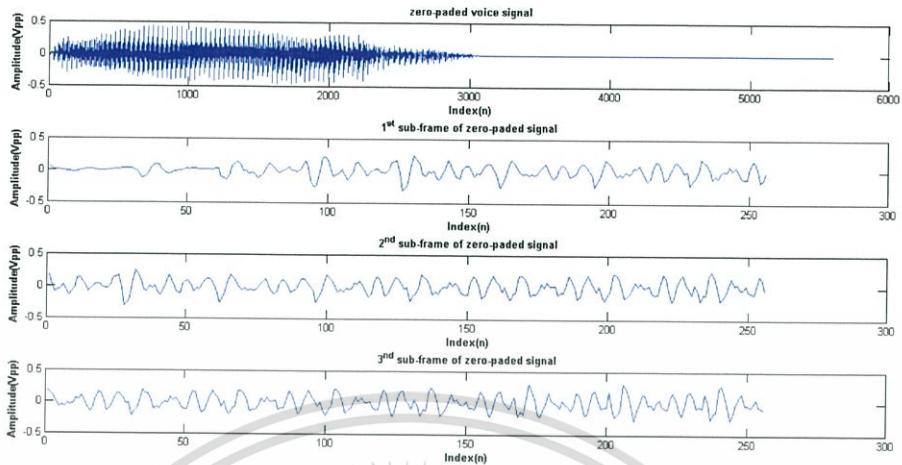
ฟรีเอมฟาซิสของสัญญาณเสียงพูดเพื่อเพิ่มปริมาณของพลังงานที่ความถี่สูงของสัญญาณเสียง ซึ่งโดยทั่วไปสัญญาณเสียงมีปริมาณพลังงานที่ความถี่ต่ำมากกว่าความถี่สูง ผลการฟรีเอมฟาซิสของสัญญาณเสียงพูดแสดงได้ดังรูปที่ 4.2 โดยรูปแถบที่หนึ่งเป็นสัญญาณเสียงที่ถูกตัดเสียงที่ไม่ใช่เสียงพูดออก แล้วนำสัญญาณเสียงนั้นมาผ่านกระบวนการฟรีเอมฟาซิสแสดงผลได้ดังแลกรูปเส้นที่สอง แต่ผลในรูปแถบที่สองเป็นการแสดงผลในโดเมนเวลาทำให้ไม่สามารถเห็นผลการทำฟรีเอมฟาซิสได้ชัดเจน จึงแสดงผลของเสียงก่อนและหลังทำฟรีเอมฟาซิสเป็นโดเมนความถี่ในรูปแถบที่สาม และสี่ตามลำดับ ในรูปแถบที่สี่จะแสดงให้เห็นว่าช่วงความถี่สูงถูกขยายขึ้น



รูปที่ 4.2 แถบความถี่ของเสียงก่อนและหลังทำกระบวนการพรีเอมฟาซีส

3) การแบ่งเฟรมเสียงพูด

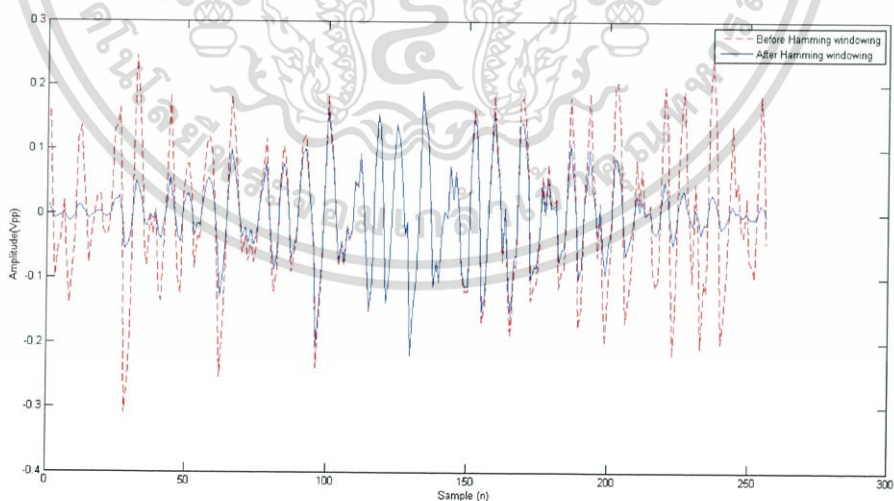
การแบ่งเฟรมจะนำเสียงที่ผ่านการทำพรีเอมฟาซีสมาแบ่งเป็นเฟรมย่อย โดยเสียงพูดที่ถูกตัดมานั้นจะมีความยาวที่ไม่เท่ากัน ซึ่งก่อนทำการแบ่งเฟรมสัญญาณเสียงจะผ่านการทำ zero-pad เพื่อปรับให้สัญญาณเสียงแต่ละสัญญาณมีความยาวเท่ากัน เมื่อทำการ zero-pad สัญญาณเสียงเรียบร้อยแล้ว เสียงที่ได้จะถูกนำมาแบ่งเฟรมให้เป็นเฟรมย่อย แสดงได้ดังรูปที่ 4.3 โดยแถบรูปแรกแสดงสัญญาณเสียงพูดที่ผ่านการทำ Zero-pad ก่อนทำการแบ่งเฟรม ส่วนแถบรูปแถบที่สอง สาม และสี่แสดงเฟรมย่อยของสัญญาณเสียงที่ถูกแบ่งเป็นเฟรมย่อยขนาด 256 จุดข้อมูล



รูปที่ 4.3 เฟรมย่อยของสัญญาณเสียง

4) การทำแฮมมิงวินโดว์

การทำแฮมมิงวินโดว์เป็นการนำแต่ละเฟรมย่อยของสัญญาณเสียงพูดมาคูณกับแฮมมิงวินโดว์ เพื่อให้แต่ละเฟรมย่อยนั้นมีลักษณะเป็นสัญญาณที่ต่อเนื่องในจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยผลของการทำแฮมมิงวินโดว์แสดงได้ดังรูปที่ 4.4 โดยสัญญาณเส้นประแสดงเฟรมย่อยของสัญญาณเสียงก่อนทำแฮมมิงวินโดว์ และสัญญาณเส้นทึบแสดงเฟรมย่อยของสัญญาณเสียงหลังจากผ่านการทำแฮมมิงวินโดว์แล้ว

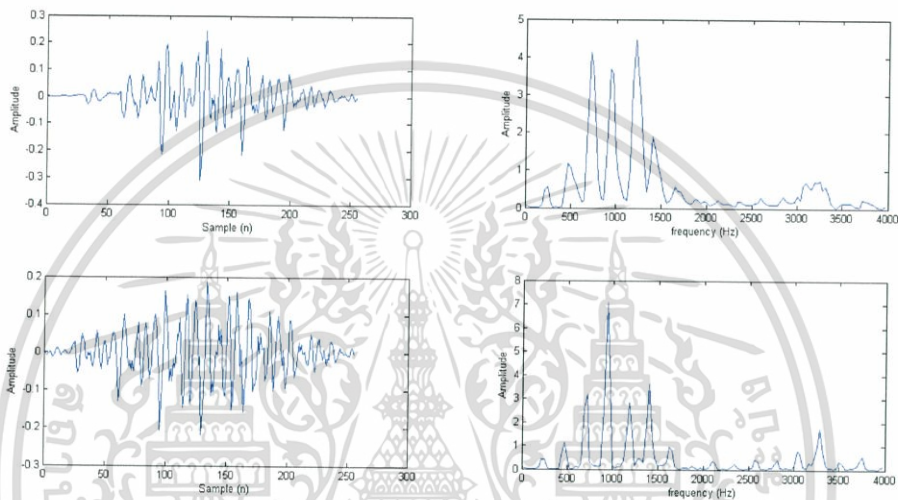


รูปที่ 4.4 สัญญาณที่ถูกทำแฮมมิงวินโดว์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง

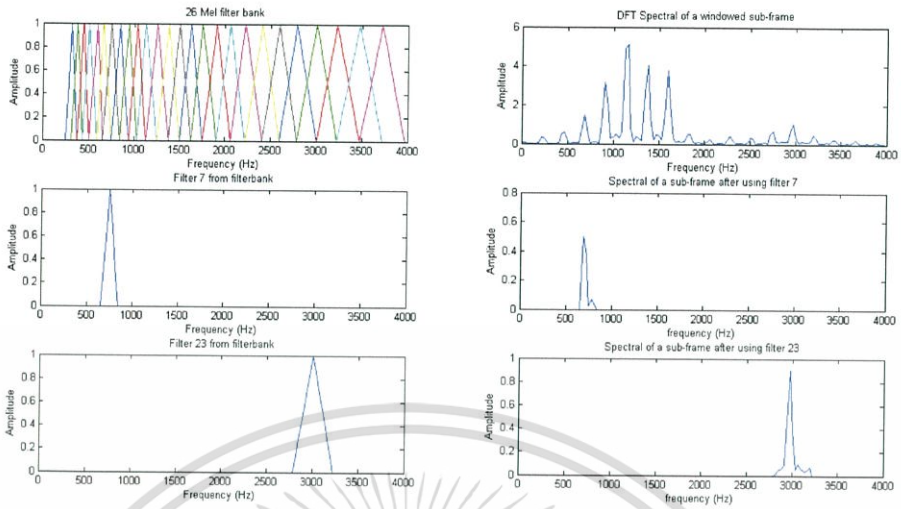
ขั้นตอนการแปลงฟูรีเยร์แบบไม่ต่อเนื่องจะนำเฟรมสัญญาณที่ผ่านการทำวินโดว์มาแปลงจากโดเมนเวลาเป็นโดเมนความถี่เพื่อทำการประมาณหาค่าสเปกตรัมของสัญญาณเสียงแต่ละเฟรม แสดงได้ดังรูปที่ 4.5 โดยแถบรูปด้านซ้ายแสดงเฟรมย่อยของสัญญาณเสียงพูดที่ผ่านการทำแฮมมิงวินโดว์และแถบรูปด้านขวาแสดงผลลัพธ์จากการแปลงฟูรีเยร์แบบไม่ต่อเนื่องของสัญญาณที่ผ่านการทำวินโดว์



รูปที่ 4.5 เฟรมย่อยของสัญญาณเสียงในโดเมนเวลา (ซ้าย) และโดเมนความถี่ (ขวา)

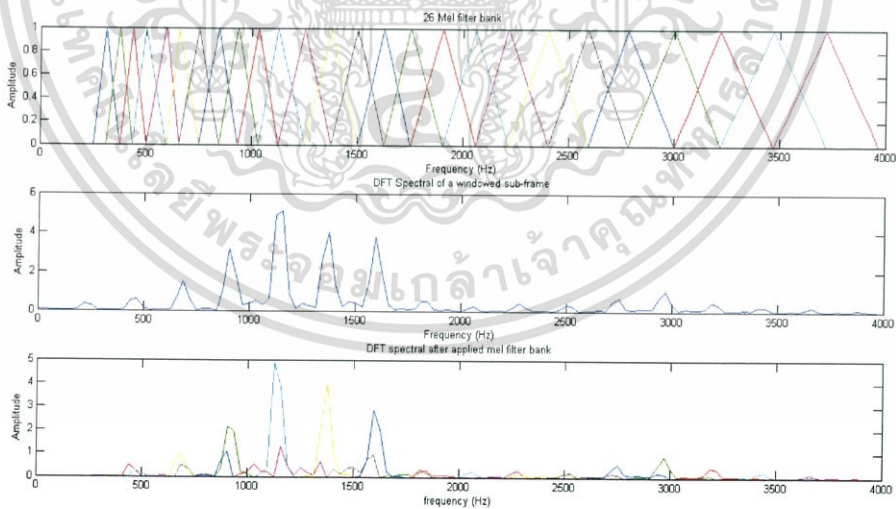
6) การทำเมลฟิลเตอร์แบงก์

การทำเมลฟิลเตอร์แบงก์เป็นการใช้ชุดของตัวกรองสามเหลี่ยม 26 ตัวคู่กับค่าสเปกตรัมของเฟรมย่อยของสัญญาณเสียง ค่าที่ได้หลังจากการใช้ตัวกรองสามเหลี่ยมแต่ละตัวเป็นค่าสเปกตรัมที่ความถี่ต่างๆ แสดงได้ดังรูปที่ 4.6 และค่าสเปกตรัมที่ได้จากการใช้ตัวกรองสามเหลี่ยมทั้งหมด 26 ตัวแสดงได้ดังรูปที่ 4.7 โดยค่าสเปกตรัมในแต่ละตัวกรองจะถูกรวมกันแล้วทำการหาค่าพลังงาน และถูกทำลอการิทึมเพื่อใช้สำหรับขั้นตอนการแปลงโคซายน์แบบไม่ต่อเนื่องต่อไป



รูปที่ 4.6 เมลฟิลเตอร์แบงก์และสเปกตรัมในแต่ละตัวกรอง

จากรูปที่ 4.6 แถบรูปด้านซ้ายบนแสดงเมลฟิลเตอร์ที่เป็นชุดของตัวกรองสามเหลี่ยม 26 ตัวกรอง แถบรูปทางด้านขวาบนแสดงแถบสเปกตรัมของเฟรมย่อยของสัญญาณเสียงพูด แถบรูปทางด้านซ้าย แถบที่สองและสามแสดงตัวกรองสามเหลี่ยมในช่วงความถี่ที่ต่างกัน และแถบรูปทางด้านขวา แถบที่สองและสามแสดงค่าสเปกตรัมในตัวกรองที่แสดงอยู่ในรูปทางด้านซ้าย



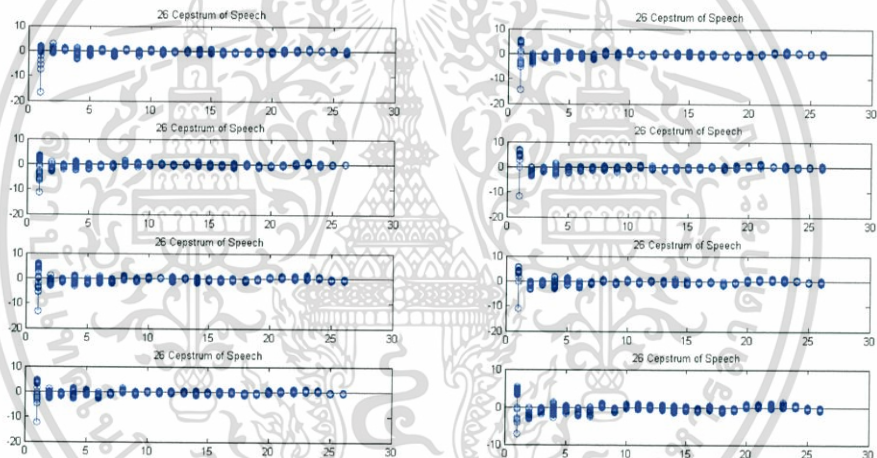
รูปที่ 4.7 เมลฟิลเตอร์แบงก์และสเปกตรัมในแต่ละตัวกรองทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.7 แลกรูปแรกแสดงเมลฟิเตอร์ที่เป็นชุดของตัวกรองสามเหลี่ยม 26 ตัวกรอง แลกรูปที่สองแสดงแถบสเปกตรัมของเฟรมย่อยของสัญญาณเสียงพูด และแลกรูปแถบสุดท้ายแสดงผลจากการคูณฟิลเตอร์แบงค์กับค่าสเปกตรัมของสัญญาณได้เป็นค่าสเปกตรัมที่ความถี่ต่างๆ ที่มีอยู่ในแต่ละตัวกรอง

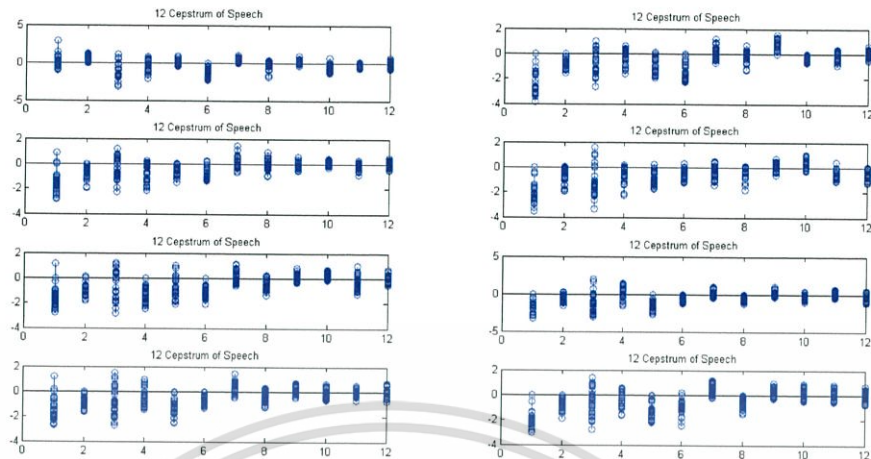
7) การแปลงโคซายน์แบบไม่ต่อเนื่อง

การแปลงโคซายน์แบบไม่ต่อเนื่องใช้สำหรับการแปลงจากค่าพลังงานลอการิทึมเมลสเปกตรัมที่อยู่ในโดเมนความถี่มาเป็นโดเมนเวลา โดยค่าสัมประสิทธิ์ที่ได้จากการแปลงโคซายน์แบบไม่ต่อเนื่องคือค่าสัมประสิทธิ์เซปสตรีม โดยค่าสัมประสิทธิ์เซปสตรีมที่ได้มีทั้งหมด 26 ค่า แสดงได้ดังรูปที่ 4.8 แต่ค่าสัมประสิทธิ์ลำดับที่ 2-13 เท่านั้นที่ถูกนำมาใช้ ค่าสัมประสิทธิ์ทั้งหมด 12 ค่า แสดงได้ดังรูปที่ 4.9



รูปที่ 4.8 ค่าสัมประสิทธิ์เซปสตรีมทั้งหมด 26 ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

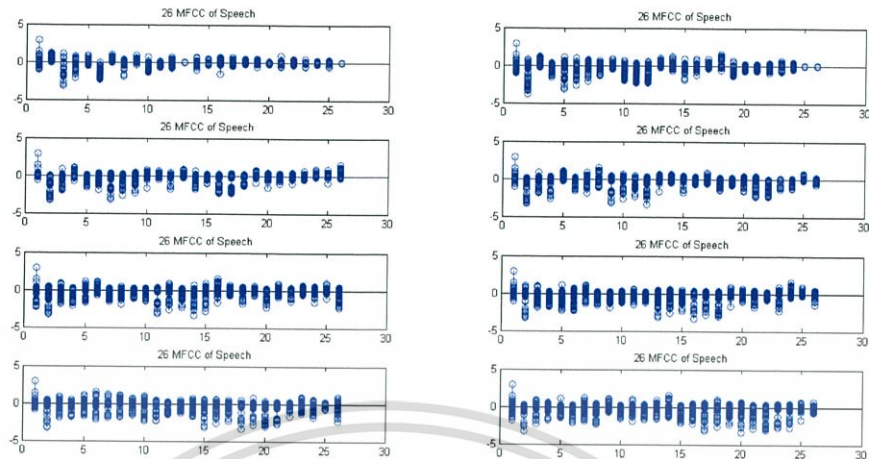


รูปที่ 4.9 ค่าสัมประสิทธิ์เซปสตรัม 12 ค่า จากตำแหน่งที่ 2 - 13

8) การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม

การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัมเป็นการคำนวณเพื่อเพิ่มค่าสัมประสิทธิ์สำหรับเฟรมย่อย โดยค่าสัมประสิทธิ์พลังงานนั้นเป็นการหาค่าพลังงานจากเฟรมเสียงย่อยที่ผ่านกระบวนการแอมมิงวินโดว์ ได้เป็นค่าสัมประสิทธิ์หนึ่งค่าออกมา ส่วนการหาสัมประสิทธิ์เดลต้าเซปสตรัมเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรัมระหว่างเฟรม จากการคำนวณนี้จะได้ค่าสัมประสิทธิ์อีก 12 ค่า และค่าสัมประสิทธิ์ตัวที่ 26 สามารถหาได้จากการคำนวณอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์พลังงานระหว่างเฟรมซึ่งเรียกว่าเดลต้าพลังงาน

จากขั้นตอนทั้งหมดจะได้ค่าสัมประสิทธิ์จากการสกัดลักษณะของเฟรมเสียงย่อยทั้งหมด 26 ค่า (26 MFCC) ซึ่งประกอบไปด้วยค่าสัมประสิทธิ์เซปสตรัม 12 ค่า ค่าสัมประสิทธิ์พลังงาน 1 ค่า ค่าสัมประสิทธิ์เดลต้าเซปสตรัม 16 ค่าและค่าเดลต้าพลังงานอีก 1 ค่า โดยค่าสัมประสิทธิ์ 26 ค่าของแต่ละเฟรมย่อยของสัญญาณเสียงหนึ่งเสียง แสดงได้ดังรูปที่ 4.10



รูปที่ 4.10 สัมประสิทธิ์ 26 ค่าจากอัลกอริทึม MFCC

4.2.1.2 ระบบโครงข่ายประสาทเทียม (Neural network)

ระบบโครงข่ายประสาทเทียมเป็นระบบที่นำมาใช้ในการรู้จำเสียงพูด เพื่อทำการแยกคำว่า ‘เปิด’ ออกจากคำอื่นๆ เมื่อเสียงผ่านกระบวนการ MFCC แล้วจะได้ค่าสัมประสิทธิ์ทั้งหมด 1404 ค่า เพื่อนำมาเป็นอินพุตของโครงข่ายประสาทเทียมที่ได้ทำการออกแบบไว้ ระบบจะแสดง ‘allow’ สำหรับการพูดคำว่า ‘เปิด’ และระบบจะแสดง ‘don’t allow’ สำหรับการพูดคำอื่นๆ ผลการทดสอบระบบสามารถแสดงได้ดังรูปที่ 4.11 การทดสอบระบบได้ทำการทดสอบทั้งหมด 100 ครั้ง โดยที่ 50 ครั้งแรกจะใช้คำว่า ‘เปิด’ ในการทดสอบ อีก 50 คำเป็นคำอื่นๆ โดยที่ผลการทดลองสามารถแสดงได้ดังตารางที่ 4.1 และตารางแสดงประสิทธิภาพของระบบสามารถแสดงได้ดังตารางที่ 4.2 ค่าความไว (Sensitivity), ค่าความจำเพาะ (Specificity), ค่าทำนายผลบวก (Positive predictive value) และค่าทำนายผลลบ (Negative predictive value) สามารถหาได้จากสมการที่ (2.22) (2.23) (2.24) และ (2.25) ตามลำดับ

>> testing	>> testing
allow	Dont allow
allow	Dont allow
allow	Dont allow
allow	Dont allow
allow	Dont allow

รูปที่ 4.11 ผลการทดสอบระบบแยกคำพูด เมื่ออินพุตเป็นสัมประสิทธิ์คำว่า ‘เปิด’ (ด้านซ้าย) และเมื่ออินพุตเป็นสัมประสิทธิ์ของคำอื่นๆ (ด้านขวา)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการทดสอบระบบโดยใช้คำ 100 คำ

คำที่ใช้ในการทดสอบ	ระบบอนุญาต (%)	ระบบไม่อนุญาต (%)	รวม (%)
‘เปิด’	96	4	100
คำอื่นๆ	4	96	100

จากตารางที่ 4.1 จะเห็นได้ว่าเมื่อพูดคำว่า ‘เปิด’ ระบบสามารถแยกได้ 96% และเมื่อพูดคำอื่นๆ ระบบก็สามารถแยกได้ 96% เช่นกัน

ตารางที่ 4.2 แสดงประสิทธิภาพของระบบจำแนกคำพูด

	Condition positive	Condition negative	
Test outcome positive	True positive (TP) = 96	False positive (FP) = 4	Positive predictive value 96%
Test outcome negative	False negative (FN) = 4	True negative (TN) = 96	Negative predictive value 96%
	Sensitivity 96%	Specificity 96%	

จากตารางที่ 4.1 นำมาหาประสิทธิภาพของระบบได้ดังตาราง 4.2 ค่าความไวของระบบ ค่าความจำเพาะ ค่าทำนายผลบวก และค่าทำนายผลลบ เท่ากับ 96% นั่นถือว่าระบบมีประสิทธิภาพที่ดี สามารถใช้ในการจำแนกคำพูดได้

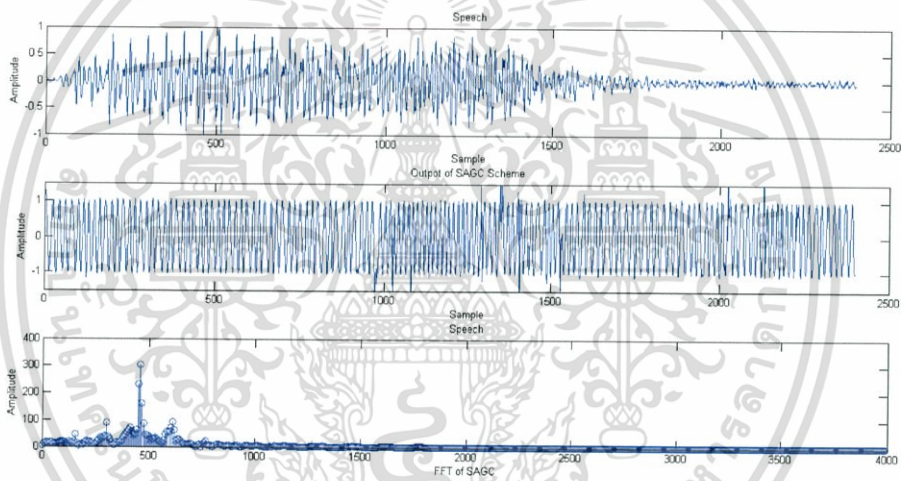
4.2.2 ระบบการจำแนกบุคคล

4.2.2.1 การสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้ในการจำแนกบุคคลโดยอาศัยขั้นตอนของ SAGC

การสกัดคุณลักษณะของสัญญาณเสียงพูดเพื่อใช้ในการจำแนกบุคคลโดยอาศัยขั้นตอนของ SAGC จากขั้นตอนนี้มีการหาค่าสัมประสิทธิ์ 3 ชุด โดยนำเอาที่พูดของ SAGC ทำการแปลงฟูเรียร์แบบไม่ต่อเนื่อง 1 ค่า การแปลงฟูเรียร์แบบไม่ต่อเนื่องของกรอบสัญญาณของพิทช์ที่ผ่านตัวกรองความถี่ต่ำผ่านเพื่อหาค่าสัมประสิทธิ์ที่มีขนาดสูงที่สุดอีก 1 ค่า การแปลงฟูเรียร์แบบ

ไม่ต่อเนื่องของกรอบสัญญาณของพิทช์ 512 ค่าโดยผลการทดลองในแต่ละขั้นตอนสามารถแสดงได้ดังต่อไปนี้

1) การแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณเอาท์พุทของ SAGC เสียงพูดที่ถูกตัดส่วนที่ไม่ใช่เสียงพูดออกจะถูกเตรียมโดยการกรองด้วยตัวกรองความถี่ต่ำผ่านสำหรับใช้เป็นอินพุทของ SAGC สัญญาณเอาท์พุทของ SAGC จะถูกแปลงฟูเรียร์แบบไม่ต่อเนื่องเพื่อหาสัมประสิทธิ์ที่มีขนาดสูงที่สุดซึ่งเป็นตัวกำหนดระดับเสียงสูงต่ำของสัญญาณเสียง และเก็บเป็นค่าสัมประสิทธิ์ไปใช้ในระบบโครงข่ายประสาทเทียมเพื่อจำแนกบุคคลต่อไป ผลการแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณเอาท์พุทของ SAGC แสดงได้ดังรูปที่ 4.12 แถบรูปเส้นแรกแสดงสัญญาณเสียงพูดที่ถูกตัดส่วนที่ไม่ใช่เสียงพูดออก แถบรูปเส้นที่สองแสดงเอาท์พุทของ SAGC และแถบรูปเส้นที่สามแสดงสเปกตรัมของสัญญาณของเอาท์พุทจาก SAGC

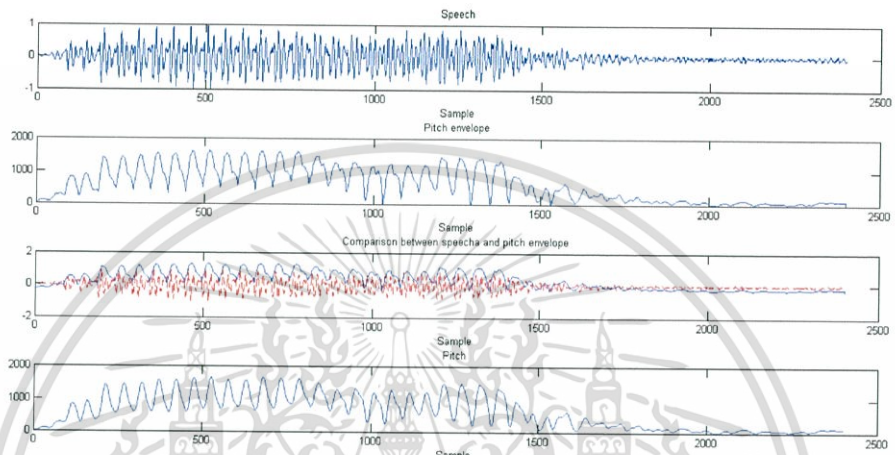


รูปที่ 4.12 การแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณเอาท์พุทของ SAGC

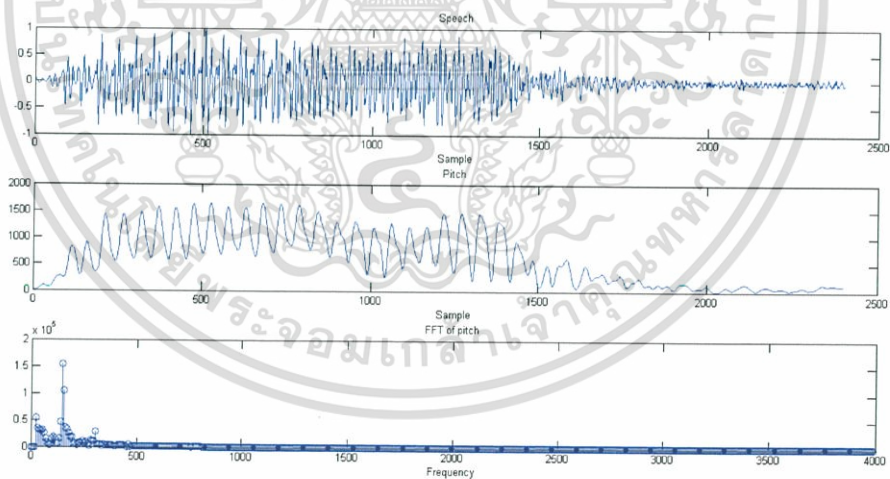
2) การแปลงฟูเรียร์แบบไม่ต่อเนื่องของพิทช์

จากการทำ SAGC สัญญาณเอาท์พุทที่ได้จากขั้นตอนการหาค่าราคาที่สองคือกรอบสัญญาณของพิทช์ (Pitch envelope) โดยสัญญาณนี้จะถูกกรองผ่านตัวกรองความถี่ต่ำผ่าน จากนั้นทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องเพื่อหาค่าความถี่ที่มีขนาดสูงที่สุดซึ่งจะเป็นตัวแทนของความถี่ของพิทช์ เพื่อเก็บเป็นค่าสัมประสิทธิ์ไปใช้ในระบบโครงข่ายประสาทเทียมเพื่อจำแนกบุคคลเช่นกัน กรอบสัญญาณของพิทช์ (Pitch envelope) แสดงได้ดังรูปที่ 4.13 แถบรูปเส้นแรกแสดงสัญญาณเสียงพูดที่ถูกตัดส่วนที่ไม่ใช่เสียงพูดออก แถบรูปเส้นที่สองแสดงกรอบสัญญาณของพิทช์ของสัญญาณเสียงพูด แถบรูปเส้นที่สามแสดงการเปรียบเทียบระหว่างสัญญาณเสียงพูดกับ

กรอบสัญญาณของพิทช์ และแถบรูปเส้นที่สี่แสดงคาบของพิทช์ (Pitch period) ส่วนผลการแปลงฟูเรียร์แบบไม่ต่อเนื่องของพิทช์ แสดงได้ดังรูปที่ 4.14 แถบรูปเส้นแรกแสดงสัญญาณเสียงพูดที่ถูกตัดส่วนที่ไม่ใช่เสียงพูดออก แถบรูปเส้นที่สองแสดงคาบของพิทช์ (Pitch period) ของสัญญาณเสียง และแถบรูปเส้นสุดท้ายแสดงผลการแปลงฟูเรียร์แบบไม่ต่อเนื่องของพิทช์



รูปที่ 4.13 กรอบสัญญาณของพิทช์

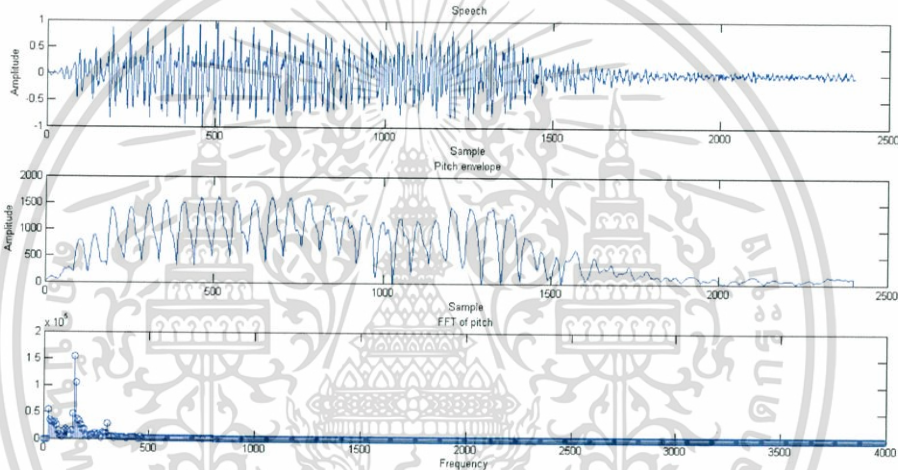


รูปที่ 4.14 การแปลงฟูเรียร์แบบไม่ต่อเนื่องของพิทช์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การแปลงฟูเรียร์แบบไม่ต่อเนื่องของกรอบสัญญาณของพิทช์ (Pitch envelope) ของสัญญาณเสียงพูด

จากการทำ SAGC สัญญาณเอาท์พุทที่ได้จากขั้นตอนการหาค่ารากที่สองซึ่งคือกรอบสัญญาณของพิทช์ (Pitch) ของสัญญาณเสียงพูด สัญญาณนี้จะถูกแปลงฟูเรียร์แบบไม่ต่อเนื่อง เพื่อนำค่าสเปกตรัมของสัญญาณไปใช้เป็นสัมประสิทธิ์อีกชุดหนึ่งของระบบโครงข่ายประสาทเทียม ผลการแปลงฟูเรียร์แบบไม่ต่อเนื่องของกรอบสัญญาณของพิทช์ แสดงได้ดังรูปที่ 4.15 แลบริปูเส้นแรกแสดงสัญญาณเสียงพูดที่ถูกตัดส่วนที่ไม่ใช่เสียงพูดออก แลบริปูเส้นที่สองแสดงกรอบสัญญาณของพิทช์ และแลบริปูเส้นที่สามแสดงสเปกตรัมของกรอบสัญญาณของพิทช์

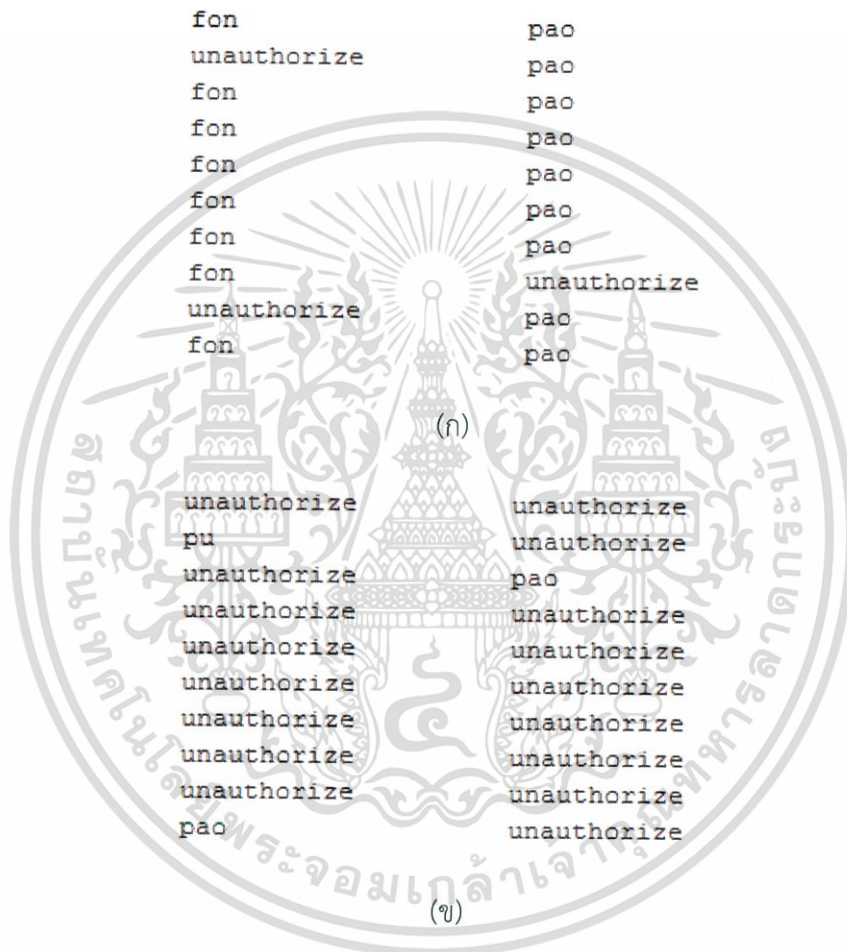


รูปที่ 4.15 การแปลงฟูเรียร์แบบไม่ต่อเนื่องของกรอบสัญญาณของพิทช์

4.2.2.2 ระบบโครงข่ายประสาทเทียม (Neural network)

ระบบโครงข่ายประสาทเทียมเป็นระบบที่นำมาใช้ในการรู้จำเสียงพูด เพื่อทำจำแนกบุคคลที่มีสิทธิ์และไม่มีสิทธิ์ใช้งานออกจากกัน เมื่อเสียงผ่านกระบวนการ SAGC และ ทำ FFT และหาตำแหน่งของค่าความถี่ที่มีขนาดของสัมประสิทธิ์ FFT ที่มากที่สุดเรียบร้อยแล้ว จะได้ค่าสัมประสิทธิ์ทั้งหมด 3 ชุด เพื่อนำมาเป็นอินพุตของโครงข่ายประสาทเทียมที่ได้ทำการออกแบบไว้ เมื่อทำการรวมระบบการตัดสินใจหลักและระบบการตัดสินใจรองที่ 1 และ 2 เข้าด้วยกัน ระบบจะแสดงชื่อผู้ที่เป็นเจ้าของเสียง สำหรับเสียงของผู้ที่ถูกอนุญาตให้ใช้ระบบได้ และระบบจะแสดง 'unauthorize' สำหรับเสียงของผู้ที่ไม่ถูกอนุญาตให้ใช้ระบบได้ ผลการทดสอบระบบสามารถแสดงได้ดังรูปที่ 4.15 การทดสอบระบบได้ทำการทดสอบทั้งหมด 100 ครั้ง โดยที่ 50 เสียงแรกจะใช้เสียง

ของผู้ที่มีสิทธิ์ใช้งานในการทดสอบ ตัวอย่างผลการทดสอบระบบสามารถแสดงได้ดังรูปที่ 4.16(ก) อีก 50 เสียงจะใช้เสียงของผู้ที่ไม่มีสิทธิ์ใช้งานในการทดสอบ ผลการทดสอบระบบสามารถแสดงได้ดังรูปที่ 4.16(ข) โดยที่ผลการทดลองสามารถแสดงได้ดังตารางที่ 4.3 และตารางแสดงประสิทธิภาพของระบบสามารถแสดงได้ดังตารางที่ 4.4 ค่าความไว ค่าความจำเพาะ ค่าทำนายผลบวก และค่าทำนายผลลบ สามารถหาได้จากสมการที่ (2.20) (2.21) (2.22) และ (2.23) ตามลำดับ



รูปที่ 4.16 ผลการทดสอบระบบจำแนกบุคคล

(ก) ตัวอย่างผลของอินพุตเป็นเสียงที่อนุญาตให้สามารถเข้าใช้ระบบได้ 2 คน

fon (ด้านซ้าย) และ pao (ด้านขวา)

(ข) ตัวอย่างผลที่เป็นอินพุตของเสียงที่ไม่ถูกอนุญาตให้เข้าใช้ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ผลการทดสอบระบบโดยใช้เสียง 100 เสียง

ประเภทบุคคล	ระบบอนุญาต (%)	ระบบไม่อนุญาต (%)	รวม (%)
มีสิทธิ์	60	40	100
ไม่มีสิทธิ์	18	82	100

จากตารางที่ 4.3 เมื่อเสียงที่เข้ามาเป็นเสียงของผู้ที่มีสิทธิ์ ระบบอนุญาต 60% และเมื่อเสียงที่เข้ามาเป็นเสียงของผู้ที่ไม่มีสิทธิ์ ระบบไม่อนุญาต 82%

ตารางที่ 4.4 แสดงประสิทธิภาพของระบบจำแนกบุคคล

	Condition positive	Condition negative	
Test outcome positive	True positive (TP) = 60	False positive (FP) = 18	Positive predictive value 77%
Test outcome negative	False negative (FN) = 40	True negative (TN) = 82	Negative predictive value 67%
	Sensitivity 60%	Specificity 82%	

จากตารางที่ 4.4 สามารถแสดงค่าประสิทธิภาพของการจำแนกบุคคล โดยค่าความไวบ่งบอกว่าระบบจำแนกได้ไวพอที่จะจับความแตกต่างที่เกิดขึ้นเพียงเล็กน้อยได้ดีเพียงใด โดยระบบมีค่าความไว 60% ถือว่าอยู่ในระดับที่พอใช้ได้ ค่าความจำเพาะบ่งบอกการจำแนกบุคคลที่ไม่มีสิทธิ์ได้ดีเพียงใด ระบบมีค่าความจำเพาะ 82% ถือว่าอยู่ในเกณฑ์ที่ดี ค่าทำนายผลบวกบ่งบอกผลบวกที่ได้จากระบบนั้นแสดงว่าเป็นผู้ที่มีสิทธิ์จริงๆ ได้ดีเพียงใด ในที่นี้คือ 77% ซึ่งถือว่าอยู่ในเกณฑ์ที่พอใช้ได้ และค่าทำนายผลลบบ่งบอกว่าผลลบที่ได้จากระบบนั้นแสดงว่าเป็นผู้ที่ไม่มีสิทธิ์จริงๆ ได้ดีเพียงใด ค่าทำนายผลลบอยู่ที่ 67% ถือว่าระบบอยู่ในระดับที่พอใช้ได้

เมื่อสร้างสองระบบ (ระบบจำแนกคำพูด และระบบจำแนกบุคคล) เสร็จเรียบร้อย จึงได้นำทั้งสองระบบมารวมเข้าด้วยกันแล้วทำการทดสอบระบบรวมโดยใช้เสียงที่พูดคำว่า 'เปิด' จากผู้ที่มีสิทธิ์ 50 เสียงและจากผู้ที่ไม่มีสิทธิ์ 50 เสียง ผลการทดสอบรวมสามารถแสดงได้ดังตารางที่ 4.5

ตารางที่ 4.5 แสดงประสิทธิภาพของการรวมระบบจำแนกคำพูด และระบบจำแนกบุคคล

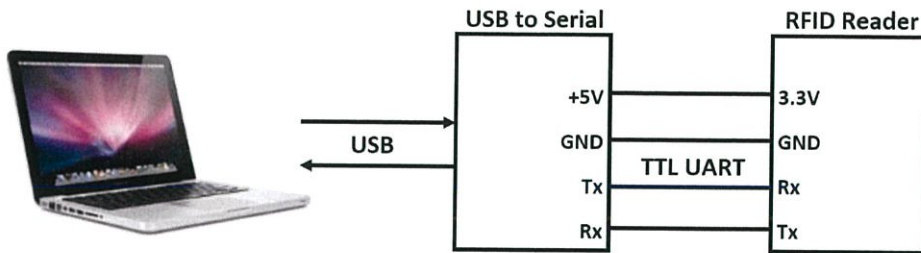
	Condition positive	Condition negative	
Test outcome positive	True positive (TP) = 52	False positive (FP) = 12	Positive predictive value 81%
Test outcome negative	False negative (FN) = 48	True negative (TN) = 88	Negative predictive value 65%
	Sensitivity 52%	Specificity 88%	

จากตารางที่ 4.5 ภาพรวมของระบบอยู่ในเกณฑ์ที่ดีพอใช้ สามารถใช้ในการจำกัดบุคคลที่มีสิทธิ์สั่งการระบบได้ ค่าความจำเพาะ และค่าทำนายผลบวกมีค่าที่ตีนั้นคือ 88% และ 81% ตามลำดับ ค่าความไว และค่าผลลบลงอยู่ในเกณฑ์ที่พอใช้ มีค่าเท่ากับ 52% และ 65% ตามลำดับ จากค่าประสิทธิภาพเหล่านี้บ่งบอกว่าระบบสามารถตรวจจับบุคคลที่มีสิทธิ์ได้ไม่ผิดนัก แต่สามารถป้องกันการเข้าถึงการควบคุมจากบุคคลที่ไม่มีสิทธิ์ได้ดีพอสมควร

4.3 การควบคุมการเปิด-ปิดประตู โดยระบบอาร์เอฟไอดี

4.3.1 การทดลองอ่านข้อมูลของตัวแท็ก โดยเครื่องอ่านอาร์เอฟไอดี

ขั้นตอนการทดลองอ่านข้อมูลของตัวแท็ก ขั้นแรกนำเครื่องอ่านอาร์เอฟไอดีเชื่อมต่อกับคอมพิวเตอร์ โดยใช้โปรแกรมทดสอบบนคอมพิวเตอร์พกพา แต่บอร์ดไม่สามารถเชื่อมต่อกับคอมพิวเตอร์พกพา ได้โดยตรง ต้องต่อวงจรเพิ่มแปลงระดับสัญญาณเป็น Serial (RS-232) เข้าพีซี สำหรับพอร์ต Serial RS-232 หรือต่อกับ USB to Serial (TTL) ในการทดลองนี้เชื่อมต่อกับ USB to Serial ดังรูปที่ 4.17

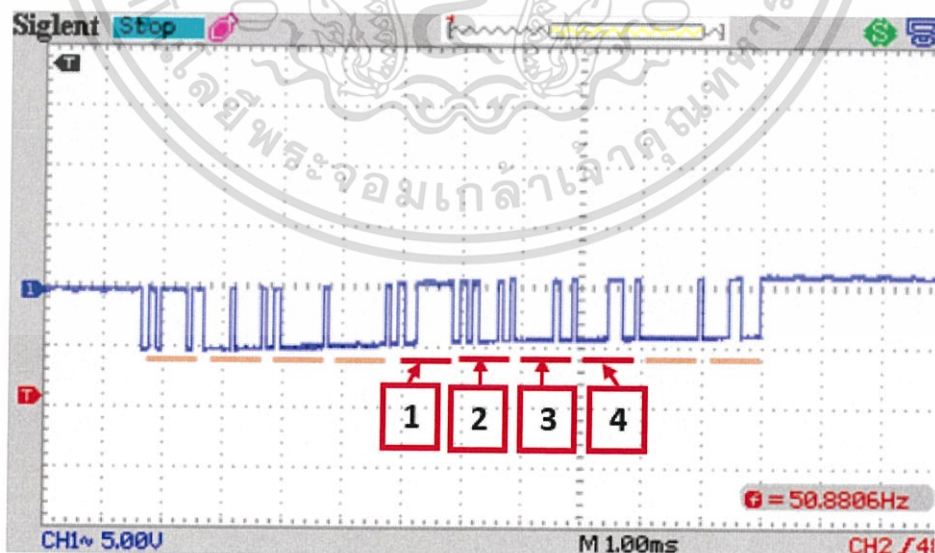


รูปที่ 4.17 การเชื่อมต่อระหว่างคอมพิวเตอร์พกพา และ เครื่องอ่านอาร์เอฟไอดี

จากนั้นทำการเชื่อมต่อบอร์ดอาร์เอฟไอดีกับ USB to Serial โดยใช้ 4 เส้นคือ

1. ขาไฟ +5V ของ USB to Serial ต่อเข้ากับ ขา VCC ของบอร์ดอาร์เอฟไอดี
2. ขากราวด์ต่อร่วมกัน
3. ขา Tx USB to Serial ต่อเข้ากับขา Rx บอร์ดอาร์เอฟไอดี
4. ขา Rx USB to Serial ต่อเข้ากับขา Tx บอร์ดอาร์เอฟไอดี

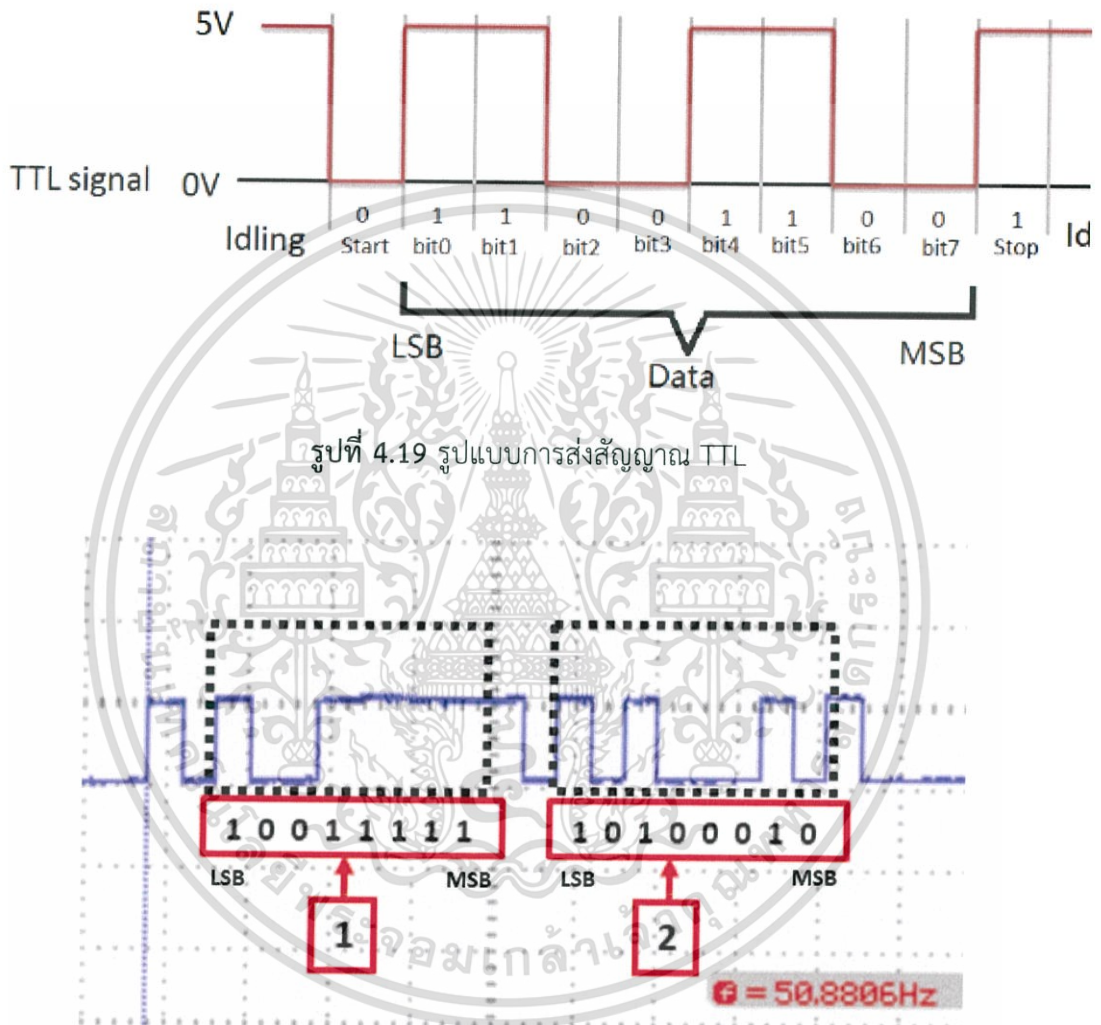
เมื่อเชื่อมต่อโมดูลเข้ากับ USB to Serial เรียบร้อยแล้ว จากนั้นทำการทดลองอ่านข้อมูลของตัวแท็ก โดยการวัดสัญญาณข้อมูลที่ส่งผ่านจากเครื่องอ่าน RFID เข้ามายังขา Rx USB to Serial ซึ่งมีการสื่อสารด้วยโปรโตคอล TTL UART ซึ่งชุดข้อมูลมีขนาดทั้งหมด 10 ไบต์ ประกอบไปด้วย Start Header 1 ไบต์, Length 1 ไบต์, Command 1 ไบต์, Status 2 ไบต์, UID 4 ไบต์ และ Checksum 1 ไบต์ เมื่อแตะแท็กที่เครื่องอ่านอาร์เอฟไอดีจะได้สัญญาณดังแสดงในรูปที่ 4.18 โดยหมายเลข 1 ถึง 4 คือชุดข้อมูล UID ที่จะทำการอ่านค่า



รูปที่ 4.18 สัญญาณข้อมูลเมื่อแตะแท็กที่เครื่องอ่านอาร์เอฟไอดี

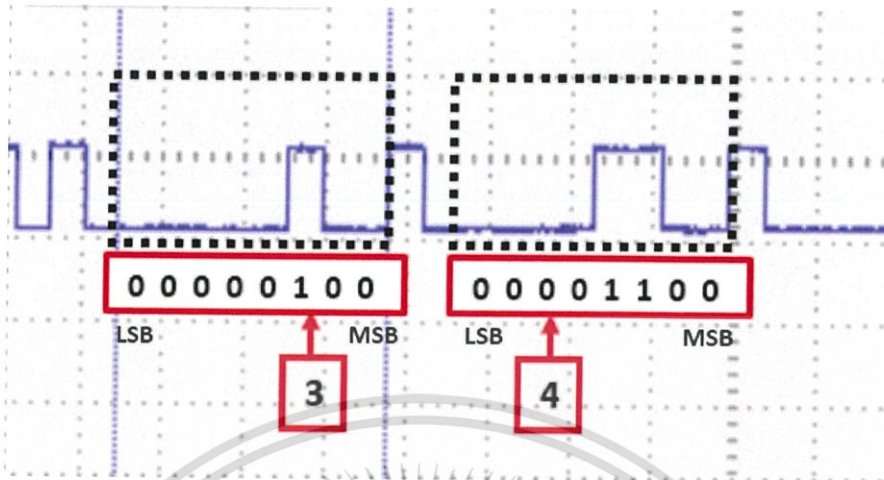
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยวิธีการอ่านชุดข้อมูล UID ของสัญญาณ TTL โปรโตคอล UART จะอ่านเป็นเลขฐานสอง โดยเริ่มที่บิตเริ่มต้นคือ บิต 0 ตามด้วยบิตข้อมูลอีก 8 บิต (การอ่านบิตข้อมูลจะอ่านจากทางด้านขวาไปซ้าย) และจบที่บิตสิ้นสุดคือบิต 1 โดยรูปแบบการอ่านสัญญาณดังกล่าวแสดงดังรูปที่ 4.19 จากนั้นทำการอ่านข้อมูล UID ของบล็อกหมายเลข 1 ถึง 4 ในรูปที่ 4.18 ซึ่งแสดงได้ดังรูปที่ 4.20 และ 4.21



รูปที่ 4.20 การอ่านบล็อกข้อมูล UID ที่ 1 และ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.21 การอ่านบล็อกข้อมูล UID ที่ 3 และ 4

การอ่านชุดข้อมูลของแต่ละบล็อกให้อ่านเป็นเลขฐานสองจากทางด้านขวามาทางด้านซ้าย จากนั้นแปลงเป็นเลขฐานสิบหก และสุดท้ายจะทำการเก็บค่าในรูปแบบเลขฐานสิบ ซึ่งจากบล็อกที่ 1 ถึง 4 สามารถแปลงค่าได้ดังนี้

บล็อกที่ 1 : 11111001 แปลงเป็นเลขฐานสิบหกเท่ากับ F9 แปลงเป็นเลขฐานสิบเท่ากับ 249

บล็อกที่ 2 : 01000101 แปลงเป็นเลขฐานสิบหกเท่ากับ 45 แปลงเป็นเลขฐานสิบเท่ากับ 69

บล็อกที่ 3 : 00100000 แปลงเป็นเลขฐานสิบหกเท่ากับ 20 แปลงเป็นเลขฐานสิบเท่ากับ 32

บล็อกที่ 4 : 00110000 แปลงเป็นเลขฐานสิบหกเท่ากับ 30 แปลงเป็นเลขฐานสิบเท่ากับ 48

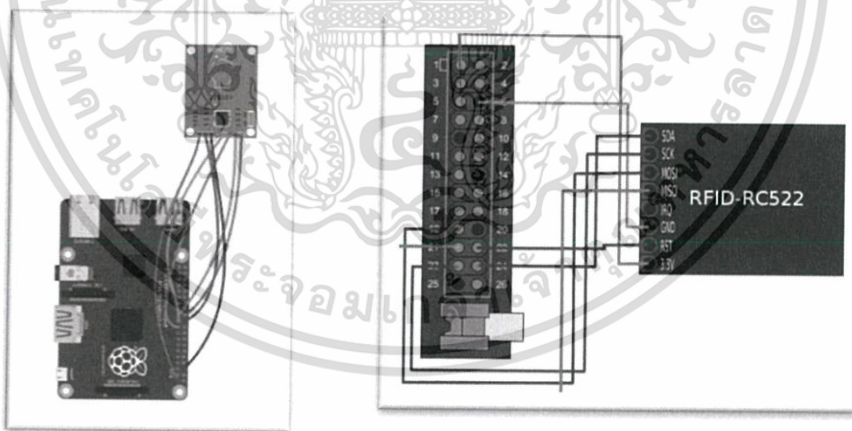
จากนั้นนำค่า UID จากบล็อกที่ 1 ถึง 4 ที่แปลงเป็นเลขฐานสิบแล้ว มาเรียงกันจะได้ 249693248 และเนื่องจากจำนวนหลักที่เป็นไปได้มากที่สุดเมื่อแปลงเป็นเลขฐานสิบคือ 12 หลัก ดังนั้นหากค่า UID ที่แปลงแล้วมีจำนวนหลักไม่ครบ 12 หลัก ให้เพิ่มเลข 0 เข้าไปจนครบเพื่อให้ค่า UID ทั้งหมดอยู่ในรูปแบบเดียวกัน จากนั้นเราจะนำค่า UID ที่ได้ไปบันทึกลงในฐานข้อมูลดังรูปที่ 14.22 เพื่อทำการเข้าถึงการเปิด-ปิดประตูต่อไป

	Name	id_voice	id_card
<input type="checkbox"/> Edit Copy Delete	FON	1	249693248000
<input type="checkbox"/> Edit Copy Delete	MUEY	2	191382091280
<input type="checkbox"/> Edit Copy Delete	PAO	3	185214164800
<input type="checkbox"/> Edit Copy Delete	PREAW	4	256444202000
<input type="checkbox"/> Edit Copy Delete	PU	5	731613048000

รูปที่ 4.22 ฐานข้อมูลที่มีการบันทึกค่า UID

4.3.2 การติดต่อสื่อสารกันระหว่างเครื่องอ่านอาร์เอฟไอดี กับ ราวสเบอร์รี่พาย

การติดต่อสื่อสารกันระหว่างเครื่องอ่านอาร์เอฟไอดี กับ ราวสเบอร์รี่พาย จะติดต่อกันผ่านระบบบัส SPI ส่งและรับข้อมูลที่ละบิต (Bit Serial) และใช้สัญญาณนาฬิกา เป็นตัวกำหนดจังหวะการทำงาน ซึ่งจะมีการเชื่อมต่อระหว่างเครื่องอ่านอาร์เอฟไอดี กับพอร์ต GPIO บนราวสเบอร์รี่พายดังรูปที่ 4.23



รูปที่ 4.23 การเชื่อมต่อระหว่างเครื่องอ่านอาร์เอฟไอดีกับราวสเบอร์รี่พาย

เมื่อทำการต่อพอร์ตและอุปกรณ์ที่จำเป็นต่อราวสเบอร์รี่พายเรียบร้อยแล้ว จะเป็นขั้นตอนอ่านและเก็บค่าไอดีจากตัวแท็ก โดยการนำแท็กไปแตะที่เครื่องอ่านอาร์เอฟไอดี จากนั้นอ่านค่าไอดีของแท็ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่ปรากฏ และนำไปเขียนโปรแกรมเก็บข้อมูล โดยใช้ระบบปฏิบัติการลินุกซ์ บนราสเบอร์รี่พายที่ หน้าเทอมินอล ดังรูปที่ 4.24 เพื่อสั่งการให้ระบบสามารถระบุตัวตนเพื่อใช้เปิดหลอดแอลอีดี

```

GNU nano 2.2.6 File: rfiddl.py
GPIO.setmode(GPIO.BOARD)
GPIO.setup(37,GPIO.OUT)
# Card Register
card1 = '1852141648'
card2 = '19138209128'

#GPIO setup
continue_reading = True

def end_read(signal,frame):
    global continue_reading
    print "Ctrl+C captured, ending read."
    continue_reading = False
    GPIO.cleanup()

# Hook the SIGINT
if UIDcode == card1:
    adminpriv = 1
else:
    adminpriv = 0

if UIDcode == card1 or UIDcode == card2 :
    if locked == '0' or adminpriv == 1:
        GPIO.setup(11, GPIO.OUT)
        print "Door open"
        GPIO.output(37,1)
        time.sleep(3)
        GPIO.setup(11, GPIO.IN)
        GPIO.output(37,0)
        print "Finished"
    else:
        print "Door locked"
    elif UIDcode == lockcard:
        counter = 0

```

รูปที่ 4.24 คำสั่งในการกำหนดหมายเลขไอดี ให้สามารถเปิดหลอดแอลอีดีได้

จากนั้นเมื่อกำหนดค่าไอดีเรียบร้อยแล้ว นำ card 1, card 2 และ card 3 ทำการแตะที่เครื่องอ่านอาร์เอฟไอดี เพื่อทดลองเปิดหลอดไฟแอลอีดี ซึ่งจากการทดลองจะพบว่า card 1 และ card 2 ที่ได้ทำการเก็บค่าไอดีไว้แล้วสามารถเปิดหลอดไฟแอลอีดี ส่วน card 3 ไม่สามารถเปิดหลอดไฟได้ ดังรูปที่ 4.25

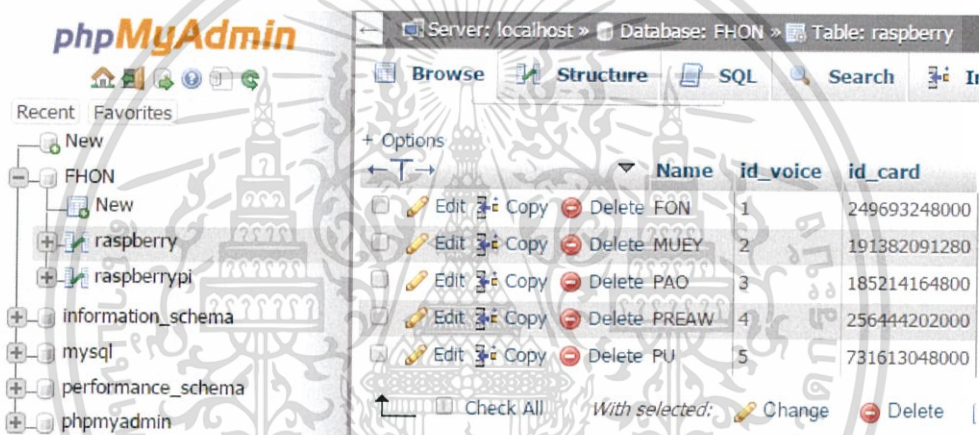


รูปที่ 4.25 ผลการทดลองการเปิดหลอดแอลอีดีโดยใช้การ์ดทั้งสามใบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3 การสร้างฐานข้อมูลบนราสเบอร์รี่พาย

ฐานข้อมูลบนราสเบอร์รี่พาย มีระบบปฏิบัติการบนพื้นฐานของลินุกซ์ทำงานเช่นเดียวกับเว็บเซิร์ฟเวอร์ ในกรณีนี้เราจะสร้างฐานข้อมูลโดยใช้ phpMyAdmin ซึ่งเป็นโปรแกรมที่ใช้ในการจัดการฐานข้อมูล MySQL ผ่านทางเบราว์เซอร์ในราสเบอร์รี่พาย โดยฐานข้อมูลที่สร้างขึ้นแสดงดังรูปที่ 4.26 จะเห็นได้ว่าในการสร้างฐานข้อมูลนี้ ได้สร้างตารางชื่อ raspberry จากนั้นทำการตั้งชื่อทั้งหมดห้าชื่อในฟิลด์ Name กำหนดหมายเลขไอดีการ์ดที่แตกต่างกันทั้งหมดห้าหมายเลขในฟิลด์ id_card และกำหนดไอดีของเสียงห้าค่าในฟิลด์ Id_voice ซึ่งหากไอดีของแท็กที่นำมาตรวจสอบตรงกับข้อมูลในฐานข้อมูล ราสเบอร์รี่พายจะดึงข้อมูลที่เหลือที่เกี่ยวข้อง แสดงผลออกมาทางหน้าจอ LCD ด้วย



รูปที่ 4.26 ฐานข้อมูลที่สร้างขึ้น

และเมื่อมีการเข้าใช้งานประตู ระบบจะทำการบันทึกข้อมูลผู้ใช้งาน โดยการนำหมายเลขไอดีของแท็ก หรือหมายเลขไอดีของเสียง ไปเปรียบเทียบกับรายชื่อที่มีอยู่ในฐานข้อมูลในตาราง raspberry ที่อยู่ในรูป 4.26 จากนั้นนำชื่อที่เปรียบเทียบได้แล้วไปบันทึกในฟิลด์ User ของตาราง Store ใน phpMyAdmin ดังตารางที่ 4.27 จากนั้นตั้งวันที่และเวลาขณะตรวจสอบได้ในขณะนั้นไปบันทึกในฟิลด์ Time เพื่อสามารถตรวจสอบผู้เข้าใช้งานย้อนหลังได้

Server: localhost » Database: FHON » Table: Store

Options: Browse, Structure, SQL, Search, Insert

	Number	User	Time
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	1	FON	2017-04-16 10:13:46
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	2	MUEY	2017-04-16 10:18:25
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	3		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	4		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	5		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	6		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	7		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	8		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	9		
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	10		

รูปที่ 4.27 ตารางสำหรับบันทึกรายชื่อผู้ใช้งานประตู

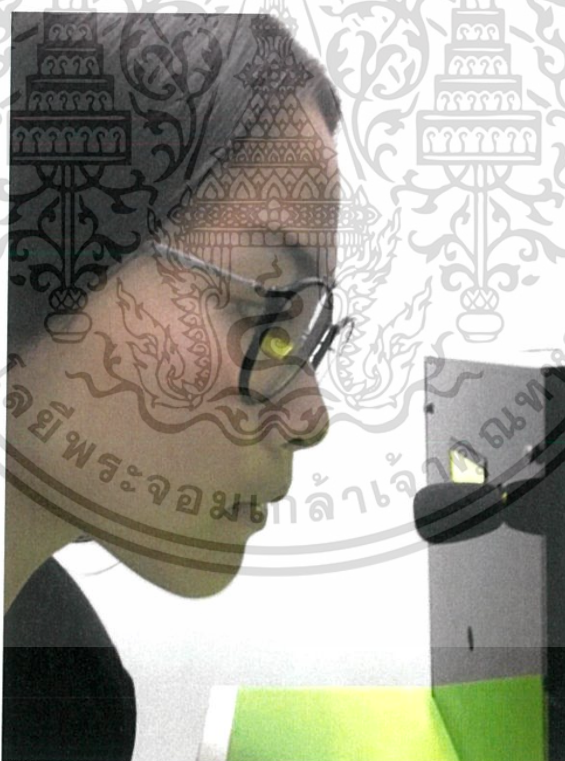
4.4 ระบบรวม (ระบบรู้จำเสียง และระบบ RFID)

ระบบรวมเกิดจากการนำระบบรู้จำเสียง และระบบ RFID มารวมเข้าด้วยกัน เพื่อให้สามารถใช้งานได้ทั้งสองระบบ โดยที่ระบบรู้จำเสียงจะใช้เป็นระบบหลักในการเข้าใช้ระบบ และมีระบบ RFID เป็นระบบเสริม หากระบบรู้จำเสียงเกิดปัญหาหรือขัดข้องไม่สามารถใช้งานได้

ในการทดสอบ จะใช้ประตูที่เปิดปิดด้วยแถบแม่เหล็ก ในสถานะปกติ ประตูจะถูกแม่เหล็กดูดเปรียบเสมือนประตูล็อกอยู่ การติดตั้งแม่เหล็กที่ประตูสามารถแสดงได้ดังรูปที่ 4.28 ระบบจะทำงานก็ต่อเมื่อมีการใช้เสียงสั่งการ โดยการสั่งการจะต้องมีเงื่อนไขว่าผู้สั่งการจะต้องพูดคำว่า 'เปิด' ที่ไมโครโฟน ดังรูปที่ 4.29 เมื่อผู้สั่งการเป็นบุคคลที่มีสิทธิ์เข้าใช้งานระบบ และพูดคำว่าเปิด จะส่งผลให้ราสเบอร์รี่จ่ายแรงดันขนาด 2.8 โวลต์ให้แก่วงจรควบคุมประตูแม่เหล็ก เมื่อมีแรงดันเข้ามาวงจรจะส่งแรงดันไปเลี้ยงแม่เหล็ก สามารถแสดงการจับสัญญาณของทั้งสองสัญญาณโดยใช้ออสซิลโลสโคปได้ดังรูปที่ 4.30 โดยที่ช่องสัญญาณที่ 1 คือการแสดงผลสัญญาณที่ควบคุมการเปิดปิดของประตูเทียบกับ ช่องสัญญาณที่ 2 ระดับความต่างศักย์ไฟฟ้าที่ใช้ในการเปิดปิดประตู และจอ LCD จะแสดงชื่อของผู้สั่งการ สามารถแสดงรูปผลของจอ LCD ได้ดังรูปที่ 4.31 และสามารถแสดงการเปิดประตูได้ดังรูปที่ 4.32

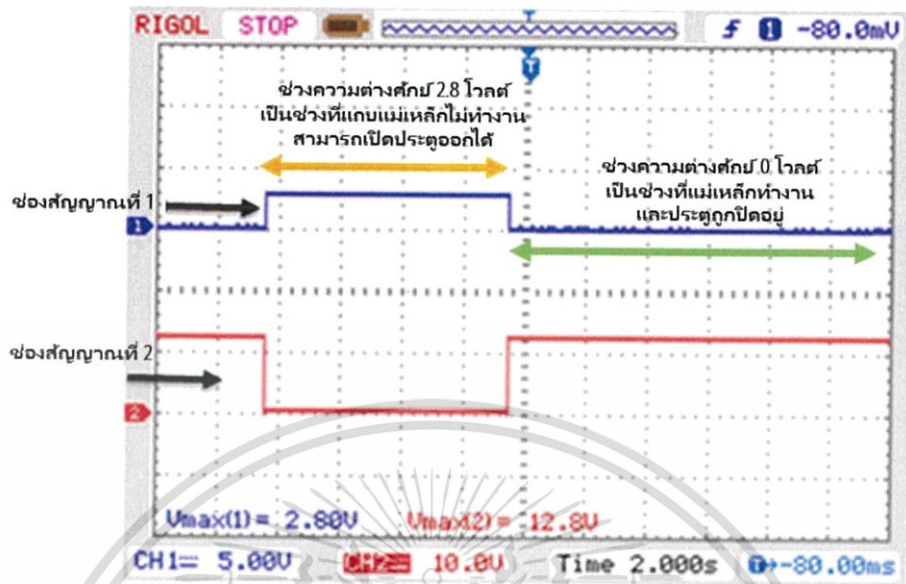


รูปที่ 4.28 การติดตั้งแม่เหล็กที่ประตูเพื่อเป็นตัวล็อค

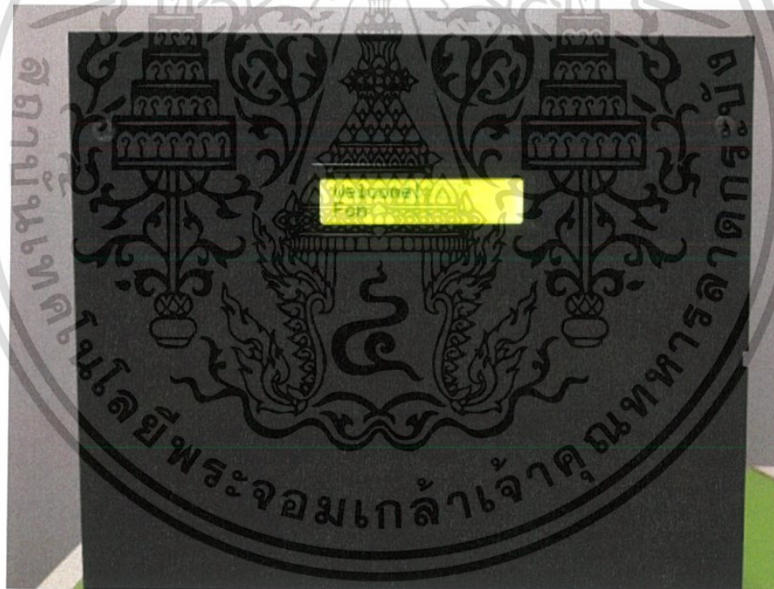


รูปที่ 4.29 วิธีการใช้งานระบบรู้จำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

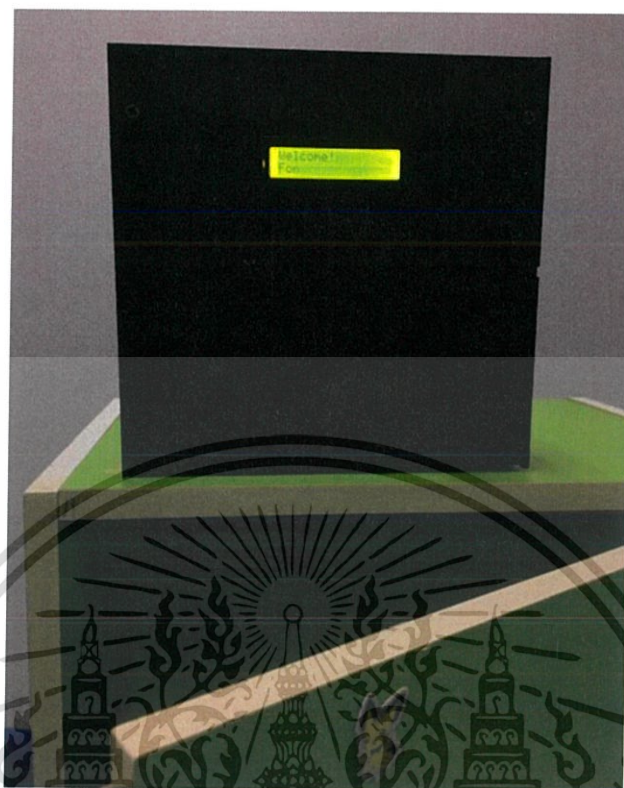


รูปที่ 4.30 การทำงานของวงจรควบคุมการทำงานของแม่เหล็กโดยการสั่งการผ่านระบบรู้จำเสียง



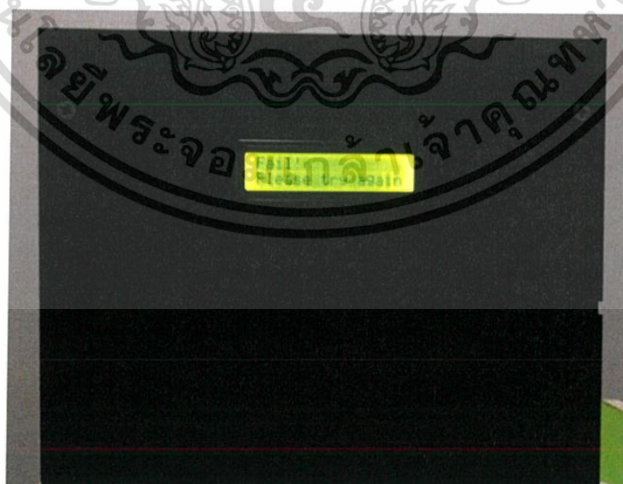
รูปที่ 4.31 จอ LCD แสดงชื่อผู้สั่งการที่อยู่ในระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.32 ไม่มีไฟเลี้ยงแม่เหล็ก ประตูสามารถเปิดออกได้

และเมื่อผู้สั่งการเป็นบุคคลที่ไม่มีสิทธิ์เข้าใช้งาน ระบบจะสามารถตรวจจับและไม่อนุญาตให้บุคคลผู้นั้นสามารถเปิดประตูได้ และจอ LCD จะขึ้นข้อความดังรูปที่ 4.33



รูปที่ 4.33 จอ LCD ขึ้นข้อความให้สั่งการใหม่

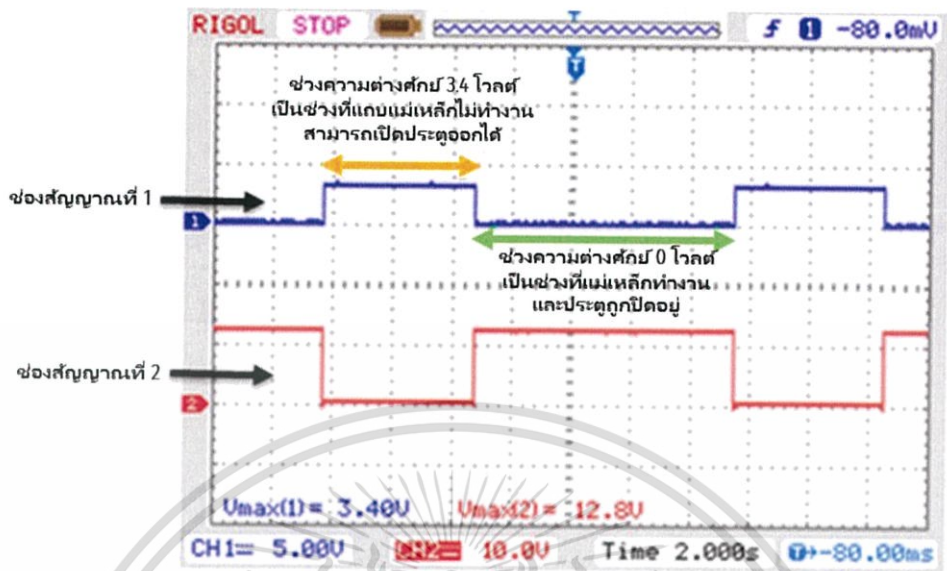
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลองระบบสั่งการด้วยการรู้จำเสียง พบว่ายังมีความผิดพลาดอยู่ เนื่องจากขณะที่ผู้สั่งการพูด เสียงรอบข้างอาจจะมีแอมพลิจูดที่สูงกว่า ทำให้เกิดเสียงรบกวน (noise) ซึ่งอาจจะส่งผลให้ระบบมีการประมวลผลที่ผิดพลาด หากเป็นเช่นนั้นแล้วจะทำให้ไม่สามารถสั่งการด้วยเสียงได้ จึงได้มีการใช้ระบบสำรองนั่นคือระบบ RFID เมื่อแตะแท็ก RFID ดังรูปที่ 4.34 หากเป็นบัตรที่ถูกบันทึกอยู่ในฐานข้อมูล ราสเบอร์รี่พายจะประมวลผลและส่งสัญญาณเอาต์พุตไปสั่งการกลไกการปลดล็อกประตู ตัดไฟออกจากชุดแม่เหล็กดังรูปที่ 4.35 แสดงให้เห็นถึงความสัมพันธ์ของช่วงเวลาในการเปิดปิดประตูในช่องสัญญาณที่ 2 ซึ่งสอดคล้องกับเอาต์พุตของราสเบอร์รี่พาย ในช่องสัญญาณที่ 1 และจอแสดงผลจะแสดงชื่อของเจ้าของบัตรที่เป็นผู้มีสิทธิ์เข้าใช้ระบบดังรูปที่ 4.36 และประตูจะเปิดออก หากใช้แท็ก RFID ที่ไม่ได้ถูกบันทึกลงในฐานข้อมูล หน้าจอจะแสดงข้อความให้ลองอีกครั้ง ดังรูปที่ 4.37 และประตูจะไม่เปิดออก

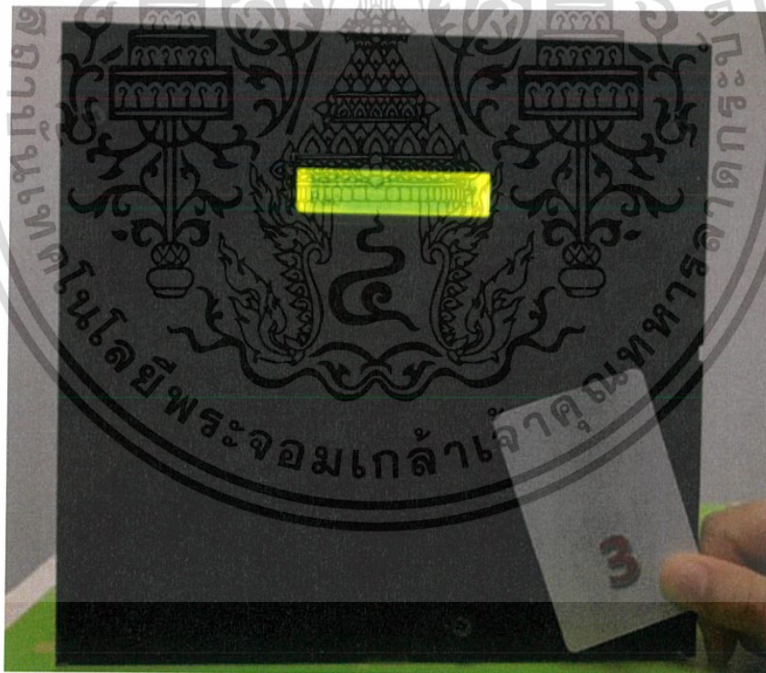


รูปที่ 4.34 วิธีการใช้งานระบบ RFID

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

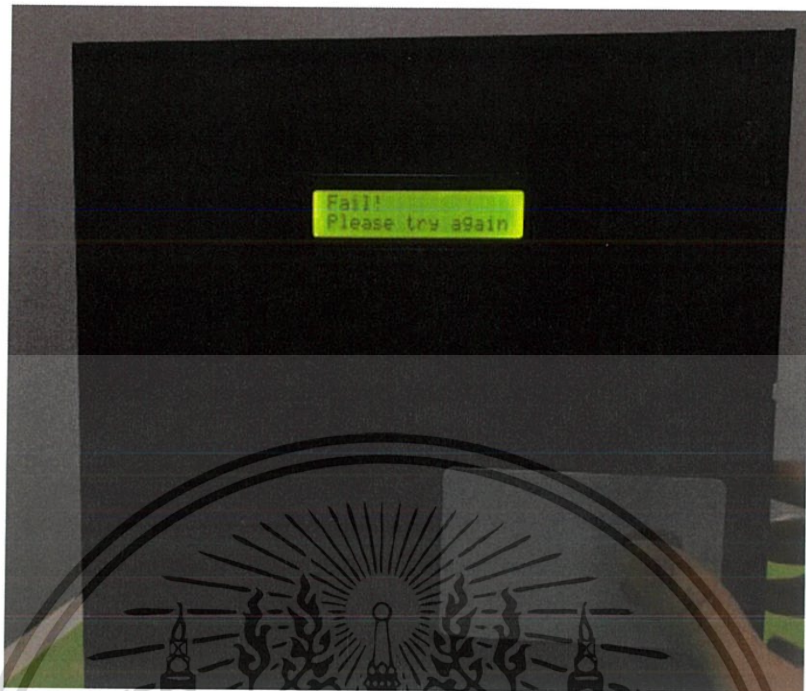


รูปที่ 4.35 การทำงานของวงจรควบคุมการทำงานของแม่เหล็กโดยการสั่งการผ่านระบบ RFID



รูปที่ 4.36 จอแสดงชื่อเจ้าของแท็ก RFID

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.37 จอแสดงผลสำหรับ RFID ที่ไม่ได้ถูกบันทึกในฐานข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผล

ปริญญานิพนธ์นำเสนอการนำเสียงมาใช้ในการระบุตัวตน สำหรับการเข้าถึงระบบเปิด-ปิดประตู ซึ่งเสียงของแต่ละบุคคล มีลักษณะเด่นที่แตกต่างกัน โดยหลักการรู้จำเสียงจะถูกนำมาใช้ในการจดจำลักษณะเด่นของเสียง และเพื่อรองรับการทำงานของระบบ การเข้าถึงระบบเปิด - ปิดประตูโดยใช้บัตรอาร์เอฟไอดี จะถูกนำมาใช้เป็นอีกทางเลือกหนึ่งของระบบที่นำเสนอด้วย โดยมีวัตถุประสงค์เพื่ออำนวยความสะดวกในการเปิดปิดประตู และระบุตัวตน และเป็นการป้องกันการเข้าถึงงานของผู้ที่ไม่พึงประสงค์ โดยแบ่งการเข้าถึงเป็น 2 รูปแบบ คือการเข้าถึงระบบโดยใช้การรู้จำเสียงและการใช้ระบบอาร์เอฟไอดี

1) การรู้จำเสียงประกอบด้วยสองส่วนคือการสกัดลักษณะของเสียงด้วยอัลกอริทึม MFCC สำหรับการจำแนกคำพูด ส่วนของการจำแนกบุคคลจากเสียงพูดโดยอาศัยขั้นตอนของ SAGC และการฝึกฝนการรู้จำเสียงในระบบโครงข่ายประสาทเทียม โดยลักษณะเด่นที่ได้ในขั้นตอนสกัดลักษณะเด่นของเสียงที่ถูกนำไปใช้ในการฝึกฝนระบบโครงข่ายประสาทเทียมนั้นสามารถระบุตัวบุคคลของเสียงนั้นได้

2) ระบบควบคุมการเข้าถึงการเปิด-ปิดประตูโดยระบบอาร์เอฟไอดี โดยใช้ไอดีแท็กในการเข้าถึง โดยเครื่องอ่านอาร์เอฟไอดีจะส่งค่าไอดีไปยังราสเบอร์รี่พาย เพื่อนำไอดีแท็กที่ได้รับไปตรวจสอบกับข้อมูลที่มีในฐานข้อมูล ถ้าข้อมูลที่ได้รับตรงกับข้อมูลในฐานข้อมูล ราสเบอร์รี่พายจะสั่งการเปิดประตูในลำดับต่อไป ซึ่งการเข้าถึงโดยระบบอาร์เอฟไอดี เป็นอีกทางเลือกหนึ่งของการเข้าถึงการควบคุมประตู

5.2 ปัญหาและข้อเสนอแนะ

5.2.1 จากการเลือกเก็บค่าสัมประสิทธิ์ 26 ค่า จากอัลกอริทึม MFCC สำหรับการจำแนกคำและค่าสัมประสิทธิ์สำหรับจำแนกบุคคล เมื่อนำค่าสัมประสิทธิ์นี้ไปเข้าโครงข่ายประสาทเทียม ผลลัพธ์ที่ได้ในการรู้จำเสียงยังมีความผิดพลาด เพื่อให้ค่าความผิดพลาดที่น้อยลงควรเพิ่มจำนวนสัมประสิทธิ์ของเสียงที่ใช้ในการฝึกฝนระบบโครงข่ายประสาทเทียม

5.2.2 สำหรับการใช้งานโครงข่ายประสาทเทียมควรปรับปรุงการออกแบบระบบ การเรียนรู้ (Training) ให้มีประสิทธิภาพและให้ผลลัพธ์ที่มีข้อผิดพลาดน้อยที่สุด

5.2.3 ควรมีหน้าจอ LCD เพื่อแสดงสถานะปัจจุบันของระบบ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

- [1] Wikipedia. “เสียง.”
<https://th.wikipedia.org/wiki/เสียง>
- [2] คมสันต์ วงศ์วรรณ. “คุณสมบัติของเสียง.”
<http://musiclib.psu.ac.th/data/western-musuc/Chapter2/chap2-1.htm>
- [3] L. Muda, M. Begam, and I. Elamvazuthi. 2010. “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques,” Journal of Computing. 2(3) : 138 - 143
- [4] P. Pal Singh and P. Rani. 2014. “An Approach to Extract Feature using MFCC,” IOSR Journal of Engineering (IOSRJEN). 04(08)
- [5] James Lyons. “Mel Frequency Cepstral Coefficient (MFCC) tutorial”
<http://practicalcryptography.com/miscellaneous/machine-learning/guide-melfrequency-cepstral-coefficients-mfccs/#eqn2>
- [6] Wikipedia. “Artificial neural network.”
https://en.wikipedia.org/wiki/Artificial_neural_network.
- [7] ถนัด ศรีบุญเรือง และคณะ. “ขนาดและรูปร่างของเซลล์.”
<http://www.trueplookpanya.com/learning/detail/2169-002640>.
- [8] Wikipedia. “ความไวและความจำเพาะ.”
<https://th.wikipedia.org/wiki/ความไวและความจำเพาะ>
- [9] Wikipedia. “คำทำนายผลบวก.”
<https://th.wikipedia.org/คำทำนายผลบวก>
- [10] ชีรศักดิ์ สาโนนสูง. “ระบบผ่านประตูอัตโนมัติโดยใช้เทคโนโลยี RFID.”
ปริญญาานิพนธ์,สาขาวิศวกรรมโทรคมนาคม คณะวิศวกรรมศาสตร์,สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,2552
- [11] ปริณันท์ วรรณสว่าง และคณะ. “RFIDเทคโนโลยีสารพัดประโยชน์.”
<http://www.lampangtc.ac.th/mnfile/branch5/file/knowledge/RFID.pdf>
- [12] ประทีป บัญญัติสินพรัตน์. (2548). “การสื่อสารข้อมูล.”
กรุงเทพฯ : สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง.

<http://cpre.kmutnb.ac.th/esl/learning/index.php?article=arduino-spi-master-slave>

[13] zunvoo. “Matlab คือ อะไร? ตอนที่ 1 และ ตอนที่ 2.”

<http://matabthai.blogspot.com/2013/04/matlab.html>

[14] Jurafsky, Daniel, Martin, James H.. Speech and Language Processing. 2nd ed. Upper Saddle River, NJ. : Prentice Hall, 2009.

[15] Harish Chander Mahendru. “Quick Review of Human Speech Production Mechanism”. 10 (January 2014)

[16] “Human Voice”

<http://mirlab.org/jang/books/audioSignalProcessing/humanVoiceProduction.asp?title=3-3%20Human%20Voice%20Production%20>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
noise = audioread('noisetesttwofon4.wav');
```

```
voice = audioread('testtwosa4.wav');
```

```
A = 3*max(noise);
```

```
L = length(voice);
```

```
xcut = zeros(L,1);
```

```
for i=1:L
```

```
    if (abs(voice(i))>=A)
```

```
        xcut(i) = voice(i);
```

```
    end
```

```
end
```

```
PF = find(xcut,1,'first');
```

```
PE = find(xcut,1,'last');
```

```
y = voice(PF:PE); % cut voice
```

```
kim = y;
```

```
if length(y)<2400
```

```
    kim = voice(PF:PF+2400);
```

```
end
```

```
y=y-mean(y);
```

```
xcut = kim;
```

```
if length(y)<= lengthpad
```

```
    y = y;
```

```
else y = y(1:lengthpad);
```

```
end
```

```
%%%%%%%%%%%%%% pre-emphasis
```

```
b = [1 -0.95]; a = 1;
```

```
preem = filter(b,a,y); % Pre-emphasised signal
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
%%%%%%%%%%%%%% pad
```

```
pad = zeros(lengthpad,1);
```

```
lpreem = length(preem);
```

```
for i=1:lpreem;
```

```
    if (i<=lpreem)
```

```
        pad(i)=preem(i);
```

```
    else pad(i)=pad(i);
```

```
    end
```

```
end
```

```
%%%%%%%%%%%%%% framing & hamming window
```

```
L = 255+1;
```

```
w = hamming(L);
```

```
numfram = round((lengthpad-L)/99)-1 ;
```

```
k = zeros(numfram+1,L);
```

```
% ham = zeros(numfram+1,L);
```

```
k(1,:)= pad(1:256);
```

```
ham = k(1,:).*w' ;
```

```
for N=1:numfram
```

```
    start = 100+(N-1)*99;
```

```
    stop = start+255;
```

```
    k(N+1,:) = pad(start:stop);
```

```
    ham(N+1,:) = k(N+1,:).*w' ;
```

```
end
```

```
%%%%%%%%%%%%%% DFT
```

```
for i=1:numfram+1
```

```
    x = ham(i,:);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

m = length(x); % Window length
n = pow2(nextpow2(m)); % Transform length
yyy = fft(x,n); % dft
f = (0:n/2-1)*(Fs/n); % Frequency range
sptr(i,:) = (abs(yyy(1:128)));
end

```

```

%%%%%%%%%%%%%% Mel-Frequency

```

```

nfft = 256;
lower = 300;
upper = Fs/2;
hfs = upper;
mN = 28;

melf = zeros(1,mN);
convf = zeros(1,mN);
fi = zeros(1,mN);
filbk = zeros(mN-2,nfft/2);

mf_upper = 1125*log(1+ (upper/700));
mf_lower = 1125*log(1+ (lower/700));

mf_gap = (mf_upper-mf_lower)/(mN - 1);

melf(1) = mf_lower;
convf(1) = lower;

fi(1)=floor(((nfft+1)*convf(1))/Fs);

fVal = Fs*[0:nfft/2-1]/nfft;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

for i=2:mN
    melf(i) = melf(i-1) + mf_gap;
    convf(i) = 700*(exp(melf(i)/1125)-1);
    fi(i)=floor(((nfft+1)*convf(i))/Fs);
end

nFb = length(melf);
H = zeros(nFb-2,nfft/2);

for k=2:nFb-1
    for j=1:nfft/2
        if (j < fi(k-1)),
            H(k,j) = 0;
        elseif ((j >= fi(k-1)) && (j <= fi(k))),
            H(k,j) = ( j - fi(k-1) ) / (fi(k) - fi(k-1));
        elseif (( j >= fi(k) ) && ( j <= fi(k+1) )),
            H(k,j) = ( fi(k+1) - j ) / ( fi(k+1) - fi(k) );
        elseif ( j > fi(k+1) ),
            H(k,j) = 0;
        end
    end
end
end

for i=2:27
    filbk(i-1,:) = H(i,:);
end

flfrm = zeros(54,128);

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

%%%%%%%%%%%%%% filter bank
for jj = 1:54
    for kk = 1:26
        fltbked = sptr(jj,:).*filbk(kk,:);
        sumspect = sum(fltbked);
        cep(jj,kk) = (sumspect)^2;
    end
end

ceplog = log10(cep);

for ii=1:54
    aa = sum(cep(ii,:));
    if (aa == 0),
        cep(ii,:) = 1;
    end
end

ceplog = log10(cep);

%%%%%%%%%%%%%% cepstral coefficient
for jj = 1:54
    mcep(jj,:) = dct(ceplog(jj,:),26);
    CT_KW(jj,:) = mcep(jj,(2:13));    % บันทึกค่าสเปส.ที่ได้สำหรับหนึ่งไฟล์เสียง
end

%%%%%%%%%%%%%% DELTA CEPSTRAL COEFFICIENT
y = CT_KW;
prep_del = zeros(1,12);
prep_del = [prep_del; y;];
bN = size(prepare_del);

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

N = bN(1);
DT_KW= zeros(bN(1)-1,bN(2));

for i=2:N-1

    DT_KW(i-1,:) = (prep_del(i+1,:) - prep_del(i-1,:)) /2;

end

%%%%%% energy
for i=1:54

    hh = ham(i,:);
    %energy(i) = sum(hh.^2);
    energy(i) = (hh(1).^2) + (hh(2).^2);
    energy = energy';

end

%%%%%% delta energy
prep_energy = zeros(1);
prep_energy = [prep_energy; energy;];
bN = size(prepare_energy);

N = bN(1);
DT_energy = zeros(bN(1)-1,bN(2));

for i=2:N-1

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
DT_energy(i-1,:) = (prep_energy(i+1,:) - prep_energy(i-1,:)) / 2;
```

```
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% 26 cepstrum
```

```
aa = CT_KW; bb=energy; cc=DT_KW; dd=DT_energy;
```

```
coeff26= [aa, bb, cc, dd];
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% merge MFCC
```

```
for m=1:54
```

```
    if (m == 1);
```

```
        stt = coeff26(m,:);
```

```
    else yy = coeff26(m,:);
```

```
        stt = [stt yy];
```

```
    end
```

```
end
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

n = 0.3;
testvoic = xcut;
testvoic = testvoic(1:n*8000);
testvoic = testvoic';
testvoic = testvoic-mean(testvoic);
testvoic = testvoic/(max(abs(testvoic)));

lp_cutoff = 500;
[bl,al] = butter(4, lp_cutoff/(Fs/2), 'low');
pre = filter(bl, al, testvoic);

tc = linspace(0,n,n*8000);
fdsig=diff(pre)./diff(tc); % [1] first derivative
% fdsig(length(testvoic))=0;
fdsig=[fdsig 0]; % [2] pad the last bit of the first
derivative
sdsig=diff(pre,2)./(diff(tc(1:length(tc)-1))).^2; % [3] second derivative
sdsig=[sdsig 0 0]; % [4] pad the last bit of the
second derivative
xsqfd=fdsig.^2; % [5] squaring the first derivative
xmul=pre.*sdsig; % [6] multiplying between signal and
second derivative
y=xsqfd-xmul; % [7] the difference of [5]and [6]
sq=sqrt(y); % [8] squaroot of [7]
sagc=fdsig./sq; % [9] [1] divided by [8]
rsq = real(sq);
rsagc =real(sagc);

%%%%%%%%%%%% DFT
nn = 1024;
fftrsagc = fft(rsagc,nn); % dft

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

f = (0:nn/2-1)*(Fs/nn);           % Frequency range
sptrsagc = (abs(fftrsagc(1:nn/2)));
maxval = max(sptrsagc);
[r,c] = find(sptrsagc == maxval);
posi_sagc = f(c);

```

%%%%%%%%%%%%%% LPF to envelope

```

lp_cutoff = 200;
[B,A] = butter(4, lp_cutoff/(Fs/2), 'low');
enve=filter(B,A,rsq);           %%% output from squaroot
nn = 1024;
ffttenve = fft(enve,nn);
f = (0:nn/2-1)*(Fs/nn);         % Frequency range
sptr_enve = (abs(ffttenve(1:nn/2)));
sptr_enve(1)=0;
sptr_enve(2)=0;
sptr_enve(3)=0;
maxval = max(sptr_enve);
[r1,c1] = find(sptr_enve == maxval);
posi_enve = f(c1);
cc = posi_sagc;
dd = posi_enve;

```

%%%%%%%%%%%%%% DFT sq

```

nn = 1024;
fftrsqsq = fft(rsqsq,nn);       % dft
f = (0:nn/2-1)*(Fs/nn);         % Frequency range
sptrrsqsq = (abs(fftrsqsq(1:nn/2)));

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

x = input_fai_nn'; % input from MFCC for training system
t = target';

% Choose a Training Function
trainFcn = 'trainbr'; % Bayesian Regularization backpropagation.

% Create a Fitting Network
hiddenLayerSize = 5;
net = fitnet(hiddenLayerSize,trainFcn);

% Setup Division of Data for Training, Validation, Testing
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% Train the Network
[net,tr] = train(net,x,t);

% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y) % Show the performance of system

Code testing neural network
clear all
close all
load('Network_complete.mat'); % Load trained network

for i=1:10
    test = coeff26_W_mos(i,:);
    n = transpose(test);

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
a = sim(net,n); % Test system
Y = round(a);

% Decision
if Y<=2
disp('allow')
end

if Y>=3
disp('Dont allow')
end
end
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

clear all
close all
load('Network_complete.mat'); % Load trained network

for i=1:10
    test = coeff26_W_mos(i,:);
    n = transpose(test);
    a = sim(net,n); % Test system
    Y = round(a);

    % Decision
    if Y<=2
        disp('allow')
    end
    if Y>=3
        disp('Dont allow')
    end
end
end

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
x = input_nn'; % input from position of peak FFT for training system
t = target';
```

```
% Choose a Training Function
```

```
trainFcn = 'trainbr'; % Bayesian Regularization backpropagation.
```

```
% Create a Fitting Network
```

```
hiddenLayerSize = 30;
```

```
net = fitnet(hiddenLayerSize,trainFcn);
```

```
% Setup Division of Data for Training, Validation, Testing
```

```
net.divideParam.trainRatio = 70/100;
```

```
net.divideParam.valRatio = 15/100;
```

```
net.divideParam.testRatio = 15/100;
```

```
% Train the Network
```

```
[net,tr] = train(net,x,t);
```

```
% Test the Network
```

```
y = net(x);
```

```
e = gsubtract(t,y);
```

```
performance = perform(net,t,y) % Show the performance of system
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

name = 'preaw'; % Name of voice
for i=1:10; % 10 inputs
    n = eval(strcat('test_',name));
    n = n(i,:);
    n = n';
    % Network of authorize perseon
    Y1 = sim(net_fon,n);
    Y2 = sim(net_muey,n);
    Y3 = sim(net_pao,n);
    Y4 = sim(net_preaw,n);
    Y5 = sim(net_pu,n);

    % Network of unauthorize perseon
    Y6 = sim(net_unauthorize,n);

    % Decision
    a = max([Y1,Y2,Y3,Y4,Y5,Y6]);
    if a == Y1
        disp 'fon'
    end

    if a == Y2
        disp 'muey'
    end

    if a == Y3
        disp 'pao'
    end

    if a == Y4
        disp 'preaw'
    end

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
end

if a == Y5
    disp 'pu'
end

if a == Y6
    disp 'unauthorize'
end

end
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้