ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

โปรแกรมควบคุมอุปกรณ์ไฟฟ้าด้วยเสียงเพื่อคนพิการ

PROGRAM ELECTRICAL EQUIPMENT CONTROLLED VIA VOICE FOR PERSONS WITH DISABILITIES

H006584

นพดล จึงเจริญสุข

รังสิมาภรณ์ เพียรทำดี

เลขหมู่.....
เลขทะเบียน 06584
วัน, เดือน, ปี 27 ก.ค. 2555

b.....
i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2553

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โปรแกรมควบคุมอุปกรณ์ไฟฟ้าด้วยเสียงเพื่อคนพิการ

PROGRAM ELECTRICAL EQUIPMENT CONTROLLED VIA VOICE FOR PERSONS WITH DISABILITIES

โดย

นพดล จึงเจริญสุข

รังสิมาภรณ์ เพียรทำดี

อาจารย์ที่ปรึกษา

ดร. ปานวิทย์ ธุวะนุติ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาคเรียนที่ 2 ปีการศึกษา 2553

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PROGRAM ELECTRICAL EQUIPMENT CONTROLLED VIA VOICE FOR PERSONS WITH DISABILITIES

NOPPADOL CHUNGCHAROENSUK

RANGSIMAPORN PIENTHUMDEE

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

BACHELOR OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2/2010

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2011

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใบรับรองปริญญานิพนธ์ ประจำปีการศึกษา 2553

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง โปรแกรมระบบควบคุมอุปกรณ์ไฟฟ้าด้วยเสียงเพื่อคนพิการ

PROGRAM ELECTRICAL EQUIPMENT CONTROLLED VIA VOICE FOR PERSONS WITH DISABILITIES

ผู้จัดทำ

1. นายนพดล จึงเจริญสุข รหัสประจำตัว 50070017
2. นางสาวรังสิมาภรณ์ เพียรทำดี รหัสประจำตัว 50070058

.....อาจารย์ที่ปรึกษา

(ดร. ปานวิทย์ ธุวะนุติ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อโครงงาน	โปรแกรมควบคุมอุปกรณ์ไฟฟ้าด้วยเสียงเพื่อคนพิการ
นักศึกษา	นายนพดล จึงเจริญสุข รหัสนักศึกษา 50070017
	นางสาวรังสิมาภรณ์ เพียรทำดี รหัสนักศึกษา 50070058
ปริญญา	วิทยาศาสตรบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
ปีการศึกษา	2553
อาจารย์ที่ปรึกษา	ดร. ปานวิทย์ ธุวะนุติ

บทคัดย่อ

โปรแกรมควบคุมอุปกรณ์ไฟฟ้าด้วยเสียงเพื่อคนพิการ เป็นโปรแกรมที่ช่วยเพิ่มความสะดวกสบายในการดำเนินชีวิตประจำวันให้แก่ผู้พิการ โดยปริญญานิพนธ์ฉบับนี้ได้นำเสนอแนวทางการศึกษาการควบคุมอุปกรณ์ด้วยการสั่งงานด้วยเสียงพูดภาษาไทย ซึ่งได้นำเอาโครงข่ายประสาทเทียมชนิดหลายเลเยอร์ โดยมีอัลกอริทึมการเรียนรู้แบบแบ็กพรอบพาเกชั่น มาประยุกต์ใช้ โปรแกรมจะรับคำสั่งเสียงผ่านทางไมโครโฟนที่ต่อเข้ากับคอมพิวเตอร์แล้วนำสัญญาณเสียงที่ได้มาประมวลผลกระบวนการรู้จำเสียง และส่งผลลัพธ์ออกไปควบคุมอุปกรณ์ไฟฟ้า เพื่อให้อุปกรณ์ดังกล่าวทำงานตามคำสั่งทางซีเรียลพอร์ตของคอมพิวเตอร์ โดยประสิทธิภาพของโปรแกรมจากการทดลองรู้จำเสียงซึ่งประกอบด้วย 4 คำ คำละ 120 เสียง จาก 40 คน ได้ค่าเปอร์เซ็นต์ความถูกต้องดังนี้ “เปิด” 25% “ปิด” 93% “ไฟ” 53% และ “รั้ว” 90%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Program electrical equipment controlled via voice for persons with disabilities.
Student	Noppadol Chungcharoensuk Student ID 50070017 Rangsimaporn Pienthumdee Student ID 50070058
Degree	Bachelor of Science
Programme	Information Technology
Year	2010
Advisor	Dr. Panwit Tuwanut

ABSTRACT

Program electrical equipment controlled via voice for persons with disabilities is a convenient program for a daily life. This project presents equipment controlled with Speech Recognition using Thai speech and using multilayer neural network technique with back-propagation learning algorithm apply in this program. Program will get voice command is input to the computer via microphone. The system will processing and send output is signal to control electronic equipment through serial port of computer and electronic circuit. Speech recognition experiments were performed on a 4 words and words per 120 voice from 40 speakers. An accuracy of the recognition rates of "open" is 25%, "close" is 93% , "light" is 53% and "gate" is 90%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จได้ ด้วยความกรุณาของ ดร. ปานวิทย์ ธุวะนุติ อาจารย์ที่ปรึกษาปริญญานิพนธ์ ซึ่งได้ให้คำปรึกษา ข้อชี้แนะ และความช่วยเหลือในหลายสิ่งหลายอย่างจนกระทั่งลุล่วงไปได้ด้วยดี จึงขอกราบขอบพระคุณเป็นอย่างสูงมา ณ ที่นี้

ขอกราบขอบพระคุณ คณาจารย์คณะเทคโนโลยีสารสนเทศทุกท่าน ที่ได้ให้ความรู้ตลอดจนให้คำแนะนำและกำลังใจตลอดการศึกษาที่ผ่านมา

ขอบคุณและขอบใจ พี่ และเพื่อนคณะเทคโนโลยีสารสนเทศทุกคน ที่คอยเป็นที่ปรึกษาในบางเรื่อง โดยเฉพาะเพื่อนผู้ทำงานร่วมกันในปริญญานิพนธ์ฉบับนี้

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่านที่ได้ประสิทประสาทวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ข้าพเจ้า

นพดล จึงเจริญสุข

รังสิมาภรณ์ เพียรทำดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ.....II

กิตติกรรมประกาศ.....III

สารบัญ..... IV

สารบัญรูป.....VI

สารบัญตาราง.....VIII

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาและความสำคัญของปัญหา.....1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....1

1.3 ขอบเขตของโครงงาน.....2

บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง.....3

2.1 เสียง.....3

2.2 ระบบการเปล่งเสียงของมนุษย์.....5

2.3 ทฤษฎีการรู้จำเสียงพูด (Speech Recognition).....8

บทที่ 3 การออกแบบระบบ.....19

3.1 โครงสร้างของระบบโดยรวม.....19

3.2 โครงสร้างของอุปกรณ์ฮาร์ดแวร์.....21

บทที่ 4 พัฒนาระบบ.....27

4.1 การเก็บข้อมูล.....27

4.2 กระบวนการเตรียมเสียงก่อนนำเข้าสู่กระบวนการเรียนรู้.....27

4.3 อุปกรณ์ Hardware ที่ต่อกับแผ่นปริ้นแล้ว.....31

บทที่ 5 สรุปผลโครงงาน.....33

5.1 สรุปผลโครงงาน.....33

5.2 ปัญหาและอุปสรรค.....33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

5.3 ข้อเสนอแนะ.....34

บรรณานุกรม.....35

ภาคผนวก ก. 36

ภาคผนวก ข. 51

ภาคผนวก ค. 54

ประวัติผู้เขียน.....63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่ 2.1 แสดงการขจัดของโมเลกุลอากาศขณะเกิดเสียง……..……3
รูปที่ 2.2 แสดงการสั่นสะเทือนของเสียงสูง – ต่ำ……..……4
รูปที่ 2.3 แสดงช่วงเสียง……..……..4
รูปที่ 2.4 แสดงภายอวัยวะที่ใช้ในการออกเสียง……..……...6
รูปที่ 2.5 แสดงตัวอย่างการวินโดว์เพื่อแบ่งเฟรมในการทำไปสกัดค่าลักษณะสำคัญของเสียงพูด.....9
รูปที่ 2.6 แสดงบล็อคไดอะแกรมขั้นตอนการบีบอัดเสียงโดยวิธีการประมาณพันธะเชิงเส้น (LPC)……....10
รูปที่ 2.7 แสดงโครงสร้างเซลล์ประสาทของมนุษย์……...……...13
รูปที่ 2.8 เซลล์ประสาทเทียม (Artificial Neuron Cell)……..……14
รูปที่ 2.9 Single layer artificial neural network……..……15
รูปที่ 2.10 Multilayer artificial neural network……..……15
รูปที่ 2.11 แสดงการเรียนรู้แบบมีการสอน (Supervised Learning)……..……...16
รูปที่ 2.12 แสดงการเรียนรู้แบบไม่มีการสอน Unsupervised Learning……..……16
รูปที่ 2.13 โครงข่าย Neural Network แบบสองชั้น ที่ใช้กระบวนการเรียนรู้แบบ Backpropagation……17
รูปที่ 3.1 แผนภาพแสดงขั้นตอนการเรียนรู้จำคุณลักษณะเสียงโดยใช้โครงข่ายประสาทเทียม…...19
รูปที่ 3.2 IC AT89C2051 (MCS51)……....21
รูปที่ 3.3 IC MAX232……...21
รูปที่ 3.4 IC 74HC541……...22
รูปที่ 3.5 IC Relay……....22
รูปที่ 3.6 IC XBee Pro……..……23
รูปที่ 3.7 มาตรฐานการสื่อสารแบบ ZigBee……...……...24
รูปที่ 3.8 การทำงานของ ZigBee……...……..……24
รูปที่ 3.9 ตัวอย่างแผ่นปริ้น ของฐานรองที่ต่อกับคอมพิวเตอร์……...……...25
รูปที่ 3.10 ตัวอย่างแผ่นปริ้น ของฐานรองที่ต่อกับ อุปกรณ์ไฟฟ้า……25
รูปที่ 3.11 ภาพแสดงตัวอย่างการต่ออุปกรณ์ฮาร์ดแวร์……26
รูปที่ 4.1 รูปแสดงตัวอย่างเสียงที่ถูกเก็บคำว่า “ปิด” 3 ครั้ง โดยยังไม่ผ่านกระบวนการใด……..…27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

หน้า

รูปที่ 4.2 รูปแสดงตัวอย่างเสียงที่ถูกเก็บคำว่า “ปิด” 3 ครั้งโดยผ่านกระบวนการกรองสัญญาณเสียง(Filter)……..28
รูปที่ 4.3 รูปแสดงตัวอย่างเสียง “ปิด” ที่ผ่านการ filter และผ่านการตัดเสียงแล้ว……29
รูปที่ 4.4 รูปแสดงตัวอย่างการแบ่งเฟรม (Windows)……...29
รูปที่ 4.5 รูปแสดงตัวอย่างการกัดหาคุณลักษณะของเสียงด้วย LPC ……..……30
รูปที่ 4.6 รูปแสดง Log-Sigmoid Transfer Function……..……..30
รูปที่ 4.7 รูปแสดงอุปกรณ์ตัวส่งที่ต่อกับคอมพิวเตอร์……31
รูปที่ 4.8 รูปแสดงอุปกรณ์ตัวรับที่ต่อกับอุปกรณ์ไฟฟ้า……...32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่ 5.1 สรุปผลการทดลอง.....33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน เทคโนโลยีได้ถูกพัฒนาและมีความก้าวหน้าอย่างรวดเร็วทำให้เกิดการพัฒนาและผลิตระบบต่างๆขึ้นมากมาย รวมไปถึงระบบรู้จำเสียงซึ่งได้เข้ามามีบทบาทต่อชีวิตประจำวันเพื่อเพิ่มและอำนวยความสะดวกสบายให้แก่ผู้ใช้งาน เพราะการสั่งการด้วยเสียงนั้นง่ายต่อการใช้งานสำหรับมนุษย์ผู้ซึ่งมีการติดต่อสื่อสารระหว่างกันด้วยการพูด รวมถึงผู้พิการที่ไม่สามารถใช้เครื่องใช้หรืออุปกรณ์ต่างๆในชีวิตประจำวันได้อย่างสะดวกนัก ซึ่งเป็นการเพิ่มโอกาสให้ผู้พิการได้สัมผัสถึงเทคโนโลยีต่างๆยิ่งขึ้น เนื่องจากอุปกรณ์อำนวยความสะดวกให้แก่คนพิการนั้นมีองค์กรที่ทำการพัฒนาโปรแกรมหรือสิ่งอำนวยความสะดวกต่างๆไม่มากนัก

โครงงานนี้จึงนำใช้หลักการรู้จำเสียงพูดโดยการนำเอาโครงข่ายประสาทเทียมซึ่งเป็นระบบจำลองการเรียนรู้ของสมองมนุษย์มาประยุกต์ใช้ในการศึกษารูปแบบของคุณลักษณะเด่นของคำพูด ซึ่งเทคนิคที่นำเสนอนี้ประกอบไป ด้วยส่วนสำคัญ สองส่วนได้แก่ ส่วนแรก คือ การสกัดหาคุณลักษณะเด่นของเสียงโดยใช้สัมประสิทธิ์ของการทำนายเชิงเส้น และส่วนที่สองคือ การเรียนรู้และตัดสินใจแบบโครงข่ายประสาทเทียมชนิดหลายเลเยอร์ โดยอัลกอริทึมการเรียนรู้แบบแบ็กพรอบพาเกชั่น ซึ่งหลักการดังกล่าวสามารถทำงานรวดเร็วและมีความถูกต้องในการแยกแยะคำสูง จากนั้นจึงนำผลลัพธ์ที่ได้จากโครงข่ายประสาทเทียม ส่งต่อไปยังแผงวงจรควบคุมอุปกรณ์ไฟฟ้าผ่านทางพอร์ตอนุกรม ของเครื่องคอมพิวเตอร์ แต่กระนั้นข้อจำกัดของโปรแกรมนี้คือผู้พิการต้องสามารถพูดหรือออกเสียงได้อย่างปกติและสามารถมองเห็นได้ จึงจะสามารถใช้งานโปรแกรมได้อย่างมีประสิทธิภาพ

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1) เพื่อสร้างโปรแกรมที่สามารถควบคุมการเปิดปิดของอุปกรณ์ไฟฟ้าภายในบ้านผ่านทางคอมพิวเตอร์ด้วยวิธีการสั่งการด้วยเสียงของมนุษย์
2) เพื่อศึกษากระบวนการรู้จำเสียง จากการแบบโครงข่ายประสาทเทียมชนิดหลายเลเยอร์
3) เพื่อศึกษาการส่งข้อมูลระหว่างอุปกรณ์ฮาร์ดแวร์รวมถึงการออกแบบวงจรไฟฟ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตของโครงงาน

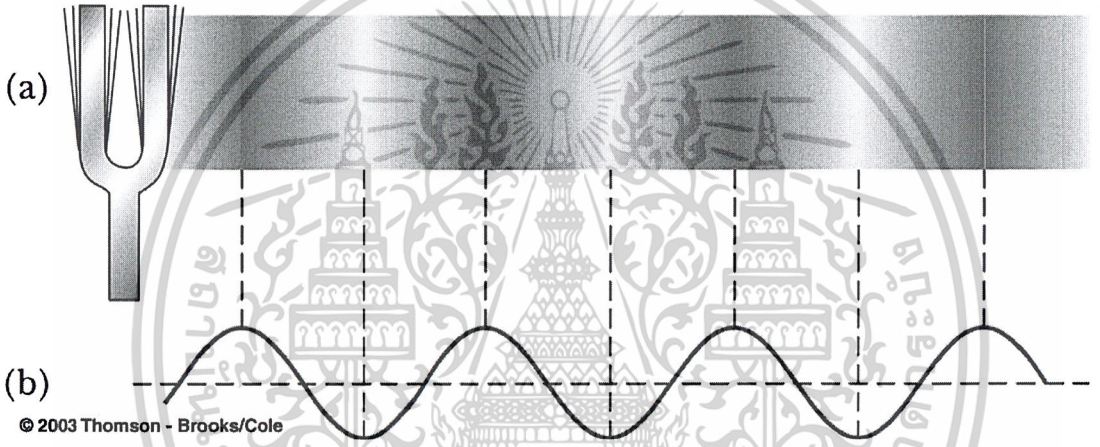
1) โปรแกรมสามารถสั่งการได้ด้วยเสียงของมนุษย์
2) โปรแกรมสามารถควบคุมอุปกรณ์ไฟฟ้าได้ โดยผ่านทางคอมพิวเตอร์
3) พัฒนาโปรแกรมด้วยภาษา Matlab
4) พัฒนาแผงวงจรที่ใช้การควบคุมอุปกรณ์ไฟฟ้า ที่เชื่อมต่อกับคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
หลักการและทฤษฎีที่เกี่ยวข้อง

2.1 เสียง (Sound)

เสียงเป็นพลังงานรูปหนึ่ง ไม่มีน้ำหนัก ไม่ต้องการที่อยู่ และไม่สามารถจับต้องได้ แต่สามารถทำงานได้ โดยเสียงเกิดจากการสั่นของวัตถุ เมื่อแหล่งกำเนิดเสียงเกิดการสั่นสะเทือน โมเลกุลของอากาศก็จะสั่นตามไปด้วยซึ่งจะทำให้เกิดพลังงานของการสั่นโดยจะแผ่ออกไปรอบๆ แหล่งกำเนิดเสียง จนกระทั้งพลังงานนั้นทำให้แก้วหูของเราสั่น เมื่อแก้วหูสั่นก็จะกระเทือนต่อเนื่องกันไปจนถึงประสาทหูเปลี่ยนเป็นสัญญาณไฟฟ้าสู่สมองทำให้มนุษย์ได้ยินเสียง



รูปที่ 2.1 แสดงการขจัดของโมเลกุลอากาศขณะเกิดเสียง

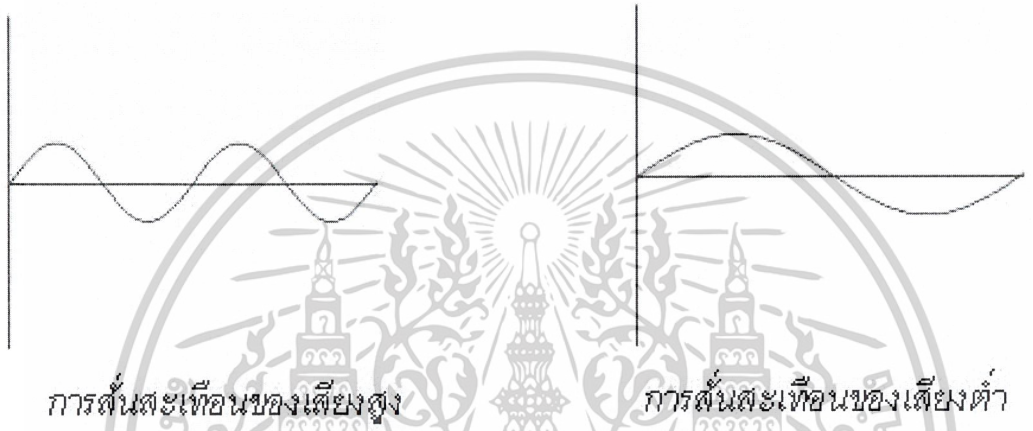
จากรูปที่ 2.1 การสั่นสะเทือนของอากาศขณะที่เกิดเสียง ทำให้การเรียงตัวของโมเลกุลอากาศเปลี่ยนแปลง ตำแหน่งโมเลกุลของอากาศที่เคลื่นที่มารวมกันอย่างหนาแน่นจะเรียกว่าช่วงอัด ส่วนโมเลกุลของอากาศที่กระจายตัวอยู่กันเบาบางจะเรียกว่าช่วงขยาย

ดังนั้นเสียงจึงคลื่นตามยาวชนิดหนึ่งซึ่งต้องอาศัยตัวกลางในการเคลื่อนที่เพราะโมเลกุลของอากาศจะสั่นในทิศทางเดียวกับทิศที่เสียงเคลื่อนที่ คุณสมบัติของเสียงจึงเหมือนคลื่นทุกประการ โดยมนุษย์จะสามารถได้ยินเสียงในช่วงความถี่ 20 เฮิร์ตซ์ (Hertz) ถึง 20,000 เฮิร์ตซ์ คลื่นเสียงที่มีความถี่สูงกว่า 20,000 เฮิร์ตซ์ นั้นเรียกว่า คลื่นเหนือเสียงหรืออัลตราโซนิก (Ultrasonic) และเสียงที่ความถี่ต่ำกว่า 20 เฮิร์ตซ์ เรียกว่าคลื่นใต้เสียงหรือ อินฟราโซนิก (Infrasonic)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

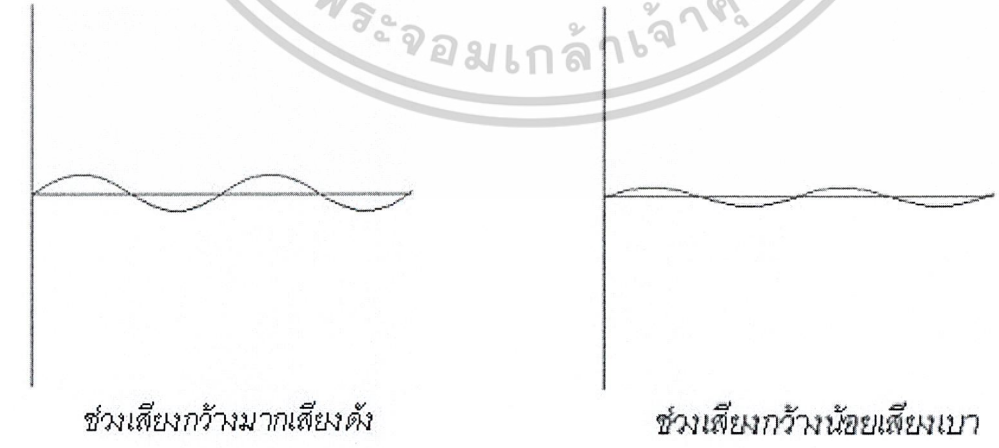
2.1.1 คุณสมบัติของเสียง

2.1.1.1) ระดับเสียง (Pitch) คือระดับความสูง – ต่ำ ของเสียง ระดับเสียงขึ้นอยู่กับความถี่ (Frequency) ของการสั่นสะเทือน เช่น ถ้าเกิดการสั่นสะเทือนอย่างเร็ว เสียงก็จะสูง แต่ถ้าเกิดการสั่นสะเทือนช้า เสียงก็จะต่ำ ความถี่ของการสั่นสะเทือนปกติจะวัดเป็น รอบต่อ วินาที (Cycles/Second) และนอกจาก วัตถุที่มีความถี่ในการสั่นสะเทือนมากกว่า จะมีเสียงที่สูงกว่าแล้ว หากความถี่มากขึ้นเท่าตัว ก็จะมีระดับเสียงสูงขึ้นเท่ากับ 1 ออกเตฟ (octave) ภาษาไทยเรียกว่า 1 ช่วงคู่แปด



รูปที่ 2.2 แสดงการสั่นสะเทือนของเสียงสูง – ต่ำ

2.1.1.2) ความดัง – ค่อย (Dynamics) ความดัง – ค่อย มีความสัมพันธ์กับช่วงกว้างของคลื่นเสียง (Amplitude) ในการสั่นสะเทือนที่ทำให้เกิดเสียงนั้นๆ ช่วงกว้างมากเสียงจะดัง และช่วงกว้างน้อยเสียงจะเบา



รูปที่ 2.3 แสดงช่วงเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การแปลงเสียงเป็นสัญญาณไฟฟ้า

เมื่อต้องการวิเคราะห์สัญญาณเสียง จำเป็นต้องแปลงสัญญาณเสียงให้อยู่ในรูปของสัญญาณไฟฟ้าก่อนโดยใช้อุปกรณ์คือ ไมโครโฟน ซึ่งหลักการทำงานของไมโครโฟน มีดังนี้

1) เมื่อคลื่นเสียงมากระทบแผ่นไดอะแฟรม (Diaphragm) ของไมโครโฟน ทำให้แผ่นไดอะแฟรมสั่นสะเทือน ขดลวดที่ติดอยู่กับแผ่นไดอะแฟรมจะสั่นสะเทือนตามไปด้วย
2) บริเวณใกล้กับขดลวดเป็นแม่เหล็กถาวร โดยขณะที่ขดลวดสั่นเข้าใกล้และสั่นออกห่างจากแม่เหล็ก จะทำให้ ฟลักซ์แม่เหล็กเปลี่ยนแปลง
3) จากกฎของฟาราเดย์ การเปลี่ยนแปลงฟลักซ์แม่เหล็กทำให้เกิดแรงเคลื่อนไฟฟ้าเหนี่ยวนำขึ้น คือสัญญาณไฟฟ้า ซึ่งเป็นสัญญาณอนาล็อค

2.2 ระบบการเปล่งเสียงของมนุษย์

เสียงที่มนุษย์เปล่งออกมานั้นเป็นเสียงที่ผลิตขึ้นโดยอวัยวะชุดหนึ่งของร่างกายโดยเฉพาะ แต่สิ่งที่น่าสนใจคือว่า มนุษย์ไม่ได้มีอวัยวะชุดใดชุดหนึ่งที่ถูกสร้างขึ้นมาเพื่อไว้ใช้ในการออกเสียงพูดโดยเฉพาะแต่เพียงอย่างเดียว แต่อวัยวะต่างๆของร่างกายที่มนุษย์ใช้ในการออกเสียงพูดในภาษานั้นมีหน้าที่หลักทางชีววิทยาเพื่อทำหน้าที่สำคัญต่างๆ ในการดำรงอยู่ของร่างกายทั้งสิ้น เช่น การหายใจ, การเคี้ยว, การกลืน เป็นต้น แต่การที่อวัยวะเหล่านั้นมาทำหน้าที่ในการออกเสียงพูดนี้เป็นเพียงหน้าที่รอง คือ ใช้ระบบการทำงานต่างๆ อยู่ แล้วดัดแปลงการทำงานเหล่านั้นให้เหมาะสมจนเกิดเป็นภาษาพูดขึ้นมา

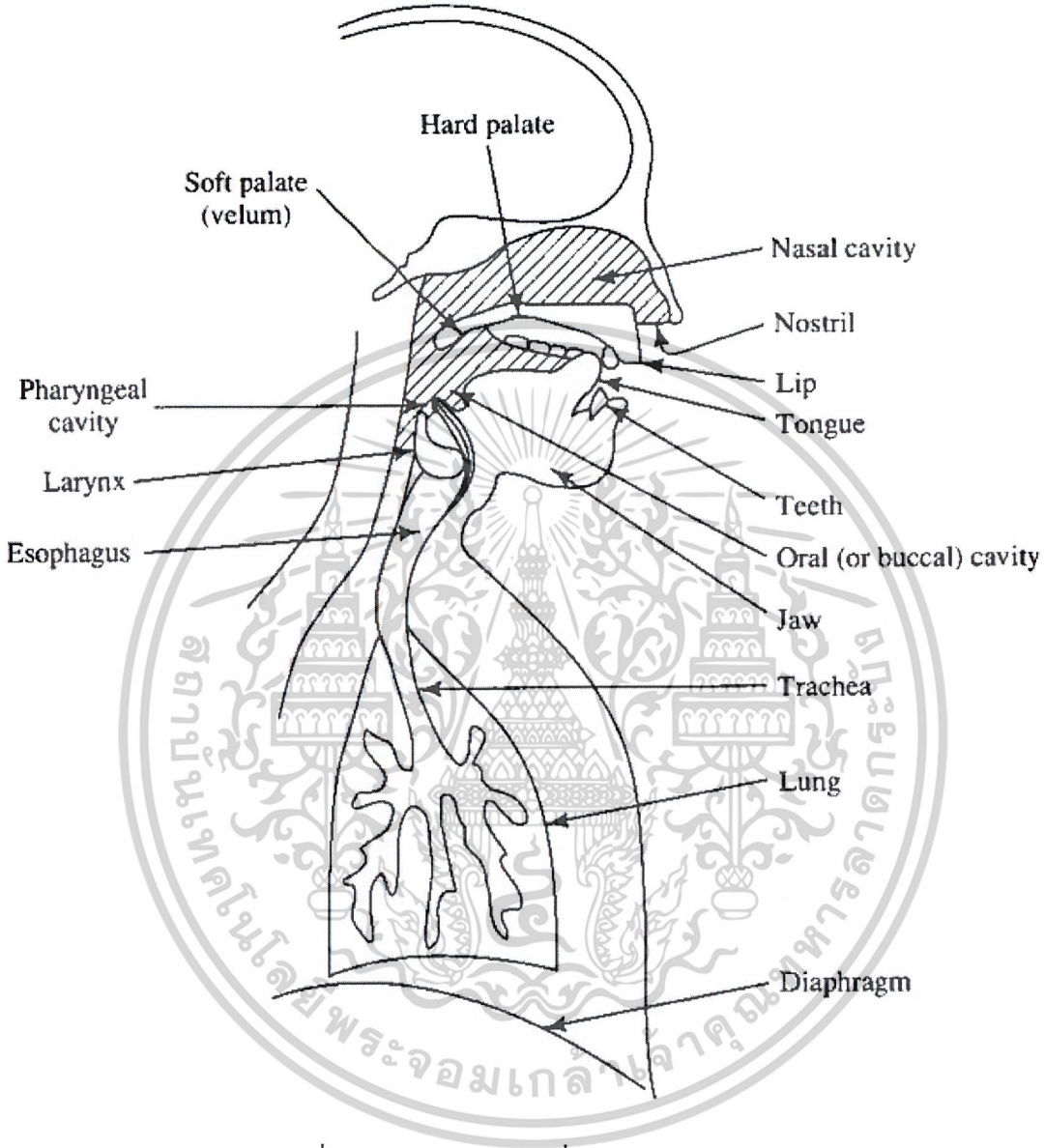
โดยทั่วไป อวัยวะที่เราใช้พูดมีลักษณะเหมือนกันทุกคน และทำงานในลักษณะเดียวกัน ผลที่ตามมา คือ ใครก็สามารถออกเสียงที่เป็นภาษาของมนุษย์ได้อย่างไม่ยุ่งยาก เพียงแต่อาจจะจำกัดในขอบเขตของอายุ เช่น การที่เด็กสามารถเรียนรู้ภาษาที่ตนเข้าไปอยู่ในสังคมนั้นๆ ได้ดี เมื่อเทียบกับผู้ใหญ่ที่ต้องเรียนรู้ภาษาต่างประเทศ ความสามารถในการเรียนรู้ภาษาใหม่ของผู้ใหญ่จะประสบกับความยากลำบากกว่าเด็ก การพัฒนาทางภาษาของเด็กๆ จะเลียนเสียงผู้ใหญ่ที่อยู่รอบตัวเขา โดยเด็กจะเลือกออกเสียงที่เหมาะสมกับตัวเอง เมื่อเด็กพ้นจากวัยรุ่นความสามารถทางภาษาจะน้อยลง นอกจากนี้อวัยวะที่ใช้ในการออกเสียงของแต่ละบุคคลนั้นยังมีขนาดไม่เท่ากัน เราจึงสามารถจดจำเสียงพูดและวิธีการพูดของคนที่คุ้นเคยกันได้เพราะความแตกต่างนี่เอง

อวัยวะที่เข้ามาเกี่ยวข้องในการออกเสียงของมนุษย์มีมากกว่าครึ่งหนึ่งของร่างกาย ตั้งแต่ศีรษะถึงช่องท้อง อวัยวะที่ใช้ในการออกเสียงนี้ สามารถแบ่งออกเป็นระบบที่สำคัญ 3 ระบบ คือ

1) **ระบบหายใจ** (Respiratory System) ที่ตั้งของระบบนี้อยู่ในช่วงอก (thorax) หมายถึง กระบังลม (diaphragm) ปอด (lungs) และหลอดลม (trachea , windpipe)
2) **ระบบการเกิดเสียง** (Phonatory System) ที่ตั้งของระบบนี้อยู่ที่ลำคอ หมายถึง กล่องเสียง และส่วนประกอบต่าง ๆ ของกล่องเสียง เช่น เส้นเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) **ระบบการเปล่งเสียง** (Articulatory System) ที่ตั้งของระบบนี้อยู่ส่วนศีรษะ หมายถึง ช่องปากและส่วนต่าง ๆ ภายในช่องปาก (Oral Cavity) และช่องจมูก (Nasal Cavity)



รูปที่ 2.4 แสดงภายอวัยวะที่ใช้ในการออกเสียง

2.2.1 อวัยวะที่ใช้ในการออกเสียงของมนุษย์

1) ปอดและกะบังลม (diaphragm and lungs) ทำหน้าที่สำคัญในการหายใจ และเป็นต้นกำเนิดการไหลของอากาศในกระบวนการผลิตเสียง
2) หลอดลม (trachea , windpipe) ทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียง และเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) กล่องเสียง (larynx) ทำหน้าที่เป็นทางเดินอากาศเวลาหายใจ และเป็นตัวผลิตพัลส์ของอากาศขณะเปล่งเสียง ประกอบด้วยเส้เสียง (Vocal cords) และช่องสายเสียง (Glottis)
4) ช่องปากและส่วนของหลอดอาหารตอนต้น อวัยวะกลุ่มนี้อยู่ต่อจากกล่องเสียง อาจเรียกว่าอวัยวะกำทอนเสียง (Vocal tract) ทำหน้าที่กำทอนเสียงทั้งที่เกิดจากกล่องเสียงและเสียงที่เกิดภายในปาก
5) โพรงจมูก เริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสอง ทำหน้าที่กำทอนเสียงร่วมกับช่องปากเมื่อมีการเปล่งเสียงที่ออกจากจมูก (Nasal sounds) เช่น เสียง ม, น, และ ง เป็นต้น

2.2.2 การทำให้เกิดเสียงพูด (Speech Production)

ถ้าจะพิจารณาในแง่การเกิดเสียงพูด ก็คือลมหายใจ ที่ถูกดัดแปลงไปโดยมีกระแสอากาศซึ่งถูกขับเคลื่อนโดยการทำงานของอวัยวะออกเสียง ซึ่งทำหน้าที่เป็นแหล่งกำเนิดพลังงานต่าง ๆ เป็นองค์ประกอบที่สำคัญที่ทำให้เกิดเป็นเสียงขึ้นมา ขั้นตอนที่ทำให้เกิดเสียงพูดมี 3 ขั้นตอน ดังนี้

1) การขับเคลื่อนกระแสอากาศ (Air-stream Mechanism)

ปัจจัยที่ทำให้เกิดเสียงขึ้นได้คือ อากาศ เพราะเสียงก็คืออากาศที่ถูกผลักดันให้เคลื่อนที่และถูกดัดแปลงหรือแปรให้เป็นเสียงประเภทต่าง ๆ โดยการทำงานของฐานกรณ์ต่าง ๆ ถ้าไม่มีอากาศก็จะไม่มีเสียงเกิดขึ้น การขับเคลื่อนกระแสอากาศมีต้นกำเนิดพลังงานจากตำแหน่งที่ต่างกัน เสียงที่เกิดขึ้นจึงแตกต่างกันไป แหล่งพลังงาน มี 3 แหล่งด้วยกันคือ แหล่งพลังงานจากปอด , แหล่งพลังงานจากกล่องเสียง และแหล่งพลังงานจากเพดานอ่อน

2) การทำให้เป็นเสียงแบบต่าง ๆ (Phonation)

การเกิดเสียงพูดนี้จะเกี่ยวข้องกับการทำงานของเส้นเสียงโดยตรงนั่นคือ การจะเกิดเสียงแบบต่าง ๆ ขึ้นอยู่กับการดัดแปลงรูปแบบการสั่นของเส้นเสียง โดยปกติภาษามีเสียง ๒ ประเภท คือ เสียงก้อง (voiced sound) จะเกิดขึ้นโดยมีการสั่นของเส้นเสียงร่วมด้วย และเสียงไม่ก้อง (voiceless sound) จะเกิดขึ้นโดยไม่มีการสั่นของเส้นเสียง

3) การแปรเสียงหรือการกล่อมเสียง (Articulation)

กระแสอากาศจากแหล่งพลังงานต่าง ๆ ซึ่งส่วนใหญ่หมายถึง ปอด เคลื่อนขึ้นสู่กล่องเสียงและถูกดัดแปลงคุณภาพเสียงให้แตกต่างไปตามรูปแบบการทำงานแบบต่าง ๆ ของเส้นเสียง แล้วต่อมาอากาศก็จะเดินทางเข้าสู่ช่องปาก ซึ่งประกอบด้วยอวัยวะแปรเสียงหรือฐานกรณ์มากมาย ซึ่งทำหน้าที่ในการกล่อมเกลาเสียงให้ออกมามีคุณลักษณะแตกต่างกัน ทั้งนี้ขึ้นอยู่กับลักษณะวิธีออกเสียงด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 ประเภทของเสียง

เสียงที่เกิดจากขั้นตอนในการเกิดเสียง จะถูกแบ่งออกเป็น 2 หัวข้อใหญ่ๆ คือ

1) เสียงก้อง (Voiced sound)

เกิดจากการออกเสียงในขณะที่เส้นเสียงถูกดึงเข้ามาใกล้กันจนเกือบปิดช่องทางลมสนิท ลมที่ดันขึ้นมาจากปอดจะทำให้เส้นเสียงสั่น ลมที่ออกมาไม่สะดวกเนื่องจากต้องบีบตัวผ่านช่องแคบเป็นจังหวะ จึงทำให้เกิดเป็นเสียงขึ้น

2) เสียงไม่ก้อง (Voiceless sound)

เป็นการออกเสียงในขณะที่เส้เสียงยังเปิดกว้าง โดยเปิดช่องระหว่างเส้นเสียงหรือคอหอยให้ลมหายใจผ่านเข้าออกสะดวกขึ้น หรืออาจกล่าวได้ว่า เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูก โดยที่อวัยวะในช่องปาก ริมฝีปาก จะมาขวางการไหลของอากาศให้ผ่านได้เป็นช้องเล็กๆ อากาศจึงไหลผ่านได้อย่างรวดเร็ว

2.3 ทฤษฎีการรู้จำเสียงพูด (Speech Recognition)

Speech recognition หมายถึง การให้คอมพิวเตอร์สามารถเข้าใจเสียงพูดของมนุษย์ได้ สาขานี้อาศัยความร่วมมือของผู้เชี่ยวชาญ 3 สาขาหลักๆ คือ ภาษาศาสตร์ (linguistics), digital signal processing (DSP), computer science โดยองค์ประกอบหลัก ประกอบด้วยขั้นตอนดังต่อไปนี้

1) การประมวลผลสัญญาณเสียงเบื้องต้น (Preprocessing)
2) การสกัดค่าลักษณะสำคัญของเสียงพูด (Speech Feature Extraction)
3) เทคนิกการรู้จำ (Speech Recognition)

2.3.1 การประมวลผลสัญญาณเสียงเบื้องต้น (Preprocessing)

เป็นขั้นตอนที่จะทำให้สัญญาณเสียงที่จะนำไปใช้หรือรับเข้ามานั้น มีความสมบูรณ์มากที่สุด ในกระบวนการนี้เป็นการปรับมาตรฐานเสียงให้อยู่ในรูปแบบเดียวกันเพราะว่า บางคนอาจจะพูดช้า บางคนพูดเร็ว บางคนเสียงเบา และบางคนเสียงดัง ดังนั้นกระบวนการนี้จะเป็นการปรับให้เสียงที่เข้ามานั้นให้อยู่ในมาตรฐานเดียวกัน โดยมีขั้นตอนในการทำดังนี้

1) การหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณ (End Point Detection)

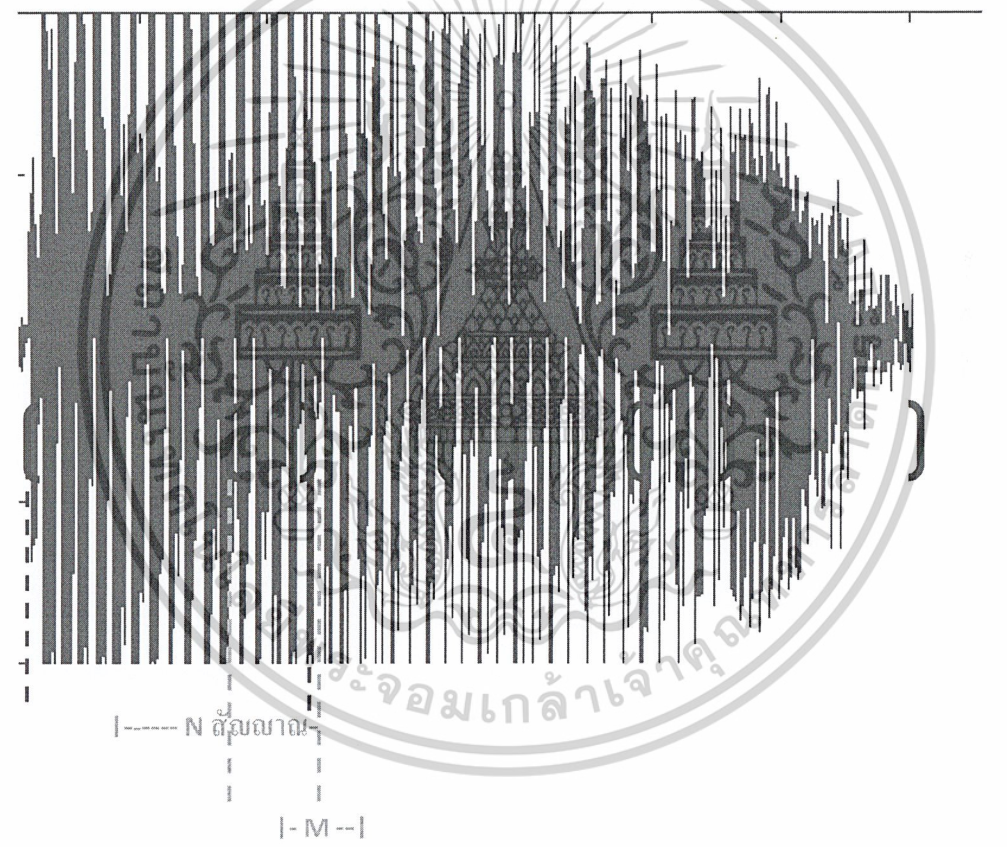
ขั้นตอนนี้เป็นขั้นตอนที่ใช้หาขอบเขตของสัญญาณซึ่งได้ผ่านกระบวนการ filter หรือการกรองเอาสัญญาณความถี่ต่ำออกจากสัญญาณเสียงไปก่อนแล้ว จากนั้นจึงนำมาทำการหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณ เพื่อนำเฉพาะสัญญาณเสียงในช่วงที่เป็นเสียงพูดมาประมวลผลเท่านั้น เนื่องจากสัญญาณเสียงที่รับเข้ามาจากไมโครโฟนหรือจากเครื่องอัดเสียงนั้นประกอบด้วยส่วนที่เป็นเสียงพูด กับส่วนที่ไม่ใช่เสียงพูดประกอบอยู่ ทั้งส่วนหน้าของเสียงพูดและต่อท้าย เช่น ช่วงที่ไม่มีเสียงหรือส่วนที่เกิดจากสิ่งแวดล้อม จึงต้องใช้กระบวนการนี้ตัดส่วนที่ไม่ต้องการออกไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณเสียงเป็นสัญญาณที่มีการเปลี่ยนแปลงตลอดเวลา และส่วนที่ไม่ใช่เสียงพูดจะมีขนาดสัญญาณเฉลี่ยต่ำกว่าส่วนที่เป็นเสียงพูด ดังนั้นในการหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณเสียง จะมีการตั้งค่าขึ้นมาค่าหนึ่ง หรือ Threshold ซึ่งจะเป็นค่าที่ใช้ตัดสินใจแยกส่วนที่เป็นเสียงและไม่ใช่เสียงออกจากกัน โดยจะคำนวณหาค่าที่ต่ำกว่า 6 % ของค่าเสียงที่มากที่สุดเพื่อเป็นค่าต่ำสุดของเสียงพูดที่ยอมรับได้ สัญญาณไหนต่ำกว่าค่านี้ จะถือว่าเป็นสัญญาณที่ไม่ต้องการ

2) การแบ่งช่วงสัญญาณ (Frame Blocking)

ข้อมูลเสียงซึ่งผ่านการตัดส่วนที่ไม่ต้องการออกไปแล้ว จะถูกทำให้มีขนาดของสัญญาณในระดับมาตราฐานเดียวกัน จากนั้นจะถูกแบ่งออกเป็นช่วงสัญญาณ ซึ่งแต่ละช่วงสัญญาณจะประกอบไปด้วยสัญญาณเสียง N สัญญาณ และมีการเหลื่อมกัน M สัญญาณ ดังภาพตัวอย่าง



รูปที่ 2.5 แสดงตัวอย่างการวินโดว์เพื่อแบ่งเฟรมในการทำไปสกัดค่าลักษณะสำคัญของเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 การสกัดค่าลักษณะสำคัญของเสียงพูด (Speech Feature Extraction)

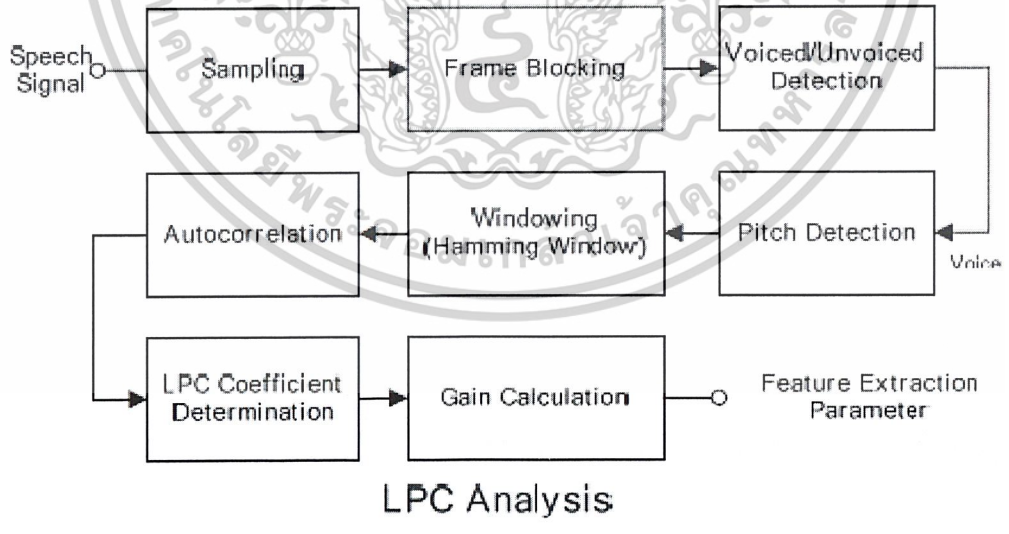
ในการรู้จำเสียง ระบบไม่ได้เอาสัญญาณเสียงไปเปรียบเทียบกันโดยตรง แต่จะดึงเฉพาะลักษณะสำคัญของเสียงออกมา เราเรียกลักษณะสำคัญนี้ว่า Speech Feature โดยปกติแล้วแม้คนเราจะพูดคำเดียวกัน แต่ก็จะมีลักษณะเด่นของคำพูดที่แตกต่างกันออกไป และแต่ละคำก็จะมีลักษณะที่แตกต่างหรืออาจมีลักษณะบางอย่างร่วมกัน ขั้นตอนนี้เป็นการหาลักษณะเด่นที่สำคัญของเสียงพูดเพื่อเป็นตัวแทนของคำแต่ละคำ การสกัดหาค่าลักษณะสำคัญของเสียงพูดสามารถทำได้หลายวิธี เช่น วิธีการประมาณพันธะเชิงเส้น (Linear Prediction Coefficient : LPC), วิธีการหาสัมประสิทธิ์เซปสตรัม (Cepstral Coefficient) และวิธีเวฟเล็ต (Wavelet) โดยโครงงานนี้จะใช้วิธีการประมาณพันธะเชิงเส้น (Linear Prediction Coefficient : LPC)

วิธีการประมาณพันธะเชิงเส้น (Linear Prediction Coding : LPC)

การประมาณพันธะเชิงเส้นจะแบ่งเป็น 2 ส่วนคือ

1) ส่วนที่เป็นการบีบอัดสัญญาณให้มีขนาดเล็กลง (Data Compression)
2) ส่วนที่เป็นการกู้สัญญาณที่ถูกบีบอัดกลับมา (Data Decompression)

ซึ่งการบีบอัดสัญญาณนั้นมี 2 แบบคือ การบีบอัดแบบสูญเสีย (Lossy Compression) กับการบีบอัดแบบไม่สูญเสีย (Lossless Compression) โดยในโครงงานนี้เป็นการนำเสนอการบีบอัดสัญญาณแบบสูญเสีย



รูปที่ 2.6 แสดงบล็อคไดอะแกรมขั้นตอนการบีบอัดเสียงโดยวิธีการประมาณพันธะเชิงเส้น (LPC)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 2.6 เป็นการวิเคราะห์ (Analysis) หาค่าสัมประสิทธิ์แอลพีซี โดยนำสัญญาณเสียงตัวอย่างมาทำการสุ่มค่าด้วยความถี่สุ่มที่เหมาะสม จากนั้นสัญญาณจะถูกแบ่งเป็นเฟรม (Frame Blocking) เมื่อทำการแบ่งเฟรมเสร็จเรียบร้อยแล้วจะตรวจหาเฉพาะส่วนที่เป็นเสียง (Voiced) เพื่อนำไปคำนวณหาค่าคาบเวลาพิทช์ (Pitch Period) จากนั้นจะนำเฟรมข้อมูลสัญญาณไปให้น้ำหนักด้วยหน้าต่างแบบแฮมมิ่งวินโดว์ (Hamming Window) เพื่อขจัดองค์ประกอบความถี่สูง ที่เกิดจากการแบ่งเฟรม นำสัญญาณเสียงได้จากข้างต้นไปหาค่าอัตสหสัมพันธ์ (Autocorrelation) ขณะเดียวกันนั้นค่าสัมประสิทธิ์แอลพีซี (LPC Coefficient) จะถูกคำนวณออกมา และระบบจะได้ค่าอัตราการขยาย (Gain) ออกมาด้วย โดยที่ค่าอัตราการขยายนั้นจะเป็นตัวที่บอกความดังค่อยของสัญญาณ

แบบจำลองทางคณิตศาสตร์ (Mathematics Model)

การประมาณพันธะเชิงเส้น (LPC) ในโครงงานนี้จะใช้อัลกอริทึมของ Levinson-Durbin มาใช้ในการประมาณค่าสัมประสิทธิ์ LPC จากสัญญาณเสียงที่ต้องการ กำหนดให้สัญญาณเดิมคือ $\mathrm{s(n)}$ ให้ค่าที่ได้จากการประมาณพันธะเชิงเส้นของสัญญาณเดิม $\mathrm{s(n)}$ คือ $\tilde{s}(\mathrm{n})$ และสมมติให้สัญญาณเสียงที่ทำการวิเคราะห์ขึ้นเกิดจากการนำเอาสัญญาณในอดีตของสัญญาณเสียงจริงมาจำนวน p ตัวอย่าง เพราะฉะนั้นจะได้ว่า

$$\tilde{s}(\mathrm{n}) = \alpha_1 s(n-1) + \alpha_2(n-2) + \ldots + \alpha_p s(n-p) = \sum_{k=1}^{p} s(n-k) \quad (2.1)$$

เมื่อ α_k คือค่าสัมประสิทธิ์ของ LPC และค่าความคลาดเคลื่อนระหว่างสัญญาณจริงกับสัญญาณที่ประมาณได้คือ

$$\varepsilon(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} s(n-k) \quad (2.2)$$

ซึ่งผลรวมของค่าความคลาดเคลื่อนกำลังสองคือ

$$E = \sum_n \varepsilon^2(n) = \sum_n \{s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)\}^2 \quad (2.3)$$

โดยค่า α_k ที่ใช้ในการประมาณพันธะเชิงเส้นจะต้องทำให้ผลรวมของค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุดด้วยการทำอนุพันธ์สมการที่ (3) เทียบกับสัมประสิทธิ์ α_k แล้วเท่ากับศูนย์

$$2\sum_n s(n-i)\{s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) = 0 \; for \; i = 1,2,3,\ldots,p\} \quad (2.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากสมการที่ (2.4) จะประกอบด้วย สมการ ตัวแปรที่ไม่ทราบค่าดังแสดง

$$\alpha_1 \sum_n s(n-i)s(n-1) + \alpha_2 \sum_n s(n-i)s(n-2) + \ldots$$
$$+\alpha_p \textstyle\sum_n s(n-i)s(n-p) = \sum_n s(n-i)s(n) \quad (2.5)$$

$$for\ i = 1,2,\ldots,p$$

และถ้ากำหนดให้

$$r(i) = \textstyle\sum_{n=0}^{n-1-i} s(n)s(n+i) \quad (2.6)$$

จากสมการที่ (2.6) ค่า $r(i)$ คือค่าอัตสหสัมพันธ์ (Autocorrelation) และเมื่อนำเอาสมการที่ (5) มาเขียนใหม่ในรูปเมตริกจะได้ดังสมการที่ (2.7)

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(p-2) & r(p-1) \\ r(1) & r(0) & \cdots & r(p-3) & r(p-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r(p-2) & r(p-3) & \cdots & r(0) & r(1) \\ r(p-1) & r(p-2) & \cdots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{p-1} \\ \alpha_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p-1) \\ r(p) \end{bmatrix} = R \cdot \alpha = r \quad (2.7)$$

ออกมาได้หนึ่งครั้ง สัมประสิทธิ์ที่ได้จะถูกแทนลงไปที่สมการที่ (2.2) แล้วจะได้ค่าความผิดพลาด ซึ่งขบวนการในสมการที่ (2.2) จะถูกเรียกว่า Analysis filter โดยมีฟังก์ชันถ่ายโอน (Transfer Function) ดังสมการที่ (2.8) สัมประสิทธิ์แต่ละตัวสามารถหาได้ด้วยวิธีการของ Levinson-Durbin แล้วค่าสัมประสิทธิ์ที่คำนวณ

$$A(z) = 1 - \textstyle\sum_{k=1}^{p} \alpha_k z^{-1} \quad (2.8)$$

2.3.3 เทคนิคการรู้จำ (Speech Recognition)

เมื่อได้ลักษณะสำคัญของเสียงแล้ว จึงนำค่าที่ได้มาวิเคราะห์ว่าเป็นคำพูดใด โดยในโครงงานนี้ได้ใช้เทคนิคการรู้จักเสียงด้วยวิธีโครงข่ายประสาทเทียม (Neural Network)

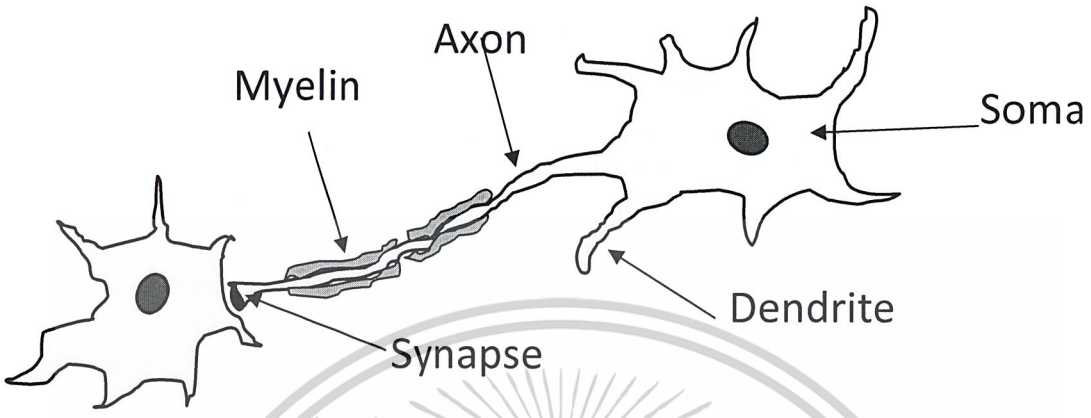
โครงข่ายประสาทเทียม (Artificial neural network)

Neural Network คือ โครงข่ายของใยประสาทที่เชื่อมต่อกัน ระหว่างเซลล์ประสาทจำนวนมาก มีความสามารถในการประมวลผลสูง บรรจุอยู่ในสมอง ในระบบการคิดของมนุษย์นั้น มีพื้นฐานจากเซลล์ประสาท ที่เรียงกันเป็นชั้นๆ อย่างซับซ้อนจำนวนมากมาย ประมาณหมื่นล้านเซลล์ ก่อนอื่นเราต้องศึกษาเซลล์ประสาทชีวภาพก่อน ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. เซลล์ประสาทชีวภาพ

1.1) โครงสร้างของเซลล์ประสาท



รูปที่ 2.7 แสดงโครงสร้างเซลล์ประสาทของมนุษย์

โครงสร้างเซลล์ประสาท ประกอบด้วย

1. ตัวเซลล์ (Soma)
2. เส้นประสาทหลัก (Axon) ทำหน้าที่เป็นส่งสัญญาณเอาท์พุต ไปยังเซลล์ประสาทอื่นๆ
3. เส้นกิ่งก้านสาขา (Dendrite) เป็นแขนงของเซลล์ประสาท ที่ทำหน้าที่รับสัญญาณเข้ามาตัวเซลล์
4. Synapse จะอยู่ส่วนปลายของเส้นประสาทหลัก ทำหน้าที่ เชื่อมต่อกับเซลล์ประสาท เซลล์ถัดไป
5. Myelin Sheath เป็นฉนวน ทำหน้าที่ป้องกันสัญญาณรบกวนจากเซลล์ข้างเคียง ทำให้นำสัญญาณได้ดีขึ้น

1.2) การทำงานของเซลล์ประสาทชีวภาพ

ทุกๆ เซลล์ประสาทจะทำหน้าที่คล้ายกันคือ รับสัญญาณอินพุต, ประมวลผล และส่งออกสัญญาณไฟฟ้าไปยังเซลล์ข้างเคียง ในสภาวะเริ่มต้นเซลล์ประสาทจะมีระดับแรงดันอ้างอิงอยู่ค่าหนึ่ง เรียกว่า Threshold voltage

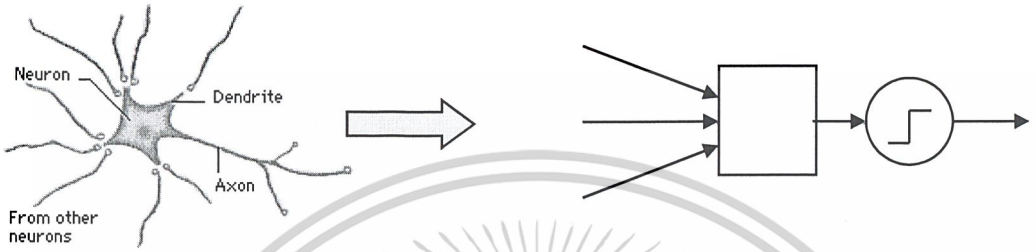
เมื่อ Dendrite หลายๆ แขนง รับสัญญาณอินพุตเข้ามามากมาย จะถูกนำมารวมกันที่เซลล์ประสาท ถ้าผลรวมของสัญญาณไฟฟ้าเคมี มีค่ามากกว่า Threshold เซลล์ประสาทนั้นๆ จะส่งสัญญาณไฟฟ้าค่าหนึ่ง ไปตาม Axon ผ่านทาง Synapse ไปยังเซลล์ประสาทข้างเคียง และสัญญาณไฟฟ้าที่ส่งไปนั้น จะถูกลดขนาดด้วย Synaptic ที่มีค่าแตกต่างกันของแต่ละเซลล์ เพราะแต่ละเซลล์ประสาท นั้นมีความเข้มของสารเคมีที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เซลล์ประสาทเทียม (Artificial Neuron Cell)

การออกแบบสร้างโครงข่ายประสาทเทียม เพื่อเลียนแบบการทำงานของโครงข่ายประสาทชีวภาพนั้น จึงต้องจำลองแบบของ เซลล์ประสาทเทียม (Artificial Neuron Cell)ขึ้นมาก่อน

จากโครงสร้างเซลล์ประสาทชีวภาพ และการทำงานที่กล่าวไปแล้วในข้อ 1. สามารถจำลองเซลล์ประสาทเทียมได้ดังรูปที่ 2.8



รูปที่ 2.8 เซลล์ประสาทเทียม (Artificial Neuron Cell)

จากรูปที่ 2.8 สัญญาณอินพุต $x_1, x_2, \ldots, x_n$ เทียบได้กับสัญญาณที่รับมาจาก Dendrite เมื่อสัญญาณผ่าน Synapse จะถูกถ่วงน้ำหนักด้วยค่า Synaptic weight (w_i) ก่อนที่จะเข้า Block summation

ผลรวมทางคณิตศาสตร์ของอินพุตที่ถูกเวจท์ เรียกว่า *Net* ดังนั้น

$$net = x_1 w_1 + x_2 w_2 + \ldots + x_n w_n \tag{2.9}$$

เมื่อได้สัญญาณ *net* แล้ว แต่ละเซลล์ต้องตัดสินใจว่าจะส่งสัญญาณเอาท์พุตออกมา เพื่อส่งไปเซลล์อื่นๆ หรือไม่ การตัดสินใจดังกล่าวนี้ กำหนดได้โดยการกำหนด ฟังก์ชั่นกระตุ้นความสนใจ (Activation Function)

$$y = K[net] \tag{2.10}$$

ตัวอย่าง Activation Function

1. Threshold Function

$$y = \begin{cases} -1, net < T \\ 1, net > T \end{cases}$$ เมื่อ T คือ Threshold level (2.11)

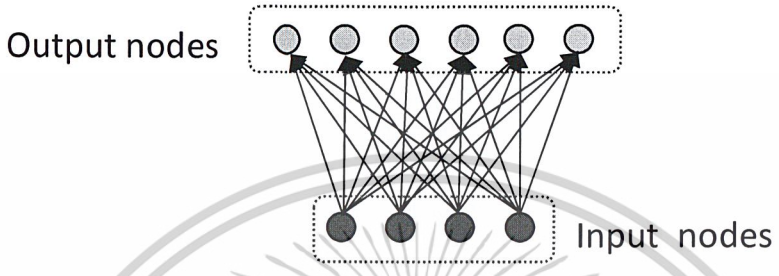
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. โครงข่ายประสาทเทียม (Artificial Neural Network)

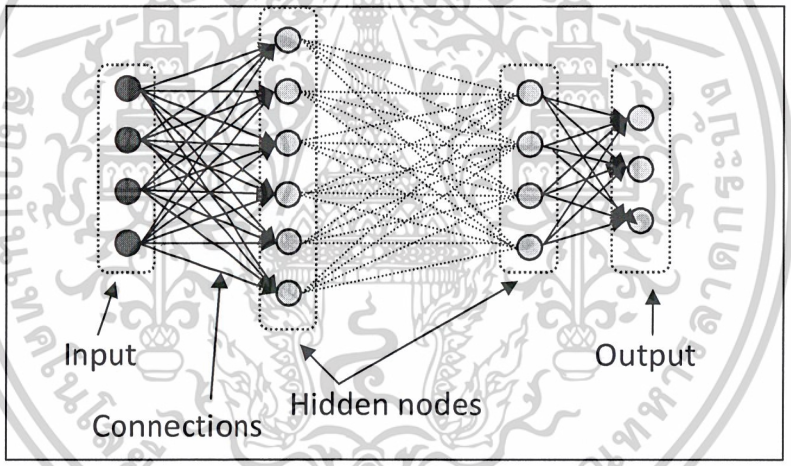
โครงข่ายประสาทเทียม คือ การนำเซลล์ประสาทเทียม (Artificial Neuron Cell) มาต่อกัน เป็นโครงข่ายในลักษณะเดียวกับเซลล์สมอง ซึ่งลักษณะการต่อกันก็มีหลายรูปแบบ แต่ละชนิดก็มีคุณสมบัติ ที่แตกต่างกันออกไป เช่น

โครงข่ายประสาทเทียมแบบชั้นเดียว (Single layer artificial neural network)

โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer artificial neural network)



รูปที่ 2.9 Single layer artificial neural network



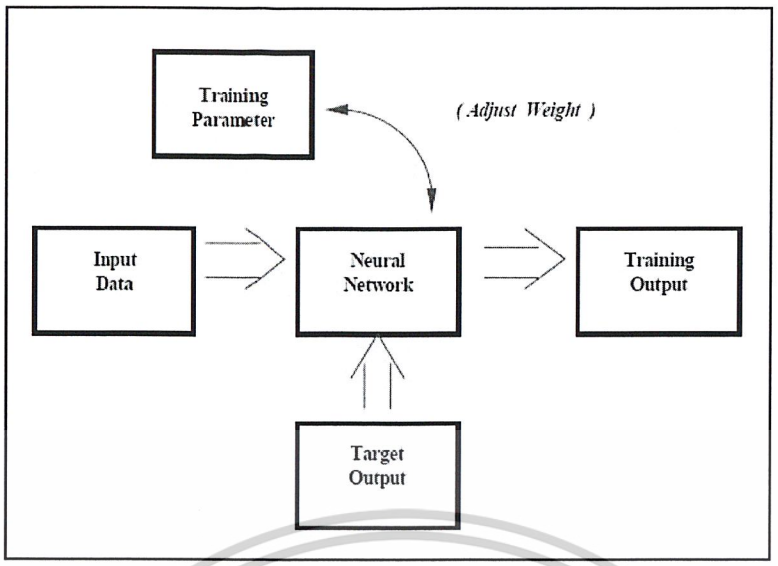
รูปที่ 2.10 Multilayer artificial neural network

4. การเรียนรู้สำหรับ Neural Network

4.1) Supervised Learning การเรียนแบบมีการสอน

เป็นการเรียนแบบที่มีการตรวจคำตอบเพื่อให้วงจรข่ายปรับตัว ชุดข้อมูลที่ใช้สอนวงจรข่ายจะมีคำตอบไว้คอยตรวจดูว่าวงจรข่ายให้คำตอบที่ถูกหรือไม่ ถ้าตอบไม่ถูก วงจรข่ายก็จะปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น (เปรียบเทียบกับคน เหมือนกับการสอนนักเรียนโดยมีครูผู้สอนคอยแนะนำ)

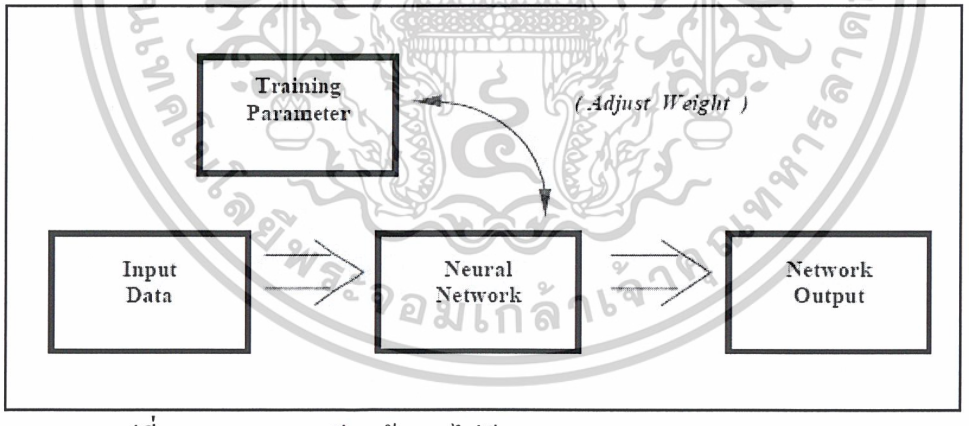
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.11 แสดงการเรียนรู้แบบมีการสอน (Supervised Learning)

4.2) Unsupervised Learning การเรียนแบบไม่มีการสอน

เป็นการเรียนแบบไม่มีผู้แนะนำ ไม่มีการตรวจคำตอบว่าถูกหรือผิด วงจรข่ายจะจัดเรียงโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้ วงจรข่ายจะสามารถจัดหมวดหมู่ของข้อมูลได้ (เปรียบเทียบกับคน เช่น การที่เราสามารถแยกแยะพันธุ์พืช พันธุ์สัตว์ตามลักษณะรูปร่างของมันได้เองโดยไม่มีใครสอน)



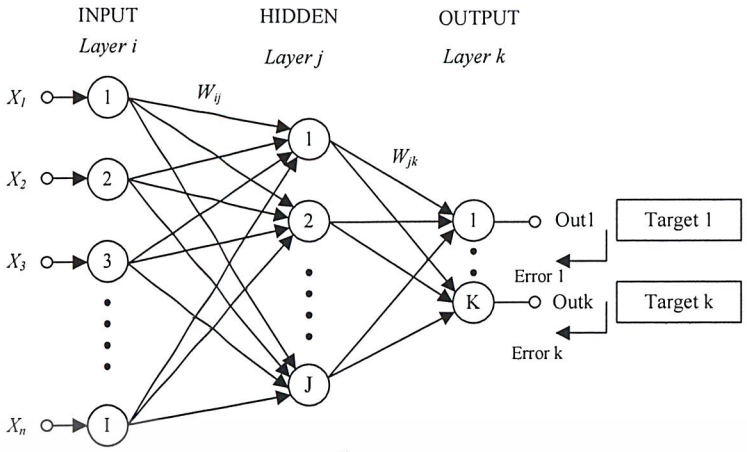
รูปที่ 2.12 แสดงการเรียนรู้แบบไม่มีการสอน Unsupervised Learning

5. โครงข่ายประสาทเทียมที่ใช้ในโครงงาน

ในโครงงานนี้ เลือกใช้โครงสร้างประสาทเทียมแบบหลายชั้น และใช้การเรียนรู้แบบ Supervised โดยใช้อัลกอริทึมการเรียนรู้แบบ Backpropagation จากรูปที่ 2.13 แสดงโครงข่ายประสาทเทียมแบบสองชั้น โดยชั้นที่ต่อกับอินพุต เรียกว่า อินพุตเลเยอร์ และชั้นที่อยู่ขวาสุดเรียกเอาท์พุตเลเยอร์ ส่วนชั้นที่อยู่ระหว่างอินพุตเลเยอร์ และ เอาท์พุตเลเยอร์ เรียกว่า ฮิดเดนเลเยอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง



รูปที่ 2.13 โครงข่าย Neural Network แบบสองชั้น ที่ใช้กระบวนการเรียนรู้แบบ Backpropagation

การฝึกสอนให้กับโครงข่ายประสาทเทียม

เป็นการปรับคุณลักษณะ ของโครงข่ายประสาทเทียม ให้มีคุณสมบัติตามต้องการ โดยใช้อัลกอรึทึม ของ Backprogpagation ซึ่งเป็นการฝึกสอนแบบ Supervise training โดยแบ่งเป็นสองส่วนคือ Forward pass กับ Reverse pass

โดยการฝึกสอนนั้น จะทำทั้ง Forward pass และ reverse pass สลับกันไปเรื่อยๆ จนกว่าจะได้ค่าความผิดพลาด (Error) ระหว่าง Output กับ Target มีค่าต่ำตามที่ต้องการ

5.1 Forward pass

จากรูปที่ 2.13 อินพุตที่ x_1 ...ถึง x_n จะถูกคูณด้วยค่า Weight และ ผลรวมของอินพุตเมื่อคูณกับค่า Weight ของแต่ละนิวรอล เรียกว่า *NET* โดยที่

$$NET_j = \sum_{i=1}^{I} x_i W_{ij} \qquad (2.12)$$

และในทำนองเดียวกัน

$$NET_k = \sum_{j=1}^{J} out_j W_{jk} \qquad (2.13)$$

ค่า *NET* ที่ได้ของแต่ละนิวรอล จะนำไปเข้า Activate Function เพื่อหาค่าเอาท์พุตที่ได้ ในที่นี้ให้ Activate Function เป็นฟังก์ชั่น Tan hyperbolic ดังนั้น

$$out_j = \tanh(NET_j) \text{ และ } out_k = \tanh(NET_k) \qquad (2.14)$$

06584
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 Reverse pass

เป็นการป้อนกลับค่า Error จากเอาท์พุตเลเยอร์ กลับมาปรับปรุงค่า Weight ของเลเยอร์ก่อนหน้านี้ โดยขั้นตอนการปรับปรุงค่า Weight สรุปได้ดังนี้

1. ปรับปรุงค่า Weight ในชั้น Hidden ไปยังชั้น Output (W_{jk})

   -หาค่า Error โดยที่

   $$error_k = f'\left(NET_k\right)\cdot\left(\text{Tarket}_k - out_k\right) \tag{2.15}$$

   - ดังนั้นจะได้ว่าค่า

   $$\Delta W_{jk} = \eta \cdot error_k \cdot out_j \text{ โดยที่ } \eta \text{ เรียกว่า อัตราการเรียนรู้} \tag{2.16}$$

2. ปรับปรุงค่า Weight ในชั้น Input ไปยังชั้น Hidden (W_{ij})

   -หาค่า Error โดยที่

   $$error_j = f'\left(NET_j\right)\cdot\sum_{k=0}^{K} error_k \cdot W_{jk} \tag{2.17}$$

   - ดังนั้นจะได้ว่าค่า

   $\Delta W_{ij} = \eta \cdot error_j \cdot x_i$ โดยที่ η เรียกว่า อัตราการเรียนรู้

   ดังนั้น $W_{jk}\left(n+1\right) = W_{jk}\left(n\right) + \Delta W_{jk}$ (2.18)

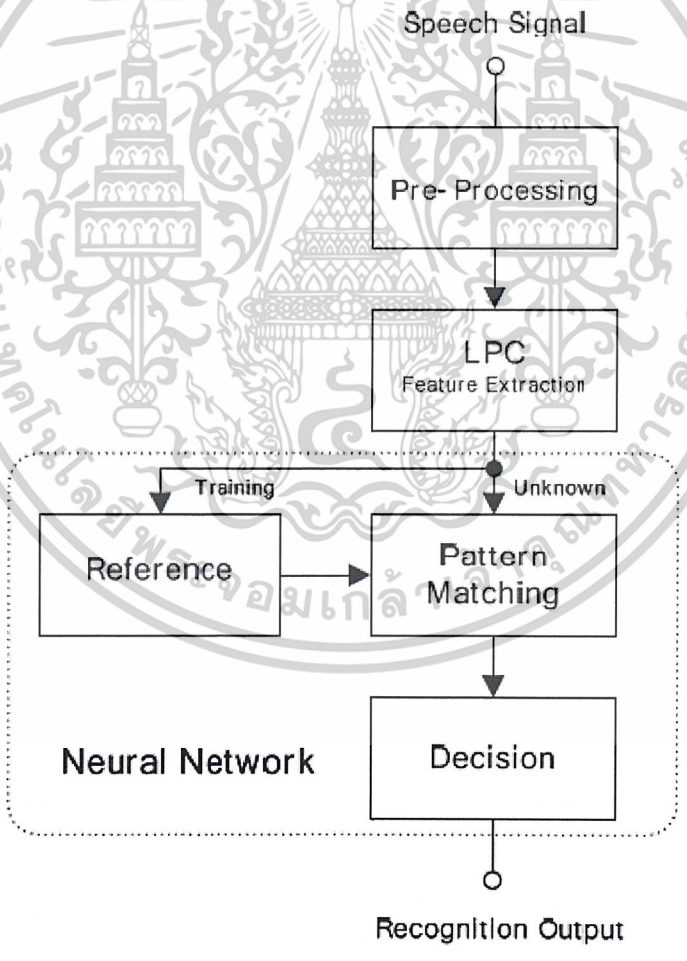
   $W_{ij}\left(n+1\right) = W_{ij}\left(n\right) + \Delta W_{ij}$ (2.19)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การออกแบบระบบ

3.1 โครงสร้างของโปรแกรม

ในขั้นแรก เราจำเป็นต้องเก็บข้อมูลเสียงต้นแบบมาให้ได้มากที่สุดเพื่อประสิทธิภาพในการเรียนรู้ของระบบ จากนั้นนำเสียงต้นแบบทั้งหมดมาผ่านกระบวนการสกัดคุณลักษณะเด่น โดยในระบบใช้การประมาณพันธะเชิงเส้น แล้วนำค่าสัมประสิทธิ์ที่ได้ไปใช้ต่อกับกระบวนการเรียนรู้โครงข่ายประสาทเทียม เนื่องจากระบบที่ทำนั้นเป็นการรู้จำเสียงแบบไม่ขึ้นกับบุคคลใด และเป็นคำพูดตายตัว ซึ่งกระบวนการต่างๆที่กล่าวมาจะแสดงดังรูปที่ 3.1



รูปที่ 3.1 แผนภาพแสดงขั้นตอนการเรียนรู้จำคุณลักษณะเสียงโดยใช้โครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

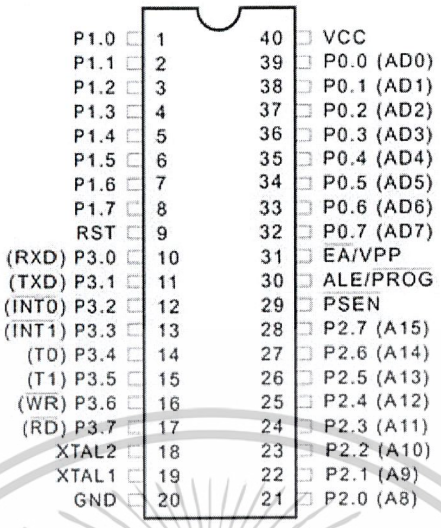
จากแผนภาพสัญญาณเสียงที่รับเข้าจะใช้อัตราการสุ่ม (Sampling Rate) 44100 Hz จากนั้นนำสัญญาณเสียงผ่านเข้ากระบวนการพรีโปรเซสซิ่ง (Pre-Processing) ในกระบวนการนี้เป็นการปรับมาตรฐานเสียงให้อยู่ในรูปแบบเดียวกันเพราะเสียงที่ได้รับมานั้นอาจมีทั้งเสียงรบกวน บางคนอาจพูดช้า บางคนพูดเร็ว ความดัง-ค่อยของเสียงไม่เท่ากันและเสียงสูงต่ำ ดังนั้นกระบวนการนี้จะเป็นการปรับให้เสียงที่เข้ามานั้นให้อยู่ในมาตรฐานเดียวกันโดยมีการทำการกรองสัญญาณรบกวนออกด้วยการ Filter จากนั้นจึงทำการค้นหาและปรับแต่งขนาดและเวลาให้อยู่ในมาตราฐานเดียวกัน

นำสัญญาณที่ผ่านกระบวนการพรีโปรเซสซิ่งไปทำการหาค่าคุณลักษณะเด่น (Feature Extraction) ของคำคำนั้นซึ่งระบบนี้ใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้นในการหาค่าคุณลักษณะเด่น โดยกระบวนการนี้จะทำการแบ่งเฟรมเสียง เฟรมละ 20 มิลลิวินาที (ประมาณ 880 samples) และมีส่วนเหลื่อมกันเฟรมละ 5 มิลลิวินาที (220 samples) และในแต่ละเฟรมจะถูกนำไปหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นของคำ

เมื่อได้ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นแล้วจะนำมาป้อนเป็นอินพุต (Input) ให้โครงข่ายประสาทเทียม เพื่อใช้ในการเรียนรู้และตัดสินใจ (Learning and Decision Process) เนื่องจากในระบบนี้โครงข่ายประสาทเทียมเป็นแบบชนิดหลายเลเยอร์แล้ว ยังเป็นแบบ Supervised Learning ดังนั้นจึงต้องมีการกำหนดเป้าหมาย (Target) ที่ต้องการไว้ ซึ่งในที่นี้จะมีการกำหนด output ทั้งหมด 5 โหนดด้วยกัน เช่นคำว่า “เปิด” จะกำหนดให้เอาต์พุต (Output) ที่ได้เป็น [0 0 0 0 1] และคำว่า “ปิด” จะกำหนดให้เอาต์พุต (Output) ที่ได้เป็น [0 0 0 1 0] เป็นต้น

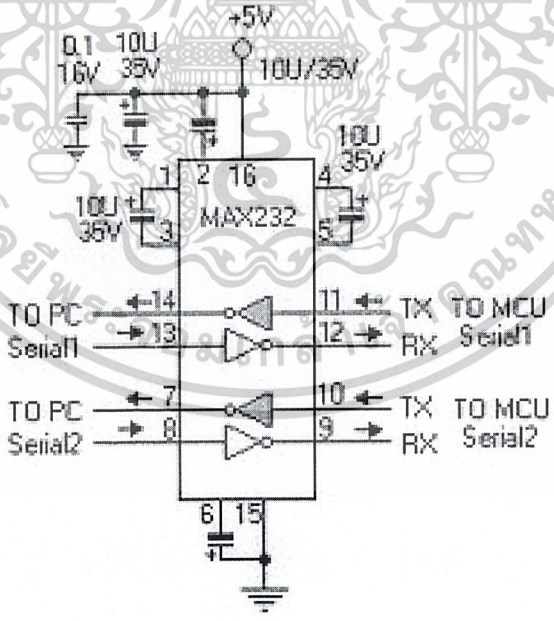
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 โครงสร้างของอุปกรณ์ฮาร์ดแวร์



รูปที่ 3.2 IC AT89C2051 (MCS51)

MCS 51 เป็น IC ที่ใช้เป็นตัวประมวลผลคำสั่งหลักโดยใช้ภาษาแอสเซมบลีในการควบคุมการทำงาน



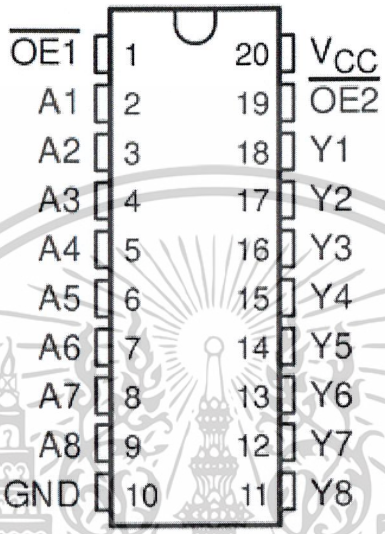
รูปที่ 3.3 IC MAX232

IC MAX232 เป็น IC ที่ใช้แปลงสัญญาณ RS-232(Computer) เป็นสัญญาณ TTL(IC) และแปลงสัญญาณ TTL(IC) เป็น สัญญาณ RS-232(Computer)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

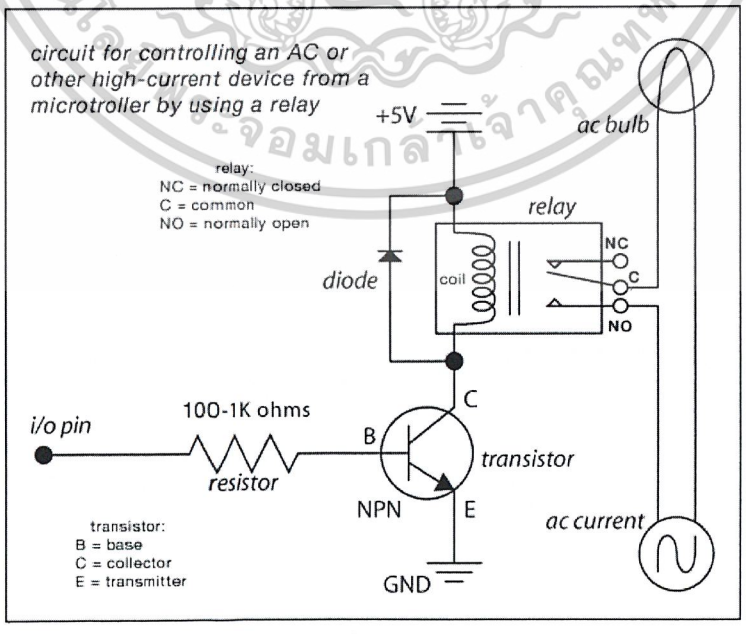
สัญญาณ RS-232 คือ มาตรฐานการเชื่อมต่อข้อมูลแบบ Serial ใช้เพื่อเพิ่มระยะทางในการส่งข้อมูล แบบ Serial ให้สามารถส่งได้ในระยะทางที่มากขึ้น โดยมีการเปลี่ยนระดับแรงดันของ Logic ให้เพิ่มมากขึ้น

สัญญาณ TTL (Transistor – Transistor Logic) เป็นระดับแรงดันที่ถูกกำหนดขึ้น เพื่อใช้ระหว่าง Transistor กับ Transistor ภายใน IC



รูปที่ 3.4 IC 74HC541

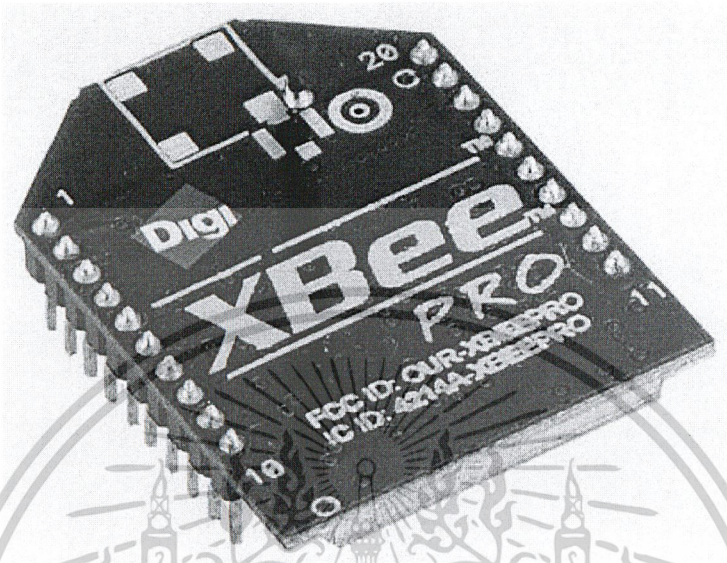
IC 74HC541 เป็น IC Buffer มีหน้าที่ขยายกระแสและป้องกันความเสียหายที่อาจเกิดกับ IC ตัวอื่นๆ



รูปที่ 3.5 IC Relay

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Relay เป็นอุปกรณ์ที่เชื่อมต่อจาก IC 74HC541 ทำหน้าที่เหมือนสวิตซ์ในการตัด/ต่อ กระแสไฟฟ้า โดยจะมีการทำงานคือ ในขณะที่ยังมีกระแสไฟเข้ามา ขา NO กับ ขา C จะอยู่ห่างกัน แต่เมื่อมีการส่งกระแสไฟถูกส่งเข้ามา ขา NO กับขา C จะเชื่อมกัน ทำให้กระแสไฟทั้ง2 ขา เชื่อมต่อกันได้

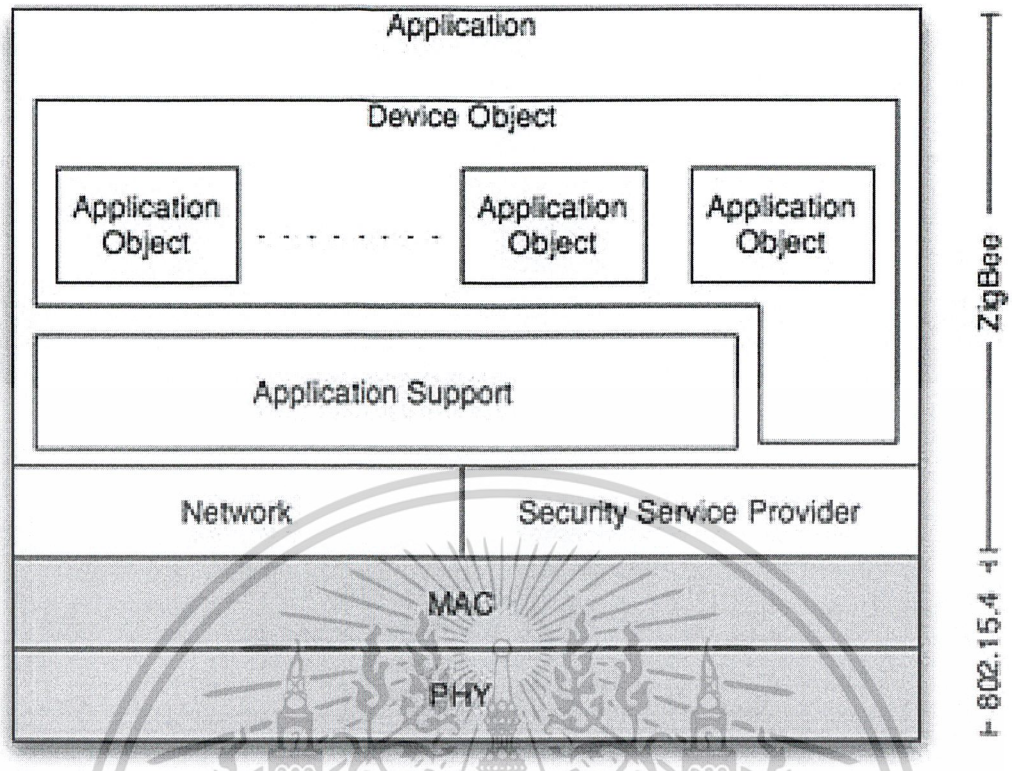


รูปที่ 3.6 IC XBee Pro

Xbee เป็น อุปกรณ์ รับ – ส่งสัญญาณแบบ Half Duplex ในย่านความถี่ 2.4 Ghz โดยใช้งานตามหลักมาตรฐาน Zigbee โดยการ Set Parameter ผ่าน Software Interface ของโปรแกรม X-TCU

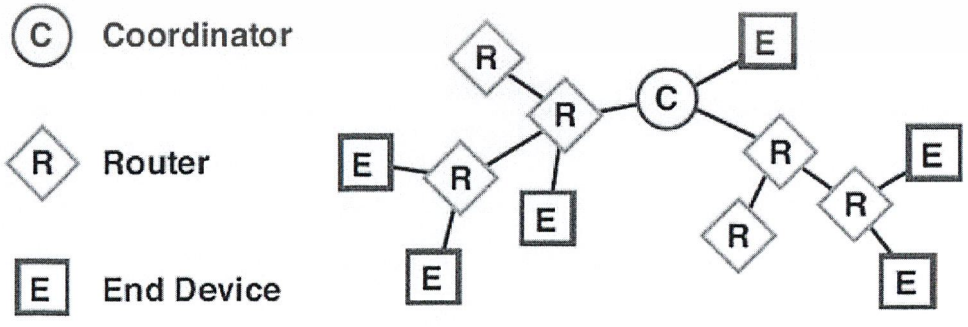
Xbee มีอยู่ 2 รุ่นคือ รุ่น series 1 (รุ่น IEEE 802.15.4) และ รุ่น series 2 (รุ่น ZNET2.5) และยังมีขนาด power ให้เลือกอีก 2 แบบ คือ แบบธรรมดา (1 mw – 2 mw) และ แบบ PRO (50mw- 60 mw) ซึ่งจะมีผลเรื่องระยะทางการรับส่งข้อมูล โดยแต่ละ Series นั้น สามารถสร้างเครือข่ายได้หลายแบบ แต่จะมีเพียง series 2 เท่านั้นที่จะทำเครือข่ายแบบ mesh ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 มาตรฐานการสื่อสารแบบ ZigBee

Zigbee คือ มาตรฐานการสื่อสารแบบไร้สาย ที่มีอัตราการรับส่งข้อมูลต่ำ ใช้พลังงานต่ำ และราคาถูก ซึ่งใช้สร้างระบบ Wireless Sensor Network ซึ่งระบบนี้จะสามารถทำงาน ได้ใน ที่ร่ม กลางแจ้ง ทนแดด ทนฝน ทั้งนี้ ZigBee ได้อ้างอิงมาตรฐานตาม IEEE 802.15.4 โดย IEEE 802.15.4 แบ่งชนิดอุปกรณ์ในเครือข่ายออกเป็น 2 ประเภท คือ FFD (Full Function Device) ซึ่งหมายถึงอุปกรณ์ที่สามารถทำงานได้ทุกอย่างในเครือข่าย และ RFD (Reduce Function Device) ซึ่งหมายถึงอุปกรณ์ที่ถูกลดความสามารถการทำงานในเครือข่าย

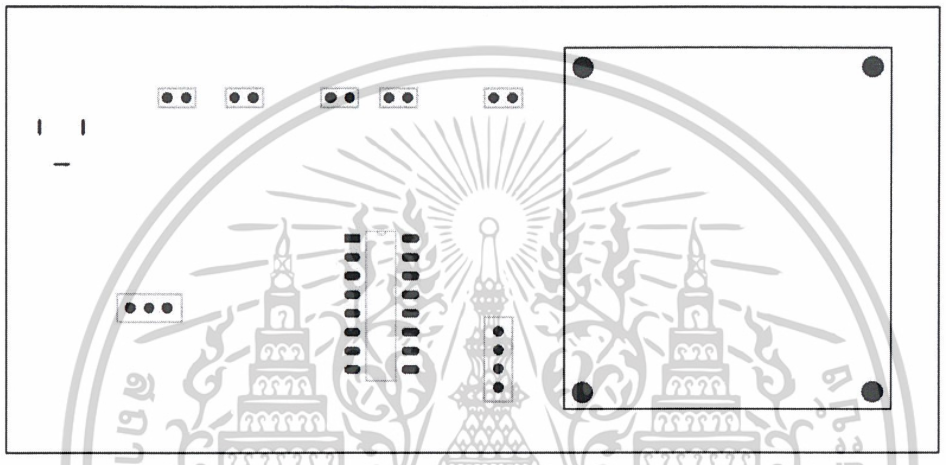


รูปที่ 3.8 การทำงานของ ZigBee

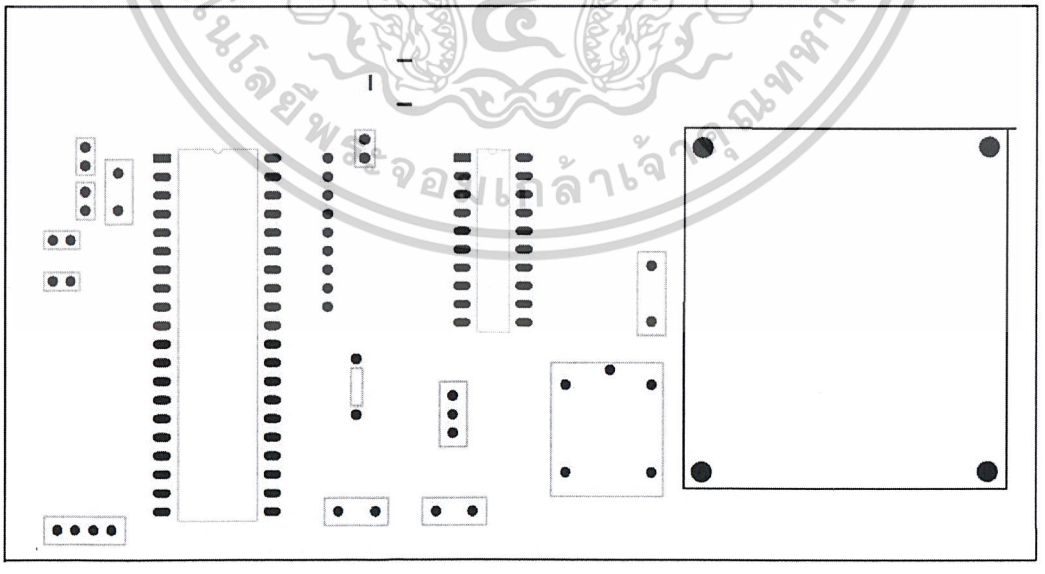
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Zigbee ได้แบ่งลักษณะการทำงาน 3 แบบ คือ

1) Coordinator มีหน้าที่สร้างการสื่อสาร เชื่อมโยงเครือข่าย ระหว่าง End Device กับ Router หรือ Coordinator กับ Coordinator ด้วยกัน หรือ Coordinator กับ Router โดยต้องกำหนด address ให้กับ device ที่อยู่ในวงเครือข่ายให้ไม่ซ้ำกัน
2) End Device เป็นอุปกรณ์ปลายทาง ซึ่งจะได้รับสัญญาณจาก Sensor ที่ปลายทาง โดยใช้พลังงานต่ำในการทำงาน
3) Router มีหน้าที่ รับส่งข้อมูลในเส้นทางต่างๆของเครือข่าย

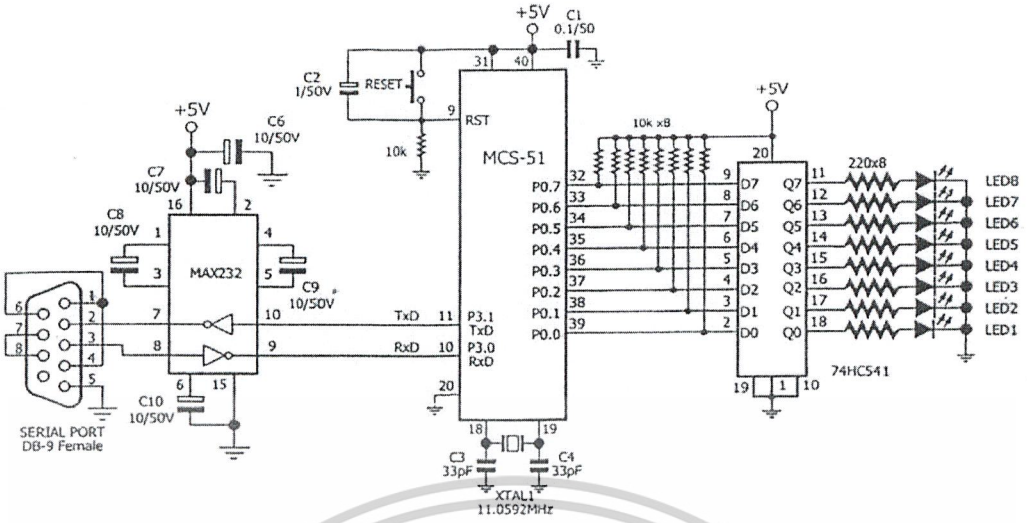


รูปที่ 3.9 ตัวอย่างแผ่นปริ้น ของฐานรองที่ต่อกับคอมพิวเตอร์



รูปที่ 3.10 ตัวอย่างแผ่นปริ้น ของฐานรองที่ต่อกับ อุปกรณ์ไฟฟ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.11 ภาพแสดงตัวอย่างการต่ออุปกรณ์ฮาร์ดแวร์

Serial port คือ port ที่ทำหน้าที่สื่อสาร แบบอนุกรม โดยที่ ถ้าต้องการส่งข้อมูลแบบอนุกรมจาก microcontroller จะส่งออกทางขา TxD หรือขา P3.1 แต่ถ้าต้องการรับข้อมูลแบบอนุกรมจาก microcontroller จะรับจากทางขา RxD หรือขา P3.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

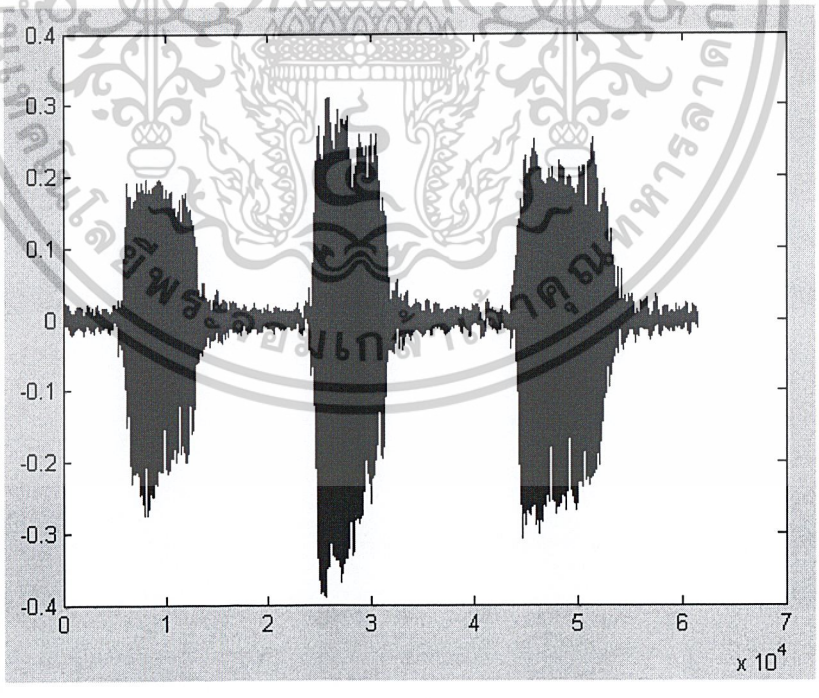
บทที่ 4
พัฒนาระบบ

4.1 การเก็บเสียงข้อมูล

การเก็บเสียงในปริญญานิพนธ์นี้ จะใช้เสียงของผู้พูดทั้งหมด 40 คน ซึ่งประกอบด้วยจำนวนคำทั้งหมด 4 คำ ได้แก่ “เปิด” “ปิด” “ไฟ” และ “รั้ว” โดยจะให้ผู้พูดแต่ละคนออกเสียงคำละ 5 ครั้ง และเลือกเสียงที่ติด noise น้อยที่สุดมาทั้งหมด 3 เสียง รวมทั้งหมด 4 คำ ได้คนละ 12 เสียง ดังนั้นปริญญานิพนธ์นี้จะมีจำนวนเสียงที่ใช้ทั้งหมด 480 เสียง โดยในการบันทึกเสียง ข้อมูลจะถูกบันทึกให้อยู่ในรูปแบบของไฟล์ข้อมูล ‘.wav’ ซึ่งข้อมูลหนึ่งตัวอย่างจะถูกแทนด้วยข้อมูลขนาด 8 บิต และใช้ความถี่ 44.1 kHz

4.2 กระบวนการเตรียมเสียงก่อนนำเข้าสู่กระบวนการการเรียนรู้

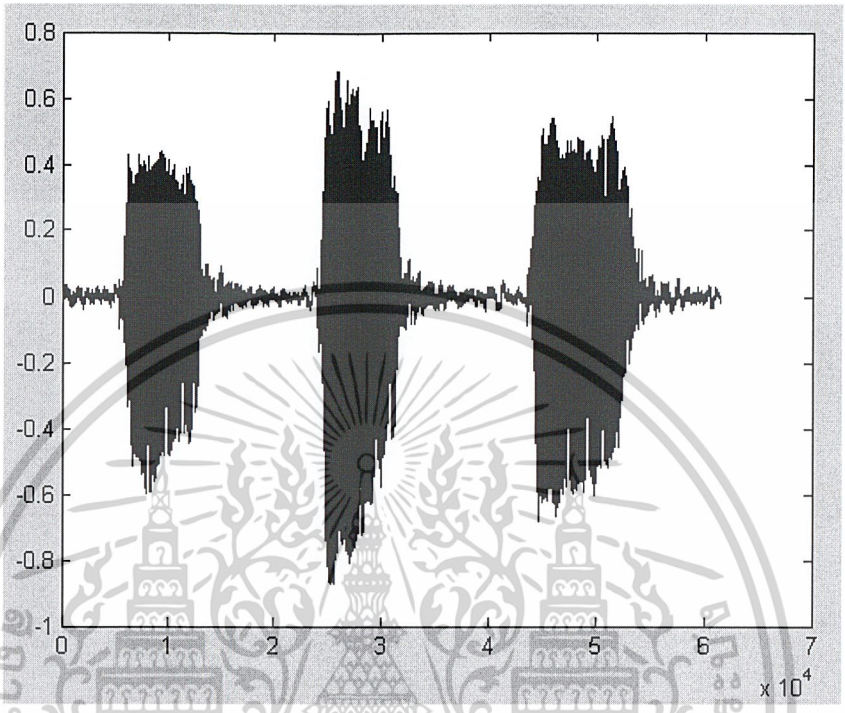
4.2.1 การตัดคำ ในครั้งแรกที่เสียงถูกบันทึกลงคอมพิวเตอร์ เสียงที่ถูกบันทึกนั้นจะมีเสียงอยู่หลายเสียง(แต่เป็นคำเดียวกัน) ดังนั้นก่อนการทำเข้าสู่การเรียนรู้ เราจึงต้องทำการตัดคำ และกำจัดสัญญาณรบกวนออกก่อน



รูปที่ 4.1 รูปแสดงตัวอย่างเสียงที่ถูกเก็บคำว่า “ปิด” 3 ครั้ง โดยยังไม่ผ่านกระบวนการใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

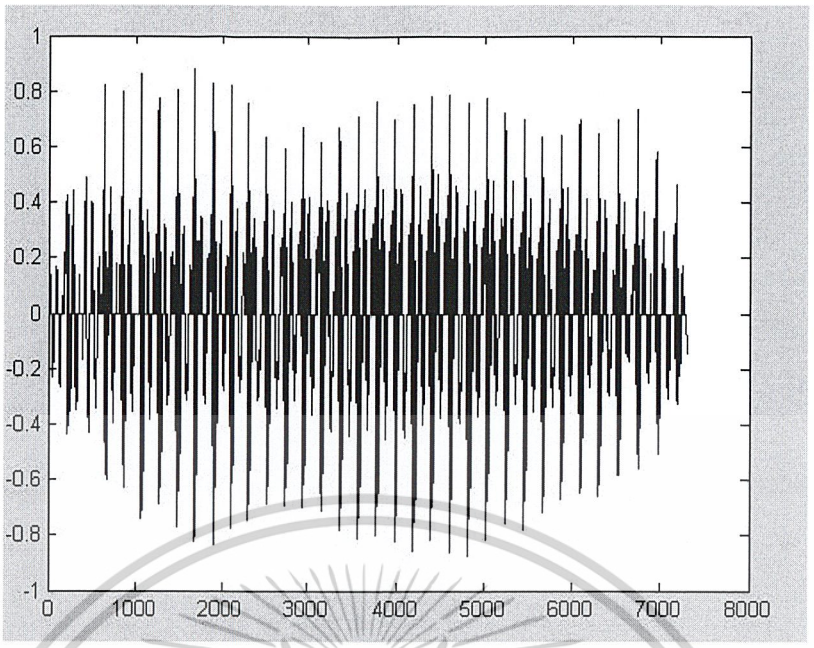
โดยกระบวนการตัดคำจะมีการทำการ Filter เสียงหรือการกรองสัญญาณเสียง เพื่อเป็นการทำให้สัญญาณเสียงที่เราไม่ต้องการถูกลดทอนให้น้อยลงหรือกำจัดออก โดยเสียงจะมีความใสมากยิ่งขึ้น



รูปที่ 4.2 รูปแสดงตัวอย่างเสียงที่ถูกเก็บคำว่า "ปิด" 3 ครั้ง โดยผ่านกระบวนการกรองสัญญาณเสียง

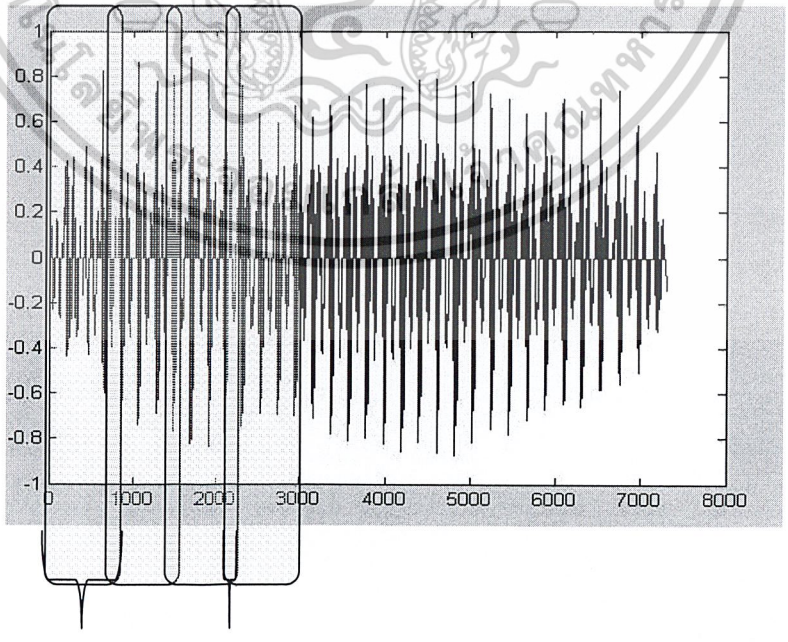
หลังจากผ่านกระบวนการกรองสัญญาณเสียง (Filter) แล้วจึงจำสัญญาณเสียงมาหาจุดอ้างอิงในการตัดคำ โดยปกติทั่วไปค่าเฉลี่ยในการตัดคำจากใช้จุดอ้างอิงที่ 5% จากค่าสูงสุดของสัญญาณเสียง ซึ่งเป็นสัญญาณเสียงที่มีสัญญาณรบกวนน้อยมาก แต่เนื่องจากในปริญญานิพนธ์นี้เสียงที่ได้รับเข้ามามีเสียงสัญญาณรบกวนค่อนข้างมาก จึงต้องทำให้กำหนดจุดอ้างอิงที่ 9% จากค่าสูงสุดของสัญญาณเสียง ซึ่งเมื่อทำนำสัญญาณเสียงเข้าสู่กระบวนการตัดเสียงเรียบร้อยแล้ว จะได้ค่าจากการตัดเสียงออกมาดังรูปที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 รูปแสดงตัวอย่างเสียง "ปิด" ที่ผ่านการ filter และผ่านการตัดเสียงแล้ว

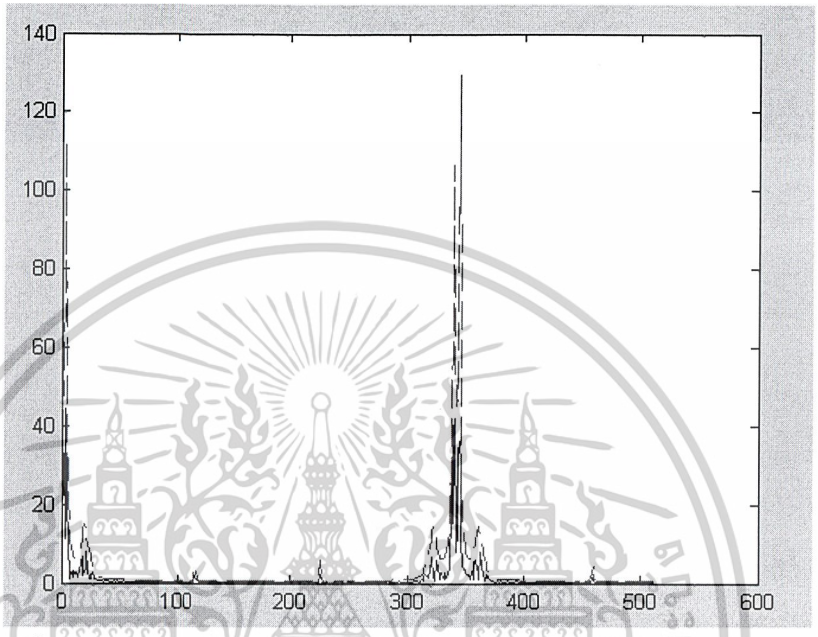
4.2.2 การเตรียมเสียง ในขั้นแรกนั้นจะมีการคำนวณเพื่อหาความยาวเสียงที่ยาวที่สุดสำหรับการนำไปใช้เป็นความยาวมาตรฐานเพื่อให้เสียงทุกเสียงที่เข้ามามีขนาดเดียวกัน รวมไปถึงคำนวณหาความยาวของ frame ที่จะใช้สำหรับการ windows แล้วนำเข้าไปสกัดหาคุณลักษณะของเสียงด้วย LPC โดยในปริญญานิพนธ์นี้ ทำการแบ่งออกเป็นเฟรมละ 20 ms หรือ 880 samples และให้คาบเกี่ยวกัน 5 ms หรือ 220 ms



รูปที่ 4.4 รูปแสดงตัวอย่างการแบ่งเฟรม (Windows)

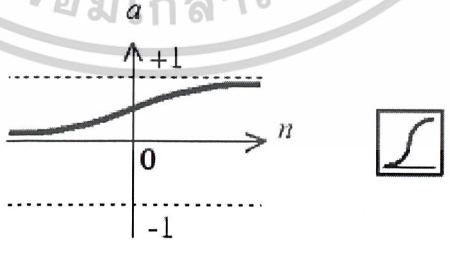
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 การสกัดหาค่าลักษณะสำคัญของเสียงด้วยการประมาณพันธะเชิงเส้น (LPC) เมื่อทำการแบ่งเฟรมเสร็จเรียบร้อยจะได้เฟรมออกมาทั้งหมด 47 เฟรม (จำนวนเฟรมที่จะใช้เป็นค่ามาตราฐานสำหรับทุกๆเสียงที่เข้ามา) จึงนำไปเป็นค่า Input ให้ LPC ซึ่งในที่นี้จะประมาณที่ค่า 90 ซึ่งสามารถทำให้ลดค่า Input ในการที่จะเข้าสู่กระบวนการเรียนรู้ด้วยเครือข่ายประสาทเทียมได้เกือบ 10 เท่า



รูปที่ 4.5 รูปแสดงตัวอย่างการกัดหาคุณลักษณะของเสียงด้วย LPC

4.2.4 เข้าสู่กระบวนการเรียนรู้ นำเสียงที่ถูกสกัดหาคุณลักษณะของเสียงด้วย LPC ทุกตัวเข้าสู่กระบวนการเรียนรู้แบบ back-propagation โดยมี Input Node ทั้งหมด 4,277 , Hidden Layer 100 และ Output 4 โหนด และใช้ Transfer function แบบ log-sigmoid ซึ่งมีลักษณะดังนี้



$a = logsig(n)$

Log-Sigmoid Transfer Function

รูปที่ 4.6 รูปแสดง Log-Sigmoid Transfer Function

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งการที่ปริญญานิพนธ์ฉบับนี้ได้เลือกการใช้ log-sigmoid function เนื่องจากoutputที่ต้องการนั้นอยู่ในช่วงระหว่าง 0-1 จึงเป็นfunction ที่มีความเหมาะสมต่อการนำมาใช้มากที่สุด

ลักษณะการกำหนด Output ในแต่ละ Node

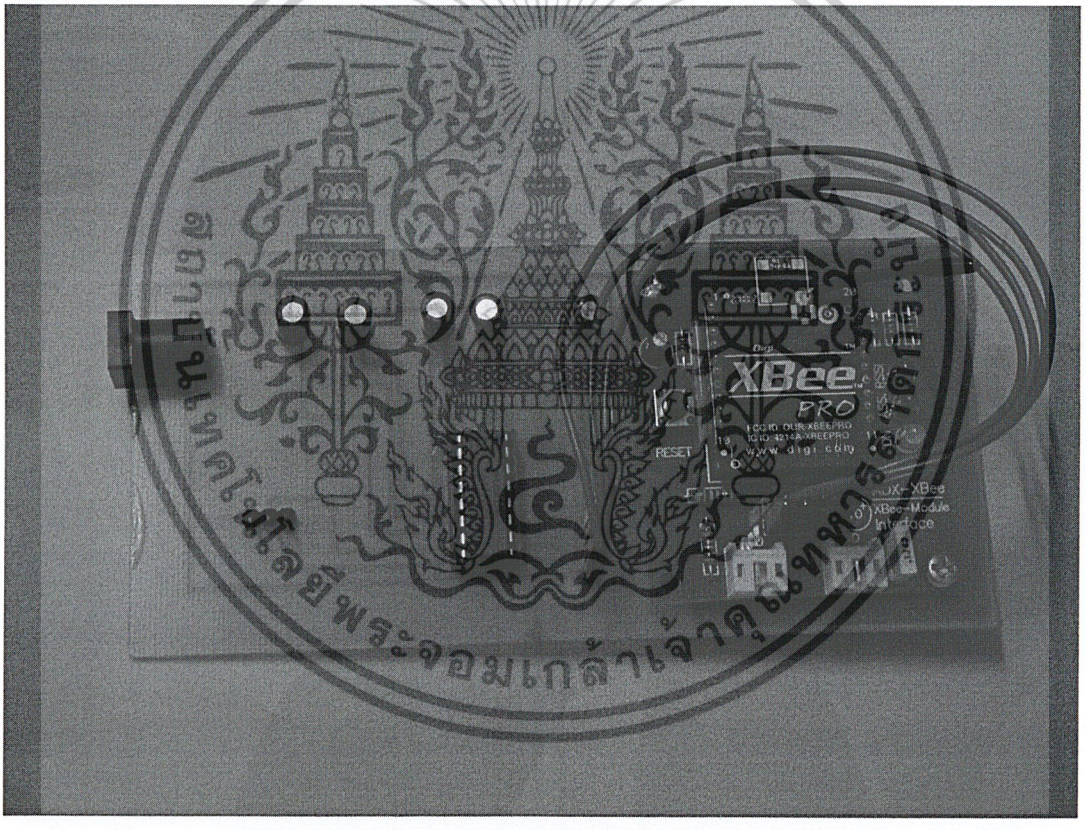
Input = 'เปิด' Output = [0;0;0;0;1]

Input = 'ปิด' Output = [0;0;0;1;0]

Input = 'ไฟ' Output = [0;0;1;0;0]

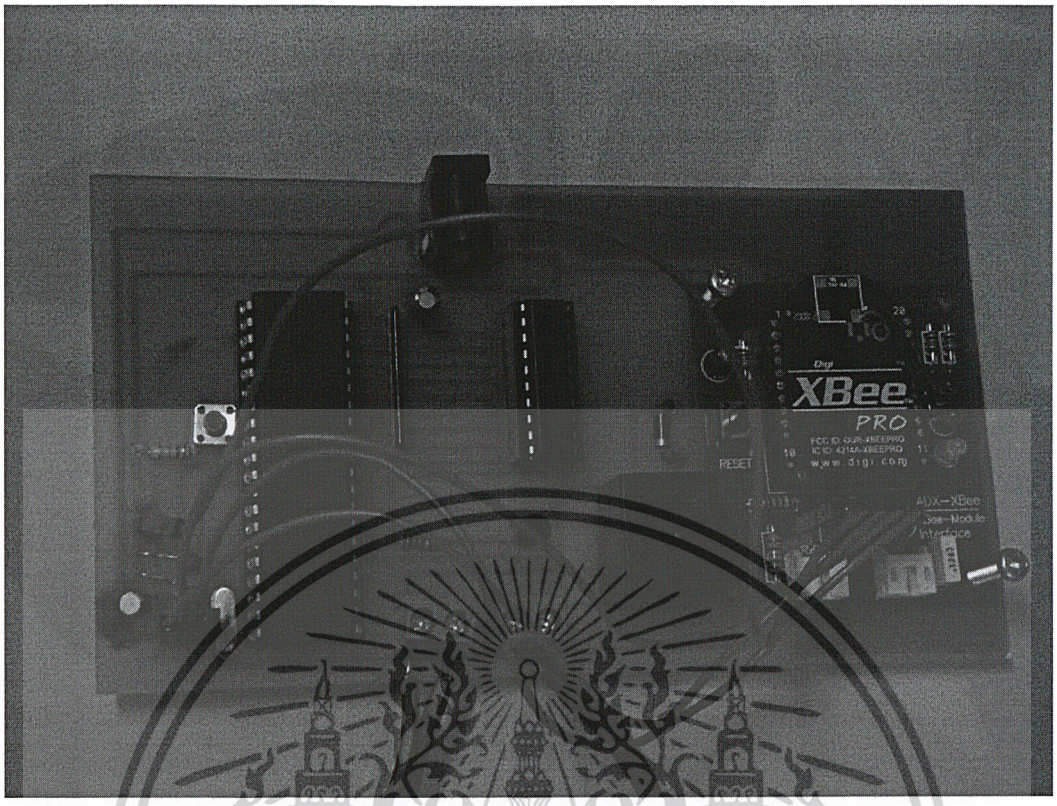
Input = 'รั้ว' Output = [0;1;0;0;0]

4.3 อุปกรณ์ Hardware ที่ทำการต่อกับแผ่นปริ้นแล้ว



รูปที่ 4.7 รูปแสดงอุปกรณ์ตัวส่งที่ต่อกับคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.8 รูปแสดงอุปกรณ์ตั วรับที่ต่อกับอุปกรณ์ไฟฟ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลโครงงาน

5.1 สรุปผลโครงงาน

ในปริญญานิพนธ์ฉบับนี้ได้นำเสนอการเรียนรู้เสียงด้วยกระบวนการเรียนรู้แบบ back-propagation โดยมีการนำสัญญาณเสียงที่ถูกบันทึกลงคอมพิวเตอร์มาทำการเตรียมเสียงสัญญาณด้วยการทำการ filter หรือการกรองสัญญาณเพื่อลดทอนหรือกำจัดสัญญาณเสียงที่ไม่ต้องการและสกัดหาค่าลักษณะเสียงสำคัญด้วยวิธีการประมาณเชิงพันธะ (LPC) โดยสรุปการเรียนรู้ออกเป็นตารางดังนี้

ตารางที่ 5.1 สรุปผลการทดลอง

เสียง(Input)	จำนวนเสียงในการทดสอบ	อัตราความถูกต้อง
เปิด	120	25%
ปิด	120	93%
ไฟ	120	53%
รั้ว	120	90%

5.2 ปัญหาและอุปสรรค

จากผลการทดลองพบว่ามีสิ่งที่ส่งผลต่อการรู้จำเสียงของระบบหลายอย่าง ดังนี้

เสียงที่นำมาประมวลผลในขั้นตอนของการเรียนรู้ มีเสียงรบกวนแทรกอยู่ในระดับปานกลาง จึงทำให้การตัดสัญญาณเสียงมีความคลาดเคลื่อน ส่งผลให้การเรียนรู้เสียงมีความผิดพลาดสูง ค่า Output ในส่วนของ Neural Network จึงผิดเพี้ยนไปจากเดิมค่อนข้างมาก

ความผิดพลาดของแต่ละคำที่นำมาเรียนรู้ อาจเกิดจากมีลักษณะของคำ โดยเฉพาะคำที่เป็น “คำเปิด” หรือ “เสียงยาว” เช่น “เปิด” หรือ “ไฟ” จะทำให้เกิดข้อผิดพลาดในการเรียนรู้สูงกว่า คำ ที่มีลักษณะเป็น “คำปิด” หรือ “เสียงสั้น” เช่น “ปิด” เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นอกจากนี้สถานที่ๆทำการทดสอบรวมถึงการเก็บเสียงในขั้นตอนการนำมาเรียนรู้มีส่วนอย่างมาก โดยการเก็บทดลองหรือเก็บเสียงใจสภาพห้องเงียบจะให้ผลที่ดีกว่าสภาพห้องที่มีเสียงรบกวน เนื่องจากสัญญาณคำสั่งเสียงภายใต้ห้องที่มีเสียงรบกวน จะถูกแทรกซ้อนด้วยสัญญาณรบกวนภายในห้องด้วย จึงทำระบบอาจประมวลผลผิดพลาด ยิ่งสัญญาณรบกวนมาก ระบบยิ่งผิดพลาดสูง

5.3 ข้อเสนอแนะ

- การเก็บเสียง ควรเก็บเสียงที่ปราศจากเสียงหรือสัญญาณรบกวนให้มากที่สุด เพื่อความถูกต้องในการเรียนรู้ด้วยเครือข่ายประสาทเทียม
- ควรเก็บเสียงให้มากที่สุดเพื่อการประมวลผลที่แม่นยำมากยิ่งขึ้น
- เสียงที่สั่งการและเสียงที่เก็บควรออกเสียงอย่างชัดเจน
- อุปกรณ์ที่ใช้บันทึกเสียงควรมีคุณภาพที่ดี เพื่อให้ได้คุณภาพเสียงที่ดีและมีความถูกต้องแม่นยำมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] S N Sivanandam, S Sumathi and S N Deepa. 2006. **Introduction to Neural Networks using Matlab 6.0**. New Delhi: Tata McGraw-Hill Publishing Company Limited.

[2] ศิริพร มณีชูเกตุ. 2542. "**ภาษาศาสตร์เบื้องต้น.**" กรุงเทพฯ : มหาวิทยาลัยนเรศวร. เอกสารประกอบการสอน

[3] ศักรินทร์ สินไชย, ปานวิทย์ ธุวะนุติ, จีรสุดา โกษียาภรณ์ และ ปราโมทย์ วาดเขียน. 2550. "การรู้จำคำพูด โดยใช้โครงข่ายประสาทเทียมด้วยการสกัดลักษณะเด่นด้วยสัมประสิทธิ์แอลพีซี." หน้า 603-610. ใน **การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 45 สาขาสถาปัตยกรรมศาสตร์และวิศวกรรมศาสตร์ สาขาทรัพยากรธรรมชาติและสิ่งแวดล้อม**. กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์.

[4] นภัส ศรียงค์ และ นิภารัตน์ ธนประโยชน์ศักดิ์. 2549. "การรู้จำเสียง." ปริญญานิพนธ์วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมสารสนเทศ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

[5] หรรษธร ธีรรัตน์รักษ์ และ กรกช เรืองฉาย. 2549. "ระบบสั่งงานด้วยเสียงพูด." ปริญญานิพนธ์ วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

[6] Phattraphong Racknoi. 2008. **Sound**. [Online]. Available: http://km.pbn1.org/research/1250214847_sound.pdf.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

(Code โปรแกรม Matlab)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. cutVoice.m ไฟล์สำหรับการกรองสัญญาณเสียงก่อนการตัดคำออกเป็นสองคำ

```
function sound = cutVoice( rec )
   [x,fs] = wavread(rec);
   y = x(:,[1]);
   signal = y;
   for i=3:length(y)
      signal(i) = y(i)+y(i-1)+0.3*y(i-2);
   end
   %-- end of filter --%
   meann = 0.0;
   meann = mean(signal);
   sigMean = signal - meann;
   maxS = max(sigMean);
   minS = min(sigMean);

   newSig = sigMean;

   maxCut = (maxS*9)/100;
   minCut = (minS*9)/100;

   for i=1:length(newSig)
      if newSig(i) > 0
         if newSig(i) < maxCut
            newSig(i) = 0;
         end
      else
         if newSig(i) > minCut
            newSig(i) = 0;
         end
      end
   end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
start = 0;j=1;dis=0;limitDis=1000;check=0;
pointS = zeros(3,2);
powerNew = newSig;

for i=1:length(powerNew)
   if start == 0
      if powerNew(i)~= 0
         start = 1;
         pointS(j,1) = i;
      end
   elseif start == 1
      if powerNew(i) == 0
         if dis < limitDis
            dis = dis+1;
            check = 1;
         elseif (i - pointS(j,1)) < 6000 %-- protect noise ,if rang sound < 6000 is noise
            pointS(j,1) = 0;
            start = 0; dis = 0; check = 0;
         else pointS(j,2) = i-dis;j = j+1; start = 0; dis = 0; check = 0;
         end
      elseif powerNew(i) ~= 0 & check == 1;
         dis = 0; check = 0;
      end
   end
end
str = 'word';
for i=1:2
   name = strcat(str,num2str(i));
   unitSig = powerNew(pointS(i,1):pointS(i,2)+1,1);
   wavwrite(unitSig,fs,name);
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
a = importdata('word1.wav');
word1 = a.data;
a = importdata('word2.wav');
word2 = a.data;

sound = struct('word1',word1,'word2',word2);

end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. findMax.m ไฟล์สำหรับการหาค่ามาตรฐานก่อนตัดไฟล์เสียง

```
clear;
load allSound.mat

n = 1;
nu = 1;
maxLength = 0;

for num=1:600 %-- amount of voice
 nu = mod(num,5);
 if nu ==1
  str = 'open';
 elseif nu ==2
  str = 'close';
 elseif nu == 3
  str = 'light';
 elseif nu == 4
  str = 'bar';
 else
  str = 'stove';
 end

 name = sound.(strcat(str,num2str(n)));

 if strcmp(str,'stove')
  n = n+1;
 end

 if maxLength < length(name)
  maxLength = length(name);
  maxSound = name;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 end
end


%-- end of find max length --%


diffMax = maxLength - 880;
over = mod(diffMax,660);
if over ~= 0
 len = (660-over)+maxLength;
 buff = zeros(len,1);
else buff = zeros(maxLength,1);
end



buff(1:maxLength,1) = maxSound;
%-- end of set length for windows --%


lenMax = length(buff);


numFrame = 1+((lenMax-880)/660);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. bpnn.m ไฟล์สำหรับการค่าน้ำหนักของเสียงแต่ละเสียงก่อนนำไปใช้งานจริง

```
clear;
load findMax.mat

n = 1;
num = 1;
nu =1;

w1 = rand([100,4277])-0.5;
w2 = rand([5,100])-0.5;

for num =1:600
 nu = mod(num,5);
 if nu ==1
 str = 'open';
 desOut = [0;0;0;0;1];
 elseif nu ==2
 str = 'close';
 desOut = [0;0;0;1;0];
 elseif nu == 3
 str = 'light';
 desOut = [0;0;1;0;0];
 elseif nu == 4
 str = 'bar';
 desOut = [0;1;0;0;0];
 else
 str = 'stove';
 desOut = [1;0;0;0;0];
 end

 signal = sound.(strcat(str,num2str(n)));
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
if strcmp(str,'stove')
 n = n+1;
end

y = signal/max(signal);
%-- end of scaling size --%
z = length(y);
des = zeros(lenMax,1);
score = 0;
j = 0;
checkP = 0;
temp = 0;
sw = 1;
if z >= lenMax/2
 disPoint = (z - mod(z,(lenMax-z)))/(lenMax-z);
  for i=1:lenMax
   if score == disPoint & checkP <= lenMax-z-1
    des(i) = 0;
    score = 0;
    checkP = checkP+1;
   else
    j = j+1;
    des(i) = y(j);
    score = score+1;
   end
  end
else
 disPoint = (z - mod(z,((lenMax-z) - z)))/((lenMax-z)-z);
 for i=1:lenMax
  if sw == 0
   if score == disPoint & checkP <= (lenMax-z)-z-1
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
      des(i) = 0;
      score = 0;
      checkP = checkP+1;
      sw = 1;
      temp = 1;
     else
      des(i) = 0;
      score = score+1;
      sw = 1;
     end
    else
     if temp == 1
      des(i) = 0;
      temp = 0;
     else
      j = j+1;
      if j > z break;
      end
      des(i) = y(j);
      sw = 0;
     end
   end
  end
 end
%-- end of scaling time --%

%-- windows --%
%-- numFrame = xxx will from findMax --%
 st = 1;
 en = 0;
 for i=1:numFrame
  if st == 1
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
  ham.(strcat('ham',num2str(i))) = des(st : 880);
  st = 880;
 else
  en = st+660;
  ham.(strcat('ham',num2str(i))) = des(st-219:en);
  st = en;
 end
end
%-- end --%

allLpc = [];
 for i=1:numFrame
  [aLpc g] = lpc(ham.(strcat('ham',num2str(i))),90);
  allLpc = [allLpc aLpc];
 end
%-- end LPC --%
for i=1:length(allLpc)
 if isnan(allLpc(1,i))
  allLpc(1,i) = 0;
 end
end

input = allLpc';
learning_rate = 0.3;
epoch = 0;
error = 1000;
while error > 0.007 & epoch < 25000
real_hide_output = logsig(w1*input);
real_output = logsig(w2*real_hide_output);
delta_output_layer = real_output.*(1-real_output).*(desOut-real_output);
new_W2 = w2+(2*learning_rate)*delta_output_layer*real_hide_output';
delta1 = ((delta_output_layer*real_hide_output')*(1-real_hide_output))'*w2;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
del_W1 = (2*learning_rate)*delta1'*input';
new_W1 = w1 + del_W1;
w1 = new_W1;
w2 = new_W2;
epoch = epoch+1;
e = abs(desOut-logsig(w2*logsig(w1*input)));
error = sum(sum(e));
end
strcat(str,num2str(n))
desOut
real_output
%-- end of NN --%
end

real_w1 = w1;
real_w2 = w2;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. nn.m ไฟล์สำหรับการเรียนรู้และตัดสินใจเมื่อนำไปใช้งานจริง

```
function outStr = nn( sound1 )
    load findMax.mat
    load real_w.mat

    out = [0;0;0;0;0];
    outStr = '';
    sav = [0;0];
    sound = sound1;
    for num =1:2

      signal = sound.(strcat('word',num2str(num)));

      y = signal/max(signal);
      %-- end of scaling size --%

      z = length(y);
      des = zeros(lenMax,1);
      score = 0;
      j = 0;
      checkP = 0;
      temp = 0;

      sw = 1;

      if z >= lenMax/2
        disPoint = (z - mod(z,(lenMax-z)))/(lenMax-z);
        for i=1:lenMax
          if score == disPoint & checkP <= lenMax-z-1
            des(i) = 0;
            score = 0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            checkP = checkP+1;
        else
            j = j+1;
            des(i) = y(j);
            score = score+1;
        end
    end
else
    disPoint = (z - mod(z,((lenMax-z) - z)))/((lenMax-z)-z);
    for i=1:lenMax
        if sw == 0
            if score == disPoint & checkP <= (lenMax-z)-z-1
                des(i) = 0;
                score = 0;
                checkP = checkP+1;
                sw = 1;
                temp = 1;
            else
                des(i) = 0;
                score = score+1;
                sw = 1;
            end
        else
            if temp == 1
                des(i) = 0;
                temp = 0;
            else
                j = j+1;
                if j > z break;
                end
            des(i) = y(j);
            sw = 0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            end
          end
        end
      end
%-- end of scaling time --%

%-- windows --%
%-- numFrame = xxx will from findMax --%
   st = 1;
   en = 0;
   for i=1:numFrame
      if st == 1
         ham.(strcat('ham',num2str(i))) = des(st : 880);
         st = 880;
      else
         en = st+660;
         ham.(strcat('ham',num2str(i))) = des(st-219:en);
         st = en;
      end
   end

   %-- end --%

   allLpc = [];

   for i=1:numFrame
      [aLpc g] = lpc(ham.(strcat('ham',num2str(i))),90);
      allLpc = [allLpc aLpc];
   end

   %-- end LPC --%
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    for i=1:length(allLpc)
      if isnan(allLpc(1,i))
        allLpc(1,i) = 0;
      end
    end

    input = allLpc';
    rHide_output = logsig(real_w1*input);
    rOutput = logsig(real_w2*rHide_output);
    an = rOutput(5,1);
    j = 4;

    while j ~=0
      if an < rOutput(j,1)
        sav(num,1) = j;
        an = rOutput(j,1);
      end
      j = j-1;
    end
  end

  if sav(1,1) == 1 | sav(1,1) == 2 | sav(1,1) == 3 | sav(2,1) == 4 | sav(2,1) == 5
    outStr = 'Please try again';
  else
    out(sav(1,1)) = 1;
    out(sav(2,1)) = 1;
    for i=1:5
      ss = num2str(out(i,1));
      outStr = strcat(outStr,ss);
    end
  end
end
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

(Code ควบคุมอุปกรณ์ฮาร์ดแวร์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. Code ควบคุมอุปกรณ์รับสัญญาณ อุปกรณ์ที่ 1

```
ORG   0000H
      MOV   P0,#00000000B
RX:   MOV   PCON,#00H
      MOV   SCON,#50H
      MOV   TMOD,#20H
      MOV   TH1,#0FDH
LOOP: SETB  TR1
WAIT: JNB   RI,WAIT
      MOV   A,SBUF

CHK1: CJNE  A,#31H,CHK2
      MOV   P0,#00000001B
      SJMP  T
CHK2: CJNE  A,#32H,CHK3
      MOV   P0,#00000010B
      SJMP  T
T:    CLR   RI
      CLR   TR1
      SJMP  LOOP
END
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. Code ควบคุมอุปกรณ์รับสัญญาณ อุปกรณ์ที่ 2

```
ORG   0000H
      MOV   P0,#00000000B
RX:   MOV   PCON,#00H
      MOV   SCON,#50H
      MOV   TMOD,#20H
      MOV   TH1,#0FDH
LOOP: SETB  TR1
WAIT: JNB   RI,WAIT
      MOV   A,SBUF

CHK1: CJNE  A,#33H,CHK2
      MOV   P0,#00000001B
      SJMP  T
CHK2: CJNE  A,#34H,CHK3
      MOV   P0,#00000010B
      SJMP  T
T:    CLR   RI
      CLR   TR1
      SJMP  LOOP
END
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.

(Data Sheet)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

XBee™/XBee-PRO™ OEM RF Modules

XBee/XBee-PRO OEM RF Modules
RF Module Operation
RF Module Configuration
Appendices

Product Manual v1.xAx - 802.15.4 Protocol
For OEM RF Module Part Numbers: XB24-...-001, XBP24-...-001

IEEE® 802.15.4 OEM RF Modules by MaxStream, Inc.

MaxStream®
355 South 520 West, Suite 180
Lindon, UT 84042
Phone: (801) 765-9885
Fax: (801) 765-9895

rf-xperts@maxstream.net
www.MaxStream.net (live chat support)

M100232
2006.10.13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

© 2006 MaxStream, Inc. All rights reserved

The contents of this manual may not be transmitted or reproduced in any form or by any means without the written permission of MaxStream, Inc.

XBee™ and XBee-PRO™ are trademarks of MaxStream, Inc.

Technical Support:

Phone: (801) 765-9885

Live Chat: www.maxstream.net

E-mail: rf-xperts@maxstream.net

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Contents

1. XBee/XBee-PRO OEM RF Modules 4

1.1. Key Features 4
1.1.1. Worldwide Acceptance 4
1.2. Specifications 5
1.3. Mechanical Drawings 6
1.4. Mounting Considerations 6
1.5. Pin Signals 7
1.6. Electrical Characteristics 8

2. RF Module Operation 9

2.1. Serial Communications 9
2.1.1. UART Data Flow 9
2.1.2. Transparent Operation 10
2.1.3. API Operation 10
2.1.4. Flow Control 11
2.2. ADC and Digital I/O Line Support 12
2.2.1. I/O Data Format 12
2.2.2. API Support 13
2.2.3. Sleep Support 13
2.2.4. DIO Pin Change Detect 13
2.2.5. Sample Rate (Interval) 13
2.2.6. I/O Line Passing 14
2.2.7. Configuration Example 14
2.3. XBee/XBee-PRO Networks 15
2.3.1. NonBeacon 15
2.3.2. NonBeacon (w/ Coordinator) 15
2.3.3. Association 16
2.4. XBee/XBee-PRO Addressing 19
2.4.1. Unicast Mode 19
2.4.2. Broadcast Mode 19
2.5. Modes of Operation 20
2.5.1. Idle Mode 20
2.5.2. Transmit/Receive Modes 20
2.5.3. Sleep Mode 22
2.5.4. Command Mode 24

3. RF Module Configuration 25

3.1. Programming the RF Module 25
3.1.1. Programming Examples 25
3.2. Command Reference Tables 26
3.3. Command Descriptions 33
3.4. API Operation 52
3.4.1. API Frame Specifications 52
3.4.2. API Types 53

Appendix A: Agency Certifications 57

United States (FCC) 57
OEM Labeling Requirements 57
FCC Notices 57
FCC-Approved Antennas (2.4 GHz) 58
Europe (ETSI) 59
OEM Labeling Requirements 59
Restrictions 59
Declarations of Conformity 59
Approved Antennas 60
Canada (IC) 60
Labeling Requirements 60
Japan 60
Labeling Requirements 60

Appendix B: Development Guide 61

Development Kit Contents 61
Interfacing Options 61
RS-232 Development Board 62
External Interface 62
RS-232 Pin Signals 63
Wiring Diagrams 64
Adapters 65
USB Development Board 66
External Interface 66
USB Pin Signals 66
X-CTU Software 67
Installation 67
Serial Communications Software 67

Appendix C: Additional Information 68

1-Year Warranty 68
Ordering Information 68
Contact MaxStream 69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. XBee/XBee-PRO OEM RF Modules

The XBee and XBee-PRO OEM RF Modules were engineered to meet IEEE 802.15.4 standards and support the unique needs of low-cost, low-power wireless sensor networks. The modules require minimal power and provide reliable delivery of data between devices.

The modules operate within the ISM 2.4 GHz frequency band and are pin-for-pin compatible with each other.

1.1. Key Features

Long Range Data Integrity

XBee
- Indoor/Urban: up to 100' (30 m)
- Outdoor line-of-sight: up to 300' (100 m)
- Transmit Power: 1 mW (0 dBm)
- Receiver Sensitivity: -92 dBm

XBee-PRO
- Indoor/Urban: up to 300' (100 m)
- Outdoor line-of-sight: up to 1 mile (1500 m)
- Transmit Power: 100 mW (20 dBm) EIRP
- Receiver Sensitivity: -100 dBm

RF Data Rate: 250,000 bps

Low Power

XBee
- TX Current: 45 mA (@3.3 V)
- RX Current: 50 mA (@3.3 V)
- Power-down Current: < 10 µA

XBee-PRO
- TX Current: 215 mA (@3.3 V)
- RX Current: 55 mA (@3.3 V)
- Power-down Current: < 10 µA

ADC and I/O line support

Analog-to-digital conversion, Digital I/O

I/O Line Passing

Advanced Networking & Security

Retries and Acknowledgements

DSSS (Direct Sequence Spread Spectrum)

Each direct sequence channels has over 65,000 unique network addresses available

Source/Destination Addressing

Unicast & Broadcast Communications

Point-to-point, point-to-multipoint and peer-to-peer topologies supported

Coordinator/End Device operations

Easy-to-Use

No configuration necessary for out-of box RF communications

Free X-CTU Software (Testing and configuration software)

AT and API Command Modes for configuring module parameters

Extensive command set

Small form factor

Free & Unlimited RF-XPert Support

1.1.1. Worldwide Acceptance

FCC Approval (USA) Refer to Appendix A [p57] for FCC Requirements.
Systems that contain XBee/XBee-PRO RF Modules inherit MaxStream Certifications.

ISM (Industrial, Scientific & Medical) **2.4 GHz frequency band**

Manufactured under **ISO 9001:2000** registered standards

XBee/XBee-PRO RF Modules are optimized for use in the **United States**, **Canada**, **Australia**, **Israel and Europe**. Contact MaxStream for complete list of government agency approvals.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2. Specifications

Table 1-01. Specifications of the XBee/XBee-PRO OEM RF Modules

Specification	XBee	XBee-PRO
Performance		
Indoor/Urban Range	up to 100 ft. (30 m)	Up to 300' (100 m)
Outdoor RF line-of-sight Range	up to 300 ft. (100 m)	Up to 1 mile (1500 m)
Transmit Power Output (software selectable)	1mW (0 dBm)	60 mW (18 dBm) conducted, 100 mW (20 dBm) EIRP*
RF Data Rate	250,000 bps	250,000 bps
Serial Interface Data Rate (software selectable)	1200 - 115200 bps (non-standard baud rates also supported)	1200 - 115200 bps (non-standard baud rates also supported)
Receiver Sensitivity	-92 dBm (1% packet error rate)	-100 dBm (1% packet error rate)
Power Requirements		
Supply Voltage	2.8 – 3.4 V	2.8 – 3.4 V
Transmit Current (typical)	45mA (@ 3.3 V)	If PL=0 (10dBm): 137mA(@3.3V), 139mA(@3.0V) PL=1 (12dBm): 155mA (@3.3V), 153mA(@3.0V) PL=2 (14dBm): 170mA (@3.3V), 171mA(@3.0V) PL=3 (16dBm): 188mA (@3.3V), 195mA(@3.0V) PL=4 (18dBm): 215mA (@3.3V), 227mA(@3.0V)
Idle / Receive Current (typical)	50mA (@ 3.3 V)	55mA (@ 3.3 V)
Power-down Current	< 10 µA	< 10 µA
General		
Operating Frequency	ISM 2.4 GHz	ISM 2.4 GHz
Dimensions	0.960" x 1.087" (2.438cm x 2.761cm)	0.960" x 1.297" (2.438cm x 3.294cm)
Operating Temperature	-40 to 85° C (industrial)	-40 to 85° C (industrial)
Antenna Options	Integrated Whip, Chip or U.FL Connector	Integrated Whip, Chip or U.FL Connector
Networking & Security		
Supported Network Topologies	Point-to-point, Point-to-multipoint & Peer-to-peer	
Number of Channels (software selectable)	16 Direct Sequence Channels	12 Direct Sequence Channels
Addressing Options	PAN ID, Channel and Addresses	PAN ID, Channel and Addresses
Agency Approvals		
United States (FCC Part 15.247)	OUR-XBEE	OUR-XBEEPRO
Industry Canada (IC)	4214A XBEE	4214A XBEEPRO
Europe (CE)	ETSI	ETSI (Max. 10 dBm transmit power output)*
Japan	n/a	005NYCA0378 (Max. 10 dBm transmit power output)**

* When operating in Europe: XBee-PRO RF Modules must be configured to operate at a maximum transmit power output level of 10 dBm. The power output level is set using the PL command. The PL parameter must equal "0" (10 dBm).

Additionally, European regulations stipulate an EIRP power maximum of 12.86 dBm (19 mW) for the XBee-PRO and 12.11 dBm for the XBee when integrating high-gain antennas.

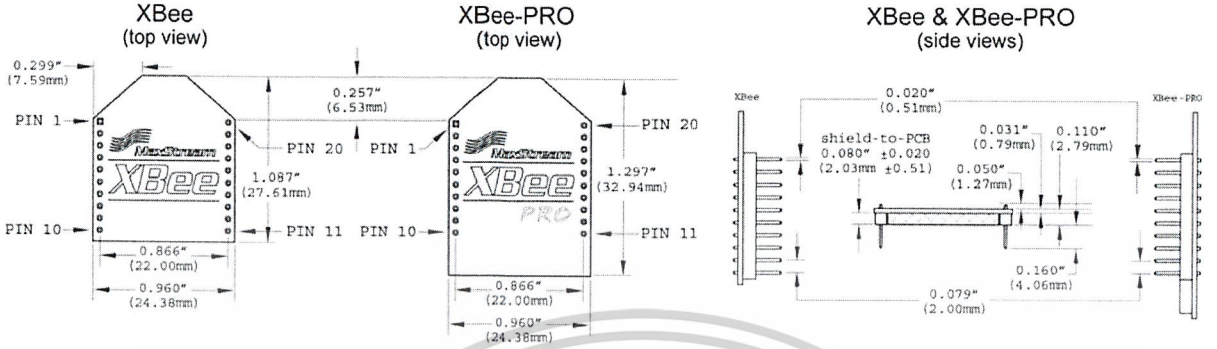
** When operating in Japan: Transmit power output is limited to 10 dBm. A special part number is required when ordering modules approved for use in Japan. Contact MaxStream for more information [call 1-801-765-9885 or send e-mails to sales@maxstream.net].

Antenna Options: The ranges specified are typical when using the integrated Whip (1.5 dBi) and Dipole (2.1 dBi) antennas. The Chip antenna option provides advantages in its form factor; however, it typically yields shorter range than the Whip and Dipole antenna options when transmitting outdoors. For more information, refer to the "XBee Antenna" application note located on MaxStream's web site (http://www.maxstream.net/support/knowledgebase/article.php?kb=153).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3. Mechanical Drawings

Figure 1-01. Mechanical drawings of the XBee/XBee-PRO OEM RF Modules (antenna options not shown)
The XBee and XBee-PRO RF Modules are pin-for-pin compatible.



1.4. Mounting Considerations

The XBee/XBee-PRO RF Module was designed to mount into a receptacle (socket) and therefore does not require any soldering when mounting it to a board. The XBee Development Kits contain RS-232 and USB interface boards which use two 20-pin receptacles to receive modules.

Figure 1-02. XBee Module Mounting to an RS-232 Interface Board.

The receptacles used on MaxStream development boards are manufactured by Century Interconnect. Several other manufacturers provide comparable mounting solutions; however, MaxStream currently uses the following receptacles:

- Through-hole single-row receptacles -
  Samtec P/N: MMS-110-01-L-SV (or equivalent)
- Surface-mount double-row receptacles -
  Century Interconnect P/N: CPRMSL20-D-0-1 (or equivalent)
- Surface-mount single-row receptacles -
  Samtec P/N: SMM-110-02-SM-S

MaxStream also recommends printing an outline of the module on the board to indicate the orientation the module should be mounted.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5. Pin Signals

Figure 1-03. XBee/XBee-PRO RF Module Pin Numbers
(top sides shown - shields on bottom)

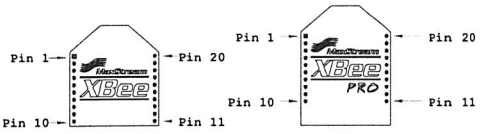


Table 1-02. Pin Assignments for the XBee and XBee-PRO Modules
(Low-asserted signals are distinguished with a horizontal line above signal name.)

Pin #	Name	Direction	Description
1	VCC	-	Power supply
2	DOUT	Output	UART Data Out
3	DIN / CONFIG	Input	UART Data In
4	DO8*	Output	Digital Output 8
5	RESET	Input	Module Reset (reset pulse must be at least 200 ns)
6	PWM0 / RSSI	Output	PWM Output 0 / RX Signal Strength Indicator
7	PWM1	Output	PWM Output 1
8	[reserved]	-	Do not connect
9	DTR / SLEEP_RQ / DI8	Input	Pin Sleep Control Line or Digital Input 8
10	GND	-	Ground
11	AD4 / DIO4	Either	Analog Input 4 or Digital I/O 4
12	CTS / DIO7	Either	Clear-to-Send Flow Control or Digital I/O 7
13	ON / SLEEP	Output	Module Status Indicator
14	VREF	Input	Voltage Reference for A/D Inputs
15	Associate / AD5 / DIO5	Either	Associated Indicator, Analog Input 5 or Digital I/O 5
16	RTS / AD6 / DIO6	Either	Request-to-Send Flow Control, Analog Input 6 or Digital I/O 6
17	AD3 / DIO3	Either	Analog Input 3 or Digital I/O 3
18	AD2 / DIO2	Either	Analog Input 2 or Digital I/O 2
19	AD1 / DIO1	Either	Analog Input 1 or Digital I/O 1
20	AD0 / DIO0	Either	Analog Input 0 or Digital I/O 0

* Function is not supported at the time of this release

Design Notes:

- Minimum connections: VCC, GND, DOUT & DIN
- Minimum connections for updating firmware: VCC, GND, DIN, DOUT, RTS & DTR
- Signal Direction is specified with respect to the module
- Module includes a 50k Ω pull-up resistor attached to RESET
- Several of the input pull-ups can be configured using the PR command
- Unused pins should be left disconnected

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6. Electrical Characteristics

Table 1-03. DC Characteristics (VCC = 2.8 - 3.4 VDC)

Symbol	Characteristic	Condition	Min	Typical	Max	Unit
V_{IL}	Input Low Voltage	All Digital Inputs	-	-	0.35 * VCC	V
V_{IH}	Input High Voltage	All Digital Inputs	0.7 * VCC	-	-	V
V_{OL}	Output Low Voltage	I_{OL} = 2 mA, VCC >= 2.7 V	-	-	0.5	V
V_{OH}	Output High Voltage	I_{OH} = -2 mA, VCC >= 2.7 V	VCC - 0.5	-	-	V
II_{IN}	Input Leakage Current	V_{IN} = VCC or GND, all inputs, per pin	-	0.025	1	µA
II_{OZ}	High Impedance Leakage Current	V_{IN} = VCC or GND, all I/O High-Z, per pin	-	0.025	1	µA
TX	Transmit Current	VCC = 3.3 V	-	45 (XBee) / 215 (PRO)	-	mA
RX	Receive Current	VCC = 3.3 V	-	50 (XBee) / 55 (PRO)	-	mA
PWR-DWN	Power-down Current	SM parameter = 1	-	< 10	-	µA

Table 1-04. ADC Characteristics (Operating)

Symbol	Characteristic	Condition	Min	Typical	Max	Unit
V_{REFH}	VREF - Analog-to-Digital converter reference range		2.08	-	V_{DDAD}	V
I_{REF}	VREF - Reference Supply Current	Enabled	-	200	-	µA
		Disabled or Sleep Mode	-	< 0.01	0.02	µA
V_{INDC}	Analog Input Voltage[1]		V_{SSAD} - 0.3	-	V_{DDAD} + 0.3	V

1. Maximum electrical operating range, not valid conversion range.

Table 1-05. ADC Timing/Performance Characteristics[1]

Symbol	Characteristic	Condition	Min	Typical	Max	Unit
R_{AS}	Source Impedance at Input[2]		-	-	10	kΩ
V_{AIN}	Analog Input Voltage[3]		V_{REFL}		V_{REFH}	V
RES	Ideal Resolution (1 LSB)[4]	$2.08V \le V_{DDAD} \le 3.6V$	2.031	-	3.516	mV
DNL	Differential Non-linearity[5]		-	±0.5	±1.0	LSB
INL	Integral Non-linearity[6]		-	±0.5	±1.0	LSB
E_{ZS}	Zero-scale Error[7]		-	±0.4	±1.0	LSB
F_{FS}	Full-scale Error[8]		-	±0.4	±1.0	LSB
E_{IL}	Input Leakage Error[9]		-	±0.05	±5.0	LSB
E_{TU}	Total Unadjusted Error[10]		-	±1.1	±2.5	LSB

1. All ACCURACY numbers are based on processor and system being in WAIT state (very little activity and no IO switching) and that adequate low-pass filtering is present on analog input pins (filter with 0.01 µF to 0.1 µF capacitor between analog input and VREFL). Failure to observe these guidelines may result in system or microcontroller noise causing accuracy errors which will vary based on board layout and the type and magnitude of the activity.
Data transmission and reception during data conversion may cause some degradation of these specifications, depending on the number and timing of packets. It is advisable to test the ADCs in your installation if best accuracy is required.
2. R_{AS} is the real portion of the impedance of the network driving the analog input pin. Values greater than this amount may not fully charge the input circuitry of the ATD resulting in accuracy error.
3. Analog input must be between V_{REFL} and V_{REFH} for valid conversion. Values greater than V_{REFH} will convert to $3FF.
4. The resolution is the ideal step size or 1LSB = $(V_{REFH}-V_{REFL})/1024$
5. Differential non-linearity is the difference between the current code width and the ideal code width (1LSB). The current code width is the difference in the transition voltages to and from the current code.
6. Integral non-linearity is the difference between the transition voltage to the current code and the adjusted ideal transition voltage for the current code. The adjusted ideal transition voltage is $(Current\ Code-1/2)*(1/((V_{REFH}+E_{FS})-(V_{REFL}+E_{ZS})))$.
7. Zero-scale error is the difference between the transition to the first valid code and the ideal transition to that code. The Ideal transition voltage to a given code is $(Code-1/2)*(1/(V_{REFH}-V_{REFL}))$.
8. Full-scale error is the difference between the transition to the last valid code and the ideal transition to that code. The ideal transition voltage to a given code is $(Code-1/2)*(1/(V_{REFH}-V_{REFL}))$.
9. Input leakage error is error due to input leakage across the real portion of the impedance of the network driving the analog pin. Reducing the impedance of the network reduces this error.
10. Total unadjusted error is the difference between the transition voltage to the current code and the ideal straight-line transfer function. This measure of error includes inherent quantization error (1/2LSB) and circuit error (differential, integral, zero-scale, and full-scale) error. The specified value of E_{TU} assumes zero E_{IL} (no leakage or zero real source impedance).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ – นามสกุล	นายนพดล จึงเจริญสุข
วัน เดือน ปีเกิด	3 มกราคม 2532 จังหวัดราชบุรี
ที่อยู่	441 หมู่ ถนนศรีสุริยวงศ์ ตำบลหน้าเมือง อำเภอเมือง จังหวัดราชบุรี 70000 โทรศัพท์ 086-516-9321
อีเมล	kurayha@gmail.com
ประวัติการศึกษา	
2553	วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ชื่อ – นามสกุล	นางสาวรังสิมาภรณ์ เพียรทำดี
วัน เดือน ปีเกิด	20 ตุลาคม 2531 จังหวัดกรุงเทพมหานคร
ที่อยู่	1951 ถนนสุขุมวิท73 แขวงพระโขนง เขตวัฒนา จังหวัดกรุงเทพมหานคร 10260 โทรศัพท์ 087-7028265
อีเมล	ume_dreamworld@hotmail.com
ประวัติการศึกษา	
2553	วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้