

การรู้จำเสียง

SPEECH RECOGNITION



โดย

นายชนกร สักดิ์ชัยยศ 3901214

นางสาวนันทนิชา หัตตะแสน 39014254

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2542

เลขที่ 36933
เลขทะเบียน 36933
วัน, เดือน, ปี 29 ส.ค. 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามเผยแพร่ต่อผู้อื่นโดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การรู้จำเสียง

SPEECH RECOGNITION



ปริญญาานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิศวกรรมอิเล็กทรอนิกส์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2542

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ ปีการศึกษา 2542


ภาควิชาอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เรื่อง การรู้จำเสียง

ผู้จัดทำ

1. นายธนากร ศักดิ์ชื่นยศ 39014214
2. นางสาวนันท์นันทิชา หัตตะแสน 39014254




อาจารย์ที่ปรึกษา
(รศ.ดร.มนัส ตั้งวรศิลป์)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การรู้จำข้อมูลเสียง

นายชนากร ศักดิ์ชื่นยศ

นางสาวนันท์นันทิชา หัตถะแสน

อาจารย์มนัส สังวรศิลป์ อาจารย์ที่ปรึกษา

ภาคการศึกษาที่ 2 ปีการศึกษา 2542

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้นำเสนอวิธีการตรวจสอบข้อมูลเสียงเลข 0-9 เพื่อให้รู้จำเสียงของผู้พูดคนเดียว (single user) โดยจะมีกระบวนการคร่าว ๆ ดังนี้คือเสียงพูดจะถูกบันทึกผ่านเข้ามาทางไมโครโฟน สัญญาณที่ได้คือ แอมพลิจูดในแกนเวลาจากนั้นจะผ่านส่วนปริโพรเซสซึ่ง เพื่อลดขนาดข้อมูลเสียง โดยการทำการปรับข้อมูลในแกนเวลา หลังจากนั้นนำข้อมูลเสียงผ่านการแปลงฟูเรียร์อย่างรวดเร็ว ได้ สัญญาณเสียงในโดเมนความถี่ จากนั้นทำการหาคุณลักษณะเฉพาะตัวของเสียงที่ต้องการทราบว่าจะ ต้องการทราบว่ามี ความเหมือนหรือสอดคล้องกับเสียงในกลุ่มเสียงฐานข้อมูลหลักมากเพียงใดโดยใช้สมการสหสัมพันธ์(Correlation) ซึ่งจะสามารถทำนายได้คร่าว ๆ ว่าเป็นเสียงใด หลังจากนั้นนำ ค่าคุณลักษณะเฉพาะของเสียงที่หาได้นำมาเข้าในเวรอลเน็ตเวิร์คเพื่อทำการตัดสินใจว่าเป็นเสียงใด ซึ่งจะเป็นขั้นตอนสุดท้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SPEECH RECOGNITION

Mr.Thanakorn Sukcheunyod

Miss Nannicha Hattasan

Assoc.Prof.Dr.Manas Sangworasil (Advisor)

1st Semester, Educationnal Year 1999

Abstract

This thesis present the application of neural, that use for numeral speech recognition. Speech signal will be detected from microphone, then convert to Digital signal using sound blaster card. All data sets are transformed to frequency Domain by FFT (Fast Fourier Transform). The transformed data are brought to find Correlation and be used as a standard profile for matching. Then the correlated data are brought to train on neural network in order to accumulate the value of weight. The accumulated weight is used as a reference to the testing data. Consequently, the neural network is able to test and classifies the numeral speech correctly

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

เรื่อง	หน้า
กิตติกรรมประกาศ	
บทคัดย่อ	
Abstract	
สารบัญ	
บทที่ 1 บทนำ	1
บทที่ 2 ลักษณะของเสียงพูด	3
2.1 อวัยวะที่ใช้ในการเปล่งเสียง	3
2.2 ลำดับการเกิดเสียง	3
บทที่ 3 ระบบการได้ยินและการวิเคราะห์เสียงของมนุษย์	5
3.1 หูและการได้ยิน	5
3.2 ธรรมชาติการรู้จำเสียง	6
บทที่ 4 การวิเคราะห์สัญญาณเสียง	8
4.1 การวิเคราะห์ในโดเมนเวลา	8
4.1.1 การหาขอบเขตของสัญญาณ	12
4.1.1.1 การวิเคราะห์โดยวัดระดับพลังงาน และขนาดเฉลี่ยโดยวัดในช่วงเวลาสั้น ๆ	12
4.1.1.2 การวิเคราะห์อัตราการตัดศูนย์	12
4.2 การวิเคราะห์ในโดเมนความถี่	13
4.2.1 การวิเคราะห์ฟูริเยร์เบื้องต้น	13
4.2.2 การแปลงฟาสฟูริเยร์	15
4.2.2.1 ขั้นตอนวิธีลดทอนทางเวลา	16
บทที่ 5 การสร้างระบบรู้จำเสียงพูดคำไทย	18
5.1 ระบบการรู้จำที่สร้างขึ้นใช้ในการทดลอง	18
5.2 การแก้ไขปัญหาการเปรียบเทียบเสียงที่มีความดัง, ระยะการเปล่งเสียงและความถี่ต่างกัน	19
5.2.1 ความดัง	19
5.2.2 ระยะเวลาที่เปล่งเสียงแต่ละพยางค์	20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรื่อง	หน้า
5.2.3 ความถี่	20
5.3 สหสัมพันธ์	21
5.3.1 การเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง สเปกตรัมเสียง 2 ชุด	22
บทที่ 6 โครงข่ายประสาทเทียม	24
6.1 โครงข่ายประสาททางชีวภาพ	24
6.2 โครงข่ายประสาทเทียม	25
บทที่ 7 แบกพรอเพกชัน	
7.1 ความหมายของแบกพรอเพกชัน	27
7.2 สถาปัตยกรรมของแบกพรอเพกชัน	28
7.3 ขั้นตอนวิธีการเรียนรู้	29
7.4 การหาค่าความผิดพลาด	30
7.5 การปรับปรุงค่าน้ำหนัก	31
7.6 ขั้นตอนการเรียนรู้	32
7.7 จำนวนชุดข้อมูลการสอน	36
7.8 รอบการสอน	37
7.9 การแปลงข้อมูลนำเข้า	37
7.10 อัลกอริธึมนิรอลเน็ตเวิร์กที่มีชั้นแฝงชั้นเดียว	37
บทที่ 8 การดำเนินการและการทดลอง	39
8.1 การจัดเก็บและจัดการข้อมูลเพื่อการประมวลผล	39
8.2 การดำเนินการทดลองและผลการทดลอง	40
8.2.1 ส่วนเตรียมข้อมูล	40
8.2.2 ส่วนตัดสินใจ	41
8.3 ผลและสรุปผลการทดลอง	43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูป 2.1 แสดงอวัยวะที่ใช้ในการออกเสียงพูด	4
รูป 3.1 แสดงหูส่วนนอก, ส่วนกลาง, ส่วนในและส่วนประกอบต่าง ๆ ของหู	6
รูป 4.1 แสดงโดเมนเวลาและสเปกตรัมของวินโดว์แต่ละชนิด	10
รูป 4.2 แสดงการเปลี่ยนแปลงเมื่อใช้ขนาดวินโดว์ต่างกัน	11
รูป 4.3 แสดงการคาบเกี่ยวของช่วงเวลาของวินโดว์	11
รูป 4.4 แสดงหน่วยดีเลย์ของการคำนวณตามขั้นตอนวิธีลดทอนเวลา	17
รูป 4.5 แสดงวิธีการของฟาสฟูเรียร์ทรานสฟอร์มแบบลดทอน ทางเวลา (DIT) สำหรับข้อมูลขนาด 8 จุด	17
รูป 5.1 แสดงโครงสร้างระบบการรู้จำแบบ Isolated	18
รูป 6.1 ตัวอย่างเซลล์ประสาทชีวภาพ	24
รูป 6.2 ไดอะแกรมของนิวรอน 1 หน่วย	25
รูป 6.3 ซิกมอยด์ฟังก์ชัน (Sigmoid function)	26
รูป 7.1 แสดงแบ็กพรอพเกชันแบบสามชั้นการส่งค่าความคิดพลาดกลับ	27
รูป 7.2 แบ็กพรอพเกชันที่มี 1 ชั้นแอบแฝง	28
รูป 7.3 แสดงความสัมพันธ์ระหว่างค่าน้ำหนักกับความผิดพลาด	31
รูป 7.4 แสดงการคำนวณที่โหนดผลลัพธ์	33

บทที่ 1

บทนำ

การที่เครื่องคอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ได้นั้นมาซึ่งประโยชน์ต่อมนุษย์มากมาย เช่น การสั่งงานด้วยเสียง การป้องกันความปลอดภัยโดยใช้เทคโนโลยีการจำเสียงผู้ที่เป็นเจ้าของ หรือการสั่งให้คอมพิวเตอร์พิมพ์ตามที่มนุษย์พูดแทนการพิมพ์ดีด ในปัจจุบันจะพบกับเครื่องมือที่สามารถรู้จำภาษามนุษย์ได้ซึ่งส่วนใหญ่จะใช้ได้กับภาษาอังกฤษเป็นหลัก แต่เนื่องจากเสียงพูดในภาษาไทยเป็นเสียงพูดที่มีระดับโทนเสียงสูงต่ำหรือภาษาเสียงวรรณยุกต์ (tonal language) ซึ่งแตกต่างจากภาษาอังกฤษที่มีเพียงแต่เพียงเสียงสระและเสียงพยัญชนะเท่านั้น

ในโครงงานนี้ออกแบบการรู้จำเสียงพูดภาษาไทยตัวเลข 0-9 แบบผู้ใช้เพียงคนเดียว โดยใช้วิธีการดังนี้คือ แปลงสัญญาณเสียงจากในโดเมนเวลา (Time domain) ไปสู่แกนความถี่ (Frequency domain) ด้วยฟาสฟูเรียร์ทรานสฟอร์ม (Fast Fourier Transform) เพื่อดึงเอาคุณสมบัติของเสียงพูดแต่ละเสียงออกมาส่งให้กับนิวรอลเน็ตเวิร์คต่อไป



รูปที่ 1.1 กระบวนการแยกแยะและรู้จำเสียง

สัญญาณเสียง จะถูกนำเข้าสู่เครื่องคอมพิวเตอร์โดยการพูดผ่านผ่านทางเซาน์ดการ์ด และจะถูกเก็บอยู่ในรูป FILE WAV ด้วยโปรแกรมบันทึกเสียง สัญญาณที่ได้จะถูกนำไปปรับแกนเวลา ให้มีขนาดข้อมูลอยู่ในช่วงที่กำหนด เพราะการพูดของคนแต่ละครั้ง แม้จะเป็นคำ ๆ เดียวกัน เวลาที่ใช้ในการออกเสียงอาจต่างกันไป ซึ่งเป็นปัญหาสำคัญอย่างหนึ่งของข้อมูล ขั้นตอนนี้เสมือนกำหนดเงื่อนไขบนแกนเวลาของเสียงแบบพยางค์เดียวทุกเสียงให้เสมอภาคกัน

โครงงานนี้ ถือกววิเคราะห์คุณสมบัติของเสียงในรูปของความถี่ ฉะนั้นสัญญาณที่ได้มา จะต้องแปลงให้อยู่ในรูปของความถี่ โดยวิเคราะห์ความถี่ของเสียงในช่วงสั้นๆ สัญญาณหลังจากที่ได้ปรับแกนเวลาแล้วจึงถูกแบ่งออกเป็นช่วงสั้นๆ แล้วนำไปแปลงให้อยู่ในรูปของแกนความถี่ด้วยการแปลงฟาสฟูเรียร์ แล้วนำสัญญาณในแกนความถี่นี้ไปวิเคราะห์คุณสมบัติต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณที่อยู่ในรูปความถี่จะถูกนำไปแปลงให้อยู่ในรูปค่ามาตรฐาน แล้วนำไปหาค่าความสัมพันธ์กับเสียงอ้างอิงที่เพื่อวิเคราะห์หาความสัมพันธ์ ด้วยการหาค่าสหสัมพันธ์เพื่อหาลักษณะที่แตกต่างกันของแต่ละเสียง ซึ่งการหาค่าสหสัมพันธ์นี้ทำให้ นอกจากจะสามารถวิเคราะห์ความสัมพันธ์ของเสียงต่างๆได้ในระดับหนึ่งแล้ว ยังเป็นการลดค่าพารามิเตอร์ที่จะนำไปเข้าเป็นอินพุทของโครงข่ายประสาทเทียม แทนที่จะนำสัญญาณเสียงทั้งเสียงเข้าเป็นอินพุท ซึ่งจะทำให้โครงข่ายประสาทเทียมมีความซับซ้อนและยุ่งยากในการตัดสินใจมากขึ้นด้วย

โครงข่ายประสาทเทียม (Neural Networks) เป็นกระบวนการในการตัดสินใจหรือหาข้อสรุปว่า เอาท์พุทของอินพุทที่ได้รับนั้นควรจะเป็นอย่างไร มีอยู่ด้วยกันหลายวิธีการ ในโครงงานนี้ จะเลือกใช้โมเดลแบบแบคพรอพาเกชัน ซึ่งเป็นการเรียนรู้แบบมีครู ซึ่งเหมาะกับการแก้ปัญหาในโครงงานนี้เป็นอย่างดี



บทที่ 2

ลักษณะของเสียงพูด

โดยทั่วไปแล้วคนเราเปล่งเสียงพูดออกมาตามดำเนินในระบบภาษาของตน แต่เสียงของคนที่เปล่งออกมาแต่ละครั้งนั้น แม้จะเป็นคำคำเดียวกัน ของคนคนเดียวกัน ที่อยู่ในสภาวะเดียวกัน จะมีลักษณะทางสัญญาณที่แตกต่างกันออกไป อย่างไรก็ตาม คนเราก็ยังสามารถรู้แยกแยะความแตกต่างของเสียงออกมาได้อย่างดี ว่าเป็นเสียงอะไร เสียงใคร พูดหรือมีความหมายว่าอย่างไร ซึ่งนับเป็นความอัจฉริยะอย่างหนึ่งสำหรับมนุษย์

แต่ถ้าการที่จะให้อุปกรณ์ทางอิเล็กทรอนิกส์ หรือคอมพิวเตอร์สามารถที่จะจดจำและแยกแยะเสียงพูดได้นั้น ไม่ใช่เรื่องง่าย เพราะอุปกรณ์ทางอิเล็กทรอนิกส์ ไม่มีความสามารถในการตัดสินใจด้วยตัวเอง ฉะนั้นในขั้นตอนแรกในการทำให้คอมพิวเตอร์จำเสียงพูดได้นั้น จึงต้องมีการศึกษา ลักษณะของเสียงพูดเสียก่อน

2.1 อวัยวะที่ใช้ในการเปล่งเสียง(Articulation)

อวัยวะที่ใช้ในการเปล่งเสียงแบ่งเป็น 3 พวกใหญ่ ๆ คือ

1. อวัยวะที่ใช้ในการสร้างลม คือส่วนที่ทำให้เกิดการเคลื่อนไหวของลม
2. อวัยวะที่เคลื่อนที่ได้(Active Articulator)
3. อวัยวะส่วนที่เคลื่อนที่ไม่ได้ (Passive Articulator) หมายถึงอวัยวะที่ติดกับกระดูกกลางส่วนบน ได้แก่ ริมฝีปากบน ฟันบน ปุ่มเหงือก เพดานแข็ง เพดานอ่อน ลิ้นไก่ และอวัยวะที่เป็นช่องว่างได้แก่ ช่องคอ ปาก ช่องจมูก

2.2 ลำดับการเกิดเสียง

ลำดับการเกิดเสียงนั้นเราแบ่งออกได้เป็น 3 ขั้นตอน ใหญ่ ๆ คือ

1. จุดเริ่มต้น(Initiation) อวัยวะที่ใช้ในขั้นตอนนี้คือ ปอด ที่ขั้นตอนนี้ลมจะถูกขับออกจากปอด
2. การตัดแปลงลมที่สั้นเสียง(Phonation) อวัยวะที่ใช้ในขั้นตอนนี้คือ อวัยวะที่อยู่ต่อจากปอดขึ้นมาจนถึงกล่องเสียง เป็นขั้นตอนที่ลมจากปอดจะผ่านมาเข้าหลอดลมและกล่องเสียงซึ่ง ณ ที่กล่องเสียงนี้สั้นเสียงจะทำหน้าที่เป็นลิ้นเปิด ปิด ทำให้เกิดเสียงได้ 2 ชนิด คือ ถ้าสั้นเสียงเปิดตลอดเวลาที่ลมผ่าน ลมจะผ่านออกมาได้สะดวก ซึ่งจะทำให้เกิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

ระบบการได้ยินและการวิเคราะห์เสียงของมนุษย์

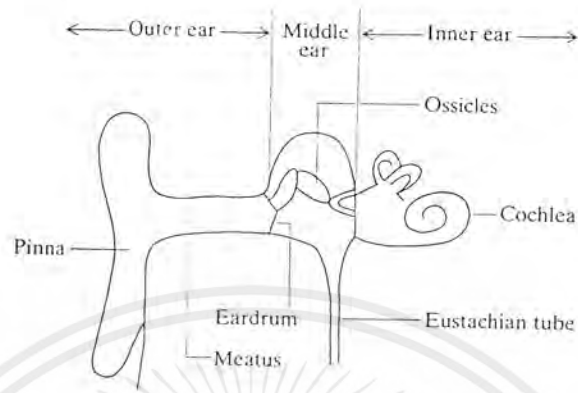
เราอาจแบ่งการรับรู้เสียงของมนุษย์ได้สองส่วน คือ ระบบการได้ยิน และ ระบบการวิเคราะห์ความหมายของเสียง ระบบการได้ยินประกอบด้วยอวัยวะต่าง ๆ ที่ใช้ในการรับเสียงทำหน้าที่รับเสียงจากสิ่งแวดล้อม แปลงเป็นสัญญาณอิมพัลส์ (Impulse) ผ่านเส้นประสาทไปยังสมองส่วนรับรู้เสียง (Auditory) ที่ทำหน้าที่เป็นระบบวิเคราะห์ความหมายของเสียง

3.1 หูและการได้ยิน

หูเป็นอวัยวะที่ใช้ในการรับเสียงทุกเสียง ประกอบด้วย 3 ส่วน ได้แก่ หูส่วนนอก หูส่วนกลาง และหูส่วนใน แรงอัดและขยายของคลื่นเสียงที่เกิดขึ้นในธรรมชาติ จะเดินทางเข้ามาทางใบหู (Pinna), ช่องหู (Auditory Canal) ซึ่งเป็นหูส่วนนอก ผ่านหูส่วนกลางที่มีโพรงอากาศประกอบด้วย Ear drum และกระดูก 3 ชิ้น ได้แก่ Hammer, Anvil และ Stirrup ทำหน้าที่ถ่ายทอดความดันสะเทือนสู่ของเหลวที่บรรจุอยู่ในอวัยวะรูปเปลือกหอยทาก (The snail-shaped cochlea) ซึ่งเป็นหูส่วนใน ของเหลวถ่ายทอดความดันสะเทือนเคลื่อนไปตามความยาวของ Cochlea (The basilar membrane) ภายในเยื่อ Basilar มีเซลล์ขน (Hair cell) ที่มีคุณลักษณะการตอบสนองความถี่แตกต่างกันนับพัน แต่อยู่ตั้งแต่ฐานถึงยอดของ Cochlea เมื่อเซลล์ขนเหล่านี้ถูกของเหลวทำให้ลักษณะเปลี่ยนไป จะเกิดเป็นสัญญาณอิมพัลส์ไฟฟ้า (Electrical impulses) กระตุ้นเป็นทอด ๆ ผ่าน เส้นประสาทเสียง (Auditory nerve) เข้าสู่สมอง คลยเซลล์ขนจะมีคุณลักษณะการตอบสนองความถี่ต่ำที่ส่วนยอดของ Cochlea และที่ส่วนฐานเซลล์ขนจะมีคุณลักษณะการตอบสนองความถี่สูง

คุณลักษณะของเสียงที่มนุษย์สามารถแยกแยะได้ประกอบด้วย ปริมาณ (Volume), พิตช์ (Pitch) และ โทน (Tone) ปริมาณความดังของเสียงขึ้นอยู่กับแอมพลิจูดหรือความเข้มของเสียง ส่วน พิตช์เป็นความสัมพันธ์ กับความถี่ของคลื่นเสียง หมายถึง จำนวนของคลื่นที่ผ่านจุดอ้างอิงต่อหนึ่งหน่วยเวลาเมื่อความถี่สูงขึ้น พิตช์ก็จะมากขึ้น, และ โทน หรือ Quality ของเสียงมีคุณสมบัติที่ซับซ้อนมากกว่า วอลุ่มและพิตช์ Quality จะแปรไปตามจำนวนและชนิดของ Overtone หรือ ฮาร์โมนิกส์ (Combination of frequencies) โดยทั่วไปมนุษย์สามารถได้ยินเสียงความถี่ตั้งแต่ 30Hz-20kHz และ ช่วงแถบความถี่ของเสียงพูดที่ใช้ในระบบสื่อสาร ที่สื่อสารกันแล้วเข้าใจ มีแบนด์วิธอยู่ในช่วง 300Hz - 3.4KHz

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.1 แสดง ส่วนนอก, ส่วนกลาง, ส่วนใน และส่วนประกอบต่าง ๆ ของหู

3.2 ธรรมชาติการรู้จำเสียง

ระบบการรู้จำเสียงของมนุษย์ มีความละเอียดซับซ้อน มีการเรียนรู้และปรับปรุงความสามารถไปพร้อมกับการพัฒนาของร่างกายตั้งแต่เยาว์วัย ระดับการรู้จำจะดำเนินไปตามขั้นตอนของมันจนกระทั่งมีความเชี่ยวชาญในการฟัง ประกอบกับมีโครงสร้างที่มีประสิทธิภาพสอดคล้องกับระบบการได้ยินทำให้สามารถแยกแยะคำและเสียงที่ซับซ้อนได้อย่างดี ความชำนาญในการฟังเกิดจากการเรียนรู้ฝึกฝนและมีการปรับปรุงอยู่ตลอดเวลา จากการศึกษาพฤติกรรมพบว่า การวิเคราะห์และรู้จำเสียงของมนุษย์มีคุณลักษณะและความสามารถหลายอย่าง พอสรุปได้ดังนี้คือ

1. การตีความหมายของคำที่ได้ยินขึ้นกับสถานการณ์, เวลา, สถานที่, ความรู้สึก, เรื่องที่กำลังสนใจ
2. สามารถรับรู้ความหมายเดียวกัน ได้แม้ความดัง, ความถี่, พิก, ความยาวของเสียงที่เปล่งแตกต่างกัน
3. ในขณะที่สนใจจะวิเคราะห์เสียงใดเสียงหนึ่ง ความสามารถในการวิเคราะห์เสียงอื่นรอบข้างจะค่อยลดลง
4. ความสามารถเลือกฟังเฉพาะเสียงที่ต้องการฟังเพื่อตีความหมายได้
5. สามารถวิเคราะห์ได้ทั้งเสียงภาษาของบุคคลทั่วไปและบุคคลเฉพาะ
6. สามารถรับรู้ทิศทางของแหล่งกำเนิดเสียงได้
7. สามารถตีความได้โดยไม่ขึ้นกับชนิดของแหล่งกำเนิดเสียงพูด
8. รับรู้อารมณ์, ความรู้สึก, น้ำเสียงของแหล่งกำเนิดเสียงพูดได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. การรับรู้มีความถูกต้องมากขึ้นเมื่อคู่สนทนาได้เห็นลีลา, สายตา, ท่าทางประกอบการพูด
10. สามารถแยกแยะคุณสมบัติของเสียงที่มีลักษณะใกล้เคียงกันอย่างมากได้อย่างมีประสิทธิภาพ
11. มีการเรียนรู้ตลอดเวลาและปรับตัวให้เข้ากับสิ่งแวดล้อมได้อย่างรวดเร็ว(ก่อนจะวิเคราะห์เสียงใด ๆ ได้ถูกต้อง ต้องมีประสบการณ์และการเรียนรู้ก่อน)
12. ปรับระดับความดังของการรับฟังให้อยู่ในระดับที่สมองสามารถนำไปวิเคราะห์ได้อย่างอัตโนมัติ โดยไม่คำนึงแหล่งกำเนิดเสียงว่าจะดังหรือค่อยเพียงใด ในขอบเขตหนึ่ง
13. มีการทำงานสอดคล้องประสานกันจิตใจที่สามารถควบคุมการวิเคราะห์จากการได้ยินได้

ดังนั้นการสร้างเครื่องจักรให้มีการรู้จำเสียงได้อย่างมีประสิทธิภาพเท่าเทียมมนุษย์เป็นเรื่องที่ยากมาก หรือหากจะทำเครื่องจักร จะต้องทำให้มีลักษณะการเลียนแบบพฤติกรรมของมนุษย์คือ “มีการเรียนรู้ ฟังฝน มีการปรับปรุงอยู่ตลอดเวลาและต้องกระทำอย่างอัตโนมัติ” การวิจัยดังกล่าวยังคงต้องอาศัยเวลาอีกพอสมควร เพราะต้องใช้คอมพิวเตอร์ที่มีความเร็วการคำนวณสูงมาก ในขั้นตอนการฝึกฝน ซึ่งคงต้องใช้เวลานานหากต้องมีการฝึกฝนบ่อย ๆ และเมื่อต้องการให้รู้จักคำมากขึ้น ประสิทธิภาพและความเร็วก็จะลดลง นอกจากนี้ระบบก็ยังต้องการอัลกอริทึมหลาย ๆ แบบร่วมกันที่ไปทำให้ความเร็วและความถูกต้องสูง หากพิจารณาคุณสมบัติการวิเคราะห์เสียงของมนุษย์พบว่ามี ความละเอียดอ่อน แม่นยำสูง ลึกซึ่งทำงานร่วมกับจิตใจซึ่งเป็นการยากที่จะสร้างระบบการวิเคราะห์ ให้มีความสามารถสูงตั้งเช่นระบบการวิเคราะห์เสียงมนุษย์ โครงการชิ้นนี้จึงมุ่งเน้นการรู้จำเสียงในจำนวนจำกัด, กลุ่มทดสอบใช้งานและกลุ่มตัวอย่างที่ค่อนข้างแคบ โดยมุ่งเน้นความถูกต้องเป็นหลัก

บทที่ 4

การวิเคราะห์สัญญาณเสียง

ในการวิเคราะห์สัญญาณเสียงซึ่งเป็นสัญญาณอนาลอกมาทำการแปลงให้อยู่ในรูปของสัญญาณดิจิทัลก่อน เมื่อทำการแปลงสัญญาณเรียบร้อยแล้วเราจะนำข้อมูลเหล่านี้ไปทำการวิเคราะห์โดยที่เราสามารถแบ่งการวิเคราะห์ได้เป็น 2 ลักษณะใหญ่ ๆ คือ การวิเคราะห์ในโดเมนเวลาและการวิเคราะห์ในโดเมนความถี่

4.1 การวิเคราะห์ในโดเมนเวลา(Time Domain Analysis)

เนื่องจากเสียงพูดมีการเปลี่ยนแปลงคุณสมบัติตามเวลา แต่ถ้าเราแบ่งเสียงพูดออกเป็นช่วง ๆ (Frame) สั้น ๆ เสียงจะมีการเปลี่ยนแปลงคุณสมบัติ น้อยมาก ทำให้การวิเคราะห์ทำได้ง่ายขึ้น การแบ่งสัญญาณเสียงพูดที่จะนำมาวิเคราะห์ออกเป็นช่วงสั้น ๆ สามารถเขียนเป็นสมการได้คือ

$$Q_n = \sum_{m=-w}^{w} T[x(m)] * w(n-m) \dots\dots\dots 4.1$$

จากสมการ 4.1 พบว่าเป็นการคอนโวลูชันระหว่าง Window ฟังก์ชัน $w(n-m)$ กับสัญญาณเสียง $T[x(m)]$ ดังนั้น Q_n ที่ได้ถือว่าเป็นค่าที่พู่ทของระบบการเปลี่ยนแปลงเวลาเชิงเส้นที่มีผลการตอบสนองเป็น $h(n) = w(n-m)$

การเลือกผลตอบสนอง $h(n)$ หรือวินโดว์ สามารถกำหนดเงื่อนไขการเลือกได้ 3 ขั้นตอนคือ

1. วินโดว์จะต้องสั้นเพียงพอที่คุณสมบัติของเสียงพูดที่เราสนใจจะต้องไม่เปลี่ยนแปลงไปมาก
2. วินโดว์จะต้องยาวเพียงพอที่จะได้รับพารามิเตอร์ของเสียงในช่วงเวลานั้นทั้งหมด
3. วินโดว์ไม่ควรสั้นเกินไปจนข้ามส่วนของเสียงพูดที่เราต้องการนำมาวิเคราะห์

วินโดว์ที่ใช้มีหลายชนิด ดังนั้นจึงควรเลือกวินโดว์ที่เหมาะสมในการนำมาใช้ในการวิเคราะห์เสียง ซึ่งมีหลักในการเลือกโดยพิจารณาลักษณะสเปกตรัม คือ

1. ความถี่เรโซลูชันสูง(high – frequency resolution) คือ มีโพลหลักแคบและแหลม (แบนด์วิทซ์แคบ)
2. การลดทอน(Attenuation) นอกช่วงความถี่ที่ผ่านได้ต่ำคือ ไซด์โพลต้องมีค่าน้อย

จากหลักการ 2 ข้างบนพบว่าจะมีลักษณะขัดแย้งกัน คือ ถ้าเมนโพลบแคบ ไซด์โพลบจะมีค่ามากหรือถ้าเมนโพลบมีค่ามาก ไซด์โพลบมีค่าน้อย ดังนั้นจึงต้องพิจารณาให้ทั้งสองข้อสอดคล้องกันคือไม่มากหรือน้อยไป ซึ่งการพิจารณาจะขึ้นอยู่กับช่วงเวลาของวินโดว์ที่ใช้โดยเราสามารถแบ่งฟังก์ชันวินโดว์ได้ดังนี้

1. วินโดว์รูปสี่เหลี่ยม(rectangular Windows)

$$w(n) = 1 \quad , 0 \leq n \leq N-1$$

2. บาร์ทเลทวินโดว์(Bartlett windows)

$$w(n) = 2n/(N-1) \quad , 0 \leq n \leq (N-1)/2$$

$$= (2-2n)/(N-1) \quad ; (N-1)/2 \leq n \leq N-1$$

3. แฮนนิ่งวินโดว์(Hanning windows)

$$w(n) = [1 - \cos(2\pi*n/(N-1))]/2 \quad ; 0 \leq n \leq N-1$$

4. แฮมมิงวินโดว์(Hamming windows)

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

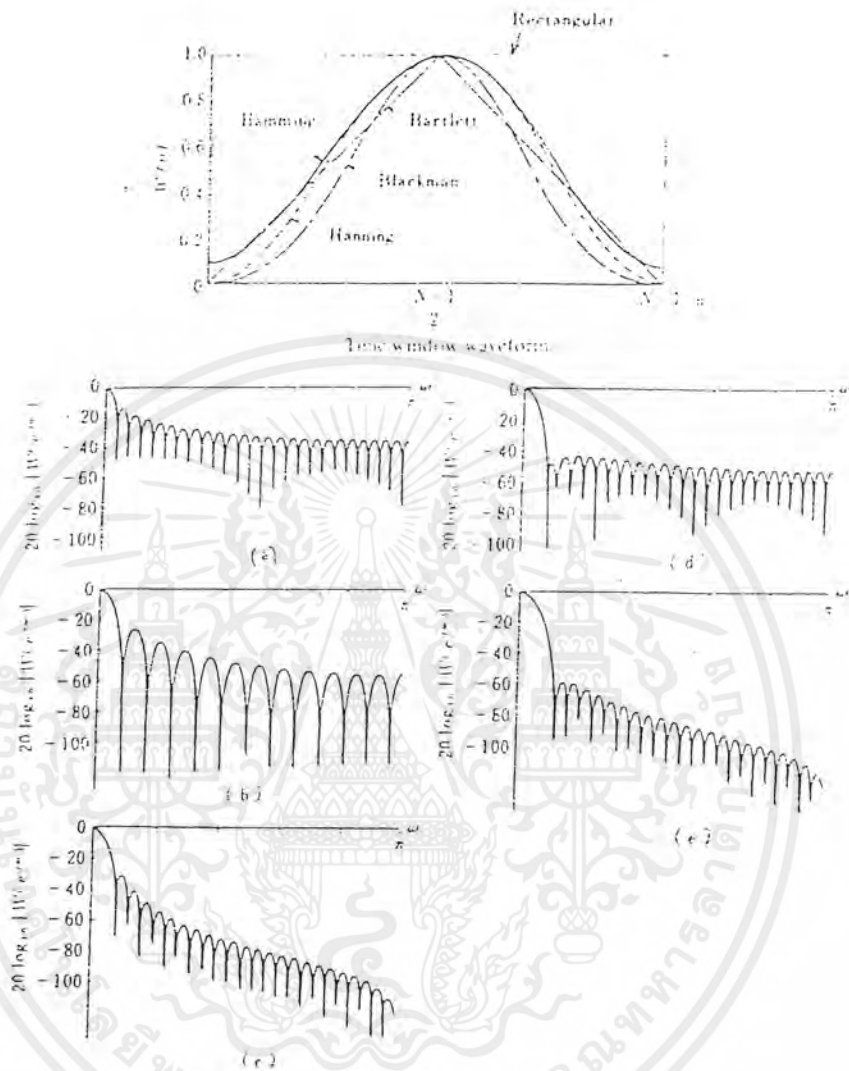
5. แบล็คแมนวินโดว์(Blackman windows)

$$w(n) = 0.42 - 0.5\cos(2\pi*n/(N-1)) \quad , 0 \leq n \leq N-1$$

จากรูปที่ 4.1 สามารถเปรียบเทียบสเปคตรัมของวินโดว์แต่ละชนิดได้และสามารถสรุปคุณลักษณะของวินโดว์แต่ละชนิดได้ดังนี้

1. วินโดว์รูปสี่เหลี่ยม พบว่ามีเมนโพลบกว้างที่สุด แต่ไซด์โพลบก็มีค่าสูงที่สุดเช่นเดียวกัน(ประมาณ -20 dB)
2. บาร์ทเลทวินโดว์ พบว่ามีเมนโพลบแคบแต่ไซด์โพลบมีค่าสูง
3. แฮนนิ่งวินโดว์ ไซด์โพลบแรกและไซด์โพลบที่สองมีค่าสูงแต่ที่ไซด์โพลบอื่นๆ มีค่าต่ำ
4. แฮมมิงวินโดว์ พบว่าไซด์โพลบมีค่าต่ำ(-40 dB) และเมนโพลบแคบพอใช้
5. แบล็คแมนวินโดว์ พบว่าไซด์โพลบมีขนาดเล็กมาก(<-60 dB) แต่เมนโพลบกว้างมาก

จากที่พิจารณาลักษณะวินโดว์แต่ละชนิดพบว่า วินโดว์ที่ควรจะถูกใช้ในการวิเคราะห์เสียงคือ แฮมมิงวินโดว์

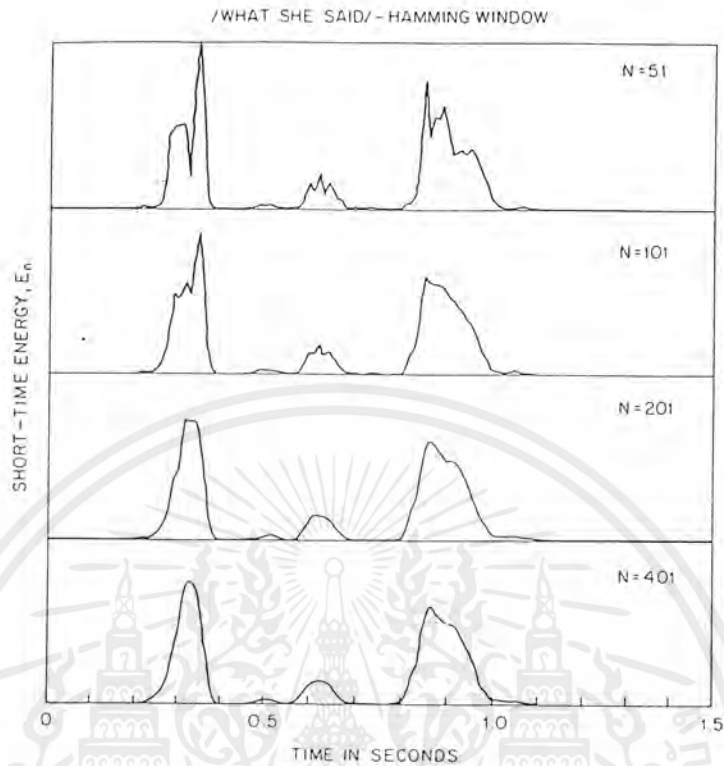


รูปที่ 4.1 แสดงโดเมนเวลาและสเปกตรัมของวินโดว์แต่ละชนิด

ในการวิเคราะห์เสียงโดยใช้วินโดว์ฟังก์ชัน จะพบว่าสัญญาณเสียงที่ผ่านการกรองโดยวินโดว์นั้นจะมีการแกว่งขึ้นลงมากขึ้นอยู่กับช่วงเวลาของวินโดว์ (ความกว้างของวินโดว์) คือถ้าช่วงเวลาของวินโดว์สั้น เอ้าท์พุทของสัญญาณจะมีการแกว่งขึ้นลงอย่างรวดเร็ว และถ้าช่วงเวลาของวินโดว์มาก เอ้าท์พุทของสัญญาณจะมีการแกว่งขึ้นลงช้า ๆ ดังแสดงในรูป 4.2

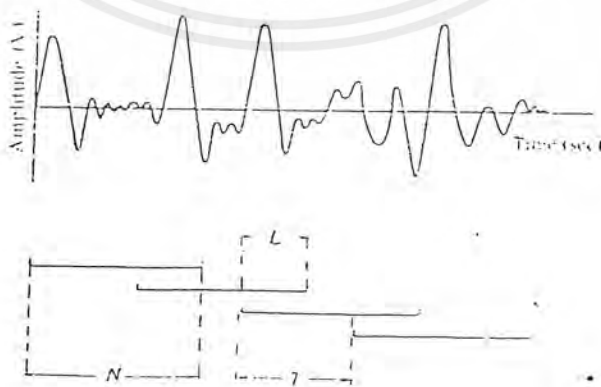
ดังนั้นการเลือกช่วงเวลาของวินโดว์จึงต้องทำการเลือกให้อยู่ในช่วงที่เหมาะสม คือไม่ให้เอ้าท์พุทของสัญญาณแกว่งช้าหรือเร็วเกินไป ซึ่งจากการทดลองพบว่าอยู่ในช่วงระหว่าง 10 -30 มิลลิวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.2 แสดงการเปลี่ยนแปลงเมื่อใช้ขนาดวินโดว์ต่างกัน

เนื่องจากช่วงเวลาที่เสียงพูดมีระยะมากกว่าช่วงเวลาที่วินโดว์ ดังนั้นการวิเคราะห์สัญญาณเสียงพูดทั้งหมด จะอาศัยหลักการคาบเกี่ยวของช่วงเวลาที่วินโดว์ แสดงดังรูปที่ 4.3 คือจุดเริ่มต้นของเฟรมปัจจุบันกับจุดเริ่มต้นของเฟรมที่เกิดขึ้นก่อน 1 เฟรม ต้องอยู่ห่างกัน 10- 20 ms ซึ่งเรียกช่วงเวลานี้ว่าช่วงเวลาในการคาบเกี่ยว(interval Shift)



รูปที่ 4.3 แสดงการคาบเกี่ยวของช่วงเวลาที่วินโดว์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1 การหาขอบเขตของสัญญาณ

เนื่องจากสัญญาณที่รับเข้ามาจากไมโครโฟน จะประกอบด้วยส่วนของสัญญาณเสียงพูด ส่วนของเสียงเงียบและสัญญาณรบกวนร่วมด้วย จึงต้องใช้กระบวนการแบ่งแยกส่วนของสัญญาณที่เป็นเสียงพูด ออกจากส่วนที่เป็นสัญญาณรบกวน เสียงเงียบ

4.1.1.1 การวิเคราะห์โดยวัดระดับพลังงานและขนาดเฉลี่ยโดยวัดในช่วงเวลาสั้น ๆ

(Short Time Energy and Average Magnitude)

การพิจารณาขนาดของเสียงในโดเมนของเวลาราวพบว่า ขนาดของสัญญาณเสียงไม่ก้องจะมีค่าต่ำกว่าขนาดของสัญญาณเสียงก้อง เราสามารถใช้ระดับพลังงานของเสียงพูดในช่วงเวลาสั้น ๆ ในการพิจารณาขนาดของสัญญาณเหล่านี้ได้ โดยจะหาค่าพลังงานจากส่วนที่ยังไม่มีเสียงพูดเข้ามาเพื่อเป็นค่าพลังงานของ Background noise (ประมาณ 1024 จุดแรก) เราสามารถกำหนดสมการพลังงานของเสียงพูดในช่วงเวลาสั้น ๆ ได้ดังนี้

พลังงานของเสียงใน 1 เฟรม ซึ่งมีจำนวน N ตัวอย่าง จะมีค่าเท่ากับ

$$E(n) = \sum_{m=0}^{N-1} |w(m)x(n-m)| \quad \dots\dots\dots 4.2$$

โดยที่ $w(m)$ คือ window function มีค่า $= 1/N$

นำค่าพลังงานที่ได้ ไปเปรียบเทียบกับส่วนที่ต่อท้ายส่วนของ 1024 จุดแรก เปรียบเทียบจุดต่อจุด จนกว่าค่าพลังงานเปรียบเทียบจะมีมากกว่า 3 เท่า ของ $E(n)$ (มาจากการตั้งเกณฑ์) ในการหาจุดเริ่มต้นของสัญญาณ ส่วนการหาจุดสิ้นสุดของสัญญาณ ก็จะเป็นทำนองเดียวกัน แต่เปรียบเทียบไล่จากทงขวาไปซ้าย จนกว่าจะพบจุดที่มีค่าพลังงานมากกว่า 3 เท่าของ $E(n)$ จุดนั้นจะเป็นจุดสิ้นสุดของสัญญาณ

4.1.1.2 การวิเคราะห์อัตราการตัดศูนย์

เนื่องจากอาจมีคำพูดบางส่วนมีค่าพลังงานน้อย การหาขอบเขตโดยเปรียบเทียบ ค่าพลังงานอาจมีข้อผิดพลาดได้ จึงต้องใช้วิธีหาค่าจุดตัดศูนย์ (Zero Crossing) เข้าช่วยพิจารณา

การเกิดจุดตัดศูนย์ เกิดขึ้นจากการที่รูปคลื่นของสัญญาณมีการตัดกับแกนเวลา นั่นคือจะเกิดเมื่อสัญญาณ Sample จุดต่อจุดมีเครื่องหมายต่างกัน อัตราการเกิดจุดตัดศูนย์จะเป็นเครื่องมืออย่างง่าย ที่ใช้ในการวัดความถี่ในการเปลี่ยนแปลงข้อมูลของสัญญาณ

ในการหาค่าจุดตัดศูนย์สามารถทำได้โดยใช้สมการ

$$Z_n = \sum_{m=-w}^w |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m) \quad \dots\dots\dots 4.3$$

โดย $\text{sgn}[x(n)] = 1$ สำหรับ $x(n) > 0$
 $= -1$ สำหรับ $x(n) < 0$

และ $w(n) = 1/2N$ สำหรับ $0 < n < N-1$
 $= 0$ ที่ใด ๆ

กำหนดให้ Z_{cm} : ค่าเฉลี่ยของจุดตัดศูนย์ในขณะที่ไม่มีการเปลี่ยนแปลง
 Z_{cs} : ค่าเฉลี่ยของจุดตัดศูนย์เปรียบเทียบ
 N_1 : เป็นจุดเริ่มต้นของสัญญาณอย่างคร่าว ๆ
 N_2 : เป็นจุดสิ้นสุดสัญญาณอย่างคร่าว ๆ

การหาขอบเขตของสัญญาณ (Endpoint detection) จะใช้วิธีเปรียบเทียบค่าพลังงานร่วมกับวิธีการหาจุดตัดศูนย์ โดยเทียบข้อมูลที่แต่ละจุด จากซ้ายไปขวาจนกว่าค่าพลังงานจุดนั้นมีค่ามากกว่า $3E(n)$ ก็จะได้จุดเริ่มต้นของสัญญาณคร่าว ๆ (N_1) เมื่อได้ตำแหน่ง N_1 แล้ว จะหาค่าเฉลี่ยของจุดตัดศูนย์เปรียบเทียบ (Z_{cs}) จากข้อมูล 25 ค่าก่อนหน้าตำแหน่ง N_1 โดยหาค่า Z_{cs} เกือบออกไปทางซ้ายของ N_1 ทีละ 25 ค่า จนกว่า $Z_{cs} < Z_{cm}$ ก็จะได้ตำแหน่งซึ่งเป็นจุดเริ่มต้นที่แท้จริงของสัญญาณในการหาจุดของสัญญาณก็เช่นเดียวกัน เริ่มด้วยหาจุดสิ้นสุดคร่าว ๆ (N_2) โดยวิธีการเปรียบเทียบค่าพลังงาน แล้วทำการคำนวณหาค่า Z_{cs} ออกไปทางขวาของ N_2 ทีละ 25 ค่า จนกว่า $Z_{cs} < Z_{cm}$ ก็จะได้ตำแหน่งของจุดสิ้นสุดที่แท้จริง

4.2 การวิเคราะห์ในโดเมนความถี่ (Frequency domain)

4.2.1 การวิเคราะห์ฟูริเยร์เบื้องต้น (Short time Fourier analysis)

การวิเคราะห์ฟูริเยร์ช่วงสั้นเป็นเทคนิคการหาความถี่ที่ใช้กันมานานแล้ว การวิเคราะห์ฟูริเยร์ให้ตัวแทนของสัญญาณเชิงพูดเป็นฟังก์ชันความถี่ในทอมของขนาดและเฟส เนื่องจากเชิงพูดไม่ได้นิ่งตลอดเวลา ดังนั้นจึงจำเป็นต้องทำการวิเคราะห์ในช่วงสั้น ๆ โดยใช้วินโดว์

การแปลงฟูริเยร์ช่วงสั้นมีสมการว่า

$$X_n(e^{jw}) = \sum_{m=-w}^w x(m)[\exp(-jwm)w(n-m)] \quad \dots\dots\dots 4.4$$

ในการคำนวณต้องใช้ DFT แทนการแปลงฟูริเยร์แบบต่อเนื่อง โดยใช้ฟังก์ชันวินโดว์ กดข้อมูลแบบไม่ต่อเนื่องทั้งหมดให้เหลือจำนวน N ตัว (N คือช่วงเวลาหรือขนาดของวินโดว์ที่ใช้ในการแปลง DFT) ข่าวสารต่าง ๆ ในจะไม่สูญหายไปจากข้อมูลเดิมถ้าการแปลงนั้นคู่มาด้วยความถี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สูงเพียงพอ(คือช่วงระยะห่างระหว่าง N)และวินโดว์ $w(n)$ ไม่มีจุดสุ่มที่เป็นศูนย์ตลอดช่วง N ตัวแปร N เป็นตัวแปรที่จะต้องระวังมากเป็นพิเศษในการวิเคราะห์ความถี่ช่วงสั้น ถ้าค่าของ N ต่ำจะทำให้ความละเอียดในโดเมนความถี่จะหยาบมาก เพราะจะทำให้ผลที่ดีในโดเมนเวลา เพราะการเฉลี่ยถูกทำเฉพาะในช่วงสั้น ๆ เท่านั้นในทางตรงกันข้ามถ้า N มีขนาดใหญ่จะทำให้ผลของความละเอียดเวลาที่อยู่ในโดเมนเวลาแต่จะทำให้โดเมนความถี่มีความละเอียดสูงกว่า

การวิเคราะห์ฟูรีเยร์ช่วงสั้นเป็นเทคนิคที่ใช้ในการหาความสัมพันธ์ระหว่าง ฟังก์ชันโดเมนเวลากับฟังก์ชันโดเมนความถี่ โดยเริ่มจากการหาการแปลงของฟูรีเยร์สำหรับสัญญาณต่อเนื่องในรูปแบบของ

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{2\pi jft} dt \quad \dots\dots\dots 4.5$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{-2\pi jft} df \quad \dots\dots\dots 4.6$$

เมื่อ $-\alpha < f < \alpha$, $-\alpha < t < \alpha$ และ $i = -1$ โดยที่ $X(f)$ จะหมายถึงฟังก์ชันโดเมนความถี่ และ $x(t)$ หมายถึงฟังก์ชันของโดเมนเวลา ในลักษณะเดียวกัน ถ้าเราทำการแปลงรูปคลื่นในโดเมนเวลาที่มีลักษณะเป็นฟังก์ชันแบบไม่ต่อเนื่อง และมีช่วงที่จำกัด ลักษณะการแปลงแบบนี้ เรียกว่า การแปลงแบบ ดิสครีตฟูรีเยร์ (Discrete Fourier Transform) สามารถเขียนได้ในรูปของ

$$X(j) = (1/N) \sum_{k=0}^{N-1} x(k) e^{-2\pi jk/N} \quad \dots\dots\dots 4.7$$

$$x(k) = \sum_{j=0}^{N-1} X(j) e^{2\pi jk/N} \quad \dots\dots\dots 4.8$$

เมื่อ $j = 0, 1, \dots, N-1$ และ $k = 0, 1, \dots, N-1$ โดยที่ทั้ง $X(j)$ และ $X(k)$ เป็นอนุกรมเชิงซ้อนเราสามารถเขียนสมการข้างต้นใหม่ โดยการให้ $W_N = e^{2\pi j/N}$ คือ

$$X(f) = (1/N) \sum_{j=0}^{N-1} x(k) W_N^{-jk} \quad \dots\dots\dots 4.9$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$X(k) = \sum_{j=0}^{N-1} X(j)W_N^{jk} \dots\dots\dots 4.10$$

ซึ่งสามารถแสดงตัวอย่างได้ คือ ถ้าเราจะแปลงข้อมูล 256 จุด (N) ของสัญญาณที่มีอัตราการสุ่มที่ 8 KHz เราจะได้ 256 ช่วงความถี่ระหว่าง 0 ถึง 8 kHz ดังแสดงในตาราง จากตารางจะเห็นว่า ทุก ๆ 31 เฮิร์ตซ์ในโดเมนความถี่ หรือ ทก ๆ 32 มิลลิวินาทีของเสียงที่จะถูกแปลงรูป โดยอาศัยความสัมพันธ์ข้างต้น

ตัวอย่างที่	โดเมนเวลา		โดเมนความถี่	
	เวลา μ sec	ตัวอย่างที่	ตัวอย่างที่	ความถี่ Hz
0	0	0	0	0
1	125	1	1	31
2	250	2	2	62
.....
254	31,750	254	254	7938
255	31,875	255	255	7969

ตารางที่ 4.1 แสดงถึงตัวอย่างความสัมพันธ์ของโดเมนความถี่กับโดเมนเวลา

4.2.2 การแปลงฟาสฟูริเยร์ (Fast Fourier Transform)

เนื่องจากการแปลงฟูริเยร์ เป็นการคำนวณที่ค่อนข้างยุ่งยากและซับซ้อน จึงต้องใช้เวลาในการคำนวณค่อนข้างมาก จึงใช้วิธีการคำนวณแบบ ฟาสฟูริเยร์(Fast Fourier Transform หรือ FFT) ทำให้การคำนวณเร็ว โดยการแปลง FFT สามารถแบ่งได้เป็น 2 ชนิดใหญ่ ๆ คือ ชนิดลดทอนทางเวลา (Decimation In Time หรือ DIT) และชนิดลดทอนทางความถี่ (Decimation In Frequency หรือ DIF) สำหรับในส่วนนี้ จะแสดงเฉพาะชนิดลดทอนทางเวลาซึ่งเกี่ยวกับในโครงงานนี้เท่านั้น

4.2.2.1 ขั้นตอนวิธีลดทอนทางเวลา(Decimation In Time)

วิธีนี้เป็นการจัดแบ่งกลุ่มลำดับสัญญาณใน โดเมนเวลา X(n) ขนาด N จุด ออกเป็น 2 ลำดับสัญญาณขนาด N/2 จุดเท่ากัน คือลำดับคู่และลำดับคี่ โดยที่ลำดับคู่เกิดจากการเอาลำดับในตำแหน่งคู่มาเรียงกัน ที่เหลือเป็นลำดับคี่ ดังนั้นจะได้

$$x_e(m) = x(2n) \quad ; m=0,1,2,\dots,(N/2)-1$$

$$x_o(m) = x(2n+1) \quad ; m=0,1,2,\dots,(N/2)-1$$

ถ้าให้ W_N เท่ากับ $\exp(-j2\pi/N)$ จะทำให้การคำนวณ DFT ของลำดับ $x(n)$ ที่ยาว N จุด สามารถเขียนใหม่ได้เป็น

$$X(k) = \sum_{m=0}^{N/2-1} x_e(2m) W_N^{2km} + \sum_{m=0}^{N/2-1} x_o(2m+1) W_N^{(2m+1)k} \quad \dots\dots\dots 4.11$$

โดยที่

$$W_N^2 = \{\exp[j2\pi / N]^2\} = \exp[j2\pi / N/2] = W_{N/2} \quad \dots\dots\dots 4.12$$

ซึ่ง $W_{N/2}$ เป็นค่า W ของลำดับความยาว $N/2$ จุด สมการ 3.5 สามารถเขียนใหม่ได้เป็น

$$X(k) = \sum_{m=0}^{N/2-1} x_e W_{N/2}^{km} + \sum_{m=0}^{N/2-1} x_o(m) W_{N/2}^{km} \quad \dots\dots\dots 4.13$$

การนำผลการแปลง DFT ขนาด 2 จุด จำนวน $N/2$ ภาคมารวมกันเพื่อให้เกิดเป็นการคำนวณ DFT ขนาด N จุด จะต้องมีหลักเกณฑ์ที่ถูกต้องด้วย จากสมการ 3.6 ถ้าเขียนให้อยู่ในช่วง $0 \leq k \leq N/2-1$ สามารถเขียนใหม่ได้เป็น

$$\begin{aligned} X(k) &= X_e + (W_N^k) X_o(k) & 0 \leq k \leq N/2-1 \\ &= X_e(k - N/2) + (W_N^k) X_o(k - N/2) & 0 \leq k \leq N/2-1 \quad \dots\dots\dots 4.14 \end{aligned}$$

เทอม W_N เรียกว่าตัวประกอบหมุน (Twiddle Factor) ซึ่งใช้ร่วมกับ DFT ขนาด 2 จุด หรือ ขนาด $N/2$ จุด ในการนำมาประกอบเป็น DFT ขนาด N จุด ได้เหมือนเดิม จากความสัมพันธ์

$$(W_N)^{k-N/2} = -(W_N)^k$$

จะได้

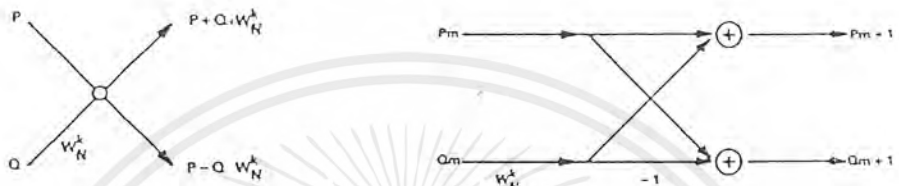
$$\begin{aligned} X(k) &= X_1(k) + (W_N)^k X_2(k) & ; 0 \leq k \leq N/2-1 \\ &= X_1(k - N/2) + (W_N)^{k-N/2} X_2(k - N/2) & ; N/2 \leq k \leq N-1 \quad \dots\dots\dots 4.15 \end{aligned}$$

ผลจากสมการนี้สามารถนำไปใช้สร้างหน่วยคำนวณที่เรียกว่า หน่วยผีเสื้อ (Butterfly Unit) โดยมีข้อมูลเข้าคือ A และ B และข้อมูลออกคือ X และ Y เป็น

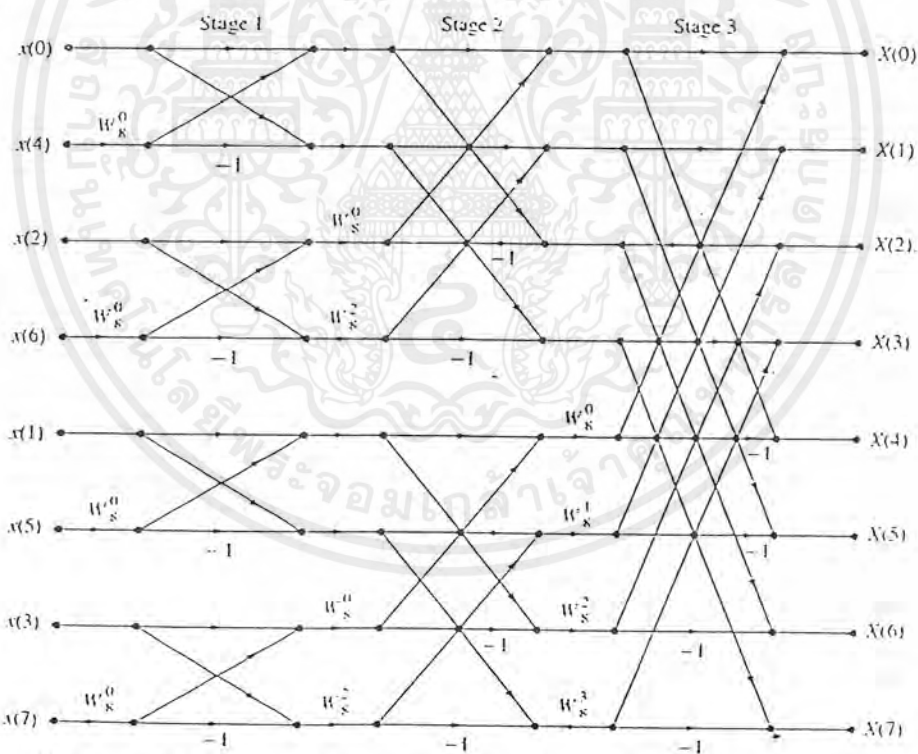
$$X = A + (W_N^k) B$$

$$Y = A - (W_N^k) B$$

ซึ่งสามารถเขียนอธิบายแทนด้วยโพลีชาร์ตดังแสดงในรูป



รูปที่ 4.4 แสดงหน่วยผีเสื้อของการคำนวณตามขั้นตอนวิธีลดทอนทางเวลา สำหรับตัวอย่างการคำนวณ DFT โดยใช้ FFT แสดงดังรูป 4.5



รูปที่ 4.5 แสดงวิธีการของ FFT แบบลดทอนทางเวลา (DIT) สำหรับข้อมูลขนาด 8 จุด

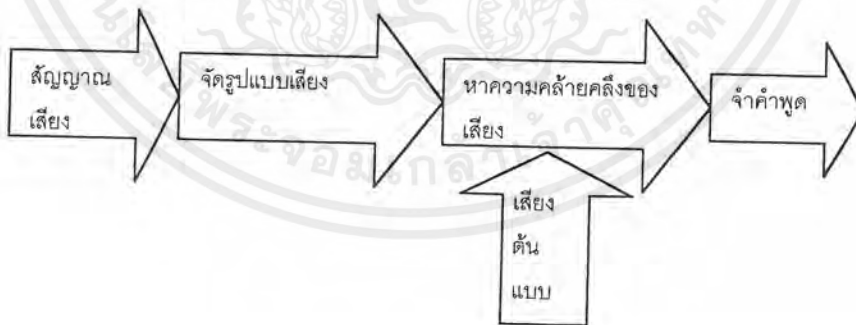
บทที่ 5

การสร้างระบบรู้จำเสียงพูดภาษาไทย

การสร้างระบบการรู้จำเสียงพูดภาษาไทยแบบไม่ขึ้นกับผู้พูด ผู้วิจัยเริ่มต้นจากการศึกษาคุณลักษณะทั่วไป คุณลักษณะเฉพาะตัวของเสียงคำพูด ระบบการได้ยิน การแปลความและวิเคราะห์ที่ความหมาย ธรรมชาติการเปล่งเสียงของมนุษย์ ตลอดจนการศึกษาหาวิธีการวิเคราะห์ที่เปรียบเทียบที่เหมาะสมเพื่อใช้ในระบบรู้จำที่จะสร้างขึ้น จากการศึกษาธรรมชาติการสื่อสารผ่านระบบเสียงและการรับฟังของมนุษย์ พอสรุปได้ว่าหากจะสร้างระบบรู้จำเสียงแบบขึ้นกับผู้พูดได้ ระบบรู้จำจะต้องสามารถวิเคราะห์เปรียบเทียบข้อมูลเสียงได้โดยไม่ขึ้นกับความดัง ความถี่ และระยะเวลาของเสียงคำหนึ่ง ๆ ที่เปล่งออกมา

5.1 ระบบการรู้จำที่สร้างขึ้นใช้ในการทดลอง

การรู้จำเสียงพูดนั้นจะเกี่ยวข้องกับการตัดสินใจระหว่าง แบบทดสอบ(Test Pattern) กับ แบบอ้างอิง(Reference pattern) หรือที่เรียกว่า Template ซึ่งจะเป็นการวิเคราะห์ถึงความสัมพันธ์ของตัวแปร (Parameter) ที่ใช้เป็นรูปแบบทั้งสอง เพื่อที่จะระบุว่าแบบทดสอบที่นำมาทดสอบนั้นมีความสัมพันธ์กับแบบอ้างอิงมากที่สุด จากนั้นจึงจะนำไปสู่การตัดสินใจต่อไป



รูปที่ 5.1 แสดงโครงสร้างระบบการรู้จำแบบ Isolated word recognition

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โครงสร้างระบบการรู้จำที่สร้างขึ้นเป็นแบบที่คล้ายกับ Isolated word recognition ดังรูปที่ 5.1 โดยในส่วนของ Feature measurement จะรวมเอา Pre-processing เข้าไปด้วยคือ A/D ,FFT เพื่อสร้างเป็นดิจิตอลสเปกตรัมเสียงที่ใช้(สเปกตรัมของเสียงใด ๆ ประกอบด้วยดิจิตอลสเปกตรัมหลาย ๆ ดิจิตอลมาวางเรียงกัน สำหรับงานวิจัยนี้บางครั้งจะเรียกดิจิตอลสเปกตรัมว่า “เฟรม”) ส่วนของ pattern similarity เป็นการเปรียบเทียบระหว่างแพตเทิร์น ที่ต้องการทราบค่ากับแพตเทิร์นเสียงหลัก ซึ่งเลือกใช้วิธีการเปรียบเทียบระดับความสัมพันธ์ของข้อมูลด้วยสมการสหสัมพันธ์ วิธีการนี้ทำให้ได้ผลลัพธ์เป็นคุณลักษณะเฉพาะตัวของเสียงที่ต้องการทราบเทียบกับฐานข้อมูลเสียงหลัก (แสดงดังภาพที่ 5.2) และขนาดของข้อมูลลดลงอย่างมาก ต่อไปผู้วิจัยขอเรียกวิธีนี้ว่า “คุณลักษณะจำเพาะสัมพันธ์” และใช้ส่วนนี้เป็นส่วนของการตัดสินใจ ในปีการศึกษานี้

5.2 การแก้ไขปัญหาการเปรียบเทียบเสียงที่มีความดัง, ระยะเวลาการเปล่งเสียงและความถี่ แตกต่างกัน

จากการศึกษาคุณลักษณะทางความถี่ของเสียงพูดพบว่า คำคำเดียวกัน ที่พูดโดยผู้พูดคนละคน มีลักษณะของแถบความถี่ที่แตกต่างกัน ทั้งนี้เนื่องจากลักษณะของช่องปาก, หลอดเสียง, และลักษณะการออกเสียงของแต่ละคนมีความแตกต่าง รวมทั้งความดังของเสียงและระยะเวลาที่เปล่งเสียงของแต่ละคน ก็เปล่งออกมาไม่เท่ากัน หรือแม้แต่คำคำเดียวกันที่ผู้พูดคนเดียวกันเปล่งออกมาแต่ละครั้ง ก็ยังมีความแตกต่างกันทุกครั้งไป ดังนั้น ในการวิเคราะห์ ควรเลือกวิธีการเปรียบเทียบที่ทำได้โดยไม่ขึ้นกับตัวแปรเหล่านี้ ได้แก่ ความดัง ระยะเวลาในการออกเสียง และความถี่ที่มีการเปลี่ยนแปลงอย่างอิสระต่อกัน โครงการนี้จะเสนอวิธีการแก้ปัญหาดังนี้

5.2.1 ความดังของเสียง จะอาศัยการเปรียบเทียบทางสถิติ ด้วยวิธีการเปรียบเทียบแบบสหสัมพันธ์ (Correlation) ซึ่งเป็นการเปรียบเทียบลักษณะความสัมพันธ์ของการเปลี่ยนแปลงชุดข้อมูล โดยไม่คำนึงถึงระดับการเปลี่ยนแปลงขนาดของข้อมูล ซึ่งคุณสมบัตินี้ ตรงกับความต้องการในโครงการ เนื่องจากในความเป็นจริง ธรรมชาติของการเปล่งเสียงของแต่ละคน มีระดับความดัง (ระดับพลังงาน) ไม่เท่ากัน ดังนั้น ถ้าจะเปรียบเทียบความสัมพันธ์ของเสียงสองเสียง ซึ่งมีความดังที่แตกต่างกัน จำเป็นต้องมีการปรับเสียงทั้งสองให้มีสเกลความดังอยู่ในระดับเดียวกันเสียก่อน (แบบ Max - Min Scale) แล้วจึงจะสามารถนำมาเปรียบเทียบกันได้ แต่ถ้าเปรียบเทียบความสัมพันธ์โดยการหาค่าสหสัมพันธ์ระหว่างข้อมูลแล้ว จะสามารถข้ามขั้นตอนนี้ไปได้ เพราะในการหาคุณลักษณะเฉพาะของสหสัมพันธ์ เป็นการหาระดับความสัมพันธ์ ของลักษณะการเปลี่ยนแปลงข้อมูล โดยไม่ขึ้นกับขนาด (Maximum) ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.2 ระยะเวลาที่เปล่งเสียงแต่ละพยางค์

เนื่องจากการเปล่งเสียงออกมาแต่ละพยางค์มีความสั้นยาวแตกต่างกันแม้เป็นคำๆ เดียวกันก็ตาม การแก้ปัญหานี้ในส่วนนี้จะทำตั้งแต่ในส่วน Normalize time เนื่องจากสัญญาณข้อมูลทั้งหมดที่ได้มา จะต้องถูกแบ่งออกเป็นเฟรมสั้นๆ เพื่อที่จะทำการวิเคราะห์ในช่วงสั้น โดยให้สัญญาณที่แบ่งได้แต่ละเฟรมมีขนาดเท่ากัน และผลลัพธ์ที่ได้จะต้องมีจำนวนเฟรมเท่ากันเพื่อสะดวกในการหาค่า Correlation และการนำไปเป็นอินพุทของโครงข่ายประสาทเทียมที่ต้องการจำนวนอินพุทคงที่ ซึ่งในการแบ่งเฟรมซึ่งจำเป็นจะต้องมีการซ้ำสัญญาณ (Overlap) เพื่อคงความต่อเนื่องของข้อมูล จึงอาศัยการ Overlap นี้ในการจำกัดให้สัญญาณที่ผ่านเข้ามา มีขนาดเท่ากัน โดยสัญญาณที่มีความยาวน้อย จะให้มีการ Overlap มาก เพื่อ ส่วนสัญญาณที่มีความยาวมากก็จะ Overlap น้อย แต่ต้องไม่น้อยเกินไปเพื่อคงความต่อเนื่องของข้อมูลไว้

ในโครงงานนี้ออกแบบให้แบ่งสัญญาณเสียงออกเป็น 10 เฟรม โดยแต่ละเฟรมมีขนาดข้อมูล 120 มิลลิวินาที เพื่อให้มีความแตกต่างกันของคุณลักษณะในแต่ละเฟรมแต่ละเสียงมากขึ้นและเพื่อให้สะดวกในการทำ FFT จึงเลือกให้แต่ละเฟรมมีขนาด 128 มิลลิวินาที จะได้ข้อมูลแต่ละเฟรมเท่ากับ 1kbyte (อัตราการแซมปลิง คูณเวลา : $8000 * 128e-3 = 1024$) จะพบว่าหลังจากสัญญาณถูกแบ่งเฟรมแล้ว จะมีข้อมูลทั้งหมด 10240 ค่าซึ่งมีความยาวยาวกว่าเสียงจริง นั่นเพื่อให้เสียงทุกเสียงต้องมีการ Overlap ด้วย ทั้งนี้หากต้องการให้มีการ Overlap ที่น้อยลงก็สามารถทำได้โดยการลดจำนวนเฟรมเอาที่พูด หรือลดขนาดข้อมูลของแต่ละเฟรมลงได้

5.2.3 ความถี่ เสียงคำๆ เดียวกันที่พูดโดยแต่ละคนมีความถี่มูลฐานไม่เท่ากัน การปรับ

ความถี่เสียงใดๆ ให้เลื่อนไปอยู่ในระดับเวลาเดียวกัน แล้วดำเนินการเปรียบเทียบ เป็นการบิดเบือนข้อมูลให้คิดเพิ่มขึ้นไปจากความเป็นจริง และมีปัญหาอยู่ที่เสียงแต่ละเสียงมีความถี่ที่เปลี่ยนแปลงตลอดเวลา เป็นการยากที่จะหาความถี่มูลฐานที่แน่นอนของแต่ละเสียงได้ จึงหลีกเลี่ยงไปใช้วิธีการนำแพตเทิร์นทดสอบไปเปรียบเทียบหาระดับความสัมพันธ์กับแพตเทิร์นอ้างอิงที่เตรียมไว้จำนวนมาก ซึ่งมีโอกาสพบกับแพตเทิร์นอ้างอิงที่ใกล้เคียงกัน แต่ต้องมีการแบ่งกลุ่มชาย-หญิง เนื่องจากความถี่มูลฐานของชาย-หญิงแตกต่างกันค่อนข้างเด่นชัด เพื่อความสะดวกในการทดลอง จะเลือกทำเฉพาะเสียงผู้ชายก่อน ส่วนการดำเนินการกับข้อมูลกลุ่มเสียงอื่นจะแยกกันเป็นโมดูล แต่ยังคงใช้วิธีการเดียวกันได้

5.3 สหสัมพันธ์(Correlation)

ค่าระดับความสัมพันธ์ของข้อมูล 2 ชุด เรียกว่า สัมประสิทธิ์ของสหสัมพันธ์ (Correlative coefficient: r) สามารถหาได้จากเทอมของ Z – Score ดังนี้

$$r_{xy} = \frac{\sum_{i=1}^N Z_{xi} Z_{yi}}{N} \dots\dots\dots 5.1$$

เมื่อ r_{xy} คือ สัมประสิทธิ์ของสหสัมพันธ์ของข้อมูล x_i และ y_i

x_i คือ ข้อมูลในแกนความถี่ของสเปกตรัมเสียงชุดที่ 1 (ทดสอบ)

y_i คือ ข้อมูลในแกนความถี่ของสเปกตรัมเสียงชุดที่ 2 (อ้างอิง)

N คือ จำนวนข้อมูลทั้งหมดที่มี x_i, y_i เป็นสมาชิก

Z_{xi} และ Z_{yi} คือค่า Z-Scores เป็นระยะห่างระหว่างค่าข้อมูลกับค่า Mean ของชุดสมการในเทอมของ Standard deviation S_{xi}

โดย
$$Z_{xi} = \frac{x_i - \bar{x}}{S_x} \dots\dots\dots 5.2$$

และ
$$Z_{yi} = \frac{y_i - \bar{y}}{S_y} \dots\dots\dots 5.3$$

เมื่อ
$$S_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \dots\dots\dots 5.4$$

และ
$$S_y = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N}} \dots\dots\dots 5.5$$

จากสมการ 3.1 ค่า r_{xy} จะมีค่าอยู่ระหว่าง -1 ถึง +1 และพอตีความหมายได้ดังนี้คือ ถ้าค่า $|r_{xy}| = 1$ หมายถึง ข้อมูลทั้ง 2 ชุด มีความสัมพันธ์กันทุกจุดข้อมูล ซึ่งในทางปฏิบัติแล้วมีโอกาสน้อยมาก

ในการทดลองจะพิจารณาเฉพาะ ค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสเปกตรัมไม่ต่ำกว่า 0.70 ซึ่งจะกล่าวในหัวข้อวิธีการหาคุณลักษณะจำเพาะสัมพันธ์ และลดขนาดข้อมูลเสียง

5.4.1 การเทียบหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสเปกตรัมเสียงสองชุด

การเทียบข้อมูลเสียงทั้งสองจะทำที่ละเฟรม(หรือ Discrete หรือ Δt) แบบเฟรมต่อเฟรม โดยเฟรมที่ 1 ของสเปกตรัมเสียงที่ 1 (ทดสอบ) ก็เปรียบเทียบกับเฟรมที่ 1 ของสเปกตรัมเสียงชุดที่ 2 (หลัก) ดังสมการ(3.6)

$$r_{xy} = \sum \frac{Z_{xj} Z_{yj}}{N} \dots\dots\dots 5.6$$

เมื่อ r_{xy} คือค่าสัมประสิทธิ์สหสัมพันธ์ที่หาได้จากหนึ่งเฟรมของสเปกตรัม x_i, y_i

Z_{xj}, Z_{yj} คือ ค่า Z-Score ได้จากสมการที่ 5.2 และ 5.3

N คือ จำนวนข้อมูลที่เป็นสมาชิกของ x_i, y_i (โดย $i = 1, 2, 3, \dots, N$)

X_i คือ ข้อมูลในแกนความถี่ของสเปกตรัมเสียงชุดที่ 1 (ทดสอบ)

Y_i คือ ข้อมูลในแกนความถี่ของสเปกตรัมเสียงชุดที่ 2 (อ้างอิง)

P คือ เฟรม(Discrete) ของสเปกตรัมที่กำลังพิจารณา

การหาค่าสัมประสิทธิ์สหสัมพันธ์ จะดำเนินการทุกเฟรมของสเปกตรัมเสียงทั้งสอง แต่เนื่องจากจำนวนเฟรมของเสียงทั้งสองไม่เท่ากันจึงกำหนดขอบเขต ของการหาค่าสัมประสิทธิ์สหสัมพันธ์โดยให้สิ้นสุดการพิจารณาที่เฟรมสุดท้ายของข้อมูลเสียงใดที่ถึงก่อน ดังสมการที่ 5.7

$$\begin{aligned} F &= I && \text{เมื่อ } J \geq I \\ F &= J && \text{กรณีอื่น} \end{aligned} \dots\dots\dots 5.7$$

เมื่อ F คือ จำนวนเฟรมของข้อมูลเสียงพยางค์ 2 ชุดที่นำมาเปรียบเทียบได้จากสมการที่ 5.7

I คือ จำนวนเฟรมของข้อมูลเสียงคำชุดที่ 1 (ข้อมูลหลัก)

J คือ จำนวนเฟรมของข้อมูลเสียงคำชุดที่สอง(ข้อมูลที่นำมาเทียบหาค่าสัมประสิทธิ์ สหสัมพันธ์)

สาเหตุที่กำหนดเช่นนี้ มาจากข้อสังเกตที่ว่า หากข้อมูลมีช่วงการเปล่งเสียงยาวนานออกไปมาก โอกาสที่จะเป็นเสียงเดียวกันก็จะน้อยลง หรือกรณีเป็นเสียงเดียวกันแต่มีช่วงเวลาการเปล่งเสียงยาวออกไปไม่มากนัก(1-2 เฟรม) ก็ไม่จำเป็นต้องนำส่วนที่เกินมาเปรียบเทียบความสัมพันธ์ตามสมการ (5.8)

$$R = r = \frac{1}{F} \sum_{p=1}^F r_{xy_p} \dots\dots\dots 5.8$$

เมื่อ $R = r$ คือค่า Mean ที่บอกถึงระดับความสัมพันธ์ของข้อมูลเสียงสองชุด

จะถูกรหัสด้วยจำนวน Discrete F หรือ จำนวน Δ ซึ่งหากเป็นเสียงเดียวกันแล้วค่าเฉลี่ยผลรวมที่ได้จะไม่ผิดพลาดมาก การหาค่าระดับของความสัมพันธ์ของข้อมูลที่ได้จะบ่งบอกถึงคุณสมบัติเฉพาะของเสียงนั้น โดยเทียบกับเสียงหลักทั้งหมด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

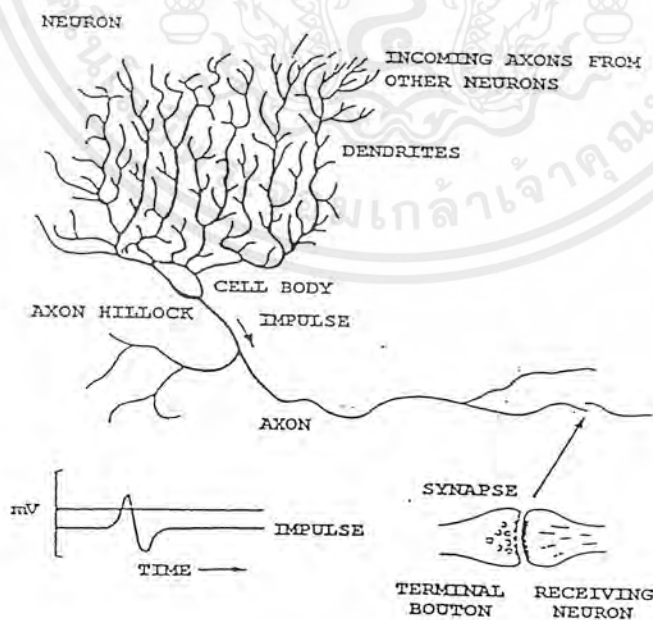
โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม หมายถึง โครงข่ายใยประสาทที่เชื่อมต่อกันระหว่างเซลล์ประสาทจำนวนมากมาขมมหาศาล บรรจุอยู่ในสมองมนุษย์ มีความสามารถในการประมวลผลสูง สมองชีวภาพเป็นศูนย์กลางควบคุมกิจกรรมการดำเนินชีวิตที่สามารถเรียนรู้และปรับตัวเองได้ตลอดเวลา

การวิจัยเพื่อพัฒนาให้คอมพิวเตอร์มีความสามารถตัดสินใจได้วิธีการหนึ่ง จึงหันมาศึกษาโครงข่ายประสาทเทียม โดยการจำลองลักษณะการทำงานของใยประสาทชีวภาพขึ้นเป็นโมเดลทางคณิตศาสตร์

6.1 โครงข่ายประสาททางชีวภาพ

ระบบการคิดของมนุษย์ มีโครงสร้างพื้นฐานจากเซลล์สมอง ที่เรียกว่านิวรอน (Neurons) เรียงเป็นชั้น ๆ จำนวนมหาศาล ประมาณหมื่นล้าน(10)นิวรอน เชื่อมโยงถึงกันประมาณพันล้านล้านจุด แต่ละนิวรอนจะมีคุณสมบัติแตกต่างกันไป แต่จะมีรูปแบบทำงานคล้าย ๆ กันคือ รับค่าเข้ามาประมวลผลและส่งสัญญาณ ไฟฟ้าเคมีออกไป



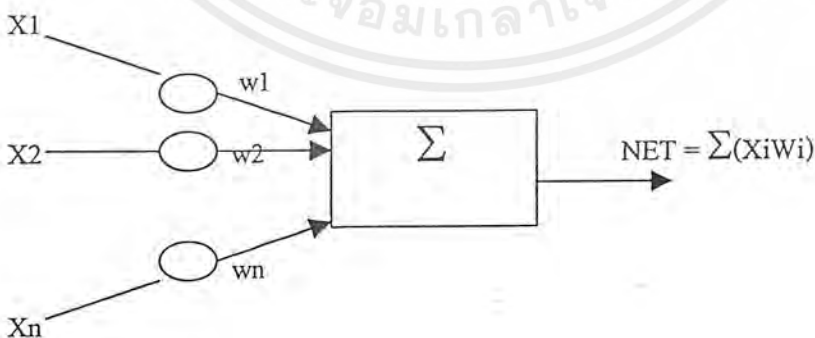
รูป 6.1 ตัวอย่างเซลล์ประสาทชีวภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากภาพแขนงที่ขยายออกไปยังเซลล์อื่น ๆ เพื่อรับสัญญาณ เรียกว่าเดนไดรด์(dendrites) จะมีจุดเชื่อมต่อกับเซลล์ประสาทอื่น ๆ เรียกว่าไซแนปส์(synapse)และส่วนที่ใช้ส่งสัญญาณออกไปเรียกว่าแอกซอน(Axon) แต่ละนิวรอนจะมีคุณสมบัติในการเพิ่มขยาย หรือลดทอนความเข้มของสัญญาณที่เข้ามาถึงเดนไดรด์ของเซลล์ อาจกระตุ้นหรือยับยั้งตัวเซลล์ก็ได้ เนื่องจากเซลล์ประสาทหนึ่งเซลล์มีเดนไดรด์จำนวนมาก ดังนั้น สัญญาณกระตุ้นเดนไดรด์ที่รับเข้ามาจากเซลล์อื่น ๆ จะถูกนำมารวมกันที่ตัวเซลล์ประสาท ที่เซลล์ประสาทจะมีค่าเทรชโฮลด์ (Threshold) ค่าหนึ่ง หากผลรวมของไฟฟ้าเคมี(electrochemical)มีค่ามากกว่าค่าเทรชโฮลด์ เซลล์ประสาทจะส่งสัญญาณค่าหนึ่งผ่านทางแอกซอนไปยังนิวรอนอื่น ๆต่อไป การจัดเรียงเป็นชั้น(Layer)และลักษณะการเชื่อมโยงระหว่างนิวรอนในสมองนั้นมีการจัดเรียงที่ซับซ้อน สอดคล้องกับหน้าที่ในการทำงานเฉพาะส่วนมีการเจริญเติบโตเรียนรู้สัมพันธ์กับสิ่งแวดล้อมตลอดเวลา จึงเป็นการยากที่จะสามารถสร้างโมเดลเลียนแบบการทำงานของสมองชีวภาพได้ทั้งหมด ปัจจุบันยังคงทำได้เพียงการจำลองเลียนแบบการทำงานของเฉพาะส่วนบางส่วนของโครงข่ายประสาทเทียม มาใช้เฉพาะกับงานใดงานหนึ่ง และมีขอบเขตจำกัดการใช้งานได้

6.2 โครงข่ายประสาทเทียม(Artificial Neural Network)

การออกแบบสร้างประสาทเทียมมีสมมูลฐานจากโครงข่ายประสาทชีวภาพ ซึ่งมีความนำสัญญาณไฟฟ้าเคมีต่างกัน โมเดลโครงข่ายประสาทเทียมจึงต้องมีการถ่วงน้ำหนักก่อนนำไปใช้งาน เรียกว่าไซแนปติกเวกท์ ปริมาตรข้อมูลจะถูกนำมารวมกัน และตัดสินใจด้วยระดับความสนใจของนิวรอน(Activation Level) แล้วส่งเอาท์พุท(Output) ออกไปยังนิวรอนอื่น ๆ



รูปที่ 6.2 โมเดลของนิวรอน 1 หน่วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากภาพเป็นเป็นไปคือแกรมที่จำลองจากแนวความคิดของเซลล์สมองทางชีวภาพ สัญญาณอินพุตคือ X_1, X_2, \dots, X_n จะถูกป้อนเข้ามาขังนืวรอก เปรียบได้กับสัญญาณไฟฟ้าเคมีเข้ามาขังไปเซเนปส์ของเซลล์ประสาท ค่าอินพุตเหล่านี้จะคูณด้วยค่าถ่วงน้ำหนัก (weight) ที่มีค่าขังมด 0.0 - 1.0 ผลรวมของสัญญาณทั้งหมดจะถูกส่งออกมาโดย

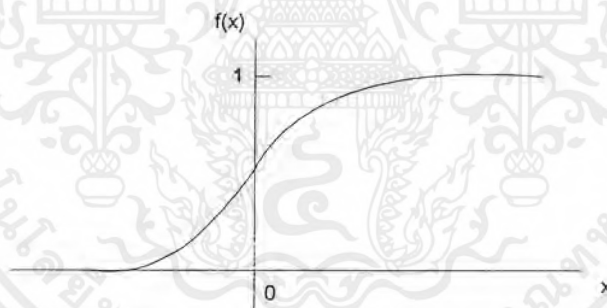
$$\begin{aligned} \text{NET} &= X_1W_1 + X_2W_2 + \dots + X_nW_n \\ &= \sum(X_iW_i) \end{aligned}$$

จากนั้นจะถูกตัดสินใจโดยหน่วยเซลล์ด้วยฟังก์ชันการตัดสินใจ (Activation Function) และได้ค่าเอาพุตออกมา

$$\text{OUT} = F(\text{NET})$$

ฟังก์ชันการตัดสินใจที่นิยมใช้กันทั่วไปคือ สควาซซิงฟังก์ชัน (Squashing function) หรือซิกมอยด์ฟังก์ชัน (Sigmoid function) ซึ่งมีรูปร่างคล้ายตัว S และมีสมการทางคณิตศาสตร์ดังนี้

$$F(x) = \frac{1}{1 + \exp(-x)}$$



รูปที่ 6.3 ซิกมอยด์ ฟังก์ชัน (Sigmoid function)

การใช้ซิกมอยด์ ฟังก์ชัน จะทำให้เทรอสโอดฟังก์ชันมีลักษณะ non linear function ทำให้ค่าอินพุตที่ไวต่อสัญญาณขนาดเล็ก ๆ และเฉลี่ยต่อสัญญาณแรง ๆ (คือสัญญาณไปทางบวกเล็กน้อยก็จะมีค่าเอาพุตทุกใกล้เคียง "1" และสัญญาณเป็นลบค่าต่ำ ๆ ใกล้เคียง "0" ขณะเดียวกันสัญญาณบวกขนาดใหญ่ ๆ ก็ยังคงให้ค่าเอาพุตทุกใกล้เคียง "1" และสัญญาณลบค่ามาก ๆ ก็ยังคงให้เอาพุตทุกใกล้เคียงศูนย์ได้ คือมีคุณสมบัติ nonlinear gain นั้นเอง ซึ่ง กลอสเบอร์ริก (Grossberg, 1973) พบว่าสามารถแก้ปัญหา Noise saturation dilemma ได้และทำให้นิวรอลสามารถทำงานได้กว้างขวางขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7

แบ็กพรอเพกชัน

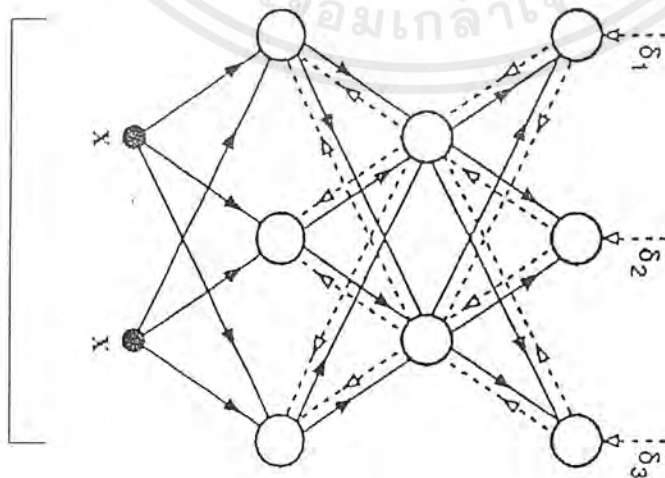
ประมาณปี 1970 เพอเซปตรอน(Perceptron) ซึ่งเป็นนิวรอลเน็ตเวิร์กประเภททำงานชั้นเดียว(Single Layer Neural Network) ถูกโจมตีว่าไม่สามารถแก้ปัญหา XOR ได้ ทำให้ความสนใจในวิทยาการในนิวรอลเน็ตเวิร์กได้ลดน้อยลงไป จนมีการเสนอวิธีการใหม่ที่สามารถแก้ไขปัญหานี้คือเปลี่ยนแปลงโครงสร้างจากทำงานชั้นเดียวเป็นทำงานหลายชั้น (Multi Layer Neural Network) ทำให้นิวรอลเน็ตเวิร์กแตกแขนงออกไปหลายรูปแบบ เพื่อใช้แก้ปัญหาในลักษณะต่างๆ

7.1 ความหมายของแบ็กพรอเพกชัน(Backpropagation)

แบ็กพรอเพกชันหรือบางครั้งเรียกว่าเพอร์เซปตรอนหลายชั้น (Multi Layer Perceptron) เนื่องจากมีความคล้ายคลึงกับเพอร์เซปตรอนมาก เรียกอย่างย่อว่าบีพี(BP) เป็นรูปแบบหนึ่งของนิวรอลเน็ตเวิร์กที่ได้รับความนิยมมาก เนื่องจากสามารถนำมาใช้แก้ปัญหาได้หลายด้าน เป็นการเรียนรู้แบบมีครู(Supervised Training) นั่นคือการมีชุดสอนที่ส่งเข้าสู่ระบบ และมีสถาปัตยกรรมแบบทำงานหลายชั้น มีลักษณะการติดต่อระหว่างนิวรอลแบบไปโนชั้นถัดไป(Feedforward) แต่ค่าความผิดพลาดจะถูกนำกลับมาปรับปรุงค่าน้ำหนักใหม่ในระหว่างการสอน(Back Propagate Error)

จากรูป 7.1 แสดง BP ในรูปทำงานสามชั้น (Three layer Network) โดยมีการส่งผ่านข้อมูลไปข้างหน้า แสดงด้วยเส้นทึบและส่งค่ากลับความผิดพลาด (δ) แสดงด้วยเส้นปะ

การประเมินผลความผิดพลาดจากการคำนวณ สำหรับแบ็กพรอเพกชันแล้วนิยมใช้หลักการหาค่าความผิดพลาดกำลังสองต่ำสุด (Least Mean Squared Error) หรือเรียกอย่างย่อว่า (LMS) เมื่อ



รูปที่ 7.1 แสดงแบ็กพรอเพกชันแบบสามชั้นที่มีการส่งค่าความผิดพลาดกลับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

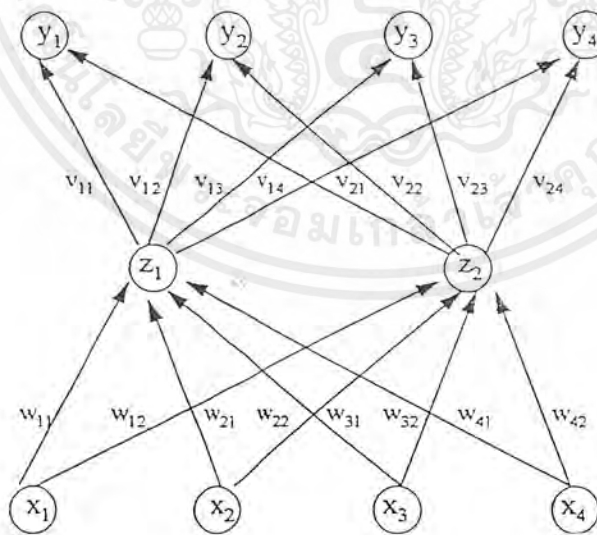
ค่าความผิดพลาดยังไม่ถึงจุดต่ำสุดก็จะทำการปรับปรุงค่าน้ำหนักใหม่แล้วผ่านการคำนวณหาผล
ลัพธ์ใหม่ในรอบการสอนถัดไป จนกระทั่งความผิดพลาดนั้นอยู่ในเกณฑ์ยอมรับได้

7.2 สถาปัตยกรรมของแบ็กพรอพेคชัน

ประกอบด้วยโครงสร้างเป็นชั้น คือ ชั้นนำเข้าข้อมูล, ชั้นแอบแฝง, และชั้นผลลัพธ์ ซึ่งชั้น
แอบแฝงอาจจะมีมากกว่า 1 ชั้นก็ได้ ภายในชั้นแต่ละชั้น จะประกอบด้วยนิวรอน(neuron) หรืออาจ
เรียกว่าโหนด (node) ในชั้นถัดไปด้วยค่าน้ำหนักชุดหนึ่ง

กรณีมี 1 ชั้นแอบแฝง จะใช้ชุดน้ำหนัก 2 ชุดคือ ชุดของค่าน้ำหนักระหว่างชั้นนำเข้าข้อมูล
และชั้นแอบแฝง และชุดของค่าน้ำหนักระหว่างชั้นแอบแฝงกับชั้นผลลัพธ์ และในกรณีมีชั้นแอบ
แฝงมากกว่า 1 ชั้นก็จะมีชุดค่าน้ำหนักระหว่างชั้นแอบแฝงกับชั้นแอบแฝงอื่นด้วยกัน ซึ่งจะกำหนด
ค่าเริ่มต้นของค่าน้ำหนักแบบสุ่ม

การสอนให้ระบบเรียนรู้ต้องมีชุดของข้อมูลนำเข้าและค่าเป้าหมายที่ต้องการ แล้วส่งค่าข้อ
มูลนำเข้าสู่กระบวนการเรียนรู้และปรับปรุงค่าน้ำหนัก ตามรูป 7.2 แสดงถึงข่ายงานแบ็กพรอพेค
ชันแบบสองชั้น และเป็นข่ายงานที่มีลักษณะการติดต่อของนิวรอนแบบติดต่อกันหมด (Fully
Connected Net) ประกอบด้วยชุดน้ำหนักสองชุด คือ V และ W ส่วนข้อมูลนำเข้าคือ X ผลลัพธ์ใน
การคำนวณจากข้อมูลนำเข้า X กับค่าน้ำหนัก V จะส่งสัญญาณไปยังชั้นถัดไปให้กับ Y ซึ่งเป็นชั้น
ผลลัพธ์



รูป 7.2 แบ็กพรอพेคชันที่มี 1 ชั้นแอบแฝง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แต่ละโหนดผลลัพธ์ Y_i ในชั้นผลลัพธ์ จะถูกเปรียบเทียบกับค่าเป้าหมายเพื่อหาความผิดพลาด ถ้าค่าความผิดพลาดยังสูงหรือยอมรับไม่ได้ ก็จะทำให้การส่งกลับค่าความผิดพลาดกลับไปยังชั้นก่อนหน้าเพื่อให้แต่ละชั้นทำการปรับปรุงค่าน้ำหนักใหม่

7.3 ขั้นตอนวิธีการเรียนรู้ (Learning Algorithm)

อาจกล่าวได้ว่า เบ็กพโรเพกชัน จะประกอบด้วย การนำชุดข้อมูลที่จะใช้สอนเข้าสู่ระบบ และทำการคำนวณหาผลลัพธ์ การวัดค่าความผิดพลาดกลับไปยังชั้นก่อนหน้าเพื่อทำการปรับปรุงค่าน้ำหนัก

ขั้นตอนการสอนของเบ็กพโรเพกชันมีลักษณะนำชุดข้อมูลที่ต้องการสอนเข้าสู่ระบบ ซึ่งประกอบด้วยชุดข้อมูลตัวอย่างที่ต้องการสอนหรือหน่วยนำเข้าสู่ข้อมูล (X_i) กับค่าเป้าหมายที่ต้องการ จากรูป 3.2 แต่ละหน่วยนำเข้าสู่ข้อมูล (X_i) จะส่งสัญญาณหรือข้อมูล ไปยังหน่วยแอมแปง Z_0 ถึง Z_p แต่ละหน่วยแอมแปงจะทำการคำนวณหาผลลัพธ์และผ่านแอกติเวชันฟังก์ชันเช่นกัน และผลลัพธ์ที่ได้จะนำมาเปรียบเทียบกับค่าเป้าหมายเพื่อหาค่าความผิดพลาด แล้วส่งค่าความผิดพลาดนั้นไปยังชั้นก่อนหน้าหรือชั้นแอมแปงเพื่อทำการปรับปรุงค่าน้ำหนักระหว่างชั้นแอมแปงและชั้นผลลัพธ์ และหาค่าความผิดพลาดที่ชั้นแอมแปงเพื่อส่งกลับไปทำการปรับปรุงค่าน้ำหนักระหว่างชั้นนำเข้าสู่ข้อมูลและชั้นแอมแปง ที่ต้องมีการคำนวณค่าความผิดพลาดที่ชั้นแอมแปงด้วย เพราะการคำนวณในชั้นแอมแปงอาจมีผลต่อความผิดพลาดในชั้นผลลัพธ์

จากขั้นตอนการสอนดังกล่าวเมื่อปรับปรุงค่าน้ำหนักเรียบร้อยแล้ว ข้อมูลนำเข้าสู่ชุดถัดไปจะถูกนำมาเข้าสู่ขั้นตอนการสอน

ซึ่งค่าหน่วยนำเข้าสู่ข้อมูล (X_i) ก็จะเปลี่ยนไปและค่าเป้าหมายก็จะเปลี่ยนไปตามชุดข้อมูลนำเข้าสู่ด้วย จนกระทั่งในที่สุดทุก ๆ ชุดข้อมูลนำเข้าสู่ถูกสอนหมด ก็จะนับเป็น 1 รอบการสอน

ปกติแล้วเบ็กพโรเพกชันจะต้องการการสอนหลาย ๆ รอบจนกว่าระบบจะเกิดการเรียนรู้ ซึ่งการเรียนรู้จะวัดได้จากค่าความผิดพลาดในชั้นผลลัพธ์ ถ้าค่าความผิดพลาดน้อยเพียงพอก็จะหยุดการสอน

ในขั้นตอนการเรียนรู้สำหรับเบ็กพโรเพกชันจะประกอบด้วย 2 หลักการสำคัญคือ การกำหนดฟังก์ชันแอกติเวชัน และการกำหนดวิธีการปรับปรุงค่าน้ำหนัก
แอกติเวชันฟังก์ชัน(Activation Function)

แอกติเวชันฟังก์ชันสำหรับเบ็กพโรเพกชันสำหรับเบ็กพโรเพกชันจะมี 3 ลักษณะสำคัญหลายประการ คือ จะต้องเป็นแบบต่อเนื่อง และมีลักษณะการหาค่าอนุพันธ์ และต้องการสูตรการคำนวณที่ง่ายมีประสิทธิภาพ และสามารถหาค่าผลลัพธ์ที่อยู่ในขอบเขตได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะของแอกติวชันฟังก์ชันสำหรับแบ็กพรอปเพกชัน จะให้ผลลัพธ์ที่สามารถสร้างกราฟออกเป็นรูปตัว S ได้ ซึ่งประกอบด้วยฟังก์ชันโลจิสติก(Logistic Function) หรือฟังก์ชันไบนารี(Binary Sigmoid) ที่ให้ผลลัพธ์อยู่ในช่วง (0,1) และ ฟังก์ชันไฮเพอริบออลิกแทนเจนต์(Hyperbolic Tangent Function)หรือฟังก์ชันไบโพลาร์(Bipolar Sigmoid) ที่ให้ผลลัพธ์อยู่ในช่วง (-1,1) หรืออาจเป็นค่าอื่นเช่น (-0.5,0.5) แต่มีลักษณะที่เป็นช่วงของค่าบวกและลบ

จากการคำนวณหาผลลัพธ์ของนิวรอนด้วยสูตร ตาม(1)คือผลบวกของผลคูณข้อมูลนำเข้ากับค่าน้ำหนัก อาจทำให้ผลลัพธ์นั้นกระจุกกระจายเนื่องจากการกำหนดค่าน้ำหนักเริ่มต้นเป็นค่าสุ่ม ดังนั้นจึงจำเป็นต้องผ่านกระบวนการวัดค่าเพื่อให้ค่าน้ำหนักมีความหมายขึ้น โดยผ่านเข้าสู่ฟังก์ชันซิกมอยด์เพื่อจำกัดขอบเขตของผลลัพธ์ ดังเช่นเมื่อใช้ฟังก์ชันโลจิสติกจะใช้สูตรตาม(2) และเมื่อใช้ฟังก์ชันฟังก์ชันไฮเพอริบออลิกแทนเจนต์จะใช้(3) เพื่อหาผลลัพธ์แล้วส่งผลลัพธ์ไปเป็นข้อมูลนำเข้าในชั้นถัดไป

$$net = \sum w_i i_i \quad \dots(7.1)$$

$$F(x) = 1 / (1 + \exp(-x)) \quad \dots(7.2)$$

$$F(x) = \tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x}) \quad \dots(7.3)$$

7.4 การหาค่าความผิดพลาด

แบ็กพรอปเพกชันดั้งเดิม จะมีการประมาณการค่าความผิดพลาดโดยใช้วิธีการคำนวณค่าเฉลี่ยของค่าความผิดพลาดกำลังสอง (Mean Squared Error) หรือ MSE ซึ่งที่จริงแล้วการใช้วิธีการคำนวณแบบนี้ไม่จำเป็นสำหรับแบ็กพรอปเพกชัน เราอาจกำหนดวิธีการหรือฟังก์ชันอื่นๆในการประมาณการค่าความผิดพลาดได้ แต่อย่างไรก็ตามวิธีประมาณการค่าความผิดพลาดก็นิยมที่จะใช้วิธีที่สามารถช่วยลดความแตกต่างระหว่างค่าเป้าหมายกับค่าผลลัพธ์ได้

สำหรับแบ็กพรอปเพกชัน การที่จะตรวจสอบว่าระบบเรียนรู้แล้วและสมควรที่จะหยุดสอนเมื่อใดนั้น วิธีที่นิยมใช้ที่สุดเพราะเข้าใจง่ายคือจะพิจารณาจากค่าเฉลี่ยความผิดพลาดกำลังสองที่น้อยที่สุด(Least Mean Squared Error) ว่าน้อยจนยอมรับได้แล้วหรือไม่มีทางค่าที่น้อยกว่านี้ได้อีก

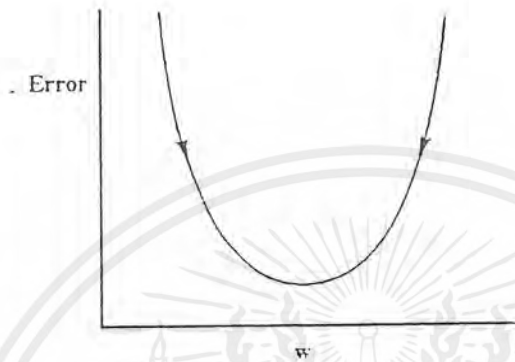
จากหลักการหาค่าเฉลี่ยความผิดพลาดกำลังสอง ตามที่รูเมลฮาร์ต และ แมคเคลลีแลน (Rumelhart and McClelland) ได้กำหนดความหมายของค่าความผิดพลาดไว้ โดยหาผลรวมของผลต่างของค่าเป้าหมายกับค่าผลลัพธ์ทุกๆหน่วยในชั้นผลลัพธ์ สำหรับทุกๆชุดข้อมูลการสอน(Training Pattern) ตาม(4) แล้วหาค่ารากของกำลังสองของผลรวมดังกล่าว ซึ่งจะเป็นการประมาณค่าความผิดพลาดทั้งหมด แล้วหารด้วยจำนวนชุดที่ใช้สอนทั้งหมด ซึ่งจะได้เป็นค่าเฉลี่ยของค่าความผิดพลาดจากการสอน

$$E_p = (\sum (t_p - o_p))^2 \quad \dots(7.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.5 การปรับปรุงค่าน้ำหนัก

ขั้นตอนวิธีการเรียนรู้ในแบ็กพรอพเพกชันชั้นกลางหลายรูปแบบ แต่ที่ได้รับการนิยมนมากที่สุดเป็นกฎเดลต้า (Generalized Delta Rule) โดยรูเมลฮาร์ทและสมาชิกของกลุ่มพีดีพี (Parallel Distribute Processing) ได้ร่วมกันวิจัยขึ้น เป็นขั้นตอนวิธีการเรียนรู้ที่นำไปใช้ได้กับข่ายงานทุกรูปแบบรวมทั้งข่ายงานหลายชั้น โดยไม่จำกัดว่าจะใช้แอกติเวชันแบบใด แต่ต้องเป็นการเรียนรู้แบบมีครู



รูปที่ 7.3 แสดงความสัมพันธ์ระหว่างค่าน้ำหนักกับค่าความผิดพลาด

การปรับปรุงค่าน้ำหนักจะใช้หลักการกราดเสชนท์ (Gradient Descent) ซึ่งเป็นกระบวนการหาหาค่าของค่าน้ำหนัก ที่ทำให้ค่าเฉลี่ยของค่าความผิดพลาดกำลังสองมีค่าน้อยที่สุด นั่นคือเมื่อข้อมูลของแต่ละชุดที่เข้าไปสู่ขั้นตอนการสอนด้วยค่าน้ำหนักชุดหนึ่ง เมื่อผลลัพธ์ไม่ตรงกับค่าเป้าหมายและยอมรับค่าผิดพลาดนั้นไม่ได้ จะทำการปรับปรุงค่าน้ำหนักใหม่โดยที่จะต้องปรับปรุงค่าน้ำหนักทุกชุด เพราะมีส่วนที่ทำให้เกิดความผิดพลาดในชั้นผลลัพธ์ จนกระทั่งการสอนในรอบถัดไปทำให้ค่าความผิดพลาดลดลงไปยังจุดต่ำสุดของแต่ละชุด ตามรูป 7.3 และการปรับปรุงค่าน้ำหนักตามแนวความคิดของกราดเสชนท์ จะได้ตามสูตร (7.5) คือจะทำการเปลี่ยนแปลงค่าน้ำหนักตามสัดส่วนของค่าลบของอนุพันธ์ของค่าความผิดพลาด

$$\Delta w_{ij} = -k(\partial E_p / \partial w_{ij}) \quad \dots (7.5)$$

เมื่อ k คือ ค่าคงที่ หรือจะทำให้อยู่ในรูปที่ง่ายขึ้นดังนี้

$$\Delta w_{ij} = \epsilon \delta_{pj} i_{ij}$$

เมื่อ ϵ = อัตราการเรียนรู้

$$\delta_{pj} = t_{pj} - o_{pj}$$

δ_{pj} เป็นความแตกต่างระหว่างค่าเป้าหมายสำหรับ โหนดที่ j ของชุดการ

สอนที่ p กับค่าผลลัพธ์จริงจากการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยมีการกำหนดค่าตัวแปรขึ้น คือ อัตราการเรียนรู้(Learning Rate) ϵ พบว่าเมื่อค่าอัตราการเรียนรู้มีค่าน้อยการเรียนรู้ช้ามาก แต่ถ้ากำหนดค่าสูงเกินไปการเรียนรู้ไวแต่อาจหาค่าตอบให้กับปัญหาไม่ได้

นอกจากค่าอัตราการเรียนรู้ เป็นตัวแปรสำคัญในการปรับปรุงค่าน้ำหนักแล้ว ยังมีค่าโมเมนตัม(Momentum) เป็นตัวแปรอีกส่วนหนึ่งที่จะทำการปรับปรุงค่าน้ำหนักตามสัดส่วนของการปรับปรุงค่าในรอบที่ผ่านมา

เมื่อแต่ละค่าน้ำหนักถูกปรับปรุงจนได้ค่าผิดพลาดน้อยที่สุดแล้ว ก็จะสามารถใช้ค่าน้ำหนักชุดนี้แก้ปัญหาได้ ถ้าข่ายงานไม่สามารถแก้ปัญหาได้อย่างถูกต้องแสดงว่าชุดของค่าน้ำหนักนี้ยังส่งผลให้มีข้อผิดพลาดอยู่ ถึงแม้ว่าผลการคำนวณหาค่าความผิดพลาดเฉลี่ยจะมีค่าน้อยที่สุดแล้วแต่มันอาจเป็นจุดต่ำสุดท้องถิ่น(Local Minimum) ก็ได้

การเลือกวิธีการกำหนดค่าเริ่มต้นให้กับน้ำหนัก จะมีผลต่อระบบในการหาค่าความผิดพลาด เพราะอาจทำให้ค่าความผิดพลาดนั้นเป็นค่าความผิดพลาดต่ำสุดโดยรวม(Global Minimum of Error) หรือเป็นค่าความผิดพลาดต่ำสุดท้องถิ่น(Local Minimum of Error)

สิ่งสำคัญคือค่าความผิดพลาดนั้นควรเป็นค่าความผิดพลาดต่ำสุดโดยรวม เพราะจะสรุปได้ว่าระบบเกิดการเรียนรู้จริง นั่นคือต้องพยายามหลีกเลี่ยงการใช้ชุดค่าน้ำหนักที่สุ่มขึ้นมา แล้วทำให้การผ่านแอคติเวชันได้ผลลัพธ์ออกมาเป็น 0 และชุดค่าน้ำหนักนั้นต้องไม่ใหญ่เกินไปหรือเล็กเกินไปและถ้าชุดค่าน้ำหนักเล็กเกินไปเมื่อผ่านแอคติเวชันก็อาจจะส่งผลให้ผลลัพธ์นั้นเข้าใกล้ 0 ด้วย ซึ่งจะทำให้การเรียนรู้ช้า

7.6 ขั้นตอนการเรียนรู้และสูตรคำนวณ

ข้อมูลเริ่มต้นที่ระบบต้องเตรียมก่อนจะสอนประกอบด้วย

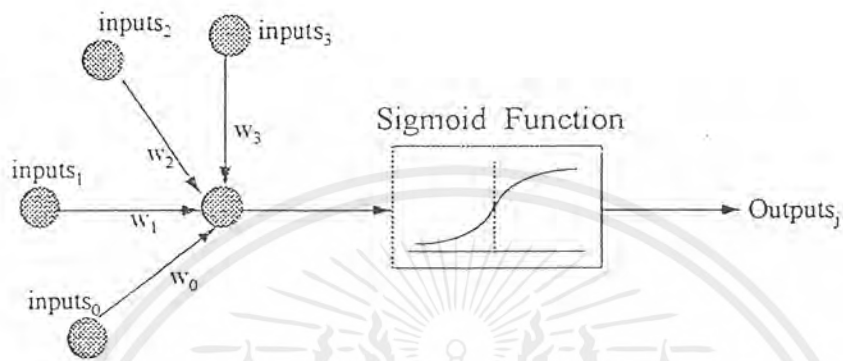
1.ชุดข้อมูลนำเข้าที่ใช้สอนแบ่งเป็น 2 ส่วนคือ ชุดค่าข้อมูลตัวอย่างและชุดค่าข้อมูลเป้าหมายที่ต้องการจากการสอน

2.การกำหนดค่าพารามิเตอร์ต่าง ๆ เช่น ค่าอัตราการเรียนรู้ ค่าโมเมนตัม ค่าความผิดพลาดที่สามารถยอมรับได้ และจำนวนรอบการสอน

3.การกำหนดค่าน้ำหนัก ซึ่งในครั้งแรกที่ทำการสอนจะกำหนดให้ทำการสุ่มขึ้นมา การสอนให้เกิดการเรียนรู้ของนิเวศลดเวกแบบเรียนรู้โดยมีครู เช่น แบ็กพรอพเทกชันจะมีลักษณะของให้ตัวอย่างของข้อมูลหรือชุดสอน(Training Pattern) จะประกอบด้วยคู่ของค่าข้อมูลนำเข้าและค่าเป้าหมายที่ถูกต้องไว้ เมื่อเราส่ง 1 ชุดการสอนไปยังข่ายงานเพื่อประมวลผลหลังการคำนวณจะมีการเปรียบเทียบผลลัพธ์ที่ได้กับค่าเป้าหมาย เมื่อยังมีค่าผิดพลาดที่ยอมรับไม่ได้จะทำการปรับปรุงค่าน้ำหนักใหม่ เพื่อที่จะหาชุดของค่าน้ำหนักที่ทำให้ได้ผลลัพธ์ใกล้เคียงค่าเป้าหมายที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการทำงานเริ่มจากส่งค่าข้อมูลนำเข้าจากชั้นนำเข้าสู่ข้อมูลไปยังชั้นถัดไป ซึ่งคือชั้นแอบแฝง ที่ชั้นนี้จะมีการคำนวณค่าผลรวมของชุดข้อมูลนำเข้าที่ติดต่อกันยัง โหนดนั้นในชั้นแอบแฝงคูณกับค่าน้ำหนักของแต่ละข้อมูลนำเข้า ($\sum x_i w_i$) จากนั้นผลลัพธ์ที่ได้จะผ่านฟังก์ชันเอกติเวชัน หรือในที่นี้คือ ฟังก์ชันซิกมอยด์ และผลลัพธ์สุดท้ายจะเป็นค่าผลลัพธ์ของโหนดหนึ่งในชั้นแอบแฝง (Y_j) และเป็นข้อมูลนำเข้าในชั้นถัดไป ตามรูป 7.4



รูป 7.4 แสดงการคำนวณที่โหนดผลลัพธ์

$$Y_j = F(x) = F(\sum x_i w_i |i|j|) \dots (7.6)$$

Y_j คือ ค่าผลลัพธ์ของโหนดที่ j ที่จะป้อนข้อมูลนำเข้าของชั้นต่อไป

$F(x)$ คือ ฟังก์ชันโลจิสติก ซึ่งเป็นประเภทหนึ่งของฟังก์ชันซิกมอยด์

$$F(x) = 1 / (1 + \exp(-x))$$

$w_i |i|j|$ คือ ค่าน้ำหนักชุดที่ i ของโหนดที่ i ในชั้นนำเข้าข้อมูลที่ส่งไปยังโหนดที่ j ในชั้นแอบแฝง ซึ่งค่าเริ่มต้นจะเป็นค่าสุ่มเล็ก ๆ เช่น ± 0.3

x_i คือ ข้อมูลนำเข้าที่ i

n คือ จำนวนข้อมูลที่นำเข้าของชั้นแอบแฝง

ค่าเริ่มต้นของค่าน้ำหนักในรอบแรกของการสอนจะเป็นค่าสุ่ม และใช้คำนวณในรอบแรกเท่านั้น เพราะในรอบถัดไปจะมีการปรับปรุงค่าน้ำหนักใหม่ และค่าผลลัพธ์ Y_j ของโหนดที่ j ในชั้นแอบแฝง จะเป็นข้อมูลนำเข้าของชั้นถัดไป ในที่นี้คือชั้นผลลัพธ์เนื่องจากเป็นช่วงงานที่มี 1 ชั้นแอบแฝง

ขั้นตอนที่กล่าวมาข้างต้นนี้จะต้องคำนวณหาผลลัพธ์ของทุก ๆ โหนดในชั้นแอบแฝงเพื่อเป็นข้อมูลนำเข้าในชั้นผลลัพธ์ก่อน หลังจากนั้นจึงจะใช้วิธีการคำนวณแบบเดียวกันในชั้นผลลัพธ์ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Z_k = F(x) = F(\sum Y_j w_2[j][k]) \quad \dots(7.7)$$

Z_k คือ ค่าผลลัพธ์ของโหนดที่ k ในชั้นผลลัพธ์ที่ใช้เปรียบเทียบกับค่าเป้าหมาย

$F(x)$ คือ ฟังก์ชันโลจิสติก ซึ่งเป็นประเภทหนึ่งของฟังก์ชันซิกมอยด์

$$F(x) = 1 / (1 + \exp(-x))$$

Y_j คือ ข้อมูลนำเข้าของชั้นผลลัพธ์ ซึ่งคือค่าผลลัพธ์ของโหนดที่ j ของชั้นแอบแฝง

$w_2[j][k]$ คือ ค่าน้ำหนักชุดที่ 2 ของโหนดที่ j ในชั้นแอบแฝงที่ส่งไปยังโหนดที่ k ในชั้นผล

ลัพธ์ซึ่งค่าเริ่มต้นจะเป็นค่าสุ่มเล็ก ๆ เช่นกัน

m คือ จำนวนข้อมูลนำเข้าของชั้นผลลัพธ์

หลังจากนั้นจะเปรียบเทียบผลลัพธ์จากการสอนกับค่าเป้าหมาย เพื่อหาค่าความผิดพลาดของแต่ละโหนดในชั้นผลลัพธ์ และรวมทั้งหาค่าความผิดพลาดของแต่ละโหนดในชั้นแอบแฝงด้วยเพื่อที่จะปรับปรุงชุดค่าน้ำหนักสำหรับขั้นตอนการคำนวณในรอบต่อไป โดยมีขั้นตอนดังนี้

1. หาค่าความผิดพลาดของผลลัพธ์ของทุกโหนดที่ k ในชั้นผลลัพธ์

$$\delta_k = (T_k - Z_k) Z_k (1 - Z_k) \quad \dots(7.8)$$

δ_k คือ ค่าความผิดพลาดของโหนดที่ k ในชั้นผลลัพธ์

T_k คือ ค่าเป้าหมาย(target value)

Z_k คือ ค่าผลลัพธ์ของ โหนดที่ k ในชั้นผลลัพธ์ที่ใช้เปรียบเทียบกับค่าเป้าหมาย

2. หาค่าความผิดพลาดของโหนดทุกโหนดในชั้นแอบแฝงที่ทำให้เกิดความผิดพลาดในชั้นผลลัพธ์ ของทุกโหนดที่ j ในชั้นแอบแฝง

$$E_j = Y_j(1 - Y_j) \sum w_2[j][k] \delta_k \quad \dots(7.9)$$

E_j คือ ค่าความผิดพลาดของโหนดที่ j ในชั้นแอบแฝง

δ_k คือ ค่าความผิดพลาดของโหนดที่ k ในชั้นผลลัพธ์

Y_j คือ ค่าผลลัพธ์ของ โหนดที่ j ในชั้นแอบแฝง หรือค่าข้อมูลนำเข้าในชั้นผลลัพธ์นั่นเอง

M คือ จำนวนโหนดของชั้นผลลัพธ์

$w_2[j][k]$ คือ ค่าน้ำหนักจาก โหนดที่ j ในชั้นแอบแฝงไปยังโหนดที่ k ในชั้นผลลัพธ์

3. การปรับปรุงค่าน้ำหนักในชั้นแอบแฝง จากค่าผิดพลาดในชั้นผลลัพธ์ ซึ่งจะปรับปรุงค่าน้ำหนักจากโหนดที่ j ในชั้นแอบแฝงไปยังทุกโหนดในชั้นผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\Delta w_2[j][k] = \varepsilon \delta_k Y_j + \alpha \Delta w_2[j][k]_{\text{lastCycle}} \quad \dots(7.10)$$

$\Delta w_2[j][k]$ คือ ค่าที่ใช้ปรับปรุงค่าน้ำหนักจากโหนดที่ j ในชั้นแอบแฝงไปยังโหนดที่ k ในชั้นผลลัพธ์

$\alpha \Delta w_2[j][k]_{\text{lastCycle}}$ คือ ค่าที่ใช้ปรับปรุงค่าน้ำหนักจากโหนดที่ j ในชั้นแอบแฝงไปยังโหนดที่ k ในชั้นผลลัพธ์ ในรอบการสอนที่ผ่านมา

δ_k คือ ค่าความผิดพลาดของโหนดที่ k ในชั้นผลลัพธ์

Y_j คือ ข้อมูลนำเข้าของชั้นผลลัพธ์ ซึ่งคือค่าของชั้นผลลัพธ์ของโหนดที่ j ของชั้นแอบแฝง

ε คือ ค่าอัตราการเรียนรู้ (learning rate) ของชั้นผลลัพธ์

α คือ ค่าโมเมนตัม

ค่าน้ำหนักใหม่คือ $w_2[j][k] + \Delta w_2[j][k]$

4. การปรับปรุงค่าน้ำหนักในชั้นนำเข้าข้อมูล จากค่าผิดพลาดของชั้นผลลัพธ์ ซึ่งจะปรับปรุงค่าน้ำหนักจากโหนดที่ l ในชั้นนำเข้าข้อมูล ไปยังทุกโหนดในชั้นแอบแฝง

$$\Delta w_1[j][k] = \varepsilon E_l x_j + \alpha \Delta w_1[j][k]_{\text{lastCycle}} \quad \dots(7.11)$$

$w_1[j][k]$ คือ ค่าน้ำหนักจากโหนดที่ l ในชั้นนำเข้าข้อมูลไปยังโหนดที่ j ชั้นแอบแฝง

$\alpha \Delta w_1[j][k]_{\text{lastCycle}}$ คือ ค่าน้ำหนักจากโหนด l ในชั้นนำเข้าข้อมูลไปยังโหนดที่ j ชั้นแอบแฝง ในรอบการสอนที่ผ่านมา

E_l คือ ค่าความผิดพลาดของโหนดที่ j ในชั้นแอบแฝง

x_j คือ ข้อมูลนำเข้าของโหนดที่ l ของชั้นแอบแฝง

ε คือ ค่าอัตราการเรียนรู้ (learning rate) ของชั้นแอบแฝง

ค่าน้ำหนักใหม่คือ

จากสูตร (7.10) และ (7.11) จะได้ว่ากระบวนการเรียนจะมีการปรับปรุงค่าน้ำหนัก ตามสัดส่วนของค่าความผิดพลาด และค่าอัตราการเรียนรู้ (learning rate) และระบบจะเรียนรู้ช้าเมื่อค่าอัตราการเรียนรู้น้อย แต่เมื่ออัตราการเรียนรู้ มีค่ามาก ๆ จะมีการปรับปรุงค่าน้ำหนักมากด้วย ทำให้เกิดการแกว่งของค่าความผิดพลาด วิธีแก้ปัญหาดังกล่าวจะเพิ่มโมเมนตัม (Momentum) เข้าไปในการคำนวณ นั่นคือ ส่วนหนึ่งของการปรับปรุงค่าน้ำหนัก จะขึ้นกับค่าโมเมนตัมของการปรับปรุงค่าน้ำหนักในรอบที่ผ่านมา หรือจะได้ว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\Delta w_{ij}[k](n+1) = \varepsilon \delta_k y_j + \alpha \Delta w_{ij}[k](n) \quad \dots(7.12)$$

$$\Delta w_{ij}[k](n+1) = \varepsilon E_{j,x_i} + \alpha \Delta w_{ij}[k](n) \quad \dots(7.13)$$

(n) เป็น ดัชนีระบุ presentation number

α เป็น ค่าโมเมนตัมมีค่าคงที่ กำหนดผลกระทบของการปรับปรุงค่า น้ำหนักที่ผ่านมา จะมีค่าระหว่าง 0 ถึง 1

ระบบจะเรียนรู้ได้เร็วขึ้น เมื่อกำหนดค่าโมเมนตัม และค่าอัตราการเรียนรู้ใหญ่ ๆ เพราะ จะทำการปรับปรุงค่าน้ำหนักมากกว่าที่กำหนดให้มีค่าเล็ก ดังนั้นมีการทดสอบค่าต่าง ๆ ที่เหมาะสมของ momentum ค่าหนึ่งคือ 0.9 แต่อย่างไรก็ตามการกำหนดค่านี้นั้นขึ้นกับแต่ละปัญหาด้วย

7.7 จำนวนชุดข้อมูลการสอน

มีการพิสูจน์แล้วว่า ในการสอนให้ระบบสามารถเรียนรู้ในงานแบ่งแยกประเภท (Pattern Classification) ความถูกต้องจากการทดสอบระบบที่เรียนรู้แล้วจะคิดเป็นเปอร์เซ็นต์ของจำนวนชุดการสอน และการทดสอบว่าระบบจะให้คำตอบถูกต้องเมื่อข้อมูลที่ใช้ทดสอบอยู่ในกลุ่มของตัวอย่างที่ใช้สอน

โดยมีความสัมพันธ์ระหว่างจำนวนชุดการสอนกับจำนวนค่าน้ำหนักกับความถูกต้องในการทดสอบดังนี้

$$P = W/c$$

เมื่อ	P	คือ จำนวนชุดการสอน
	W	คือ จำนวนค่าน้ำหนัก
	1-c	คือ ความต้องการความถูกต้องในการทดสอบระบบที่เรียนแล้ว

ถ้าจำนวนค่าน้ำหนัก 80 ค่า ถ้าต้องการความถูกต้อง 90% คือ c มีค่า 0.1 และจะต้องมีจำนวนชุดการสอน 800 ชุด

7.8 รอบการสอน (Cycle or Epoch)

ปัญหาอย่างหนึ่งของการสอนคือควรจะกำหนดให้มีการสอนกี่รอบ เนื่องจากการสอนมากเกินไปไม่ได้ทำให้ระบบเรียนรู้ได้มากขึ้น เราจึงจำเป็นต้องประมาณการจำนวนน้อยที่สุดของการสอนที่จำเป็นสำหรับการแก้ปัญหาหนึ่ง ๆ ที่ทำให้ระบบเกิดการเรียนรู้ และสามารถหยุดการสอนได้ พบว่าระบบควรหยุดการสอนเมื่อเฉลี่ยของค่าความผิดพลาดแต่ละชุดข้อมูลถึงจุดต่ำสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อระบบอยู่ในขั้นตอนการเรียนรู้ แต่ละชุดการสอนจะมีการคำนวณค่าเฉลี่ยของค่าความผิดพลาด และค่านี้จะมีค่าลดลงเมื่อระบบเริ่มที่จะเรียนรู้ เมื่อใดก็ตามที่ค่านี้เริ่มสูงขึ้นแสดงว่าการเรียนรู้ นั้นเริ่มจะสูญเสียการเรียนรู้แล้ว

ในทางปฏิบัติเราต้องลองคิดลองดู ด้วยนี้ทุกอย่างขึ้นอยู่กับประสบการณ์ของผู้ใช้ คือ อาจกำหนดรอบการสอนให้มีค่ามาก ๆ และอาจหยุดการสอนเมื่อเห็นว่าค่าเฉลี่ยของค่าความผิดพลาด อยู่บนขอบเขตที่ยอมรับได้ ดังนั้นหลังการสอนจำเป็นจะต้องมีการตรวจสอบความสามารถของระบบ

7.9 การแปลงข้อมูลนำเข้า

สิ่งสำคัญอย่างหนึ่งการสอนคือการแปลงข้อมูลนำเข้าจากปัญหาที่ต้องการแก้ไขให้มีประสิทธิภาพ นั่นคือข้อมูลนำเข้าควรจะมีรูปแบบใดลงจะทำให้ระบบเรียนรู้ได้รวดเร็ว และสามารถแก้ปัญหาได้

พบว่าเมื่อ ผลลัพธ์จากการผ่านแอกติเวชันแล้วมีค่าใกล้ 0 มักจะเรียนรู้ช้า หรืออาจจะไม่เรียนเลย ดังนั้นการปรับปรุงการเรียนรู้ให้ดีขึ้นข้อมูลนำเข้าไม่ควรอยู่ในรูปไบนารี คือ ค่า 0 และ 1 เท่านั้น นั่นคือควรจะเป็นค่าที่เป็นค่าที่เข้าใกล้ 0 และ 1 และมีทั้งค่าลบ และค่าบวก

จากการทดสอบการแก้ปัญหา XOR ที่มีโครงสร้าง 2-4-1 คือ จำนวนโหนดข้อมูลนำเข้ามีมีขนาด 2 และจำนวนโหนดข้อมูลแอบแฝงมีขนาด 4 และจำนวนโหนดข้อมูลผลลัพธ์มีขนาด 1 นั่นคือต้องผลลัพธ์เพียงถูกหรือผิดเท่านั้น โดยกำหนด ค่าน้ำหนักเริ่มต้นอยู่ในขอบเขต -0.5 ถึง 0.5 และเมื่อกำหนดข้อมูลนำเข้าแบบ ไบนารีก็จะใช้แอกติเวชันแบบไฮโปลา

เมื่อแทนค่าข้อมูลนำเข้าแบบ ไบนารีพบว่าต้องสอนถึง 3000 รอบถึงจะเรียน และในการแก้ปัญหาเดียวกันเมื่อแทนค่าแบบไฮโปลา ระบบเรียนรู้เร็วขึ้นสอนเพียง 390 รอบ ระบบก็เรียนรู้ได้แล้ว

7.10 อัลกอริธึม นิวรอลเน็ตเวิร์กที่มีชั้นแฝงชั้นเดียว

(one hidden layer neural network)

ขั้นที่ 0 เริ่มต้นค่าน้ำหนัก(Initialize weight)

(กำหนดให้เป็นค่าที่เกิดจากการสุ่มตัวลขน้อย ๆ)

ขั้นที่ 1 ขณะที่จะอินพุตชุดเป็นเท็จ ทำขั้นที่ 2 ถึง 9

ขั้นที่ 2 สำหรับแต่ละคู่การฝึก, ทำขั้นที่ 3-8

Feedforward :

ขั้นที่ 3 แต่ละหน่วยอินพุต ($X_i, i=1, \dots, n$) รับสัญญาณอินพุต X_i และกระจาย

สัญญาณนี้ไปยังทุกหน่วยในชั้นข้างบน(หน่วยแฝง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นที่ 4 แต่ละหน่วยแฝง ($Z_j, j=1, \dots, p$) รวมค่าสัญญาณอินพุตที่ถ่วงน้ำหนัก

$$z_{in_j} = v_{0j} + \sum x_i v_{ij}$$

นำแอกติวชันฟังก์ชันมาใช้เพื่อคำนวณสัญญาณเอาต์พุต

$$z_j = f(z_{in_j})$$

และส่งสัญญาณไปยังหน่วยทั้งหมดในชั้นข้างบน (หน่วยเอาต์พุต)

ขั้นที่ 5 แต่ละหน่วยเอาต์พุต ($Y_k, k=1, \dots, m$) รวมค่าสัญญาณอินพุตที่ถูกถ่วงน้ำหนัก

หน้า

$$y_{in_k} = w_{0k} + \sum z_j w_{jk}$$

นำแอกติวชันฟังก์ชันมาใช้เพื่อคำนวณสัญญาณเอาต์พุต

$$y_k = f(y_{in_k})$$

Backpropagation of error

ขั้นที่ 6 แต่ละหน่วยเอาต์พุต ($Y_k, k=1, \dots, m$) ระบุรูปแบบเป้าหมาย (target pattern)

ให้สอดคล้องกับรูปแบบการฝึกอินพุต (input training pattern) ทำการคำนวณในรูปแบบ

$$\delta_k = (t_k - y_k) f'(y_{in_k})$$

คำนวณเทอรัมการปรับแก้หน้าหนัก (เพื่อใช้ปรับปรุง w_{jk} ภายหลัง)

$$\Delta w_{jk} = \eta \delta_k z_j$$

และส่ง δ_k ไปยังหน่วยในระดับล่าง

ขั้นที่ 7 แต่ละหน่วยแฝง ($Z_j, j=1, \dots, p$) รวมเดลต้าอินพุต (จากหน่วยในชั้นข้าง

บน)

$$\delta_{in_j} = \sum \delta_k w_{jk}$$

คูณด้วยค่าอนุพันธ์ของแอกติวชันฟังก์ชันเพื่อที่จะคำนวณค่าเทอมของความผิดพลาด

$$\delta_j = \delta_{in_j} f'(z_{in_j})$$

คำนวณค่าเทอรัมแก้ไขหน้าหนัก

$$\Delta v_{ij} = \eta \delta_j x_i$$

Update weights:

ขั้นที่ 8 แต่ละหน่วยเอาต์พุต ($Y_k, k=1, \dots, m$) ปรับปรุงค่าน้ำหนัก ($j=0, \dots, p$):

$$w_{jk}(\text{new}) = w_{jk}(\text{old}) + \Delta w_{jk}$$

แต่ละหน่วยแฝง ($Z_j, j=1, \dots, p$) ปรับปรุงค่าน้ำหนัก ($i=0, \dots, n$):

$$v_{ij}(\text{new}) = v_{ij}(\text{old}) + \Delta v_{ij}$$

ขั้นที่ 9 ทดสอบเงื่อนไขหยุด (Test stopping condition)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 8

การทดลองและผลการทดลอง

ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลเสียงคำไทย โดยเลือกเฉพาะคำที่เป็นคำโดด ๆ ชนิดหนึ่งคำหนึ่งพยางค์ ที่มีความหมายในตัวที่สอดคล้องกับการนำไปประยุกต์ใช้กับการสั่งงานหรือควบคุมในจำนวนที่ไม่มากนักคือ 10 คำ

ผู้วิจัยเลือกวิเคราะห์เสียงผู้ชายคนเดียว เพื่อความง่ายในการทำการวิจัยแล้วจึงค่อยทำการวิจัยเพื่อขยายให้สามารถใช้ได้กับเสียงผู้ชายหลาย ๆ คนต่อไป

8.1 วิธีการเก็บและจัดการข้อมูลเพื่อการประมวลผล

ผู้วิจัยเก็บข้อมูลเสียง โดยแบ่งเป็น 3 กลุ่ม ดังนี้

กลุ่ม A ใช้สร้างข้อมูลหลักสำหรับหาค่าสัมประสิทธิ์สัมพันธ์ กับกลุ่ม B โดยผู้พูดจะพูด “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า”, รวม 10 คำ เรียงตามลำดับ 1 รอบ จะได้ตัวอย่างเสียงทั้งหมด 10 ตัวอย่างเสียง ข้อมูลเสียงในกลุ่ม A นี้จะถูกใช้เป็นฐานข้อมูลหลักสำหรับคำนวณหาสัมประสิทธิ์สหสัมพันธ์ ในกระบวนการ Preprocessing จะได้คุณลักษณะจำเพาะสัมพันธ์ ของข้อมูลเสียงทุกเสียงในกลุ่ม B

กลุ่ม B วิธีการเก็บข้อมูลใช้วิธีเดียวกับการเก็บตัวอย่างเสียงในกลุ่ม A เป็นจำนวน 100 ตัวอย่าง ข้อมูลเสียงกลุ่ม B นี้จะถูกนำไปผ่านกระบวนการ Preprocessing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ ซึ่งข้อมูลในส่วนนี้จะถูกนำไปใช้ในการสอนโครงข่ายประสาทเทียม

กลุ่ม C วิธีการเก็บข้อมูลใช้เทคนิคเดียวกับกลุ่ม A จะได้ 80 ตัวอย่างเสียง ข้อมูลกลุ่ม B นี้จะถูกนำไปผ่านกระบวนการ Preprocessing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ ข้อมูลส่วนนี้จะนำไปใช้ในการทำนายว่าเป็นเสียงใด

ข้อมูลทุกข้อมูลจะได้มาจากการอัดเสียงโดยใช้การ์ดชาวบลาสเตอร์ ซึ่งผลิตโดยบริษัท Creative Labs, Inc. กำหนดอัตราสุ่มที่ 8 KHz แบบ 8 บิตโมโนจะได้เป็นแฟ้มข้อมูลเสียงของแต่ละเสียงจากนั้นนำข้อมูลเสียงไปผ่านส่วนตัดข้อมูลก่อนแล้วก็นำไปวิเคราะห์สัญญาณในเชิงความถี่ ข้อมูลที่ได้นี้จะนำไปผ่านส่วน Preprocessing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ และจะนำไปใช้ในระบบรู้จำที่สร้างขึ้นจากโครงข่ายประสาทเทียม ต่อไป

8.2 การดำเนินการทดลองและผลการทดลอง

ในการดำเนินการทดลองนั้น จะถูกแบ่งออกเป็น 2 ขั้นตอนด้วยการคือส่วนการเตรียมข้อมูลและ ส่วนตัดสินใจ ในการเตรียมข้อมูลนั้นจะแปลงสัญญาณให้อยู่ในรูปค่าสแอมป์พันธ์ และในส่วนการตัดสินใจ จะใช้วิธีโครงข่ายประสาทเทียมแบบแบคพร็อบเพกชัน

8.2.1 ส่วนเตรียมข้อมูล

1). เก็บตัวอย่างข้อมูลและกลุ่มข้อมูลที่ใช้บันทึกเสียงตัวอย่างที่ต้องการโดยผ่านเซนเซอร์ การ์ดชาวบลาสเตอร์ โดยกำหนดอัตราการสุ่มตัวอย่างที่ 8 KHz แบบ 8 บิตโมโน จะได้เป็นแฟ้มข้อมูลเสียงของแต่ละเสียง ซึ่งบันทึกเป็นข้อมูลในรูปแบบ WAV FILE (*.WAV) และแบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่มคือ กลุ่มเสียงหลัก กลุ่มเสียงอ้างอิง กลุ่มเสียงทดสอบซึ่งทั้ง สามกลุ่มมาจากเสียงคน ๆ เดียวกัน

2). หาขอบเขตของสัญญาณเสียง และทำ window การหาขอบเขตของสัญญาณเสียงใน Time domain หรือการตัดเสียงที่เป็นช่วงของเสียงเงียบออก จะใช้วิธีการวิเคราะห์โดยวัดระดับพลังงานและขนาดเฉลี่ยโดยวัดในช่วงเวลาสั้นตามที่กล่าวไว้ในบทที่ 4 และการวิเคราะห์อัตราการตัดศูนย์ของสัญญาณเสียง โดยมีสมมติฐานว่า อัตราการตัดศูนย์เป็นลักษณะทางความถี่ที่แสดงอยู่ใน Time Domain ฉะนั้นสัญญาณเสียงที่เกิดจากลมปากขณะเริ่มออกเสียงนั้นจะมีค่าพลังงานในจุดเริ่มต้นของเสียงมีสูง เช่นเสียงหก ซึ่งส่วนของเสียงสมนี้ ไม่ถือว่าเป็นคุณลักษณะที่สำคัญของเสียง ส่วนเสียง สอง ตาม ที่ ศูนย์ เสียงเหล่านี้ ในตอนเริ่มต้นเสียง สัญญาณจะมีความถี่ หรือค่าอัตราการตัดศูนย์ในตอนเริ่มต้นสูงมาก แต่ที่ตำแหน่งนี้สัญญาณจะมีค่าพลังงานต่ำมาก ซึ่งทำให้จำเป็นต้องใช้ทั้ง อัตราการตัดศูนย์ และการวัดระดับพลังงานเฉลี่ยในช่วงสั้นช่วยพิจารณา จากนั้นนำสัญญาณเสียงที่ทำการแบ่งออกเป็นช่วงย่อย ซึ่งในการนี้ ต้องการ Output ทั้งหมดของแต่ละสัญญาณไปวิเคราะห์ด้วยโครงข่ายประสาทเทียมอีกครั้ง จึงต้องการ Output จาก Preprocessing จำนวนเท่ากัน แต่เนื่องจาก เสียงพูดแต่ละสัญญาณมีขนาดความยาวแตกต่างกัน จึงใช้วิธีการเข้าข้อมูล โดยข้อมูลเสียงที่มีความยาวมาก จะถูกเข้าข้อมูลน้อย แต่ในข้อมูลที่สั้น จะเข้าข้อมูลมากขึ้น ซึ่งวิธีนี้จะยังคงรักษาคุณลักษณะของเสียงได้ทั้งหมด วิธีการนี้จะทำโดยการกำหนดให้เสียงทุกเสียงจะต้องถูกแบ่งออกเป็นจำนวนเฟรมคงที่ และแต่ละเฟรมต้องมีขนาดเท่ากัน ฉะนั้นแต่ละเสียงจะมีการเข้าข้อมูล (Overlap) ที่ไม่เท่ากัน ในที่นี้เลือกการแบ่งข้อมูลเสียงแต่ละเสียงออกเป็น 10 เฟรม โดยแต่ละเฟรมจะต้องมีข้อมูลอยู่ 1024 ค่า นอกจากนี้ กำหนดให้ค่า การเข้าข้อมูลของแต่ละเฟรมอย่างน้อยที่สุด 256 ค่า จากนั้นนำข้อมูลแต่ละเฟรมที่ได้ มาคูณด้วย ฮานนิงวินโดว์

3) นำข้อมูลที่ได้มาจากขั้นตอนที่ 2 ทำการแปลงข้อมูลเสียงทุกเสียงซึ่งเป็นข้อมูลในเชิงเวลา ให้เป็นข้อมูลเชิงความถี่ ด้วยวิธีการแปลงแบบฟูเรียร์อย่างรวดเร็ว ฉะนั้น แต่ละเสียง จะได้ค่าใน Frequency ออกมาเสียงละ 10 เฟรม เช่นจะได้สเปกตรัมของเสียง แต่ในการแปลงฟูเรียร์อย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรีวนั้น จะได้ข้อมูลออกมา ในรูปของ Magnitude ที่มีลักษณะเท่ากันทุกประการ 2 ด้าน ฉะนั้น จะเลือกข้อมูลส่วนที่ต้องการออกมาเพียงครั้งเดียว คือ 512 ค่าเท่านั้นในการที่จะนำไปหาค่ามาตรฐานต่อไป

4) นำข้อมูลที่ได้จากในขั้นตอนที่ 3 มาแปลงให้อยู่ในรูปมาตรฐาน หรืออยู่ในรูปค่าคุณลักษณะจำเพาะสัมพัทธ์ของกลุ่มเสียงทดสอบ เทียบกับเสียงหลัก (ข้อมูลชุด A)

5.) เปรียบเทียบค่าคุณลักษณะจำเพาะสัมพัทธ์ที่ได้มาจากการเทียบเสียงทดสอบกับเสียงหลัก ทั้ง 10 เสียง จะได้ค่าสัมประสิทธิ์สหสัมพัทธ์ 10 ค่า จะสามารถบอกขั้นต้นได้ว่าเสียงที่นำมาทดสอบนั้นน่าจะเป็นเสียงใด ซึ่งสำหรับกรณีที่มีเสียงอ้างอิง หรือกลุ่มข้อมูลหลักจำนวนไม่มาก การหาค่าสัมประสิทธิ์สหสัมพัทธ์ จะสามารถตรวจสอบและตัดสินใจได้บ้างก็จริง แต่ เปอร์เซนต์ความถูกต้อง ยังไม่เหมาะที่จะนำมาใช้งาน จึงควรมีการหาวิธีการตัดสินใจด้วยที่ดีขึ้น สำหรับสมมติฐานของการทดลองนี้ คือ ค่าสัมประสิทธิ์สหสัมพัทธ์ของเสียงหนึ่งเทียบกับเสียงใดใด จะมีลักษณะที่คล้ายคลึงกัน จึงนำค่าที่ได้นี้ไปเป็น Input ให้กับโครงข่ายประสาทเทียมต่อไป

8.2.2 ส่วนตัดสินใจ

นิวรอนเน็ตเวิร์คที่ใช้ประกอบด้วย 3 ชั้น ดังต่อไปนี้

- ชั้นอินพุทเลเยอร์(input layer) จำนวน 10 หน่วย
 - ชั้นซ่อน(hidden layer) จำนวน 6 หน่วย
 - ชั้นเอาต์พุท เลเยอร์ (Output Layer) จำนวน 4 หน่วย
- จำนวน 1 ฟิล์เสียงที่ใช้ในการเทรนคือ 100 เสียง
จำนวน 1 ฟิล์เสียงที่ใช้ในการทดสอบคือ 80 เสียง

8.2.2.1 ขั้นตอนการสอน

จะประกอบด้วยขั้นตอนดังต่อไปนี้

1.การจองเนื้อที่สำหรับตัวแปรพารามิเตอร์และโครงสร้างของระบบในชั้นต่าง ๆ และกำหนดชุดของค่าน้ำหนักที่ใช้สอนหรือจะสุ่มค่าขึ้นมาใหม่

2.การรับค่าชุดข้อมูลนำเข้าที่ประกอบด้วยกลุ่มข้อมูล 2 กลุ่มคือ กลุ่มข้อมูลนำเข้าที่เป็นค่าของแต่ละโหนดในชั้นนำเข้าข้อมูล และกลุ่มข้อมูลเป้าหมายที่เป็นค่าเป้าหมายของแต่ละโหนดในชั้นผลลัพธ์

3.ทำการคำนวณหาค่าผลลัพธ์จากค่าน้ำหนักและค่าข้อมูลนำเข้าในชั้นนำเข้าข้อมูล แล้วส่งผลลัพธ์นั้นไปเป็นข้อมูลนำเข้าในแต่ละโหนดในชั้นแอบแฝง และในทำนองเดียวกันก็จะคำนวณหาผลลัพธ์จากชุดข้อมูลนำเข้าหรือผลลัพธ์จากชั้นที่ผ่านมากับค่าน้ำหนักในชั้นนี้ แล้วส่งไปเป็นค่าข้อมูลนำเข้าในชั้นถัดไป จนกระทั่งถึงชั้นผลลัพธ์

4.เปรียบเทียบผลลัพธ์จากการคำนวณจาก (3) ของแต่ละโหนดในชั้นผลลัพธ์กับค่าเป้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หมายถึงแต่ละโหนด ในชั้นผลลัพธ์ เพื่อคำนวณค่าเฉลี่ยความผิดพลาดกำลังสอง

5. จาก(4) หากค่าเฉลี่ยความผิดพลาดกำลังสองที่คำนวณจาก (4) มีค่าน้อยกว่าเงื่อนไขที่กำหนด หรือหากจำนวนรอบในการสอนครบตามเงื่อนไขที่กำหนด ก็จะหยุดการสอนและบันทึกชุดของค่าน้ำหนักทั้งหมดลงเพิ่มข้อมูล

6. จาก(4) หากค่าเฉลี่ยความผิดพลาดกำลังสองที่คำนวณยังมากอยู่ จะทำการคำนวณหาค่าความผิดพลาดแล้วส่งให้กับชั้นก่อนหน้าทุก ๆ ชั้น กรคำนวณหาค่าความผิดพลาดได้จาก 3 ตัวแปร คือ ชุดค่าน้ำหนักในชั้นนั้น , ชุดของข้อมูลนำเข้าในชั้นนั้น และชุดค่าความผิดพลาดของชั้นก่อนหน้า

7. การปรับปรุงค่าน้ำหนักจากค่าความผิดพลาดที่คำนวณได้ใน(6) และทำการปรับปรุงค่าน้ำหนักในทุก ๆ ชั้นที่มีค่าน้ำหนัก

8. นำชุดข้อมูลนำเข้าชุดถัดไปเข้าสู่ขั้นตอนการสอน และทำการประมวลผลตาม(3) ถึง (6) จนกระทั่งชุดข้อมูลนำเข้าทุกชุดผ่านผ่านเข้าสู่ขั้นตอนการสอนหมดแล้ว จะเรียกว่า 1 รอบการสอน และจะทำการวนเอาข้อมูลชุดแรกเข้าสู่ขั้นตอนการสอนอีกครั้ง จนกว่าค่าเฉลี่ยความผิดพลาดกำลังสองที่คำนวณในชั้นผลลัพธ์จาก (3) มีค่าน้อยกว่าเงื่อนไขที่กำหนด ก็จะหยุดการสอนและบันทึกชุดของค่าน้ำหนักทั้งหมดลงเพิ่มข้อมูล

เงื่อนไขการหยุดสอนของ โครงข่ายจะกำหนดจากค่าความผิดพลาดรวมทั้ง โครงข่ายซึ่งหาได้จาก

$$\text{Train Error} = (\sum \text{Erf}[n]) / \text{Numfile}$$

ค่า Erf คือค่าความผิดพลาดที่คำนวณจากแต่ละไฟล์อินพุท และค่าเป้าหมายของไฟล์

$$\text{Erf}[\text{ไฟล์ที่ } n] = 0.5 * (\sum (\text{target}_m - \text{output}_m)^2)$$

จะเห็นว่าการสอนให้ระบบเรียนรู้จะมีการประมวลผลกับชุดข้อมูลตัวอย่างทั้งหมดและจะมีการสอนหลาย ๆ รอบ จนกระทั่งค่าเฉลี่ยความผิดพลาดกำลังสองจะลดน้อยลงจนเป็นที่พอใจและกระบวนการสอนอาจจัดกลุ่มได้ 2 ประเภท คือ การคำนวณหาผลลัพธ์ในชั้นต่าง ๆ (forward Pass) และการส่งค่าความผิดพลาดกลับและทำการปรับปรุงค่าน้ำหนัก(Backward Pass)

8.2.2.2 ขั้นตอนการทดสอบ

จะประกอบด้วยขั้นตอนดังต่อไปนี้

1. การจองเนื้อที่สำหรับตัวแปรพารามิเตอร์และโครงสร้างของระบบในชั้นต่าง ๆ รวมทั้งการกำหนดชุดของค่าน้ำหนักที่ใช้ทดสอบ

2. การรับค่าชุดข้อมูลนำเข้าที่ใช้ทดสอบซึ่งจะประกอบด้วยกลุ่มข้อมูลเพียงกลุ่มเดียว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คือกลุ่มข้อมูลนำเข้าที่เป็นค่าของแต่ละโหนดในชั้นนำเข้าข้อมูลเท่านั้น เพราะในชั้นตอนนี้จะประมวลผลเพียงรอบเดียว โดยไม่มีการปรับปรุงค่าน้ำหนักและไม่มีการเปรียบเทียบกับชุดข้อมูลเป้าหมาย

3. การคำนวณหาผลลัพธ์จากชุดข้อมูลนำเข้าในชั้นนั้นกับชุดของค่าน้ำหนักในชั้นนั้นเช่นกัน และส่งผลลัพธ์ให้กับชั้นถัดไปซึ่งผลลัพธ์นี้จะเป็นชุดข้อมูลนำเข้าให้กับชั้นถัดไปด้วย แล้วทำการคำนวณหาผลลัพธ์จนกระทั่งถึงชั้นผลลัพธ์ จะถือว่าเป็นผลลัพธ์ของชุดข้อมูลที่ใช้ทดสอบ

8.3 ผลการทดลอง

ตารางที่ 8.1 แสดงค่าอัตราการเรียนรู้จากการทดสอบนิเวศน์เน็ตเวิร์ก ที่ค่า Momentum =0.3, Learning Rate =0.7

แบบทดสอบ	แบบอ้างอิง										%ความถูกต้อง
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	
ศูนย์	6		2								75
หนึ่ง		8									100
สอง			5				3				62.5
สาม				8							100
สี่	2				6						75
ห้า		1				3			1	3	37.5
หก							8				100
เจ็ด						2		6			75
แปด									8		100
เก้า		2		1						5	62.5
										เฉลี่ย	78.75
										สูงสุด	100
										ต่ำสุด	37.5

จากการทดสอบการทำงานของโปรแกรมโครงข่ายประสาทเทียม (โดยใช้ข้อมูลเสียงที่ผ่านการหาค่าสัมประสิทธิ์สหสัมพันธ์ แล้ว) พบว่า เปรอร์เซนต์การจดจำของโครงข่ายไม่ค่อยดีนักถ้าเทียบกับการตัดสินใจปัญหาโดยใช้วิธี สัมประสิทธิ์สหสัมพันธ์ เนื่องมาจากเหตุผลหลายประการที่มีผลต่อการจดจำของโครงข่ายประสาทเทียม ซึ่งก็คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8.3.1 รูปแบบของข้อมูลมีผลต่อโครงข่ายประสาทเทียมคือ ยิ่งข้อมูลมีความแตกต่างกันมากเท่าไรจะยิ่งทำให้การเรียนรู้เร็วขึ้นเท่านั้นสังเกตได้จากเมื่อนำค่าสหสัมพันธ์ที่ยังไม่สามารถตัดสินใจได้ในขั้นแรกและมีความคล้ายคลึงกันกับหลาย ๆ เสียงมาสอนให้กับโครงข่ายประสาทเทียม พบว่าโครงข่ายประสาทเทียมมีความสามารถในการจดจำได้ต่ำมากและเมื่อทำการเปลี่ยนข้อมูลให้มีความแตกต่างกันในแต่ละเสียง ทำให้โครงข่ายมีความสามารถในการจดจำได้เร็วขึ้นและมากขึ้น ดังนั้นการออกแบบส่วนฟรี-โพรเซสจึงจึงมีความสำคัญต่อโครงข่ายอย่างมาก โดยควรสามารถดึงเอาลักษณะที่แตกต่างกันของแต่ละเสียงออกมาให้ได้มากที่สุด ซึ่งในโครงข่ายนี้เลือกใช้คุณสมบัติเสียงด้านค่าสหสัมพันธ์ของข้อมูลในโดเมนความถี่ เป็นข้อมูลที่จะนำเข้าโครงข่ายประสาทเทียม

8.3.2 การเพิ่มขนาดชั้นซ่อน (Hidden Layer) จาก 2 หน่วย จนมีขนาด 9 หน่วย (อินพุทเลเยอร์มีขนาด 10 หน่วย) มีผลการทดลองดังกราฟที่ 8.1 (ภาคผนวก) แสดงการเปลี่ยนแปลงของค่า sum square error ที่จำนวน ฮิดเดน โหนด ต่าง ๆ โดยการเทรนจะใช้ค่า โมเมนตัม=0.3, Learning Rate = 0.7 จำนวนรอบ 7500 รอบ

พบว่าเมื่อมีการเพิ่มโหนดในชั้นซ่อนขึ้นจะทำให้ค่า error ที่เกิดขึ้นค่อย ๆ ลดลงจนถึงช่วงที่มีโหนดในชั้นซ่อนประมาณ 6-8 โหนดก็คือจุดที่มีค่า error น้อยที่สุดแล้วค่า error จะค่อย ๆ เพิ่มขึ้นอีกครั้งเราจึงสามารถสรุปได้ว่าการเพิ่มขนาดของโหนดในชั้นฮิดเดน จะทำให้ความเร็วในการรู้จำเพิ่มขึ้นในช่วงแรก ๆ เท่านั้น เมื่อเลยจุดที่ดีที่สุดของปัญหาไปแล้วความเร็วในการรู้จำก็จะเพิ่มขึ้นเหมือนเดิมซึ่งเป็นเพราะว่าขนาดของชั้นซ่อนเกินพอสำหรับตัดสินใจปัญหาที่ความยากระดับนั้น ๆ แล้วและการเพิ่มขนาดจะทำให้ต้องคำนวณปรับค่าจำนวนมากเกินความจำเป็น ที่ปัญหาต่างระดับกันจะทำให้มีขนาดชั้นซ่อนที่เหมาะสมต่างกัน

ขนาดชั้นซ่อน	ค่า min sum square error
2	0.268589
3	0.17
4	0.074680
5	0.051550
6	0.006983
7	0.015268
8	0.005098
9	0.015

ตารางที่ 8.2 แสดงการเปลี่ยนแปลงของค่า minimum sum square error ที่ขนาดชั้นซ่อนต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8.3.3 จากการปรับค่า Learning Rate และค่าโมเมนตัม จะพบว่า มีผลต่อการรู้จำ เป็นอย่างมากดังผลจากการทดลองกราฟ 8.2 (ภาคผนวก)จะพบว่า

8.3.3.1 ที่ค่า M ต่ำ ๆ และค่า LR ต่ำ ๆ จะทำให้การเรียนรู้ของโครงข่ายประสาทเทียมช้ามากแต่มีข้อดีคือทำให้โครงข่ายมีเสถียรภาพคือไม่เกิดการขึ้นลงของค่า error

8.3.3.2 ที่ค่า M ต่ำ ๆ และค่า LR สูง ๆ จะทำให้การเรียนรู้ของโครงข่ายประสาทเร็วขึ้นมากและสามารถได้ค่า Error ที่ค่อนข้างต่ำ แต่มีข้อเสียคือโครงข่ายไม่มีเสถียรภาพคือเกิดการขึ้นลงของค่า error จะทำให้เกิดการผิดพลาดในการรู้จำได้ซึ่งถึงแม้ว่า ค่า error จะต่ำแล้วก็ตาม

8.3.3.3 ที่ค่า M สูง ๆ และค่า LR ต่ำ ๆ จะพบว่าการเรียนรู้ของโครงข่ายค่อนข้างเร็วแต่เกิดการไม่มีเสถียรภาพของโครงข่ายค่อนข้างสูง

8.3.3.4 ที่ค่า M สูง ๆ และค่า LR สูง ๆ จะพบว่าเกิดการไม่มีเสถียรภาพของโครงข่ายสูงมากและมีความเร็วในการเรียนรู้สูงมาก

ดังนั้นจากการทดลองจึงได้จุดที่เหมาะสมคือค่าโมเมนตัมเท่ากับ 0.3 และค่า Learning Rate เท่ากับ 0.7 ที่ได้ค่า error ต่ำที่สุดและมีเสถียรภาพพอประมาณ ค่าโมเมนตัมและค่า Learning Rate จะมีค่าเท่าใดนั้นจะขึ้นอยู่กับลักษณะปัญหา

2.4 การเลือกค่าทาร์เก็ต(target)สำหรับการสอนโครงข่ายก็มีผลทำให้โครงข่ายไม่สามารถจำได้คือการทดลองคือ

ตัวเลข	ค่าทาร์เก็ตครั้งที่1	ค่าทาร์เก็ตครั้งที่2
0	1111	0000
1	0001	0001
2	0010	0010
3	0011	0011
4	0100	0100
5	0101	0101
6	0110	0110
7	0111	0111
8	1000	1000
9	1001	1001

ตารางที่ 8.3 แสดงการเปลี่ยนค่าทาร์เก็ตของแต่ละ output

จากตารางจะสังเกตเห็นว่าค่าทาร์เก็ตของศูนย์มีค่าเป็น 1111 ในครั้งแรกซึ่งเมื่อนำไปทำการเทรนจะทำให้โครงข่ายไม่สามารถจำเลขศูนย์ได้โดยจะจำเป็นเลขเจ็ดแทนทั้งหมด และเมื่อทำการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เปลี่ยนค่าทาร์เก็ตของศูนย์เป็น 0000 แทนจะทำให้โครงข่ายสามารถจดจำเลขศูนย์ได้ทั้งหมดและสามารถจำได้ดีด้วย ซึ่งเนื่องจากอาจเป็นเพราะการกำหนดให้ทาร์เก็ตของ ศูนย์เป็น 1111 จะทำให้โครงข่ายต้องมองเห็นว่าศูนย์เหมือนกับเลขทุก ๆ ตัวทำให้การตัดสินใจของศูนย์ผิดพลาดไป

8.3.4 ค่าเวลาที่สุ่มขึ้นมาครั้งแรกมีผลต่อการเรียนรู้ของโครงข่ายเป็นอย่างมาก โดยจากทดลองพบว่า ถ้าค่าเวทครั้งแรกมีค่าเป็นศูนย์หรือมีค่าเท่ากันทั้งหมดจะทำให้การเรียนรู้ช้ามากและไม่สามารถตัดสินใจได้ถูกต้องดังนั้นค่าเวทเริ่มต้นที่ดีจะต้องเป็นค่าน้อยๆ อยู่ในช่วง-0.5-0.5 และเป็นค่าที่แตกต่างกันจึงจะทำให้การเรียนรู้ของโครงข่ายเป็นไปอย่างรวดเร็วถ้าค่าเวทครั้งแรกมีความเหมาะสมมากเท่าใดจะยิ่งทำให้โครงข่ายเรียนรู้ได้เร็วและดีขึ้นเท่านั้น และถ้าค่าเวทครั้งแรกไม่เหมาะสมแล้วไม่ว่าจะให้เรียนรู้มากเท่าใดโครงข่ายก็ไม่สามารถจำได้ดี



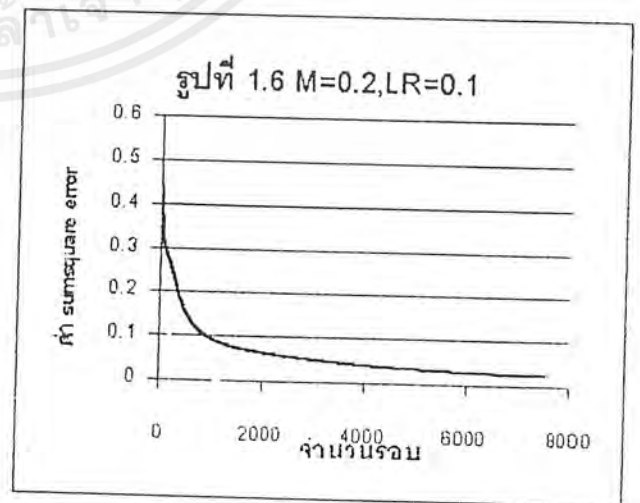
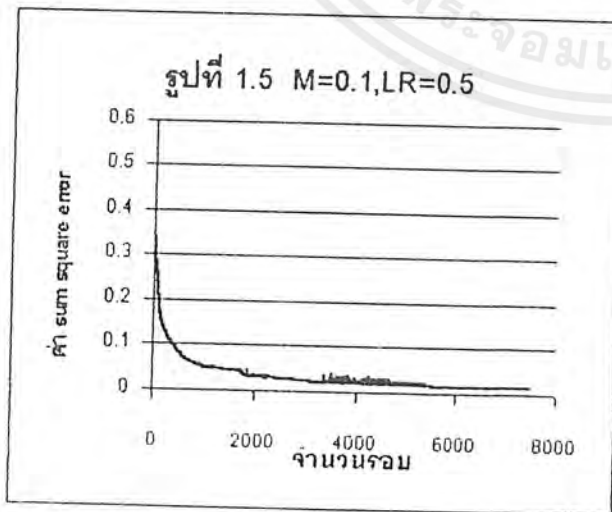
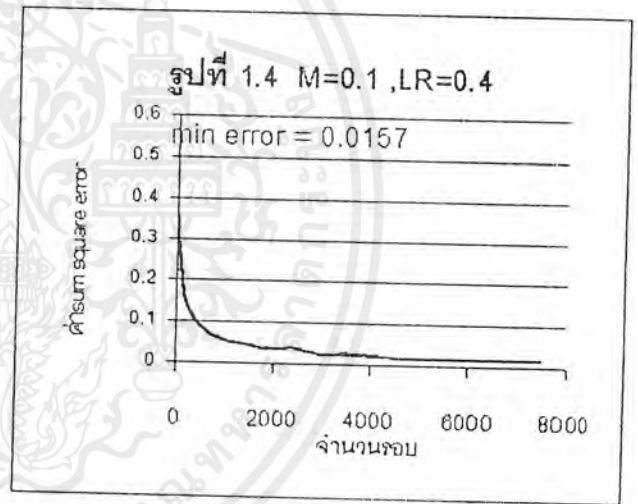
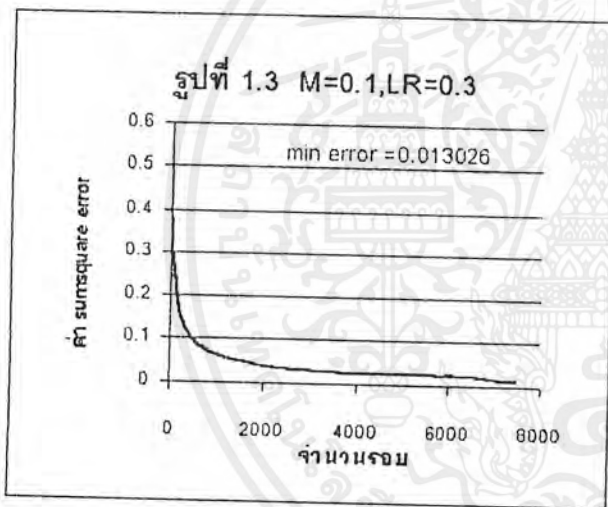
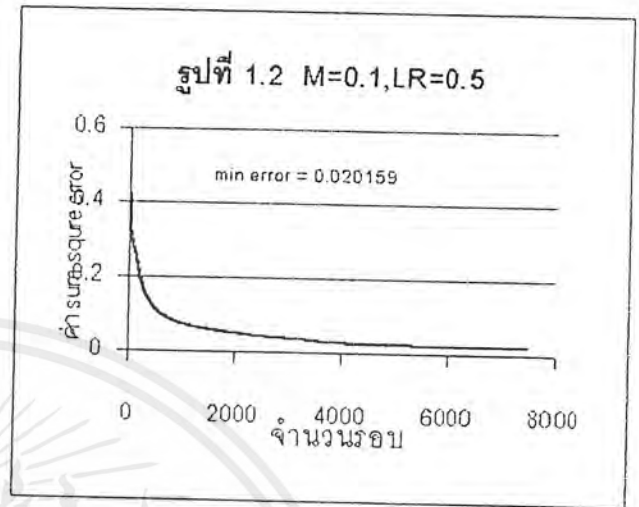
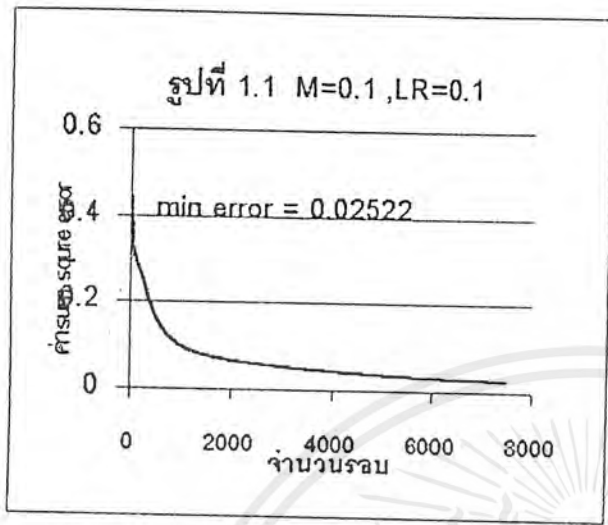
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



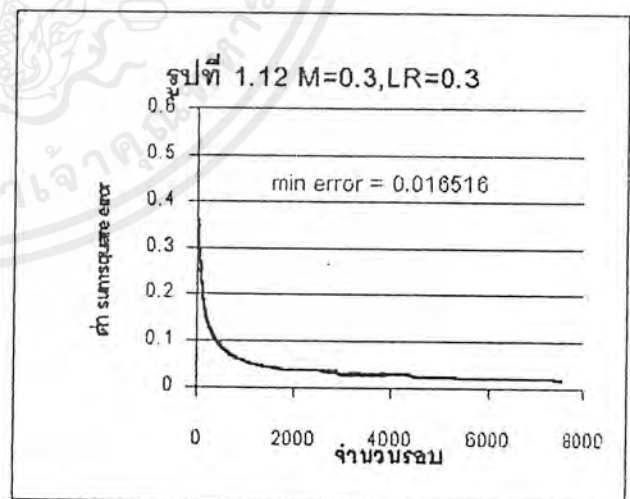
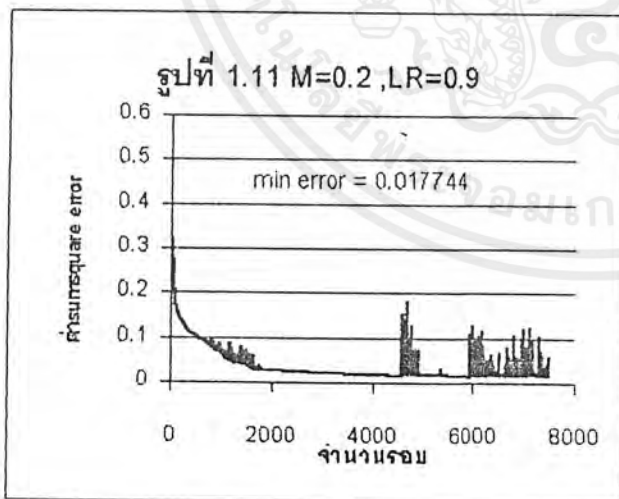
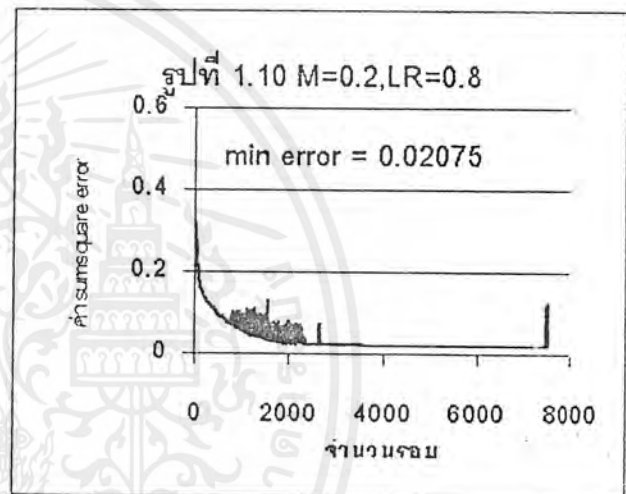
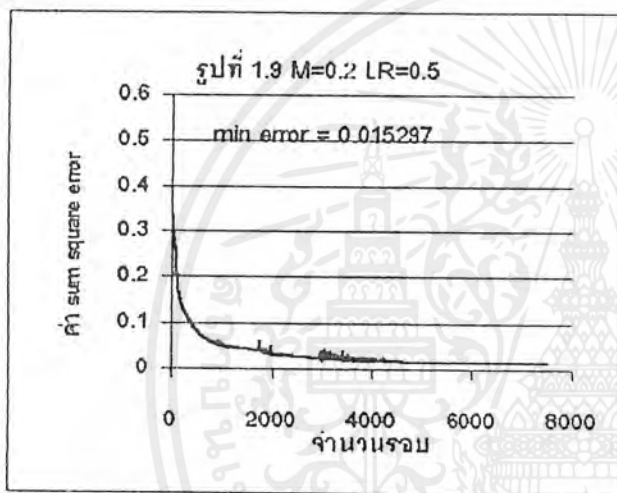
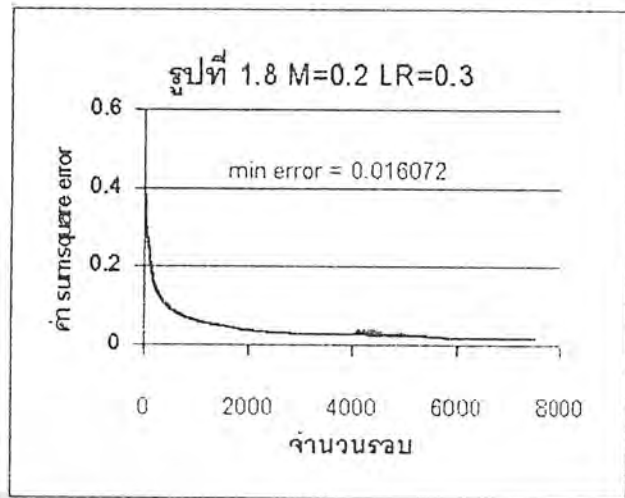
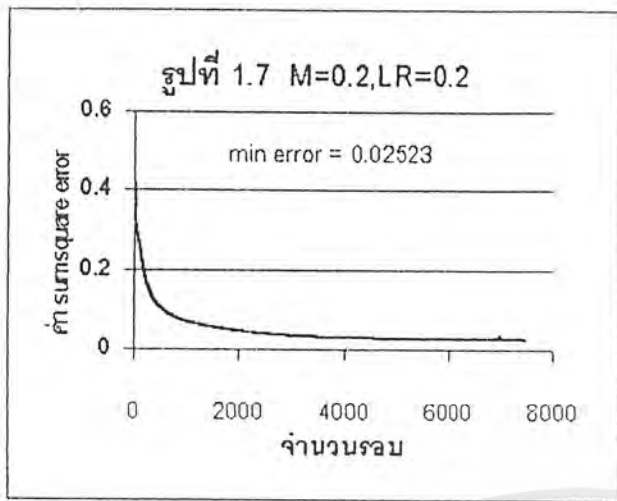
ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

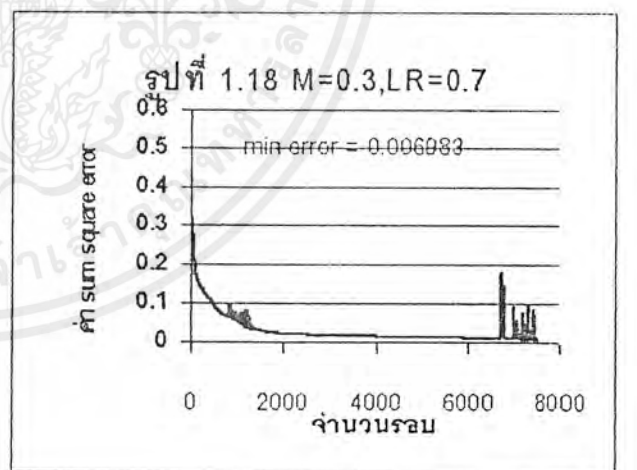
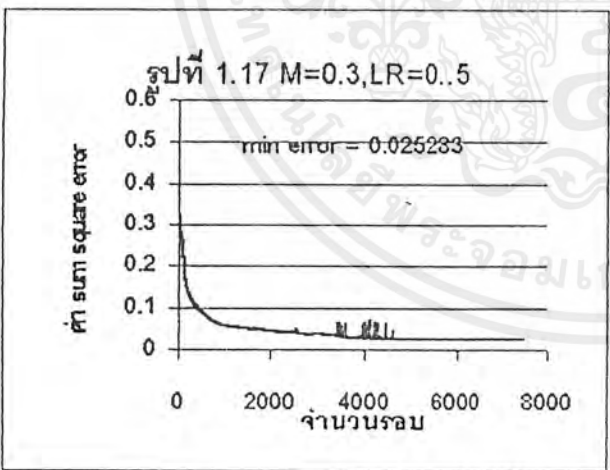
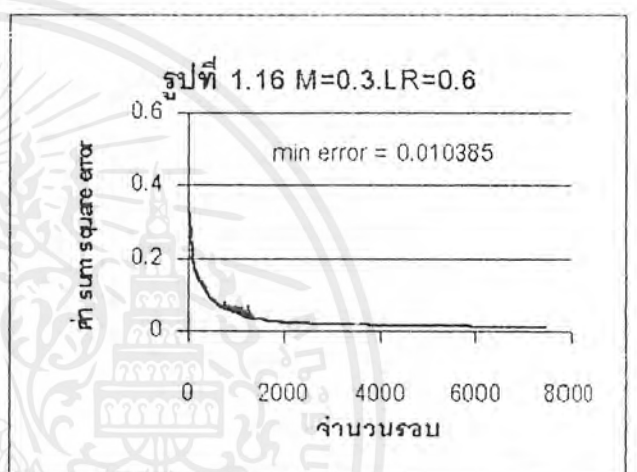
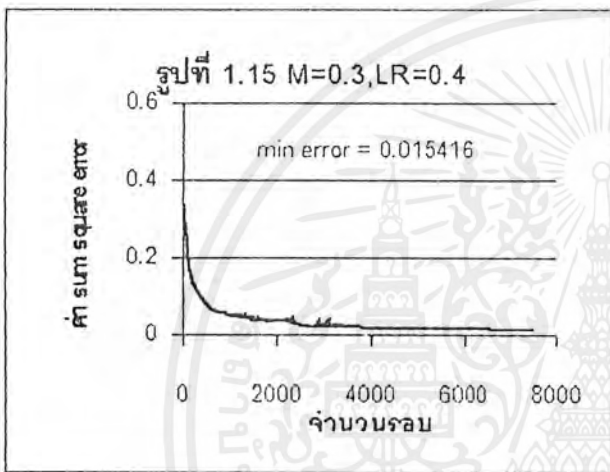
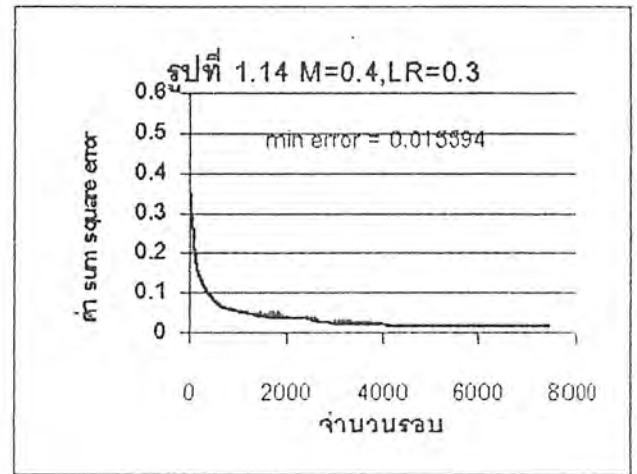
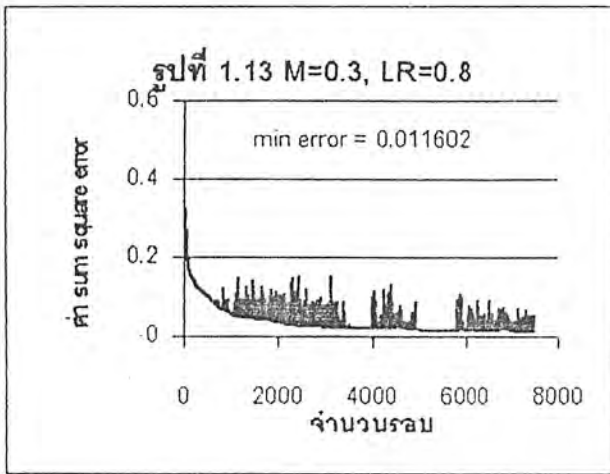
กราฟที่ 1 แสดงค่า sumsquare error ที่ค่าโมเมนต์ และ Learning Rate ที่ค่าต่าง ๆ โดยมีจำนวนรอบคงที่คือ 7500 รอบ



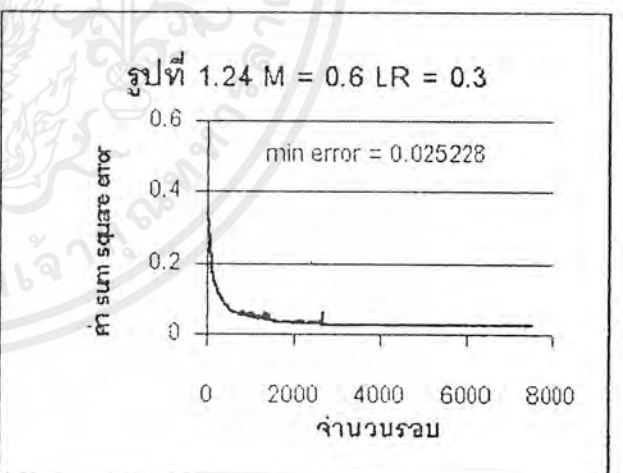
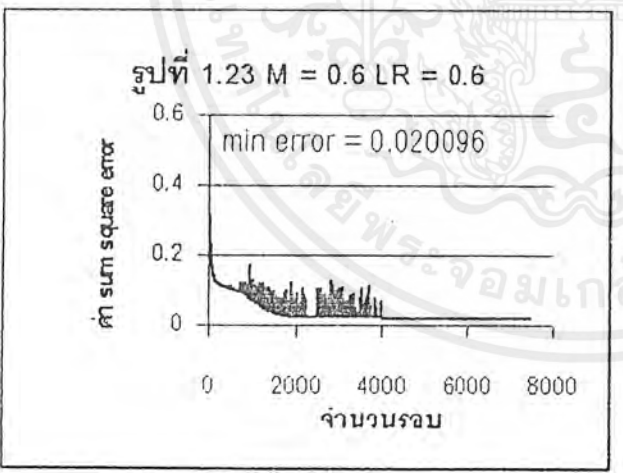
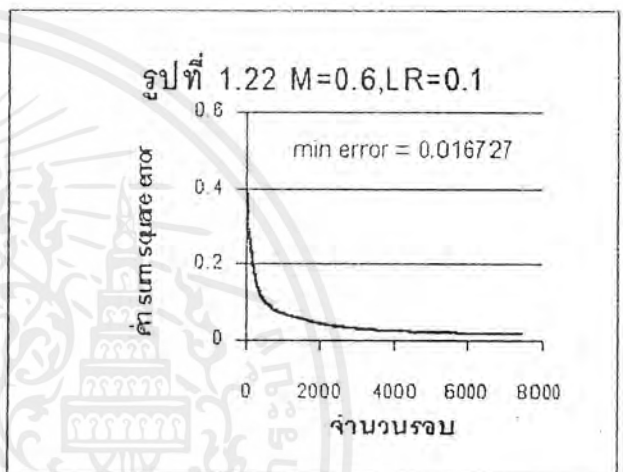
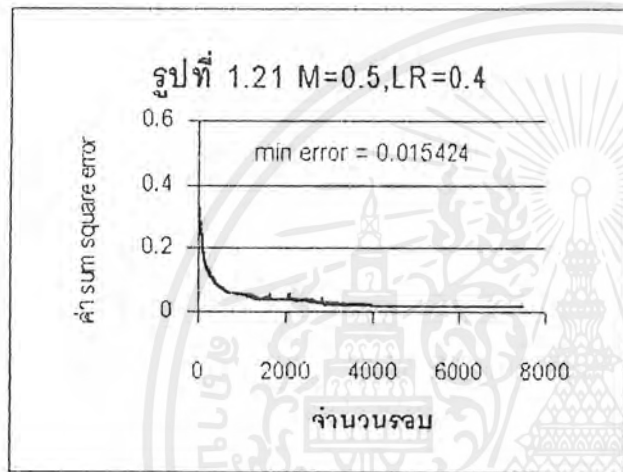
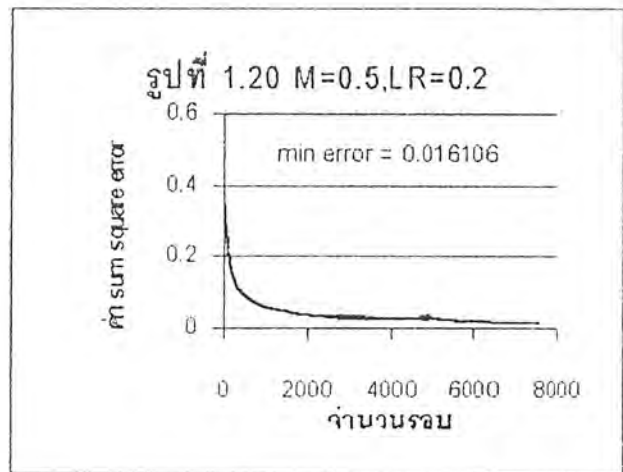
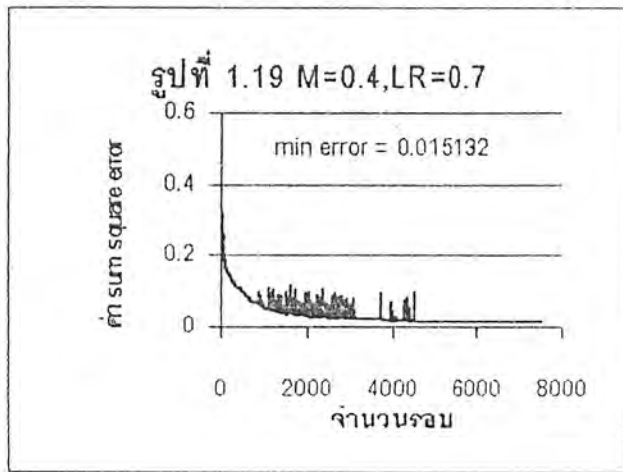
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



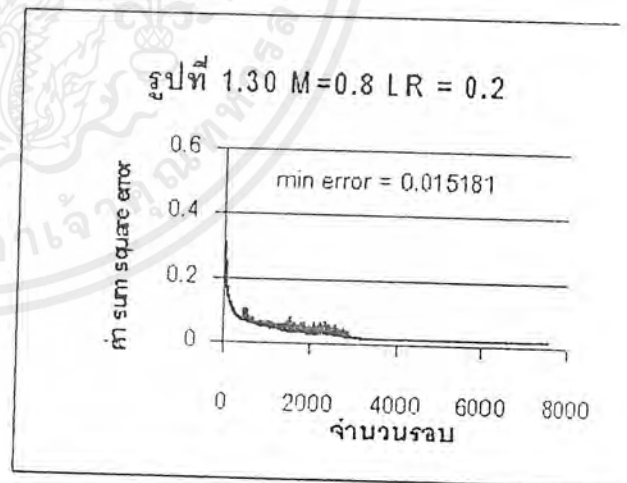
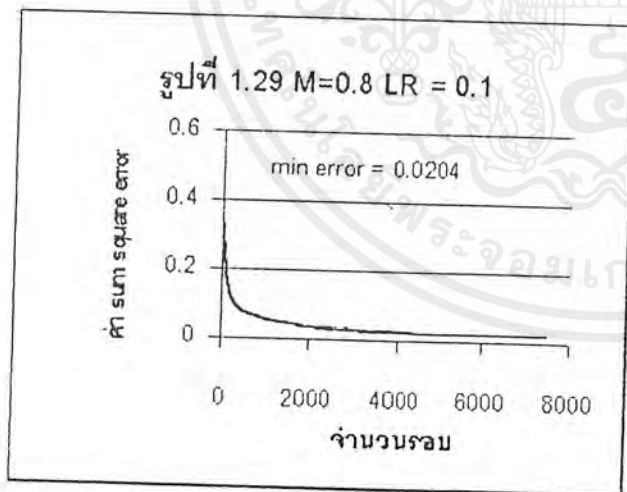
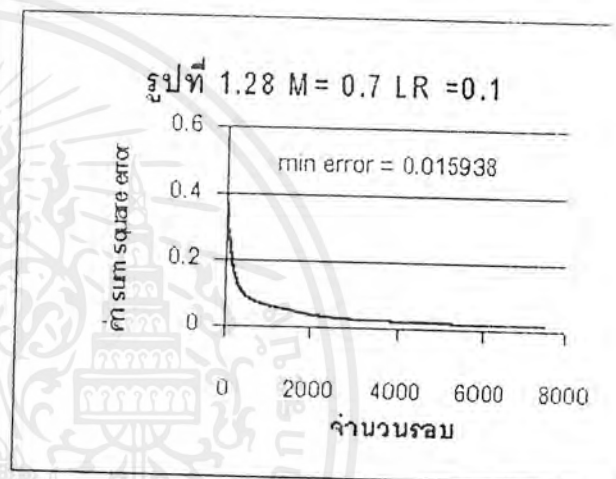
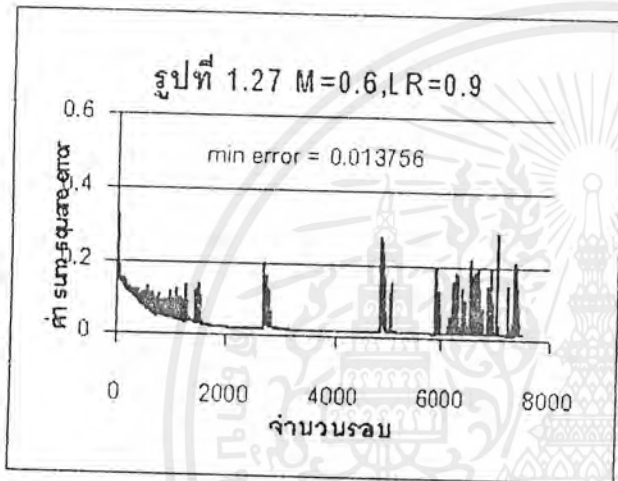
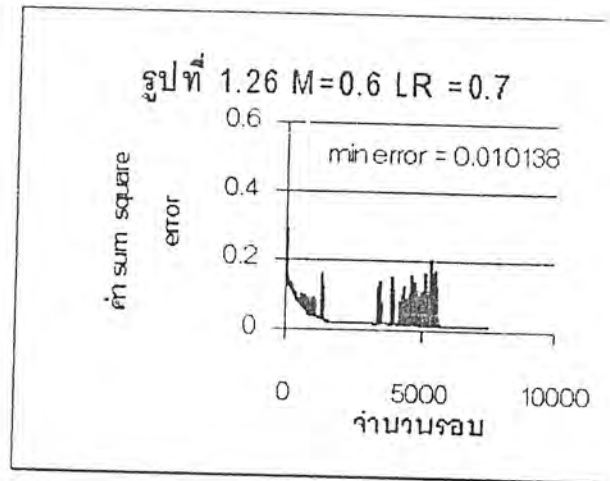
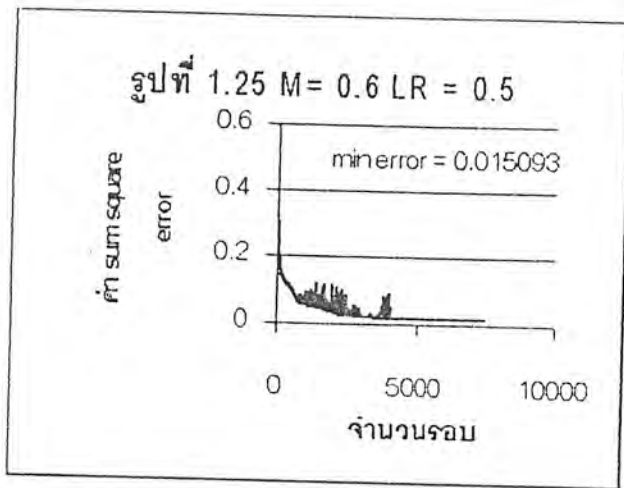
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



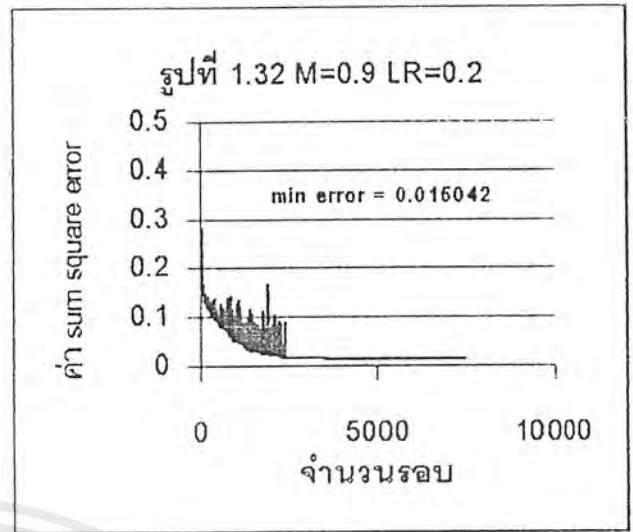
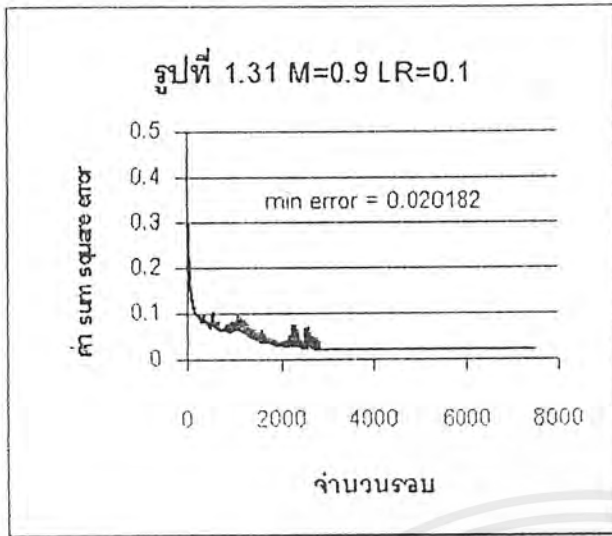
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

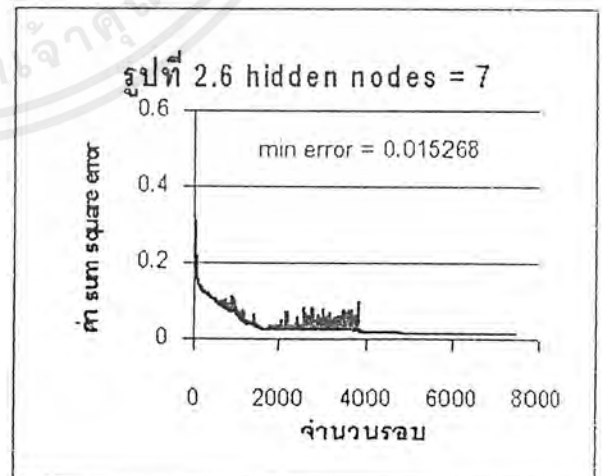
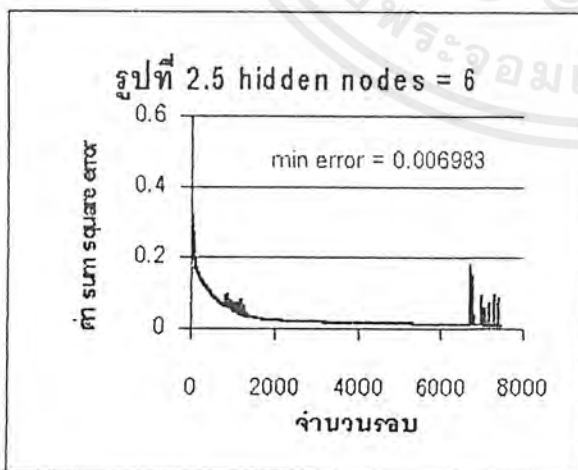
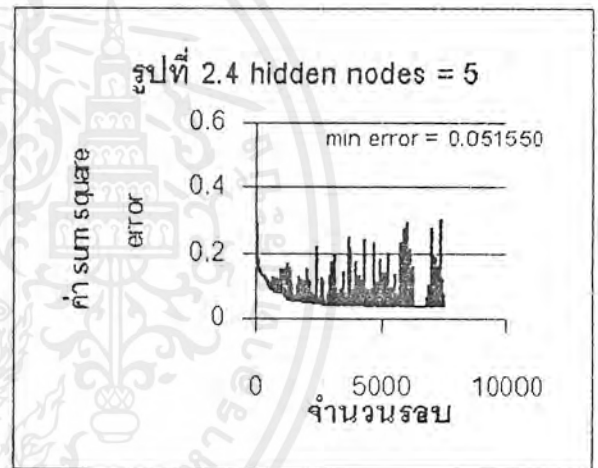
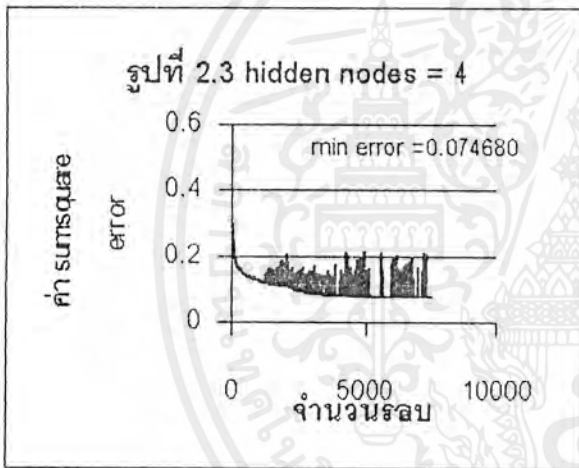
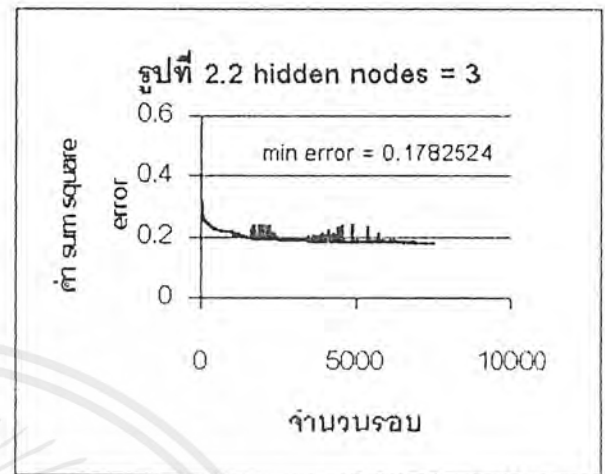
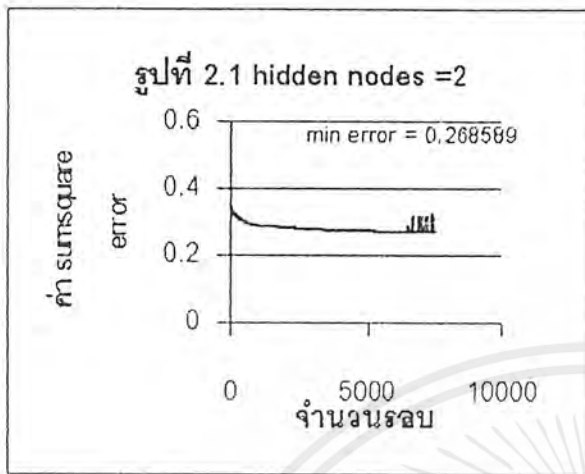


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

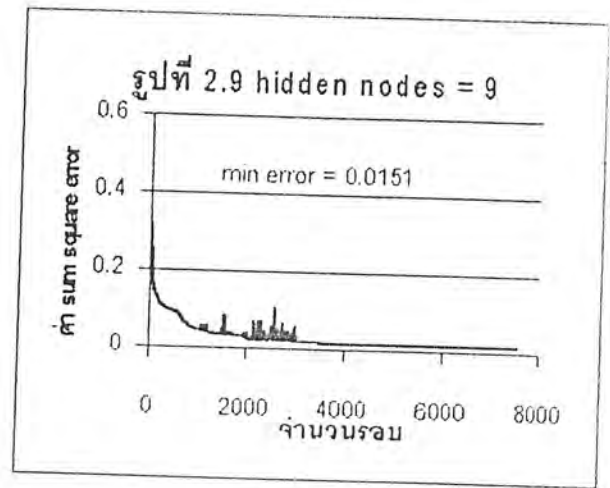
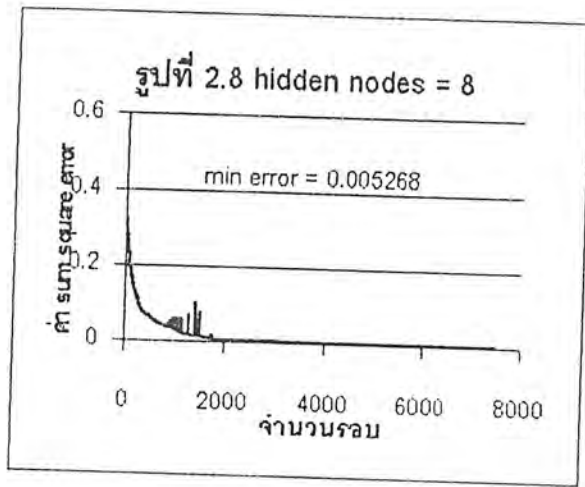


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กราฟที่ 2 แสดงการเปลี่ยนแปลงของค่า sum square error ที่จำนวน hidden nodes ต่าง ๆ โดยมีค่า $M=0.3, LR=0.7$ และจำนวนรอบ 7500 รอบ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงที่ใช้งานทั้งหมด

	Sound 0	Sound 1	Sound 2	Sound 3	Sound 4	Sound 5	Sound 6	Sound 7	Sound 8	Sound 9
Sound 0_1	0.781753	0.467697	0.72992	0.580533	0.573144	0.348733	0.722636	0.597919	0.474445	0.4619
Sound 0_2	0.740985	0.552079	0.694843	0.577913	0.554056	0.381032	0.763277	0.641748	0.52043	0.496204
Sound 0_3	0.72679	0.398584	0.713449	0.546013	0.518452	0.266616	0.685844	0.528026	0.407139	0.438017
Sound 0_4	0.739687	0.468433	0.678144	0.574325	0.474938	0.323015	0.738832	0.597241	0.4	0.439451
Sound 0_5	0.692778	0.491794	0.661538	0.570763	0.492576	0.350301	0.677074	0.597169	0.42788	0.41834
Sound 0_6	0.759522	0.57128	0.739203	0.585091	0.52601	0.340344	0.714957	0.652988	0.528897	0.449342
Sound 0_7	0.77017	0.475717	0.727822	0.562223	0.537215	0.348393	0.735086	0.600612	0.462867	0.482365
Sound 0_8	0.749928	0.442902	0.683919	0.537876	0.476106	0.3346	0.699581	0.699581	0.699581	0.699581
Sound 0_9	0.714862	0.478516	0.688786	0.587786	0.450001	0.372762	0.735405	0.612804	0.459401	0.463301

ตารางที่ 1.1 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงที่คู่มือ

Sound 1_1	0.270133	0.431818	0.231053	0.188066	0.282257	0.431682	0.208964	0.413701	0.326185	0.157244
Sound 1_2	0.541861	0.427919	0.608358	0.497902	0.36109	0.333007	0.662235	0.494592	0.412455	0.463538
Sound 1_3	0.24768	0.379424	0.204902	0.152338	0.270445	0.125707	0.294875	0.344633	0.14975	0.086069
Sound 1_4	0.423477	0.791469	0.339769	0.264555	0.442269	0.277566	0.360617	0.609525	0.484314	0.256617
Sound 1_5	0.274314	0.60039	0.182645	0.099952	0.375052	0.120071	0.223679	0.431112	0.288206	0.078801
Sound 1_6	0.462999	0.582547	0.330704	0.251341	0.386056	0.176601	0.449511	0.515762	0.258199	0.201645
Sound 1_7	0.176212	0.501779	0.143409	0.113306	0.25036	0.382775	0.103388	0.419395	0.332494	0.185389
Sound 1_8	0.535892	0.780238	0.489954	0.380974	0.502659	0.368975	0.482258	0.683912	0.550891	0.357965
Sound 1_9	0.318875	0.636177	0.288127	0.259071	0.365894	0.39421	0.257237	0.549424	0.464976	0.289662

ตารางที่ 1.2 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงหนึ่ง

Sound 2_1	0.580502	0.391568	0.65856	0.588546	0.316162	0.399201	0.683566	0.499033	0.494719	0.485642
Sound 2_2	0.542051	0.387352	0.629336	0.491196	0.451436	0.304928	0.594679	0.505247	0.477809	0.403198
Sound 2_3	0.540675	0.322128	0.610551	0.50747	0.335541	0.279975	0.625263	0.473751	0.366654	0.371728
Sound 2_4	0.733826	0.4617	0.802856	0.664513	0.490518	0.399495	0.722248	0.601621	0.544417	0.480211
Sound 2_5	0.654853	0.415088	0.728442	0.615082	0.416723	0.332815	0.71353	0.565705	0.456029	0.445171
Sound 2_6	0.645936	0.425786	0.744563	0.641033	0.444461	0.347336	0.652929	0.548271	0.484035	0.408888
Sound 2_7	0.676561	0.375371	0.743375	0.603518	0.414835	0.299136	0.655613	0.496886	0.44931	0.40136
Sound 2_8	0.57356	0.389412	0.657407	0.581014	0.462365	0.385237	0.647744	0.540344	0.453444	0.445171
Sound 2_9	0.672319	0.401252	0.759488	0.610991	0.530228	0.379452	0.658201	0.568086	0.575387	0.466778

ตารางที่ 1.3 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงสอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sound 3_1	0.445068	0.284694	0.525229	0.562274	0.30407	0.353339	0.466392	0.417097	0.554967	0.461881
Sound 3_2	0.556915	0.32987	0.578811	0.661013	0.386836	0.419033	0.563254	0.461367	0.533732	0.491767
Sound 3_3	0.39209	0.238546	0.465255	0.55949	0.187603	0.31547	0.454288	0.338331	0.441705	0.406284
Sound 3_4	0.549997	0.338639	0.616119	0.662687	0.297182	0.408289	0.604886	0.493008	0.518821	0.492764
Sound 3_5	0.361027	0.395026	0.411166	0.514123	0.275738	0.364615	0.368615	0.406679	0.569425	0.424567
Sound 3_6	0.604735	0.441641	0.649922	0.745531	0.340112	0.42688	0.588736	0.51297	0.601985	0.48663
Sound 3_7	0.651371	0.422773	0.682522	0.770481	0.352838	0.405062	0.638638	0.53224	0.570519	0.492859
Sound 3_8	0.645041	0.39918	0.72445	0.739357	0.37243	0.424244	0.639247	0.503775	0.61349	0.52372
Sound 3_9	0.617882	0.404211	0.666068	0.673143	0.406553	0.403858	0.593637	0.510497	0.622065	0.483258

ตารางที่ 1.4 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงสาม

Sound 4_1	0.32704	0.395263	0.266677	0.156382	0.544005	0.102542	0.272917	0.390311	0.158957	0.172857
Sound 4_2	0.385929	0.425157	0.36592	0.200019	0.633611	0.172744	0.308279	0.423126	0.316496	0.142223
Sound 4_3	0.407277	0.485136	0.4051	0.282932	0.600341	0.155117	0.323707	0.491024	0.344382	0.165603
Sound 4_4	0.567738	0.486865	0.520887	0.389174	0.601456	0.239986	0.557983	0.556625	0.32657	0.312177
Sound 4_5	0.433935	0.423916	0.381299	0.297467	0.440947	0.223306	0.426754	0.457023	0.264021	0.25305
Sound 4_6	0.633728	0.541704	0.603768	0.453147	0.592055	0.299145	0.619448	0.616914	0.40174	0.359487
Sound 4_7	0.645534	0.516419	0.602179	0.482207	0.625828	0.337624	0.565744	0.613524	0.510231	0.366735
Sound 4_8	0.676137	0.603187	0.648656	0.497546	0.674968	0.350554	0.593981	0.663619	0.58462	0.454723
Sound 4_9	0.512081	0.539396	0.458916	0.377796	0.577784	0.289187	0.512736	0.577597	0.403512	0.30635

ตารางที่ 1.5 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงสี่

Sound 5_1	0.380229	0.431995	0.355576	0.316578	0.305241	0.327937	0.301907	0.43455	0.413939	0.219038
Sound 5_2	0.203419	0.237441	0.265226	0.336991	0.045504	0.531059	0.253161	0.264416	0.378874	0.292501
Sound 5_3	0.169334	0.347066	0.235843	0.307022	0.124532	0.349509	0.186343	0.329414	0.447711	0.468294
Sound 5_4	0.323494	0.360152	0.369157	0.398816	0.202015	0.420458	0.354972	0.362809	0.453093	0.4054
Sound 5_5	0.262587	0.280023	0.294413	0.342592	0.178511	0.522091	0.277161	0.305184	0.389366	0.228013
Sound 5_6	0.368057	0.449548	0.411559	0.431956	0.223862	0.436845	0.396736	0.444723	0.519835	0.565031
Sound 5_7	0.55461	0.489771	0.541791	0.528186	0.382505	0.442784	0.530551	0.503139	0.609292	0.485584
Sound 5_8	0.242642	0.373464	0.284224	0.342057	0.105732	0.575586	0.257537	0.359984	0.450367	0.330113
Sound 5_9	0.17464	0.260062	0.248095	0.293246	0.105863	0.640113	0.224795	0.255374	0.381241	0.297977

ตารางที่ 1.6 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงห้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sound 6_1	0.334382	0.28788	0.377787	0.316838	0.287232	0.199643	0.399451	0.356155	0.296182	0.228218
Sound 6_2	0.660709	0.502332	0.644129	0.522458	0.457326	0.300821	0.814589	0.519316	0.413031	0.444662
Sound 6_3	0.715785	0.461736	0.702265	0.580846	0.530044	0.299597	0.787788	0.654133	0.435749	0.43261
Sound 6_4	0.721717	0.618845	0.733914	0.593779	0.531951	0.422688	0.814621	0.657969	0.56411	0.523805
Sound 6_5	0.730324	0.536531	0.705815	0.593047	0.534612	0.361669	0.820856	0.62593	0.458437	0.507328
Sound 6_6	0.73057	0.470941	0.706727	0.57234	0.460535	0.340695	0.856475	0.577906	0.410433	0.487014
Sound 6_7	0.771231	0.500594	0.758475	0.600921	0.533198	0.373271	0.855418	0.655256	0.448102	0.498869
Sound 6_8	0.749433	0.468011	0.720377	0.606837	0.470945	0.356717	0.865048	0.613531	0.436646	0.506168
Sound 6_9	0.711564	0.439532	0.715311	0.564421	0.457201	0.336553	0.840477	0.510455	0.412081	0.505516

ตารางที่ 1.7 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงนก

Sound 7_1	0.488255	0.591195	0.453028	0.31901	0.562562	0.232771	0.430858	0.719427	0.341628	0.211433
Sound 7_2	0.437926	0.674755	0.41344	0.322518	0.46224	0.25053	0.444734	0.647239	0.382678	0.244807
Sound 7_3	0.582662	0.615057	0.578198	0.434089	0.551288	0.295183	0.573381	0.73386	0.443459	0.315253
Sound 7_4	0.56306	0.679236	0.50203	0.404812	0.54938	0.308946	0.47672	0.750973	0.471119	0.299988
Sound 7_5	0.540474	0.631483	0.528771	0.441329	0.48729	0.378143	0.541892	0.698928	0.474624	0.359329
Sound 7_6	0.683913	0.681556	0.636788	0.559895	0.567256	0.412724	0.632857	0.806427	0.573116	0.417382
Sound 7_7	0.622305	0.692051	0.576234	0.477661	0.566345	0.349126	0.560791	0.783721	0.508812	0.343422
Sound 7_8	0.654115	0.709586	0.57235	0.456504	0.595332	0.347035	0.632382	0.770543	0.46833	0.397674
Sound 7_9	0.566085	0.623078	0.508888	0.395315	0.480888	0.346351	0.60543	0.661888	0.354857	0.290338

ตารางที่ 1.8 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงเจ็ด

Sound 8_1	0.479287	0.556222	0.527661	0.577623	0.327343	0.489276	0.488776	0.583697	0.724819	0.540981
Sound 8_2	0.289552	0.426197	0.360672	0.430292	0.257048	0.396262	0.333611	0.411256	0.605385	0.467482
Sound 8_3	0.497081	0.473702	0.63923	0.672793	0.290372	0.47107	0.466523	0.507871	0.774415	0.510516
Sound 8_4	0.451259	0.495693	0.578653	0.578329	0.287158	0.45816	0.403594	0.513152	0.794534	0.487171
Sound 8_5	0.491797	0.483682	0.60224	0.577557	0.370397	0.437355	0.420455	0.536292	0.791471	0.468708
Sound 8_6	0.449521	0.538119	0.589955	0.568785	0.325276	0.461983	0.397063	0.489815	0.787348	0.461995
Sound 8_7	0.443609	0.458757	0.6044	0.584195	0.261267	0.439823	0.414298	0.43706	0.762324	0.495549
Sound 8_8	0.517895	0.448219	0.652307	0.633093	0.292192	0.44688	0.512721	0.457727	0.760548	0.537562
Sound 8_9	0.349156	0.459593	0.406779	0.416263	0.365023	0.389659	0.279285	0.436098	0.736891	0.42141

ตารางที่ 1.9 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงแปด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sound 9_1	0.349697	0.370207	0.404073	0.48373	0.161159	0.351906	0.429567	0.35284	0.520367	0.607433
Sound 9_2	0.264134	0.174711	0.324111	0.440585	0.0525	0.416134	0.366926	0.22732	0.395878	0.379115
Sound 9_3	0.438352	0.389641	0.500666	0.573001	0.214136	0.47379	0.484089	0.404477	0.598522	0.721585
Sound 9_4	0.367448	0.324611	0.424327	0.452239	0.187271	0.328547	0.445288	0.328904	0.476124	0.740727
Sound 9_5	0.307593	0.308231	0.34085	0.429509	0.120439	0.500274	0.382141	0.272902	0.442989	0.4079
Sound 9_6	0.304399	0.290772	0.359496	0.405409	0.136399	0.322142	0.415337	0.277521	0.459913	0.536888
Sound 9_7	0.48812	0.392043	0.549935	0.55207	0.319478	0.354414	0.550751	0.426831	0.524182	0.613105
Sound 9_8	0.382393	0.320905	0.425629	0.463194	0.160426	0.312146	0.470748	0.329978	0.454235	0.743993
Sound 9_9	0.375243	0.335386	0.42653	0.462171	0.171096	0.405748	0.445832	0.359221	0.475509	0.668977

ตารางที่ 1.10 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงเก้า

Wav 0_1	0.721018	0.49106	0.678587	0.565364	0.483288	0.371647	0.770877	0.61718	0.399419	0.488988
Wav 0_2	0.750588	0.528512	0.725626	0.583155	0.554693	0.368069	0.727753	0.66419	0.537455	0.473114
Wav 0_3	0.784586	0.564863	0.718641	0.596134	0.569428	0.405988	0.747181	0.674565	0.541541	0.47584
Wav 0_4	0.681409	0.506333	0.647809	0.561807	0.469178	0.334612	0.751159	0.59835	0.416936	0.524634
Wav 0_5	0.732704	0.469184	0.739992	0.595779	0.503436	0.329372	0.729322	0.602704	0.498541	0.459461
Wav 0_6	0.759685	0.47947	0.678656	0.558586	0.494867	0.315246	0.740444	0.601747	0.385146	0.451278
Wav 0_7	0.729008	0.521716	0.694469	0.627462	0.480011	0.393567	0.703127	0.649953	0.482129	0.450726
Wav 0_8	0.702157	0.525624	0.680651	0.593174	0.443922	0.360968	0.731802	0.622756	0.469919	0.45022
Wav 0_9	0.712178	0.424769	0.682067	0.536888	0.476907	0.276148	0.733815	0.554608	0.352431	0.380499

ตารางที่ 1.11 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงศูนย์

Wav 1_1	0.208039	0.858696	0.121726	0.050283	0.352765	0.0979	0.170838	0.40083	0.307636	0.087505
Wav 1_2	0.353466	0.721205	0.242504	0.204274	0.41353	0.225058	0.326527	0.552274	0.378091	0.16622
Wav 1_3	0.401452	0.622543	0.356599	0.316513	0.441289	0.316124	0.349916	0.614247	0.475635	0.370918
Wav 1_4	0.581179	0.712991	0.480208	0.375404	0.502071	0.327837	0.507325	0.652875	0.478768	0.348613
Wav 1_5	0.479956	0.456708	0.436669	0.296958	0.409081	0.240527	0.456659	0.474472	0.364497	0.281521
Wav 1_6	0.530405	0.651702	0.483465	0.371994	0.503081	0.324385	0.456497	0.68874	0.487993	0.326081
Wav 1_7	0.386142	0.800962	0.31488	0.237541	0.427448	0.243666	0.341903	0.576217	0.48018	0.246344
Wav 1_8	0.380042	0.754188	0.286988	0.221874	0.416916	0.219518	0.294187	0.571491	0.463485	0.210706
Wav 1_9	0.422563	0.748023	0.375884	0.273101	0.446073	0.277481	0.370405	0.609875	0.503497	0.285752

ตารางที่ 1.12 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wav 2_1	0.62517	0.372631	0.807146	0.620077	0.414962	0.363848	0.649391	0.553307	0.552669	0.462173
Wav 2_2	0.618191	0.379587	0.77097	0.64255	0.384611	0.365906	0.65434	0.52224	0.515531	0.450902
Wav 2_3	0.621338	0.422124	0.709998	0.591081	0.468145	0.395045	0.681039	0.578541	0.497778	0.504979
Wav 2_4	0.699277	0.380458	0.823804	0.63083	0.489922	0.376729	0.711035	0.551878	0.565103	0.510007
Wav 2_5	0.638403	0.413796	0.756038	0.657871	0.385732	0.39079	0.700439	0.592053	0.518546	0.45745
Wav 2_6	0.586652	0.316086	0.654557	0.515229	0.383049	0.282077	0.691042	0.471726	0.354121	0.381176
Wav 2_7	0.681671	0.340727	0.7316	0.615599	0.422526	0.314171	0.729927	0.541664	0.356618	0.420378
Wav 2_8	0.639683	0.396565	0.737101	0.642416	0.386804	0.382127	0.707393	0.572592	0.495057	0.471646
Wav 2_9	0.684708	0.428587	0.775896	0.667277	0.44358	0.391521	0.742427	0.58893	0.524368	0.493449

ตารางที่ 1.13 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงลง

Wav 3_1	0.59784	0.357558	0.721253	0.712542	0.352545	0.431498	0.60967	0.472926	0.662526	0.552085
Wav 3_2	0.504838	0.366043	0.601984	0.691288	0.205679	0.421146	0.526441	0.427771	0.616206	0.54305
Wav 3_3	0.641322	0.384146	0.710824	0.680391	0.382257	0.410779	0.629303	0.488323	0.63692	0.531434
Wav 3_4	0.500028	0.228709	0.668299	0.633851	0.291078	0.328775	0.559335	0.38543	0.537828	0.45417
Wav 3_5	0.578503	0.369876	0.641078	0.689586	0.264002	0.40981	0.637913	0.508621	0.555068	0.514809
Wav 3_6	0.577196	0.33782	0.635233	0.68646	0.282522	0.392456	0.639459	0.639459	0.639459	0.639459
Wav 3_7	0.52707	0.360465	0.62422	0.698175	0.23895	0.45876	0.577081	0.485415	0.598183	0.520059
Wav 3_8	0.520117	0.330563	0.595451	0.635372	0.235733	0.42023	0.588521	0.451413	0.556437	0.514936
Wav 3_9	0.545336	0.327101	0.64064	0.684101	0.252027	0.397418	0.624609	0.463364	0.527802	0.495459

ตารางที่ 1.14 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงตาม

Wav 4_1	0.366388	0.496132	0.307899	0.167793	0.614584	0.092854	0.287488	0.45026	0.253424	0.109383
Wav 4_2	0.447592	0.535953	0.420564	0.252571	0.639565	0.186616	0.31707	0.50441	0.469303	0.228139
Wav 4_3	0.570091	0.52722	0.522121	0.391273	0.616839	0.285816	0.525428	0.59071	0.410163	0.374802
Wav 4_4	0.440542	0.511558	0.410134	0.273126	0.552467	0.155988	0.460831	0.488322	0.318059	0.305943
Wav 4_5	0.436928	0.528331	0.3871	0.309479	0.60806	0.179798	0.396696	0.537009	0.337067	0.208095
Wav 4_6	0.623601	0.51767	0.583243	0.422953	0.655825	0.251611	0.588443	0.591518	0.350986	0.318765
Wav 4_7	0.330784	0.632113	0.287892	0.213857	0.561526	0.119524	0.30306	0.555923	0.381641	0.184678
Wav 4_8	0.35946	0.50345	0.355206	0.228584	0.687437	0.152101	0.264039	0.496262	0.386498	0.167882
Wav 4_9	0.549983	0.54931	0.488167	0.371984	0.680957	0.221398	0.48587	0.593667	0.379279	0.269958

ตารางที่ 1.15 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงลิ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wav 5_1	0.222309	0.255518	0.275458	0.285475	0.092308	0.523166	0.248507	0.290531	0.354087	0.278067
Wav 5_2	0.174986	0.214239	0.238647	0.300887	0.088387	0.543162	0.204723	0.254021	0.360306	0.196353
Wav 5_3	0.332936	0.342044	0.341774	0.370096	0.227135	0.329906	0.315501	0.361389	0.400781	0.242271
Wav 5_4	0.307997	0.219832	0.350637	0.385588	0.102451	0.428219	0.357454	0.294351	0.402982	0.323016
Wav 5_5	0.392979	0.513673	0.439654	0.512205	0.287466	0.424381	0.420192	0.511065	0.574448	0.460788
Wav 5_6	0.384144	0.410225	0.433116	0.460841	0.196172	0.590912	0.388628	0.43977	0.545009	0.366763
Wav 5_7	0.342831	0.438689	0.39759	0.4389	0.2443	0.349626	0.446457	0.416613	0.438017	0.396374
Wav 5_8	0.179034	0.265385	0.231485	0.258001	0.106721	0.259668	0.223999	0.260283	0.35302	0.350765
Wav 5_9	0.235799	0.358506	0.321945	0.358935	0.165717	0.351174	0.320438	0.316332	0.451019	0.482471

ตารางที่ 1.16 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงห้า

Wav 6_1	0.742449	0.519291	0.702217	0.589709	0.529889	0.358892	0.860173	0.652221	0.437654	0.48549
Wav 6_2	0.696206	0.404012	0.698879	0.558491	0.520935	0.315174	0.807407	0.603764	0.410572	0.447956
Wav 6_3	0.70557	0.418286	0.666368	0.522855	0.455814	0.308143	0.775858	0.589298	0.339429	0.407375
Wav 6_4	0.697644	0.430162	0.699634	0.608571	0.449384	0.365631	0.853967	0.54972	0.427096	0.540255
Wav 6_5	0.696493	0.488459	0.72214	0.619957	0.519142	0.366754	0.793951	0.623679	0.503689	0.495149
Wav 6_6	0.719971	0.4257	0.697841	0.520183	0.522708	0.283029	0.805136	0.605479	0.345272	0.418001
Wav 6_7	0.708291	0.363251	0.685573	0.554394	0.485818	0.318897	0.782006	0.610865	0.334314	0.40152
Wav 6_8	0.765024	0.496414	0.750206	0.605836	0.521773	0.412244	0.823089	0.674954	0.46753	0.492493
Wav 6_9	0.736343	0.469759	0.724795	0.611063	0.512283	0.357836	0.833888	0.623227	0.461159	0.513552

ตารางที่ 1.17 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงหก

Wav 7_1	0.595864	0.704794	0.565346	0.474889	0.58894	0.324895	0.554696	0.791764	0.497433	0.331344
Wav 7_2	0.43287	0.627269	0.406973	0.30406	0.506322	0.251887	0.456412	0.629925	0.370153	0.258082
Wav 7_3	0.528922	0.645512	0.484299	0.424778	0.503908	0.291822	0.495929	0.681071	0.449323	0.302267
Wav 7_4	0.642114	0.715933	0.62188	0.518059	0.532662	0.409078	0.643368	0.789541	0.534264	0.414406
Wav 7_5	0.597485	0.703669	0.552097	0.47364	0.564881	0.366378	0.523226	0.79173	0.549373	0.35772
Wav 7_6	0.639892	0.69987	0.592476	0.520117	0.573201	0.401774	0.587808	0.845772	0.55665	0.384558
Wav 7_7	0.589906	0.685643	0.553357	0.476176	0.568554	0.375514	0.518696	0.794776	0.575342	0.368885
Wav 7_8	0.592884	0.729955	0.576666	0.50029	0.589288	0.350101	0.547641	0.789741	0.586752	0.377957
Wav 7_9	0.56168	0.708761	0.532655	0.447109	0.574142	0.321547	0.511998	0.78273	0.531259	0.334325

ตารางที่ 1.18 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงเจ็ด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wav 8_1	0.405307	0.478858	0.56795	0.580469	0.284897	0.453531	0.381032	0.466002	0.823817	0.480401
Wav 8_2	0.418155	0.456776	0.548684	0.566496	0.302915	0.438572	0.365563	0.45567	0.816233	0.488236
Wav 8_3	0.492282	0.508816	0.627729	0.580353	0.356611	0.44977	0.460826	0.546024	0.812469	0.516408
Wav 8_4	0.526775	0.51901	0.58154	0.573657	0.39697	0.465019	0.532739	0.556454	0.731834	0.549828
Wav 8_5	0.539986	0.558723	0.635592	0.657514	0.335707	0.46002	0.516561	0.560527	0.78543	0.532379
Wav 8_6	0.547658	0.528855	0.645136	0.617984	0.403425	0.481536	0.534934	0.562458	0.795592	0.55313
Wav 8_7	0.581849	0.534896	0.674246	0.674	0.328643	0.496761	0.553744	0.551569	0.752443	0.531062
Wav 8_8	0.508039	0.533437	0.606708	0.597124	0.295899	0.455448	0.506496	0.565586	0.728559	0.524312
Wav 8_9	0.533889	0.522687	0.66395	0.652042	0.35041	0.472806	0.514726	0.554748	0.775826	0.520474

ตารางที่ 1.19 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงแปด

Wav 9_1	0.291768	0.264167	0.353097	0.45571	0.060416	0.437931	0.354284	0.272455	0.494133	0.42769
Wav 9_2	0.338867	0.267361	0.396181	0.465399	0.13675	0.411745	0.414381	0.313747	0.479029	0.468824
Wav 9_3	0.347711	0.247275	0.375664	0.430433	0.129256	0.357952	0.387765	0.265707	0.455387	0.428625
Wav 9_4	0.389387	0.285914	0.420033	0.4388	0.178523	0.375231	0.437284	0.330456	0.425224	0.444001
Wav 9_5	0.375613	0.383986	0.433159	0.499331	0.19996	0.506109	0.428419	0.405612	0.53127	0.543527
Wav 9_6	0.30834	0.283949	0.369459	0.439735	0.102868	0.52045	0.404958	0.298032	0.445178	0.448823
Wav 9_7	0.509242	0.419587	0.523536	0.518077	0.333842	0.356018	0.536938	0.396474	0.542989	0.533363
Wav 9_8	0.402851	0.408215	0.482518	0.550942	0.262749	0.318588	0.458613	0.382517	0.5196	0.528273
Wav 9_9	0.268589	0.246917	0.338274	0.424449	0.078534	0.356155	0.370993	0.265655	0.440263	0.553856

ตารางที่ 1.20 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเสียงเก้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#include<conio.h>
#include<time.h>
float Win[20][20],Wout[20][20],input[100][15],show[10000],error_buff;
float Wout_buff[20][20],Win_buff[20][20],LR1,LR2,Wout_buff1[20][20],Win_buff1[20][20];
float inputh[10],outputh[10],output[10],output1[10],invfn[10],invfn1[10],t;
int target[100][10],z;
float error[100],data,sum[20],moment1,moment2;
int number_input,number_output,number_inputh,num_loop;
int number_outputh,x,y,number_file,target1,a,b;
FILE *pointer,*pointer1,*pointer2,*pointer3,*pointer4,*pointer5,*file_pointer;
char str1[20],str2[20],str3[20],str4[20],str5[20],str6[20];
void buff_zero()
{
for (x=0;x<number_output;x++)
{
for (y=0;y<number_inputh;y++)
{
Wout_buff1[x][y]=0;
Wout_buff[x][y]=0;
}
}
for (x=0;x<number_outputh;x++)
{
for (y=0;y<number_input;y++)
{
Win_buff[x][y]=0;
Win_buff1[x][y]=0;
}
}
}
void prepare()
{
//////////prepare data//////////
printf("please enter file name:");
scanf("%s",&str1);
pointer2= fopen(str1,"r+t");
for (a=0;a<number_file;a++)
{
for(b=0;b<number_input;b++)
{
fscanf(pointer2,"%f",&input[a][b]);
printf("input [%d] [%d]=%f",a,b,input[a][b]);
/*fread(input,sizeof(input),1,pointer2);
if(ferror(pointer2)
{clrscr;printf("can not read data file");exit(1);}*/
}
}
//getch();
fclose(pointer2);
printf("\nplease enter target file:");
scanf("%s",&str2);
/*pointer3= fopen(str2,"w");
for (a=0;a<number_file;a++)
{
for(b=0;b<number_output;b++)
{ printf("file %d target no. (%d)=",a,b);
scanf("%d",&target_buff[a][b]);
fprintf(pointer3,"%d",target_buff[a][b]);
//fscanf(pointer3,"%d",&target[a][b]);
/*fread(target,sizeof(target),1,pointer3);
if(ferror(pointer3)
{clrscr;printf("can not read data file");exit(1);}*/
}
}

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    //printf("target [%d] [%d]=%d\n",a,b,target [a] [b] );
    }
}
fclose(pointer3);*/
pointer5= fopen(str2,"r");
for (a=0;a<number_file;a++)
{
    for(b=0;b<number_output;b++)
    {
        fscanf(pointer5,"%d\n",&target [a] [b] );
        printf("target [%d] [%d]=%d\n",a,b,target [a] [b] );
    }
}
fclose(pointer5);
getch();
}
void rando()
{
    file_pointer= fopen(str5,"r");
    for (x=0;x<number_inpuh;x++)
    {
        for (y=0;y<number_input;y++)
            fscanf(file_pointer,"%f ",&Win[x] [y] );
    }
    fclose(file_pointer);
    //load weight for second layer
    pointer=fopen(str6,"r");
    for (x=0;x<number_output;x++)
    {
        for (y=0;y<number_output;y++)
            fscanf(pointer,"%f ",&Wout [x] [y] );
    }
    fclose(pointer);
}
void multi1()
{
    //multiplier and summation for first layer
    for (x=0;x<number_inpuh;x++)
    {
        inpuh[x]=0;
        for(y=0;y<number_input;y++)
            inpuh [x] =inpuh [x] +Win [x] [y] *input [a] [y] ;
    }
}
void func1()
{
    //use function
    for (x=0;x<number_inpuh;x++)
        output[x] = (1 / (1 + exp (- (inpuh [x] ))));
}
void multi2()
{
    // multiplier and summation for second layer
    for (x=0;x<number_output;x++)
    {
        output1[x]=0;
        for(y=0;y<number_output;y++)
            output1 [x] =output1 [x] +Wout [x] [y] *output [y] ;
    }
}
void func2()
{
    //use function
    for (x=0;x<number_output;x++)
        output [x] = (1 / (1 + exp (- (output1 [x] ))));
}

```

เอกสารนี้เป็นเอกสารที่สถาบันใช้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

}
void ferrorf()
{
//find error all output
error[a]=0;
error_buff=0;
for (x=0;x<number_output;x++)
{
t=target [a] [x]-output [x];
error_buff=((target [a] [x]-output [x])*(target [a] [x]-output [x]))+error_buff;
}
error[a]=error_buff/2;
}

void new_weight()
{
//find new weight
// find new weight for second layer
for (x=0;x<number_output;x++)
{
invfn[x]=(target [a] [x]-output [x])*output [x]*(1-output [x]);
for (y=0;y<number_inpuh;y++)
{
Wout_buff[x] [y]=output[y]*LR2*invfn[x]+moment2*Wout_buff1[x] [y];
Wout_buff1[x] [y]=Wout_buff[x] [y];
}
}

//find summation of Spi
for (x=0;x<number_output;x++)
{
sum[x]=0;
for (y=0;y<number_output;y++)
{
sum[x]=sum[x]+(invfn[y]*Wout [y] [x]);
}
invfn1[x]=output[x]*(1-output[x])*sum[x];
}

for (x=0;x<number_output;x++)
{
for (y=0;y<number_inpuh;y++)
{
Wout [x] [y]=Wout [x] [y]+Wout_buff[x] [y];
}
}

//find new weight for first layer
//find Spi and new weight
for (x=0;x<number_outpuh;x++)
{
for (y=0;y<number_input;y++)
{
Win_buff[x] [y]=LR1*invfn1[x]*input [a] [y]+moment1*Win_buff1[x] [y];
Win_buff1[x] [y]=Win_buff[x] [y];
}
}

for (x=0;x<number_outpuh;x++)
{
for (y=0;y<number_input;y++)
{
Win [x] [y]=Win [x] [y]+Win_buff[x] [y];
}
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

}

////////////////////////////////////
void main()
{
clrscr();
printf("please enter number_input(integer):");
scanf("%d",&number_input);
printf("\nplease enter number_neuron(integer):");
scanf("%d",&number_inpuh);
number_outpuh=number_inpuh;
printf("\nplease enter number_output(integer):");
scanf("%d",&number_output);
/*number_input=10;
number_inpuh=number_outpuh=6;
number_output=4;*/
printf("\nplease enter number_file(integer):");
scanf("%d",&number_file);
printf("\nplease enter moment1:");
scanf("%f",&moment1);
printf("\nplease enter moment2:");
scanf("%f",&moment2);
printf("\nplease enter learning ratel:");
scanf("%f",&LR1);
printf("\nplease enter learning rate2:");
scanf("%f",&LR2);
printf("\nplease enter number of loop:");
scanf("%d",&num_loop);
printf("\nplease enter weightin name file:");
scanf("%s",&str5);
printf("\nplease enter weightout name file:");
scanf("%s",&str6);
prepare();
rando();
buff_zero();
// start program
z=0;
do
{ z=z+1;
for (a=0;a<number_file;a++)
{
multi1();
func1();
multi2();
func2();
ferrorf();
new_weight();
}
}
////end for loop/////
////check error for loop/////
show[z-1]=0;
for (a=0;a<number_file;a++)
{
show[z-1]=error[a]+show[z-1];
}
show[z-1]=show[z-1]/number_file;
clrscr();
printf("show[z-1]=%f\n",show[z-1]);
if (show[z-1]<0.0001){break;}
}
while (z<num_loop); ////check
//end program

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

// save weightin to file
printf("\nplease enter weightin name file:");
scanf("%s",&str3);
file_pointer=fopen(str3,"w");
for (x=0;x<number_input;x++)
{
    for(y=0;y<number_input;y++)
    {
        fprintf(file_pointer,"%f    ",Win[x][y]);
    }
}
fclose(file_pointer);

// save weightout to file
printf("\nplease enter weightout name file:");
scanf("%s",&str4);
file_pointer=fopen(str4,"w");
for (x=0;x<number_output;x++)
{
    for(y=0;y<number_output;y++)
    {
        fprintf(file_pointer,"%f    ",Wout[x][y]);
    }
}
fclose(file_pointer);
printf("\nplease enter show name file:");
scanf("%s",&str5);
pointer4=fopen(str5,"w");
for (x=0;x<num_loop;x++)
{
    fprintf(pointer4,"%f\n",show[x]);
}
fclose(pointer4);
//////////show output//////////

printf("z=%d\n",z);
for(x=0;x<number_output;x++)
{
    printf("output no.%d=%f",x,output[x]);
}
getch();
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

//-----
#include <vcl.h>
#pragma hdrstop

#include <conio.h>
#include <errno.h>
#include <fcntl.h>
#include <io.h>
#include <iostream.h>
#include <fstream.h>
#include <malloc.h>
#include <math.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

#include "Dynamee.h"
//-----
#pragma resource "*.dfm"

#define twopi 3.1415926535 * 2.0
#define pi 3.14592654

int Sign(long a);
void bit_reversal(unsigned int *, int, int);
void WTS(float *, float *, int, int);
void FFT(float *xr, float *xi, float *, float *, int, int);

TPagesDlg *PagesDlg;
//-----
__fastcall TPagesDlg::TPagesDlg(TComponent* AOwner)
: TForm(AOwner)
{
}
//-----
void __fastcall TPagesDlg::OpenFileClick(TObject *Sender)
{
    OpenDialog1->Execute();
    ProgressBar1->Position = 0;
    MediaPlayer1->FileName = OpenDialog1->FileName;
    MediaPlayer1->Open();
}
//-----
void __fastcall TPagesDlg::TESTClick(TObject *Sender)
{
    ProgressBar1->Enabled = true;

    FILE *file_pointer, *pointer, *fptr;
    int i, j, k, m, n, N, n2, begin, end, nFrame, overlap, NOutput;
    int Power[32000], XPower[30000], ZeroX[30000], ZCross[30000];
    long Parameter[32000];
    double Freq[12][1024], mean[1024], Xigma2[12], Sd[12];
    float window[12][1024], Zcore[10][500], w[1024];
    float *wr, *wi, *xr, *xi;
    float Corr, save[10], AutoCor, z1[5000], z2[5000];
    char infile[10], string[5];
    long handle, length;
    unsigned char sound[36000], Frame[12][1024];
    unsigned int *L;

/*===== Load Input Sound =====*/
    for (i=0; i<36000; i++) sound[i]=0;

    MediaPlayer1->FileName = OpenDialog1->FileName;
    file_pointer = fopen(MediaPlayer1->FileName.c_str(), "r");
    handle = fileno (file_pointer);
    length = filelength (handle);
    fread (sound, length, 1, file_pointer);
    fclose(file_pointer);

/*===== Save Input Sound to file for the use of the program =====*/
    pointer = fopen("InputSound.out", "w");
    for (i=0; i<length; i++) {
        fprintf(pointer, "%d\n", sound[i]);
    }
    fclose(pointer);
}

```

*เอกสารนี้เป็นเอกสารที่เผยแพร่โดยทางโรงเรียนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าในรูปแบบใดก็ตาม หากมีข้อผิดพลาดประการใดขออภัยเป็นอย่างสูงและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

begin = 256;
end = length -128;
length = end - begin;

/*===== Find Begin & End Point =====*/
for (i=0; i<length; i++)
    Power[i] = abs((int) sound[i+begin]-128);
XPower[0] = 0;
for (i=0; i<256; i++)
    XPower[0] = XPower[0]+Power[i];
for (i=0; i<length-256; i++)
    XPower[i+1] = XPower[i]-Power[i]+Power[i+256];

for (i=0; i<length; i++)
    ZeroX[i] = abs(Sign(long(sound[i+begin+1]))-Sign(long(sound[i+begin])));
ZCross[0] = 0;
for (i=0; i<256; i++)
    ZCross[0] = ZCross[0] + ZeroX[i];
for (i=0; i<length-256; i++)
    ZCross[i+1] = ZCross[i]-ZeroX[i]+ZeroX[i+256];

for (i=0; i<length-256; i++)
    Parameter[i]=long(XPower[i])*long(ZCross[i]);

for (i=0; i<length; i++)
    Series1->Add(sound[i], ' ', clYellow);
for (i=0; i<length; i++)
    Series4->Add(sound[i], ' ', clRed);
for (i=0; i<length-256; i++)
    Series5->Add(XPower[i], ' ', clBlue);
for (i=0; i<length-256; i++)
    Series6->Add(ZCross[i], ' ', clGreen);
for (i=0; i<length-256; i++)
    Series7->Add(double(Parameter[i]), ' ', clRed);

for (i=0; i<length-256; i++)
    if ((Parameter[i] > 10000)&&(XPower[i] > 600))
    {
        begin = i+256;
        break;
    }
for (i=length-256; i>begin; i--)
    if ((Parameter[i] > 10000)&&(XPower[i] > 600))
    {
        end = i+256;
        break;
    }

length = end-begin;

/*===== Cut Quiet Sound & Store Only Voice =====*/
for(i=0; i<length; i++) // Cut Voice
    sound[i] = sound[i+begin];

pointer = fopen("Voice.out", "w");
for(i=0; i<length; i++)
    fprintf(pointer, "%d\n", sound[i]);
fclose(pointer);

for (i=0; i<length; i++)
    Series2->Add(sound[i], ' ', clRed);

/*===== Set Overlap =====*/
if (length < 1024) overlap = 1024;
    overlap = 1024 - (length-1024)/9;

/*===== Frame Score and Hamming Window=====*/
n = 0;
nFrame = 10; // fix the number of Frame
for(i=0; i<=nFrame; i++)
    for(j=0; j<1024; j++)
        Frame[i][j] = sound[n];
        w[j] = 0.54-0.46*cos(twopi*j/(1024));

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

        window[i][j] = w[j]*(float)sound[n];
        n = n+1;
    )
    n = n-overlap;
}
for (i=0; i<1024; i++)
    Series8->Add(int(window[4][i]), ' ', clRed);

/*===== Fourier Transform =====*/
N = 1024;
m = (int)(log10((double)N)/log10((double)2.0));
k = 1;
for(i = 0; i < m; i++) k <<= 1;

L = (unsigned int *)malloc(N*sizeof(unsigned int));

bit_reversal(L,m,N);

xr = (float *)malloc(N*sizeof(float));
xi = (float *)malloc(N*sizeof(float));

n2 = (N>>1)-1;
wr = (float *)malloc(n2*sizeof(float));
wi = (float *)malloc(n2*sizeof(float));

for(i=0; i<nFrame; i++)
{
    for(j = 0; j < N; j++)
    {
        k = L[j];
        xr[k] = window[i][j];
        xi[k] = 0.0;
    }
    WTS(wr, wi, N, -1);
    FFT(xr, xi, wr, wi, m, N);
    for(k=0; k<512; k++)
    {
        if((xr[k]==0)&&(xi[k]==0))
        {
            ShowMessage( AnsiString("Couldn't read file") );
            exit(1);
        }
    }
    for (k=0; k<512; k++)
        Freq[i][k] = 20*log(sqrt(xr[k]*xr[k]+xi[k]*xi[k]));
}
for (i=0; i<512; i++)
    Series9->Add(int(Freq[4][i]), ' ', clBlue);

/*===== Normalize to Z Parameter =====*/
for(i=0; i<nFrame; i++)
{
    mean[i] = 0.0;
    Xigma2[i] = 0.0;
    for(j=0; j<500; j++)
    {
        mean[i] = mean[i]+Freq[i][j+12];
        Xigma2[i] = Xigma2[i]+Freq[i][j+12]*Freq[i][j+12];
    }
    mean[i] = mean[i]/500.0;
    Sd[i] = sqrt(Xigma2[i]/500.0-mean[i]*mean[i]);

    for(m=0; m<500; m++)
        Zcore[i][m] = (Freq[i][m+12]-mean[i])/Sd[i];
}
for (i=0; i<nFrame; i++)
    Series10->Add(int(1000*Zcore[4][i]), ' ', clGreen);

/*===== Set Data to find Z Parameter =====*/
for (i=0; i<10; i++)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับภาคใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุดเปลี่ยนแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

        for (j=0; j<500; j++)
            z1[i*500+j] = Zcore[i][j];

strcpy(infile, "z00.out");

for(k=0; k<10; k++)
{
    switch(k)
    {
        case 0: infile[1] = '0';break;
        case 1: infile[1] = '1';break;
        case 2: infile[1] = '2';break;
        case 3: infile[1] = '3';break;
        case 4: infile[1] = '4';break;
        case 5: infile[1] = '5';break;
        case 6: infile[1] = '6';break;
        case 7: infile[1] = '7';break;
        case 8: infile[1] = '8';break;
        case 9: infile[1] = '9';break;
    }

    fptr = fopen(infile, "r");
        for (i=0; i<5000; i++)
            fscanf(fptr, "%f", &z2[i]);
    fclose(fptr);

    Corr = 0;
    for (i=0; i<5000; i++)
        Corr = Corr + z1[i]*z2[i];
        AutoCor = Corr/5000;

    save[k] = AutoCor;
}

pointer = fopen("Correlation.out", "w");
for(i=0; i<nFrame; i++)
    fprintf(pointer, "%f\n", save[i]);
fclose(pointer);

Corr = 0;
for ( i=0; i<10; i++)
    Corr = max(Corr, save[i]);
for ( i=0; i<10; i++)
    if ( Corr == save[i] )
    {
        itoa(i, string, 10);
        Edit1->Text = string;
    }
for (i=0; i<10; i++)
    Series3->Add(save[i], ' ', clBlue);

/*===== Begin Neural =====*/
float Win[20][20], Wout[20][20], input[20], inpuh[20],
outputh[20], output[20];
int number_input, number_output, number_inpuh,
number_outpuh, x, y;

number_input = 10;
number_outpuh = number_inpuh = 6;
number_output = 4;

for(i=0; i<number_input; i++) input[i] = save[i];

/*===== load weight for first layer =====*/
pointer= fopen("Win1", "r");
for (x=0; x<number_inpuh; x++)
{
    for (y=0; y<number_input; y++)
        fscanf(pointer, "%f", &Win[x][y]);
} fclose(pointer);

/*===== load weight for second layer =====*/
pointer= fopen("Win2", "r");
for (x=0; x<number_output; x++)
{
    for (y=0; y<number_outpuh; y++)
        fscanf(pointer, "%f", &Wout[x][y]);
} fclose(pointer);

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามเผยแพร่ต่อผู้อื่นและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

/*===== start program =====*/
/*===== Multiply and Summation for First layer =====*/
for (x=0; x<number_inpuh; x++)
{
    inputh[x]=0;
    for(y=0; y<number_input; y++)
        inputh[x]=inputh[x]+Win[x][y]*input[y];
}
/*===== Sigmoid function =====*/
for (x=0; x<number_inpuh; x++)
    outputh[x]=(float)(1/(1+exp(-inputh[x])));

/*===== multiplier and summation for second layer =====*/
for (x=0; x<number_output; x++)
{
    output1[x]=0;
    for(y=0;y<number_outpuh;y++)
        output1[x]=output1[x]+Wout[x][y]*outputh[y];
}

/*===== SigMoid function =====*/
for (x=0; x<number_output; x++)
    output[x]=(float)(1/(1+exp(-output1[x])));

for (i=0; i<number_output; i++)
{
    if (output[i]>=0.5) output[i] = 1.0;
    else output[i] = 0.0;
}

NOutput = (int)output[0]*8
+ (int)output[1]*4
+ (int)output[2]*2
+ (int)output[3];

if (NOutput > 9 ) Edit2->Text = '???';
else
{
    itoa(NOutput, string, 10);
    Edit2->Text = string;
}

/*===== End program =====*/
}
//-----

```

```

void __fastcall TPagesDlg::MediaPlayer1Click(TObject *Sender,
TMPBtnType Button, bool &DoDefault)
{
    if(Button == btRecord)MediaPlayer1->Save();
}
//-----

```

```

int Sign(long a)
{
    if (a > 128) return(1);
    if (a == 128) return(0);
    if (a < 128) return(-1);
}

```

/******

```

void bit_reversal(unsigned int *L, int m, int N)
/* Routine for generating LUT for bit reversal.
Note: N = (2 to the power of m). LUT will reside in L[] */
{
    unsigned int MASK, C, A, j, k, i;

    for(k = 0; k < N; k++){
        MASK = 1;
        C = 0;
        for(i = 0; i < m; i++){
            A = (k&MASK) >> i;
            j = (A << (m-i)) | C;
            C |= A;
            MASK = MASK << 1;

```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และห้ามเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าในรูปแบบใดก็ตาม อีกทั้งห้ามมีเหตุดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

    }
    L[k] = C;
}
/*****

void WTS(float *wr, float *wi, int N, int sign)
{
    int n2, i;
    float theta;

    n2 = (N>>1)-1;

    theta = 2.0*pi/((float)N);
    for(i = 0; i < n2; i++){
        wr[i] = (float)cos((double)((i+1)*theta));
        wi[i] = (float)sin((double)((i+1)*theta));
        if(sign==(int)(-1))wi[i] =- wi[i];
    }
}
/*****

void FFT(float *xr, float *xi, float *wr, float *wi, int m, int N)
{
    int ip, k, kk, l, incr, iter, j, i, loop;
    float Tr, Ti;

    for(loop=0; loop < N; loop++)
    {
        xi[loop]/=(float)N;
        xr[loop]/=(float)N;
    }

    ip = 1;
    kk = (N>>1);
    incr = 2;

    for(iter=0; iter < m; iter++){
        for(j=0; j < N; j+= incr){
            i = j + ip;
            Tr = xr[i];
            Ti = xi[i];
            xr[i] = xr[j] - Tr;
            xi[i] = xi[j] - Ti;

            xr[j] = xr[j] + Tr;
            xi[j] = xi[j] + Ti;
        }
        if(iter != 0){
            for(k=1; k<ip; k++){
                l=k*kk-1;
                for(j=k; j<N; j+=incr){
                    i=j+ip;
                    Tr = xr[i]*wr[l]-xi[i]*wi[l];
                    Ti = xr[i]*wi[l]+xi[i]*wr[l];

                    xr[i]=xr[j]-Tr;
                    xi[i]=xi[j]-Ti;

                    xr[j]=xr[j]+Tr;
                    xi[j]=xi[j]+Ti;
                }
            }
        }
        kk >>= 1;
        ip <<= 1;
        incr <<= 1;
    }
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
void __fastcall TPagesDlg::ClearClick(TObject *Sender)
```

```
{  
    MediaPlayer1->Close();  
}  
//-----
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

1. Lawrence R. Rabiner / Ronald W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632 1978
3. Panos E. Papamichalis, Ph.D., Practical Approaches to speech coding, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 1987
4. J.R. Parker, Algorithms for image processing and computer vision, Wiley Computer Publishing, 1996
5. Thomas W. Parsons, Voice and Speech Processing, McGraw-Hill Book Company, 1986
6. Mazin G. Rahim, Artificial Neural Networks for Speech Analysis/Synthesis, Chapman & Hall
7. Ben Gold & Nelson Morgan, Speech And Audio Signal Processing, Jhon Wiley & Sons, Inc.
8. ศ. ดร. วิมลภ คชระกำพลธร, Digital Signal Processing, บริษัทไตนาว-พรีนท์ จำกัด, ตุลาคม 2533
9. ธันวา ศรีประโมง, การเขียนโปรแกรมภาษาซี สำหรับวิศวกรรม พิมพ์ครั้งที่สี่, โครงการตำราวิชา การมหาวิทยาลัยเทคโนโลยีมหานคร, 2538

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ในการทำโครงการครั้งนี้ สามารถสำเร็จลุล่วงไปได้ด้วยดี ทางคณะผู้จัดทำโครงการขอขอบคุณ รศ.ดร. มนัส สัจจวรศิลป์ เป็นอย่างสูง ที่ท่านคอยดูแลเอาใจใส่และให้คำปรึกษาอย่างใกล้ชิด และขอขอบคุณพี่จิ่ง พี่นนท์ พี่ปุ๋ และเพื่อน ๆ ในห้อง B204 ที่ให้คำปรึกษาในด้านต่าง ๆ และให้กำลังใจจนโครงการสามารถสำเร็จลุล่วงลงได้



ธนากร สักดิ์ชื่นยศ
(ธนากร สักดิ์ชื่นยศ)

นันทินี นัททะแสน

(นันทินี นัททะแสน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้