

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การรู้จำเสียงพูดคำไทยแบบไม่ขึ้นกับผู้พูดด้วยนิวรัลเน็ตเวิร์ค
INDEPENDENT SPEAKER THAI WORDS SPEECH RECOGNITION
USING NEURAL NETWORKS



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2540

ISBN 974-622-061-6

ลิขสิทธิ์ของบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เลขหมู่.....
เลขทะเบียน.....
วัน, เดือน, ปี.....

29396

26 ส.ค. 2541

สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

INDEPENDENT SPEAKER THAI WORDS SPEECH RECOGNITION
USING NEURAL NETWORKS



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING
SCHOOL OF GRADUATE STUDIES
KING MONGKUT 'S INSTITUTE OF TECHNOLOGY LADKRABANG
1997

ISBN 974-622-061-6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์

การรู้จำเสียงพูดคำไทยแบบไม่ขึ้นกับผู้พูดด้วยนิวรัลเน็ตเวิร์ค

นักศึกษา

นายเสรี ปานซาง

อาจารย์ผู้ควบคุมวิทยานิพนธ์

รศ.ดร. ชม กิมปาน

ระดับการศึกษา

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.

2540

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้มีจุดมุ่งหมายเพื่อศึกษาพฤติกรรมและหาอัตราการรู้จำเสียงพูดคำไทยโดยจำนวน 20 เสียงทดสอบ แบบไม่ขึ้นกับผู้พูด เสียงทดสอบที่ใช้ทดสอบคือ "ศูนย์", "หนึ่ง", "สอง", "สาม", "สี่", "ห้า", "หก", "เจ็ด", "แปด", "เก้า", "สิบ", "เปิด", "ปิด", "หมุน", "ยก", "วาง", "ช่อง", "ซ้าย", "ขวา", "ไฟ" ในขบวนการ ปริ-โปร เซลซึ่ง สัญญาณในแกนเวลาของเสียงที่ใช้ทดสอบจะถูกแปลงให้เป็นสัญญาณดิจิทัล แล้วใช้การแปลงฟูเรียร์อย่างรวดเร็ว ร่วมกับการแปลงข้อมูลในแกนความถี่ ให้อยู่ในแกนฮาร์โมนิคของความถี่ [3] เพื่อลดขนาดของข้อมูลและขยายรายละเอียดในช่วงความถี่ที่มีการเปลี่ยนแปลงมาก ได้เป็นสเปกโตรแกรมของเสียงซึ่งจะนำไปใช้ในการหาคุณลักษณะจำเพาะ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างสเปกโตรแกรมของเสียงที่ต้องการ กับ กลุ่มสเปกโตรแกรมเสียงหลัก แล้วนำไปจัดเรียงแบบมากไปน้อย แล้วเลือกออกมา 10 เปอร์เซนต์แรก เพื่อนับคะแนนความถี่ของเสียงแต่ละเสียง ผลที่ได้จะแสดงคุณลักษณะจำเพาะของเสียงทดสอบนั้น เมื่อนำคุณลักษณะจำเพาะของเสียงทดสอบกลุ่มหนึ่ง มาฝึกสอนให้กับนิวรัลเน็ตเวิร์คแบบแบคพรอบพาซัน โดยกำหนดเป้าหมายที่สอดคล้องกับเสียงที่นำมาฝึกสอน จนกระทั่งค่าผิดพลาดลดลงถึงค่าที่ยอมรับได้ แล้วนำข้อมูลคุณลักษณะจำเพาะของเสียงอีกกลุ่มหนึ่งมาทดสอบ นิวรัลเน็ตเวิร์คจะสามารถแยกแยะระดับความสัมพันธ์ของเสียงที่นำมาทดสอบนั้นสอดคล้องกับเป้าหมายที่เคยเรียนรู้มา

จากผลการดำเนินงาน อัตราการรู้จำ คุณลักษณะจำเพาะของเสียงทดสอบโดยไม่ขึ้นกับผู้พูด 20 เสียง พูดยเสียงทดสอบละ 3 ครั้ง จากกลุ่มเสียงที่ใช้ฝึกสอน 10 คน โดยใช้กลุ่มเสียงทดสอบ 12 คน เป็น 73.89 % เมื่อลดจำนวนเสียงทดสอบลงมา เหลือ 11 เสียง คือ "ศูนย์", "หนึ่ง", "สอง", "สาม", "สี่", "ห้า", "หก", "เจ็ด", "แปด", "เก้า", "สิบ" อัตราการรู้จำเป็น 80.05 % อัตราการรู้จำคุณลักษณะจำเพาะของเสียงทดสอบโดยขึ้นกับกลุ่มผู้พูด 20 เสียง ภายในกลุ่ม 22 คน เป็น 84.32 % และอัตราการรู้จำเสียงทดสอบ 11 เสียง ภายในกลุ่มผู้พูด 12 คน เป็น 92.42 % จากผลการทดลองพบว่าอัตราการรู้จำจะแปรผกผันตามจำนวนผู้พูดและเสียงทดสอบ การวิเคราะห์แพตเทิร์นของเสียงแบบเทียบกับแบบอ้างอิงที่ใช้อัลกอริทึมแบบเดียวกัน ไม่สามารถนำไปใช้รู้จำเสียงทดสอบจำนวนมาก แบบไม่ขึ้นกับผู้พูดได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Independent Speaker Thai words speech Recognition using Neural Networks
Student	Mr. Seri Pansang.
Thesis Advisor	Assoc. Prof. Chom Kimpan. D.Eng.
Level of Study	Master of Engineering in Electrical Engineering King Mongkut's Institute of Technology Ladkrabang
Year	1997

Abstract

This thesis was aimed at studying behaviour and finding speech recognition rate of Thai words in isolation of independent speaker by using Back Propagation Neural Network. In pre-processing unit to find voices pattern in harmonic frequency domain by using Fast Fourier Transform and find characteristic relation of voices by correlation from the level of relationship between test pattern spectrogram and all of the reference spectrum. All of voice characteristic relation are using for training and testing process in neural network

· this purposed method results independent voice recognition rate :

73.89 % which 20 Thai words by 20 training and 12 testing persons

80.05 % which 11 Thai words by 20 training and 12 testing persons

and

92.42 % which 11 Thai words by 10 training and 10 testing persons for dependent group.

The results of experiment showing that recognition rate is inverse which number of testing word and number of persons. This method could not use for many testing word and speaker.

กิตติกรรมประกาศ

การจัดทำวิทยานิพนธ์นี้ สำเร็จลุล่วงลงได้ด้วยความอนุเคราะห์ของ รศ.ดร.ชม กัมปาน อาจารย์ที่ปรึกษา ที่กรุณาให้คำปรึกษา คอยให้กำลังใจ คำแนะนำและชี้แนะแนวทางวิธีการซึ่งเป็นประโยชน์อย่างมากในการดำเนินการวิจัย ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ พี่ เพื่อน และน้องๆ ในสถาบันราชภัฏเชียงใหม่ ที่ให้การสนับสนุน คอยให้กำลังใจ และขอบคุณเจ้าของตัวอย่างเสียงทุกท่านที่สละเวลาในการจัดเก็บข้อมูลเสียง วิทยานิพนธ์ฉบับนี้จะสำเร็จลุล่วงไม่ได้เลยหากขาดการสนับสนุนจากภรรยาคุณวนิดาและบุตรสาวเด็กหญิงวาสนี ปานซาง ข้าพเจ้าขอขอบคุณมา ณ.ที่นี้ด้วย และขอขอบพระคุณทุนอุดหนุนการทำวิทยานิพนธ์ จากบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง

เสรี ปานซาง



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญภาพ.....	VIII
บทที่	
1 บทนำ.....	1
ความเป็นมาของงานวิจัย.....	1
จุดมุ่งหมายของการวิจัย.....	1
สมมติฐานและการประยุกต์ความรู้ด้านเสียงมาใช้กับงานวิจัย.....	2
หลักการที่นำมาใช้ในการรู้จำเสียงพูดในงานวิจัย.....	3
ขอบเขต.....	4
ขั้นตอนและวิธีการดำเนินงาน.....	5
เครื่องมือที่ใช้ทำการทดลอง.....	6
ข้อตกลงเบื้องต้น.....	7
2 ระบบการได้ยินและการวิเคราะห์เสียงของมนุษย์.....	8
หูและการได้ยิน.....	8
ธรรมชาติการรู้จำเสียง.....	9
3 การสร้างระบบการรู้จำเสียงพูดคำไทย.....	11
การวิเคราะห์เงื่อนไขการสร้างระบบการรู้จำที่ใช้ในการทดลอง.....	11
- ระบบการรู้จำที่สร้างขึ้นใช้ในการทดลอง.....	11
- การแก้ไขปัญหาการเปรียบเทียบเสียงที่มีความดัง, ระยะเวลาการเปล่งเสียงและความถี่ที่แตกต่างกัน.....	12
การนำนิวรัลเน็ตเวิร์คมาใช้ในการระบบการรู้จำเสียงพูดคำไทย.....	16
การสร้างแพตเทิร์นอ้างอิงจากสเปกโตรแกรมเสียงกลุ่มอ้างอิง.....	16
- สหสัมพันธ์ (Correlation).....	17
- การเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสเปกโตรแกรมเสียงสองชุด.....	18

สารบัญ (ต่อ)

บทที่	หน้า
	วิธีการหาคุณลักษณะจำเพาะสัมพันธ์ของเสียง.....23
	การแยกคำออกจากสเปกโตรแกรม.....24
	วิธีการหาคุณลักษณะจำเพาะสัมพันธ์และลดขนาดข้อมูลเสียง.....27
	การแบ่งแบนด์ของสเปกโตรแกรมเพื่อเพิ่มความแม่นยำ.....32
4	นิวรัลเน็ตเวิร์ค (Neural Network).....36
	ความรู้เบื้องต้นเกี่ยวกับนิวรัลเน็ตเวิร์ค.....36
	- นิวรัลเน็ตเวิร์คชีวภาพ.....37
	- โครงข่ายประสาทเทียม (Artificial Neural Network).....38
	- ฟังก์ชันกระตุ้นความสนใจ (Activation Function).....39
	- โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Artificial Neural Network).....42
	- โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Artificial Neural Network).....44
	- ฟังก์ชันกระตุ้นความสนใจแบบไม่เป็นเชิงเส้น (The Nonlinear Activation Function).....45
	- การฝึกสอนให้กับโครงข่ายประสาทเทียม (Training of Artificial Neural Network) 45
	- วัตถุประสงค์ของการเทรนนิ่ง (Objective of Training).....46
	- การเทรนนิ่งแบบควบคุม (Supervised Training).....46
	- การเทรนนิ่งแบบอิสระ (Unsupervised Training).....47
	- วิธีการแก้ปัญหการฝึกสอน (Traning lgorithm).....47
	การประยุกต์ใช้นิวรัลเน็ตเวิร์คกับงานวิจัย.....48
	- ขบวนการเทรนนิ่ง.....49
	- FORWARD PASS.....50
	- REVERSE PASS.....51
	- เงื่อนไขการหยุดการฝึกสอน.....56
	- การป้อนข้อมูลอินพุตให้กับนิวรัลเน็ตเวิร์คในกระบวนการฝึกสอน.....56
	- ขบวนการทดสอบ.....57
	- การป้อนข้อมูลอินพุตให้กับนิวรัลเน็ตเวิร์คในกระบวนการทดสอบ.....58

สารบัญ (ต่อ)

บทที่	หน้า
5	การดำเนินการและการทดลอง.....59
	การเลือกข้อมูลเพื่อใช้ในการทดลอง.....59
	วิธีการเก็บและจัดการข้อมูลเพื่อการประมวลผล.....59
	ขั้นตอนการดำเนินการทดลอง.....60
	วิธีการทดลอง.....62
	- การทดลองและผลการทดลองแบบที่ 1.....62
	- การทดลองและผลการทดลองแบบที่ 2.....70
	- การทดลองและผลการทดลองแบบที่ 3.....73
	- การทดลองและผลการทดลองแบบที่ 4.....75
6	การลดขนาดของข้อมูลด้วยวิธีการหาค่าเฉลี่ยร่วมกับการประมาณค่า.....79
	การปรับขนาดของข้อมูลให้เหมาะสมในกระบวนการปริโปรเซสซิง.....79
	วิธีการปรับระดับพลังงานสูงสุด-ต่ำสุดของข้อมูลสเปกโตรแกรมเสียงให้อยู่ในช่วงเดียวกัน...79
7	สรุปผลการวิจัยและข้อเสนอแนะ.....87
	- สรุปผลการทดสอบ.....87
	- การรู้จำคำไทยโดยไม่ขึ้นต่อผู้พูด.....87
	- การรู้จำคำไทยโดยขึ้นต่อกลุ่มของผู้พูด.....88
	- ผลการทดลองที่ดำเนินการในบทที่ 6.....89
	สรุปผลการวิจัย.....90
	ข้อเสนอแนะ.....92
	บรรณานุกรม.....93
	ภาคผนวก.....95
	ภาคผนวก ก การจัดแฟ้มข้อมูล.....96
	ภาคผนวก ข กราฟแสดงค่า Sum square error ที่ได้จากการฝึกสอน.....100
	ภาคผนวก ค โปรแกรมที่พัฒนาขึ้นในงานวิจัย.....102
	ภาคผนวก ง ผลงานวิจัยที่ได้รับการตีพิมพ์.....126
	ประวัติผู้เขียน.....133

สารบัญตาราง

ตารางที่	หน้า
1 ตารางแสดงผลการตรวจสอบนิวัตรเน็ตเวอร์ค ด้วยข้อมูลเดิมที่ใช้ฝึกสอน ค่าในตารางเป็นค่าเฉลี่ยของแต่ละเสียงทดสอบที่ได้จากคน 10 คน ในกลุ่ม A ซึ่งทดสอบหลังจากการฝึกสอนแล้ว จำนวน 2500 รอบ.....	65
2 แสดงผลการทดสอบการรู้จำของนิวัตรเน็ตเวอร์ค ด้วยข้อมูลเสียงทดสอบกลุ่ม C เฉพาะข้อมูลเสียงทดสอบที่พูดครั้งแรก	66
3 แสดงผลการทดสอบการรู้จำของนิวัตรเน็ตเวอร์ค ด้วยข้อมูลเสียงจากผู้พูด 12 คน ในกลุ่ม C เฉพาะข้อมูลที่พูดครั้งที่สองของแต่ละเสียงทดสอบ.....	67
4 แสดงผลการทดสอบการรู้จำของนิวัตรเน็ตเวอร์ค ด้วยข้อมูลเสียงทดสอบจากผู้พูด 12 คน ในกลุ่ม C เฉพาะข้อมูลเสียงทดสอบที่พูดครั้งที่สามของแต่ละเสียงทดสอบ.....	68
5 แสดงผลการรู้จำของเน็ตเวอร์คที่ ฝึกสอนด้วยข้อมูลกลุ่ม B และ C 22 คน ด้วยข้อมูลเสียงที่เปล่งครั้งที่ 1 และ 2 แล้วทดสอบ ด้วย ข้อมูลกลุ่ม B และ C 22 คน เช่นกัน แต่เป็นข้อมูลเสียงที่เปล่งครั้งที่ 3 ตามแนวคิดของการทดลองแบบที่ 2.....	72
6 แสดงผลการตรวจสอบการรู้จำของนิวัตรเน็ตเวอร์คที่มีต่อ ชุดเสียงกลุ่มเสียงตัวเลข ศูนย์ ถึง สิบ 11 เสียง จากผู้ออกเสียง 12 คน ออกเสียงคนละ 3 ครั้งต่อ 1 เสียง ตัวเลขทดสอบ รวม 36 ครั้งโดยเน็ตเวอร์คไม่เคยเรียนรู้เสียงจาก 12 คนนี้มาก่อน.....	73
7 แสดงผลการตรวจสอบการรู้จำของนิวัตรเน็ตเวอร์คที่มีต่อชุดเสียงกลุ่มคำที่ใช้สั่งงาน 9 คำสั่ง จากผู้ออกเสียง 12 คน ออกเสียงคนละ 3 ครั้งต่อ 1 เสียง คำสั่ง ทดสอบ โดยเน็ตเวอร์คไม่เคยเรียนรู้เสียงจาก 12 คนนี้มาก่อน.....	75
8 แสดงผลการทดสอบนิวัตรเน็ตเวอร์ค ด้วยข้อมูลกลุ่มในเดียวกัน ในส่วนที่เหลือจาก การฝึกสอน ได้จากจำนวนผู้พูด 32 คน การพูด 2 ครั้งแรก ของทุกเสียงทดสอบจะ นำมาใช้ฝึกสอน การพูดครั้งที่ 3 นำมาใช้ทดสอบ.....	76
9 แสดงผลการทดสอบการรู้จำของ นิวัตรเน็ตเวอร์ค ด้วยข้อมูลกลุ่มในเดียวกัน ในส่วนที่ เหลือจากการฝึกสอน ที่ได้จาก ผู้พูดคนที่ 21 ถึง 32 จำนวน 12 คน การพูด 2 ครั้งแรก ของทุกเสียงทดสอบจะนำมาใช้ฝึกสอน การพูดครั้งที่ 3 นำมาใช้ทดสอบ.....	77
10 ตารางสรุปผลการทดลองจากตารางที่ 2-9 การทดลองแบบที่ 1, 3-1, 3-2 เป็นการหา อัตราการเรียนรู้ของระบบแบบไม่ขึ้นกับผู้พูด การทดลองแบบที่ 2, 4-1, 4-2 เป็นการหา อัตราการเรียนรู้ของระบบแบบขึ้นกับกลุ่มของผู้พูด.....	78

สารบัญตาราง (ต่อ)

ตารางที่

หน้า

11	แสดง อัตราการรู้จำโดยใช้วิธีการลดขนาดของสเปกโตรแกรม และปรับให้ทุก สเปกโตรแกรมให้มีจำนวนสเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 20 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 20 คน และทดสอบด้วยเสียงทดสอบ จำนวน 20 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12 คน.....	83
12	แสดง อัตราการรู้จำโดยใช้วิธีการลดขนาดของสเปกโตรแกรม และปรับให้ทุก สเปกโตรแกรมให้มีจำนวนสเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 11 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 20 คน และทดสอบด้วยเสียงทดสอบ จำนวน 11 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12 คน.....	85
13	แสดง อัตราการรู้จำโดยใช้วิธีการลดขนาดของสเปกโตรแกรม และปรับให้ทุก สเปกโตรแกรมให้มีจำนวนสเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 9 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 20 คน และทดสอบด้วยเสียงทดสอบ จำนวน 9 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12 คน.....	86

สารบัญภาพ

		หน้า
1	แสดงทูลส่วนนอก, ส่วนกลาง, ส่วนใน และส่วนประกอบต่างๆของทูล.....	8
2	แสดงโครงสร้างระบบการรู้จำแบบ Isolated word recognition.....	11
3	แสดงสเปกโตรแกรมของเสียง “เก้า” ที่แปลงเสียงโดยผู้พูดคนเดียว 6 ครั้ง.....	13
4	แสดงสเปกโตรแกรมของเสียง “เก้า” ที่แปลงเสียงโดยผู้พูด 10 คน.....	14
5	แสดงสเปกโตรแกรมของชุดข้อมูลเสียง “หนึ่ง” ถึง “ห้า” ที่แปลงจากคนเดียว จากซ้ายไปขวาตามลำดับ.....	19
6	แสดงสเปกตรัมของแถบความถี่(เฟรม) ของเสียงพูด “หนึ่ง” ช่วงเวลา Δt ที่ 6.....	20
7	แผนภูมิแสดงผังงานการหาระดับความสัมพันธ์ระหว่างสเปกโตรแกรมเสียงทดสอบ กับ เสียงอ้างอิง.....	22
8	แสดงบล็อกไดอะแกรมของวิธีการหาคุณลักษณะจำเพาะสัมพันธ์ของเสียงโดยใช้ สัมประสิทธิ์สหสัมพันธ์.....	23
9	แสดงผลการแยกข้อมูลช่วงที่มีการแปลงเสียงออกจากสเปกโตรแกรมโดยใช้ สมการที่ 3.12.....	26
10	แสดงวิธีการหาค่าคุณสมบัติจำเพาะสัมพันธ์โดยใช้สหสัมพันธ์ที่สร้างขึ้น.....	27
11	แสดงวิธีการนำข้อมูลสเปกโตรแกรมอ้างอิง(B) หนึ่งเสียง(ก) ไปหาระดับความสอดคล้องของข้อมูล กับชุดข้อมูลสเปกโตรแกรมหลัก(A) ทุกเสียง ซึ่งจะได้ ค่าระดับความสัมพันธ์ 600 ค่า ต่อ 1เสียงจากกลุ่ม (B).....	29
12	แสดงการหาคุณสมบัติจำเพาะสัมพันธ์ ของชุดข้อมูลเสียงอ้างอิง(B) กับชุดข้อมูลเสียงหลัก (A) แล้วนำไปฝึกสอนให้กับนิรลเน็ตเวิร์ค.....	30
13	แสดงการหาคุณลักษณะจำเพาะสัมพันธ์ ของชุดข้อมูลเสียงทดสอบ C กับชุดข้อมูลหลัก A แล้วนำไปทดสอบนิรลเน็ตเวิร์ค.....	31
14	แสดงการจัดแบ่งสเปกโตรแกรมของเสียงหนึ่งออกเป็น 3 แบนด์ แบบซ้อนทับกัน 50% ของข้อมูล ซึ่งนำไปใช้เปรียบเทียบกับสเปกโตรแกรมของเสียงอ้างอิงต่าง ๆ แบบแบนด์ต่อแบนด์	34
15	แผนภูมิแสดงผังงานการหาค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสามแบนด์ ของ สเปกโตรแกรมเสียงที่ต้องการหาค่ากับเสียงหลัก.....	29
16	แสดงโครงสร้างตัวอย่างของเซลล์ประสาทชีวภาพ.....	37
17	แสดงไดอะแกรมของนิรลเน็ตเวิร์คที่สร้างขึ้น (Artificial Neuron).....	39
18	แสดงโมดูลนิรลเน็ตเวิร์คที่สร้างขึ้นร่วมกับ Activation Function.....	40

สารบัญภาพ (ต่อ)

		หน้า
19	แสดงกราฟที่ได้จากสมการซิกมอยด์ลอจิสติกฟังก์ชัน.....	41
20	แสดง Hyperbolic Tangent Function.....	42
21	แสดงลักษณะโครงข่ายประสาทเทียมแบบชั้นเดียว.....	43
22	แสดงไดอะแกรมของ Backpropagation Neural Network แบบสองชั้น.....	44
23	แสดงไดอะแกรมของ Two Layer Feed-forward.....	49
24	แสดงไดอะแกรมเชื่อมโยงระหว่างชั้นของนิรวลแบบ 2 ชั้น ของ Backpropagation Network.....	50
25	แสดงไดอะแกรมของการปรับค่า Weight ในชั้น Output layer และ Hidden layer.....	52
26	ไดอะแกรมแสดงวิธีการหาค่าเวกซ์ที่เหมาะสมของการเทรนนิ่งโดยใช้ อัลกอริทึมแบบเบคพรอบพาเกชั่น.....	54
27	แผนภูมิแสดงFlow chart แสดงขั้นตอนการเทรนนิ่ง เพื่อหาค่าเวกซ์ ที่เหมาะสมในขั้นตอนการทดสอบ.....	55
28	แสดงการจัดข้อมูลเพื่อฝึกสอนให้กับนิรวลเน็ตเวิร์ค มีลักษณะเป็น Data pair ที่ให้เป้าหมายมีลักษณะเป็น "One of end coding".....	57
29	แสดงการจัดข้อมูลเพื่อทดสอบนิรวลเน็ตเวิร์ค เพื่อหาอัตราการเรียนรู้.....	58
30	แผนภูมิแสดงไดอะแกรมการใช้เบคพรอบพาเกชั่น นิรวลเน็ตเวิร์ค เพื่อให้รู้จำเสียงแบบไม่ขึ้นกับผู้พูด.....	63
31	กราฟแสดงการเปรียบเทียบอัตราการเรียนรู้ของนิรวลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด 3 ครั้ง จากตารางที่ 2,3 และ 4 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุดกัน.....	69
32	แสดงไดอะแกรมการใช้เบคพรอบพาเกชั่น นิรวลเน็ตเวิร์ค เพื่อให้รู้จำเสียงแบบ ขึ้นกับผู้พูดในกลุ่มเดียวกัน.....	70
33	กราฟแสดงการเปรียบเทียบอัตราการเรียนรู้ของนิรวลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด จำนวนเสียงทดสอบ 11 เสียง จากผู้พูด 12 คน จากตารางที่ 6 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุด.....	74
34	กราฟแสดงการเปรียบเทียบอัตราการเรียนรู้ของนิรวลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด จำนวนเสียงทดสอบ 9 เสียง จากผู้พูด 12 คน จากตารางที่ 7 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุด.....	74

35	แสดงการลดขนาดของข้อมูลแกนความถี่จาก 128 ลงเหลือ 16 ข้อมูล และปรับจำนวนเฟรมของสเปกโตรแกรมซึ่งแต่ละเสียงมีจำนวนไม่เท่ากัน ให้มีจำนวนเท่ากันที่ 10 เฟรมต่อสเปกตรัมเสมอ.....	80
36	ภาพแสดงการปรับข้อมูลในแวลล่างสุด ซึ่งมีข้อมูลจำนวน 13 ข้อมูลต่อแวล นำมาประมาณค่าใหม่ ให้มีข้อมูลลดลงเหลือ 10 ข้อมูลต่อแวล ด้วยการประมาณค่าแบบลากรองอินเตอร์โพลิตัง โพลีโนเมียล.....	81
37	แสดงชื่อแฟ้มที่จัดแบ่งในกลุ่ม A, B และ C.....	98
38	แสดงแฟ้มที่ใช้ และแฟ้มที่ได้ในแต่ละขั้นตอน ในส่วนของ PRE-PROCESSING.....	99
39	กราฟแสดงค่า Sum square error ที่ได้จากการฝึกสอน นิวรัลเน็ตเวิร์ค ด้วยข้อมูลคุณลักษณะจำเพาะ ของเสียงในกลุ่ม A.....	100
40	กราฟแสดงค่า Average Sum square error ที่ได้จากการฝึกสอน ด้วยเสียงในกลุ่มคำสั่งควบคุม 9 เสียง 20 คน คนละ 3 ครั้ง.....	101

บทที่ 1

บทนำ

ความเป็นมาของงานวิจัย

การวิเคราะห์เสียงที่ไม่ขึ้นกับผู้พูด เป็นการพัฒนาระบบการรับคำสั่งของคอมพิวเตอร์ โดยไม่เจาะจงผู้ออกเสียงคำสั่ง เป็นการเปิดกว้างให้กับคนทั่วไปสามารถออกคำสั่งกับคอมพิวเตอร์ได้ เป็นจุดเริ่มต้นของการพัฒนาระบบการสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์โดยใช้เสียงคำพูดเป็นสื่อ ทำให้เกิดความสะดวกในกลุ่มสังคมที่ใช้ประโยชน์จากคอมพิวเตอร์ ซึ่งสามารถประยุกต์ใช้งานได้มากมายเช่น ระบบการป้อนข้อมูลด้วยเสียง, ระบบสั่งงานด้วยเสียงพูดทางโทรศัพท์, การควบคุมอุปกรณ์ไฟฟ้าของคนพิการและผู้ป่วยอัมพาต โดยใช้เสียง เป็นต้น

จุดมุ่งหมายของการวิจัย

1. ศึกษาหาความเป็นไปได้วิธีการรู้จำเสียงพูดคำไทยโดดๆ แบบไม่ขึ้นกับผู้พูด โดยวิธีการเปรียบเทียบระหว่างแบบทดสอบกับแบบอ้างอิงด้วยเทคนิค Correlation และ Neural Networks
2. หาอัตราการรู้จำ แบบไม่ขึ้นกับผู้พูด และ แบบขึ้นกับกลุ่มของผู้พูด
3. เพื่อหาวิธีการวิเคราะห์เสียงพูดคำไทยที่มีความแม่นยำสูงขึ้น
4. ศึกษาวิธีการที่เหมาะสมของการนำ Backpropagation Neural Networks มาประยุกต์ใช้กับงานจดจำเสียงพูด
5. หาความสัมพันธ์ระหว่างอัตราการรู้จำ กับจำนวนผู้พูดและเสียงที่ทดสอบ
6. เพื่อใช้เป็นแนวทางและพื้นฐานความรู้ในการพัฒนาระบบรู้จำเสียงพูดคำไทยโดด ๆ โดยวิธีการอื่นที่ให้ประสิทธิภาพสูงกว่าในอนาคต

สมมติฐานและการประยุกต์ความรู้ด้านเสียงมาใช้กับงานวิจัย

สมมติฐานที่ตั้งขึ้นสรุปจาก การสังเกตพฤติกรรมการได้ยิน, การวิเคราะห์เสียง แล้วพยายามหาขอบเขตที่เหมาะสมในการทำวิจัย มีรายละเอียดดังนี้ การรู้จำเสียงเริ่มขึ้นตั้งแต่วัยทารก มารดาของเด็กก็จะพูดกับลูกบ่อยๆ จนเด็กเคยชินกับเสียงนั้น ในจำนวนคำน้อยๆ ก่อน เช่น แม่ หม่ำๆ จำ พ่อ หรืออะไรก็ได้ เด็กจะรู้จักคำและความหมายง่ายๆ จำนวนคำที่รู้จักจะเพิ่มมากขึ้นตามวัยและพัฒนาการ การเรียนรู้ภาษาต้องเรียนรู้จากผู้อื่นที่อยู่รอบตัว เด็กที่อยู่ตามลำพังไม่ได้พบกับคนอื่นเลยหรือคนที่เสียงไม่พูดกับเด็ก และไม่เปิดโอกาสให้ได้ยินใครพูดอะไรเลย เด็กจะพูดเป็นภาษาไม่ได้

ดร.กาญจนา นาคสกุล (2520:4) ได้ยกตัวอย่าง คำนิยามของคำว่า “ภาษา” ของชาเพียร์ว่า ภาษาเป็นวิธีที่เป็นของมนุษย์โดยเฉพาะเท่านั้นไม่ใช่สิ่งที่เป็นสัญชาตญาณ แต่มีไว้เพื่อเป็นเครื่องมือสื่อสารความคิด ความรู้สึก ความปรารถนา และการกระทำโดยการเปล่งระบบของสัญลักษณ์เป็นเสียงออกมาโดยตั้งใจ Bloch และ Trager(1942:5) อธิบายว่า ภาษาคือ ระบบของสัญลักษณ์ซึ่งแทนด้วยเสียงพูด, สัญลักษณ์เหล่านี้กำหนดขึ้นอย่างไม่มีการเกณฑ์หรือเหตุผล และสังคมมนุษย์ติดต่อกันได้ด้วยระบบนี้ .

จากคำนิยามดังกล่าวพอสรุปได้ว่า เสียงพูดและคำพูดเท่านั้นนับเป็นภาษาที่แท้จริง การสื่อความหมายด้วยเสียงพูดและคำพูดต้องควบคู่ไปกับการฟัง ดังนั้นการเรียนรู้ภาษาพูดจึงเป็นการเรียนรู้ความหมายเสียงพูดหรือคำพูดและเกิดการรู้จำขึ้นในระบบการแปลความหมาย เมื่อได้ยินเสียงพูดหรือคำพูดที่มีลักษณะคล้ายกับที่เคยได้ยินมาก็จะสามารถแปลความหมายของเสียงนั้นว่า เป็นความหมายเดียวกับประสบการณ์ที่เคยเรียนรู้มา

เสียงที่มนุษย์เปล่งออกมาแต่ละคน แม้จะเป็นภาษาเดียวกันและพูดอย่างเดียวกันก็ตาม หากพิจารณาโดยหลักเกณฑ์ทางวิทยาศาสตร์จะพบว่า มีองค์ประกอบแตกต่างกันเช่น ความดัง, ลักษณะพฤติกรรมทางความถี่และความเข้มต่อเวลา ทั้งนี้ขึ้นกับความแตกต่างกันของอวัยวะที่ใช้ออกเสียงและวิธีการเปล่งเสียงของแต่ละคน อารมณ์และความรู้สึกก็มีส่วนทำให้ลักษณะการออกเสียงคำๆเดียวกันมีความแตกต่างกันออกไปเช่นกัน ลักษณะการพูดของแต่ละคนจะมีเอกลักษณ์เฉพาะตัว ด้วยเหตุนี้เราจึงสามารถจดจำเสียงและวิธีพูดของบุคคลที่เราคุ้นเคยได้เป็นอย่างดี และนอกจากนี้ยังพบว่า มนุษย์สามารถแยกแยะคำ ๆ เดียวกัน จากผู้พูดหลาย ๆ คนได้ทั้ง ๆ ที่มีความแตกต่างกันทั้งความดัง, ความถี่ และลักษณะการเปล่งเสียงซึ่งตัวแปรทั้งสาม มีการเปลี่ยนแปลงอย่างอิสระต่อกัน ซึ่งระบบที่สร้างขึ้นควรเป็นระบบที่สามารถรู้จำเสียงได้โดยไม่ขึ้นกับความดัง, ความถี่และระยะเวลาการเปล่งเสียง

วิธีการประยุกต์ความรู้ต่างๆที่เกี่ยวข้องกับเสียงมาใช้กับงานวิจัยเห็นจะไม่พ้นวิธีการสร้างโมเดลเลียนแบบการทำงานของอวัยวะที่ใช้ในการรับเสียง และตั้งสมมติฐานเพื่อการตีความหมายของเสียงที่รับเข้ามา น่าจะเป็นวิธีที่ดีที่สุดถึงแม้ว่าจะเป็นวิธีการพื้นฐานง่าย ๆ แต่ก็ใช้ได้ผลกับงานวิจัยอื่นๆมาแล้วมากมาย วิธีที่เลือกใช้ในการทำวิจัยเป็นวิธีนำผลจากการแปลงเสียงไปวิเคราะห์ โดยใช้อุปกรณ์ตรวจจับเสียงเพียงชิ้นเดียว คือ ไมโครโฟน สัญญาณที่ได้คือ แอมพลิจูดในแกนเวลา นำไปแปลงเป็นสัญญาณดิจิทัลแล้วผ่านโมเดลที่สร้างขึ้นจากสมมติฐานที่พยายามเลียนแบบระบบรับรู้จำของมนุษย์ ที่มีความถูกต้องตามแนวความคิด ของการรับฟัง และระบบนี้จะต้องสามารถเรียนรู้เสียงจากแหล่งกำเนิดเสียงอื่นๆ ได้โดยอิสระ

หลักการที่นำมาใช้รู้จำเสียงพูดในงานวิจัย

หลักการพื้นฐานอย่างมากอย่างหนึ่งที่เป็นธรรมชาติของมนุษย์ คือการเปรียบเทียบซึ่งถือเป็นพฤติกรรมที่เป็นสัญชาตญาณ การเปรียบเทียบเป็นการหาความแตกต่างระหว่างสองสิ่งอาจเป็นรูปธรรมหรือนามธรรมก็ได้ และสามารถเปรียบเทียบระหว่างสิ่งที่เหมือนหรือไม่เหมือนกันก็ได้ อาจเป็น : รูปธรรมกับรูปธรรม , นามธรรมกับนามธรรม , นามธรรมกับรูปธรรม หรือ รูปธรรมกับนามธรรม การเปรียบเทียบระหว่างสองสิ่ง จะต้องให้สิ่งใดสิ่งหนึ่งเป็นบรรทัดฐานจึงจะบอกข้อแตกต่างของอีกสิ่งหนึ่งได้ ฉะนั้นผู้วิจัยจึงเลือกวิธีการบันทึกโครงสร้างทางความถี่ต่อเวลาของเสียงพูดจำนวนหนึ่ง แล้วนำมาเปรียบเทียบกับคำพูดของผู้ที่ส่งงานจริง (ใช้หลักการเทียบหาคุณสมบัติจำเพาะที่สัมพันธ์กันระหว่างฐานข้อมูลหลัก(A)กับข้อมูลอ้างอิง(B) และฐานข้อมูลหลัก(A)กับข้อมูลเสียงทดสอบ(C)ดูภาพที่33) ที่ใช้กันอย่างกว้างขวางในการวิจัยในสาขาการประมวลผลสัญญาณดิจิทัล หมวดการรู้จำรูปแบบ (Pattern recognition) ก็คือ วิธีการแมชชีนโดยยึดหลักการในแนวความคิดที่ง่ายแต่ได้ผลเพื่อหลีกเลี่ยงความซับซ้อน

งานวิจัยนี้เป็นการเทียบหาความสัมพันธ์ของข้อมูลด้วยสมการทางสถิติร่วมกับนิวรัลเน็ตเวิร์ค มีการดำเนินการกับข้อมูลสเปกโตรแกรมเสียง 2 ส่วนคือ

1. กระบวนการปริโปรเซสซึ่ง กระทำเพื่อลดขนาดของข้อมูลสเปกโตรแกรมของเสียงที่ต้องการรู้จำและปรับข้อมูลให้เหมาะสมสำหรับใช้เป็นอินพุทของนิวรัลเน็ตเวิร์ค โดยพยายามหาคุณลักษณะบิตเฉพาะตัวของเสียงที่ต้องการทราบว่ามีความเหมือนหรือสอดคล้องกับเสียงนั้นเสียงนี้ในกลุ่มเสียงฐานข้อมูลหลัก(A) ก็เปอร์เซนต์(รูป ช. ในภาพที่10) ซึ่งต่อไปจะขอเรียกว่า “คุณลักษณะจำเพาะสัมพันธ์” โดยใช้สมการสหสัมพันธ์ (Correlation) หาคุณลักษณะจำเพาะสัมพันธ์ระหว่างข้อมูลเสียงอ้างอิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อิงกับฐานข้อมูลหลักเพื่อใช้เป็นข้อมูลสำหรับฝึกสอนให้กับนิวรัลเน็ตเวิร์ค และระหว่างข้อมูลเสียงทดสอบกับฐานข้อมูลหลักเพื่อใช้เป็นข้อมูลสำหรับทดสอบนิวรัลเน็ตเวิร์คในการหาอัตราการรู้จำ จุดเด่นของงานวิจัยนี้อยู่ที่วิธีการที่ใช้ในกระบวนการปริโปรเซสซึ่ง นั่นเอง ณ. จุดนี้ขออธิบายเหตุผลของการแบ่งกลุ่มข้อมูลเป็น 3 ชุด คือ ชุดข้อมูลเสียงอ้างอิง ชุดข้อมูลที่เป็นฐานข้อมูลหลักและชุดข้อมูลเสียงทดสอบ โดยชุดข้อมูลเสียงอ้างอิง เป็นชุดข้อมูลเสียงต่างๆ ที่ต้องการให้นิวรัลเน็ตเวิร์ครู้จัก แต่การนำไปฝึกสอนให้กับนิวรัลเน็ตเวิร์คโดยตรงนั้น ทำไม่ได้เนื่องจากขนาดของอินพุตแปรปรวนไปตามเวลา การแปลงเสียงแต่ละเสียงซึ่งไม่เท่ากันประการหนึ่ง และอีกประการหนึ่งสเปกโตรแกรมเสียงที่ใช้มีขนาดของข้อมูลมากเกินไปไม่เหมาะกับเครื่องมือที่ใช้และการนำมาใช้งานจริง จึงจำเป็นต้องลดขนาดของข้อมูลลง และปรับให้จำนวนข้อมูลไม่แปรตามช่วงระยะเวลาการแปลงเสียงโดยการหาคุณลักษณะจำเพาะที่สัมพันธ์กับกลุ่มเสียงกลุ่มหนึ่ง กลุ่มเสียงกลุ่มนั้นคือฐานข้อมูลหลักนั่นเอง(วิธีการหาคุณลักษณะจำเพาะสัมพันธ์ดังกล่าวเปรียบเสมือนเป็นฟังก์ชันหนึ่ง ซึ่งเมื่อป้อนข้อมูลอินพุตไปแล้ว จะตอบสนองออกมาเป็นแพตเทิร์นของข้อมูลชุดหนึ่งที่มีจำนวนน้อยลงและคงที่) การหาคุณลักษณะจำเพาะที่สัมพันธ์ ของกลุ่มเสียงทดสอบก็มีวิธีการเช่นเดียวกัน แต่คุณลักษณะจำเพาะสัมพันธ์ที่ได้จะนำไปทดสอบอัตราการรู้จำของนิวรัลเน็ตเวิร์ค

2. กระบวนการตัดสินใจ ใช้กรรมวิธีเลียนแบบโครงข่ายเซลล์สมองที่เรียกว่า โครงข่ายประสาทเทียม (Artificial Neural Networks; ANN) โดยจะฝึกสอนเน็ตเวิร์คในลักษณะ Supervised training เพื่อให้รู้จำแพตเทิร์นอ้างอิงในส่วนของ Training mode ตรวจสอบและตัดสินใจคุณลักษณะของเสียงที่นำมาทดสอบกับแพตเทิร์นอ้างอิงว่าเสียงที่ตรวจสอบเป็นเสียงใด ในส่วนของ Testing mode ซึ่งผู้วิจัยเลือกใช้อัลกอริทึมแบบเบคพรอบพาเก้น เนื่องจากมีลักษณะเหมาะสมกับงานที่ใช้ดังกล่าวในหัวข้อที่เกี่ยวข้องถัดไป

ขอบเขต

วิทยานิพนธ์ฉบับนี้ มีขอบเขตมุ่งศึกษาวิธีการรู้จำเสียงพูดที่เป็นกลุ่มของคำไทยโดดๆ หนึ่งคำ-หนึ่งพยางค์แบบไม่ขึ้นกับผู้พูด มีการกำหนดเสียงพูดให้ระบบรู้จำ ไม่เกิน 20 คำ สันนิษฐานในงานควบคุมด้วยเสียง โดยกำหนดเสียงตัวอย่างที่ใช้คือ “ศูนย์” , “หนึ่ง” , “สอง” , “สาม” , “สี่” , “ห้า” , “หก” , “เจ็ด” , “แปด” , “เก้า” , “สิบ” , “เปิด” , “ปิด” , “หมุน” , “ยก” , “วาง” , “ช่อง” , “ซ้าย” , “ขวา” , “ไฟ” โดยพยายามหาคุณสมบัติของเสียงในรูปของแพตเทิร์นของเสียงที่เรียกว่า “คุณลักษณะจำเพาะสัมพันธ์” และใช้กระบวนการแปลงข้อมูลสัญญาณเสียงเชิงเวลา ไปเป็นสเปกโตรแกรมสามมิติในแกน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของความถี่-เวลา-พลังงาน ขนาด Band width 125Hz-4KHz (ธันวา ศรีประโม่ง 2537:13-21) ใช้สมการสหสัมพันธ์ช่วยเปรียบเทียบหาระดับความสัมพันธ์ระหว่างสเปกโตรแกรมเสียงที่ต้องการทราบกับกลุ่มสเปกโตรแกรมเสียงหลัก เพื่อลดขนาดข้อมูลและหาคุณลักษณะจำเพาะสัมพันธ์ของเสียงว่าเสียงที่ต้องการทราบมีความสอดคล้องกับเสียงหลักเสียงนั้นกี่เปอร์เซ็นต์เสียงนี้กี่เปอร์เซ็นต์มีลักษณะเป็นแพตเทิร์นที่มีขนาดคงที่ มีการแบ่งข้อมูลสเปกโตรแกรมเสียงของผู้ออกเสียงออกเป็น 3 กลุ่มคือสเปกโตรแกรมเสียงหลัก สเปกโตรแกรมเสียงอ้างอิง และสเปกโตรแกรมเสียงทดสอบ กลุ่มข้อมูลคุณลักษณะจำเพาะสัมพันธ์(แพตเทิร์น)ที่ทำได้จากกลุ่มสเปกโตรแกรมเสียงอ้างอิงกับกลุ่มสเปกโตรแกรมเสียงหลักนำไปใช้ฝึกสอนให้นิวรัลเน็ตเวิร์ค(Neural Networks) ร่วมกับเป้าหมายที่ต้องการ กลุ่มข้อมูลคุณลักษณะจำเพาะสัมพันธ์(แพตเทิร์น) ที่ทำได้จากกลุ่มสเปกโตรแกรมเสียงทดสอบกับกลุ่มสเปกโตรแกรมเสียงหลักนำไปใช้ทดสอบนิวรัลเน็ตเวิร์ค เอ้าพุทที่ได้จะแสดงความถูกต้องของข้อมูลที่บ่งบอกถึงอัตราการรู้จำของระบบที่สร้างขึ้น แสดงโดยตารางสรุปอัตราการรู้จำที่ทดลองแต่ละแบบ การทดลองทำเป็นขั้นตอนในลักษณะที่เรียกว่า ออฟไลน์(Off line)

ขั้นตอนและวิธีการดำเนินงาน

1. เริ่มจากการศึกษาหาแนวทางจากแหล่งข้อมูลต่างๆที่เกี่ยวข้อง
2. ศึกษากระบวนการพูดและการฟังของมนุษย์โดยวิเคราะห์และสังเกตจากพฤติกรรม
3. กำหนดขอบเขตของงานวิจัย
4. เก็บข้อมูลเสียงเพื่อใช้สำหรับการวิจัย
5. ตั้งสมมติฐานวิธีการวิเคราะห์ และลองทดสอบด้วยโปรแกรมช่วยการคำนวณเพื่อหาความเป็นไปได้ แล้วเลือกวิธีการที่เหมาะสม
6. เขียนโปรแกรมเพื่อการวิเคราะห์และทดสอบ
7. ปรับปรุงแก้ไขให้เหมาะกับการใช้งาน
8. บันทึกและสรุปผลการทดสอบ

เครื่องมือที่ใช้ทำการทดลอง

ฮาร์ดแวร์

1. ไมโครคอมพิวเตอร์ส่วนบุคคล IBM PC/AT Compatible ใช้โปรเซสเซอร์ รุ่น 486 DX4-100MHz, หน่วยความจำหลัก 16Mbytes
2. Sound Card ขนาด 16 bit ยี่ห้อ Creative Sound Blaster รุ่น Discovery CD
3. Hard Disk Drive ขนาด 1024 Mbytes
4. Hi-fidelity audio tape recorder desk ของ Sony รุ่น TC-U30 และ Microphone

ซอฟต์แวร์

1. WINDOWS95 และ Microsoft office
2. โปรแกรมเทอร์โบซี
3. โปรแกรมสำเร็จรูป Dynamind 4.0

ข้อตกลงเบื้องต้น

สำหรับการวิเคราะห์เสียงคำไทยในงานวิจัยนี้ มิได้วิเคราะห์ในเชิงภาษาศาสตร์ หากแต่เป็นการวิเคราะห์ คุณลักษณะทางกายภาพของเสียงในเชิงวิทยาศาสตร์ฟิสิกส์ ที่เกี่ยวกับความดัง, ความถี่และคาบเวลาของเสียง เสียงคำที่ถูกวิเคราะห์แล้วจะได้ผลลัพธ์มีลักษณะเป็นคะแนนความสัมพันธ์กับกลุ่มเสียงตัวอย่าง(เหมือนกับเสียงนั้นเสียงนี้ในกลุ่มเสียงหลักก็เปอร์เซ็นต์) เป็นแพตเทิร์นที่มี จำนวนข้อมูลคงที่ไม่แปรปรวนไปตามระยะเวลาการเปล่ง กลุ่มตัวอย่างเสียงที่ใช้ในการทดลองเป็นเสียงของผู้ชายทั้งหมด ไม่จำกัดวัย ออกเสียงอิสระในสำเนียงหลากหลาย โดยไม่มีการฝึกฝนมาก่อนและเป็นธรรมชาติตามเอกลักษณ์ของการออกเสียงของแต่ละคน และการดำเนินการทดลองมีวัตถุประสงค์เพื่อหาเปอร์เซ็นต์ความถูกต้องในการรู้จำเสียงแบบไม่ขึ้นกับผู้พูดและแบบขึ้นกับผู้พูดเป็นกลุ่ม โดยใช้วิธีการหาคุณลักษณะจำเพาะสัมพันธ์ของเสียง(หาคุณลักษณะจำเพาะของเสียงที่นำมาสอบเทียบ(B-C)กับฐานข้อมูลเสียงหลัก(A)) ในกระบวนการปริโปรเซสซึ่ง ร่วมกับนิวรัล เน็ตเวอร์คในกระบวนการตัดสินใจ ในบทที่ 6 เป็นการนำเสนอวิธีการลดขนาดของข้อมูลสเปกโตรแกรมเสียงแต่ละเสียงในกระบวนการปริโปรเซสซึ่ง โดยใช้วิธีการหาค่าเฉลี่ยตามแกนความถี่จาก 128ค่า ให้ลดลงเหลือ 16ค่า และปรับลดระยะเวลาการเปล่งเสียงในแกนเวลาของสเปกโตรแกรมเสียงทุกเสียงให้เท่ากันโดยประมาณค่า ด้วยวิธีการกรองอินเตอร์โพลดิ้งโพลีโนเมียล(Lagrange interpolating polynomials) ปรับจำนวนดิสครีตสเปกตรัม(เฟรม) ให้มีจำนวน 10 ดิสครีตสเปกตรัมต่อหนึ่งสเปกโตรแกรมเสียง เท่ากันทุกสเปกโตรแกรม ได้เป็นข้อมูลขนาด 16x10 ข้อมูลสเปกโตรแกรมของทุกเสียงที่ผ่านกระบวนการปริโปรเซสซึ่งดังกล่าว จะถูกแบ่งเป็นสองส่วน ส่วนหนึ่งจะใช้ฝึกสอน และอีกส่วนหนึ่งใช้ทดสอบนิวรัลเน็ตเวอร์ค อัตราการรู้จำจะนำมาวิเคราะห์และเปรียบเทียบกับแบบวิธีหาคุณลักษณะสัมพันธ์ อย่างไรก็ตามผู้วิจัยขอทำความเข้าใจกับผู้อ่านว่า รายละเอียดและวิธีการที่นำเสนอในบทที่ 6 นั้น ไม่ใช่สาระสำคัญและวัตถุประสงค์ของการทำวิจัยนี้ หากแต่เป็นเพียงวิธีการหนึ่งที่สร้างขึ้นเพื่อนำผลลัพธ์มาเปรียบเทียบกับนั้น

บทที่ 2

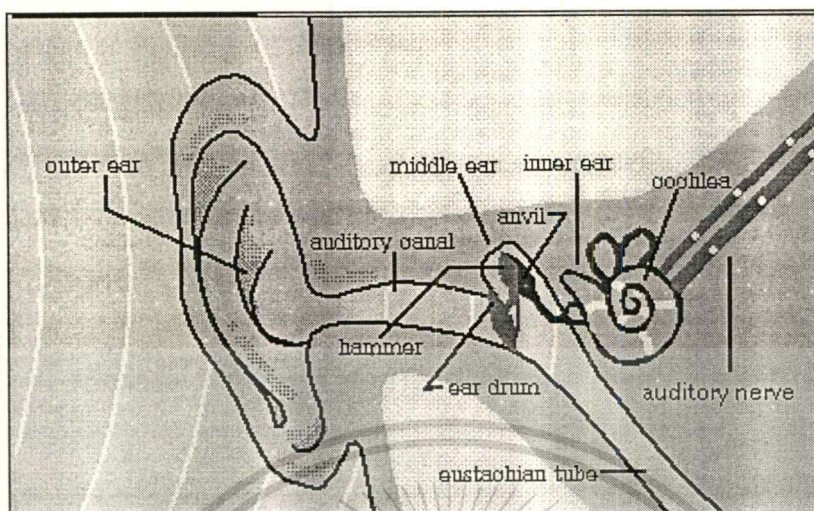
ระบบการได้ยินและการวิเคราะห์เสียงของมนุษย์

เราอาจแบ่งระบบการรับรู้ด้านเสียงของมนุษย์ได้สองส่วน คือ ระบบการได้ยิน และ ระบบการวิเคราะห์ความหมายของเสียง ระบบการได้ยินประกอบด้วยอวัยวะต่าง ๆ ที่ใช้ในการรับเสียง ทำหน้าที่รับเสียงจากสิ่งแวดล้อม แปลงเป็นสัญญาณอิมพัลส์ (Impulse) ผ่านเส้นประสาทเสียงไปยัง สมองส่วนรับรู้เสียง (Auditory) ที่ทำหน้าที่เป็นระบบวิเคราะห์ความหมายของเสียง

หูและการได้ยิน

หู เป็นอวัยวะที่ใช้ในการรับเสียงทุกเสียง ประกอบด้วย 3 ส่วน ได้แก่ หูส่วนนอก หูส่วนกลาง และหูส่วนใน แรงอัดและขยายของคลื่นเสียงที่เกิดขึ้นในธรรมชาติ จะเดินทางเข้ามาทางใบหู (Pinna), ช่องหู (Auditory cannal) ซึ่งเป็นหูส่วนนอก ผ่านหูส่วนกลางที่มีโพรงอากาศประกอบด้วย Ear drum และกระดูก 3 ชิ้นได้แก่ Hammer, Anvil และ Stirrup ทำหน้าที่ถ่ายทอดความสั่นสะเทือนสู่ของเหลวที่บรรจุอยู่ในอวัยวะรูปเปลือกหอยทาก (The snail-shaped cochlea) ซึ่งเป็นหูส่วนใน ของเหลวถ่ายทอดความสั่นสะเทือนเคลื่อนไปตามความยาวของ Cochlea (The basilar membrane) ภายในเยื่อ Basilar มีเซลล์ขน (Hair cell) ที่มีคุณลักษณะการตอบสนองความถี่แตกต่างกันนับพัน แผลอยู่ตั้งแต่ว่าจนถึงยอดของ Cochlea เมื่อเซลล์ขนเหล่านี้ถูกของเหลวทำให้ลักษณะเปลี่ยนไป จะเกิดเป็นสัญญาณอิมพัลส์ไฟฟ้า (Electrical impulses) กระตุ้นเป็นทอด ๆ ผ่านเส้นประสาทเสียง (Auditory nerve) เข้าสู่สมอง โดยเซลล์ขนจะมีคุณลักษณะการตอบสนองความถี่ต่ำที่ส่วนยอดของ Cochlea และที่ส่วนฐาน เซลล์ขนจะมีคุณลักษณะการตอบสนองความถี่สูง

ภาพที่ 1



แสดงหูส่วนนอก, ส่วนกลาง, ส่วนใน และส่วนประกอบต่างๆของหู

(GROLIER ELECTRONICS PUBLISHING, INC. 1995 Multimedia: "EAR")

คุณลักษณะของเสียงที่มนุษย์สามารถแยกแยะได้ประกอบด้วย ปริมาณ (Volume), พิต (Pitch) และ โทน (Tone) ปริมาณความดังของเสียงขึ้นอยู่กับแอมพลิจูดหรือความเข้มของคลื่นเสียง พิตเป็นความสัมพันธ์กับความถี่ของคลื่นเสียง หมายถึง จำนวนของคลื่นที่ผ่านจุดอ้างอิงต่อหนึ่งหน่วยเวลา เมื่อความถี่สูงขึ้น พิตก็จะมากขึ้น โทนหรือ Quality ของเสียงมีคุณสมบัติที่ซับซ้อนมากกว่า วอลลุ่ม และ พิต Quality จะแปรไปตามจำนวนและชนิดของ Overtone หรือ ฮาร์โมนิกส์ (Combinations of frequencies) โดยทั่วไปมนุษย์สามารถได้ยินเสียงความถี่ตั้งแต่ 30Hz-20KHz และช่วงแถบความถี่ที่ของเสียงพูดที่ใช้ในระบบสื่อสาร ที่สื่อสารกันแล้วเข้าใจ มีแบนด์วิดท์อยู่ในช่วง 300Hz-3.4KHz

ธรรมชาติการรู้จำเสียง

ระบบการรู้จำเสียงของมนุษย์ มีความละเอียดอ่อนซับซ้อน มีการเรียนรู้และปรับปรุงความสามารถไปพร้อมกับการพัฒนาของร่างกายตั้งแต่เยาว์วัย ระดับการรู้จำจะดำเนินไปตามขั้นตอนของมันจนกระทั่งมีความเชี่ยวชาญในการฟัง ประกอบกับมีโครงสร้างที่มีประสิทธิภาพสอดคล้องกับระบบการได้ยินทำให้สามารถแยกแยะคำและเสียงที่ซับซ้อนได้อย่างดี ความชำนาญในการฟังเกิดจากการ เรียนรู้

ฝึกฝนและมีการปรับปรุงอยู่ตลอดเวลา จากการศึกษาพฤติกรรมพบว่า การวิเคราะห์และรู้จำเสียงของมนุษย์มีคุณลักษณะและความสามารถหลายอย่าง พอสรุปได้ดังนี้คือ

1. การตีความหมายของคำที่ได้ยินขึ้นกับสถานการณ์, เวลา, สถานที่, ความรู้สึก, เรื่องที่กำลังสนใจ
2. สามารถรับรู้ความหมายเดียวกันได้แม้ความดัง, ความถี่, พิต, ความยาวของเสียงที่เปลี่ยนแปลงต่างกัน
3. ขณะสนใจจะวิเคราะห์เสียงใดเสียงหนึ่ง ความสามารถวิเคราะห์เสียงอื่นรอบข้างจะด้อยลง
4. สามารถเลือกฟังเฉพาะเสียงที่ต้องการฟังเพื่อตีความหมายได้
5. สามารถวิเคราะห์ได้ทั้งเสียงภาษาของบุคคลทั่วไปและบุคคลเฉพาะ
6. สามารถรับรู้ทิศทางของแหล่งกำเนิดเสียงได้
7. สามารถตีความได้โดยไม่ขึ้นอยู่กัขนาดของแหล่งกำเนิดเสียงพูด
8. รับรู้อารมณ์, ความรู้สึก, น้ำเสียงของแหล่งกำเนิดเสียงพูดได้
9. การรับรู้มีความถูกต้องมากขึ้นเมื่อคู่สนทนาได้เห็นลีลา, สายตา ท่าทางประกอบการพูด
10. สามารถแยกแยะคุณสมบัติของเสียงที่มีลักษณะใกล้เคียงกันอย่างมากได้อย่างมีประสิทธิภาพ
11. มีการเรียนรู้ตลอดเวลาและปรับตัวให้เข้ากับสิ่งแวดล้อมได้อย่างรวดเร็ว (ก่อนจะวิเคราะห์เสียงใด ๆ ได้ถูกต้อง ต้องมีประสบการณ์และการเรียนรู้ก่อน)
12. ปรับระดับความดังของการรับฟังให้อยู่ในระดับที่สมองสามารถนำไปวิเคราะห์ได้อย่างอัตโนมัติ โดยไม่คำนึงแหล่งกำเนิดเสียงว่าจะดังหรือค่อยเพียงใด ในขอบเขตหนึ่ง
13. มีการทำงานสอดคล้องประสานกันจิตใจที่สามารถควบคุมการวิเคราะห์จากการได้ยินได้

ดังนั้นการสร้างเครื่องจักรให้มีความสามารถรู้จำเสียงได้อย่างมีประสิทธิภาพเท่าเทียมมนุษย์เป็นเรื่องค่อนข้างยาก หรือหากจะทำเครื่องจักร จะต้องทำให้มีลักษณะการเลียนแบบพฤติกรรมของมนุษย์คือ “เรียนรู้ ฝึกฝน มีการปรับปรุงอยู่ตลอดเวลาและต้องกระทำอย่างอัตโนมัติ” การวิจัยดังกล่าวยังคงต้องอาศัยเวลาอีกพอสมควร เพราะต้องใช้คอมพิวเตอร์ที่มีความเร็วการคำนวณสูงมาก ในขั้นตอนการฝึกฝน ซึ่งคงต้องใช้เวลามากหากต้องมีการฝึกฝนบ่อย ๆ และเมื่อต้องการให้ระบบรู้จักคำมากขึ้น ประสิทธิภาพและความเร็วก็จะลดลง นอกจากนี้ระบบก็ยังต้องการอัลกอริทึมหลายๆแบบร่วมกันที่ให้ความเร็วและความถูกต้องสูง หากพิจารณาคุณลักษณะสมบัติการวิเคราะห์เสียงของมนุษย์พบว่ามีความละเอียดอ่อน แม่นยำสูง ลึกซึ้งทำงานร่วมกับจิตใจซึ่งเป็นการยากที่จะสร้างระบบการวิเคราะห์ให้มีความสามารถสูงดังเช่นระบบการวิเคราะห์เสียงของมนุษย์ การวิจัยจึงมุ่งประเด็นการรู้จำเสียงในจำนวนจำกัด, กลุ่มทดสอบใช้งานและกลุ่มตัวอย่างที่ค่อนข้างแคบแต่กำหนดให้เพียงพอต่อการใช้งานแบบเฉพาะกิจ

โดย มู ง เน น ความ ถูก ต ้อง เป็ น ห ลั ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การสร้างระบบการรู้จำเสียงพูดคำไทย

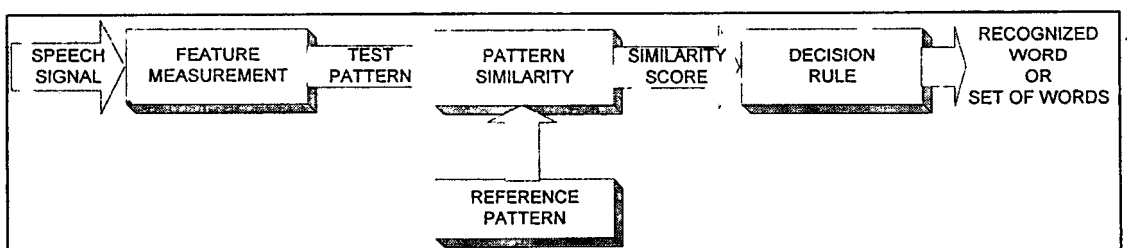
การวิเคราะห์เงื่อนไขการสร้างระบบการรู้จำที่ใช้ในการทดลอง

การสร้างระบบการรู้จำเสียงพูดคำไทยแบบไม่ขึ้นกับผู้พูด ผู้วิจัยเริ่มต้นจากการศึกษาคุณลักษณะทั่วไป คุณลักษณะเฉพาะตัวของเสียงคำพูด ระบบการได้ยิน การแปลความและวิเคราะห์ความหมาย ธรรมชาติการเปล่งเสียงของมนุษย์ ตลอดจนการศึกษาหาวิธีการวิเคราะห์เปรียบเทียบที่เหมาะสมเพื่อใช้ในแบบรู้จำที่จะสร้างขึ้น จากการศึกษาธรรมชาติการสื่อสารผ่านระบบเสียงและการรับฟังของมนุษย์ พอสรุปได้ว่าหากจะสร้างระบบรู้จำที่สามารถรู้จำเสียงแบบไม่ขึ้นกับผู้พูดได้ ระบบรู้จำจะต้องสามารถวิเคราะห์เปรียบเทียบข้อมูลเสียงได้โดยไม่ขึ้นกับความดัง ความถี่ และระยะเวลาของเสียงคำหนึ่งๆ ที่เปล่งออกมา ซึ่งต้องทำได้ดังดังที่กล่าวเป็นอย่างย่อ

ระบบการรู้จำที่สร้างขึ้นใช้ในการทดลอง

การรู้จำเสียงพูดนั้นจะเกี่ยวข้องกับการตัดสินใจระหว่างแบบทดสอบ (Test pattern) กับแบบอ้างอิง (Reference pattern) หรือที่เรียกว่า Template ซึ่งจะเป็นการวิเคราะห์ถึงความสัมพันธ์ของตัวแปร (Parameter) ที่ใช้เป็นรูปแบบทั้งสอง เพื่อที่จะระบุว่าแบบทดสอบที่นำมาทดสอบนั้นมีความสัมพันธ์กับแบบอ้างอิงใดมากที่สุด จากนั้นจึงจะนำไปสู่การตัดสินใจต่อไป... (ระพีพัฒน์ เพ็ญศิริ 2538:16-18)

ภาพที่ 2



แสดงโครงสร้างระบบการรู้จำแบบ Isolated word recognition (Rabiner and Levinson, 1981)

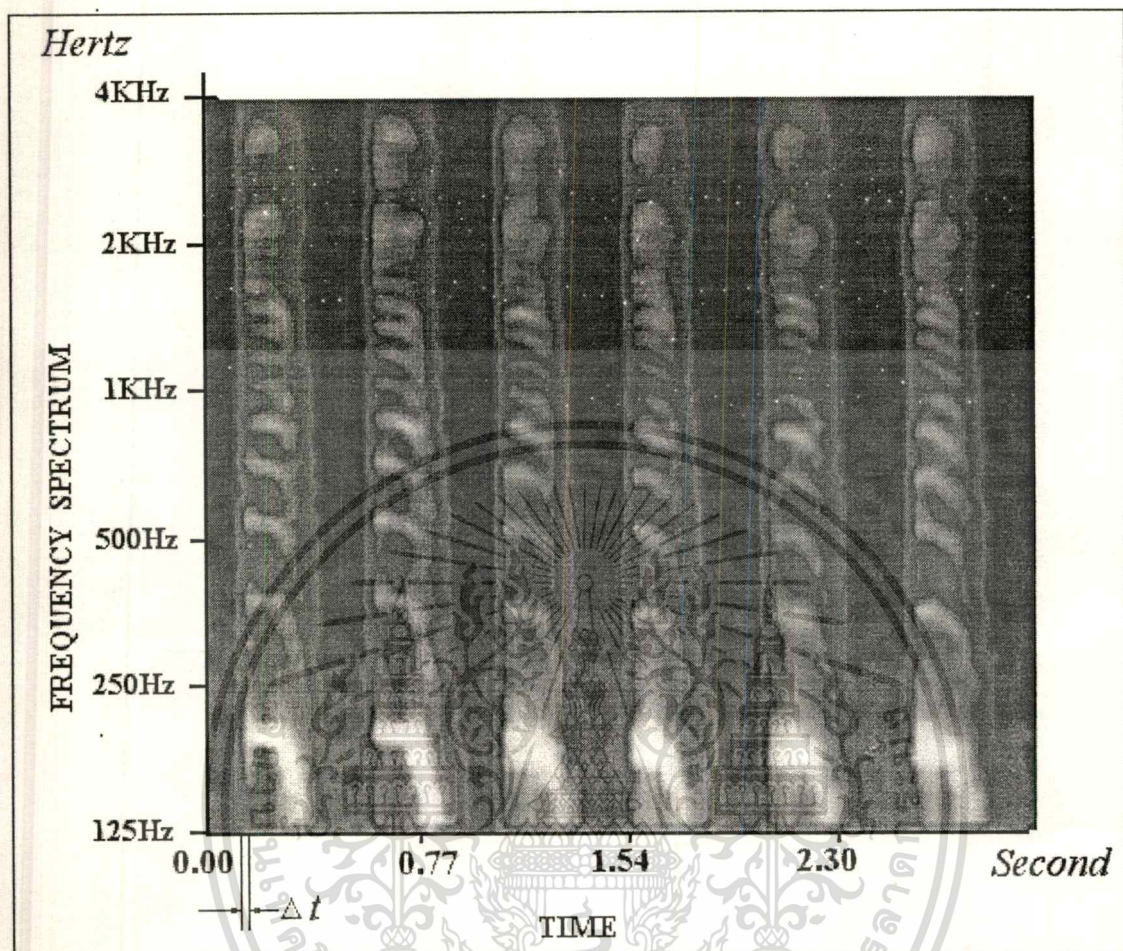
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โครงสร้างระบบการรู้จำที่สร้างขึ้นเป็นแบบที่คล้ายกับ Isolated word recognition (Rabiner and Levinson, 1981) ดังภาพที่ 2 โดยในส่วนของ Feature measurement จะรวมเอา Pre-processing เข้าไปด้วยคือ A/D, FFT เพื่อสร้างเป็นดีสคริตสเปกตรัมของสเปกโตรแกรมเสียงที่ใช้ (สเปกโตรแกรมของเสียงใดๆ ประกอบด้วยดีสคริตสเปกตรัมหลายๆ ดีสคริตมาวางเรียงกัน สำหรับงานวิจัยนี้บางครั้งจะดีสคริตสเปกตรัมว่า “เฟรม”) ส่วนของ Pattern similarity เป็นการเปรียบเทียบระหว่างแพตเทิร์นที่ต้องการทราบค่ากับแพตเทิร์นเสียงหลัก ซึ่งเลือกใช้วิธีการเปรียบเทียบระดับความสัมพันธ์ของข้อมูลด้วยสมการสหสัมพันธ์ วิธีการนี้ทำให้ได้ผลลัพธ์เป็นคุณลักษณะเฉพาะตัวของเสียงที่ต้องการทราบเทียบกับฐานข้อมูลเสียงหลัก(แสดงดังภาพที่10) และขนาดของข้อมูลลดลงอย่างมาก ต่อไปผู้วิจัยขอเรียกวิธีนี้ว่า “คุณลักษณะจำเพาะสัมพันธ์” และในส่วนของ Decision rule จะใช้การตัดสินใจด้วยนิรวัลเน็ตเวิร์คที่ผ่านการฝึกสอนด้วยแพตเทิร์นคุณลักษณะจำเพาะสัมพันธ์อ้างอิงมาแล้ว ด้วยอัลกอริทึมแบคพรอบพาเก้น สาเหตุที่เลือกอัลกอริทึมแบบ แบคพรอบพาเก้น เพราะมีการฝึกสอนแบบ Supervised Training ซึ่งเป็นอัลกอริทึมที่สอดคล้องกับงานที่ใช้ คือการรู้จำเสียงคำใดๆ ต้องมีแม่แบบเสียงก่อน แล้วจึงเรียนรู้ไปตามแม่แบบนั้น เมื่อมีเสียงทดสอบที่มีลักษณะคล้ายกับลักษณะที่เคยเรียนรู้เข้ามาก็สามารถตัดสินใจได้ว่าเสียงทดสอบนั้นน่าจะเป็นเสียงใด

การแก้ไขปัญหามหาการเปรียบเทียบเสียงที่มีความดัง,ระยะเวลาการเปล่งเสียงและความถี่ แตกต่างกัน

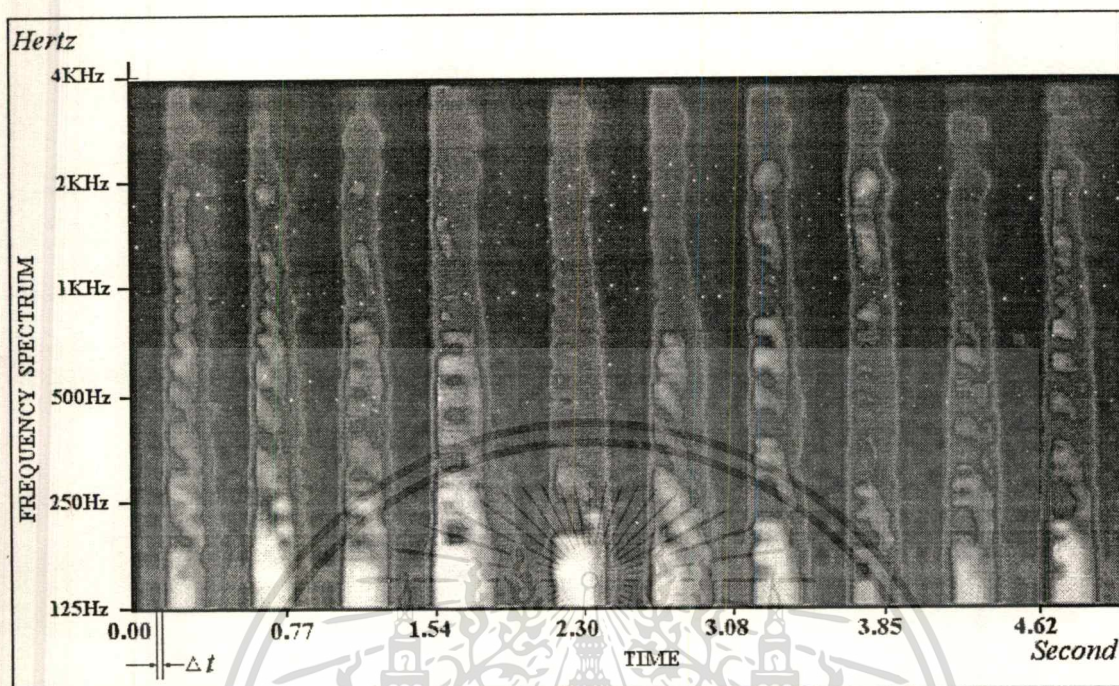
จากการศึกษาคุณลักษณะทางความถี่ของเสียงพูดตัวอย่างจากสเปกโตรแกรมพบว่าคำ คำเดียวกัน ที่พูดโดยผู้พูดคนละคน มีลักษณะของแถบความถี่แตกต่างกัน ทั้งนี้เนื่องจาก ลักษณะของช่องปาก, หลอดเสียง, และลักษณะการออกเสียง ของแต่ละคนมีความแตกต่างกัน รวมทั้งความดังของเสียง และระยะเวลาที่เปล่งเสียงของแต่ละคนเปล่งออกมาก็ไม่เท่ากัน หรือแม้แต่คำคำเดียวกันที่ผู้พูดคนเดียวเปล่งออกมาแต่ละครั้ง ก็ไม่เหมือนกันเสียทีเดียวดังสังเกตได้จากภาพที่ 3 และ 4

ภาพที่ 3



แสดงสเปกโตรแกรมของเสียง “เก๋า” ที่เปล่งเสียงโดยผู้พูดคนเดียวกัน 6 ครั้ง

ภาพที่ 4



แสดงสเปกโตรแกรมของเสียง “เก๋” ที่เปล่งเสียงโดยผู้พูด 10 คน

ในการวิเคราะห์การเลือกวิธีการเปรียบเทียบที่ทำได้โดยไม่ขึ้นกับตัวแปรทั้งสาม ได้แก่ ความดัง ระยะเวลาของการออกเสียง และความถี่ ซึ่งมีการเปลี่ยนแปลงอย่างอิสระต่อกัน งานวิจัยนี้ขอเสนอวิธีการแก้ปัญหาดังกล่าวดังต่อไปนี้

1 ความดังของเสียง มีวิธีการเปรียบเทียบทางสถิติวิธีหนึ่งที่น่าสนใจคือ วิธีการเปรียบเทียบแบบสหสัมพันธ์ (Correlation) ซึ่งเป็นการเปรียบเทียบลักษณะความสัมพันธ์ของการเปลี่ยนแปลงของชุดข้อมูล โดยไม่คำนึงถึงระดับการเปลี่ยนแปลงของขนาดของข้อมูล คุณสมบัตินี้ตรงกับความต้องการเพราะสอดคล้องกับความเป็นจริง คือ ธรรมชาติการเปล่งเสียงของแต่ละคน มีระดับพลังงาน (ความดัง) ไม่เท่ากัน ดังนั้น ถ้าจะเปรียบเทียบความสัมพันธ์ของสเปกตรัมเสียงสองเสียง ซึ่งมีความดังแตกต่างกันด้วยวิธีปกติ จะต้องปรับเสียงทั้งสองให้มีสเกลความดังอยู่ในระดับเดียวกันเสียก่อน (แบบ Max-Min Scale) แล้วจึงทรานส์ฟอร์มไปเป็นสเปกตรัมของเสียงแล้วเปรียบเทียบ แต่ถ้าเปรียบเทียบความสัมพันธ์โดยการหาค่าสหสัมพันธ์ระหว่างข้อมูล แล้วจะสามารถข้ามขั้นตอนดังกล่าวไปได้ เพราะคุณสมบัติเฉพาะของวิธีการ เป็นการหาระดับความสัมพันธ์ ของลักษณะการเปลี่ยนแปลงของข้อมูล โดยไม่ขึ้นกับขนาด (Maxnitude) ของข้อมูล ดังนั้นหากใช้วิธีการดังกล่าวจะสามารถหาระดับความสัมพันธ์ได้โดยไม่ต้องคำนึงถึงระดับความดังของข้อมูลเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2 **ระยะเวลาที่แปลงเสียงแต่ละพยางค์** การแก้ปัญหาที่เกิดจากระยะเวลาการแปลงเสียงระหว่างเสียงทดสอบและเสียงอ้างอิงไม่เท่ากันนั้นอาศัยประโยชน์จากวิธีการเปรียบเทียบแบบสหสัมพันธ์ระหว่างแพตเทอร์น ของสเปกตรัมเสียงทดสอบกับเสียงอ้างอิง โดยการหาระดับความสัมพันธ์ที่ทำการแบบสเปกตรัมต่อสเปกตรัม (เนื่องจากระยะเวลาการแปลงเสียงแต่ละพยางค์ไม่เท่ากัน จำนวนสเปกตรัมจึงไม่เท่ากัน) และหยุดเมื่อการเปรียบเทียบกระทำไปถึงสเปกตรัมสุดท้ายของเสียงที่มีระยะเวลาการแปลงเสียงที่น้อยกว่า จากนั้นจึงหาค่าเฉลี่ยสหสัมพันธ์ของแต่ละสเปกตรัมของเสียงทั้งสอง เหตุผลที่กำหนดจุดสิ้นสุดของการเปรียบเทียบดังกล่าวคือ

2.1 การแปลงเสียงของเสียงเดียวกันแต่ละครั้งมีลักษณะคล้ายคลึงกัน และระยะเวลาการแปลงเสียงใกล้เคียงกัน หากการเปรียบเทียบสเปกตรัมขาดไปบ้างบางส่วนจะส่งผลกระทบต่อผลลัพธ์ในแต่ละเฟรมยังต้องนำไปหาค่าเฉลี่ยสหสัมพันธ์รวมของเสียงที่เปรียบเทียบอีกครั้งสำหรับกรณีเสียงทดสอบและเสียงอ้างอิงเป็นเสียงเดียวกัน หากระยะเวลาการแปลงเสียงแตกต่างกันมากขึ้นก็จะส่งผลกระทบต่อมากขึ้น แต่เนื่องจากลักษณะความคล้ายคลึงกันยังคงมีอยู่จึงไม่กระทบกระเทือนต่อค่าเฉลี่ย ทำให้ผลการเปรียบเทียบยังอยู่ในเกณฑ์เชื่อถือได้

2.2 กรณีเสียงทดสอบและเสียงอ้างอิงเป็นคนละเสียง แต่มีระยะเวลาการแปลงเสียงใกล้เคียงกัน ผลการเปรียบเทียบก็จะชี้วัดว่าระดับความสัมพันธ์ของข้อมูลต่ำเนื่องจากข้อมูลขาดความคล้ายคลึงกันตั้งแต่ต้น และหากระยะเวลาการแปลงเสียงแตกต่างกันก็จะยิ่งส่งผลกระทบต่อระดับความสัมพันธ์ของข้อมูลต่ำลงไปอีก

3 **ความถี่** เสียงคำๆ เดียวกันที่พูดโดยแต่ละคนมีความถี่มูลฐานไม่เท่ากัน การปรับความถี่เสียงใด ๆ ให้เลื่อนไปอยู่ในระดับเดียวกัน แล้วดำเนินการเปรียบเทียบ เป็นการบิดเบือนข้อมูลให้ผิดเพี้ยนไปจากความเป็นจริง และมีปัญหาอยู่ที่เสียงแต่ละเสียงมีความถี่ที่เปลี่ยนแปลงตลอดเวลา เป็นการยากที่จะหาความถี่มูลฐานที่แน่นอนของแต่ละเสียงได้ จึงหลีกเลี่ยงไปใช้วิธีการการนำแพตเทอร์นทดสอบไปเปรียบเทียบหาระดับความสัมพันธ์กับแพตเทอร์นอ้างอิงที่เตรียมไว้จำนวนมาก ซึ่งมีโอกาสพบกับแพตเทอร์นอ้างอิงที่ใกล้เคียงกัน แต่ต้องมีการแบ่งกลุ่มชาย-หญิง เนื่องจากความถี่มูลฐานของชาย-หญิงแตกต่างกันค่อนข้างเด่นชัด เพื่อความสะดวกในการทดลองจะเลือกทำเฉพาะเสียงผู้ชายไทยทั่วไปก่อน ส่วนการดำเนินการกับข้อมูลกลุ่มเสียงอื่นจะแยกกันเป็นโมดูล แต่ยังคงใช้วิธีการเดียวกันได้

การนำนิรลเน็ตเวิร์คมาใช้ในระบบการรู้จำเสียงพูดภาษาไทย

อัลกอริทึมการเทรนนิ่ง แบบแบคพรอบพาเกชันของนิรลเน็ตเวิร์ค (Backpropagation neural networks;BPNN) เป็นการเรียนรู้แบบ Supervised training ซึ่งต้องมีเป้าหมายที่ต้องการ เช่นเดียวกับการเรียนรู้ภาษาของมนุษย์ ต้องมีภาษาใดภาษาหนึ่งเป็นแม่แบบ และหากต้องการรู้ภาษาอื่น ก็ต้องเริ่มเรียนรู้กันใหม่เพราะภาษาไม่ใช่สัญชาตญาณ ด้วยเหตุนี้จึงเลือกนำเอาอัลกอริทึมนี้มาใช้เฉพาะ ส่วนการตัดสินใจ คุณลักษณะจำเพาะของเสียงที่เป็นอินพุตว่า น่าจะเป็นเสียงใดเท่านั้น การจะนำ BPNN มาใช้รู้จำโดยตรงทั้งระบบ ยังติดปัญหาอยู่ที่ความแปรปรวนของขนาดของแพตเทิร์นข้อมูลเสียง และขนาดของเน็ตเวิร์คก็มีขนาดใหญ่เกินไปจึงจำเป็นต้องปรับข้อมูลให้เหมาะสมเสียก่อนเพื่อให้เน็ตเวิร์ค ทำงานได้อย่างรวดเร็วและมีประสิทธิภาพ สอดคล้องเหมาะสมกับเครื่องมือที่ใช้

การสร้างแพตเทิร์นอ้างอิงจากสเปกโตรแกรมเสียงกลุ่มอ้างอิง

ข้อมูลเสียงที่ใช้ในการทดลองจะถูกแบ่งเป็น 3 ส่วนคือส่วนที่นำไปใช้เป็นฐานข้อมูลหลัก(A), ข้อมูลอ้างอิง(B) และส่วนที่เป็นข้อมูลทดสอบ(C) ข้อมูลเสียงที่ได้จากการ Quantizing และ Coding จากสัญญาณอนาล็อกทุกชุด จะถูกทรานส์ฟอร์มไปเป็นสเปกโตรแกรม ความถี่-เวลา-พลังงาน (ธันวา ศรีประโม 2537: 13-21) ได้เป็นกลุ่มสเปกโตรแกรมเสียงอ้างอิงและสเปกโตรแกรมเสียงทดสอบ ในการสร้างแพตเทิร์นอ้างอิงจะนำสเปกโตรแกรม เสียงกลุ่มอ้างอิงกลุ่มเดียวกันมาเทียบหาความสัมพันธ์ของข้อมูลกับเสียงหลัก แบบเสียงหนึ่งเสียงนำไปเทียบหาระดับความสัมพันธ์โดยเทียบกับเสียงหลักทั้งหมด โดยใช้สมการสหสัมพันธ์ (Correlative equation) จะได้ผลลัพธ์เป็นสัมประสิทธิ์สหสัมพันธ์ของเสียงนั้นกับเสียงทั้งหมด มีจำนวนเท่ากับเสียงอ้างอิงที่นำไปเทียบ จากนั้นจะจัดเรียงสัมประสิทธิ์สหสัมพันธ์เสียใหม่ แบบมากไปน้อยโดยมีฟิลต์ชื่อประจำเสียงนั้นๆ เป็นตัวแปรตาม เมื่อจัดเรียงเสร็จจะเลือกสัมประสิทธิ์สหสัมพันธ์ที่มีค่ามากที่สุดตามลำดับออกมาประมาณ 10 เปอร์เซนต์ ของจำนวนเสียงทั้งหมด แล้วนับคะแนนความถี่ในฟิลต์ชื่อประจำเสียงแต่ละชื่อที่ซ้ำกัน จะได้เป็นแพตเทิร์นอ้างอิง (หรือคุณสมบัติเฉพาะตัวของเสียงนั้นเทียบกับเสียงตัวอย่างทั้งหมด เรียกว่า "คุณลักษณะจำเพาะสัมพันธ์") จำนวนของแพตเทิร์นอ้างอิงที่ได้จึงมีจำนวนเท่ากับจำนวนสเปกโตรแกรมเสียงอ้างอิงนั่นเอง พิจารณาที่ (๗) ในภาพที่ 10

สหสัมพันธ์ (Correlation)

ค่าระดับความสัมพันธ์ของข้อมูล 2 ชุด เรียกว่า สัมประสิทธิ์ของสหสัมพันธ์ (Correlative coefficient ; r) สามารถหาได้จากเทอมของ Z-Scores ดังนี้

$$r_{xy} = \sum_{i=1}^N \frac{Zx_i Zy_i}{N} \dots\dots\dots(3.1)$$

เมื่อ r_{xy} คือ สัมประสิทธิ์ของสหสัมพันธ์ของข้อมูล x_i และ y_i

x_i คือ ข้อมูลในแกนความถี่ของสเปกโตรแกรมเสียงชุดที่ 1 (ทดสอบ)

y_i คือ ข้อมูลในแกนความถี่ของสเปกโตรแกรมเสียงชุดที่ 2 (อ้างอิง)

N คือ จำนวนข้อมูลทั้งหมดที่มี x_i, y_i เป็นสมาชิก (โดยข้อมูล x_i, y_i ทั้ง 2 ชุดมีจำนวนเท่ากัน ในการทดลองมีข้อมูลจำนวน 128 ค่า)

Zx_i และ Zy_i คือค่า Z-Scores เป็นระยะทางระหว่างค่าข้อมูลกับค่า Mean ของชุดสมการในเทอมของ Standard deviation Sx_i

$$\text{โดย } Zx_i = \frac{x_i - \bar{X}}{Sx} \dots\dots\dots(3.2)$$

$$\text{และ } Zy_i = \frac{y_i - \bar{Y}}{Sy} \dots\dots\dots(3.3)$$

$$\text{เมื่อ } Sx = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}} \dots\dots\dots(3.4)$$

$$\text{และ } Sy = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N}} \dots\dots\dots(3.5)$$

จากสมการ 3.1 ค่า r_{xy} จะมีค่าอยู่ในระหว่าง -1 ถึง +1 และ

$$0 \leq |r_{xy}| \leq 1$$

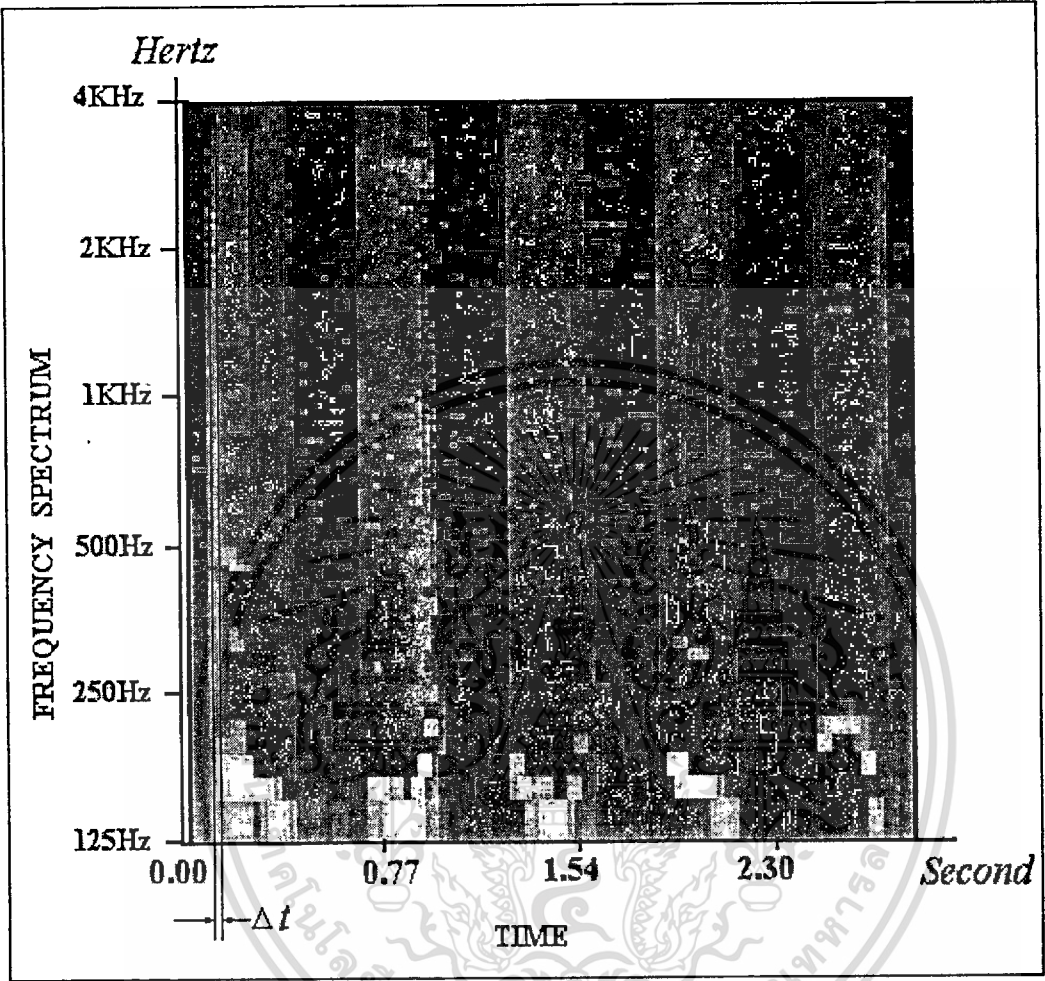
พอตีความหมายได้ดังนี้คือ ถ้าค่า $|r_{xy}| = 1$ หมายถึง ข้อมูลทั้ง 2 ชุด มีความสัมพันธ์กันทุกจุดข้อมูล ซึ่งในทางปฏิบัติแล้วมีโอกาสน้อยมาก

ในการทดลองจะพิจารณาเฉพาะ ค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสเปกโตรแกรมไม่ต่ำกว่า 0.70 ซึ่งจะกล่าวในหัวข้อวิธีการหาคุณลักษณะจำเพาะสัมพันธ์ และลดขนาดข้อมูลเสียง

การเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสเปกโตรแกรมเสียงสองชุด

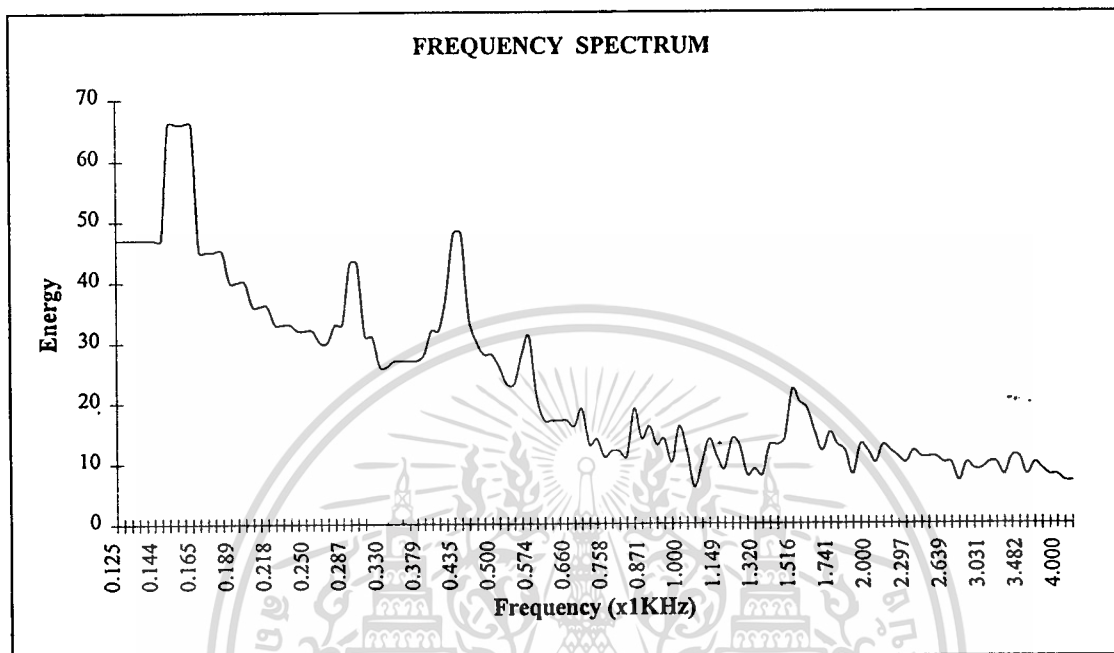
สเปกโตรแกรมเสียง หมายถึง การนำสเปกตรัมของเสียงที่ได้จากการทรานส์ฟอร์มสัญญาณเสียงในเชิงเวลา ที่เวลา Δt ใด ๆ มาเรียงต่อกัน ในกรณีนี้เราจะสนใจ สเปกโตรแกรมเสียงนับตั้งแต่เริ่มเปล่งเสียงจนสิ้นสุดการเปล่งเสียง ซึ่งเสียงที่ใช้ทดสอบกำหนดเป็นเสียงคำไทยแบบ 1 คำ 1 พยางค์ หากพิจารณาสเปกโตรแกรม 3 มิติ จากภาพที่ 5 พบว่า ชุดข้อมูลเสียงใด ๆ ประกอบด้วย Discrete ของระดับพลังงาน เป็นสเปกตรัมความถี่เสียง (แสดงเป็นแถบสี) มาเรียงต่อกัน โดยแต่ละ Discrete เป็นค่าเฉลี่ยในช่วงเวลา Δt แกนทางแนวนอน คือแกนเวลา โดยแสดงตั้งแต่เวลาที่ 0 ถึง 3.07 วินาที แต่ละ 1 ช่วงเวลา $\Delta t = n$ แกนแนวตั้งเป็นแกนของแถบความถี่ ตั้งแต่ 125 Hz ไปจนถึง 4 KHz จำนวน 128 ค่า, จัดเรียงสเกลแถบความถี่แบบ Logarithm ส่วนระดับพลังงาน แสดงโดยระดับความเข้มของสีเทาโดยเรียงจากความสว่างมาก(ขาว)ไปจนถึงดำ จากระดับพลังงานสูงไปหาดำตามลำดับ จากภาพที่ 5 แสดงความเข้ม 16 ระดับ (โดยแปลงจากค่าที่คำนวณได้ผ่าน Fitting scale ที่ 0-16 ระดับ แล้วนำมาแสดงผลโดยย่อสีแทนระดับพลังงาน) เมื่อเปรียบเทียบระยะเวลาการเปล่งเสียงของชุดข้อมูลเสียงแต่ละชุด จะพบว่ามีเวลาการเปล่งเสียง แตกต่างกันไปบ้างตามแต่ผู้พูด ในช่วงเวลาส่วนย่อย (Discrete) อาจเขียนเป็นกราฟของแถบความถี่ได้ภาพที่ 6

ภาพที่ 5



แสดงสเปกโตรแกรมของชุดข้อมูลเสียง “หนึ่ง” ถึง “ห้า” ที่เปล่งจากคนเดียวจากซ้ายไปขวาตามลำดับ

ภาพที่ 6



แสดงสเปกตรัมของแถบความถี่(เฟรม) ของเสียงพูด “หนึ่ง” ช่วงเวลา Δt ที่ 6

การเทียบชุดข้อมูลเสียงทั้งสองจะทำทีละเฟรม (หรือ Discrete หรือ Δt) แบบเฟรมต่อเฟรม โดยเฟรมที่ 1 ของสเปกโตรแกรมเสียงชุดที่1 (ทดสอบ) ก็เปรียบเทียบกับเฟรมที่1 ของสเปกโตรแกรมเสียงชุดที่ 2 (หลัก) ดังสมการ (3.6)

$$r_{xy}_p = \sum_{i=1}^N \frac{Zx_i Zy_i}{N} \dots\dots\dots(3.6)$$

เมื่อ r_{xy}_p คือค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากหนึ่งเฟรมของสเปกโตรแกรม x_i, y_i

Zx_i, Zy_i คือ ค่า Z-Scores ได้จากสมการที่ (3.2) และ (3.3)

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

x_i คือ ข้อมูลในแกนความถี่ของสเปกโตรแกรมเสียงชุดที่ 1 (ทดสอบ)

y_i คือ ข้อมูลในแกนความถี่ของสเปกโตรแกรมเสียงชุดที่ 2 (อ้างอิง)

p คือ เฟรม(Discrete) ของสเปกโตรแกรมที่กำลังพิจารณา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การหาค่าสัมประสิทธิ์สหสัมพันธ์ จะดำเนินการทุกเฟรมของสเปกโตรแกรมเสียงทั้งสอง แต่เนื่องจากจำนวนเฟรมของเสียงทั้งสองไม่เท่ากันจึงกำหนดขอบเขต ของการหาค่าสัมประสิทธิ์สหสัมพันธ์ โดยให้สิ้นสุดการพิจารณาที่เฟรมสุดท้ายของข้อมูลเสียงใดที่ถึงก่อน ดังสมการที่ (3.7)

$$F = I \text{ เมื่อ } J \geq I$$
$$F = J \text{ กรณีอื่น(3.7)}$$

- เมื่อ F คือ จำนวนเฟรมของข้อมูลเสียงพยางค์ 2 ชุดที่นำมาเปรียบเทียบได้จากสมการที่ 3.7
- I คือ จำนวนเฟรมของข้อมูลเสียงคำชุดที่หนึ่ง (ข้อมูลหลัก)
- J คือ จำนวนเฟรมของข้อมูลเสียงคำชุดที่สอง (ข้อมูลที่นำมาเทียบหาสัมประสิทธิ์ สหสัมพันธ์)

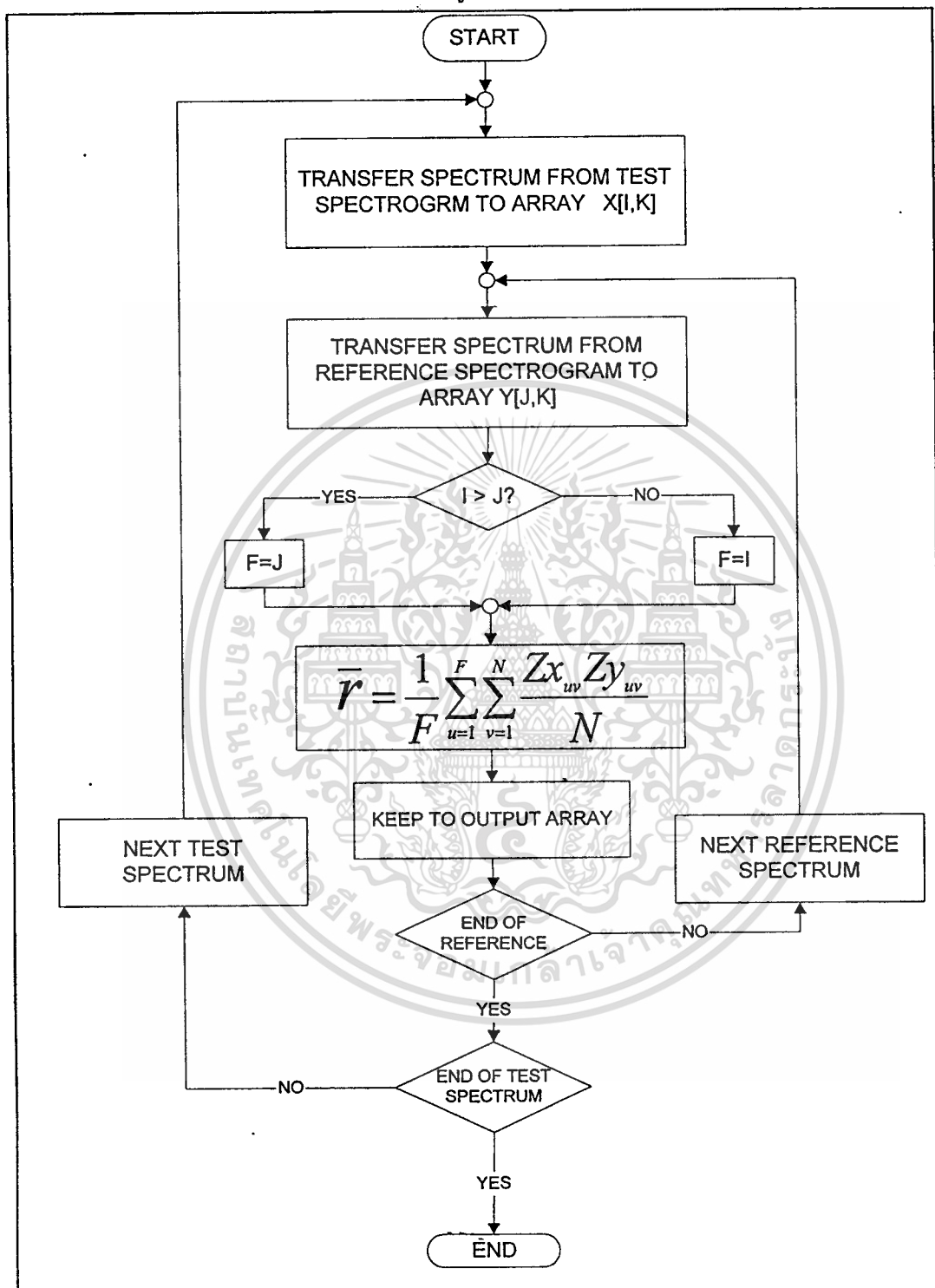
สาเหตุที่กำหนดเช่นนี้ มาจากข้อสังเกตที่ว่า หากข้อมูลมีช่วงเวลาการเปล่งเสียงยาวนานออกไปมาก โอกาสที่จะเป็นเสียงเดียวกันก็น้อยลง หรือกรณีเป็นเสียงเดียวกันแต่มีช่วงเวลาการเปล่งเสียงยาวออกไปไม่มากนัก(1-2เฟรม) ก็ไม่จำเป็นต้องนำส่วนที่เกินมาเปรียบเทียบความสัมพันธ์ เพราะเมื่อค่าเฉลี่ย (Mean) ของระดับสัมประสิทธิ์สหสัมพันธ์ ตามสมการที่ (3.8)

$$R = \bar{r} = \frac{1}{F} \sum_{p=1}^F r_{xy_p} \text{(3.8)}$$

เมื่อ R หรือ $\bar{r}$ คือค่า Mean ที่บอกถึงระดับความสัมพันธ์ของข้อมูลเสียงสองชุด

จะถูกหารด้วยจำนวน Discrete F หรือ จำนวน Δt ซึ่งหากเป็นเสียงเดียวกันแล้วค่าเฉลี่ยผลรวมที่ได้จะไม่ผิดพลาดมาก การหาค่าระดับของความสัมพันธ์ของข้อมูลแสดงเป็นผังงานได้ดังภาพที่ 7 ค่าระดับความสัมพันธ์ที่ได้จะบ่งบอกถึงคุณสมบัติเฉพาะของเสียงนั้นโดยเทียบกับเสียงหลักทั้งหมด

แผนภูมิที่ 7



แผนภูมิแสดงผังงานการหาระดับความสัมพันธ์ระหว่าง สเปกโตรแกรมเสียงที่ต้องการหา กับเสียงหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผังงาน ดังแผนภูมิที่ 7 จะแสดงการหาจุดเริ่มต้นและจุดสิ้นสุดของชุดข้อมูลเสียงทั้งสอง โดยใช้สมการการหาค่าโครีเลชันเฉลี่ยของเสียงทดสอบและเสียงอ้างอิง จากสมการที่ 3.7 และ 3.8 จะได้

$$\bar{r} = \frac{1}{F} \sum_{u=1}^F \sum_{v=1}^N \frac{Zx_{uv} Zy_{uv}}{N} \dots\dots\dots(3.9)$$

โดยกำหนด $\bar{r}$ คือ ค่า Mean ของโครีเลชันเฉลี่ย ของเสียงทดสอบและเสียงอ้างอิง

N คือจำนวนของข้อมูลในแกนความถี่ (ในการทดลองใช้ 128 ข้อมูลต่อสเปกตรัม)

F คือจำนวนสเปกตรัมเฟรมที่สนใจที่จะนำมาหาค่าโครีเลชันเฉลี่ย ของเสียงทดสอบและเสียงอ้างอิง โดยกำหนดให้มีจำนวนเท่ากับจำนวนสเปกตรัมของของพยางค์ที่เป็นเสียงทดสอบหรือเสียงอ้างอิง ที่มีจำนวนน้อยที่สุด

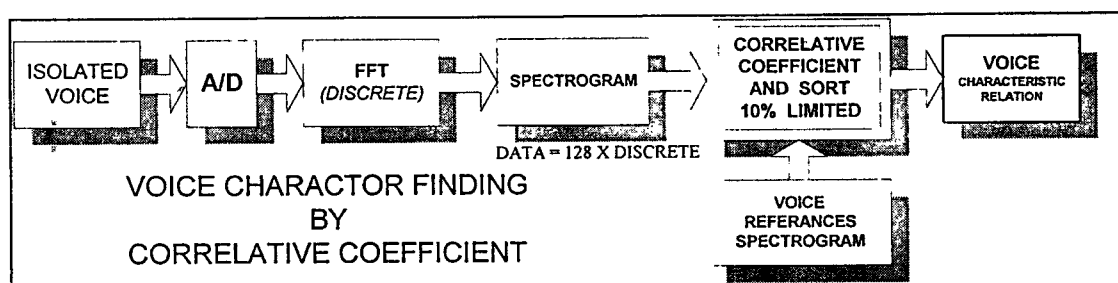
v คือ ข้อมูลที่เป็นสมาชิกของ N เริ่มตั้งแต่ 1 ถึง N

u คือ สเปกตรัมที่เป็นสมาชิกของเสียงหนึ่งพยางค์ในสเปกโตรแกรม เริ่มที่สเปกตรัมที่ 1 ถึง F

วิธีการหาคุณลักษณะจำเพาะสัมพันธ์ของเสียง

การวิเคราะห์เสียงที่ใช้ในการทดลอง ทำโดยนำสเปกโตรแกรมเสียงใดๆ มาเทียบหาคุณลักษณะจำเพาะกับสเปกโตรแกรมเสียงหลักว่า เสียงที่นำมาทดสอบมีคุณสมบัติใกล้เคียงเสียงหลักใด เพียงใด โดยใช้วิธีหาคุณลักษณะจำเพาะสัมพันธ์ ด้วยการหาสัมประสิทธิ์สหสัมพันธ์ ของเสียงที่ต้องการทดสอบ ตามบล็อกไดอะแกรม ดังภาพที่ 8

ภาพที่ 8



แสดงบล็อกไดอะแกรมของวิธีการหาคุณลักษณะจำเพาะสัมพันธ์ของเสียงโดยใช้สัมประสิทธิ์สหสัมพันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากบล็อกไดอะแกรม ในภาพที่ 8 ชุดสัญญาณเสียงอนาล็อกที่เป็นเสียงคำโดดที่ใช้เป็นตัวอย่าง จะถูกสุ่มเข้ามาโดยกำหนดอัตราสุ่มที่ 10KHz แล้วแปลงให้เป็นสัญญาณดิจิทัล ด้วยรายละเอียดของการแปลงข้อมูลขนาด 8 บิต ที่บล็อก A/D ทาสเปกโตรแกรมด้วยการรวมวิธีการแปลงฟูเรียร์อย่างรวดเร็วแล้วปรับสเปกโตรแกรมเพื่อให้เหมาะสมสำหรับการวิเคราะห์เสียง (ธันวา ศรีประโม่ง, 2537:13-21) สเปกโตรแกรมที่ได้มีลักษณะเช่นเดียวกับภาพที่ 5 งานที่ทำต่อมาก็คือ การแยกค่าเฉพาะของเสียงแต่ละพยางค์ที่ต้องการออกมาจากสเปกโตรแกรม แล้วนำมาหาคุณลักษณะจำเพาะของเสียงด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlative Coefficient) แล้วจัดเรียงใหม่ในขอบเขตจำกัด ที่ 10% ของเสียงทั้งหมด คุณลักษณะจำเพาะของเสียงที่หาได้นี้เป็นข้อมูลที่แสดงค่าระดับความสัมพันธ์กับสเปกโตรแกรมของเสียงอ้างอิง(ดูภาพที่10) มีขนาดข้อมูลลดลงอย่างมาก ซึ่งจะนำไปใช้ในกระบวนการวิเคราะห์เสียงด้วยนิวรัลเน็ตเวิร์คในลำดับถัดไป

การแยกค่าออกจากสเปกโตรแกรม

กระทำเพื่อแยกข้อมูลสเปกโตรแกรมเฉพาะส่วนที่เป็นเนื้อเสียงออกจากแบคกราวด์นอยส์ (Background noise) เพื่อการสร้างเป็นฐานข้อมูลหลัก, ข้อมูลทดสอบ และข้อมูลอ้างอิง วิธีการแยกเนื้อเสียงที่ต้องการนำมาวิเคราะห์ อาจกระทำได้หลายแบบดังนี้

1.แยกจากสัญญาณเสียงในขั้นตอนที่เป็นอนาล็อกโดยตรวจสอบระดับแอมพลิจูดของเสียงแล้วนำมาแยกเป็นส่วน ๆ ตามระดับพลังงาน แล้วแปลงเป็นสเปกโตรแกรม

(ระพีพัฒน์ เพ็ญศิริ 2538:12-17)

2.แยกเนื้อของเสียงคำต่างๆ ในสเปกโตรแกรม ออกมาหลังจากแปลงฟูเรียร์แล้ว

(ธันวา ศรีประโม่ง 2537:24-25)

สำหรับงานวิจัยที่น่าเสนอเลือกวิธีการที่สองเพราะสะดวกและรวดเร็วกว่า แต่ต้องพิถีพิถันเรื่องสัญญาณรบกวนในการบันทึกเสียงมากพอสมควร

เนื่องจากงานวิจัยมุ่งไปที่การวิเคราะห์เสียงคำโดด ๆ ประเภทหนึ่งคำหนึ่งพยางค์ ซึ่งมีลักษณะพิเศษในการออกเสียงคือ การออกเสียงแต่ละคำมีช่องว่างค่อนข้างเด่นชัด ทำให้สามารถแยกค่าออกมาได้โดยไม่ยากนักหากสังเกตจากรูปสเปกโตรแกรม จะเห็นขอบเขตของแต่ละเสียงได้ชัดเจน โดยช่องว่างระหว่างเสียงจะมีระดับพลังงานต่ำแสดงด้วยสีน้ำเงิน การแยกพยางค์ออกจากสเปกโตรแกรมอาศัยวิธีการหาผลรวมเฉลี่ยของระดับพลังงานในแถบความถี่เสียงแต่ละเฟรมทุกเฟรม แล้วหา เทอร์สโวลที่ 20% จากค่าต่ำสุดไปถึงสูงสุด จากนั้นจึงพิจารณาในแต่ละสเปกตรัม(เฟรม) หากผลรวมเฉลี่ยของระดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พลังงานเสียงในสเปกตรัมได(เฟรม)มีค่ามากกว่าเทรสโฮล ก็จะนำข้อมูลเฟรมไปเก็บไว้ที่บัฟเฟอร์ โดยถือเป็นเนื้อเสียงของพยางค์ที่จะนำไปพิจารณา วิธีการดังกล่าวเขียนเป็นสมการได้ดังนี้

$$Eav_{(ti)} = \frac{\sum_{f=1}^M Ef}{M} \quad | \quad ti = 1 \text{ to } n \dots \dots \dots (3.10)$$

และ

$$Tref = K(Eav_{\max} - Eav_{\min}) + Eav_{\min} \dots \dots \dots (3.11)$$

ข้อมูลแต่ละพยางค์จะถูกแยกออกมาเมื่อ

$$Eav_{(ti)} > Tref \dots \dots \dots (3.12)$$

กำหนด $Eav_{(ti)}$ คือ ค่าระดับพลังงานเฉลี่ย ณ ช่วงเวลา Δt ไต ๆ

Ef คือ ค่าระดับพลังงานของแต่ละความถี่ในสเปกตรัม โดย f คือ

ลำดับข้อมูลของสเปกตรัมในแกนความถี่ มีลำดับตั้งแต่ 1 ถึง M

M คือ จำนวนข้อมูลในแกนความถี่ (การทดลองใช้ 128 ข้อมูล)

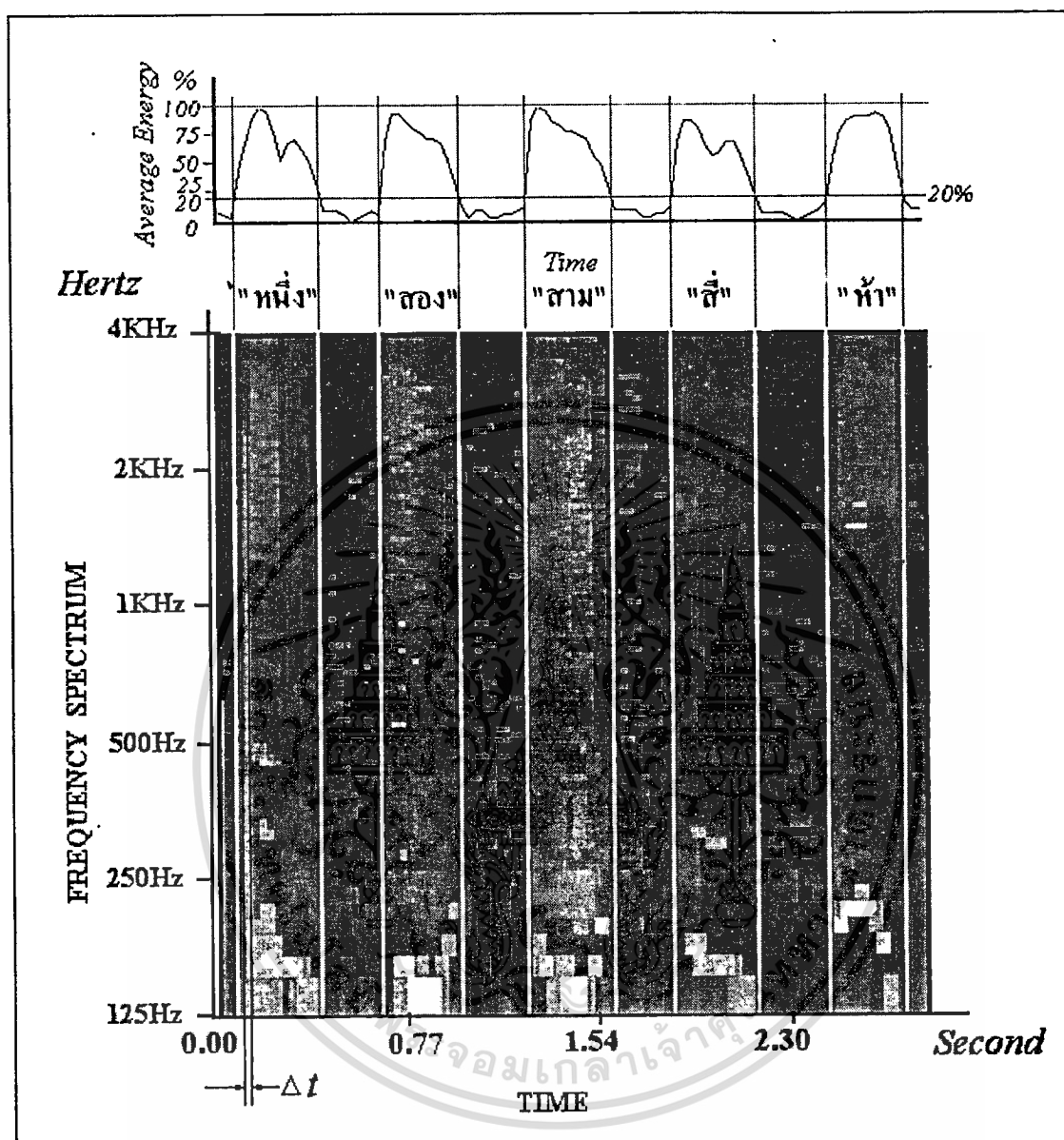
ti คือ จำนวนสเปกตรัม ไต ๆ ที่กำลังสนใจ, ti มีค่าตั้งแต่ 1 ถึง n

n คือ จำนวนสเปกตรัม ไต ๆ (Δt) ทั้งหมดที่ประกอบเป็นสเปกโตรแกรม

$Tref$ คือ ค่าเทรสโฮล ที่กำหนดขึ้นเพื่อแยกพยางค์ออกมาพิจารณา

K คือ สัมประสิทธิ์ ที่ใช้กำหนดจุดแยกพยางค์ออกมาพิจารณา มีค่าตั้งแต่ 0.0-1.0 (จากการทดลองค่าที่ดีที่สุด คือ 0.2)

ภาพที่ 9



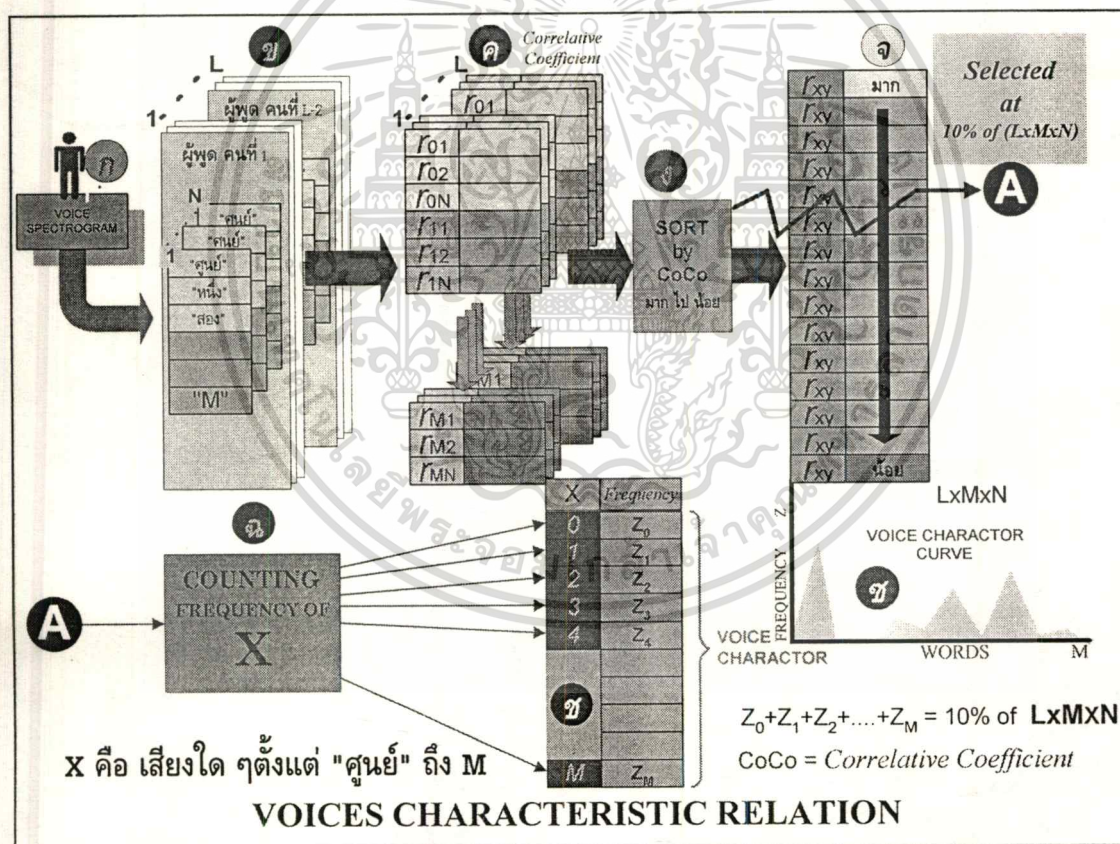
แสดงผลการแยกข้อมูลช่วงที่มีการเปล่งเสียงออกจากสเปกโตรแกรมโดยใช้สมการที่ 3.12

จากภาพที่ 9 กราฟด้านบนเป็นกราฟที่ได้จากสมการที่ 3.10 เช่นกันแต่นำมาปรับแสดงผลเป็นเปอร์เซ็นต์ จากการทดลองพบว่า ค่าเทรชโฮลที่เหมาะสมอยู่ที่ 20 % (สมการที่ 3.10) จากจุดต่ำสุดไปจุดสูงสุด การตัดค่าแบบนี้มีความแม่นยำพอสมควร (มากกว่า 92%) โดยมีเงื่อนไขคือ ขณะบันทึกเสียงต้องไม่มีเสียงสอดแทรกรบกวนเกินไปในช่องว่างระหว่างของพยางค์จึงเพียงพอที่จะเลือกไปใช้แยกพยางค์เพื่อการทดลองได้

วิธีการหาคุณลักษณะจำเพาะสัมพันธ์และลดขนาดข้อมูลเสียง

การหาคุณลักษณะจำเพาะสัมพันธ์ (Voice Characteristic Relation) ที่ใช้ในการทดลองเป็นการแปลงข้อมูล สเปกโตรแกรมของเสียงที่ต้องการวิเคราะห์ให้เป็นข้อมูลคุณลักษณะจำเพาะประจำตัวของเสียงนั้นว่าสัมพันธ์กับเสียงนั้นเสียงนี้กี่เปอร์เซ็นต์ในกลุ่มเสียงหลัก ใช้วิธีการเทียบระดับความสัมพันธ์ของสหสัมพันธ์ ระหว่างสเปกโตรแกรมเสียงที่ต้องการหาค่ากับเสียงฐานข้อมูลหลัก โดยคุณลักษณะประจำตัวของเสียงนั้นเป็นคุณสมบัติที่เทียบกับเสียงฐานข้อมูลหลักเท่านั้น และจำเป็นต้องสร้างคุณลักษณะประจำตัวของเสียงทุกเสียงก่อนเพื่อใช้ในการขบวนการอ้างอิงเปรียบเทียบที่ (ข)

ภาพที่ 10



แสดงวิธีการหาคุณลักษณะจำเพาะสัมพันธ์โดยใช้สหสัมพันธ์ที่สร้างขึ้น

จากภาพที่ 10 (ก) คือสเปกโตรแกรมของเสียงที่ต้องการหาคุณลักษณะ (ข) คือกลุ่มสเปกโตรแกรมของเสียงที่ใช้เป็นฐานข้อมูลหลัก สร้างจากผู้พูด L คน ที่พูด "ศูนย์", "หนึ่ง", ถึง M จำนวน N

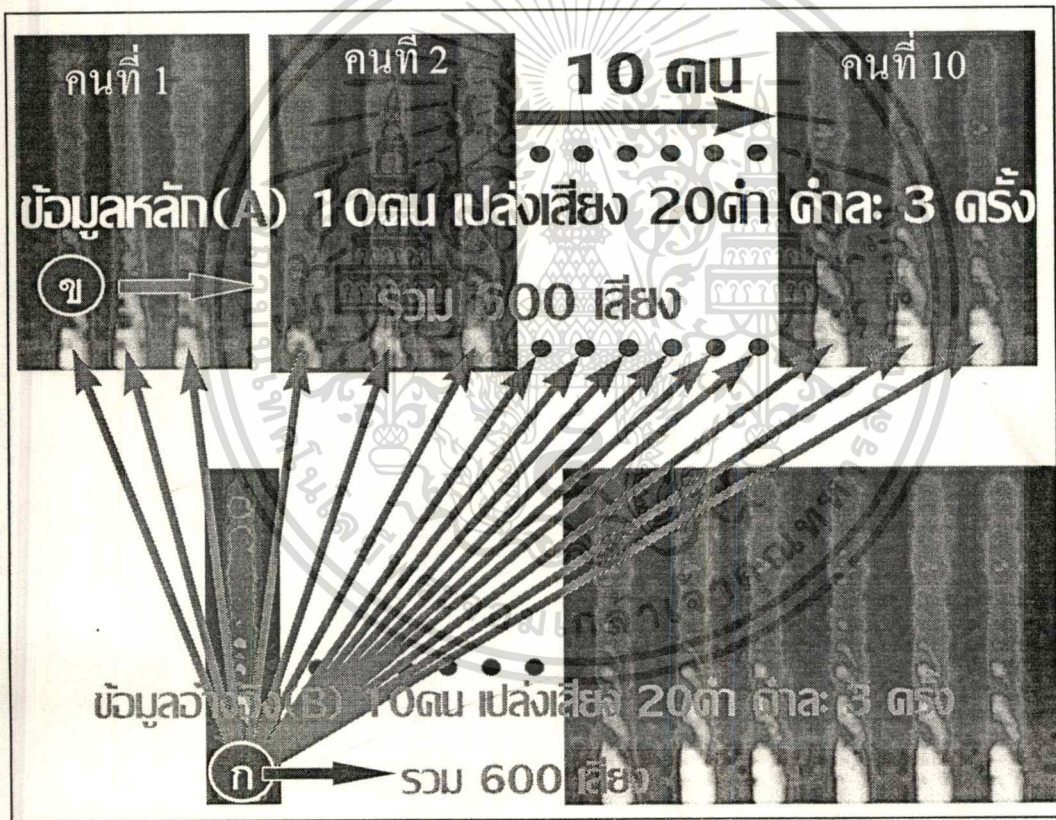
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ครั้ง มีการคำนวณหาค่าระดับสหสัมพันธ์ของข้อมูลทั้งสองของทุกคน โดยใช้สมการที่ 3.6, 3.7 และ 3.8 ข้อมูลสเปกโตรแกรมเสียงที่ต้องการทดสอบ (ก) 1ชุด จะนำไปเทียบหาค่าสหสัมพันธ์กับข้อมูลสเปกโตรแกรม เสียงอ้างอิง (ข) ทุกชุด ได้เป็นค่าสัมประสิทธิ์สหสัมพันธ์กับชุดสเปกโตรแกรมเสียงอ้างอิงทุกชุดของทุกคน (ค) ได้จำนวนเป็น r_{11} ถึง r_{MN} เมื่อ M คือจำนวนเสียงคำใดๆ ที่กำหนดขึ้นเพื่อทดลอง N คือจำนวนครั้งที่พูดซ้ำของผู้พูด จากนั้น นำข้อมูล r_{11} ถึง r_{MN} ของผู้พูด L คน ที่ได้จาก (ค) ไปจัดเรียงในบล็อก SORT ที่ (ง) ซึ่งเป็นการจัดเรียงแบบมากไปน้อย ตามค่าสัมประสิทธิ์สหสัมพันธ์ซึ่งมีชื่อของค่า(M)เป็นตัวแปรตาม ได้เป็นตาราง (จ) โดย r_{xy} คือตัวแปรตาม ที่แปรไปตามค่า Coefficient (มากไปน้อย x คือ เสียงของคำใดๆที่ใช้ทดสอบ y คือ ครั้งของคำที่พูด จำนวนค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้นี้จะมีความเท่ากับ $L \times M \times N$ การสรุปคุณลักษณะของเสียงทดสอบจะเลือกค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการจัดเรียงมาจาก 10% แรกของจำนวนข้อมูลสัมประสิทธิ์สหสัมพันธ์ที่ได้ ($L \times M \times N$) ที่ (ฉ) ณ จุดนี้จะมีการตรวจสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้ด้วยคือ ต้องมีค่า ไม่ต่ำกว่า 0.7 หากมีค่าต่ำกว่าจะคัดออก จากนั้นจึงมานับความถี่ของแต่ละเสียงคำที่กำหนดขึ้นเพื่อทดลอง ที่อยู่ใน 10% แรก ความถี่ของแต่ละคำจะมีตัวเลขได้มากที่สุดไม่เกิน $L \times N$ (ช) ซึ่งเสียงคำที่กำหนดขึ้นเพื่อทดลอง มีจำนวน M คำ นั่นคือ คุณลักษณะจำเพาะสัมพันธ์ของเสียงนั้น ซึ่งสหสัมพันธ์กับฐานข้อมูลหลักอาจเขียนเป็นกราฟได้ดังรูป(ซ) ณ จุดนี้จะสามารถบอกความเป็นไปได้ของเสียงที่น่าจะเป็นได้แล้วในระดับหนึ่งคือค่า (X) ที่มีระดับความถี่สูงที่สุดนั่นเอง (ซ) จากการทดลองพบว่า คุณลักษณะจำเพาะสัมพันธ์ที่ได้พอที่จะบอกหรือตัดสินได้แล้วว่าเสียงที่ต้องการทราบนั้นคือเสียงใดจากคุณลักษณะจำเพาะสัมพันธ์ที่บอกว่าเหมือนกับเสียงนั้นเสียงนี้ก็เปอร์เซ็นต์ในฐานข้อมูลเสียงหลักโดยสังเกตจากเปอร์เซ็นต์ที่มากที่สุดว่าตรงกับเสียงใด แต่เมื่อนำผลที่ได้ไปหาอัตราการรู้จำ ที่เป็นเปอร์เซ็นต์ความถูกต้องแล้ว มีความถูกต้องเฉลี่ยประมาณ 63 เปอร์เซ็นต์ ซึ่งไม่เพียงพอต่อการนำไปใช้งานจริง แต่เนื่องจากค่าคุณลักษณะจำเพาะสัมพันธ์ที่ได้มีลักษณะเป็นแพตเทิร์นที่มีจำนวนข้อมูลลงที่คือ 20ข้อมูล ที่ไม่ขึ้นกับความดังและระยะเวลาการแปลงเสียงทำให้สามารถนำไปฝึกสอนให้นิวรัลเน็ตเวิร์คได้

จากภาพที่ 10 เป็นภาพที่อธิบายกระบวนการอย่างหยาบ ๆ ภาพที่ 11, 12 และ 13 จะแสดงรายละเอียดของภาพที่ 10 เพิ่มขึ้น ในภาพที่ 11 แสดงให้เห็นว่า สเปกโตรแกรม ในกลุ่ม B ที่ (ก) หนึ่งสเปกโตรแกรมจะนำไปเทียบหาความสอดคล้องกับสเปกโตรแกรมในกลุ่มเสียงหลัก A ที่ (ข) ทั้งหมด จะได้เป็นค่าระดับความสัมพันธ์ จำนวน 600 ค่า โดยแต่ละค่าจะมีเสียงกำกับ เช่น เสียง "หนึ่ง" ของคนที่ 11 (คนแรกในกลุ่ม B) ค่าระดับความสัมพันธ์ กับกลุ่มเสียงหลัก เสียง "หนึ่ง" ของคนที่ 1 = 0.87, "หนึ่ง" ของคนที่ 2 = 0.76, "หนึ่ง" ของคนที่ 3 = 0.80, "หนึ่ง" ของคนที่ 4 = 0.77, "หนึ่ง" ของคนที่ 1 = 0.79, "สอง" ของคนที่ 5 = 0.65, "สี่" ของคนที่ 4 = 0.71 และเสียงอื่นๆ ค่าอื่นๆ จะเห็นว่า มีชื่อของเสียงกำกับคู่กันไปกับค่าระดับความสัมพันธ์ ค่าดังกล่าวนี้ คือค่า B1A1, B1A2, B1A3, ... , B1Am ในภาพที่ 12 หรือ (ค) ในภาพที่ 10 จากนั้น นำไปจัดเรียงตามค่าระดับความสัมพันธ์แบบมากไปหาน้อย โดยมีเสียงเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำกับเป็นตัวแปรตาม แล้วเลือกออกมา 10 เปอร์เซ็นต์แรกจากค่าที่มากที่สุด แล้วนับจำนวนของเสียงกำกับ ว่าในจำนวนเสียง 10 เปอร์เซ็นต์ มีเสียง "หนึ่ง" กี่คะแนน, เสียง "สอง" กี่คะแนน, ... , เสียง "ไฟ" กี่คะแนน จะเห็นว่า มีข้อมูลจำนวนตายตัว เท่ากับจำนวนแบบเสียงที่ใช้ในการทดลองคือ 20 เสียงเสมอ ระดับคะแนนตรงนี้เองที่เรียกว่า คุณสมบัติจำเพาะสัมพันธ์ การดำเนินการจะนำทุกสเปกโตรแกรมในกลุ่มอ้างอิง(B) ไปหาค่าคุณลักษณะจำเพาะสัมพันธ์กับกลุ่มสเปกโตรแกรมเสียงหลัก เมื่อครบแล้วจะนำไปฝึกสอนให้กับนิวรัลเน็ตเวิร์ค ดังภาพที่ 12 ส่วนภาพที่ 13 แสดงถึงกระบวนการทดสอบนิวรัลเน็ตเวิร์ค ซึ่งมีกระบวนการปรีโพรเซสซึ่งเหมือนกับกระบวนการปรีโพรเซสซึ่งของขั้นตอนฝึกสอนให้นิวรัลเน็ตเวิร์ค

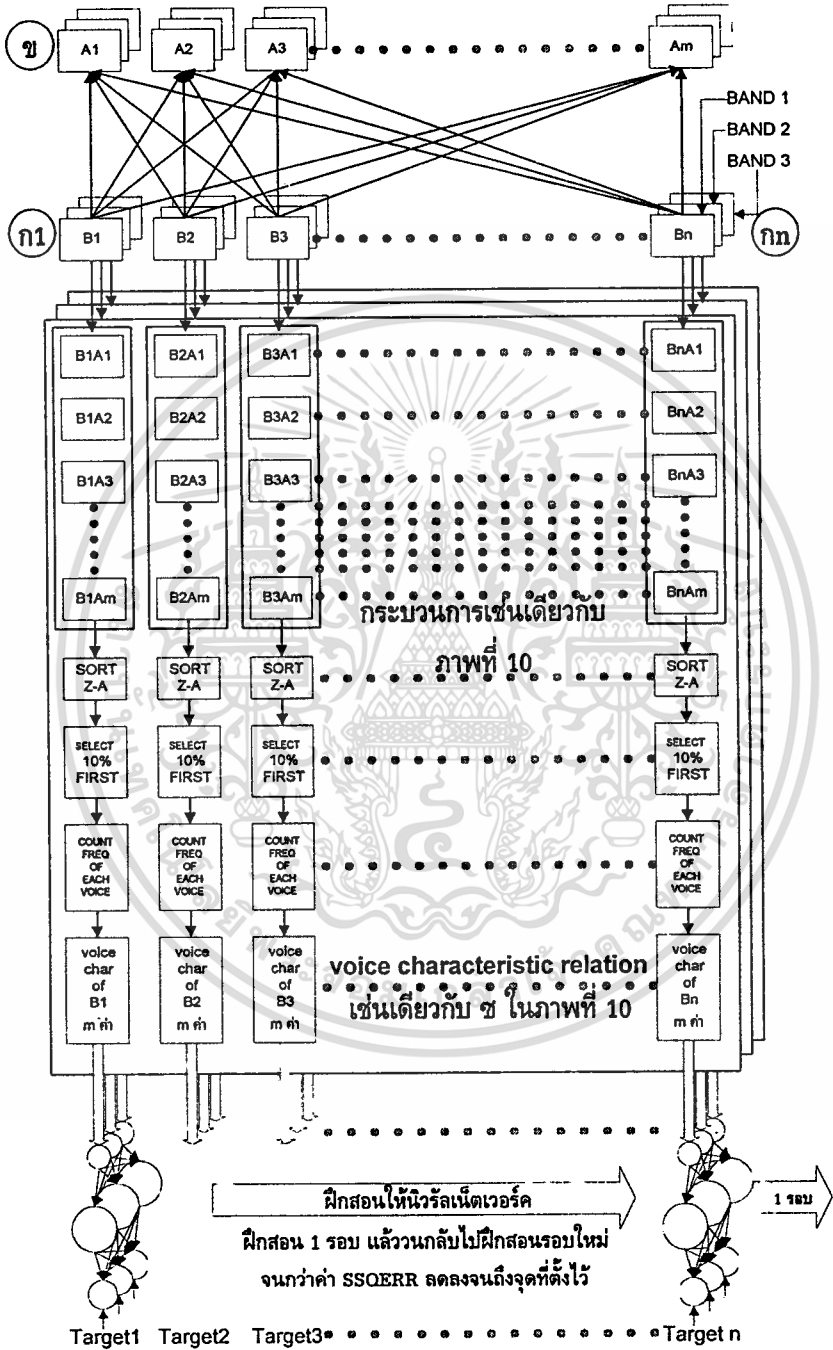
ภาพที่ 11



แสดงวิธีการนำข้อมูลสเปกโตรแกรมอ้างอิง(B) หนึ่งเสียง(ก) ไปหาระดับความสอดคล้องของข้อมูล กับชุดข้อมูลสเปกโตรแกรมหลัก(A) ทุกเสียง ซึ่งจะได้ ค่าระดับความสัมพันธ์ 600 ค่า ต่อ 1 เสียง จากกลุ่ม (B)

ภาพที่ 12

แสดงการ หาคุณลักษณะจำเพาะสัมพันธ์ ของชุดข้อมูลเสียงอ้างอิง(B)กับชุดข้อมูลเสียงหลัก(A) แล้วนำไปฝึกสอนให้กับนิวรัลเน็ตเวิร์ค

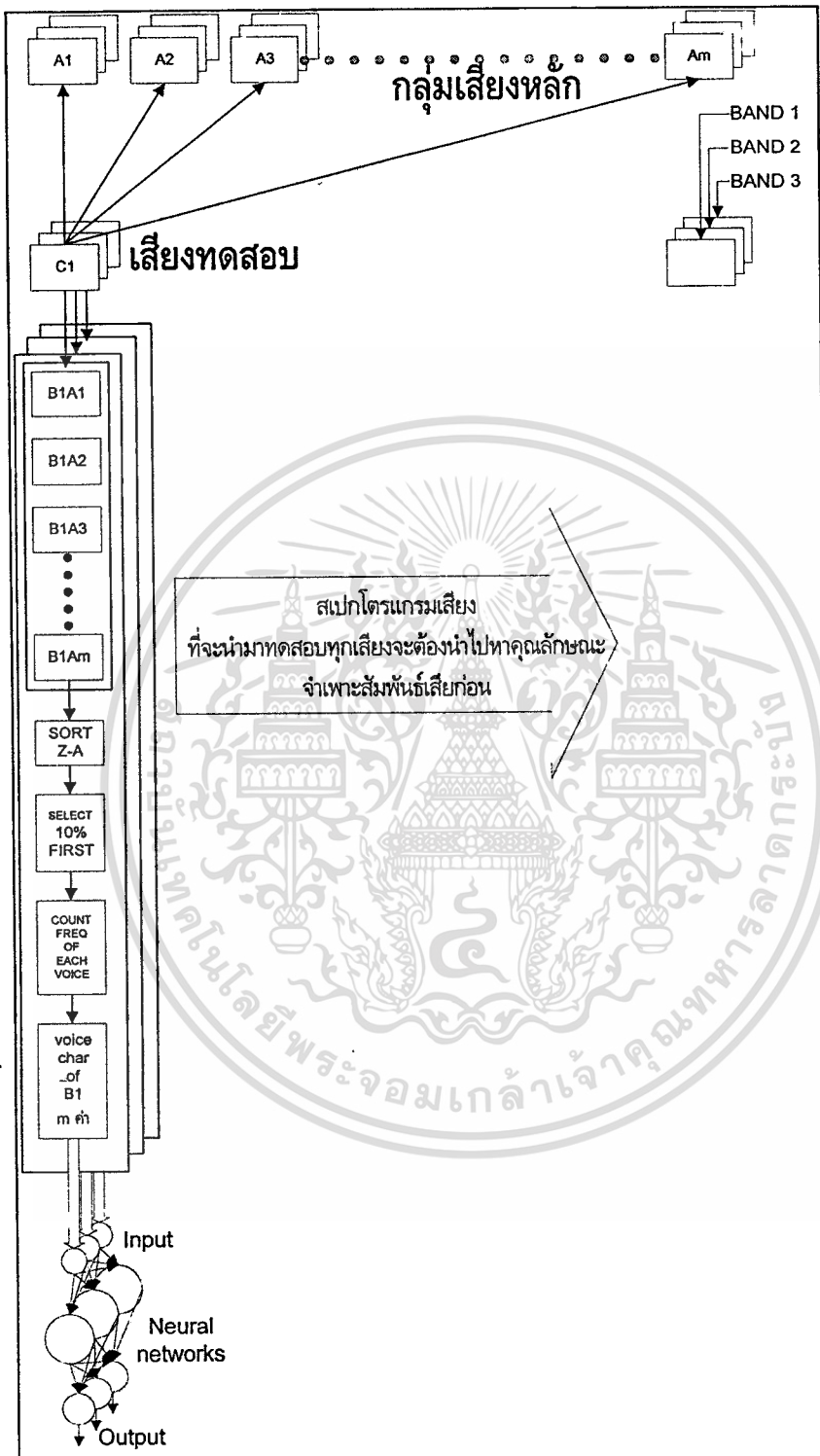


แสดงการหาคุณสมบัติจำเพาะสัมพันธ์ ของชุดข้อมูลเสียงอ้างอิง(B) กับชุดข้อมูลเสียงหลัก (A)

แล้วนำไปฝึกสอนให้กับนิวรัลเน็ตเวิร์ค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 13



แสดงการหาคุณลักษณะจำเพาะสัมพันธ์ ของชุดข้อมูลเสียงทดสอบ C กับชุดข้อมูลหลัก A

แล้วนำไปทดสอบนิวรัลเน็ตเวิร์ค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การแบ่งแบนด์ของสเปกโตรแกรมเพื่อเพิ่มความแม่นยำ

การแบ่งแบนด์ของสเปกโตรแกรม ผู้วิจัยตั้งใจไม่กล่าวถึงในลำดับแรก เพื่อหลีกเลี่ยงความสับสน แต่ขอกล่าวสรุปก่อนว่า วิธีการและขั้นตอนยังคงเป็นเช่นเดิม แต่มีการแบ่งพิจารณาเพิ่มขึ้นเป็นช่วง ๆ สามช่วง ดังจะกล่าวต่อไปนี้

เนื่องจากมีข้อมูลของสเปกตรัมในสเปกโตรแกรมถึง 128 ข้อมูลต่อ 1 สเปกตรัม (1 สเปกตรัม = 1 ดีกรีของสเปกโตรแกรม) แสดงค่าตามสเกล Logarithm ครอบคลุมแถบความถี่ตั้งแต่ 125Hz ถึง 4KHz ซึ่งข้อมูลมากพอที่ทำให้เกิดข้อผิดพลาดของการเปรียบเทียบหาระดับความสัมพันธ์ระหว่างสเปกตรัมทดสอบกับสเปกตรัมหลักได้ ในกรณีได้ค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับกลาง ค่านี้อาจเกิดได้หลายกรณีเช่น

1. ค่าช่วงต้นของข้อมูลสัมพันธ์กันมากแต่ช่วงปลายสัมพันธ์กันน้อย
2. ค่าช่วงต้นของข้อมูลสัมพันธ์กันน้อย แต่ช่วงปลายสัมพันธ์กันมาก
3. ค่าช่วงกลางของข้อมูลสัมพันธ์กันมาก แต่ช่วงต้นกับปลายสัมพันธ์กันน้อย
4. ค่าความสัมพันธ์อยู่ในระดับกลางตลอด

ดังนั้นจึงแก้ปัญหานี้โดยใช้วิธีแบ่งช่วงพิจารณาโดยแบ่งสเปกตรัมความถี่ของแต่ละช่วงเวลาออกเป็น 3 ย่าน คือ

ย่านแบนด์ต่ำ ครอบคลุมความถี่ 125 Hz - 1 KHz

รวม 75 ข้อมูล ในช่วงข้อมูลที่ 0-75 ของข้อมูลในสเปกตรัม

ย่านแบนด์ต่ำ-กลางครอบคลุมความถี่ 500 Hz - 2 KHz

รวม 50 ข้อมูลในช่วงข้อมูลที่ 50 - 100 ของข้อมูลในสเปกตรัม

ย่านแบนด์กลางครอบคลุมความถี่ 1 KHz - 4 KHz

รวม 50 ข้อมูล ในช่วงข้อมูลที่ 75 - 128 ของข้อมูลในสเปกตรัม

ลักษณะของแบนด์จะมีลักษณะเหลื่อมซ้อนกัน 50% เพื่อลดข้อผิดพลาดที่เกิดขึ้นในช่วงรอยต่อของแบนด์ที่แบ่ง, ลดขนาดของข้อมูลให้เหมาะสมกับวิธีการ และต้องการให้ผลการหาความสัมพันธ์ระหว่างสเปกโตรแกรมแต่ละแบนด์ ได้ค่าสอดคล้องกัน หากพิจารณาลงไปจะพบว่า มีแบนด์ต่อกันอยู่แล้วคือแบนด์1กับแบนด์3 โดยมีแบนด์2 มาโอเวอร์แลป ดังนั้นไม่ว่าในกระบวนการเตรียมข้อมูลแบบอ้างอิง หรือกระบวนการทดสอบก็ตาม ต้องดำเนินการทำทั้ง 3 แบนด์ หมายถึงต้องทำตามวิธีการหาคุณลักษณะจำเพาะของเสียงดังกล่าวถึง 3 ครั้ง ทำให้ได้ค่าแพต เทอร์นอ้างอิง 3 ชุด ข้อมูลที่นำมาทดสอบก็ทำเช่นเดียวกัน จากนั้นจึงหาค่าระดับความสัมพันธ์ระหว่างแบบทดสอบกับแบบอ้างอิงทั้ง 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบด์ แบบแบบด์ต่อแบบด์ เมื่อได้ค่าระดับความสัมพันธ์ระหว่างแบบดังกล่าวทั้ง 3 แบบด์ แล้วจึงนำ มาหาค่าเฉลี่ยของความสัมพันธ์ทั้ง 3 แบบด์ อีกครั้งหนึ่งด้วยสมการ 3.13

$$CORR_{AVG} = \frac{1}{F} \sum_{j=1}^F \sum_{i=BSTRT}^{BSTOP} \frac{Zx_{ji} Zy_{ji}}{BSTOP} \dots\dots\dots(3.13)$$

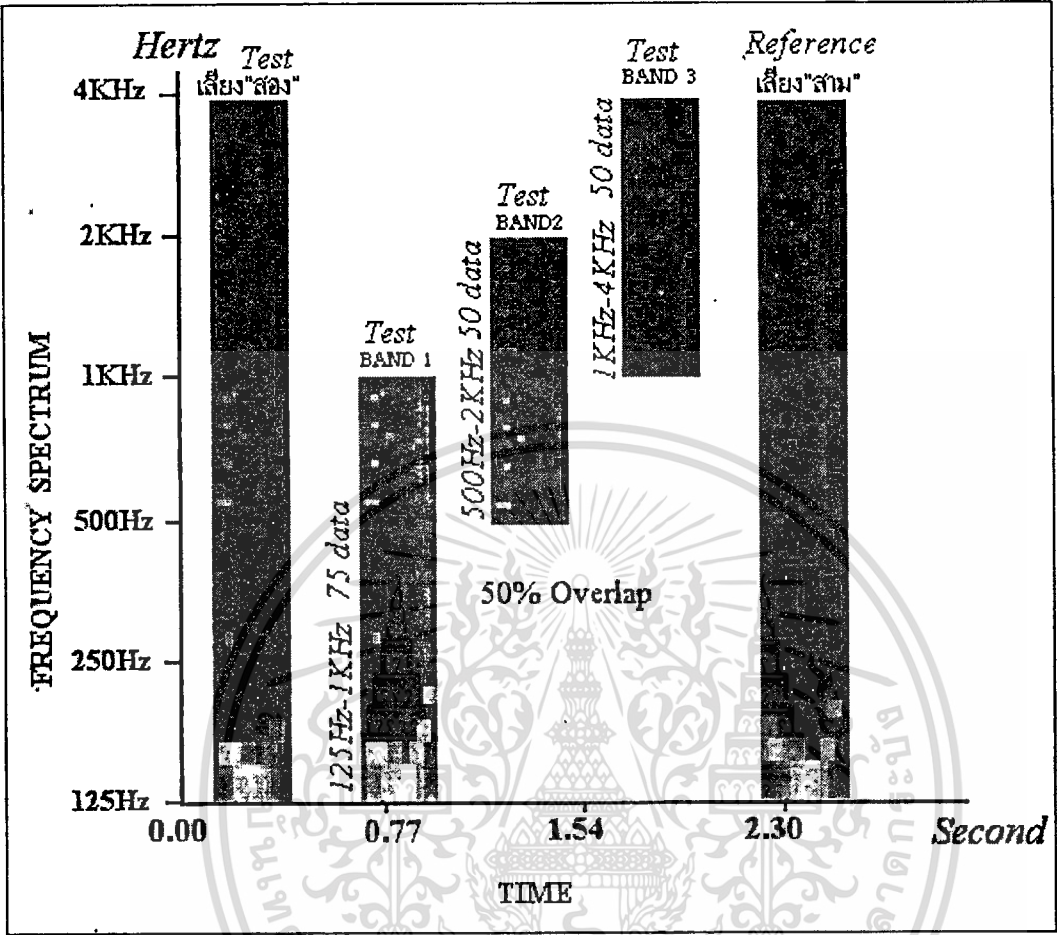
$CORR_{AVG}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยในแบบด์ที่กำลังสนใจ
จากสมการที่ 3.13 ค่า Summation ในเทอมหลัง เป็นค่าความสัมพันธ์ของข้อมูล แบบโครีรีเลชันระหว่างเฟรมในสเปคโตรแกรมเสียงทดสอบกับเสียงอ้างอิง ในแบบด์ที่กำลังพิจารณา

F คือจำนวนเฟรมของเนื้อเสียงที่กำลังพิจารณา(1เฟรม = $1\Delta t$ ในภาพที่ 9) อาจมี จำนวนเฟรมเท่าเสียงอ้างอิงหรือเสียงทดสอบก็ได้โดยพิจารณาจากจำนวนเฟรมของเสียงใดที่น้อยที่สุด โดยให้

$F = m$ เมื่อ $m-n \leq 0$
 $F = n$ เมื่อ $m-n > 0$

เมื่อ m คือ จำนวนเฟรมของสเปคโตรแกรมเสียงที่ใช้ทดสอบ
 n คือ จำนวนเฟรมของสเปคโตรแกรมเสียงอ้างอิง
 $BSTRT$ และ $BSTOP$ คือตำแหน่งเริ่มต้นและสิ้นสุดของข้อมูลของแบบด์ที่กำลังพิจารณา
 Zx_{ji} และ Zy_{ji} คือค่า Z-Score ของเฟรมทดสอบกับเฟรมอ้างอิง เฟรมที่ j
 j คือเฟรมที่กำลังพิจารณา

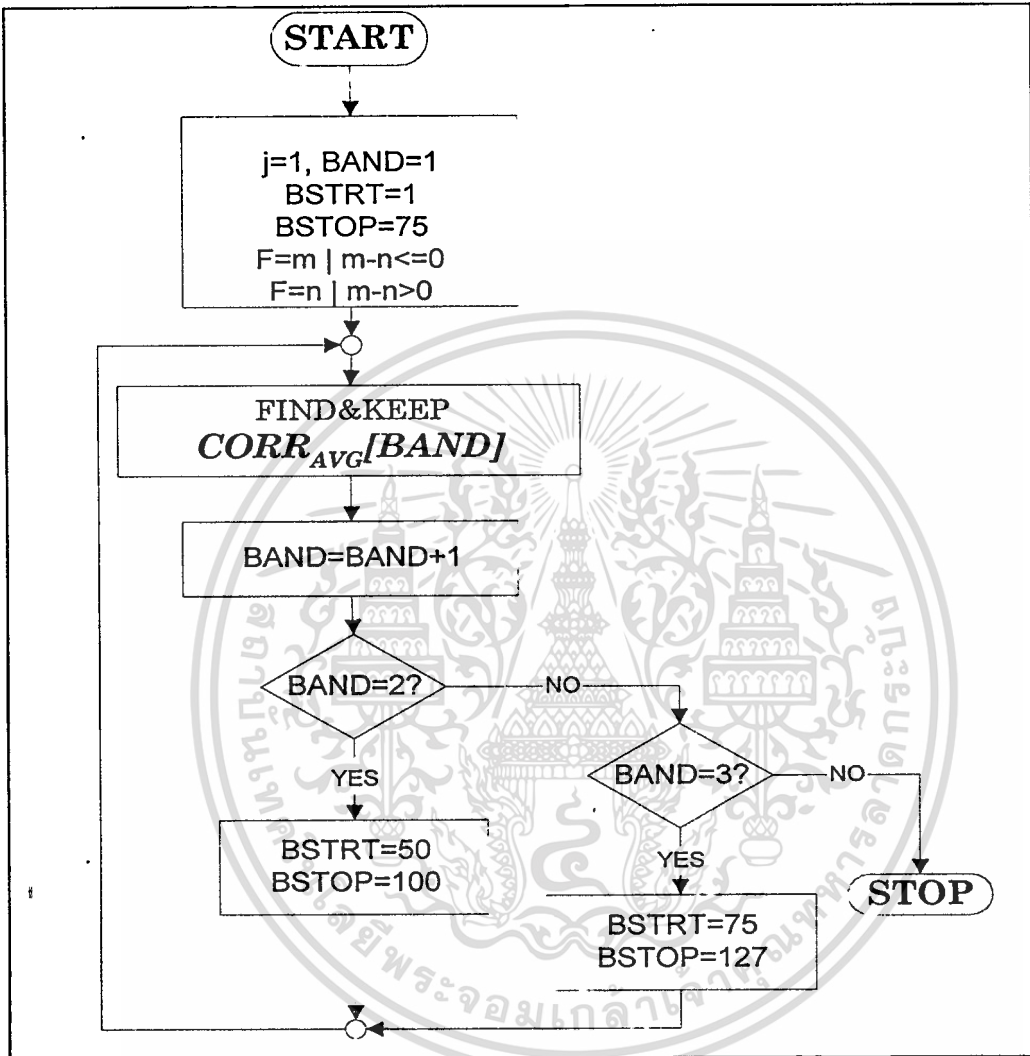
ภาพที่ 14



แสดงการตัดแบ่งสเปกโตรแกรมของเสียงหนึ่งออกเป็น 3 แบนด์ แบบซ้อนทับกัน 50% ของข้อมูล ซึ่งนำไปใช้เปรียบเทียบกับสเปกโตรแกรมของเสียงอ้างอิงต่าง ๆ แบบแบนด์ต่อแบนด์

การหาค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสามแบนด์ พอแสดงเป็นผังงานได้ตามแผนภูมิที่ 15 ดังนี้

แผนภูมิที่ 15



แสดงผังงานการหาค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสามแบนด์ ของสเปกโตรแกรมเสียงที่ต้องการ
กับเสียงหลัก

หลังจากเสร็จสิ้นการหาค่าสัมประสิทธิ์สหสัมพันธ์เฉลี่ยทั้งสามแบนด์แล้ว จะนำค่า
 $CORR_{AVG}$ ที่ได้ใช้เป็นแบบอ้างอิงเพื่อใช้ในการคำนวณการต่อไป

บทที่ 4

นิวรัลเน็ตเวิร์ค (Neural Networks)

ความรู้เบื้องต้นเกี่ยวกับนิวรัลเน็ตเวิร์ค

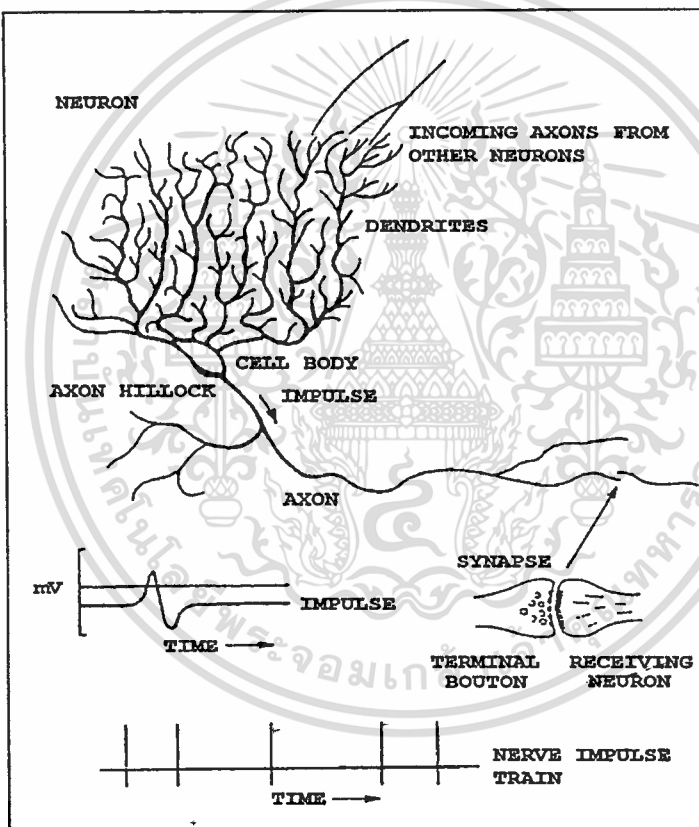
ในช่วงระยะเวลา 8-9 ปีที่ผ่านมา ในต่างประเทศมีการตื่นตัวในการวิจัยและพัฒนาเกี่ยวกับโครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) อย่างกว้างขวาง ทั้งทางด้านทฤษฎีและการประยุกต์ใช้งาน จนกระทั่งถึงปัจจุบันได้มีการประยุกต์นำนิวรัลเน็ตเวิร์ค มาใช้ในอุปกรณ์เครื่องใช้ต่างๆ มากขึ้น เช่น เครื่องมือหาปลา (Sonar) ที่มีความฉลาดมากขึ้น เช่น สามารถบอกได้ว่า ฝูงปลาที่กำลังตรวจจบบอยู่นั้นเป็นปลาชนิดใด, จำนวนเท่าไร เครื่องโทรศัพท์แบบที่สามารถเรียกเลขหมายปลายทางให้อัตโนมัติ (Voice phone) เพียงยกหูแล้วพูดชื่อของผู้ที่จะติดต่อเท่านั้น เครื่องอ่านตัวอักษร (OCR) ที่สามารถเปลี่ยนภาพตัวอักษรให้เป็นรหัสตัวอักษรแบบแอสกี ระบบนักบินอัตโนมัติ (Auto pilot aircraft) ระบบการคาดเดาอนาคตจากข้อมูลในอดีต (Forecasting, Prediction) ฯลฯ ซึ่งเครื่องมือและอุปกรณ์ต่างๆ ที่นำเอา ANN มาช่วยวิเคราะห์นั้น จะมีความฉลาดมากขึ้นและมีระบบความคิดที่ทำงานในลักษณะคล้ายกับของมนุษย์

นิวรัลเน็ตเวิร์ค หมายถึงโครงข่ายประสาทที่เชื่อมต่อกันระหว่างเซลล์ประสาทจำนวนมากมายมหาศาล มีความสามารถประมวลผลสูงบรรจุอยู่ในสมอง สมองชีวภาพที่เป็นจุดศูนย์กลางการควบคุมกิจกรรมของการดำเนินชีวิต การวิจัยสร้างโครงข่ายประสาทเทียม (Artificial Neural Networks) มีแนวคิดเลียนแบบการทำงานของสมองชีวภาพ โดยเรียนรู้และศึกษาการทำงานของสมองชีวภาพเพื่อกำหนดแนวทางสำหรับการสร้างแบบจำลองขึ้นมา แล้วพยายามสมมติฐานลักษณะการทำงานโดยจำลองเป็นโมเดลคณิตศาสตร์ที่มีลักษณะเดียวกันแล้วดำเนินการคำนวณโดยใช้คอมพิวเตอร์

นิวรัลเน็ตเวิร์คชีวภาพ

ระบบคิดคำนึงของมนุษย์ มีโครงสร้างพื้นฐานจากเซลล์สมองที่เรียกว่านิวรอน (Neurons) เรียงเป็นชั้นๆ อย่างซับซ้อนจำนวนมหาศาล ประมาณหมื่นล้าน(10^{11}) นิวรอนและอาจมีจุดเชื่อมโยงส่งผ่านจุดเชื่อมโยงภายในถึงพันล้านล้าน(10^{15}) จุด (Philip D. Wasserman, 1989:12) แต่ละนิวรอนจะมีคุณลักษณะแตกต่างกันไปโดยมีการทำงานคล้ายกันคือ รับเข้า, ประมวลผล, ส่งออกสัญญาณไฟฟ้าเคมีผ่านไปยังนิวรอน ซึ่งจะส่งสื่อสารไปตามระบบของสมอง

ภาพที่ 16



แสดงโครงสร้างตัวอย่างของเซลล์ประสาทชีวภาพ

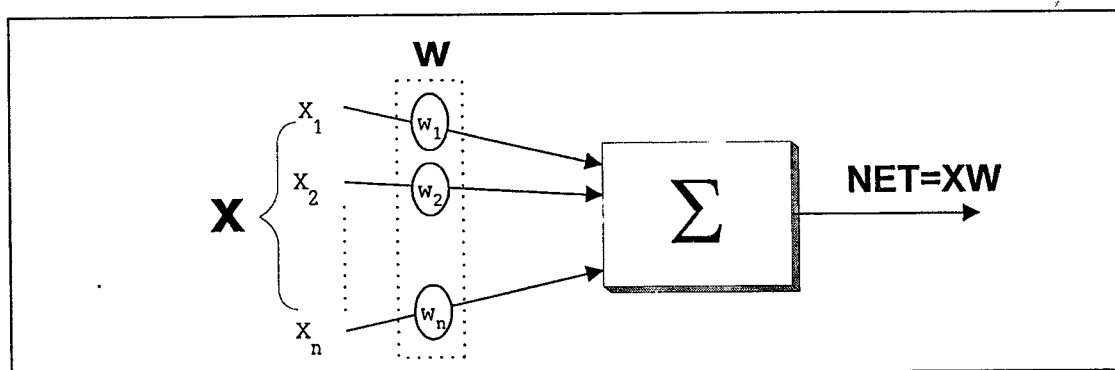
จากภาพที่ 16 ส่วนแขนที่ขยายแยกออกไปจากตัวเซลล์ต่อไปยังเซลล์อื่น ๆ เพื่อรับสัญญาณ เรียกว่า เดนไดรต์ (Dendrites) จุดรับสัญญาณจากเซลล์อื่นเข้ามายังตัวเซลล์จะผ่านมาทางจุดเชื่อมต่อที่เรียกว่า ซินแนปส์ (Synapse) ซึ่งแอกซอน (Axon) จะเป็นตัวส่งสัญญาณเอาที่พุ่งออกไปยังนิวรอนอื่น จากผลการวิจัยพบว่า แต่ละนิวรอนจะเชื่อมต่อออกไปยังนิวรอนอื่นๆ ซึ่งแต่ละนิวรอนจะมีคุณสมบัติในการเพิ่มขยายหรือลดทอนความเข้มของสัญญาณบางสัญญาณที่เข้ามาทางเดนไดรต์ ของเซลล์ (ซึ่งมี

แขนงมากมาย) อาจสามารถกระตุ้นตัวเซลล์ แต่บางสัญญาณก็อาจจะยับยั้งตัวเซลล์ เนื่องจากเซลล์ประสาทหนึ่งเซลล์มีเดนไดรต์มาก ฉะนั้น สัญญาณ กระตุ้นจากเดนไดรต์ ที่รับเข้ามาจากเซลล์ประสาทอื่นๆ จะถูกนำมารวมกันที่ตัวเซลล์ประสาทที่เซลล์ประสาทจะมีค่าเทรชโฮลด์ (Threshold) ค่าหนึ่งหากผลรวมของสัญญาณไฟฟ้าเคมี (Electrochemical) มีค่ามากกว่า เทรชโฮลด์เซลล์ เซลล์ประสาทก็จะส่งสัญญาณขนาดหนึ่งผ่านทางแอกซอนไปยังนิวรอลอื่นๆ การจัดเรียงชั้น (Layer) และลักษณะการเชื่อมโยงระหว่างนิวรอลในสมองนั้นมีการจัดเรียงที่ซับซ้อน สอดคล้องกับหน้าที่การทำงานเฉพาะส่วน มีการเจริญเติบโตสัมพันธ์กับสิ่งแวดล้อมและมีการเรียนรู้ตลอดเวลาซึ่งใช้เวลานานนับปีดังนั้นจึงยากที่จะสร้างโมเดลขึ้นมาเพื่อเลียนแบบให้มีคุณลักษณะคล้ายสมองชีวภาพได้ทั้งหมด ผลงานที่ได้จากการท้าววิจัยในปัจจุบันเป็นเพียงการจำลอง การเลียนแบบและการทำงานเฉพาะบางส่วนของโครงข่ายประสาท มาใช้เฉพาะกับงานใดงานหนึ่ง ซึ่งมีการวิจัยลักษณะของโครงข่ายแบบต่างๆ ขึ้นมา โดยแต่ละแบบจะเหมาะสมกับงานประเภทหนึ่ง ๆ เท่านั้น

โครงข่ายประสาทเทียม (Artificial Neural Networks)

การออกแบบสร้างประสาทเทียมนั้นมีสมมติฐานขั้นแรกจากคุณสมบัติของระบบประสาทชีวภาพ ดังที่กล่าวมา กล่าวคือ ชุดรับสัญญาณข้อมูล อินพุตของเซลล์ประสาทหนึ่งได้จากสัญญาณเอาต์พุตของเซลล์ประสาทอื่นๆ ผ่านทางซินแนปส์และเดนไดรต์ ข้อมูลแต่ละค่าที่รับมาจะถูกลดขนาดด้วย ซินแนปติกส์ ซึ่งภายในประกอบด้วยสารเคมีประเภท K^+, Ca^{++}, Na^+, Cl^- ซึ่งจะมีลักษณะทางความนำพัลส์ (Pulse) สัญญาณไฟฟ้าเคมีที่แตกต่างกัน (James A. Freeman and David M. Skapura, 1991:8-9) ด้วยเหตุนี้ โมเดลประสาทเทียมที่สร้างขึ้น จะต้องมีการถ่วงน้ำหนักให้กับโมเดลก่อนที่จะนำเข้าสู่โมเดลประสาทเทียม จุดนี้เรียกว่า ซินแนปติกส์เวกท์ ปริมาณของข้อมูลที่เข้าสู่นิวรอลจะถูกนำมารวมกัน และตัดสินใจด้วยระดับความสนใจของ นิวรอล (Activation level) แล้วจะส่งเป็นเอาต์พุตออกที่แอกซอนไปยังนิวรอลอื่นๆ

ภาพที่ 17



แสดงโดยแกรมของนิวรอลที่สร้างขึ้น (Artificial Neuron)

จากภาพที่ 17 แสดงถึงโมเดลที่สร้างขึ้นโดยแนวความคิดจากเซลล์สมองชีวภาพ สัญญาณอินพุต คือ $X_1, X_2, \dots, X_n$ จะถูกป้อนเข้าไปยังนิวรอลที่สร้างขึ้น ซึ่งเปรียบเทียบกับสัญญาณที่ป้อนเข้ายังซินแนปส์ของ นิวรอลชีวภาพ สัญญาณอินพุตนี้จะนำไปคูณกับค่าซินแนปติกส์เวกซ์ที่มีค่าตั้งแต่ 0.00-1 (Weight: ค่าที่ใช้ถ่วงน้ำหนัก) $W_1, W_2, \dots, W_n$ ก่อนที่จะเข้าสู่บล็อกซัมเมชัน (Σ : Summation) ซึ่งค่าถ่วงน้ำหนักนี้จะสอดคล้องกับค่าสเตรงท์ (Strength) ของจุดต่อซินแนปส์ชีวภาพแต่ละจุด (Single biological synaptic connection) บล็อกซัมเมชันนี้จะทำหน้าที่สอดคล้องคล้ายกับตัวเซลล์สมองชีวภาพ ผลรวมทางคณิตศาสตร์ของอินพุตและเวกซ์จะได้เป็นเอาต์พุต เราเรียกว่า เน็ต (NET) ซึ่งเราจะรวมกันในรูปของเวกเตอร์ได้ดังนี้

$$NET = X_1W_1 + X_2W_2 + \dots + X_nW_n \dots\dots\dots(4.1)$$

จะได้

$$NET = XW \dots\dots\dots(4.2)$$

ฟังก์ชันกระตุ้นความสนใจ (Activation Function)

เมื่อได้สัญญาณ NET แล้วกระบวนการต่อมาที่นิวรอลต้องทำคือตัดสินใจเราจึงต้องกำหนดฟังก์ชันการ ตัดสินใจ เพื่อใช้เป็นระดับของการตัดสินใจให้กับนิวรอล เพื่อให้ได้สัญญาณเอาต์พุตของ

นิวรอลออกมา ซึ่งเชื่อมต่อไปยังนิวรอลตัวอื่น ๆ เป็นโครงข่าย OUT ที่ได้อาจเป็น Simple linear function โดย

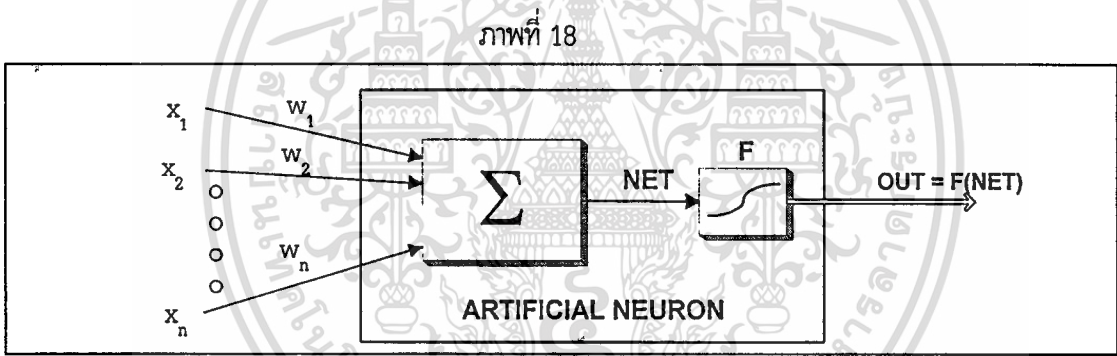
$$OUT = K[NET] \dots\dots\dots(4.3)$$

โดย K เป็นค่าคงที่ ที่เรียกว่า Threshold function
ตัวอย่างเช่น

$$OUT = 1 \text{ ถ้า } NET > T$$

$$OUT = 0 \text{ เมื่อเป็นกรณีอื่น} \dots\dots\dots(4.4)$$

และ T เป็นค่าเทรชโฮลคงที่ หรืออาจเป็น Function อื่น ๆ ที่เลียนแบบคุณสมบัติที่ไม่เป็นเชิงเส้นของเซลล์ประสาทชีวภาพได้อย่างใกล้เคียงกว่า และใช้เป็นฟังก์ชันให้กับโครงข่ายทั่วไปได้

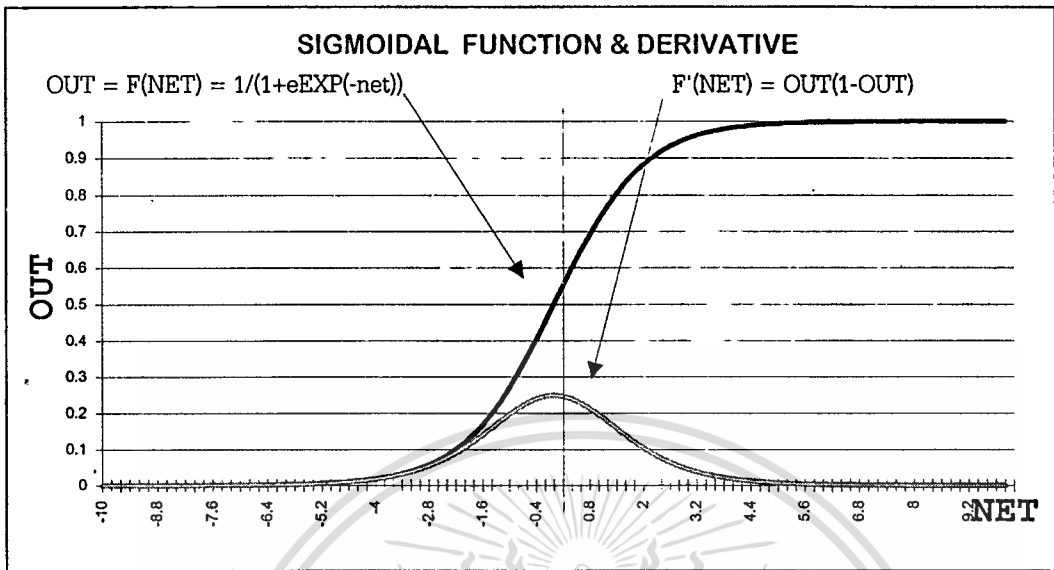


แสดงโมเดลนิวรอลที่สร้างขึ้นร่วมกับ Activation Function

ในภาพที่ 18 บล็อก F จะรับผลที่ได้จาก NET มาสร้างเป็นสัญญาณแอาร์ทพุทที่ OUT โดยการขบวนการภายในบล็อก F จะบีบช่วงของ OUT ให้อยู่ในขอบเขตจำกัด ตามต้องการ ดังนั้น ค่า OUT จะมีค่าไม่ต่ำกว่าช่วงที่กำหนดโดยค่าของ NET เราเรียกบล็อก F นี้ว่าสแควชิ่งฟังก์ชัน (Squashing function) และโดยทั่วไป สแควชิ่งฟังก์ชันที่ใช้เป็นแบบลอจิสติกฟังก์ชัน หรือซิกมอยด์ (Logistic function or “Sigmoid”) ซึ่งมีรูปร่างคล้ายตัว S โดยเขียนเป็นสมการคณิตศาสตร์ได้ดังนี้ คือ

$$F(x) = \frac{1}{(1 - e^{-x})} \dots\dots\dots(4.5)$$

ภาพที่ 19

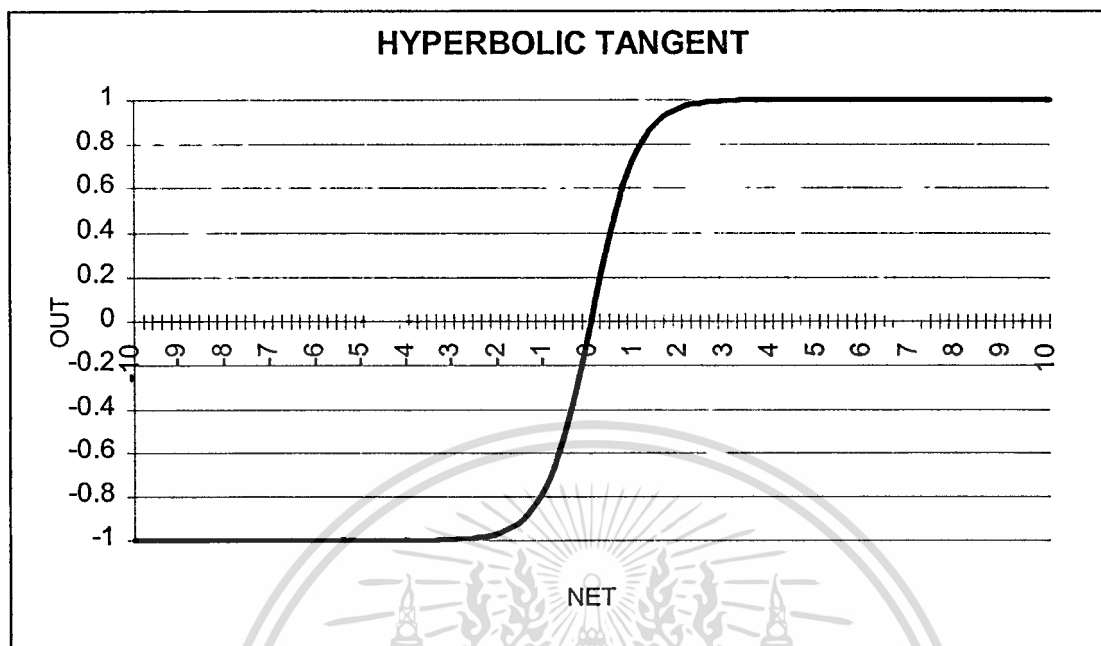


แสดงกราฟที่ได้จากสมการซิกมอยด์ลอจิสติกฟังก์ชัน(Sigmoidal logistic function)

ลักษณะของเทรโซลฟังก์ชันมีลักษณะเป็น Non-linear function เช่น S-Curve เราจะได้ค่าเอาต์พุต ที่มีความไวต่อสัญญาณอินพุตที่มีขนาดเล็กๆ และเฉื่อยต่อสัญญาณแรงๆ ซึ่งสัญญาณอ่อนๆ ไปทางบวกเพียงเล็กน้อยก็จะให้ OUT ใกล้เคียง "1" กระตุ้นหรือสัญญาณอ่อนๆ ทางลบเพียงเล็กน้อยก็จะทำให้ Output ใกล้เคียง "0" (ยับยั้ง) ขณะที่สัญญาณแรงๆ ทางบวกก็ยังคงให้ Output ใกล้เคียง "1" และสัญญาณทางลบแรงๆ ก็คงให้ Output ใกล้เคียง "0" เช่นกัน คุณลักษณะแบบนี้ เป็นแบบ NON-LINEAR GAIN ซึ่งคลอสส์เบอร์ก (Grossberg,1973) พบว่า คุณลักษณะที่เป็น Non-linear gain นี้สามารถแก้ปัญหา Noise-saturation dilemma ได้ และทำให้นิวรอลเทียมที่สร้างขึ้นสามารถทำงานกับขนาดของอินพุตได้กว้างมากขึ้น

ยังมีฟังก์ชันอื่นๆ อีกคือ ไฮเพอร์โบลิค แทนเจนท์ (Hyperbolic tangent) มันจะมีลักษณะคล้ายกับ Logistic function และนิยมใช้บ่อยๆ ในการสร้างโมเดลคณิตศาสตร์ การกระตุ้นแล้วความสนใจของเซลล์สมองเทียม ซึ่งมีคุณสมบัติคล้ายชีวภาพของเซลล์สมอง คือ $OUT = \tanh(X)$

ภาพที่ 20



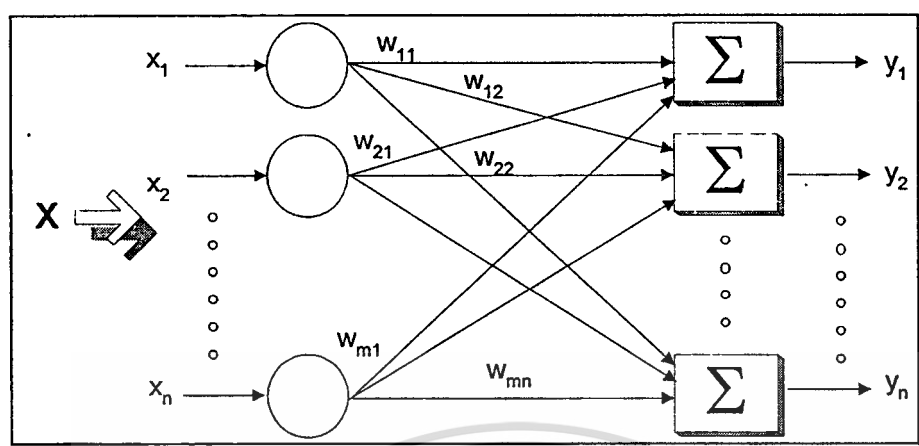
แสดง Hyperbolic Tangent Function

จากภาพที่ 20 ส่วนที่เหมือนกับ ซิกมอยด์ ลอจิสติก ฟังก์ชัน คือมีลักษณะเป็น S แต่เนื่องจากมันจะมีความสมมาตรจึงให้ OUTPUT อยู่ระหว่าง “-1” ถึง “1” OUTPUT จะเป็น “0” เมื่อ NET เป็น “0” OUTPUT เข้าใกล้ “1” เมื่ออินพุตไปทางบวกและเข้าใกล้ “-1” เมื่อ อินพุตมีทิศทางไปทางลบ

โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Artificial Neural Networks)

ที่กล่าวมาจนถึงจุดนี้ เป็นการกล่าวถึงหลักการและเหตุผลในการสร้างเซลล์ประสาทเทียมเพียงหนึ่งเซลล์ โดยใช้แนวความคิดจากเซลล์ประสาทชีวภาพ การจะนำเซลล์ประสาทเทียมมาใช้งานได้นั้น ต้องใช้เซลล์ประสาทเทียมที่มีคุณลักษณะต่างๆ กัน (ค่า Weight จะทำให้คุณสมบัติของเซลล์ประสาทเทียมแต่ละเซลล์มีคุณลักษณะแตกต่างกันไป) มาเชื่อมโยงเป็นโครงข่ายในลักษณะเดียวกับเซลล์สมองชีวภาพเสียก่อน ซึ่งลักษณะการเชื่อมโยงมีหลายชนิด แต่ละชนิดก็มีคุณลักษณะเด่นที่แตกต่างกันไป

ภาพที่ 21



แสดงลักษณะโครงข่ายประสาทเทียมแบบชั้นเดียว (Single-Layer Neural Networks)

จากภาพที่ 21 เป็นโครงข่ายประสาทเทียมแบบชั้นเดียว ที่ประกอบด้วยเซลล์ประสาทเทียมง่ายๆ หลายๆชุด ความสามารถในการคำนวณของโครงข่ายประสาทเทียมได้มาจากลักษณะการเชื่อมต่อเป็นโครงข่ายประสาทเทียมโครงข่ายง่ายๆ เป็นกลุ่มโมดูลประสาทเทียมที่เชื่อมต่อกัน เป็นชั้น ๆ (Layer) ในภาพที่ 22 เป็นโครงข่ายประสาทเทียม แบบชั้นเดียว (Single layer) ที่ประกอบด้วยเอาต์พุทเลเยอร์ (กลุ่มของบล็อกซิมเมชันที่อยู่ทางขวามือ) และอินพุทเลเยอร์(วงกลมทางซ้ายมือ)โดยไม่พิจารณาอินพุทเลเยอร์ว่าเป็นนิวรอลเลเยอร์ เนื่องจากอินพุทเลเยอร์จะทำหน้าที่เชื่อมต่ออินพุทที่รับมาและส่งออก ไปให้ยังแต่ละอินพุทนิวรอลเลเยอร์ (ในที่นี้คือ Output layer) ในชั้นถัดไป โดยแต่ละอินพุทจะถูกคูณโดยค่าเวกต์เฉพาะแต่ละอินพุท โครงข่ายประสาทเทียมที่สร้างขึ้นในชั้นแรกไม่ซับซ้อน โดยแต่ละนิวรอลจะได้เอาต์พุทจาก

$$OUT = \text{Logistic Function} \text{ คูณ (ผลรวมของ Input } X \text{ กับ Weight)}$$

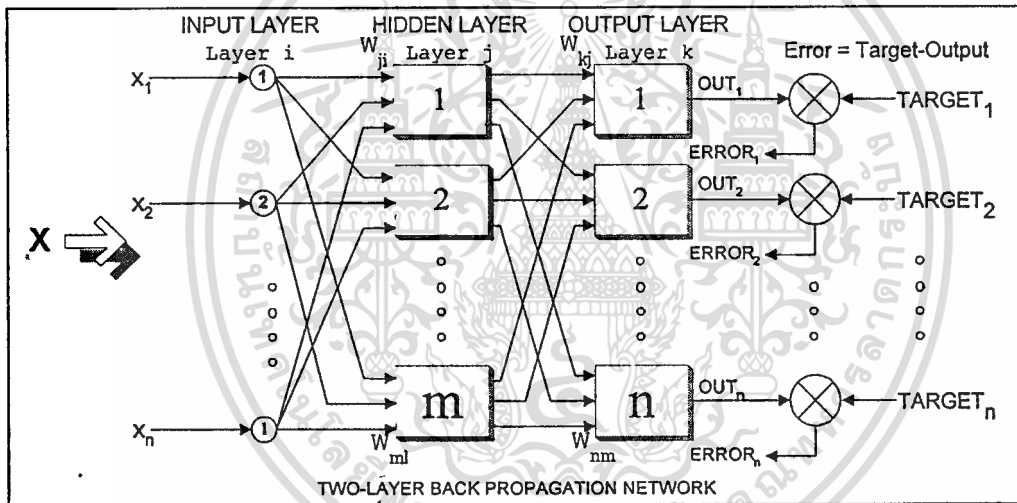
หรือ
$$OUT = F(NET) \dots\dots\dots(4.6)$$

อย่างไรก็ดีลักษณะการเชื่อมโยงระหว่างโครงข่ายไม่ได้มีแบบเดียว การเชื่อมโยงระหว่างเลเยอร์อาจมีการเชื่อมโยงย้อนกลับมาที่อินพุทเลเยอร์อีก ซึ่งโครงข่ายประสาทชีวภาพก็มีลักษณะดังกล่าวเช่นกัน สำหรับค่า Weight ในภาพที่ 22 มีวิธีการพิจารณาในรูปของ เวกต์เมตริก (Weight matrix) ซึ่งหากโครงข่ายมีหลายชั้น จะช่วยให้ระบุค่าเวกต์ได้ง่ายขึ้น และเพื่อหลีกเลี่ยงความสับสนจะกำหนดเป็นไดเมนชัน (Dimensions) ของเมตริก โดยให้ m แทนจำนวนแถว หรือจำนวนของอินพุทและ n แทนจำนวนของนิวรอล ที่สร้างขึ้น ตัวอย่างเช่น เวกต์ ที่เชื่อมระหว่างอินพุท ตัวที่ 4 กับนิวรอลตัวที่ 2 คือ $W_{4,2}$

โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Artificial Neural Networks)

โครงข่ายที่ซับซ้อนจะมีความสามารถในการคำนวณที่ดีขึ้นแน่นอนจะเป็นโครงข่ายที่มีโครงสร้างเป็นจินตนาการที่น่าเป็นไปได้โดยการจัดการเชื่อมโยงนิวรอนมีโครงสร้างเป็นชั้นๆ คล้ายส่วนหนึ่งของสมอง และมีการพัฒนาอัลกอริทึมเกี่ยวกับการฝึกสอนให้โครงข่ายแบบหลายชั้นทำงานได้ ตามความต้องการแล้วเมื่อไม่นานมานี้ โครงข่ายแบบหลายชั้น อาจจะสร้างจาก กลุ่มของโครงข่ายแบบชั้นเดียวเอาท์พุทของ Layer หนึ่ง จะใช้เป็นอินพุทของ Layer ถัดไป ในภาพที่ 22 แสดงเน็ตเวิร์คที่มีการเชื่อมต่อแบบสองชั้น

ภาพที่ 22



แสดงโดยแกรมของ Backpropagation Neural Networks แบบสองชั้น

ภาพที่ 22 แสดงโครงข่ายประสาทเทียมแบบหลายชั้น ที่ต่อเชื่อมโยงแบบเต็มชั้น ในโครงข่ายแบบหลายชั้นมีการเรียกชื่อชั้นต่างๆ ดังนี้ คือ ชั้นที่ต่อโดยตรงกับอินพุท เรียกว่า อินพุทเลเยอร์ (Input layer) ชั้นนี้จะไม่มีการคำนวณ แต่จะทำหน้าที่ต่อเชื่อมข้อมูลไปยังชั้นถัดไป ชั้นที่อยู่ท้ายสุดทางขวามือ เรียกว่า เอาท์พุทเลเยอร์ (Output layer) เป็นชั้นที่โครงข่ายจะให้ผลลัพธ์ ส่วนชั้นที่อยู่ระหว่างอินพุทเลเยอร์และเอาท์พุทเลเยอร์ จะมีกี่ชั้นก็ตามจะเรียกว่า ฮิดเดนเลเยอร์ (Hidden layer) หากฮิดเดนเลเยอร์มีหลาย ๆ ชั้นก็จะมีการตั้งชื่อเฉพาะลงไปที่กับแต่ละชั้น

ฟังก์ชันกระตุ้นความสนใจแบบไม่เป็นเชิงเส้น (The Nonlinear Activation Function)

การนำเอาทพุทของเลเยอร์หนึ่ง มาเชื่อมกับอินพุทของเลเยอร์ชั้นถัดไป โดยผ่านฟังก์ชันกระตุ้นความสนใจแบบไม่เป็นเชิงเส้น จะทำให้โครงข่ายมีความสามารถในการคำนวณเพิ่มขึ้น (หากไม่ผ่านฟังก์ชันดังกล่าวความสามารถการคำนวณจะไม่เพิ่มขึ้นและจะมีความสามารถไม่แตกต่างไปจาก Single layer networks) และสามารถกำหนดขอบเขตของเอาท์พุทให้อยู่ในช่วงที่ต้องการได้

การฝึกสอนให้กับโครงข่ายประสาทเทียม (Training of Artificial Neural Networks)

ค่าเวกซ์ มีความสัมพันธ์กับอะไร ? เปลี่ยนแปลงอย่างไร ? นั่นก็เช่นเดียวกับเด็กที่คลอดออกมา ก็มีสมองแล้วแต่สมองยังไม่เจริญเติบโตเพียงพอและยังไม่ได้รับการฝึกสอนและเรียนรู้ เด็กจึงไม่สามารถทำกิจกรรมใดๆ ด้วยตนเอง เว้นแต่กิจกรรมที่ธรรมชาติสร้างมาพร้อมกับการกำเนิดที่เรียกว่า “สัญชาตญาณ” ซึ่งธรรมชาติใส่คุณลักษณะบางอย่างให้เซลล์สมองบางส่วนตั้งแต่ทารกเจริญเติบโตอยู่ในครรภ์มารดา เช่น ระบบควบคุมการหายใจ, ความรู้สึก, การเรียกร้องเมื่อหิว, การตอบสนองต่อสิ่งเร้า ฯลฯ เด็กจะพัฒนาการเรียนรู้ไปตามขั้นตอน หลังจากนั้นสมองของเขาจะได้รับการฝึกสอน และเจริญเติบโตไปพร้อมๆ กัน เซลล์สมองจะได้รับการปรับคุณลักษณะสอดคล้องกับการฝึกสอน และจะเจริญเป็นโครงข่ายสอดคล้องกัน

โครงข่ายประสาทเทียมที่สร้างขึ้นมีลักษณะเช่นเดียวกัน คือ เมื่อสร้างเสร็จ แต่ละเซลล์ประสาทที่สร้างขึ้นมานั้น จะยังไม่มีคุณลักษณะใดเลย เนื่องจากยังไม่มีกำหนดค่าซินแนปติกส์เวกซ์ที่เหมาะสมกับงานที่ต้องการให้กับมัน จึงต้องมีการฝึกสอนเพื่อให้เน็ตเวิร์คที่สร้างขึ้นมีคุณลักษณะตามที่ต้องการ การฝึกสอนของโครงข่ายประสาทเทียม จะกระทำโดยการปรับเปลี่ยนค่าซินแนปติกส์เวกซ์เพื่อให้โครงข่ายจดจำ แพตเทิร์นความสัมพันธ์ระหว่างอินพุทกับเอาท์พุทได้ โดยในขั้นแรกอาจกำหนดเป็นค่าสุ่มใดๆ (Random weight) ก่อนแล้วจึงปรับเปลี่ยนเวกซ์ไปตามอัลกอริทึมสมมติฐาน หลากๆ รอบจนกว่าจะได้เอาท์พุทของเน็ตเวิร์ค เหมือนกับเอาท์พุทที่ต้องการ ในเงื่อนไขความผิดพลาดที่ยอมรับได้

วัตถุประสงค์ของการเทรนนิ่ง (Objective of Training)

เนื่องจากค่าเวกต์ที่ให้เป็นค่าสุ่มใดๆ โครงข่ายจึงไม่แสดงคุณลักษณะได้ออกมา การฝึกสอน (Training) ให้โครงข่ายก็คือการปรับค่าเวกต์ทุก ๆ จุดให้สอดคล้องกับอินพุตหลายๆแบบ เพื่อให้ได้เอาต์พุตตามต้องการนั่นเอง การฝึกสอนโครงข่าย จะต้องบรรลุถึงกระบวนการเข้าใจพื้นฐานเสียก่อน คือ การเรียนรู้ในโครงข่ายประสาทเทียมนั้นมีขีดจำกัด ปัญหาต่างๆ ผู้ใช้คงต้องแก้ไขมันก่อน แล้วนำผลนั้นไปอ้างอิงสำหรับการปรับปรุงค่าเวกต์ หลังจากปรับเวกต์จนได้ค่าผิดพลาดที่เอาต์พุตเทียบกับเป้าหมาย น้อยลงเป็นที่พอใจแล้ว โครงข่ายประสาทเทียมนั้นก็พร้อมที่จะวิเคราะห์อินพุตและให้เอาต์พุตตามลักษณะตัวอย่างที่มันเคยเรียนรู้มา การเรียนรู้จะมีการปรับเวกต์หลายๆรอบ จนค่าเวกต์สอดคล้องกับตัวอย่างหลายๆ ตัวอย่าง และให้เอาต์พุตตามต้องการ พบว่าโครงข่ายได้ตัวอย่างสำหรับการเทรนนิ่งมากๆ โครงข่ายก็จะมีคามแม่นยำสูงขึ้น แต่ก็ใช้เวลาในการเทรนนิ่งเพิ่มขึ้นเช่นกัน หากพิจารณาต่อไปจะพบว่า โครงข่ายประสาทเทียมที่สร้างขึ้นจะมีพฤติกรรมคล้ายกับระบบการเรียนรู้ของมนุษย์มาก เป็นเพราะมีต้นแบบมาจากระบบประสาทชีวภาพนั่นเอง

การเทรนนิ่งแบบควบคุม (Supervised Training)

เทรนนิ่งอัลกอริทึมถูกจัดเป็น 2 ประเภท คือ แบบควบคุม (Supervised training) และแบบอิสระ (Unsupervised training) โดย การเทรนนิ่งแบบควบคุม จะต้องการคู่ของการเทรนนิ่งระหว่างอินพุตกับเป้าหมายที่ต้องการ ที่เรียกว่า เทรนนิ่งแพร์ (Training pairs) โครงข่ายจะถูกเทรนไปตามจำนวนของคู่ที่เทรนนิ่ง (จำนวนคู่ของ Input กับ Output ที่ต้องการให้โครงข่ายรู้จัก) เอาต์พุตที่คำนวณได้จากโครงข่ายจะถูกเปรียบเทียบกับความสอดคล้องกับเป้าหมาย ค่าผิดพลาดที่เกิดขึ้นจะถูกป้อนกลับไปยังโครงข่ายและเปลี่ยนแปลงค่าเวกต์ให้สอดคล้องกับอัลกอริทึม ที่ทำให้แนวโน้มของค่าผิดพลาดที่เกิดขึ้นระหว่างเอาต์พุตกับเป้าหมายโดยเฉลี่ยมีค่าลดต่ำลง ตัวอย่างการเทรนนิ่งแบบนี้ ได้แก่ การเทรนนิ่งแบบแพร่กลับ (Back propagation)

การเทรนนิ่งแบบอิสระ (Unsupervised Training)

ถึงแม้ว่าอัลกอริทึมแบบควบคุม (Supervised training) สามารถจะประยุกต์ใช้เพื่อปรับคุณลักษณะของโครงข่ายได้สำเร็จ แต่ก็ยังมีข้อวิจารณ์อยู่ คือ มันเป็นไปอย่างแบบชีวภาพไม่ได้ และยากที่จะเชื่อได้ว่า กลไกการเทรนนิ่งของสมองต้องการ การเปรียบเทียบระหว่างค่าที่ต้องการกับเอาต์พุตจริง โดยกระบวนการป้อนกลับไปแก้ไขคุณลักษณะของโครงข่าย และถ้าสมมติว่า ถ้าสมองมีกลไกเช่นนี้ ต้องมีผู้หาเอาต์พุตที่ต้องการเพื่อนำมาเป็นเป้าหมายตลอดเวลา และจะเอามาจากที่ใด ? สรุปคือ ต้องมีผู้คิดเป้าหมายให้กับโครงข่ายก่อน โครงข่ายไม่สามารถคิดและปรับคุณลักษณะได้ก่อนด้วยตนเอง ในทางตรงกันข้ามหากพิจารณาทารกแรกเกิดสมองของเขาสามารถจัดระบบเองได้อย่างไร การเทรนนิ่งแบบอิสระ (Unsupervised learning) ที่สร้างขึ้นคงยังห่างไกลความเป็นไปได้ ที่จะมีลักษณะการเทรนนิ่งแบบระบบของสมอง จนกระทั่งมีการพัฒนาการ เทรนนิ่งแบบอิสระนี้ขึ้นราวปี 1984 โดย โคโฮเนน (Kohonen) และบุคคลอื่นๆ โดยได้เสนอแนวคิดที่เป็นการเทรนนิ่งแบบไม่ต้องการเป้าหมาย ไม่มีการตัดสินใจด้วยเหตุผลในอุดมคติมาก่อน ชุดของการเทรนนิ่ง จะมีเพียงอินพุต เวกเตอร์เท่านั้น เทรนนิ่งอัลกอริทึมจะเปลี่ยนแปลงค่าเวกซ์ของโครงข่าย เพื่อสร้างเอาต์พุตที่มีความมั่นคง ยกตัวอย่าง เช่น หากให้โครงข่ายรู้จำภาพหน้าคนหนึ่ง หากภาพหน้าคนคนนั้น เปลี่ยนแปลงไปเล็กน้อย (Image อาจมี Noise ร่วมอยู่บ้าง) โครงข่ายนั้นก็ยังสามารถบอกได้ว่า คนคนนั้นเป็นคนเดิมเป็นต้น การเทรนนิ่งจะไม่มีการตัดสินใจมาก่อน ไม่มีการกำหนดแบบเอาต์พุตมาก่อน (อาจกล่าวได้ว่าแบบเอาต์พุตจะถูกกำหนดโดยอินพุตเวกเตอร์นั่นเอง) ดังนั้น เอาต์พุตของโครงข่ายก็เช่นกัน ส่วนใหญ่จะถูกแปรรูปซึ่งจะเข้าใจได้ภายหลังกระบวนการเทรนนิ่ง ดังนั้นจึงไม่สามารถแก้ปัญหาที่เคร่งครัดสำคัญได้ แต่มักนิยมใช้โครงข่ายแบบนี้กับงานง่ายๆ ประเภทการเปรียบเทียบเอกลักษณ์, รูปแบบที่สัมพันธ์กันระหว่างอินพุต-เอาต์พุต ที่ถูกกำหนดโดยโครงข่าย

วิธีการแก้ปัญหาการฝึกสอน (Traning Algorithm)

ส่วนใหญ่แล้วทุกวันนี้ การแก้ปัญหาฝึกสอนของโครงข่ายค่อยๆ พัฒนาก้าวหน้าขึ้นจากแนวความคิดของ ดี โอ เฮบบ์ (ปี 1961) เขาได้เสนอโมเดลของ การเทรนนิ่ง แบบอิสระ (Unsupervised training) ในแบบซินแนปติกส์สเตรงท์ หรือเวกซ์ ซึ่งจะเพิ่มขึ้น ถ้าทั้งแหล่งกำเนิด (Input Source) และจุดหมายปลายทาง (Destination) ของนิวรอนได้รับการสนใจ กรณีนี้ถ้ามีการใช้งานทางเส้นนี้บ่อยๆ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก็จะทำให้ซินแนปติกส์สเตรงจ์(หรือเวกต์)แข็งแรงขึ้น (เซลล์สมองที่ใช้งานมากบ่อยๆก็จะทำให้ ซินแนปส์ใหญ่ขึ้น การส่งผ่านข้อมูลพัลส์ไฟฟ้าทำได้ดีขึ้น ทำให้มีประสิทธิภาพดีขึ้น เช่นสามารถคิดหรือจดจำได้เร็วและดีขึ้น) โครงข่ายประสาทเทียมนี้ ใช้การเรียนรู้แบบเฮบบ์ (Hebbian learning) จะเพิ่มค่าเวกต์ของโครงข่ายอย่างสอดคล้องกับผลคูณของระดับความสนใจของแหล่งกำเนิดและจุดหมายของนิวรอนตามสมการดังนี้

$$W_{ij}(n+1) = W_{ij}(n) + \alpha OUT_i OUT_j \dots\dots\dots(4.7)$$

โดย $W_{ij}(n)$ คือค่าเวกต์ จากนิวรอน i ไปยังนิวรอน j ก่อนปรับปรุงค่า

$W_{ij}(n+1)$ คือค่าเวกต์ จากนิวรอน i ไปยังนิวรอน j หลังปรับปรุงค่า

α คือค่าคงที่ของการเรียนรู้ (Learning rate coefficient)

OUT_i คือเอาต์พุตของนิวรอน i และเป็นอินพุตของนิวรอน j

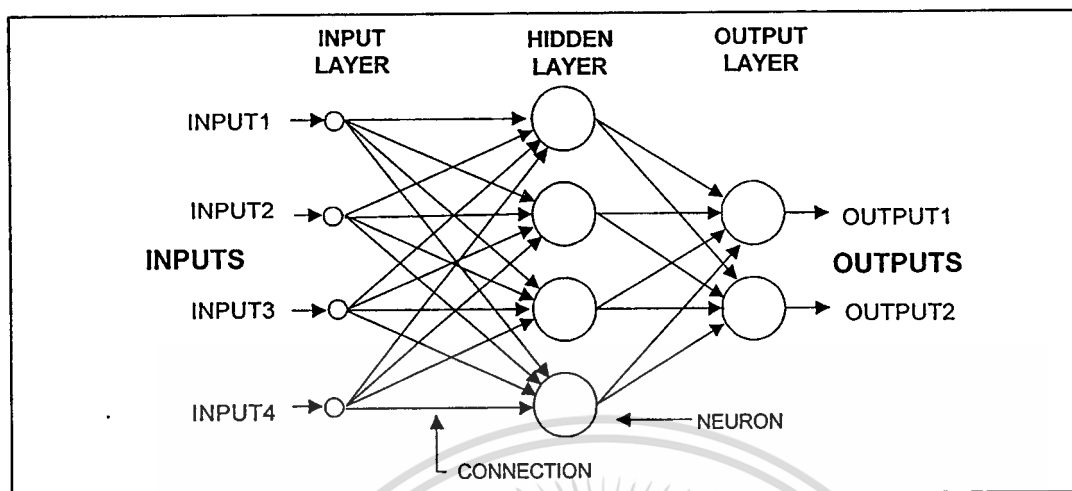
OUT_j คือเอาต์พุตของนิวรอน j

โครงข่ายที่มีลักษณะการเทรนนิ่งแบบ เฮบบ์ นั้น เป็นผลมาจากการพัฒนามาแล้วกว่า 20-30 ปี โดยเฉพาะงานของโรเซนเบลท์ (Rosenblatt:1962), วิโดว์(Widrow:1959), วิโดว์และฮอฟฟ์ (Widrow&Hoff:1960) และอีกหลายๆ คนที่พยายามพัฒนาระบบการเทรนนิ่งแบบควบคุม ที่สร้างโครงข่ายที่สามารถเรียนรู้แบบของอินพุตได้อย่างกว้างขวางและมีอัตราการเรียนรู้สูง ที่บรรลุผลได้จากหลักการพื้นฐานของการเรียนรู้หรือเทรนนิ่งให้กับโครงข่ายเช่น Perceptrons, Hopfieldnets, Backpropagation Networks และCounter propagation.

การประยุกต์ใช้นิวรอนตเวอ์กับงานวิจัย

จากการศึกษาพฤติกรรม การสื่อสารของมนุษย์พบว่า การเรียนรู้ภาษาต้องเรียนรู้จากผู้อื่น ภาษาไม่ใช่สิ่งที่ป็นสัญชาตญาณ ดังนั้น จึงต้องเรียนรู้จากภาษาที่สังคมนั้นกำหนดขึ้นซึ่งเป็น แพตเทิร์นของภาษาที่บรรพบุรุษได้สร้างขึ้นสั่งสมกันมาก่อน จากเหตุผลดังกล่าว พบว่าสอดคล้องกับอัลกอริทึม การเทรนนิ่งแบบควบคุม (Supervised training) ผู้วิจัยเลือกใช้วิธี Backpropagation มัลติเลเยอร์เป็นแบบ 2 เลเยอร์ แสดงดังภาพที่ 23

ภาพที่ 23



แสดงไดอะแกรมของ Two Layer Feed-forward

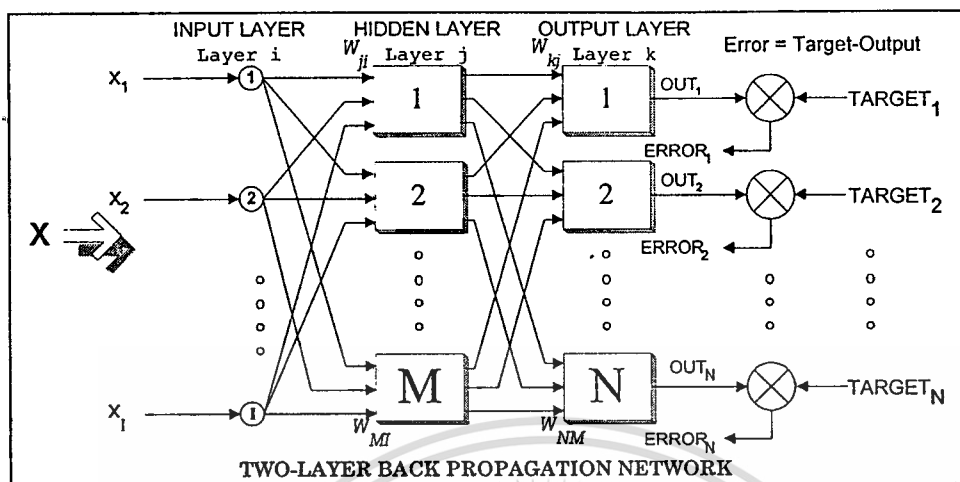
พิจารณาภาพที่ 23 ประกอบ จะเห็นว่าประกอบด้วย input array, hidden layer และ output layer ในส่วนของ Input array จะไม่ถือเป็น layer เพราะไม่มีการคำนวณใด ๆ แต่มันจะทำหน้าที่เชื่อมโยงข้อมูลให้กับ Hidden layer ทุกๆนิวรอน output ของ hidden layer ก็จะถูกเชื่อมโยงไปเป็น input ให้กับ output layer ทุกอินพุตของนิวรอนทุกตัว จะมีค่า Weight ประจำตัว ค่า Weight นี้ทำให้ระดับของอินพุตผ่านเข้าไปภายในนิวรอนแต่ละตัวไม่เท่ากัน ซึ่งเป็นตัวกำหนดคุณลักษณะสมบัติของนิวรอนแต่ละตัวในเน็ตเวิร์ค วิธีการกำหนดคุณสมบัติดังกล่าวทำได้โดยการเทรนนิ่งโดยในขั้นแรกจะสมมุติว่า ค่าแบบสุ่มน้อย ๆ ให้(ช่วงที่ใช้คือตั้งแต่-0.5ถึง+0.5)แล้วจึงค่อยๆ ปรับเปลี่ยนค่าเวกทีในแต่และ layer ให้สอดคล้องกับคุณสมบัติที่ต้องการ

กระบวนการเทรนนิ่ง

กระบวนการเทรนนิ่ง คือ กระบวนการปรับคุณลักษณะของนิวรัลเน็ตเวิร์ค ให้มีคุณสมบัติตามที่ต้องการ การเทรนนิ่งของอัลกอริทึม Backpropagation เป็นแบบ Supervised Training แบ่งเป็น 2 ส่วนคือ Forward pass และ Reversed pass การเทรนนิ่ง จะกระทำทั้ง Forward pass และ Reverse pass สลับกันไป จนกว่า ค่าผิดพลาดระหว่าง Output กับ Target จะมีค่าต่ำกว่าจุดที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 24



แสดงโดยแกรมเชื่อมโยงระหว่างชั้นของนิวรอลแบบ 2 ชั้นของ Backpropagation Networks

FORWARD PASS

จากภาพที่ 24 อินพุต X ตั้งแต่ X_1 ถึง X_I จะถูกคูณด้วย Weight matrix ผลรวมของอินพุตเมื่อคูณกับค่า Weight ของแต่ละนิวรอลเรียกว่า เน็ต (NET) เขียนเป็นสมการได้คือ

$$NET_j = X_1 W_{j1} + X_2 W_{j2} + \dots + X_I W_{ji} \dots (4.8)$$

หรือ

$$NET_j = \sum_{i=1}^I X_i W_{ji} \dots (4.9)$$

โดย X คือ ค่าที่รับเข้ามาของ Input node

W คือ ค่าเวรท์ประจำตัวของแต่ละอินพุตโนด

i คือ หมายเลขโนดโดยเริ่มที่โนด 1 ถึง โนด I

j คือ หมายเลขนิวรอลใน Hidden layer ตั้งแต่ นิวรอลที่ 1 ถึง M

k คือ หมายเลขนิวรอลใน Output layer ตั้งแต่ นิวรอลที่ 1 ถึง N

และทำนองเดียวกัน

$$NET_k = \sum_{j=1}^M Op_j W_{kj} \dots (4.10)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อได้ค่า NET ของแต่ละนิวรอนแล้วจะปรับให้ได้ Output อยู่ในช่วงที่ต้องการเป็นค่าที่จะนำไปยับยั้งหรือกระตุ้นนิวรอนตัวถัดไปโดย Squashing Function ที่ใช้เป็น Logistic Function หรือ Sigmoid จะได้ตามสมการ 4.11 และ 4.12

$$Op_j = F_j(NEt_j) = \frac{1}{1 + e^{-(net/\theta_j)}} \dots\dots\dots(4.11)$$

$$Op_k = F_k(NEt_k) = \frac{1}{1 + e^{-(net/\theta_k)}} \dots\dots\dots(4.12)$$

กำหนด $\theta_j = 0.05$ และ $\theta_k = 0.05$

θ_j และ θ_k คือ ค่าที่กำหนดความชันของ Sigmoid curve

Op_i คือ Output ที่ Neuron j ของ Hidden Layer (เป็น Input ของ Output Layer)

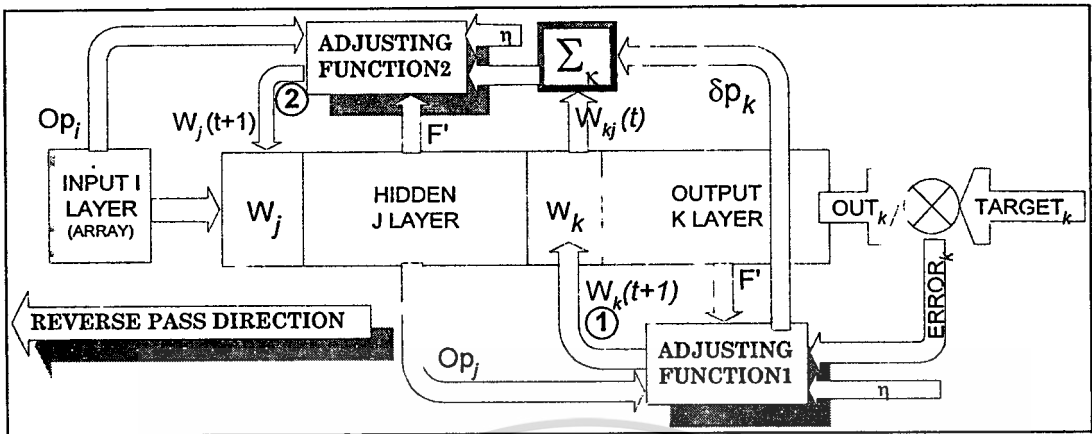
Op_k คือ Output ที่ Neuron k ของ Output Layer (เป็น Output ของ Output Layer)

การคำนวณหาค่า Output จะเริ่มที่ Hidden layer ที่ชั้น j ก่อน เมื่อได้ค่า output ของ Hidden layer ครบทุกตัวแล้ว จึงคำนวณหาค่า Output ที่ Output layer โดย Input ของ Output layer ก็คือ Output ของ Hidden layer นั้นเอง โดยใช้การคำนวณเหมือนกันกับ Hidden layer Output ที่ได้จะนำไปเปรียบเทียบกับ Target ที่ต้องการเกิดเป็น Error ขึ้น ค่า Error นี้ จะถูกนำไปใช้ร่วมกับค่าดิฟเฟอเรนเชียล ของ $F'(F')$ ใน Output layer เพื่อปรับค่า Weight ให้กับ Output layer ดังจะกล่าวต่อไปในส่วนของ Reverse pass

REVERSE PASS

เป็นส่วนที่ป้อนกลับค่า Error(Target-Output) จากเอาท์พุทเลเยอร์ย้อนกลับมาปรับปรุงค่าเวกซ์ที่ เลเยอร์ก่อนหน้า สำหรับนิวรอนเน็ตเวิร์คที่มีมากกว่า 1 เลเยอร์การปรับปรุงค่าเวกซ์ที่ชั้นถัดถอยมาจะใช้ค่าเวกซ์เดิมของเลเยอร์ที่สูงกว่า ร่วมกับตัวร่วมทางด้าน Output ของเลเยอร์มาคำนวณ (เนื่องจากไม่มีค่า Error มาใช้เหมือน Output layer) ดังจะกล่าวในรายละเอียดดังต่อไปนี้

ภาพที่ 25



แสดงไดอะแกรมของการปรับค่า Weight ในชั้น Output layer และ Hidden layer

จากภาพที่ 25 การปรับเปลี่ยนค่าเวทใหม่ จะเป็นไปตามทิศทางของลูกศร และได้ข้อมูลจากค่า Error ที่ Output layer เป็นต้นกำเนิดเพื่อกำหนดทิศทางปรับค่าเวทให้ในนิวรัลเน็ตเวิร์ค มีคุณลักษณะสมบัติตามต้องการ ลำดับขั้นตอนของ Reverse pass พอสรุปได้ดังนี้คือ

1. นำ Output ที่ได้จาก Output Layer ไปลบออกจาก Target ได้เป็น error
2. นำค่า Error มาคูณกับ F' ที่ได้จาก Output layer เรียกว่า δp_k
3. นำค่า δp_k คูณกับ Op_i แล้วคูณกับ η (อัตราการเรียนรู้ 0.01-1)
4. ค่า Weight ค่าใหม่ที่ได้ คือ ค่า Weight เดิมบวกกับผลลัพธ์ ที่ได้จากข้อ 3

พอเขียนเป็นสมการใหม่ได้คือ

$$W_{kj}(t+1) = W_{kj}(t) + \eta \delta p_k Op_j \quad (4.13)$$

$$\text{เพราะว่า } \eta \delta p_k Op_j = \Delta W_{kj} \quad (4.14)$$

$$\text{ฉะนั้น } W_{kj}(t+1) = W_{kj}(t) + \Delta W_{kj} \quad (4.15)$$

เมื่อ $W_{kj}(t)$ คือ ค่าเวทเดิมก่อนปรับ

$W_{kj}(t+1)$ คือ ค่าเวทหลังปรับ

η คือ อัตราการเรียนรู้

Op_j คือ Output ของ Layer j

δp_k คือ ค่าที่ได้จาก Error คูณกับดิฟเฟอเรนเชียลของฟังก์ชัน F

การปรับค่าของ Layer ถัดมา (Hidden layer) มีลักษณะคล้ายคลึงกับ Output layer คือ

$$W_{ji}(t+1) = W_{ji}(t) + \eta \delta p_i Op_i \quad (4.16)$$

$$\text{เพราะ } \eta \delta p_i Op_i = \Delta W_{ji} \quad (4.17)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้น
$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji} \dots\dots\dots(4.18)$$

สมการที่ 4.17 แตกต่างกับสมการ 4.14 ตรงการได้มาซึ่ง δp_i เพราะไม่มีค่า Output กับ Target ที่ Hidden layer ค่าดังกล่าวสามารถทดแทนได้ด้วย $\sum_k (\delta p_k W_{kj})$

โดย
$$\delta p_j = Op_j(1 - Op_j)(\sum_k (\delta p_k W_{kj})) \dots\dots\dots(4.19)$$

เมื่อ $W_{ki}(t)$ คือ ค่าเวกต์เดิมก่อนปรับ

$W_{ki}(t+1)$ คือ ค่าเวกต์หลังปรับ

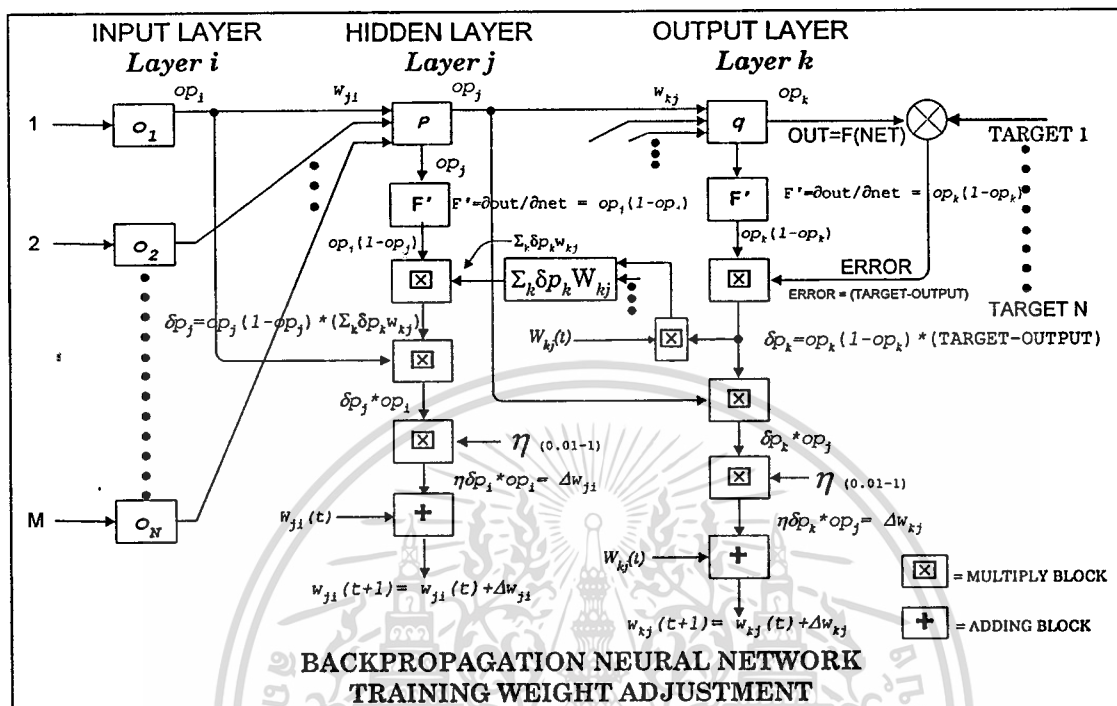
η คือ อัตราการเรียนรู้

Op_i คือ ข้อมูลที่ได้จาก Input layer

δp_i คือ ค่าความคุมให้เวกต์เปลี่ยนแปลงเข้าสู่จุดที่ดีที่สุด

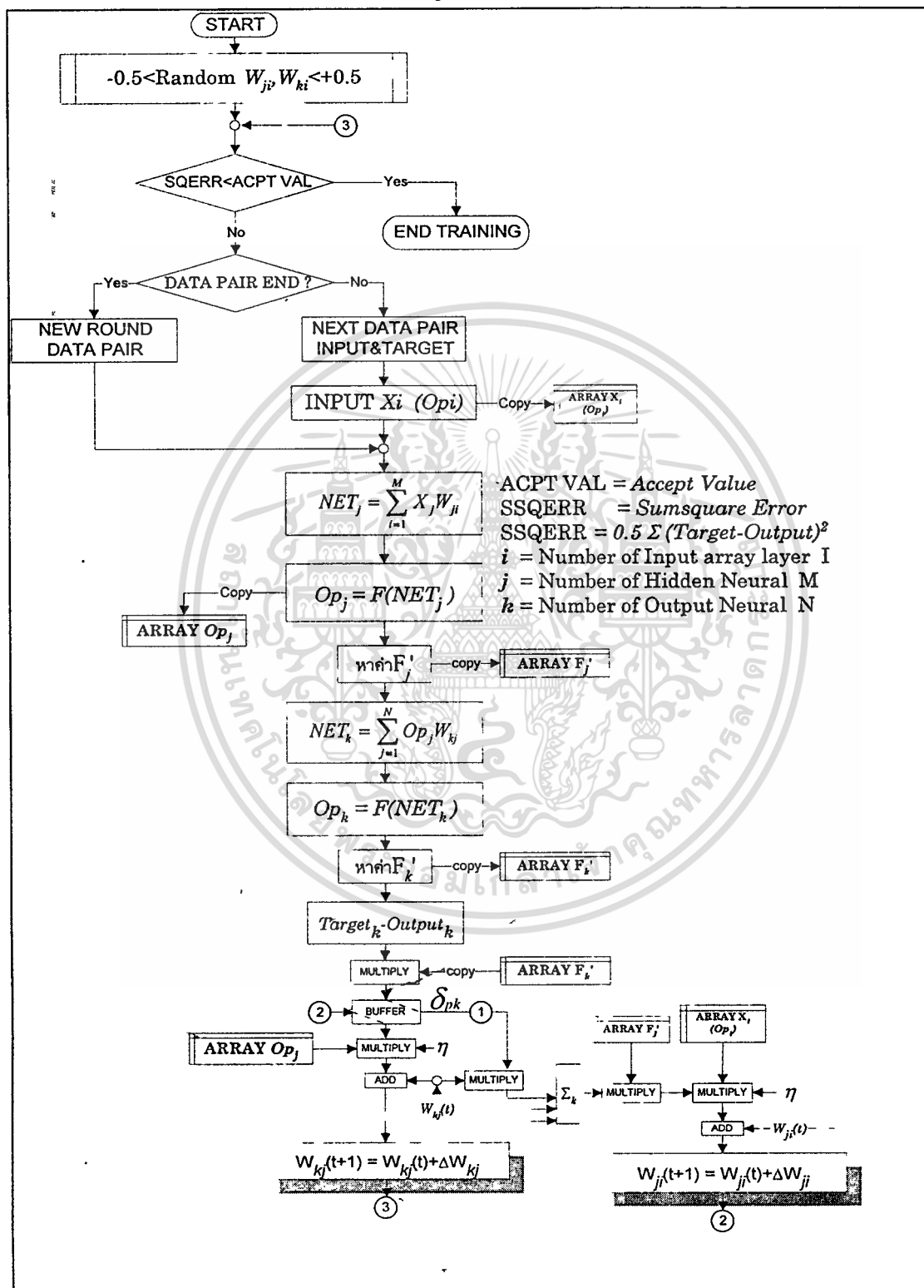
การปรับค่าเวกต์จะปรับทีละเลเยอร์จาก Output layer กลับไปยัง Input layer เมื่อเสร็จสิ้นก็จะกลับสู่ส่วนของ Forward pass อีกครั้ง เพื่อหา Output มาเปรียบเทียบกับ Target และดำเนินการปรับค่าเวกต์ในส่วนของ Reverse pass เช่นนี้สลับกันไปจนกว่าค่า Error ที่ได้จะลดลงต่ำกว่าค่าที่ต้องการจึงหยุดกระบวนการ ค่าเวกต์ที่ได้จะเป็นค่าที่เหมาะสมสำหรับทุกคู่ของ Input และ Target ซึ่งพร้อมที่จะนำไปใช้ในกระบวนการตรวจสอบ การปรับค่าเทรนนิ่งเวกต์ของ Backpropagation Networks พออธิบายได้ด้วยไดอะแกรมดังนี้

ภาพที่ 26



ไดอะแกรมแสดงวิธีการหาค่าเกณฑ์ที่เหมาะสมของการเทรนนิ่งโดยใช้อัลกอริทึมแบบแบคพรอปพาคชั่น และพอเขียนเป็น Flow Chart ได้ดังภาพที่ 27

แผนภูมิที่ 27



แสดงFlow chart แสดงกระบวนการเทรนนิ่ง เพื่อหาค่าเวกซ์ ที่เหมาะสมในกระบวนการทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เงื่อนไขการหยุดการฝึกสอน

การกำหนด เงื่อนไขการหยุดฝึกสอนทำได้สองกรณีคือ

1. เมื่อผลรวมของค่าผิดพลาดเฉลี่ยระหว่างเอ้าท์พุทกับเป้าหมายทั้งหมด(SSOERR)ลดลง น้อยกว่าค่าที่กำหนด โดยกำหนดสมการหาค่าผิดพลาดเฉลี่ยของเอ้าท์พุททั้งหมดดังนี้

$$SSQERR = \frac{1}{2} \sum (Target - Output)^2 \dots\dots\dots(4.20)$$

SSOERR คือ ค่า Sumsquare error ที่ได้จาก Output neurons เทียบกับ Target ใช้สำหรับบ่งชี้

ผลของการฝึกสอนว่าเพียงพอกับการใช้งาน หรือเงื่อนไขที่กำหนดหรือไม่

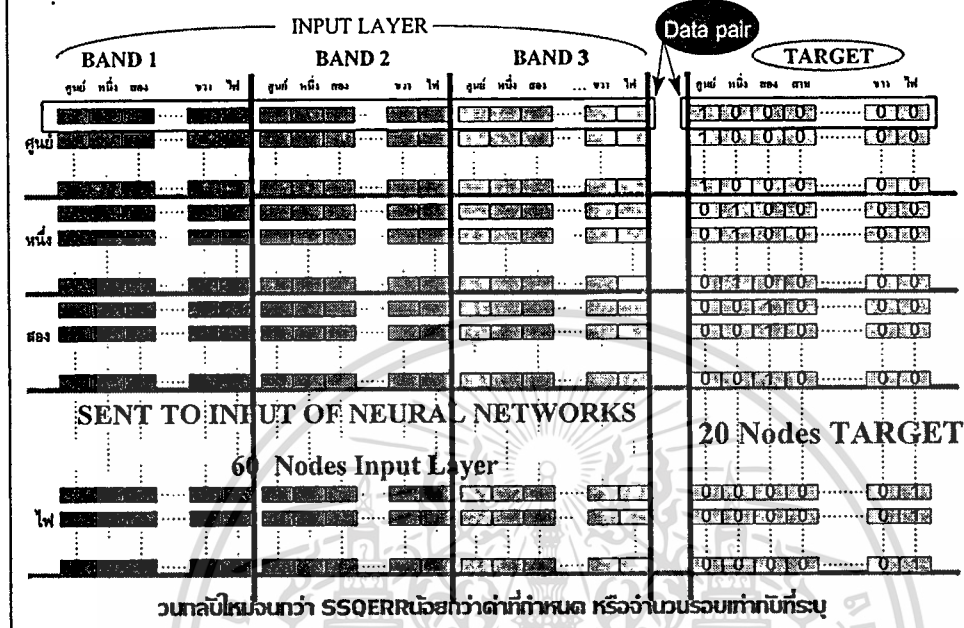
2. จำนวนรอบของการฝึกสอนดำเนินไปถึงค่าที่ตั้งไว้ ซึ่งจำนวนรอบที่เหมาะสมได้จากการทดลอง (ในการทดลองจำนวนรอบที่เหมาะสมอยู่ในช่วง 2,500-5,000 รอบ)

การป้อนข้อมูลอินพุทให้กับนิวรัลเน็ตเวิร์คในกระบวนการฝึกสอน

ข้อมูลที่ได้จากกระบวนการปรีโพรเซสซิ่ง ดังแสดงในภาพที่ 10(ซ) จะได้ผลลัพธ์ เป็นค่าความสอดคล้องกันระหว่างเสียงที่ต้องการทราบกับกลุ่มเสียงหลัก ผลลัพธ์ที่ได้ทั้ง 3 แบนด์ จะถูกนำมาป้อนให้กับอินพุทของนิวรัลเน็ตเวิร์ค ร่วมกับผลลัพธ์ที่ต้องการที่เอ้าท์พุทของนิวรัลเน็ตเวิร์ค (Target) เป็นคู่ๆ เรียกว่า Data pair กลุ่มของคุณลักษณะจำเพาะสัมพันธ์ที่ได้ จากกระบวนการปรีโพรเซสซิ่ง จะค่อยๆ ถูกป้อนควบคู่ไปกับเอ้าท์พุทที่ต้องการทีละคู่ แล้วปรับค่าซินแนปติกเวกซ์ของนิวรัลเน็ตเวิร์ค จะกระทำเช่นนี้ไปจนหมดข้อมูล เรียกว่า หนึ่งรอบการฝึกสอน เมื่อรอบดังกล่าวจะมีการตรวจสอบการหยุดการเทรนนิ่ง ตามเงื่อนไขการหยุดการฝึกสอนที่กล่าวมาข้างต้น ลักษณะของเอ้าท์พุทที่ต้องการ มีจำนวนเท่ากับเสียงที่ต้องการให้นิวรัลเน็ตเวิร์ครู้จัก และถ้าให้คุณลักษณะจำเพาะของเสียงอินพุท ตรงกับเอ้าท์พุทใด ก็จะทำให้เอ้าท์พุทนั้นมีค่าเป็น "1" ส่วนที่เหลือเป็น "0" ทั้งหมด การกำหนดลักษณะนี้เรียกว่า "One of end coding" ดังแสดงในภาพที่ 28

ภาพที่ 28

แสดงการจัดข้อมูลเพื่อฝึกสอนให้กับ BPNN แบบไม่ขึ้นกับพู่พุด



วนกลับใหม่กว่า SSQERRน้อยกว่าค่าที่กำหนด หรือจำนวนรอบเท่ากับระบุ

แสดงการจัดข้อมูลเพื่อฝึกสอนให้กับนิวรัลเน็ตเวิร์ค มีลักษณะเป็น Data pair ที่ให้เป้าหมายมีลักษณะเป็น "One of end coding"

กระบวนการทดสอบ

ในกระบวนการทดสอบข้อมูล จะมีวิธีการคำนวณเพื่อให้ได้ Output ในลักษณะเช่นเดียวกับการฝึกสอนในส่วนของ Forward pass ทุกประการ แต่แตกต่างกันตรงที่ข้อมูลที่ใช้เป็นข้อมูลทดสอบและ Output ที่ได้ คือผลลัพธ์ของการตรวจสอบซึ่งการตรวจสอบนี้จะใช้เวลาน้อยกว่ากระบวนการเทรนนิ่งมากเพราะการคำนวณเป็นเพียง Forward pass เพียงรอบเดียวโดยใช้ค่า Weight ที่ได้จากการเทรนนิ่งเป็นตัวถอดรหัสความสัมพันธ์

บทที่ 5

การดำเนินการและการทดลอง

การเลือกข้อมูลเพื่อใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลเสียงพูดคำไทย โดยเลือกเฉพาะคำที่เป็นคำโดดๆ ชนิดหนึ่ง คำหนึ่งพยางค์ ที่มีความหมายในตัวที่สอดคล้องกับการนำไปประยุกต์ใช้กับการสั่งงานหรือควบคุมใน จำนวนคำที่ไม่มากนักคือ 20 คำ

ผู้วิจัยเลือกกลุ่มข้อมูลเฉพาะเสียงผู้ชาย ในช่วงวัยหนุ่ม คือ ตั้งแต่อายุ 20 ปี ถึง 45 ปี สาเหตุ ที่เลือกกลุ่มข้อมูลเฉพาะ เนื่องจากต้องการบิบบขอบเขตของการวิเคราะห์ให้แคบลง เพื่อให้เหมาะกับขีด ความสามารถของเครื่องมือ และอุปกรณ์ที่ใช้ ซึ่งหากต้องการให้ระบบสามารถวิเคราะห์เสียงได้ทุกเพศ ทุกวัย ก็เพียง จัดกลุ่มข้อมูลเพิ่มขึ้น และในกระบวนการทดสอบให้แยกวิเคราะห์ทีละกลุ่ม

วิธีการเก็บและจัดการข้อมูลเพื่อการประมวลผล

ผู้วิจัยเก็บข้อมูลเสียง จากกลุ่มผู้พูดเสียงตัวอย่างที่เป็นชายหนุ่ม 32 คน โดยแบ่งเป็น 3 กลุ่ม ดังนี้

กลุ่ม A มีจำนวนผู้พูด 10 คน เพื่อพูดสร้างชุดข้อมูลเสียงหลักสำหรับหาค่าสัมประสิทธิ์สห สัมพันธ์ กับกลุ่ม B และ C โดยผู้พูดแต่ละคนจะพูด “ศูนย์” , “หนึ่ง” , “สอง” , “สาม” , “สี่” , “ห้า” , “หก” , “เจ็ด” , “แปด” , “เก้า” , “สิบ” , “เปิด” , “ปิด” , “หมุน” , “ยก” , “วาง” , “ช่อง” , “ซ้าย” , “ขวา” , “ไฟ” รวม 20 คำที่เป็นเสียงที่จะทดสอบเรียงตามลำดับ 3 รอบ (ผู้พูด จะพูด 3 ครั้งต่อ 1 เสียง ทดสอบ) จะได้ 60 ตัวอย่างเสียงต่อ 1 คน รวมกลุ่ม A มี 600 ตัวอย่างเสียง ข้อมูลเสียงผู้พูดกลุ่ม A นี้ จะถูกใช้เป็นฐานข้อมูลหลักสำหรับคำนวณหาสัมประสิทธิ์สหสัมพันธ์ ในกระบวนการ Preprocessing จะได้เป็นคุณลักษณะจำเพาะสัมพันธ์ ของข้อมูลเสียงทุกเสียงในกลุ่ม B และ C ข้อมูลส่วนนี้จะนำไปใช้ ในระบบรู้จำที่สร้างขึ้นจากนิวรัลเน็ตเวิร์ค

กลุ่ม B มีจำนวน 10 คน วิธีการเก็บข้อมูลใช้เทคนิคเดียวกับกลุ่ม A จะได้ 60 ตัวอย่างเสียง ต่อ 1 คน เช่นกัน รวมกลุ่ม B ได้ 600 ตัวอย่างเสียง ข้อมูลกลุ่ม B นี้จะถูกนำไปผ่านกระบวนการ Pre-

Processing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ ข้อมูลส่วนนี้จะนำไปใช้ในระบบรู้จำที่สร้างขึ้นจากนิวรัลเน็ตเวิร์ค

กลุ่ม C มีจำนวน 12 คน วิธีการเก็บข้อมูลใช้เทคนิคเดียวกับกลุ่ม A จะได้ 60 ตัวอย่างเสียง ต่อ 1 คน เช่นกัน รวมกลุ่ม C ได้ 720 ตัวอย่างเสียง ข้อมูลกลุ่ม C นี้จะถูกนำไปผ่านกระบวนการ Pre-Processing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ ข้อมูลส่วนนี้จะนำไปใช้ในระบบรู้จำที่สร้างขึ้นจากนิวรัลเน็ตเวิร์ค เช่นกัน

สาเหตุที่เลือกเก็บข้อมูลตามวิธีการที่ระบุในกลุ่ม A ก็เพื่อหลีกเลี่ยงการออกเสียง เสียงเดียวกันซ้ำต่อเนื่องกันของผู้ออกเสียง ซึ่งจะทำให้เสียงเสียงเดียวกันมีความแตกต่างกันน้อยไม่เป็นธรรมชาติ เมื่อเก็บข้อมูลครบ 32 คน จะได้ตัวอย่างเสียงทั้งสิ้น 1,920 ตัวอย่างเสียง (32 คน คนละ 3 ตัวอย่างเสียง ต่อ หนึ่งเสียงทดสอบ) โดยบันทึกลงเครื่องบันทึกเทปคาสเซตต์เดคแบบไฮ-ไฟ จากนั้นจึงนำไปจัดกลุ่มคำใหม่โดยให้เสียงที่ใช้ทดสอบ ที่เป็นคำคำเดียวกันอยู่ในกลุ่มเดียวกัน จะได้ 20 กลุ่มเสียงทดสอบ ดังนั้นในแต่ละกลุ่มเสียงทดสอบ จะมีกลุ่มของผู้พูด 3 กลุ่มย่อยคือ กลุ่ม A 10 คน 30 เสียง กลุ่ม B 10 คน 30 เสียง และกลุ่ม C 12 คน 36 เสียง รวม 96 เสียงตัวอย่างต่อ 1 กลุ่มเสียงทดสอบ ($96 \times 20 = 1,920$) การจัดกลุ่มข้อมูลแบบนี้จะทำให้ เสียงทดสอบเสียงเดียวกันที่เปล่งโดยคนคนเดียวกันทั้ง 3 ครั้ง มารวมกันเป็นหนึ่งแฟ้ม แฟ้มทุกแฟ้มจะถูกนำไปแปลงให้เป็นสัญญาณดิจิทัลด้วย A/D Converter ของการ์ดซาวด์บลาสเตอร์ 16 ซึ่งผลิตโดยบริษัท Creative Labs, Inc. กำหนดอัตราสุ่มที่ 8 KHz แบบ 8 บิต โมโน จะได้แฟ้มข้อมูลเสียงดิจิทัลที่เป็นเสียงทดสอบเดียวกันเป็น 3 ตัวอย่างเสียงต่อแฟ้ม จากนั้นจึงนำแฟ้มข้อมูลไปวิเคราะห์สัญญาณในเชิงความถี่เพื่อให้ได้เป็นแฟ้มสเปกโตรแกรม 3 มิติของความถี่-พลังงาน-เวลาซึ่งจะทำทุกแฟ้ม แฟ้มข้อมูลสเปกโตรแกรม เสียงที่ได้นี้ จะถูกนำไปผ่านกระบวนการ Pre-Processing เพื่อหาคุณลักษณะจำเพาะสัมพันธ์ และนำไปใช้ในระบบรู้จำที่สร้างขึ้นจากนิวรัลเน็ตเวิร์ค ต่อไป

ขั้นตอนการดำเนินการทดลอง

การดำเนินการทดลองของงานวิจัยนี้ มีขั้นตอนและเหตุผลดังนี้

1. กำหนดวิธีการเก็บข้อมูล และกลุ่มข้อมูลที่ใช้
2. บันทึกข้อมูลของบุคคลที่เปล่งเสียงตัวอย่างลงบนเครื่องบันทึกเสียงที่มีความเชื่อถือได้ โดยมีการจดบันทึกตำแหน่งของเสียงในเนื้อเทป จาก Tape counter เพื่ออ้างอิงกับผู้ออกเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. นำมาตัดต่อเสียงใหม่ให้เสียงคำๆเดียวกันอยู่กลุ่มเดียวกัน โดยใช้คอมพิวเตอร์แปลงข้อมูลนาลอกให้เป็นสัญญาณดิจิทัล ด้วยอัตราสุ่ม 22.050 KHz จากนั้นนำกลับไปบันทึกลงเทปอีกครั้ง

4. สุ่มข้อมูลเข้ามาใหม่ โดยใช้อัตราสุ่ม 8 KHz นำมาเก็บเป็นแฟ้มเสียงแบบฟอร์แมต VOC File มาตรฐาน โดยแต่ละแฟ้มจะบรรจุข้อมูลเสียงดิจิทัล จำนวน 1 เสียงที่แปลง 3 ครั้ง จาก 1 คน โดยจะได้แฟ้มทั้งหมด 640 แฟ้ม จากบุคคลผู้แปลงเสียง 32 คน แปลงเสียงคนละ 20 เสียงตัวอย่าง(ตั้งแต่ "หนึ่ง, สอง, สาม,... , ไฟ" ตัวอย่างละ 3 ครั้ง)

5. แบ่งกลุ่มของข้อมูลออกเป็น 3 กลุ่มคือ กลุ่มเสียงหลักเลือกจากผู้แปลงเสียงคนที่ 1ถึง10 กลุ่มเสียงอ้างอิงเลือกจากผู้แปลงเสียงคนที่ 11ถึง20 และกลุ่มเสียงทดสอบเลือกจากผู้แปลงเสียงคนที่ 21ถึง32 จากนั้นแปลงข้อมูลเสียงซึ่งเป็นข้อมูลเชิงเวลาทุกแฟ้มให้เป็นข้อมูลเชิงความถี่ ด้วยวิธีการแปลงแบบฟูเรียร์อย่างรวดเร็ว จะได้เป็นสเปกโตรแกรมของเสียง(*.FRQ) แล้วปรับแกนเวลาใหม่ให้เป็น Log scale เพื่อให้สอดคล้องกับธรรมชาติของการได้ยิน(*.OTV)

6. แยกค่าทุกค่าออกมาจากสเปกโตรแกรม นำมาสร้างเป็นแฟ้มใหม่ ขณะนี้จะได้แฟ้มข้อมูลทั้งหมด เท่ากับ $640 \times 3 = 1920$ แฟ้ม ที่มีนามสกุลเป็น *.OTV

7. ดำเนินการหาคุณลักษณะจำเพาะสัมพันธ์ของกลุ่มเสียงอ้างอิง ที่สัมพันธ์กับเสียงหลัก และหาคุณลักษณะจำเพาะสัมพันธ์ของกลุ่มเสียงทดสอบ ที่สัมพันธ์กับเสียงหลัก

8. นำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงอ้างอิง มาฝึกสอนให้กับนิวรัลเน็ตเวิร์ค โดยป้อนควบคู่ไปกับเป้าหมายที่เอาท์พุทที่ต้องการ จนเน็ตเวิร์คมีค่า SSOERR ลดลงต่ำกว่าจุดที่ยอมรับได้ หรือ จำนวนรอบครบตามที่กำหนด ในการทดลองกำหนดให้หยุดตามจำนวนรอบของการฝึกสอน โดยกำหนดไว้ที่ 4,000-6000 รอบ

9. นำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงทดสอบมาทดสอบเน็ตเวิร์คที่ผ่านการฝึกสอนแล้ว แล้วนำเอาท์พุทที่ได้หลายๆค่ามาหาอัตราการเรียนรู้ของนิวรัลเน็ตเวิร์ค

10. ทำการทดลองเพิ่มเติมในบทที่ 6 เป็นวิธีการลดขนาดของข้อมูลแบบหนึ่ง ที่พยายามปรับข้อมูลสเปกโตรแกรมของทุกเสียง ให้มีจำนวนเท่ากัน และระดับพลังงานอยู่ในระดับเดียวกันโดยใช้วิธี Max-min scale ปรับระดับพลังงานในสเปกโตรแกรมให้อยู่ในระดับเดียวกัน ลดจำนวนข้อมูลในแกนความถี่จาก 128 ลงเหลือ 16 โดยหาค่าเฉลี่ยทุกๆ 8 ข้อมูล และปรับจำนวนสเปกตรัมของทุกสเปกโตรแกรม ที่มีจำนวนแตกต่างกันไปตามระยะเวลาการออกเสียงแต่ละคน ให้มีจำนวนเท่าคือ 10 สเปกตรัม ต่อหนึ่งสเปกโตรแกรมเสมอ โดยใช้วิธีประมาณค่าแบบลากรองอินเตอร์โพลลิง โพลีโนเมียล ซึ่งจะกล่าวในบทที่ 6

วิธีการทดลอง

ผู้วิจัยแบ่งวิธีการทดลองออกเป็น 4 แบบคือ

แบบที่ 1 เพื่อหาอัตราการรู้จำของนิรวัลเน็ตเวอร์ค แบบไม่ขึ้นกับผู้พูดโดยใช้ข้อมูลทดสอบคนละกลุ่มกับชุดข้อมูลที่ใช้ฝึกสอน

แบบที่ 2 เพื่อหาอัตราการรู้จำของนิรวัลเน็ตเวอร์ค แบบขึ้นกับผู้พูดในกลุ่มเดียวกัน โดยแบ่งข้อมูลส่วนหนึ่งมาใช้ฝึกสอน และอีกส่วนหนึ่งเพื่อการทดสอบ โดยเน็ตเวอร์คจะรู้จักคุณสมบัติจำเพาะของกลุ่มเสียงที่จะทดสอบมาก่อน

แบบที่ 3 เป็นวิธีการแบบเดียวกับแบบที่ 1 แต่ลดจำนวนของเสียงทดสอบลง โดยทำการทดลองสองครั้ง คือการทดลองที่ 3-1 จะสร้างเน็ตเวอร์คให้รู้จำเสียง 11 เสียงทดสอบ คือ “ศูนย์” , “หนึ่ง” , “สอง” , “สาม” , “สี่” , “ห้า” , “หก” , “เจ็ด” , “แปด” , “เก้า” , “สิบ” และการทดลองที่ 3-2 จะสร้างเน็ตเวอร์คให้รู้จำเสียง 9 เสียงทดสอบ คือ “เปิด” , “ปิด” , “หมุน” , “ยก” , “วาง” , “ช่อง” , “ซ้าย” , “ขวา” , “ไฟ” ทั้งสองครั้งใช้ผู้พูด 12 คน พุด 3 ครั้ง ต่อ 1 เสียงทดสอบ

แบบที่ 4 เป็นวิธีการแบบเดียวกับแบบที่ 2 แต่ลดจำนวนของเสียงทดสอบลงเหลือ 11 เสียง ทดลองทำสองครั้ง การทดลองที่ 4-1 ใช้ผู้พูด 32 คน การทดลองที่ 4-2 จะลดจำนวนผู้พูดเหลือ 12 คน เพื่อเปรียบเทียบผลการรู้จำ

การทดลองและผลการทดลองแบบที่ 1

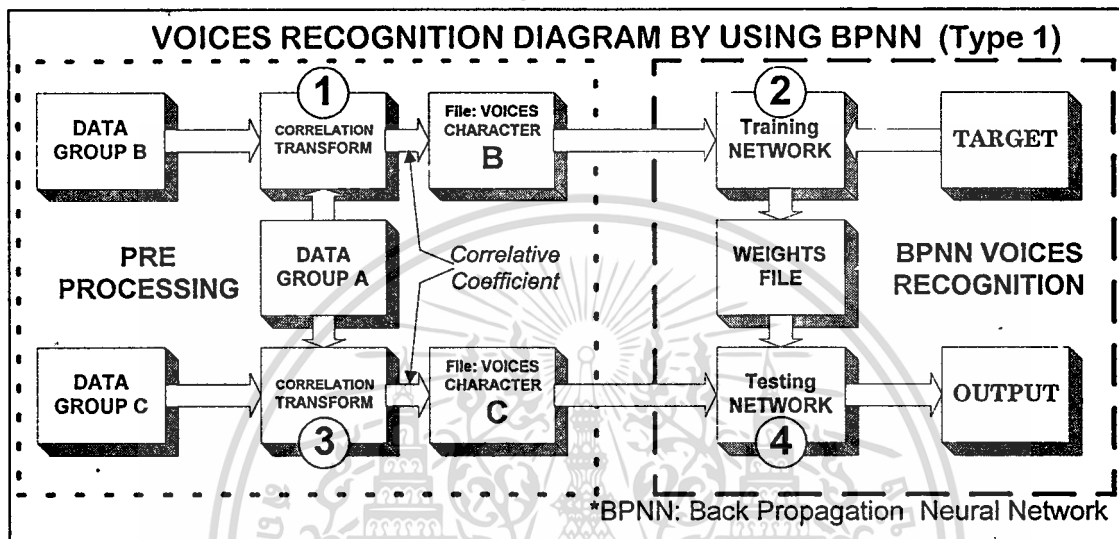
การทดลองแบบที่ 1 แบ่งเป็น 3 ขั้นตอนคือ ปริโปรเซสซิง, ฝึกสอน และทดสอบ มีรายละเอียดดังนี้

ขั้นตอนที่ 1 ปริโปรเซสซิง เป็นขั้นตอนที่ ปรับเตรียมข้อมูลดิบ ให้สอดคล้องกับวิธีการประมวลผล กำหนดให้ข้อมูลกลุ่ม B ใช้ในกระบวนการฝึกสอนและข้อมูลกลุ่ม C ใช้ในกระบวนการทดสอบ ของนิรวัลเน็ตเวอร์ค ข้อมูลกลุ่ม C นี้ในทางปฏิบัติเมื่อนำไปใช้งานจริงๆ จะมีข้อมูลของเสียงที่ต้องการวิเคราะห์ครั้งละเสียงเดียว แต่ในการทดลองใช้หลายๆ ชุดข้อมูลเสียงเพราะต้องการหาอัตราการรู้จำของเน็ตเวอร์ค ข้อมูลเสียงทดสอบทุกเสียงของกลุ่ม B และ C จะถูกนำไปหาความสัมพันธ์กับข้อมูลกลุ่ม A (ซึ่งใช้เป็นฐานข้อมูลหลัก) แบบสหสัมพันธ์ ดังนั้นจึงต้องหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง กลุ่ม B กับกลุ่ม A และ กลุ่ม C กับกลุ่ม A ผลที่ได้คือคุณลักษณะจำเพาะของเสียงแต่ละเสียงของกลุ่ม B

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และ C' เมื่อเทียบกับกลุ่ม A, ขนาดของข้อมูลจะลดลงจำนวนมากและฟอร์แมตเป็นมาตรฐานทุกเสียง จึงเพียงพอที่นำมาใช้เป็นข้อมูลอินพุตให้กับนิวรัล เน็ตเวอร์คได้ โครงสร้างของการทดลองแสดงดัง โดยแแกรมที่แสดงในภาพที่ 30

แผนภูมิที่ 30



แสดงโดยแแกรมการใช้แบบครอบพาเกชั่น นิวรัลเน็ตเวอร์ค เพื่อใช้รู้จำเสียงแบบไม่ขึ้นกับผู้พูด

พิจารณาโดยแแกรมในภาพที่ 30 ที่กระบวนการ Preprocessing จะพบว่าข้อมูลเสียงทดสอบ

กลุ่ม B แต่ละเสียง จะถูกนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์กับข้อมูลเสียงทดสอบกลุ่ม A ที่ ① โดยแบ่งข้อมูลแต่ละเสียงทดสอบทั้งสองกลุ่มเป็น 3 แบนด์ และจัดคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ที่ละแบนด์ เพื่อเพิ่มความแม่นยำ ผลลัพธ์ที่ได้เป็นคุณลักษณะจำเพาะที่คำนวณได้ในแต่ละแบนด์ของเสียงทดสอบแต่ละเสียงของกลุ่ม B เมื่อเทียบกับกลุ่ม A จะเก็บไว้ในแฟ้ม Voice Character B การดำเนินการแบบนี้เรียกว่า การหาคุณลักษณะจำเพาะสัมพันธ์ของเสียง (Voice Characteristic Correlative Relation), ข้อมูลเสียงทดสอบกลุ่ม C ก็เช่นเดียวกันจะถูกนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlative Coefficient) กับข้อมูลเสียงทดสอบกลุ่ม A ทั้ง 3 แบนด์ที่ ③ และเก็บผลไว้ในแฟ้ม Voice Character C เมื่อเสร็จสิ้นกระบวนการ Preprocessing แล้ว จึงเข้าสู่กระบวนการ ฝึกสอน ที่ ② โดยนำข้อมูลในแฟ้ม Voice Character B ที่คำนวณได้ทั้ง 3 แบนด์ เป็นอินพุตให้กับอินพุตเลเยอร์ของเน็ตเวอร์คเพื่อฝึกสอนแบบแบบครอบพาเกชั่น ที่นิวรัล เน็ตเวอร์คที่สร้างขึ้น ซึ่งมีแฟ้ม Target บรรจุข้อมูลเอาต์พุตที่ต้องการ มาใช้ในกระบวนการฝึกสอน เพื่อปรับค่าเวกซ์ของแตล่นิวรัลในเน็ตเวอร์คให้ตอบสนองข้อมูลในแฟ้มอินพุต Voice Character B สอดคล้องกับเป้าหมายที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจาก Error ที่เกิดขึ้นจากการฝึกสอนลดลงถึงค่าที่ยอมรับได้ (ในการทดลองกำหนดที่ Average Square Error ไม่เกิน 0.05 หรือ 5000 รอบการฝึกสอน หยุดเมื่ออย่างใดถึงก่อน) ค่าเวกซ์ที่ได้จะถูกบรรจุลงเพิ่มเวกซ์ ซึ่งจะนำไปใช้ในกระบวนการ Testing ที่ ④ หากพิจารณาต่อไปในกระบวนการ Testing ข้อมูลเพิ่ม Voice Character C จะใช้เป็น Input ร่วมกับเพิ่มเวกซ์ให้กับเน็ตเวิร์ค โดยเน็ตเวิร์คจะตรวจสอบข้อมูลในเพิ่ม Voice Character C แบบ Feed Forward โดยอ้างอิงกับข้อมูลในเพิ่มเวกซ์ที่ได้จากกระบวนการ ฝึกสอน ผลลัพธ์ที่ได้ จึงสอดคล้องกับข้อมูลที่ใช้ ฝึกสอน ซึ่งก็คือ คำตอบของเน็ตเวิร์คที่เป็นผลการรู้จำเสียงคำไทยแบบที่ไม่ขึ้นกับผู้พูดนั่นเอง หลังจากการฝึกสอนจนถึงค่า Average Sum Square Error น้อยกว่าค่าที่กำหนดหรือจำนวนรอบของการฝึกสอนถึงค่าที่กำหนดแล้ว ผู้วิจัยทำการทดลองทดสอบเน็ตเวิร์คด้วยข้อมูลเดิมที่เคยใช้ฝึกสอน เพื่อทดสอบประสิทธิภาพของเน็ตเวิร์ค ได้ผลลัพธ์แสดงดัง ตารางที่ 1

พิจารณาค่าในตารางที่ 1 เป็นการสนองตอบต่อข้อมูลที่ใช้เป็นแบบทดสอบของนิวรัลเน็ตเวิร์ค ที่ผ่านการฝึกสอนแล้ว 2500รอบ โดยข้อมูลที่ใช้ทดสอบก็คือข้อมูลชุดเดียวกับที่ใช้ฝึกสอนเป็นการทดสอบการทำงานและพิสูจน์ความถูกต้องของโปรแกรมนิวรัลเน็ตเวิร์คที่สร้างขึ้น ซึ่งนิวรัลเน็ตเวิร์คสามารถตอบสนองได้อย่างถูกต้อง โดยมีเอ็ทพุทอยู่ระหว่าง 0.887-0.994 มีค่าเฉลี่ยอยู่ที่ 0.961 จากเป้าหมาย 1.000 และค่า Maxmitude ของเอ็ทพุทเลเยอร์ ที่ตอบสนองสามารถแยกแยะเสียงได้ถูกต้องทั้งหมด

การทดสอบการรู้จำของเน็ตเวิร์ค ในแบบที่1 ผู้วิจัยแบ่งชุดข้อมูลเสียงทดสอบในกลุ่ม C เป็น 3 กลุ่ม ตามครั้งของการออกเสียง เช่น กลุ่มที่ 1 จะมีข้อมูลเสียงทดสอบ “ศูนย์” ครั้งแรก ของคนที่ 1 ถึงคนที่ 12 และตามด้วย “สอง” ครั้งแรก ของคนที่ 1 ถึงคนที่ 12 ไปจนถึงเสียงทดสอบ “ไฟ” ครั้งแรก ของคนที่ 1 ถึงคนที่ 12 รวม $20 \times 12 = 240$ เสียงทดสอบต่อกลุ่ม กลุ่มที่ 2 และ 3 ก็มีลักษณะเช่นเดียวกับกลุ่มที่ 1 แต่ใช้ข้อมูลของการออกเสียงครั้งที่ 2 และ 3 ตามลำดับ รวมเสียงทดสอบในกลุ่ม C $240 \times 3 = 720$ ตัวอย่างเสียง การทดสอบจะทำครั้งละกลุ่ม รวม 3 ครั้ง ได้เป็นตารางที่ 2, 3 และ 4 ตามลำดับ

ตารางที่ 1

ตารางแสดงผลการตรวจสอบนิรวัลเน็ตเวิร์ค ด้วยข้อมูลเดิมที่ใช้ฝึกสอน ค่าในตารางเป็นค่าเฉลี่ยของแต่ละเสียงทดสอบที่ได้จากคน 10 คน ในกลุ่ม A ซึ่งทดสอบหลังจากการฝึกสอนแล้ว
จำนวน 2500 รอบ

แบบทดสอบ A	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม A																		
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา
ศูนย์	0.988	0	0.022	0.005	0.001	0	0	0	0	0	0	0	0	0.105	0	0	0	0	0.001
หนึ่ง	0	0.991	0	0	0.002	0.001	0	0	0.002	0.004	0	0.002	0.001	0.006	0	0	0	0	0.007
สอง	0.005	0	0.983	0.001	0	0	0	0	0	0.001	0	0.002	0	0	0.001	0	0.006	0	0.001
สาม	0	0	0	0.993	0	0.008	0	0	0	0	0	0	0	0.006	0	0	0	0	0.003
สี่	0.001	0.006	0	0.003	0.982	0	0	0	0	0	0.013	0	0.011	0.001	0	0	0.001	0	0.002
ห้า	0	0	0	0	0	0.991	0	0	0.005	0.001	0	0.003	0	0	0.003	0.001	0	0.002	0.001
หก	0	0	0	0	0	0	0.987	0	0.001	0	0	0.002	0	0.012	0.004	0	0	0	0.005
เจ็ด	0	0.002	0	0	0.001	0	0.006	0.977	0	0	0.01	0.002	0.014	0	0	0	0	0.001	0
แปด	0	0.001	0	0.009	0	0.005	0.002	0.002	0.991	0	0	0	0	0	0	0.004	0	0.002	0.001
เก้า	0	0	0	0.003	0	0.016	0	0	0	0.995	0	0	0	0	0.002	0.007	0.001	0	0
สิบ	0	0	0	0.006	0.018	0	0	0.005	0	0	0.985	0	0	0	0	0	0	0	0
เปิด	0	0.004	0.004	0	0	0	0.005	0.004	0.001	0.001	0	0.888	0	0.001	0.003	0.005	0	0	0
ปิด	0	0.004	0	0	0.015	0	0.003	0.108	0.003	0	0.009	0	0.978	0.002	0	0	0	0	0.002
หมุน	0.008	0.011	0	0.001	0	0	0.006	0	0	0	0	0.003	0.005	0.89	0	0.001	0.001	0	0.011
ยก	0.001	0	0	0	0	0	0.007	0	0	0.001	0	0.001	0	0.001	0.984	0	0.007	0.001	0.006
วาง	0	0	0.001	0.006	0	0.004	0	0	0.001	0.003	0	0	0	0	0	0.887	0.001	0.002	0.015
ช่อง	0.006	0	0.004	0	0	0	0	0	0	0	0	0	0	0	0	0	0.994	0	0.001
ซ้าย	0.002	0	0	0.009	0	0.004	0	0	0.001	0	0	0.002	0	0	0	0	0.002	0.984	0
ขวา	0	0	0.001	0.009	0	0	0	0	0	0	0	0	0	0.011	0	0.009	0	0.004	0.975
ไฟ	0	0.002	0.003	0	0.009	0	0.001	0	0.004	0.001	0	0.001	0	0	0.004	0	0.005	0.006	0.984

พิจารณาตารางที่ 1 ค่าที่อยู่ในตารางเป็นผลการทดสอบนิรวัลเน็ตเวิร์คที่ผ่านการฝึกสอนแล้ว ด้วยข้อมูลชุดเดียวกับชุดที่ใช้ฝึกสอน ซึ่งนิรวัลเน็ตเวิร์คสามารถแยกแยะได้ถูกต้องทั้งหมด โดยสังเกตจากค่าสูงสุด ที่ปรากฏที่เอ้าท์พุททั้ง 20 ของนิรวัลเน็ตเวิร์ค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2

แสดงผลการทดสอบการรู้จำของนิรวัลเน็ตเวิร์ค ด้วยข้อมูลเสียงทดสอบกลุ่ม C เฉพาะข้อมูลเสียง
ทดสอบที่พูดครั้งแรก

แบบทดสอบ C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B พูดครั้งแรก ผู้พูด 12 คน																			% ความถูกต้อง
ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา	ไฟ	ถูกต้อง
ศูนย์	8												1			1		2		66.67
หนึ่ง		9							1		1								1	75.00
สอง			9													1		2		75.00
สาม				9			1								1			1		75.00
สี่					7					1		3							1	58.33
ห้า				3		7									2					58.33
หก							10					2								83.33
เจ็ด								10	1		1									83.33
แปด						1	1		10											83.33
เก้า				1						11										91.67
สิบ								1			6		5							50.00
เปิด							1					9			1	1				75.00
ปิด		1											10						1	83.33
หมุน													1	9	1			1		75.00
ยก												2			10					83.33
วาง				1								1				9		1		75.00
ช่อง			1														11			91.67
ซ้าย															2		10			83.33
ขวา				1														11		91.67
ไฟ							1								1				10	83.33
																			เฉลี่ย	77.08
																			สูงสุด	91.67
																			ต่ำสุด	50.00

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3

แสดงผลการทดสอบการรู้จำของนิรวัลเน็ตเวิร์ค ด้วยข้อมูลเสียงจากผู้พูด 12 คนในกลุ่ม C เฉพาะ
ข้อมูลที่พูดครั้งที่สองของแต่ละเสียงทดสอบ

แบบทดสอบ C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B พูดครั้งที่สอง ผู้พูด 12 คน																			% ความถูกต้อง
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา	ไฟ
ศูนย์	9													2	1					75.00
หนึ่ง		9				1								1		1				75.00
สอง			5						1	1							1		4	41.67
สาม	1			10			1													83.33
สี่		1			5								4	1						41.67
ห้า				4		5				1						2				41.67
หก							9					1					1		1	75.00
เจ็ด							1	9	1										1	75.00
แปด						2	2		7							1				58.33
เก้า									1	11										91.67
สิบ	1				1				1		6		3							50.00
เปิด		1					2					8				1				66.67
ปิด		1						1					10							83.33
หมุน		1												10		1				83.33
ยก					1										9		1		1	75.00
วาง							1		1							9		1		75.00
ช่อง																	11		1	91.67
ซ้าย							2											10		83.33
ขวา																1			11	91.67
ไฟ					1	1	1												9	75.00
เฉลี่ย																				71.67
สูงสุด																				91.67
ต่ำสุด																				41.67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

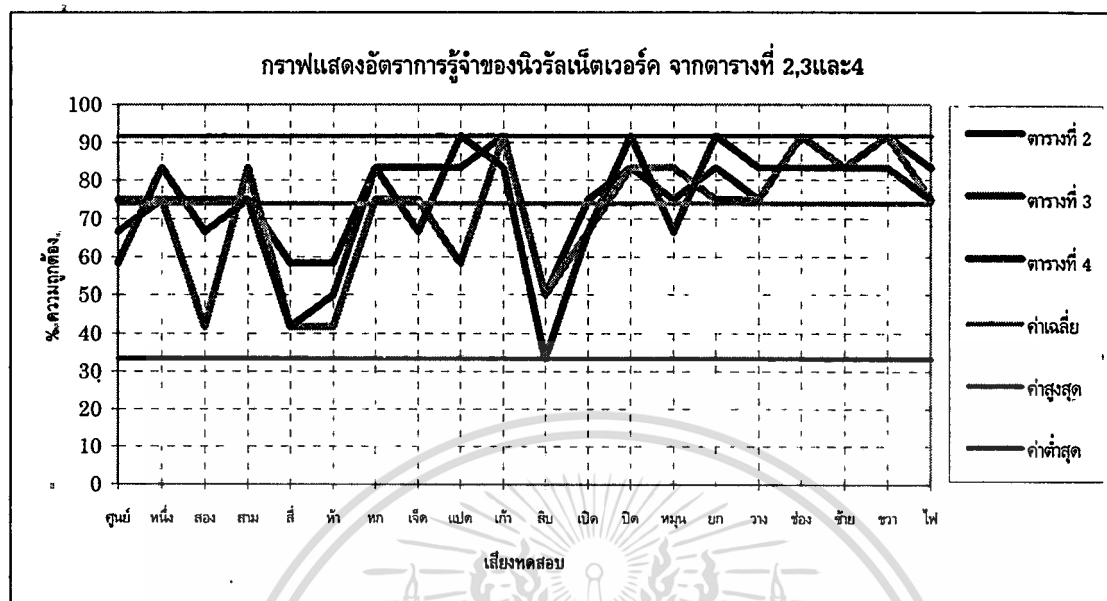
ตารางที่ 4

แสดงผลการทดสอบการรู้จำของนิรวัลเน็ตเวิร์ค ด้วยข้อมูลเสียงทดสอบจากผู้พูด 12 คน ในกลุ่ม C
เฉพาะข้อมูลเสียงทดสอบที่พูดครั้งที่สามของแต่ละเสียงทดสอบ

แบบทดสอบ B	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B พูดครั้งที่สาม ผู้พูด 12 คน																			% ความถูกต้อง	
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา	ไฟ	ถูกต้อง
ศูนย์	7													1	1	1			2		58.33
หนึ่ง		10				1													1		83.33
สอง			8														1	2	1		66.67
สาม			1	9										1					1		75.00
สี่		1			5		1				1		3	1							41.67
ห้า				2		6					3					1					50.00
หก							10		1									1			83.33
เจ็ด		1						8			1		1							1	66.67
แปด									11							1					91.67
เก้า										10							1	1			83.33
สิบ	2						2				4		4								33.33
เปิด		1					2						8		1						66.67
ปิด							1						11								91.67
หมุน		1					2							8						1	66.67
ยก							1								11						91.67
วาง																10		1	1		83.33
ช่อง			1														10			1	83.33
ซ้าย							1											10		1	83.33
ขวา				1														1	10		83.33
ไฟ							2	1												9	75.00
เฉลี่ย																					72.92
สูงสุด																					91.67
ต่ำสุด																					33.33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาพที่ 31



กราฟแสดงการเปรียบเทียบอัตราการเรียนรู้ของนิรลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด 3 ครั้ง จากตารางที่ 2,3,4 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุดกัน

จากกราฟภาพที่ 31 พบว่าเสียง “สี่” “ห้า” และ “สิบ” ระบบรู้จำมีอัตราการเรียนรู้ต่ำผิดพลาดมากกว่าเสียงอื่น ๆ หากพิจารณา ร่วมกับตารางที่ 2, 3, 4 จะพบว่า เสียงตัวเลขที่ทำให้ระบบรู้จำผิดพลาด มี สามคู่คือ “สี่” กับ “บิต”, “ห้า” กับ “สาม” และ “สิบ” กับ “บิต”

เสียง “สี่” จะผิดพลาดเป็นเสียง “สิบ” ทั้งนี้เนื่องจากเสียงทั้งคู่เป็นเสียง เอก แดก ต่างกันตรงที่ สี่ เป็นคำเป็น ส่วน สิบ เป็นคำตาย มีความผิดพลาด 10 ครั้ง จากการพูด 36 ครั้ง

เสียง “ห้า” จะผิดพลาดเป็นเสียง “สาม” ทั้งนี้เนื่องจากเสียงทั้งคู่มีลักษณะการเปลี่ยนแปลงของข้อมูลแกนความถี่ในสเปกตรัมโดยรวมใกล้เคียงกัน มีความผิดพลาด 9 ครั้ง จากการพูด 36 ครั้ง

เสียง “สิบ” จะผิดพลาดเป็นเสียง “บิต” ทั้งนี้เนื่องจากเสียงทั้งคู่มีความถี่มูลฐานจากหลอดเสียงคล้ายคลึงกันมาก หากพยายามออกเสียงทั้งสอง โดยไม่ให้มีการเปลี่ยนแปลงของอวัยวะในช่องปาก เสียงที่เกิดขึ้นจะเหมือนกันทุกประการ ซึ่งอัลกอริทึมที่ผู้วิจัยใช้ไม่สามารถแยกแยะได้ถูกต้อง มีความผิดพลาด 12 ครั้ง จากการพูด 36 ครั้ง

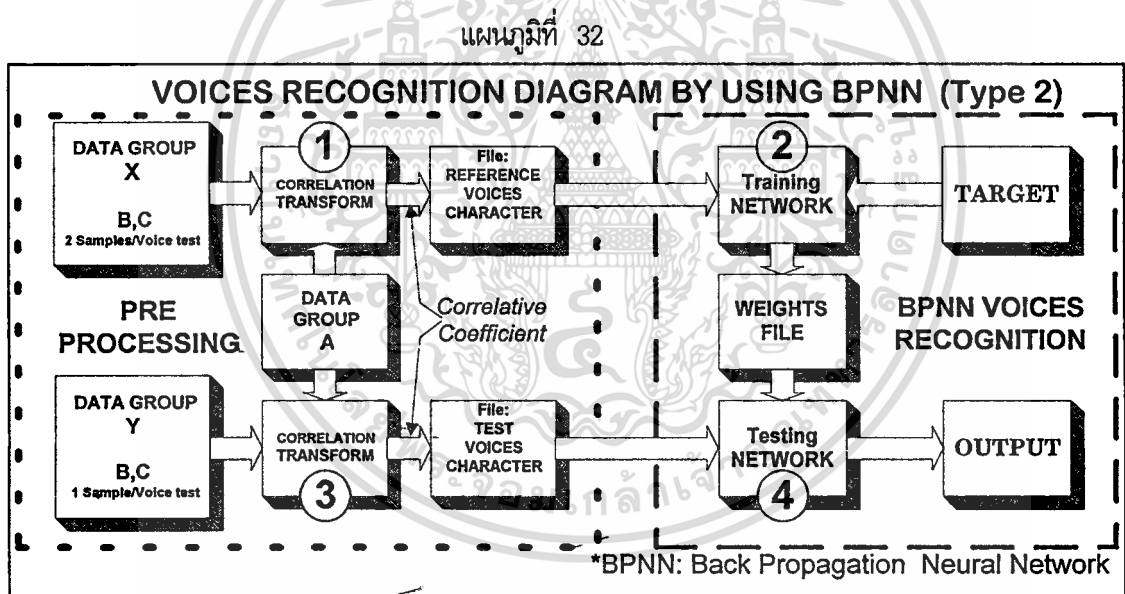
ความผิดพลาดของเสียงทั้งสามคู่นี้ มีผลทำให้อัตราการเรียนรู้โดยรวมของนิรลเน็ตเวิร์คลดลงอย่างมาก ณ จุดนี้พอสรุปได้ว่าอัตราการเรียนรู้ของนิรลเน็ตเวิร์คตามวิธีที่ทดลอง ขึ้นกับความคล้ายคลึงของเสียงคำที่ถูกเลือกมาใช้ในระบบด้วย

เอเจนซีการศึกษานานาชาติเพื่อการศึกษานานาชาติ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองและผลการทดลองแบบที่ 2

การทดลองแบบที่ 2 มีลำดับขั้นตอนเหมือนกับขั้นตอนที่ 1 แตกต่างกันตรงที่ ข้อมูล 3 กลุ่ม คือ X, และ Y ดังภาพที่ 32 ข้อมูลที่ X และ Y นี้เป็นข้อมูลที่จัดให้สอดคล้องกับแนวคิดในการทดลองแบบที่ 2 คือ ฝึกสอนให้เน็ตเวิร์คให้รู้จักข้อมูลเสียงชุดค่าส่วนหนึ่งจากกลุ่มของผู้พูดตัวอย่าง แล้วทดสอบด้วยข้อมูลเสียงค่าชุดเดิม ในส่วนที่เหลือ วิธีการเป็นดังนี้ ข้อมูลชุด B และ C ซึ่งประกอบด้วยเสียงทดสอบจำนวน 20 แบบ จากผู้พูด 10 คน (11 คนสำหรับข้อมูลชุด C) คนละ 3 ครั้ง ต่อ 1 เสียงทดสอบ ข้อมูลชุด B และ C แต่ละชุด จะถูกแบ่งเป็น 2 ส่วน แบบตามครั้งที่พูด กล่าวคือ เสียงทดสอบ (ทุกเสียงที่พูดครั้งที่ 1 และ 2 จะใช้สำหรับกระบวนการฝึกสอน ส่วนข้อมูลเสียงที่พูดครั้งที่ 3 จะนำมาใช้ในกระบวนการทดสอบ พิจารณาแผนภูมิที่ 32



แสดงไดอะแกรมการใช้แบบครอบพาเก้น นิวรัลเน็ตเวิร์ค เพื่อใช้รู้จำเสียงแบบขึ้นกับผู้พูดในกลุ่มเดียวกัน

พิจารณาไดอะแกรมในภาพที่ 32 ที่กระบวนการ **Preprocessing** จะพบว่าข้อมูลเสียงทดสอบกลุ่ม X คือข้อมูลในกลุ่ม B และ C จากผู้พูด 22 คน ที่เลือกเฉพาะข้อมูลเสียงที่เปล่งครั้งที่ 1 และ 2 แต่ละเสียง จะถูกนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์กับข้อมูลเสียงทดสอบกลุ่ม A กระบวนการหลังจากนี้ จะเหมือนการทดลองในแบบที่ 1 ในส่วนของข้อมูลทดสอบกลุ่ม Y ก็ยังคงเป็นข้อมูลในกลุ่ม B และ C จากผู้พูด 22 คน แต่เลือกเฉพาะข้อมูลเสียงที่เปล่งครั้งที่ 3 แต่ละเสียง จะถูกนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์กับข้อมูลเสียงทดสอบกลุ่ม A

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัมพันธ์กับข้อมูลเสียงทดสอบกลุ่ม A เช่นกันเพื่อสร้างคุณลักษณะจำเพาะ เก็บไว้ Test Voice Character เพื่อสำหรับการประมวลผลการทดสอบ คำตอบของเน็ตเวิร์คที่เป็นผลการรู้จำเสียงคำไทยแบบขึ้นกับผู้พูดในกลุ่มเดียวกันนั่นเอง ผลการทดลองการหาอัตราการรู้จำในแบบที่ 2 แสดงได้ดังตารางที่ 5



ตารางที่ 5

แสดงผลการรู้จำของเน็ตเวิร์คที่ ฝึกสอนด้วยข้อมูลกลุ่ม BและC 22 คน ด้วยข้อมูลเสียงที่แปลงครั้งที่ 1 และ 2 แล้วทดสอบ ด้วย ข้อมูลกลุ่ม BและC 22 คน เช่นกัน แต่เป็นข้อมูลเสียงที่แปลงครั้งที่ 3 ตามแนวคิดของการทดลองแบบที่ 2

แบบทดสอบB-C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B-C พุดครั้งที่หนึ่ง-สอง ผู้พูด 22 คน																				% ความถูกต้อง
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา	ไฟ	ถูกต้อง
ศูนย์	18														1	2		1			81.82
หนึ่ง		19	1					1						1							86.36
สอง			20						1									1			90.91
สาม			1	17					1	1	1						1				77.27
สี่		2			15			1			3		1								68.18
ห้า			1			18			1	1									1		81.82
หก							19					2	1								86.36
เจ็ด		1					1	18					2								81.82
แปด		1							21												95.45
เก้า										21					1						95.45
สิบ	2							2			16		2								72.73
เปิด		2						1				18			1						81.82
ปิด								1			1		20								90.91
หมุน	2				1		1							17	1						77.27
ยก												1			21						95.45
วาง														1		19			2		86.36
ช่อง			1														21				95.45
ซ้าย				1														19		2	86.36
ขวา			1	1												1		1	18		81.82
ไฟ		1		1	1		1							1			1			16	72.73
พุดครั้งที่ สาม																				เฉลี่ย	84.32
ผู้พูด 22 คน																				สูงสุด	95.45
oytabc.txt																				ต่ำสุด	68.18

การทดลองและผลการทดลองแบบที่ 3

การทดลองแบบที่ 3 มีวิธีการแบบเดียวกับการทดลองแบบที่ 1 แตกต่างกันตรงที่ จำนวนเสียงทดสอบที่ต้องการให้ระบบรู้จำมีจำนวนน้อยลง จากเดิมมีจำนวน 20 ตัวอย่างเสียงทดสอบ ลดจำนวนลงเหลือ 11 และ 9 เสียงทดสอบ ตามลำดับ โดยทำการทดสอบแยกกัน ได้ผลการทดสอบดังตารางที่ 6 และ 7

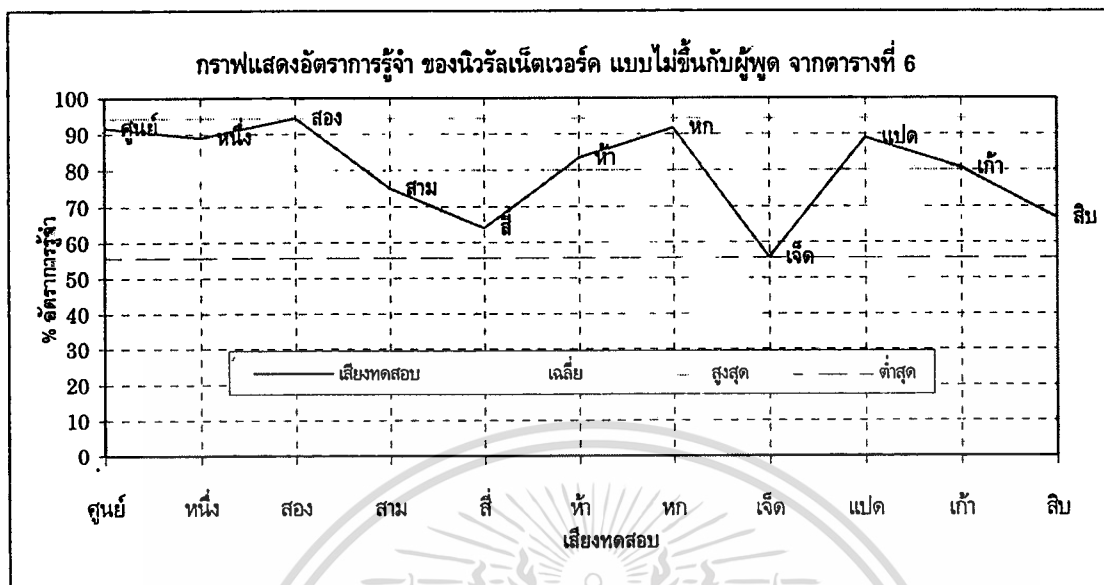
ตารางที่ 6

แสดงผลการตรวจสอบการรู้จำของนิรวัลเน็ตเวิร์คที่มีต่อ ชุดเสียงกลุ่มเสียงตัวเลขศูนย์ ถึง สิบ 11 เสียง จากผู้พูด 12 คน ออกเสียงคนละ 3 ครั้งต่อ 1 เสียงตัวเลขทดสอบ รวม 36 ครั้งโดยเน็ตเวิร์คไม่เคยเรียนรู้เสียงจาก 12 คนนี้มาก่อน

แบบทดสอบ C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B พูด 3 ครั้ง/เสียงทดสอบ ผู้พูด 12 คน											% ความถูกต้อง
	1.ศูนย์	2.หนึ่ง	3.สอง	4.สาม	5.สี่	6.ห้า	7.หก	8.เจ็ด	9.แปด	10.เก้า	11.สิบ	
1.ศูนย์	33			1			2					91.67
2.หนึ่ง		32			2	1	1					88.89
3.สอง			34	1						1		94.44
4.สาม				27	2	3	2			2		75.00
5.สี่	1	5			23			1			6	63.89
6.ห้า				3		30				3		83.33
7.หก		1			1		33		1			91.67
8.เจ็ด		13			1			20	1		1	55.56
9.แปด					1	2			32	1		88.89
10.เก้า						4	3			29		80.56
11.สิบ		9						3			24	66.67
เฉลี่ย												80.05
สูงสุด												91.67
ต่ำสุด												55.56

TABLEV11.XLS

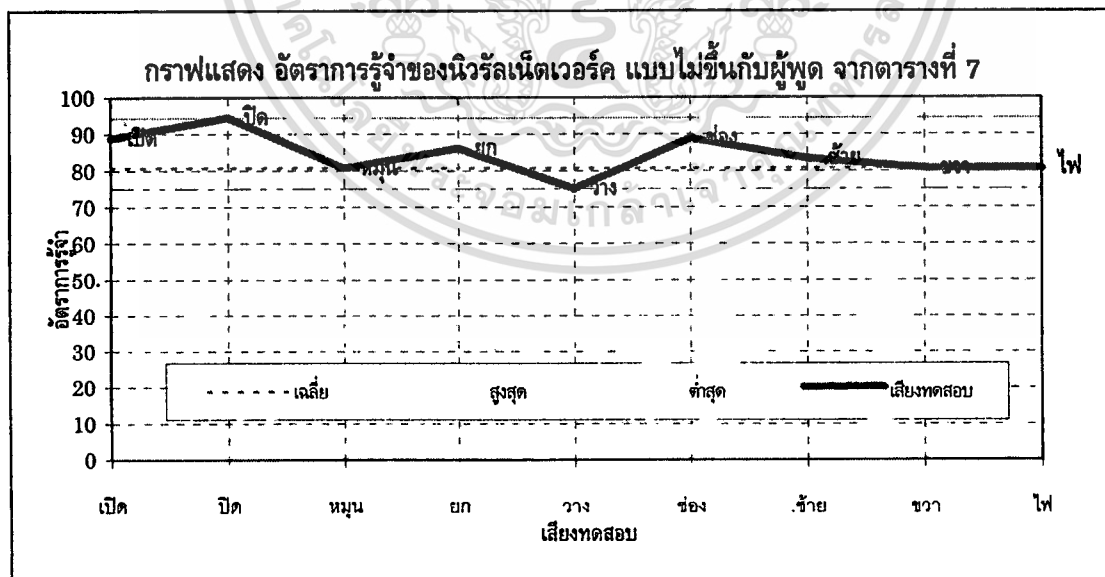
ภาพที่ 33



กราฟแสดงการเปรียบเทียบอัตราการรู้จำของนิรวัลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด จำนวนเสียงทดสอบ

11 เสียง จากผู้พูด 12 คน จากตารางที่ 6 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุด

ภาพที่ 34



กราฟแสดงการเปรียบเทียบอัตราการรู้จำของนิรวัลเน็ตเวิร์ค แบบไม่ขึ้นกับผู้พูด จำนวนเสียงทดสอบ

9 เสียง จากผู้พูด 12 คน จากตารางที่ 7 ข้อมูลที่ใช้ฝึกสอน และทดสอบคนละชุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 7

แสดงผลการตรวจสอบการรู้จำของนิรวัลเน็ตเวิร์คที่มีต่อชุดเสียงกลุ่มคำที่ใช้สั่งงาน 9 คำสั่ง จากผู้ออกเสียง 12 คน ออกเสียงคนละ 3 ครั้งต่อ 1 เสียง คำสั่ง ทดสอบโดยเน็ตเวิร์คไม่เคยเรียนรู้เสียงจาก12คนนี้มาก่อน

แบบทดสอบ C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม B พุด 3 ครั้ง/เสียงทดสอบ ผู้พุด 12 คน									% ความถูกต้อง
	1.เปิด	2.ปิด	3.หมุน	4.ยก	5.วาง	6.ช่อง	7.ซ้าย	8.ขวา	9.ไฟ	
1.เปิด	32				1				3	88.89
2.ปิด	1	34	1							94.44
3.หมุน			29	6		1				80.56
4.ยก	3	1	1	31						86.11
5.วาง	3				27			2	4	75.00
6.ช่อง			1	3		32				88.89
7.ซ้าย	1					2	30		3	83.33
8.ขวา			2		5			29		80.56
9.ไฟ	1		2			2	1	1	29	80.56
เฉลี่ย										80.56
สูงสุด										94.44
ต่ำสุด										75.00

TABLEV9.XLS

การทดลองและผลการทดลองแบบที่ 4

การทดลองแบบที่ 4 มีวิธีการแบบเดียวกับการทดลองแบบที่ 2-แตกต่างกันตรงที่ จำนวนเสียงทดสอบที่ต้องการให้ระบบรู้จำมีจำนวนน้อยลง จากเดิมมีจำนวน 20 ตัวอย่างเสียงทดสอบ ลดจำนวนลงเหลือ 11 ตัวอย่างเสียงทดสอบ (ตั้งแต่ ศูนย์ ถึง สิบล) โดยแบ่งการทดสอบออกเป็น 2 ส่วน ส่วนแรกใช้ข้อมูลจากผู้พุดทดสอบ ทั้ง 32 คน ผลการทดสอบแสดงดังตารางที่ 8 และ ส่วนที่สอง ลดจำนวนผู้พุดทดสอบลงมาเหลือ 12 คน ผลการทดสอบแสดงดังตารางที่ 9

ตารางที่ 8

แสดงผลการทดสอบนิรวัลเน็ตเวอร์ค ด้วยข้อมูลกลุ่มในเดียวกัน ในส่วนที่เหลือจากการฝึกสอน ได้จากจำนวนผู้พูด 32 คน การพูด 2 ครั้งแรก ของทุกเสียงทดสอบจะนำมาใช้ฝึกสอน การพูดครั้งที่ 3 นำมาใช้ทดสอบ

แบบทดสอบ A-B-C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม A-B-C พูดครั้งที่หนึ่ง-สอง ผู้พูด 32 คน											%ความถูกต้อง
	1.ศูนย์	2.หนึ่ง	3.สอง	4.สาม	5.สี่	6.ห้า	7.หก	8.เจ็ด	9.แปด	10.เก้า	11.สิบ	
1.ศูนย์	32											100
2.หนึ่ง		26			1	1			1		3	81.3
3.สอง	2		30									93.8
4.สาม	4			6		15	3			4		46.9
5.สี่	7	2			17						6	53.1
6.ห้า						30			1	1		93.8
7.หก	2	1					27	1	1			84.4
8.เจ็ด	4	3						19			6	59.4
9.แปด			1		1	4			25		1	78.1
10.เก้า	1					2				29		90.6
11.สิบ	5				1					1	25	78.1
พูดครั้งที่ 3											เฉลี่ย	76.1
ผู้พูด 32 คน											สูงสุด	100
tableg11.txt											ต่ำสุด	46.9

ตารางที่ 9

แสดงผลการทดสอบการรู้จำของนิวัลเน็ตเวิร์ค ด้วยข้อมูลกลุ่มในเดียวกัน ในส่วนที่เหลือจากการฝึกสอน ที่ได้จาก ผู้พูดคนที่ 21 ถึง 32 จำนวน 12 คน การพูด 2 ครั้งแรก ของทุกเสียงทดสอบจะนำมาใช้ฝึกสอน การพูดครั้งที่ 3 นำมาใช้ทดสอบ

แบบทดสอบ A-B-C	แบบอ้างอิงที่ใช้ฝึกสอน กลุ่ม A-B-C พูดครั้งที่หนึ่ง-สอง ผู้พูด 12 คน											%ความถูกต้อง
	1.ศูนย์	2.หนึ่ง	3.สอง	4.สาม	5.สี่	6.ห้า	7.หก	8.เจ็ด	9.แปด	10.เก้า	11.สิบ	
1.ศูนย์	12											100
2.หนึ่ง		10							1		1	83.33
3.สอง			11	1								91.67
4.สาม				12								100
5.สี่					12							100
6.ห้า						12						100
7.หก							11	1				91.67
8.เจ็ด								11			1	91.67
9.แปด									12			100
10.เก้า						1				11		91.67
11.สิบ	1				2			1			8	66.67
พูดครั้งที่ 3											เฉลี่ย	92.42
ผู้พูด 32 คน											สูงสุด	100
outgv12.txt											ต่ำสุด	66.67

ตารางที่ 10

ตารางสรุปผลการทดลองจากตารางที่ 2-9

การทดลองแบบที่ 1, 3-1, 3-2 เป็นการหาอัตราการเรียนรู้ของระบบแบบไม่ขึ้นกับผู้พูด

การทดลองแบบที่ 2, 4-1, 4-2 เป็นการหาอัตราการเรียนรู้ของระบบแบบขึ้นกับกลุ่มของผู้พูด

แบบ การทดสอบ	PRE PROCESSING	จำนวน ผู้พูด	กลุ่มเสียง ที่ใช้ฝึกสอน	กลุ่มเสียง ที่ใช้ทดสอบ	อัตราการเรียนรู้ (%)		
					เฉลี่ย	สูงสุด	ต่ำสุด
แบบที่ 1 20เสียงทดสอบ ศูนย์-หนึ่ง-.....-ไฟ	กลุ่ม A 10คน คนที่1-10 Independent	ฝึกสอน 10 ทดสอบ 12	B ผู้พูด คนที่ 11-20 เปล่งครั้งที่ 1-2-3	C ผู้พูด คนที่ 21-32 เปล่งครั้งที่ 1-2-3	73.89	91.67	41.66
แบบที่ 2 20เสียงทดสอบ ศูนย์-หนึ่ง-.....-ไฟ	กลุ่ม A 10คน คนที่1-10 Dependent Group	ฝึกสอน 12 ทดสอบ 12	B-C ผู้พูด คนที่ 11-32 เปล่งครั้งที่ 1-2	B-C ผู้พูด คนที่ 11-32 เปล่งครั้งที่ 3	84.32	95.45	68.18
แบบที่ 3-1 11เสียงทดสอบ ศูนย์-หนึ่ง-.....-สิบ	กลุ่ม A 10คน คนที่1-10 Independent	ฝึกสอน 10 ทดสอบ 12	B ผู้พูด คนที่ 11-20 เปล่งครั้งที่ 1-2-3	C ผู้พูด คนที่ 21-32 เปล่งครั้งที่ 1-2-3	80.05	91.67	55.56
แบบที่ 3-2 9เสียงทดสอบ เปิด-ปิด-.....-ไฟ	กลุ่ม A 10คน คนที่1-10 Independent	ฝึกสอน 10 ทดสอบ 12	B ผู้พูด คนที่ 11-20 เปล่งครั้งที่ 1-2-3	C ผู้พูด คนที่ 21-32 เปล่งครั้งที่ 1-2-3	80.56	94.44	75.00
แบบที่ 4-1 11เสียงทดสอบ ศูนย์-หนึ่ง-.....-สิบ	กลุ่ม A 10คน คนที่1-10 Dependent Group	ฝึกสอน 32 ทดสอบ 32	A-B-C ผู้พูด คนที่ 1-32 เปล่งครั้งที่ 1-2	A-B-C ผู้พูด คนที่ 1-32 เปล่งครั้งที่ 3	76.10	100.0	46.90
แบบที่ 4-2 11เสียงทดสอบ ศูนย์-หนึ่ง-.....-สิบ	กลุ่ม A 10คน คนที่1-10 Dependent Group	ฝึกสอน 12 ทดสอบ 12	C ผู้พูด คนที่ 21-32 เปล่งครั้งที่ 1-2	C ผู้พูด คนที่ 21-32 เปล่งครั้งที่ 3	92.42	100.0	66.67

บทที่ 6

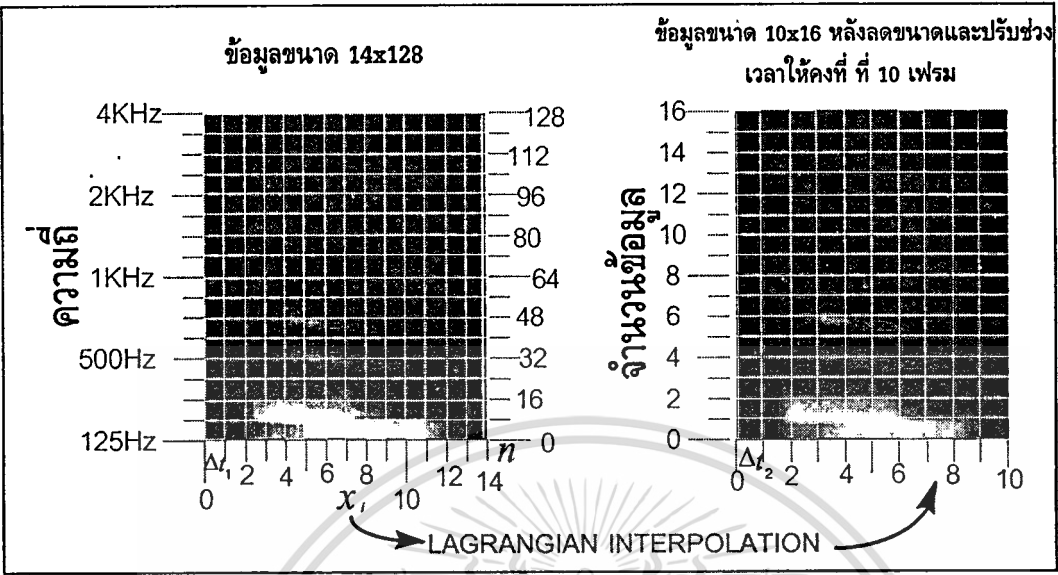
การลดขนาดของข้อมูลด้วยวิธีหาค่าเฉลี่ยร่วมกับการประมาณค่า การปรับขนาดของข้อมูลให้เหมาะสมในกระบวนการปริโปรเซสซิง

ในบทนี้ จะกล่าวถึงอีกวิธีหนึ่งในการปรับ-ลดจำนวนของข้อมูลของสเปกโตรแกรมเสียงใดๆ ที่มีจำนวนข้อมูลต่างกัน แปรปรวนไปตามลักษณะการออกเสียงของแต่ละคน โดยปรับให้มีระดับพลังงานสูงสุด-ต่ำสุดของข้อมูลที่เป็นองค์ประกอบของสเปกโตรแกรมเสียงทุกเสียงให้อยู่ในช่วงเดียวกันเสียก่อน จากนั้นจึงลดข้อมูลในแกนความถี่ ของทุกดีสครีตสเปกตรัมของสเปกโตรแกรมเสียงให้มีจำนวนน้อยลง และปรับจำนวนดีสครีตของสเปกตรัมของสเปกโตรแกรมเสียงทุกเสียงให้มีจำนวนเท่ากัน ส่วนวิธีการนำไปให้นิรวัลเน็ตเวิร์ค รู้จำนั้นจะแบ่งข้อมูลเป็น 2 ชุด ชุดหนึ่งใช้สำหรับฝึกสอน อีกชุดหนึ่งใช้สำหรับทดสอบกับนิรวัลเน็ตเวิร์คเพื่อใช้สรุปหาอัตราการรู้จำของวิธีการ รายละเอียดและวิธีการทั้งหมดที่กล่าวในบทนี้ไม่ใช่สาระสำคัญหรือวัตถุประสงค์หลักของงานวิจัยนี้ หากเป็นเพียงการนำเสนอวิธีการหนึ่งที่น่าสนใจ ในกระบวนการปริโปรเซสซิง มีวัตถุประสงค์เพื่อหาอัตราการรู้จำ นำผลลัพธ์มาเปรียบเทียบกับวิธีการแบบหาคุณลักษณะจำเพาะสัมพันธ์

วิธีการปรับระดับพลังงานสูงสุด-ต่ำสุดของข้อมูลสเปกโตรแกรมเสียงให้อยู่ในช่วงเดียวกัน

วิธีที่เสนอนี้ เป็นวิธีลดข้อมูลโดยลดรายละเอียดทางแกนความถี่ของสเปกโตรแกรมลง จาก 128 ข้อมูล เป็น 16 ข้อมูล โดยหาค่าเฉลี่ยทุก 8 ค่า ใช้วิธีนี้เนื่องจากข้อมูลของแกนความถี่มีขนาดคงที่ ส่วนทางด้านแกนเวลา จะปรับจำนวนสเปกตรัม (องค์ประกอบของสเปกโตรแกรม) ของสเปกโตรแกรมที่แตกต่างกันให้มีจำนวนเท่ากัน ในการทดลองจะปรับให้ทุกสเปกโตรแกรมประกอบด้วย 10 สเปกตรัม(เฟรม) โดยใช้วิธีการลากรองอินเตอร์โพลลิงโพลีโนเมียล (Lagrange interpolating polynomials) จะได้ข้อมูลขนาด $16 \times 10 = 160$ ไนตอินพุท นำไปเป็นอินพุทให้กับนิรวัล เน็ตเวิร์ค โดยตรง ดังภาพที่ 35

ภาพที่ 35



แสดงการลดขนาดของข้อมูลแกนความถี่จาก 128 ลงเหลือ 16 ข้อมูล และปรับจำนวนเฟรมของสเปกโตรแกรมซึ่งแต่ละเสียงมีจำนวนไม่เท่ากัน ให้มีจำนวนเท่ากันที่ 10 เฟรมต่อสเปกตรัมเสมอ

จากภาพที่ 35 วิธีการลดขนาดข้อมูลของสเปกโตรแกรม ใช้วิธีการประมาณค่าแบบ Lagrangian Interpolation (ดร.จเร สุวพันธ์ปัญญา 2531:96-101) โดยใช้สมการ

$$P(x) = \sum_{j=1}^n y_j \prod_{\substack{j=1 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \dots\dots\dots(6.1)$$

- โดย $P(x)$ คือ อินเตอร์โพลิติงโพลีโนเมียล ของจุดที่ต้องการประมาณค่า
- x คือ ช่วงเวลาแต่ละช่วงที่เป็นจุดที่ต้องการทราบค่าที่ได้จากการประมาณค่า
- n คือ จำนวนเฟรมของสเปกโตรแกรม ที่ต้องการปรับขนาดข้อมูล มีจำนวนเท่ากับ Δt_1 (จำนวน Sampling ของตัวอย่างข้อมูล)
- x_i คือ ลำดับข้อมูลที่ Sampling โดยมีค่า $x_i = 1, 2, 3, \dots, n$
- y_i คือ ระดับพลังงานของข้อมูลที่ Sampling x_i ในแถวข้อมูลที่กำลังสนใจ

จะดำเนินการประมาณค่า $P(x)$ ในแต่ละแถวข้อมูลของสเปกโตรแกรมที่ประมาณค่าขึ้นมาใหม่ โดยจะเพิ่ม x ไปทีละ Δt_2 10 ครั้งจนครบในแต่ละแถวข้อมูลโดย

$$x = x + \Delta t_2 \dots\dots\dots(6.2)$$

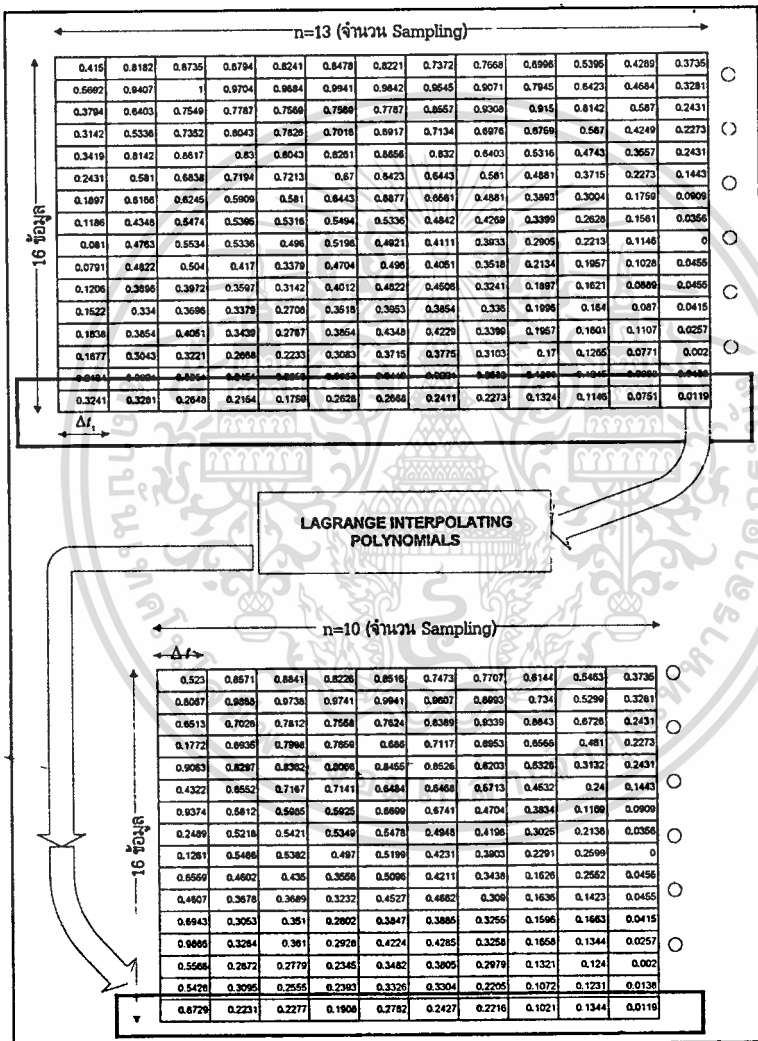
โดย

$$\Delta t_2 = \frac{n}{10} \Delta t_1 \dots\dots\dots(6.3)$$

Δt_1 คือ ช่วงเวลาการ Sampling ของข้อมูลของสเปกโตรแกรม

Δt_2 คือ ช่วงห่างเวลาที่ต้องการทราบค่าการประมาณ โดยจะแบ่งเวลาของสเปกโตรแกรมใดๆ ออกเป็น 10 ส่วนเสมอ

ภาพที่ 36



ภาพแสดงการปรับข้อมูลในแถวล่างสุด ซึ่งมีข้อมูลจำนวน 13 ข้อมูลต่อแถว นำมาประมาณค่าใหม่ ให้มีข้อมูลลดลงเหลือ 10 ข้อมูลต่อแถว ด้วยการประมาณค่าแบบลากรองอินเตอร์โพลลิง โพลีโนเมียล

เมื่อประมาณค่าข้อมูลเสร็จสิ้นทุกแถว จะได้สเปกโตรแกรมที่มีขนาด 10×16 ข้อมูล ก็พร้อมจะนำไปป้อนให้ อินพุตโนด ขนาด 10×16 ของนิวรัลเน็ตเวิร์ค เพื่อทำตามกระบวนการฝึกสอนและทดสอบโดยมีวิธีการดังนี้

1. ฝึกสอนด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน แปลงเสียง 20 เสียง เสียงละ 3 ครั้ง จำนวน 1,200 สเปกโตรแกรม โดยมีรายละเอียดโมดูลของนิวรัลเน็ตเวิร์คดังนี้

INPUT NEURONS = 160

HIDDEN NEURONS = 40

OUTPUT NEURONS = 20

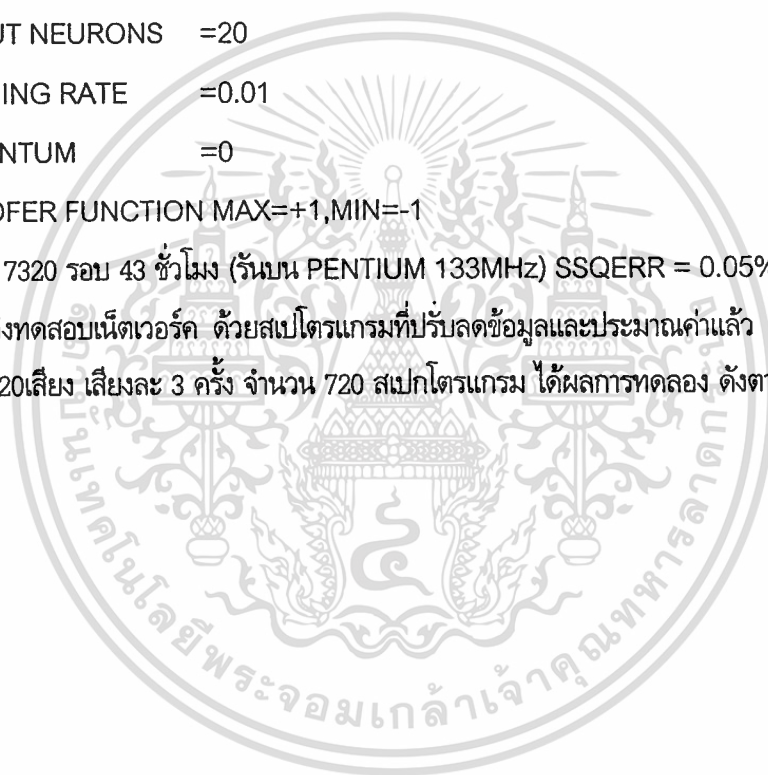
LEARNING RATE = 0.01

MOMENTUM = 0

TRANSFER FUNCTION MAX=+1, MIN=-1

เทรนนิ่ง 7320 รอบ 43 ชั่วโมง (รันบน PENTIUM 133MHz) SSQERR = 0.05%

จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12 คน แปลงเสียง 20 เสียง เสียงละ 3 ครั้ง จำนวน 720 สเปกโตรแกรม ได้ผลการทดลอง ดังตารางที่ 11



ตารางที่ 11

แสดง อัตราการรับรู้โดยใช้วิธีการลดขนาดของสเปกโตรแกรม และปรับให้ทุกสเปกโตรแกรมให้มีจำนวน
สเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 20 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 20 คน และทดสอบ
ด้วยเสียงทดสอบ จำนวน 20 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12 คน

แบบ	แบบอ้างอิง																			%ความ			
ทดสอบ	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	เปิด	ปิด	หมุน	ยก	วาง	ช่อง	ซ้าย	ขวา	ไฟ	รวม	ถูกต้อง	
ศูนย์	29		1				2					1	1						2		36	80.56	
หนึ่ง	2	26			1	2		1			1		1	1						1	36	72.22	
สอง			29							3								1		2	1	36	80.56
สาม			1	29									1			1	1		3		36	80.56	
สี่					29						1		5		1						36	80.56	
ห้า						26	1		4				3				1		1		36	72.22	
หก							30		1				5								36	83.33	
เจ็ด								34	1				1								36	94.44	
แปด						3			31				2								36	86.11	
เก้า			1						1	33			1								36	91.67	
สิบ								4				26	6								36	72.22	
เปิด						2					1		30			1	1		1		36	83.33	
ปิด	2	2			5	1	1	2					23								36	63.89	
หมุน	1						2						4	29							36	80.56	
ยก	1						1								34						36	94.44	
วาง				1									1				28		2	3	1	36	77.78
ช่อง			3										1					30	1		1	36	83.33
ซ้าย			1			1										1			30		3	36	83.33
ขวา												1								34	1	36	94.44
ไฟ	1								1				1					1	1		31	36	86.11
																					เฉลี่ย	82.08	
																					สูงสุด	94.44	
																					ต่ำสุด	63.89	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ฝึกสอนด้วยสเปโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน เปล่งเสียง 11 เสียง เสียงละ 3 ครั้ง จำนวน 660 สเปกโตรแกรม โดยมีรายละเอียดโมดูลของนิวรัลเน็ตเวิร์ค ดังนี้

INPUT NEURONS =160

HIDDEN NEURONS =40

OUTPUT NEURONS =11

LEARNING RATE =0.01

MOMENTUM =0

TRANSFER FUNCTION MAX=+1,MIN=-1

เทรนนิ่ง 6300 รอบ 20 ชั่วโมง (รันบน PENTIUM 133MHz) SSOERR = 0.04% จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12 คน เปล่งเสียง 11 เสียง เสียงละ 3 ครั้ง จำนวน 396 สเปกโตรแกรม ได้ผลดัง ตารางที่ 12

3. ฝึกสอนด้วยสเปโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน เปล่งเสียง 9 เสียง เสียงละ 3 ครั้ง จำนวน 540 สเปกโตรแกรม โดยมีรายละเอียดโมดูลของนิวรัลเน็ตเวิร์ค ดังนี้

INPUT NEURONS =160

HIDDEN NEURONS =40

OUTPUT NEURONS =9

LEARNING RATE =0.01

MOMENTUM =0

TRANSFER FUNCTION MAX=+1,MIN=-1

เทรนนิ่ง 6530 รอบ 21:35 ชั่วโมง (รันบน PENTIUM 133MHz) SSOERR = 8.06×10^{-5} จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12 คน เปล่งเสียง 9 เสียง เสียงละ 3 ครั้ง จำนวน 324 สเปกโตรแกรม ได้ผลดัง ตารางที่ 13

ตารางที่12

แสดง อัตราการรู้จำโดยใช้วิธีการลดขนาดของสเปกโตรแกรม และรับให้ทุกสเปกโตรแกรมให้มี
จำนวนสเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 11 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง20คน และ
ทดสอบด้วยเสียงทดสอบ จำนวน 11 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12คน

แบบทดสอบ	แบบอ้างอิง												%ความถูกต้อง
	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า	สิบ	รวม	
ศูนย์	31	1			2	2						36	86.11
หนึ่ง		30			3		1			1	1	36	83.33
สอง		1	32		1					2		36	88.89
สาม				33	1					1	1	36	89.19
สี่		1			35							36	97.22
ห้า						32			2	1	1	36	88.89
หก							36					36	100
เจ็ด		1						33	1		1	36	91.67
แปด									35	1		36	97.22
เก้า			1			1			1	33		36	91.67
สิบ			3		2		3				28	36	77.78
เฉลี่ย													90.18
สูงสุด													100
ต่ำสุด													77.78

ตารางที่13

แสดง อัตราการรู้จำโดยใช้วิธีการลดขนาดของสเปกโตรแกรม และปรับให้ทุกสเปกโตรแกรมให้มี
จำนวนสเปกตรัมเท่ากัน ของเสียงอ้างอิง จำนวน 9 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง20คน และ
ทดสอบด้วยเสียงทดสอบ จำนวน 9 เสียง เสียงละ 3 ครั้ง จากผู้ออกเสียง 12คน

แบบ ทดสอบ	แบบอ้างอิง										%ความ ถูกต้อง
	เปิด	ปิด	ยก	หมุน	วาง	ช่อง	ซ้าย	ขวา	ไฟ	รวม	
เปิด	33						2	1		36	89.19
ปิด	1	35								36	97.22
ยก	1		35							36	97.22
หมุน				36						36	100
วาง		1			31		2	2		36	86.11
ช่อง			1			34			1	36	94.44
ซ้าย	1				2	1	29		3	36	80.56
ขวา		1						33	2	36	91.67
ไฟ						1	1		34	36	94.44
เฉลี่ย											92.32
สูงสุด											100
ต่ำสุด											80.56

บทที่ 7

สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการทดสอบ

งานวิจัยนี้ ยังคงยึดวิธีการรู้จำเสียง ด้วยวิธีจัดจำและทดสอบแพตเทิร์นของเสียงแต่ละเสียงที่สนใจจะให้ระบบทำการรู้จำ กระบวนการ Pre-Processing จะลดขนาดและหาคุณสมบัติจำเพาะสัมพันธ์ของเสียงนั้นจากกลุ่มเสียงอ้างอิง โดยวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ แล้วฝึกสอนให้กับนิวรัลเน็ตเวิร์ค ด้วยอัลกอริทึมแบบครอบงำกัน เพื่อปรับค่าซินแนปติกส์เวกท์ภายใน Hidden layer ให้มีค่าเหมาะสม สอดคล้องกับอินพุต และเป้าหมาย เมื่อสิ้นสุดการฝึกสอน จะนำข้อมูลทดสอบมาหาคุณสมบัติจำเพาะสัมพันธ์ของเสียงแล้วป้อนเป็นอินพุตให้กับเน็ตเวิร์ค เน็ตเวิร์คก็จะสามารถให้เอาท์พุตได้สอดคล้อง กับเป้าหมายและตัวอย่างเสียงที่เคยฝึกสอนมา จากผลการทดสอบหาอัตราการรู้จำของระบบรู้จำที่สร้างขึ้น สรุปผลได้ดังนี้

การรู้จำคำไทยโดยไม่ขึ้นต่อผู้พูด

ข้อมูลเสียงผู้พูดกลุ่ม A จำนวน 10 คน จะถูกใช้เป็นฐานข้อมูลหลัก สำหรับคำนวณหาสัมประสิทธิ์สหสัมพันธ์กับข้อมูลเสียงในกลุ่ม B และ C ในกระบวนการ Preprocessing จะได้เป็นคุณลักษณะจำเพาะสัมพันธ์ ของเสียงในกลุ่ม B และ C เมื่อเทียบกับข้อมูลเสียงผู้พูดกลุ่ม A ข้อมูลส่วนนี้จะนำไปใช้ในระบบรู้จำที่สร้างขึ้นโดยใช้นิวรัลเน็ตเวิร์ค พอสรุปผลการทดลองได้ดังนี้

1. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า”, “สิบ”, “เปิด”, “ปิด”, “หมุน”, “ยก”, “วาง”, “ช่อง”, “ซ้าย”, “ขวา” และ “ไฟ” ที่ผ่านกระบวนการ Preprocessing แล้วจำนวน 20 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม B ผู้พูด 10 คน และทดสอบด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม C ผู้พูด 12 คน ได้อัตราการรู้จำร้อยละ 73.89

2. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า” และ “สิบ” ที่ผ่านกระบวนการ Preprocessing แล้วจำนวน 11 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม B ผู้พูด 10 คน และทดสอบด้วย ข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม C ผู้พูด 12 คน ได้อัตราการรู้จำร้อยละ 80.05

3. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “เปิด”, “ปิด”, “หมุน”, “ยก”, “วาง”, “ช่อง”, “ซ้าย”, “ขวา” และ “ไฟ” ผ่านกระบวนการ Preprocessing แล้วจำนวน 9 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม B ผู้พูด 10 คน และทดสอบด้วย ข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม C ผู้พูด 12 คน ได้อัตราการรู้จำร้อยละ 80.50

การรู้จำคำไทยโดยขึ้นต่อกลุ่มของผู้พูด

1. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า”, “สิบ”, “เปิด”, “ปิด”, “หมุน”, “ยก”, “วาง”, “ช่อง”, “ซ้าย”, “ขวา” และ “ไฟ” ที่ผ่านกระบวนการ Preprocessing แล้ว จำนวน 20 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม B และ C ที่พูดครั้งที่ 1 และ 2 ของผู้พูด 22 คน และทดสอบด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ พูดครั้งที่ 3 ของเสียงกลุ่มเดิม ได้อัตราการรู้จำร้อยละ 84.32

2. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า” และ “สิบ” ที่ผ่านกระบวนการ Preprocessing แล้วจำนวน 11 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม A-B และ C ที่พูดครั้งที่ 1 และ 2 ของผู้พูด 32 คน และทดสอบด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ พูดครั้งที่ 3 ของเสียงกลุ่มเดิม ได้อัตราการรู้จำร้อยละ 76.10

3. การรู้จำคุณลักษณะจำเพาะสัมพันธ์ ของเสียงคำ “ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า” และ “สิบ” ที่ผ่านกระบวนการ Preprocessing แล้วจำนวน 11 เสียงทดสอบ โดยฝึกสอนด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ ของเสียงกลุ่ม B และ C ที่พูดครั้งที่ 1 และ 2 ของผู้พูด 12 คน และทดสอบด้วยข้อมูลคุณลักษณะจำเพาะสัมพันธ์ พูดครั้งที่ 3 ของเสียงกลุ่มเดิม ได้อัตราการรู้จำร้อยละ 92.42

หากต้องการเปรียบเทียบรายละเอียดเพิ่มเติม ให้ลองพิจารณาตารางที่ 10

ผลการทดลองที่ดำเนินการในบทที่ 6

การทดลองในบทที่ 6 เป็นการทดลองลดข้อมูลและเตรียมข้อมูลที่จะป้อนให้นิวรัลเน็ตเวิร์ค โดย หาค่าเฉลี่ยของระดับพลังงานในแกนความถี่ 125 Hz-4KHz จาก 128 ข้อมูล ลดลงเหลือ 16 ข้อมูล โดยการหาค่าเฉลี่ยทุกๆ 8 ข้อมูล และปรับจำนวนสเปกตรัมในสเปกโตรแกรมที่แปรปรวนตามระยะเวลาการเปล่งเสียง ให้มีจำนวนเท่ากับด้วยวิธีประมาณค่าแบบลากรอง อินเตอร์โพลชิ้นโพลีโนเมียล ปรับให้ทุกสเปกโตรแกรมมีจำนวนสเปกตรัมเท่ากับ 10 สเปกตรัม ได้เป็นข้อมูลสเปกโตรแกรมละ 16x10 ข้อมูล แล้วนำไปเป็นอินพุตให้กับนิวรัลเน็ตเวิร์คโดยตรง การทดลองทำเพียงหาอัตราการรู้จำคำไทย โดยไม่ขึ้นต่อผู้พูด เพียงอย่างเดียว ส่วนการหาอัตราการรู้จำคำไทยโดยขึ้นต่อกลุ่มของผู้พูดนั้น ไม่ได้ทำการทดลองเนื่องจากการทดลองบนนี้เป็นเพียงการทดลองเพิ่มเติม และวัตถุประสงค์ของการทดลองเป็นเพียงการหาความเป็นไปได้ของวิธีการ ไม่ใช่วัตถุประสงค์หลักของงานวิจัยนี้ ผลการทดลองแสดงดังต่อไปนี้.

1. ฝึกสอนด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน เปล่งเสียง 20 เสียง(“ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า”, “สิบ”, “เปิด”, “ปิด”, “หมุน”, “ยก”, “วาง”, “ช่อง”, “ซ้าย”, “ขวา” และ “ไฟ”) เสียงละ 3 ครั้ง จำนวน 1,200 สเปกโตรแกรม จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12คน เปล่งเสียง 20 เสียง เสียงละ 3 ครั้ง จำนวน 720 สเปกโตรแกรม อัตราการรู้จำเฉลี่ยร้อยละ 82.08

2. ฝึกสอนด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน เปล่งเสียง 11เสียง(“ศูนย์”, “หนึ่ง”, “สอง”, “สาม”, “สี่”, “ห้า”, “หก”, “เจ็ด”, “แปด”, “เก้า” และ “สิบ”) เสียงละ 3 ครั้ง จำนวน 660 สเปกโตรแกรม จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12คน เปล่งเสียง 11เสียง เสียงละ 3 ครั้ง จำนวน 396 สเปกโตรแกรม อัตราการรู้จำเฉลี่ยร้อยละ 90.18

3. ฝึกสอนด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 20 คน เปล่งเสียง 9 เสียง(“เปิด”, “ปิด”, “หมุน”, “ยก”, “วาง”, “ช่อง”, “ซ้าย”, “ขวา” และ “ไฟ”) เสียงละ 3 ครั้ง จำนวน 540 สเปกโตรแกรม จากนั้นจึงทดสอบเน็ตเวิร์ค ด้วยสเปกโตรแกรมที่ปรับลดข้อมูลและประมาณค่าแล้ว ของผู้พูด 12คน เปล่งเสียง 9 เสียง เสียงละ 3 ครั้ง จำนวน 324 สเปกโตรแกรม อัตราการรู้จำเฉลี่ยร้อยละ 92.32

จึงพอสรุปได้ว่าวิธีการทดลองแบบนี้ เป็นวิธีการที่เป็นไปได้และอัตราการรู้จำสูงกว่าแบบที่ใช้วิธีการหาค่าคุณลักษณะจำเพาะสัมพันธ์ในกระบวนการปริโปรเซสซึ่ง ซึ่งใช้เวลาดำเนินการนานกว่าและอัตราการรู้จำต่ำกว่า

สรุปผลการวิจัย

ผลสรุปการวิจัยที่ได้จากการดำเนินการวิจัยและทดลองพอสรุปลได้ดังนี้

1. อัตราการรู้จำของระบบรู้จำ จะแปรผกผันกับ จำนวนผู้พูด และ จำนวนเสียง

ทดสอบ

2. จากผลการทดลองพบว่าอัลกอริทึมที่ใช้จะเกิดความผิดพลาดในการรู้จำสูง หากแบบเสียงที่ต้องการให้ระบบรู้จำรู้จัก มีลักษณะพ้องเสียงกัน หรือเสียงวรรณยุกต์เดียวกัน

3. อัตราการรู้จำขึ้นกับคุณสมบัติและลักษณะสมบัติทางฟิสิกส์ของเสียงที่ต้องการให้ระบบรู้จำ ถ้าหากแบบเสียงที่ต้องการให้ระบบรู้จำรู้จักมีไม่มากนัก ระบบรู้จำแบบที่สร้างขึ้นจะมีอัตราการรู้จำสูง ดังผลการทดลองที่ได้จากการทดลองแบบที่ 3-2 และ 3-1 (ตารางที่ 10)

4. การหาคุณลักษณะจำเพาะสัมพันธ์ ระหว่างกลุ่มเสียงนั้นสามารถลดขนาดของข้อมูลลงได้ก็จริง แต่หมายถึงรายละเอียดความสัมพันธ์ และคุณลักษณะอื่น ๆ จะสูญเสียไปด้วยดังนั้น การสร้างระบบรู้จำที่ต้องการอัตราการรู้จำสูง ๆ จำเป็นต้องมีการหาคุณลักษณะในแบบอื่น ๆ หลาย ๆ แบบควบคู่กันไป เช่นเดียวกับมนุษย์ที่ใช้วิธีการหลายแบบ เพื่อแปลความหมายของการสื่อสาร เช่น เสียง, น้ำเสียง, กริยาท่าทาง, สีหน้า, อารมณ์, ความต่อเนื่องของเรื่องราวที่กำลังสื่อสาร เพื่อตรวจย้ำความถูกต้อง ซึ่งก็ยังมีย่อยครั้งที่การสื่อสารล้มเหลว

5. รายละเอียดของข้อมูล และจำนวนวิธีการที่ใช้ทดสอบมีผลต่ออัตราการรู้จำเป็นอย่างมาก แต่ก็หมายถึงเวลาที่ใช้ประมวลผลจะต้องเพิ่มขึ้นตามด้วย

6. วิธีการตัดคำที่ใช้ กำหนดค่าเทรโดโลด์ ที่ 20% ของค่า Magnitude ใน 1 ชุดเสียง กรณีนี้ เสียงเสียดแทรกซึ่งปรากฏในช่วงต้นของการออกเสียง มี แอมพลิจูดต่ำและมีช่วงเวลาไม่ยาวนาน ทำให้ช่วงต้นสเปกโตรแกรมซึ่งเป็นช่วงเสียงเสียดแทรกซึ่งมีระดับพลังงานต่ำถูกตัดออกไปไม่ถูกนำมาพิจารณาในระบบ วิธีการตัดคำที่มีประสิทธิภาพจะช่วยให้อัตราการเรียนรู้ดีขึ้น

7. เวลาส่วนใหญ่ที่ใช้ในกระบวนการทดสอบเสียง สูญเสียไปในกระบวนการ Pre-Processing ซึ่งขึ้นกับจำนวนเสียงที่ใช้เป็นฐานในกลุ่ม A ดังตัวอย่างที่ทดลอง เสียงที่ต้องการทดสอบเพียง 1 เสียง จะถูกนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์ กับทุกเสียงในกลุ่มฐานข้อมูลหลัก A ทั้ง 600 ตัวอย่างเสียง ทั้งสามแบนด์ แล้วนำไปจัดเรียง เพื่อบันทึกคะแนนความถี่ ได้เป็นคุณลักษณะจำเพาะสัมพันธ์ของเสียงนั้น แล้วจึงนำไปทดสอบกับ นิวรัลเน็ตเวิร์ค ซึ่งใช้สัดส่วนเวลาไม่ต่ำกว่า 45:1 ตามลำดับ วิธีการที่สามารถลดเวลาในช่วง Pre-Processing ลงได้คือการ Simulating การหาค่าสัมประสิทธิ์สหสัมพันธ์ ด้วยนิวรัลเน็ตเวิร์ค ซึ่งจะทำให้ สัดส่วนของ Pre-Processing ต่อ Testing

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใกล้เคียง 1:1 แต่ถึงอย่างไรก็แล้วแต่ การหาค่าเพื่อใช้ฝึกสอนให้กับระบบยังต้องใช้ค่าคุณลักษณะจำเพาะสัมพันธ์ ที่ทำด้วยวิธีสัมพันธ์สหสัมพันธ์ แบบเดิมก่อน ซึ่งทำเพียงครั้งเดียวในขั้นตอนการฝึกสอน

8. หัวข้อการแก้ไขปัญหาการเปรียบเทียบเสียงที่มีความดัง, ระยะเวลาการเปล่งเสียง และความถี่ แตกต่างกันในบทที่ 3 ในส่วนของการแก้ไขระยะเวลาการเปล่งเสียง สามารถใช้วิธีการแบบอื่นที่มีประสิทธิภาพกว่าแทนได้เช่น วิธีปรับจำนวนเฟรมของสเปกโตรแกรมของทุกเสียงให้มีจำนวนเท่ากันทุกเฟรม โดยใช้วิธีการประมาณค่าระหว่างจุดข้อมูลด้วย Lagrang Interpolating Polynomials (ซึ่งกล่าวในบทที่ 6) แล้วคำนวณหาข้อมูลระหว่างข้อมูลแรกกับข้อมูลสุดท้าย 10 ค่าเท่าๆกัน จะได้เป็นสเปกโตรแกรมเสียงใหม่ที่มีจำนวนเฟรมเท่ากัน แล้วจึงเปรียบเทียบความคล้ายหรือความสัมพันธ์ด้วยสมการสหสัมพันธ์ต่อไป

9. อาจกล่าวได้ว่าการวิจัยที่ดำเนินงานมา วิเคราะห์ในแกนความฮาร์โมนิคโดยวิเคราะห์ที่ตีสคริตสเปกตรัมในสเปกโตรแกรม หากสามารถวิเคราะห์ลักษณะแอมพลิจูดของเสียงในแกนเวลา ด้วยหลักการเดียวกันกับงานวิจัยนี้ และนำมาวิเคราะห์ร่วมกันด้วยนิวรอลเน็ตเวิร์คอีกครึ่งหนึ่งก็เป็นแนวทางหนึ่งที่จะพัฒนาต่อไป แต่วิธีนี้ก็ยังไม่เหมาะกับแพตเทอร์นจำนวนมาก ๆ

10. โดยภาพรวมแล้วผลการวิจัยสอดคล้องกับงานวิจัยที่ทำมาก่อนในลักษณะเดียวกัน อัตราการรู้จำใกล้เคียงกัน โดยเทียบการทดลองที่ 3-1 กับ การทดลองการรู้จำเสียงคำไทย 0-9 แบบไม่ขึ้นต่อผู้พูด (ระพีพัฒน์ เพ็ญศิริ 2538: 44-45) อยู่ในระดับ 80.50% กับ 79.25% ตามลำดับ ในกรณีนี้สรุป ได้ว่าการรู้จำเสียงคำโดดใดๆ แบบไม่ขึ้นกับผู้พูดโดยใช้วิธีการเทียบพารามิเตอร์กับแบบอ้างอิง จะมีอัตราการรู้จำใกล้เคียงกัน และ อัตราการรู้จำยังขึ้นตรงกับความแตกต่างคุณสมบัติประจำตัวของเสียงที่นำมาให้ระบบรู้จำอีกด้วย

หัวใจของปัญหาของงานวิจัยการรู้จำเสียงแบบที่ดำเนินการ มีอยู่สามประการ คือ

ประการที่หนึ่ง ระดับพลังงานของเสียงเสียงหนึ่ง ที่เปล่งออกจากบุคคลใด ๆ มีความแตกต่างกัน ผู้วิจัยแก้ไขโดยใช้วิธีสหสัมพันธ์ เป็นวิธีการหาความสัมพันธ์ของการเปลี่ยนแปลงข้อมูลสองชุด โดยไม่ขึ้นกับแอมพลิจูดข้อมูล(หาลักษณะความสอดคล้องของการเปลี่ยนแปลงของแอมพลิจูดเสียงพูด มิได้คำนึงถึงขนาดของแอมพลิจูดข้อมูล) ซึ่งคุณสมบัติสอดคล้องกับความต้องการพอดี

ประการที่สอง ระยะเวลาของการเปล่งเสียงเสียงหนึ่ง ที่เปล่งออกจากบุคคลใด ๆ มีความแตกต่างกัน ผู้วิจัยทดลองแก้ปัญหาโดย วิเคราะห์เฉพาะช่วงระยะเวลาที่เปล่งเสียง ของเสียงคู่ทดสอบที่มีระยะเวลาเปล่งเสียงที่สั้นกว่า

ประการที่สาม ความถี่ และการเปลี่ยนแปลงของความถี่เสียงเสียงหนึ่ง ที่เปล่งออกจากบุคคลใด ๆ มีความแตกต่างกัน ผู้วิจัยแก้ปัญหานี้โดยใช้เสียงตัวอย่างจำนวนมากมาเป็นแบบอ้างอิงในไลบรารี แล้วนำเสียงที่ต้องการทดสอบไปหาความสัมพันธ์ แล้วพัฒนามาใช้นิวรัลเน็ตเวิร์คเพื่อลดเวลา เพิ่มความสะดวกและความแม่นยำ

หากพิจารณาปัญหาประการที่สาม ถ้าสามารถหาคุณลักษณะจำเพาะสัมพันธ์ (ลักษณะการเปลี่ยนแปลงของความถี่เสียงพูด ที่ไม่ขึ้นกับความถี่) ของการเปลี่ยนแปลงของความถี่ได้ ลักษณะการวิเคราะห์จะไม่ ผูกติดกับจำนวนแบบอ้างอิง ความแม่นยำ, ความเร็ว และความถูกต้องจะตามมา อาจดัดแปลงจากวิธีการแยกพยางค์โดยอาศัยความแตกต่างของสัญญาณในทางความถี่ของช่วงเวลาที่ติดกัน (ธันวา ศรีประโม่ง 2537:26-38) วิธีการนี้ยังสามารถนำไปใช้พัฒนา ปัญหาประการที่สองได้อีกด้วย

ข้อเสนอแนะ

ปัจจุบันมีเทคนิคที่สามารถดัดแปลงมาใช้กับงานวิจัยมากมาย ในการลดขนาดของข้อมูล อาจใช้วิธีอื่นเช่น ใช้เทคนิค Vector Quantization(VQ) หรือทดลองแบบอื่น ๆ เพื่อหาความเหมาะสม หรืออาจพิจารณา นิวรัลเน็ตเวิร์ค อัลกอริทึมอื่น ๆ ที่เหมาะสมกับการใช้งาน ลดขนาดข้อมูล ในการใช้นิวรัล เน็ตเวิร์ค รู้จำเสียงยังมีอัลกอริทึมอื่นที่เหมาะสมกับการรู้จำเสียงพูดที่น่าทดลองเช่น Spatio Temporal Neural Networks (STNN) และสามารถหาข้อมูลความรู้และการวิจัยนิวรัลเน็ตเวิร์คที่ทันสมัยได้จาก อินเทอร์เน็ต http://ftp.sas.com/pub/neural/FAQ2.htm#A_bias หรืออาจใช้วิธีการอื่นนอกเหนือจากที่กล่าวเช่นลดราบละเอียดของสเปกโตรแกรมทางแกนความถี่ลงและปรับจำนวนเฟรมของสเปกโตรแกรมของทุกเสียงให้มีจำนวนเท่ากันทุกเฟรมด้วยวิธีการเช่นเดียวกับที่กล่าวในข้อ 8 แล้วนำไปฝึกสอนให้กับนิวรัลเน็ตเวิร์คโดยตรง ซึ่งวิธีการนี้จะรวดเร็วกว่ามากเพราะไม่ต้องเสียเวลาหาคุณลักษณะจำเพาะสัมพันธ์ อัตราการเรียนรู้ดีขึ้น ผู้วิจัยได้ทำการทดลองเบื้องต้นแล้วยืนยันว่าดีกว่าจริงดังหลักฐานผลการทดลองเพิ่มเติมแสดงในภาคผนวก ง. สาเหตุที่ไม่นำมารวมกันในเนื้อหาของวิทยานิพนธ์ เพราะเป็นการทดลองต่างวิธีกัน นอกจากนี้อาจใช้วิธีการฝึกสอนและทดสอบแบบมัลติเน็ตเวิร์ค โดยฝึกสอนแต่ละเน็ตเวิร์คให้รู้จำรายละเอียดแตกต่างกัน เช่น เสียงวรรณยุกต์, คำเป็น-คำตาย แล้วนำเอาที่พุดแต่ละเน็ตเวิร์ค ไปโปรเซสด้วยนิวรัลเน็ตเวิร์คอีกครั้งหนึ่ง

บรรณานุกรม

เสรี ปานทาง และชม กัมปาน. การรู้จำเสียงพูดคำไทยเฉพาะบุคคลด้วยนิวรัลเน็ตเวิร์ค.

การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 18. ชลบุรี: มหาวิทยาลัยมหานคร, 982-987, 2538.
ระพีพัฒน์ เพ็ญศิริ และสมชาย จิตะพันธ์กุล. การรู้จำเสียงตัวเลขภาษาไทยโดยไม่ขึ้นต่อผู้พูดโดยการ
ใช้ไดนามิกไทม์วาร์ปิง., การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 18 . ชลบุรี: มหาวิทยาลัย
มหานคร, 977-981, 2538.

ธันวา ศรีประโม่ง และชม กัมปาน. การวิเคราะห์เสียงภาษาไทยในแกนความถี่ฮาร์โมนิค.

การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 15 , 2535.

ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปิง.

ปริญญาณิพนธ์ วิศวกรรมศาสตรมหาบัณฑิต.กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์
มหาวิทยาลัย. 2538.

ธันวา ศรีประโม่ง. การวิเคราะห์เสียงพูดภาษาไทยในแกนความถี่ฮาร์โมนิค. ปริญญาณิพนธ์
วิศวกรรมศาสตรมหาบัณฑิต.กรุงเทพฯ: บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณ
ทหาร ลาดกระบัง. 2538.

จเร สุวรรณปัญญา. การคำนวณเชิงตัวเลขด้วย BASIC. กรุงเทพฯ: คณะวิศวกรรมศาสตร์ สถาบัน
เทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง. 2531.

กาญจนา นาคสกุล. ระบบเสียงภาษาไทย. กรุงเทพฯ: คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
2520.

ประคอง กรรณสูต. สถิติศาสตร์ประยุกต์สำหรับครู. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์
มหาวิทยาลัย. 2513.

ประสิทธิ์ ภาพยักลอน. การศึกษาภาษาไทยตามแนวภาษาศาสตร์. กรุงเทพฯ: โรงพิมพ์ไทยวัฒนาพานิช,
2516.

จิตติธ มาโควิทซ์. "ทอร์คกิง ทู แมชีนส์" แปลจาก วารสาร ไบท์ สหรัฐอเมริกา โดย กิตติรัตน์ พรหม
รัตน์ ในชื่อ. "ใช้เสียงสั่งคอมพิวเตอร์". วารสาร ไบท์ ไทยแลนด์. ฉบับที่ 22 (กุมภาพันธ์ 2539):
39-44.

สุเทพ จีระกุล, ผู้แปล " "หู" ตัวกรองความถี่ที่ธรรมชาติให้มา" วารสาร เคมิคอนดักเตอร์อิเล็กทรอนิกส์
ฉบับที่ 88 (กันยายน-ตุลาคม 2531): 196-205.

เศรษฐพล ลั่นปราชญา และ กิตติ ไพฑูรย์วัฒนกิจ. "การใช้นิวรอลเน็ตเวิร์กเพื่อจดจำแพตเทิร์น"
วารสาร สำนักงานคณะกรรมการวิจัยแห่งชาติ, ปีที่ 25 ฉบับที่ 1 (2536): 45-67.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม (ต่อ)

- Schmandt, Christopher. Voice Communication with Computers,
New York: Van Nostrand, 1994.
- Freeman, James A and David M. Skapura., Neural Network Algorithms, Applications,
and Programming Techniques. New York: Addison-Wesley, 1991.
- Wasserman, Philip D. Neural Computing Theory and Practice.
New York: Van Nostrand, 1989.
- Dayhoff, Judith E. Neural Network Architectures : an Introduction.
New York: Van Nostrand, 1990.
- Hanly, Jeri R., Elliot B. Koffman and Joan C. Horvath. C program design for engineers.
New York: Addison Wesley, 1995.
- Mammone, Richard J. Artificial Neural Networks for Speech and Vision.
Chapter 17-18. London: Chapman&Hall., 1993.
- D. Brook and R. j. Wynne. Signal Processing Principles and Applications. Edward
Arnold. 1989.
- Sejnowski, Terrence J and Charls R. Rosenberg. "Parallel network that learns to
pronounce English text". Journal of Complex System. 1:145-168. 1987.
- DynaMind Developer User's Guide. Version 4.0.1. Neuro DynamX, Inc. 1992.
- Turbo C User's guide and Reference guide. Version 2.0, Boland Internation, Inc.
- "EAR" Multimedia Encyclopedia. Grolier Electronic Publishing, Inc. 1995.



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
การแบ่งกลุ่มแฟ้มข้อมูล

หลังจากบันทึกเสียงและจัดเสียงพูดคำเดียวกันของผู้พูดหนึ่งคนให้ในแฟ้มเดียวกันแล้ว ในหนึ่งแฟ้มเสียง จะประกอบด้วยข้อมูลของเสียงคำ ใดๆเดียวกัน 3 ครั้ง คำที่ใช้เป็นเสียงทดสอบ จำนวน 20 คำ ผู้พูดทั้งหมด 32 คน รวมเป็นเสียงที่ใช้ในการทดลองทั้งสิ้น $3 \times 20 \times 32 = 1,920$ คำ ผู้วิจัยมีวิธีตั้งชื่อแฟ้มตามรูปแบบดังนี้

1	2	3	4	5	6	7
---	---	---	---	---	---	---

อักขรตัวแรก หมายถึงตัวอักษรที่ใช้แทนลำดับกระบวนการที่ผู้ทดลองตั้งขึ้นSเป็นตัวย่อมาจากSound

อักขรตัวที่ 2 หมายถึง เสียงทดสอบ ที่ใช้ในการทดลอง มี 20 แบบเสียงทดสอบ จะใช้อักขรและตัวเลขแทนเสียงทดสอบมีรายละเอียดดังนี้

เสียง	ศูนย์	หนึ่ง	สอง	สาม	สี่	ห้า	หก	เจ็ด	แปด	เก้า
อักขระ	0	1	2	3	4	5	6	7	8	9

เสียง	สิบ	เปิด	ปิด	ยก	หมุน	วาง	ช่อง	ซ้าย	ขวา	ไฟ
อักขระ	A	B	C	D	E	F	G	H	I	J

อักขรตัวที่ 3-4 หมายถึง ลำดับที่ของผู้พูดเป็นคนไหน ในการทดลองใช้ตัวอย่างเสียงจากผู้พูด 32 คน คือคนที่ 01, 02, 03,.....32

อักขรตัวที่ 5-6-7 เป็นส่วนขยายของแฟ้ม มีรายละเอียดดังนี้

.VOC คือแฟ้มเสียง ที่ได้จากการบันทึกเสียงโดยผ่านทางช่อง AUX ของ Sound Card จากเทปคาสเซตที่บันทึกเสียงทดสอบ ที่ตัดต่อจัดเรียงแล้ว โดยเล่นผ่านคาสเซตดีเคแบบไฮฟิดลิตี

.FRQ คือแฟ้มสเปกโตรแกรมที่ได้จากการนำแฟ้ม .VOC มาผ่านการแปลงฟูเรียร์อย่างรวดเร็ว (ธันวา ศรีประโม่ง 2537:14-17)มีลักษณะเป็นไบนารีไฟล์

.OTV คือแฟ้มที่นำข้อมูลสเปกโตรแกรมในแฟ้ม .FRQ มาจัดสเกลใหม่ให้อยู่ในสเกล Logarithm เพื่อให้สอดคล้องกับระบบความเป็นจริงในการรับฟังเสียงของมนุษย์ มีลักษณะเป็นไบนารีไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

.TXT คือแฟ้มที่ใช้เป็นอินพุทและเอาต์พุท ของระบบรู้จำที่ใช้ในการวิจัย เนื่องจากมีความสะดวกในการทดสอบข้อมูลและผลลัพธ์ที่ได้จากการทดลอง

ตัวอย่างเช่น แฟ้ม SA24.OTV หมายถึง แฟ้มข้อมูลสเปกโตรแกรมเสียง "สิบ" ของผู้พูดคนที่ 24 มีสเปกโตรแกรมของเสียง "สิบ" ที่พูด 3 ครั้งอยู่ภายใน และในกระบวนการทดสอบจะมีการแบ่งเสียง จากแฟ้มที่ขึ้นต้นด้วย S ออกเป็น 3 แฟ้ม โดยตั้งชื่อขึ้นต้นด้วย Lxxx.OTV, Mxxx.OTV, Rxxx.OTV เพื่อความสะดวกในการดำเนินการทดลอง ดังแสดงในภาพที่ 38



การจัดกลุ่มแฟ้มข้อมูลเพื่อการวิจัย

ในการดำเนินงาน แบ่งข้อมูลสเปกโตรแกรมเสียงที่ใช้เป็น 3 กลุ่มคือ กลุ่ม A, B และ C
กลุ่ม A ใช้เป็นกลุ่มอ้างอิงเพื่อใช้หา คุณลักษณะจำเพาะสัมพันธ์ ของเสียง

กลุ่ม B-C นำไปหาคุณลักษณะจำเพาะสัมพันธ์ โดยใช้ กลุ่ม A เป็นกลุ่มอ้างอิง ผลที่ได้นำไป
ฝึกสอนและทดสอบด้วยนิรวัลเน็ตเวิร์ค มีรายละเอียดดังภาพที่ 37

ภาพที่ 37

S001.OTV	S101.OTV	S201.OTV		SI01.OTV	SJ01.OTV
S002.OTV	S102.OTV	S202.OTV		SI02.OTV	SJ02.OTV
⋮	⋮	⋮	GROUP A	⋮	⋮
S009.OTV	S109.OTV	S209.OTV		SI09.OTV	SJ09.OTV
S010.OTV	S110.OTV	S210.OTV		SI10.OTV	SJ10.OTV

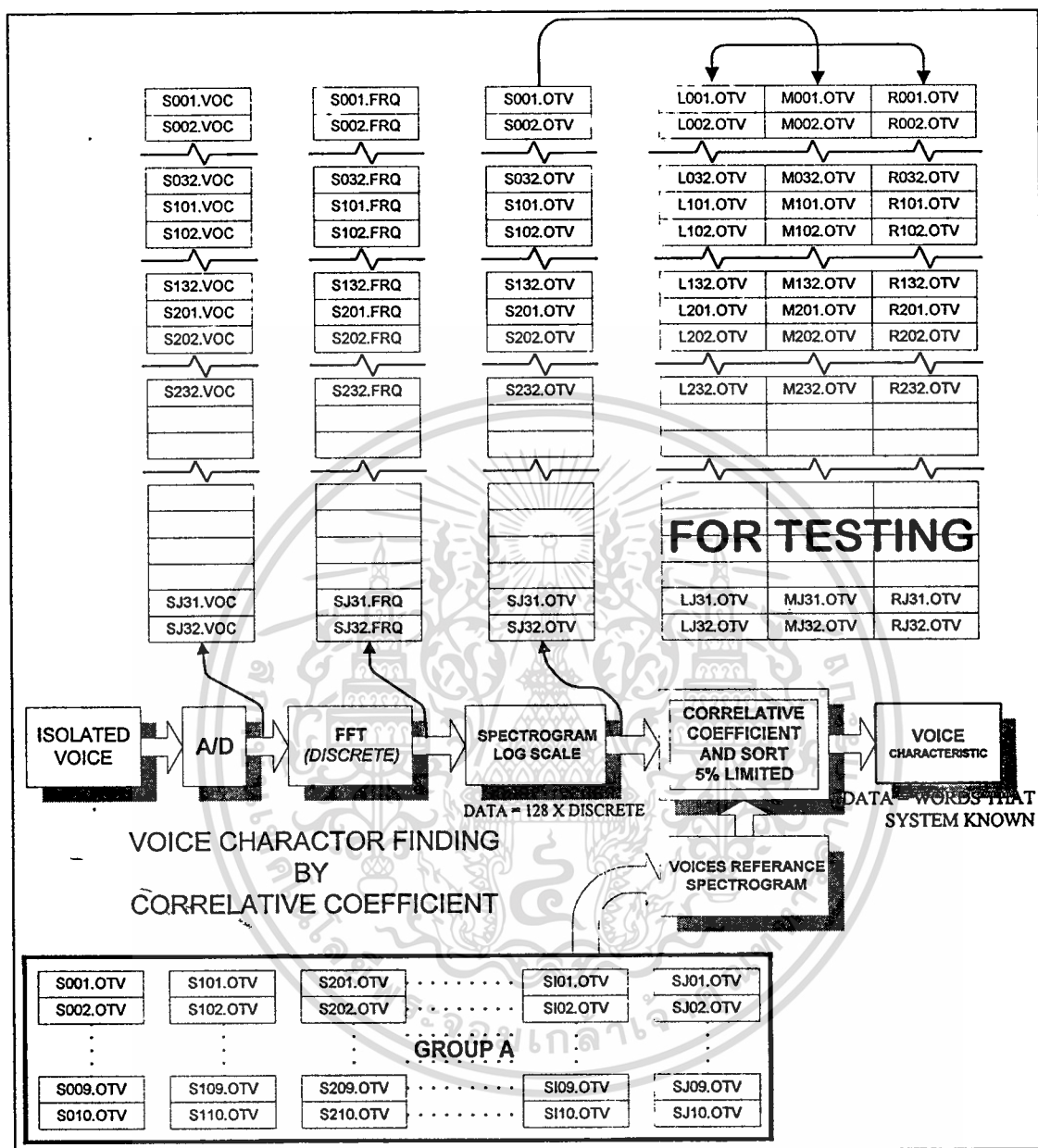
S011.OTV	S111.OTV	S211.OTV		SI11.OTV	SJ11.OTV
S012.OTV	S112.OTV	S212.OTV		SI12.OTV	SJ12.OTV
⋮	⋮	⋮	GROUP B	⋮	⋮
S019.OTV	S119.OTV	S219.OTV		SI19.OTV	SJ19.OTV
S020.OTV	S120.OTV	S220.OTV		SI20.OTV	SJ20.OTV

S021.OTV	S121.OTV	S221.OTV		SI21.OTV	SJ21.OTV
S022.OTV	S122.OTV	S222.OTV		SI22.OTV	SJ22.OTV
⋮	⋮	⋮	GROUP C	⋮	⋮
S031.OTV	S131.OTV	S231.OTV		SI31.OTV	SJ31.OTV
S032.OTV	S132.OTV	S232.OTV		SI32.OTV	SJ32.OTV

แสดงชื่อแฟ้มที่จัดแบ่งในกลุ่ม A, B และ C

ในการทดสอบ จำเป็นต้องตัดแบ่งข้อมูลสเปกโตรแกรมแต่ละเสียงจากแฟ้ม Sxxx.OTV ออก
เป็น 3 แฟ้ม คือ Lxxx.OTV(เสียงที่พูดครั้งที่ 1), Mxxx.OTV(เสียงที่พูดครั้งที่ 2), และ Rxxx.OTV
(เสียงที่พูดครั้งที่ 3) หนึ่งแฟ้มต่อหนึ่งเสียง เพื่อความสะดวกในการทดลองแบบที่ 2, 4-1 และ 4-2 ใน
บทที่ 5

ภาพที่ 38



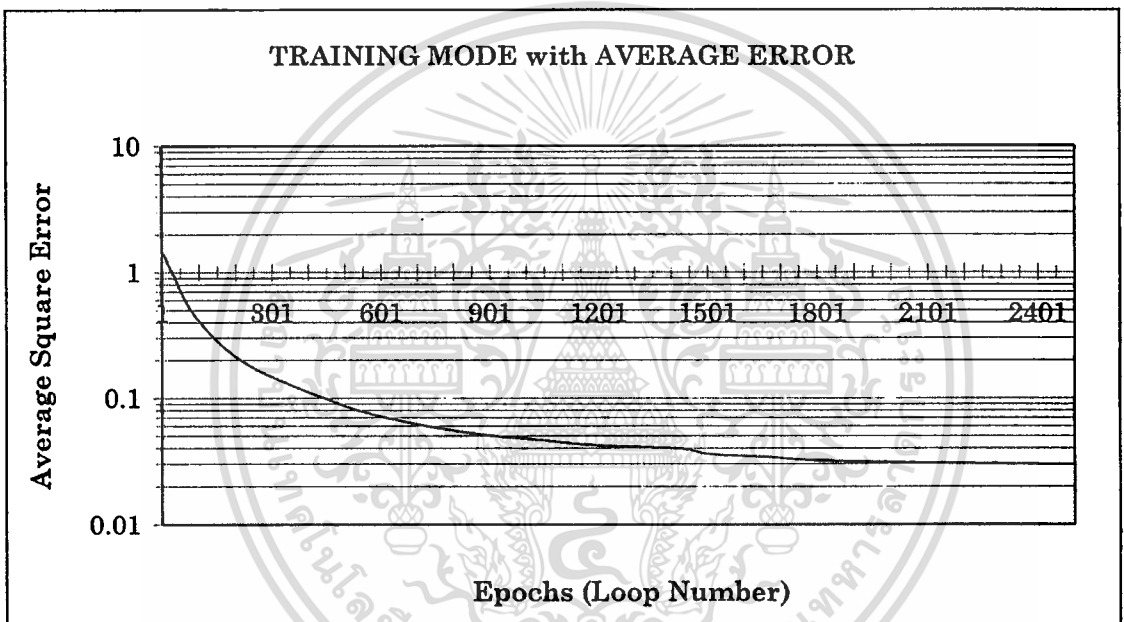
แสดงแฟ้มที่ใช้ และเพิ่มผลลัพธ์ที่ได้แต่ละขั้นตอน ในส่วนของ PRE-PROCESSING

ภาคผนวก ข

กราฟแสดงค่า SUM SQUARE ERROR ที่ได้จากการฝึกสอน

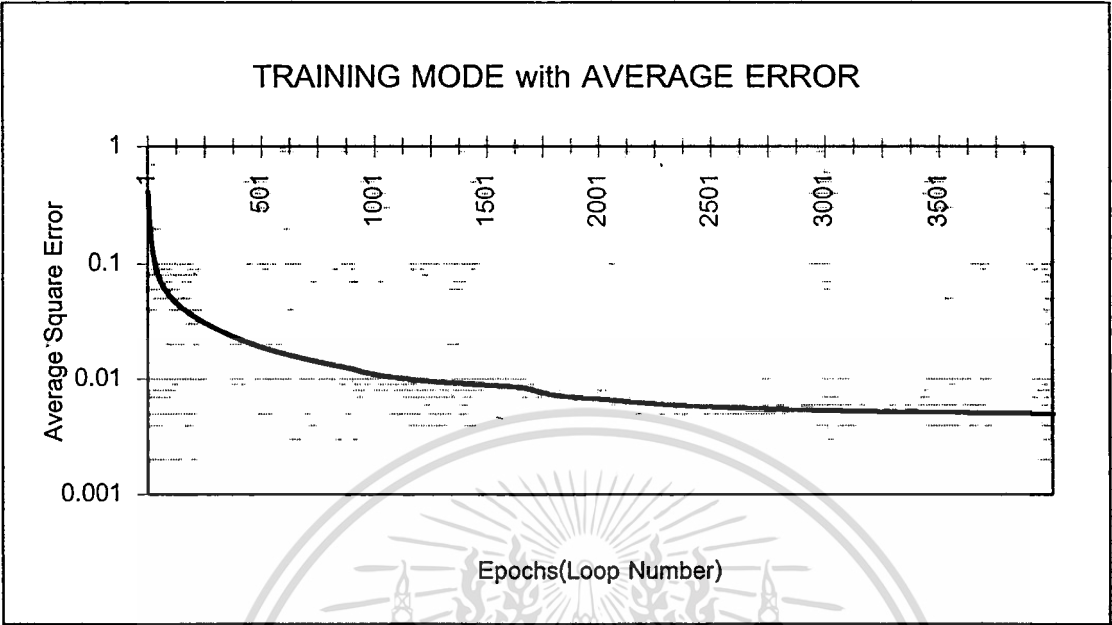
กราฟแสดงค่า Sum square error ที่แสดงในภาคผนวกนี้ ไม่ได้นำมาแสดงทุกการทดลอง
กราฟในส่วน อื่นๆ จะมีลักษณะการลดลงของค่า Sum square error กราฟที่คล้ายกัน

ภาพที่ 39



กราฟแสดงค่า Sum square error ที่ได้จากการฝึกสอน นิวรัลเน็ตเวิร์ค ด้วยข้อมูลคุณลักษณะ
จำเพาะของเสียงในกลุ่ม A

ภาพที่ 40



กราฟแสดง ค่า Average Sum Square Error ของการฝึกสอน ด้วยเสียงในกลุ่มคำสั่งควบคุม 9 เสียง
20 คน คนละ 3 ครั้ง

ภาคผนวก ค

โปรแกรมที่พัฒนาขึ้นในการวิจัย

โปรแกรมหาค่าคุณลักษณะจำเพาะสัมพันธ์ของเสียง

แบบอ้างอิงจากผู้พูด 10 คน

```

/* P/R/O/G/R/A/M/ F/O/R/ C/R/E/A/T/E/I/N/G/ */

/* V/O/I/C/E/ C/H/A/R/A/C/T/E/R/ R/E/L/A/T/I/O/N/ */

/* filename: VCHACO.C */
/* AUTOMATIC input test file by use 3 band overlapping correlation */
/* Band1 :125Hz-1KHz 0-75 */
/* Band2 :500Hz-2kHz 50-100 */
/* Band3 : 1KHz-4KHz 75-125 */

/* ##### */
/* ##### T/H/I/S/ F/I/L/E/ U/S/E/ F/O/R ./O/T/V/ O/N/L/Y/ ##### */
/* ##### */

/* FIND CORRELATION BETWEEN TESTING VOICE WITH REFERENCE PATTERN VOICES */
/* December,20,1996. B10 persons with A10 persons in c:\cvoc\ref\ */

#include<math.h>
#include<conio.h>
#include<stdarg.h>
#include<stdlib.h>
#include<dos.h>
#include<io.h>
#include<stdio.h>
#include<graphics.h>

#define Voice 4
#define Pitch 24
#define Channel 75

struct octavetype {
    char name[7];
    char version;
    char information[15];
    char endofname;
    long datalength;
    char filetype;
    char bit_per_sampling;
    int max_frequency;
    int FFTgroup;
    int FFTstep;
    char segmentno;
    char windowno;
    int step_per_octave;
    int base_frequency;
    int otvcount;
    char reserved[20];
}otv;

/* GLOBAL VARIABLE */
int basenergy[Voice][Pitch][Channel];
int energy[Voice][Pitch][Channel];
float corrl[3][16][3][16];
int speechbase[Channel],speech[Channel];
int ibase,i,outcount,seq,seq2=0;
char fname[15],filename[15],xfile[15];
char arame[400][9];

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

float   aval[252];
int     quake,quake2;
int     vname[2000];
float   vsim[2000],test;
int     t[20];                      /* use in summary */
int     bandstar,bandstop;          /* keep start-stop band */
int     chain;                      /* use in energy detec */
int     bandcntr;                   /* use in main */

FILE    *fp;
FILE    *fpw;
FILE    *fpx;

FILE    *fp1;
FILE    *fp2;
FILE    *fp3;

FILE *openfile (char *name, char *format)
{
    FILE *fp;
    if ( (fp=fopen(name, format)) == (FILE *) NULL )
    {
        perror("Open file-read failed \n");
        exit(1);
    }
    return fp;
}

void energydetec()
{
    int   a,b,c,d,e,f,g,h,c1,c2,p,dcount=0;
    int   k,k1,k2,k3,k4,k5,k6,fz,plot,sumax,avg,avghi,avglo;
    int   s1,s4,s5,s6,s7,s8,w,x,y,z,width=1;
    int   j,l,m,n,o,xslice=0,slice=4;
    int   map,mark;
    int   max[32],min[32];
    float maxminscale,fy;
    float c3;
    char   l[3],ah[3],al[3],dc[5];

    for(i=0;i<32;i++) /* Clear max&min array to zero */
    {
        max[i]=0;
        min[i]=1000;
    }

    fp = openfile(fname, "rb");
    avghi=avglo=0;
    s4=s7=0;
    x=d=f=o=0;
    y=e=224;
    n=128; m=50;
    k5=400;
    sumax=avg=0;

    /* ##### */
    /* ##### R/E/A/D/   H/E/A/D/E/R/   F/I/L/E/S/   ##### */
    /* ##### */
    fread(&otv,sizeof(struct octavetype),1,fp);

    while ( (c=getc(fp)) != EOF ) /* Get value from file */
    {
        if (x >= 128*width) /* Check if complete 1 line */
        {
            x=d=0;
        }
        sumax = sumax+c;
        dcount=dcount+1;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

f=f+1;
if(f>=128)
{
f=0;
avg = sumax/128;
if(s4<1)
{
avghi=avglo=avg;
s4=2;
}
if(avg>avghi)
avghi = avg;          /* keep maximum value */
if(avg<avglo)
avglo = avg;          /* keep minimum value */
sumax = 0;
o=(k5-(avg*2));
n=n+4;
xslice=xslice+slice;
}
x=x+width;
}
fclose(fp);
/* ***** */
s7= ((avghi-avglo)*80/100)+avglo;
/* ***** */
/* ##### F/I/N/D/ U/S/I/N/G/ A/R/E/A/ ##### */
/* ##### */
e=x=y=i=o=0;
map=sumax=0;

/* Open file again by reset fp */
fp = fopen(fname, "rb");
fread(&otv,sizeof(struct octavetype),1,fp);
d=(otv.datalength/(otv.FFTstep/2))*-10*/;
i = m = o = p = mark = 0;
n = 1;
while ((c=getc(fp))!=EOF) /* Get character from file */
{
if(x>=128)
{
x=0;
p=0;
m++;
y++;
o++;
sumax=sumax/128;
if(sumax<s7)
{
map=1;
mark++;
}
else
{
map=0;
e=1;
}
if((e>0) && (map>0))
{
i++;
m=0;
e=0;
}

if((mark>=1) && (map<1))
{
speech[i]=mark;
mark=0;
}
}
}

```

```

        sumax=0;
    }
    if(y<d);
    {
        if(map>0 && (x>=bandstar && x<bandstop))
        {
            energy[i][m][p] = c;
            if(energy[i][m][p] > max[i])
                max[i]=energy[i][m][p];
            if(energy[i][m][p] < min[i])
                min[i]=energy[i][m][p];
            p++;
        }
    }
    sumax=sumax+c;
    x++;
}
fclose(fp);
}

```

```

void basetransfer()
{
    int    w,m,p;
    ibase=i;
    for(w=1;w<ibase+1;w++)
    {
        speechbase[w]=speech[w]; /* speechbase are base of the first
files */
        for(m=0;m<speechbase[w];m++)
        {
            for(p=0;p<(bandstop-bandstar);p++)
            {
                basenergy[w][m][p]=energy[w][m][p];
            }
        }
    }
}

```

/* Find Voices Correlation Coefficient*/

```

void correl()
{
    int          base[32],opr[32],dbase[32],dopr[32];
    long          sumbas,sumopr;
    float         rxy,ssdbase,ssdopr,ssqx,ssqy,ssqxy;
    int           count;
    int           wbase,mbase,pbase,wopr,mopr,popr;
    float         meanx,meany;
    sumbas=sumopr=ssdbase=ssdopr=ssqx=ssqy=ssqxy=0;
    /* ***** */
    /* **** PEASON'S PRODUCT MOMENT CORRELATION COEFFICIENT **** */
    /* ***** */
    for(wbase=1;wbase<ibase+1;wbase++)
    {
        for(mbase=0;mbase<speechbase[wbase];mbase++)
        {
            for(pbase=0;pbase<(bandstop-bandstar);pbase++)
            {
                sumbas =
sumbas+basenergy[wbase][mbase][pbase];
            }
            meanx = sumbas/((bandstop-bandstar)*1.0);
            sumbas=0;
            for(wopr=1;wopr<i+1;wopr++)
            {
                for(mopr=0;mopr<speech[wopr];mopr++)
                {
                    for(count=0;count<(bandstop-bandstar);count++)

```

```

    {
        sumopr = sumopr+energy[wopr][mopr][count];
    }
    meany = sumopr/((bandstop-bandstar)*1.0);
    sumopr=0;
    for(popr=0;popr<(bandstop-bandstar);popr++)
    {
        ssdbase = basenergy[wbase][mbase][popr]-
meanx;

        ssdopr = energy[wopr][mopr][popr]-meany;
        ssqx = ssqx+(ssdbase*ssdbase);
        ssqy = ssqy+(ssdopr*ssdopr);
        ssqxy= ssqxy+(ssdbase*ssdopr);
    }
    rxy = ssqxy/sqrt(ssqx*ssqy);

    sumbas=sumopr=ssdbase=ssdopr=ssqx=ssqy=ssqxy=0;
    corrl[wbase][mbase][wopr][mopr]=rxy;
    }
}
meanx=meany=0;
}
}

/* ***** */
/* search rxy again for voice(Base) by voice(Distribution) correlation */
/* ***** */

void resort()
{
    int    Voicebase,Subvoicebase,voicedis,subvoicedis,unimat;
    float  cor2;
    float  similar,average;
    float  maxim,maxim2;
    maxim=maxim2=0;
    average=0;
    unimat =0;
    for(Voicebase=1;Voicebase<ibase+1;Voicebase++)
    {
        for(voicedis=1;voicedis<i+1;voicedis++)
        {
            for(Subvoicebase=0;Subvoicebase<speechbase[Voicebase];Subvoicebase++)
            {
                for(subvoicedis=0;subvoicedis<speech[voicedis];subvoicedis++)
                {
                    /* [%d] [%d] [%d] [%d]= [number of Base voice][number of Subbase voice]
                    [number of Distribution voice][number of SubDistribution
                    voice]
                    consequently */
                    if(Subvoicebase==subvoicedis)
                    {
                        unimat++;
                        cor2=corrl[Voicebase][Subvoicebase][voicedis][subvoicedis];
                        average=average+cor2;
                    }
                }
            }
            similar=average/unimat;
            quake2 = quake/10;          /* here "10 are number of persons"
*/
            vname[seq2]= quake2;
            vsim[seq2]=similar;
            seq2++;
            if(maxim<similar)
                maxim=similar;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

        average=0;
        unimat=0;
    }
    if (maxim2<maxim)
        maxim2=maxim;
    }
    aval[quake]=maxim2;
    printf(">");
seq++;
}

sortint_1()
{
    int      a,b,c,n,temp2;
    float    temp;

    n=seq2;
    a=b=0;

    printf("\n JUST A MINUTE FOR SORTING %d SIMILARITY DATA ",n);
    for(a=1;a<n;a++)
    {
        for(b=0;b<n-1;b++)
        {
            if((vsim[b+1]-vsim[b]) > 0)
            {
                temp=vsim[b];
                temp2=vname[b];

                vsim[b]=vsim[b+1];
                vname[b]=vname[b+1];

                vsim[b+1]=temp;
                vname[b+1]=temp2;
            }
        }
    }

}

summary()
{
    int      count1,count2,voice,tempo,m,n;
    int      pattern[20]=
        {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19};

    m=20;
    n=60;

    /* Limited Sort by select 10% of every voices correlative coefficient */
    tempo=0;

    for(count2=0;count2<m;count2++)
    {
        for(count1=0;count1<n;count1++)
        {
            if(pattern[count2]==vname[count1])
                tempo++;
        }
    }

    printf("\n      Band:%d, Voice: %d, Amount:%d\n",bandcntr+1,pattern[count2],tempo);

    if(bandcntr==0)
        fprintf(fp1,"%d ",tempo);

    if(bandcntr==1)
        fprintf(fp2,"%d ",tempo);

    if(bandcntr==2)
        fprintf(fp3,"%d ",tempo);
}

```

```

        tempo=0;
    }
}

void main()
{
    char    name;
    char    string[15],testfile[15];
    char    fnamebuf[15],voctest[15],vocbase[15];
    char    BAND1[15],BAND2[15],BAND3[15];

    int     w,m,p,q;
    FILE    *fpy,*fpz;

    clrscr();
    printf("\nIndex Table File for BASE REFERENCE (Text file format) ");
    printf("\nFILE NAME (base1.txt) : ");
    gets(vocbase);
    printf("\nIndex Table File for TESTING DATA (Text file format) ");
    printf("\nFILE NAME (data1.txt) : ");
    gets(voctest);
    printf("\nVoices Characteristic Relation BAND1 : 125Hz-1KHz ");
    printf("\nOutput File Name BAND1 : ");
    gets(BAND1);
    printf("\nVoices Characteristic Relation BAND2 : 500Hz-2KHz ");
    printf("\nOutput File Name BAND2 : ");
    gets(BAND2);
    printf("\nVoices Characteristic Relation BAND3 : 1KHz-4KHz ");
    printf("\nOutput File Name BAND3 : ");
    gets(BAND3);
    fpz = fopen(voctest,"r"); /* TESTING FILE THAT KEEP FILES NAME *.OTV FOR TEST
*/
    fp1 = fopen(BAND1,"w"); /* ("coeff1.txt","w"); */
    if(fp1 == NULL)
    {
        fprintf(stderr,"Error opening file %s.",xfile);
        exit(1);
    }
    fp2 = fopen(BAND2,"w"); /* ("coeff2.txt","w"); */
    if(fp2 == NULL)
    {
        fprintf(stderr,"Error opening file %s.",xfile);
        exit(1);
    }
    fp3 = fopen(BAND3,"w"); /* ("coeff3.txt","w"); */
    if(fp3 == NULL)
    {
        fprintf(stderr,"Error opening file %s.",xfile);
        exit(1);
    }
    while((fgets(testfile,9,fpz)) != NULL)
    {
        strcpy(fnamebuf,testfile);
        for (bandcntr=0;bandcntr<3;bandcntr++)
        {
            switch (bandcntr)
            {
                case 0:
                    bandstar=0;
                    bandstop=75;
                    break;

                case 1:
                    bandstar=50;
                    bandstop=100;
                    break;
            }

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

                                case 2:
                                bandstar=75;
                                bandstop=127;
                                break;
                                }
printf("\nPLEASE WAIT FOR PROCESSING \n");
if (bandcntr==0)
    fprintf(fp1,"\n%s ",fnamebuf);
if (bandcntr==1)
    fprintf(fp2,"\n%s ",fnamebuf);
if (bandcntr==2)
    fprintf(fp3,"\n%s ",fnamebuf);
strcpy(fname,fnamebuf);
printf(" File: %s, BAND : %d || bandstart= %d , bandstop= %d \n",fname,
bandcntr+1,bandstar,bandstop);
energydetec();
basetransfer();
fpy = fopen(vocbase, "r");/* REFERENCE FILE THAT KEEP FILES NAME *.OTV FOR
REFERENCE */
quake=0;
while ( (fgets(string,9,fpy)) != NULL ) /* Get string from file */
{
    strcpy(fname,string);
    strcpy(arame[quake],string);
    energydetec();/* output of this function are 3d array
'energy[] [] []' */
    correl();
    resort();
    quake++;
    fgets(string,9,fpy); /* compensate */
} /* end while for auto files filling */
fclose(fpy);
sortint_1();
summary();
seq=seq2=0;
}/* 's for(bandcntr.....) loop */
fgets(testfile,9,fpz); /* compensate */
} /* end of the first while */
fclose(fpz);
fclose(fp3);
fclose(fp2);
fclose(fp1);
printf("\n\n OK THAT COMPARING ARE COMPLETED ");
getch();
}
/*----- END OF MAIN PROGRAM -----*/

```

โปรแกรมฝึกสอนให้หาวัลเน็ตเวิร์ค

BPN.C

```

/* N/E/U/R/A/L/ N/E/T/W/O/R/K/ P/R/O/G/R/A/M/I/N/G/ */

/* B/A/C/K/P/R/O/P/A/G/A/T/I/O/N/ T/R/A/I/N/I/N/G/ P/R/O/C/E/S/S/ */

/* W/I/T/H/ O/P/T/I/O/N/ M/O/M/E/N/T/U/M/ */

/* FILE: BPN.C */
/* The last modify: January,8,1997. */
/* PROGRAMMING BY SERI PANSANG */
/* For input range = wide 0-1 output range = 0 to 1 */

#include<math.h>
#include<conio.h>
#include<stdarg.h>
#include<stdlib.h>
#include<dos.h>
#include<io.h>
#include<stdio.h>
#include<graphics.h>

/* GLOBAL VARIABLE DECLAIRATION */
double Opi[40];
double Wji[40][40];
double Opj[128];
double Wkj[40][40];
double Deltabuff1[20][30];/* Delta Wkj(t) for momentum option */
double Deltabuff2[20][30];/* Delta Wkj(t-1) for momentum option */
double Opk[20];
double Tpk[20];
double Error[30];
double Deltapk[30],Deltapj[40],SummationK[40];
double netpj,netpk;
double difnetpj[40],difnetpk[30];
int round=0,setpoint;
int seq=0;
int sequence=0;
long ep=0;
double sssqerr,ssqerr,sqerr;
FILE *fpx,*fpo;
FILE *inputfp,*outputfp,*target,*wlayer1,*wlayer2;

void opengraph ()
{
    struct palettetype pal;
    int i, ht, y, xmax;
    int g, driver, mode;

    detectgraph(&driver,&mode);
    driver = VGA;
    mode = VGAHI;
    initgraph(&driver,&mode,"");
    if((g=graphresult()) !=grOk)
    {
        printf("Graphics System Error:%d\n",g);
        exit(0);
    }
    getpalette(&pal);

    /* create gray scale */
    for (i=0; i<pal.size; i++)
        setrgbpalette(pal.colors[i], i*4, i*4, i*4);
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

FILE *openfile (char *name, char *format)
{
    FILE *fp;
    if ( (fp=fopen(name, format)) == (FILE *) NULL )
    {
        perror("Open file-read failed \n");
        exit(1);
    }
    return fp;
}

rndweight (HID, INL, OUT)
/*
    THIS ROUTINE USE FOR RANDOM WEIGHT FOR HIDDEN NEURALS AND OUTPUT NEURALS
    weightw : use for counting loops the weight "W" of neural
    inputx  : use for counting loops the input "X" of neural
    */

    int *HID,*INL,*OUT;
{
    int    inputx,weightw;
    int HIDNEURAL=*HID;
    int INLAYER=*INL;
    int OUTNEURAL=*OUT;
    /*****
    /***** PUT IN RANDOM WEIGHT FOR HIDDEN NEURALS AND OUTPUT NEURALS *****/
    /*****
    /* For hidden neurals */
        for(weightw=0;weightw<HIDNEURAL;weightw++)
            for(inputx=0;inputx<INLAYER;inputx++)
            {
                Wji [weightw] [inputx]= 0.5-(random(9999)/10000.0);
            }
    /* For output neurals */
        for(weightw=0;weightw<OUTNEURAL;weightw++)
            for(inputx=0;inputx<HIDNEURAL;inputx++)
                Wkj [weightw] [inputx]=0.5-(random(9999)/10000.0);
    /* For prepare & initial Wkjbuff2 in the first time */
        for(weightw=0;weightw<OUTNEURAL;weightw++)
        {
            for(inputx=0;inputx<HIDNEURAL;inputx++)
            {
                Deltabuff2 [weightw] [inputx]=0;
            }
        }
    /*****
    }
    read_input_file(int *INL)
    {
        int    INLAYER=*INL;
        int    q,x;
        float  ival;
        x=q=0;
        while ( (fscanf(inputfp,"%f", &ival)) != EOF ) /* Get value from file */
        {
            if (x >= INLAYER-1)
            {
                Opi [x]=ival;
                x=0;
                q++;
                break;
            }
            Opi [x]=ival;
            x++;
        }
    }
    void read_target_file(int *OUT)
    {

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

int    OUTNEURAL=*OUT;
int    q,x;
float  tarval;
x=q=0;
while ( (fscanf(target,"%f", &tarval)) != EOF ) /* Get value from file
*/
{
    if (x >= OUTNEURAL-1)
    {
        Tpk[x]=tarval;
        x=0;
        q++;
        break;
    }
    Tpk[x]=tarval;
    x++;
}

```

```

void propagate_forward(int *HID,int *INL,int *OUT)
{
    int    HIDNEURAL=*HID;
    int    INLAYER=*INL;
    int    OUTNEURAL=*OUT;
    int    j,i;
    for(j=0;j<HIDNEURAL;j++) /* j is the number of hidden neural */
    {
        netpj=0;
        for(i=0;i<INLAYER;i++)
        {
            netpj=netpj+(Opi[i]*Wji[j][i]);/* update point */
        }
        Opj[j]=1.0/(1.0+exp(-netpj/*+THETAJ*/));
    }
    ssqerr=sqerr=0;
    for(j=0;j<OUTNEURAL;j++) /* j is the number of output neural */
    {
        netpk=0;
        for(i=0;i<HIDNEURAL;i++)
        {
            netpk=netpk+(Opj[i]*Wkj[j][i]); /* NET=Opj*Wkj */
        }
        Opk[j]=1.0/(1.0+exp(-netpk/*+THETAK*/)); /* OUT=F(NET) */
        sqerr=sqerr+( (Tpk[j]-Opk[j])*(Tpk[j]-Opk[j]));
    }
    ssqerr=sqerr/2;
}

```

```

void compute_output_error(int *OUT)
{
    int    OUTNEURAL=*OUT;
    int    i,j;
    for(j=0;j<OUTNEURAL;j++) /* length(outputs) */
    {
        Deltapk[j]=Opk[j]*(1.0-Opk[j])*(Tpk[j]-Opk[j]);
    }
}

```

```

void backpropagate_error(int *HID,int *OUT) /* backpropagate error from an
upper to lower layer */
{
    int    HIDNEURAL=*HID;
    int    OUTNEURAL=*OUT;
    int    i,j;
    for(i=0;i<HIDNEURAL;i++)
    {
        SummationK[i]=0;
        for(j=0;j<OUTNEURAL;j++)

```

```

        {
            SummationK[i]=SummationK[i]+(Deltap[k]*Wkj[j][i]);
        }
        SummationK[i]=SummationK[i]*Opj[i]*(1.0-Opj[i]);
    }
}

void adjust_weightji(HID,INL,OUT,L,ALP)
int  *HID, *INL, *OUT;
float *L, *ALP;
{
    int    HIDNEURAL=*HID;
    int    INLAYER=*INL;
    int    OUTNEURAL=*OUT;
    float  LR=*L;
    float  ALPHA=*ALP;
    int    i,j,k;

    /* adjust Wji */
    for(j=0;j<HIDNEURAL;j++)
    {
        for(k=0;k<INLAYER;k++)
        {
            Wji[j][k]=Wji[j][k]+(Opi[k]*LR*SummationK[j]);
        }
    }
    /* adjust Wkj with option momentum */
    for(j=0;j<OUTNEURAL;j++)
    {
        for(k=0;k<HIDNEURAL;k++) /* prepare for momentum */
        {
            Deltabuff1[j][k]=Opj[k]*LR*Deltap[k];
        }
    }
    for(j=0;j<OUTNEURAL;j++) /* adjust Wkj */
    {
        for(k=0;k<HIDNEURAL;k++)
        {
            Wkj[j][k]=Wkj[j][k]+Deltabuff1[j][k]+(ALPHA*Deltabuff2[j][k]);
            Deltabuff2[j][k]=Deltabuff1[j][k]; /* Keep old delta Wkj for
momentum */
        }
    }
}

void main()
{
    int    INRECORD;
    int    INLAYER;
    int    OUTNEURAL;
    int    HIDNEURAL;

    float  LR;
    float  THETAJ=0.05;
    float  THETAK=0.05; /*
    float  ALPHA;
    int    frontend,i,j,k;
    char    xfile[15];
    char    inputa[15],targeta[15],error[15],weight[15],output[15];
    FILE    *fpo;
    k=0;
    round=0;
    clrscr();
    printf("\n          WELCOME TO TRAINING MODE \n");
    printf("\n");
    printf("N e t w o r k   A l i g n m e n t   \n");

```

```

printf("\n");

printf("2-LAYER NEURAL NETWORKS\n ");
printf("\n");
printf("\nNumber of INPUT NEURONS(MAX=40) : ");
scanf("%d",&INLAYER);
printf("\nNumber of HIDDEN NEURONS(MAX=40) : ");
scanf("%d",&HIDNEURAL);
printf("\nNumber of OUTPUT NEURONS(MAX=20) : ");
scanf("%d",&OUTNEURAL);
printf("\nLEARNING RATE (Between 0-1, Normal= 0.1) : ");
scanf("%f",&LR);
printf("\nMOMENTUM : (Between 0-1, Normal= 0.90) : ");
scanf("%f",&ALPHA);

printf("\nSUCCESSFULL ADJUSTMENT \n PRESS ANY KEY TO SET FILES FOR PROCESSING:
");
getch();
clrscr();
printf("\n          SET FILES FOR PROCESSING:  press any key");
getch();

printf("\nInput File      :  ");
gets(inputa);
printf("\n\nNumber of RECORD of INPUT FILE :  ");
scanf("%d",&INRECORD);
gets(targeta);
printf("Target File      :  ");
gets(targeta);
printf("Error File         :  ");
gets(error);
printf("Number of Training round :  ");
scanf("%d",&setpoint);

printf("Weight File       :  ");
gets(weight);

fpo = fopen(error,"w");
if(fpo == NULL)
{
    fprintf(stderr,"Error oprning files %s.",xfile);
    exit(1);
}
fpx = fopen(weight,"w");
if(fpx == NULL)
{
    fprintf(stderr,"Error oprning files %s.",xfile);
    exit(1);
}

rndweight(&HIDNEURAL,&INLAYER,&OUTNEURAL);
for(round=0;round<=setpoint;round++)
{
    inputfp = fopen(inputa,"r"); /* KEEP INPUT FILE NAME FOR PROCESS */
    target = fopen(targeta,"r"); /* KEEP TARGET FILE NAME FOR PROCESS */

    sssqerr=0;
    for(sequence=0;sequence<INRECORD;sequence++)
    {
        read_input_file(&INLAYER);
        read_target_file(&OUTNEURAL);

        propagate_forward(&HIDNEURAL,&INLAYER,&OUTNEURAL);
        sssqerr=sssqerr+ssqerr;
        compute_output_error(&OUTNEURAL);
        backpropagate_error(&HIDNEURAL,&OUTNEURAL);
    }
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

        adjust_weightji(&HIDNEURAL,&INLAYER,&OUTNEURAL,&LR,&ALPHA);
    }
    fclose(target);
    fclose(inputfp);
    printf("\n %d ",ep);
    printf(" Avg ssqerr = %1.8f ",sssqerr/INRECORD);
    fprintf(fpo,"%1.8f\n",sssqerr/INRECORD);
    ep++;
}
printf("\nPress any key for save weight Wji and Wkj \n");
getch();
/***** Keep the last weight(Wji and Wkj) to Weight file *****/
/***** KEEP THE LAST Wji */
/*
    fprintf(fpx,"\nWji\n"); */
    for(j=0;j<HIDNEURAL;j++)
    {
        fprintf(fpx,"\n");
        for(i=0;i<INLAYER;i++)
            fprintf(fpx,"%2.6f ",Wji[j][i]);
        fprintf(fpx,"\n");
        fprintf(fpx,"\n");
/* KEEP THE LAST Wkj */
/*
    fprintf(fpx,"\nWkj\n"); */
    for(k=0;k<OUTNEURAL;k++)
    {
        fprintf(fpx,"\n");
        for(j=0;j<HIDNEURAL;j++)
            fprintf(fpx,"%2.6f ",Wkj[k][j]);
        fprintf(fpx,"\n");
    }
/***** End of process: press any key \n*****/
    printf("\nEnd of process: press any key \n");
    fclose(fpo);
    fclose(fpx);
    getch();
}

```

โปรแกรมทดสอบเสียงด้วยนิวรัลเน็ตเวิร์ค

```

/* TESTING NETWORKS PROGRAM */

/* File: Testnet.c */
/* January,12,1997. */
/* This program made for testing neural networks*/
/* For input range = -1 to 1   output range = 0 to 1 */
#include<math.h>
#include<conio.h>
#include<stdarg.h>
#include<stdlib.h>
#include<dos.h>
#include<io.h>
#include<stdio.h>
#include<graphics.h>

/* GLOBAL VARIABLE DECLAIRATION */

float  Opi[40];
float  Wji[40][40];
double Opj[128];
double Wkj[40][40];
double Opk[20];
double Tpk[20];
double Error[30];
double Deltapk[30],Deltapj[40],SummationK[40];
double netpj,netpk;
double difnetpj[40],difnetpk[30];
int    round=0;
int    seq=0;
int    sequence=0;
long   ep=0;
float  sssqerr,ssqerr,sqerr;
FILE   *fpx; /* now not use */
FILE   *fpo; /* for index  output file "output.out" */
FILE   *inputfp,*outputfp,*target,*wlayer1,*wlayer2;
FILE   *inputWji,*inputWkj;

void opengraph ()
{
    struct palettetype pal;
    int    i, ht, y, xmax;
    int    g, driver, mode;
    detectgraph(&driver,&mode);
    driver = VGA;
    mode = VGAHI;
    initgraph(&driver,&mode,"");
    if((g=graphresult()) !=grOk)
    {
        printf("Graphics System Error:%d\n",g);
        exit(0);
    }
    getpalette(&pal);
    /* create gray scale */
    for (i=0; i<pal.size; i++)
        setrgbpalette(pal.colors[i], i*4, i*4, i*4);
}

FILE *openfile (char *name, char *format)
{
    FILE *fp;
    if ( (fp=fopen(name, format)) == (FILE *) NULL )
    {
        perror("Open file-read failed \n");
        exit(1);
    }
    return fp;
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

void inputweight(int *INL,int *HID,int *OUT) /* Read input weight from file */
{
    int  INLAYER=*INL;
    int   HIDNEURAL=*HID;
    int  OUTNEURAL=*OUT;
    /*
    THIS ROUTINE USE FOR INPUT WEIGHT FOR HIDDEN NEURALS AND OUTPUT NEURALS
    weightw : use for counting loops the weight "W" of neural
    inputx   : use for counting loops the input "X" of neural
    */

    int  p,q,x,y;
    float weightvalue;
    x=y=p=q=0;
    while ( (fscanf(inputWji,"%f", &weightvalue)) != EOF ) /* Get value from file
    */
    {
        if (q<HIDNEURAL)
        {
            if (x >=INLAYER)
            {
                x=0;
                q++;
            }
            if(q != HIDNEURAL)
            {
                Wji[q][x] = weightvalue;
                printf("Wji[%d][%d] = %f \n",q,x,Wji[q][x]);
            }
            x++;
        }
        if (p<OUTNEURAL && q>=HIDNEURAL)
        {
            if (y >= HIDNEURAL)
            {
                y=0;
                p++;
            }
            if(p != OUTNEURAL)
            {
                Wkj[p][y] = weightvalue;
                printf("Wkj[%d][%d] = %f \n",p,y,Wkj[p][y]);
            }
            y++;
        }
    }
    fclose(inputWji);
}

void read_input_file(int *INL)
{
    int  INLAYER=*INL;
    int   q,x;
    float ival;
    x=q=0;
    while ( (fscanf(inputfp,"%f", &ival)) != EOF ) /* Get value from file */
    {
        if (x >= INLAYER-1)
        {
            Opi[x]=ival;
            x=0;
            q++;
            break;
        }
        Opi[x]=ival;
        x++;
    }
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

}

void propagate_forward(int *HID,int *INL,int *OUT)
{
    int    HIDNEURAL=*HID;
    int    INLAYER=*INL;
    int    OUTNEURAL=*OUT;
    int    j,i;
    for(j=0;j<HIDNEURAL;j++)    /* j is the number of hidden neural */
    {
        netpj=0;
        for(i=0;i<INLAYER;i++)
        {
            netpj=netpj+(Opi[i]*Wji[j][i]);/* update point */
        }
        Opj[j]=1.0/(1.0+exp(-netpj));
    }

    for(j=0;j<OUTNEURAL;j++)    /* j is the number of output neural */
    {
        netpk=0;
        for(i=0;i<HIDNEURAL;i++)
        {
            netpk=netpk+(Opj[i]*Wkj[j][i]); /* NET=Opj*Wkj */
        }
        Opk[j]=1.0/(1.0+exp(-netpk)); /* OUT=F(NET) */
        printf("%2.4f  ",Opk[j]);
        fprintf(fpo,"%f ",Opk[j]); /* keep data to file: output.out
*/
    }
}

void main()
{
    int    INRECORD;
    int    INLAYER;
    int    OUTNEURAL;
    int    HIDNEURAL;
    float  LR;
    float  ALPHA;
    char    input[15],output[15],weight[15];
    int    frontend,i,j,k;
    char    xfile[15];
    round=0;
    clrscr();
    printf("\n      WELCOME TO NEURAL NETWORKS TESTING MODE  \n");
    printf("\n");
    printf("N e u r a l   N e t w o r k   C r e a t i n g  \n");
    printf("SETTING PARAMETER \n");
    printf("\n");
    printf("\n 2-LAYER NEURAL NETWORKS \n");
    printf("\n");
    printf("Number of INPUT NEURONS(MAX=40)      :  ");
    scanf("%d",&INLAYER);
    printf("\nNumber of HIDDEN NEURAL(MAX=40)  :  ");
    scanf("%d",&HIDNEURAL);
    printf("\nNumber of OUTPUT NEURONS(MAX=20) :  ");
    scanf("%d",&OUTNEURAL);
    printf("\nLEARNING RATE (Between 0.0-1.0,Normal= 0.1)  :  ");
    scanf("%f",&LR);
    printf("\n MOMENTUM (Between 0.0-1.0,Normal= 0.90)  :  ");
    scanf("%f",&ALPHA);
    printf("\n SUCESSFULL SETTING  NEXT STEP FOR FILL FILE NAME  ");
    printf("\n PRESS ANY KEY :  ");
    getch();

    clrscr();
    printf("\n SET FILES NAME FOR PROCESSING : \n");

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

printf("\n");
gets(input);
printf("\nInput File      :    ");
gets(input);
printf("\n\nNumber of RECORD of INPUT FILE :    ");
scanf("%d",&INRECORD);
gets(weight);
printf("\nWeight File      :    ");
gets(weight);
printf("\nOutput File      :    ");
gets(output);

inputWji = fopen(weight,"r");
inputfp = fopen(input,"r");/* for read_input_file */
fpo = openfile(output,"w");
if(fpo == NULL)
{
    fprintf(stderr,"Error oprning files %s.",xfile);
    exit(1);
}

inputweight(&INLAYER,&HIDNEURAL,&OUTNEURAL);
getch();
for(sequence=0;sequence<=INRECORD;sequence++)
{
    read_input_file(&INLAYER);
    propagate_forward(&HIDNEURAL,&INLAYER,&OUTNEURAL);
    printf("\n"); /* new line at display on screen */
    fprintf(fpo,"\n"); /* new line for output.out */
}
printf("\n\n                                End of process:  press any key
\n");
printf("\n\n                                Output keep in file: %s \n",output);
getch();
fclose(fpo);
fclose(inputfp);
}
END.

```

โปรแกรมลดจำนวนข้อมูลของสเปกโตรแกรมและปรับให้มีจำนวนสเปกตรัมคงที่ที่ 10สเปกตรัมโดยใช้ วิธีลากรอง อินเตอร์โพลแทน

```

/* OTVAVG4.C
/* Reduce data enough for input neural network  develop No.1*/
/* By using average and Lagrange interpolation polynomials */
/* ##### */
/* ##### T/H/I/S/ F/I/L/E/ U/S/E/ F/O/R  ./O/T/V/ O/N/L/Y/ ##### */
/* ##### */
/* DISPLAY SPECTROGRAM IN FREQUENCY DOMAIN AND AVERAGE ENERGY POWER */
/* FIND MAX & MIN VALUE */
/* AND CONVERSE DATA TO  A/S/C/I/I  F/I/L/E/S */

#include<conio.h>
#include<stdarg.h>
#include<stdlib.h>
#include<dos.h>
#include<io.h>
#include<stdio.h>
#include<graphics.h>

/* void.opengraph () */
struct octavetype {
    char    name[7];
    char    version;
    char    information[15];
    char    endofname;
    long    datalength;
    char    filetype;
    char    bit_per_sampling;
    int     max_frequency;
    int     FFTgroup;
    int     FFTstep;
    char    segmentno;
    char    windowno;
    int     step_per_octave;
    int     base_frequency;
    int     otvcount;
    char    reserved[20];
}otv;

void.opengraph ()
{
    struct palettetype pal;
    int    i, ht, y, xmax;
    int    g, driver, mode;

    detectgraph(&driver,&mode);
    driver = VGA;
    mode = VGAHI;
    initgraph(&driver,&mode,"");
    if((g=graphresult()) !=grOk)
    {
        printf("Graphics System Error:%d\n",g);
        exit(0);
    }
    getpalette(&pal);

    /* create gray scale */
    for (i=0; i<pal.size; i++)
        setrgbpalette(pal.colors[i], i*4, i*4, i*4);
}

FILE *openfile (char *name, char *format)

```

```

{
    FILE *fp;
    if ( (fp=fopen(name, format)) == (FILE *) NULL )
    {
        perror("Open file-read failed \n");
        exit(1);
    }
    return fp;
}

void main()
{
    int    a,b,c,d,e,f,g,h,c1,c2,p,dblue=0,dred=0,dcount=0;
    int    k,k1,k2,k3,k4,k5,k6,fx,fz,plot,sumax,avg,avghi,avglo;
    int    s1,s2,s3,s4,s5,s6,s7,s8,w,x,y,z,width=1;
    int    i,j,l,m,n,o,xslice=0,slice=4,vslice=3;
    char    fname[15],filename[15],xfile[15];
    int    map,mark;

    int    energy[16][24][32];
    /* max 16 voices,24 spectrum,32 frq-channel per 1 spectrogram */
    int    icount,jcount,kcount,frame;
    float    voicespec[4][20][16];
    float    outspec[4][20][16],seqx[20];
    float    exampleinput,testinput,lagoutput,dvar,step,check;
    char    string[15],vobase[15];
    int    power[128][128];
    int    avgpow4[128][32];
    float    maxminscale,fy;
    int    max[32],min[32],speech[16];

    float    c3;
    char    al[3],ah[3],al[3],dc[5];
    FILE    *fp;
    FILE    *fpw;
    FILE    *fpx;
    FILE    *fpy;

    for(i=0;i<32;i++) /* Clear max&min array to zero */
    {
        max[i]=0;
        min[i]=1000;
    }
    clrscr();
    printf("\nIndex Table for generating standard format for input ANN");
    printf("\nFILE NAME (TEXTTEST.TXT) :      ");
    gets(vobase);

    printf("\n THIS STEP FOR FIND AVERAGE OF FREQUENCY AXIS 16 CH.");
    printf("\n ENTER .TXT FILES NAME :  ");
    gets(xfile);

    fpx = openfile(xfile, "w");
    if(fpx == NULL)
    {
        fprintf(stderr,"Error opening file %s.", xfile);
        exit(1);
    }
    fpy=openfile(vobase, "r"); /* Keep data for navigtion to RUN */
    while((fgets(string,9,fpy)) != NULL) /* Get String from file */
    {
        strcpy(fname,string);
        fp = openfile(fname, "rb");
    }
    opengraph();
    avghi=avglo=0;
    s4=s5=s6=s7=s8=0;
    x=d=f=o=0;

```

```

y=e=224;
z=480;
l=n=128; m=s2=s3=50;
k5=400;
sumax=avg=0;
/* ##### */
/* ##### R/E/A/D/ H/E/A/D/E/R/ F/I/L/E/S/ ##### */
/* ##### */
fread(&otv,sizeof(struct octavetype),1,fp);
setcolor(13);
outtextxy(250,5,fname);
while ( (c=getc(fp)) != EOF ) /* Get value from file */
{
    if (x >= 128*width) /* Check if complete 1 line old= 640 */
    {
        x=d=0;
    }
    sumax = sumax+c;
    dcount=dcount+1;
    f=f+1;
    for(g=0;g<width;g++)
        putpixel((x+g)-width,dblue,9);

    for(g=0;g<width;g++)
        putpixel((x+g)-width,dred,6);

    if(f>=128)
    {
        for(y=300;y<480;y++)
            putpixel(x,y,6);
        f=0;
        avg = sumax/128;
        if(s4<1)
        {
            avghi=avglo=avg;
            s4=2;
            /* sound(500); delay(300); nosound(); */
        }
        if(avg>avghi)
            avghi = avg; /* keep maximum value */
            s5= (k5-(avghi*2));
        if(avg<avglo)
            avglo = avg; /* keep minimum value */
            s6= (k5-(avglo*2));
        sumax = 0;
        o=(k5-(avg*2));
        setcolor(13);
        /* curve */
        line(1,m,n,o);
        /* average hi */
        line(1,s2,n,s5);
        /* average lo */
        line(1,s3,n,s6);

        l=n;
        m=o;
        s2=s5;
        s3=s6;
        n=n+4;
        xslice=xslice+slice;
    }

/*
    plot=15-(c/6);
    for(k2=(z-(f*vslice));k2>((z-(f*vslice))-vslice);k2--)
        for(k1=(128+xslice);k1<(128+xslice+slice);k1++)
            putpixel(k1,k2,plot);

*/
    x=x+width;

```

```

    }
    xslice=0;

    fclose(fp);

/*      sound(350);      delay(800);      nosound();      */
/* ***** */
    s7= ((avghi-avglo)*80/100)+avglo;
/*      printf("\navghi = %d ",avghi);
      printf("\navglo = %d ",avglo);
      printf("\ncutting at: %d ",s7);      */
        s8=(k5-(s7*2));
        for(i=128;i<640;i++)
            putpixel(i,s8,15);
/* ***** */
/*      F/I/N/D/  U/S/I/N/G/  A/R/E/A/      */
/* ***** */
/*      printf("\n Press any key to continue\n");
      getch();      */
    closegraph();
    for(i=0;i<16;i++) /* clear array "energy" 16voices*/
        for(w=0;w<24;w++)
            for(x=0;x<32;x++)
                energy[i][w][x]=0;

    e=w=x=y=i=o=0;
    map=sumax=0;
    printf("\n Now we read: %s.\n",fname); /*
/* Open file */
    fp = fopen(fname, "rb");
    fread(&otv,sizeof(struct octavetype),1,fp);
    d=(otv.datalength/(otv.FFTstep/2))-10;
/* for limit quantity ofdata getting in time axis */
/*      printf("d=%d \n",d);      */
        i = m = o = p = fx = fy = mark = map= 0;
        n = 1;
    while ((c=getc(fp))!=EOF) /* Get character from file */
    {
        if(x>=128)
        {
            /* fprintf(fpx, "\n"); */ /* new line for average 32 channel */
            x=0;
            p=0;
            m++;
            y++;
            o++;
            sumax=sumax/128;
            if(sumax<s7)
            {
                map=1;
                mark++;
                printf("\n");
            }
            else
            {
                map=0;
                e=1;
            }
            if((e>0) && (map>0))
            {
                i++;
                m=0;
                e=0;
            }
            if((mark>=1) && (map<1))
            {
                speech[i]=mark;

```

```

/* printf("\nspeech[%d]=%d max[%d]=%d min[%d]=%d  \n\n"
,i,speech[i],i,max[i],i,min[i]);
getch(); */
mark=0;
}
sumax=0;
}

if(y<d);
{
if(map>0)
{
fx = fx+c;
if(n>=8) /* average come from sum of 8 values div by 8 */
{
energy[i][m][p] = fx;
if(energy[i][m][p] > max[i])
max[i]=energy[i][m][p];
if(energy[i][m][p] < min[i])
min[i]=energy[i][m][p];
/* printf("%d ",energy[i][m][p]); */
p++;
n = 0;
fx = 0;
}
n++;
}
else
{
energy[i][m][p] = 0;
}
}
sumax=sumax+c;
x++;
}

/* Now we are finding average speech word */
for(w=1;w<i+1;w++)
{
/*printf("\nspeech[%d]=%d max[%d]=%d min[%d]=%d  \n\n ",w,
speech[w],w,max[w],w,min[w]);*/
/* printf("\n");
printf("speech[%d]=%d \n",w,speech[w]); */
for(m=0;m<speech[w];m++)
{
/* fprintf(fpx,"\n"); */ /* new line after 16 ch sub word */
/* printf("\n"); */
for(p=0;p<16;p++)
{
fy = energy[w][m][p]-min[w];
fz = max[w]-min[w];
maxminscale = 1.0-(fy/fz);
voicespec[w][m][p]=maxminscale;
/* fprintf(fpx,"%1.4f ",maxminscale); */
/* printf("%1.2f ",voicespec[w][m][p]);*/ /* maxminscale);*/
/* printf("energy[%d] [%d] [%d]=%d fy=%1.1f,fz=%d
\n",w,m,p,energy[w][m][p],fy,fz); */
}
}

/*****
/* ***** for find value at point of require *****/
/* " Lagrangian Interpolation " **** */
/*****/
for(kcount=0;kcount<speech[w];kcount++)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

        seqx[kcount]=kcount+1;
        step=speech[w]/10.0; /* converse to 10 frame */
        for(kcount=0;kcount<16;kcount++)
        {
            /*      fprintf(fpx,"\n"); */
            /*      printf("\n"); */
            testinput=0;
            for(frame=0;frame<10;frame++) /* converse to 10 frame */
            {
                testinput=testinput+step;
                lagoutput=0;
                for(icount=0;icount<speech[w];icount++)
                {
                    dvar=1;
                    for(jcount=0;jcount<speech[w];jcount++)
                        if(jcount != icount)
                            dvar=dvar*((testinput-seqx[jcount])/(seqx[icount]-seqx[jcount]));
                    lagoutput=lagoutput+voicespec[w][icount][kcount]*dvar;
                }
                outspec[w][frame][kcount]=lagoutput;
                fprintf(fpx,"%1.4f ",lagoutput);
                printf("%1.4f ",lagoutput); /*
            }
        }
    }
    fprintf(fpx,"%s\n",fname); /* New line for new speech word */
}
fclose(fp);
fgets(string,9,fpy); /* Compensate */
} /* End while loop auto files filling */
fclose(fpy);
fclose(fpx);
printf("\nPress any key to end \n");
getch();
}
/*----- END OF MAIN PROGRAM -----*/

```

ภาคผนวก ง

ผลงานวิจัยที่ได้รับการตีพิมพ์

ภาคผนวกนี้ เป็นผลงานที่ได้รับการตีพิมพ์ ลงในหนังสือการประชุมวิชาการทางวิศวกรรมไฟฟ้า

ครั้งที่ 18 พ.ศ. 2538



การรู้จำเสียงพูดคำไทยเฉพาะบุคคลด้วยนิวรัลเน็ตเวิร์ค

THAI WORD DEPENDENT SPEAKER SPEECH RECOGNITION BY NEURAL NETWORK

รศ.ดร.ชม กัมปาน*

Assoc.Prof.Dr. Chom Kimpan*

เสรี ปานซาง**

Mr. Seri Pansang**

บทคัดย่อ

บทความนี้นำเสนอวิธีการตรวจสอบข้อมูลเสียง ด้วยนิวรัลเน็ตเวิร์คชนิด Back Propagation เพื่อให้รู้จำเสียงคำพูดเฉพาะบุคคลโดยใช้ข้อมูลจากค่าเฉลี่ยของระดับพลังงานของแต่ละความถี่ในแกนเวลา จากสเปกโตรแกรมความถี่ฮาร์โมนิคส์เวลาพลังงาน (ของเสียงตัวอย่าง) ข้อมูลทุกชุดจะถูกแปลงให้มีลักษณะเป็น Uniform ก่อนเพื่อใช้เป็นกรอบมาตรฐานสำหรับเปรียบเทียบแบบแมชชีน แล้วนำไปใช้เทรนนิ่งบนเน็ตเวิร์คเพื่อคำนวณค่าเวกซ์ ที่เหมาะสมกับข้อมูลตัวอย่าง ที่สอดคล้องกับเอาท์พุตตามต้องการ เมื่อนำข้อมูลทดสอบป้อนให้กับอินพุตของเน็ตเวิร์ค แล้วใช้เวกซ์ที่คำนวณได้จากการเทรนนิ่ง มาอ้างกับข้อมูลทดสอบ จะทำให้เน็ตเวิร์คตรวจสอบและแยกแยะเสียงได้อย่างถูกต้อง เน็ตเวิร์ค จะพัฒนาความสามารถรู้จำทุกครั้งที่ผู้ให้พบบทพร้อมแล้วเทรนกลับให้ใหม่ จึงมีความยืดหยุ่นสูง เพราะสามารถปรับตัวให้เหมาะกับขอบเขตของงานที่ใช้

Abstract

This paper presents the method of voice data for Back Propagation Neural Network in order to recognize the voices produce. The data of this project was obtained from average energy power level of each frequency in time axis which was derived from Harmonic-time-energy spectrogram of the sampling voice. All data sets were transformed in to a particular uniform to be used as a standard profile for matching. Then, the transformed data were brought to trained on Neural Network in order to accumulate the value of weight. The accumulated weight was then used as a reference to the testing data. In consequence, the Neural Network was able to test and classified the voiced correctly. The network has

*รองศาสตราจารย์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์

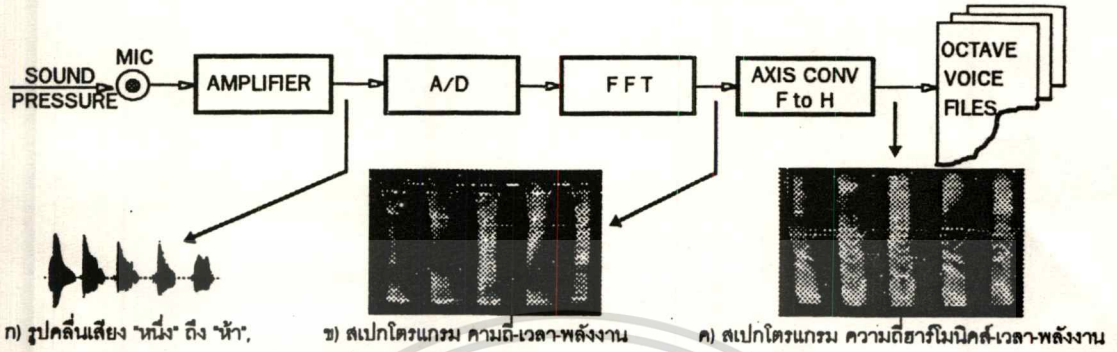
**นักศึกษาปริญญาโท คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

capacity to develop ability in measuring every time that the user find an error. In vice versar, the user could bring the error to be retrained. According to that the network contains high flexibility to which it could develop to a wide range of works.

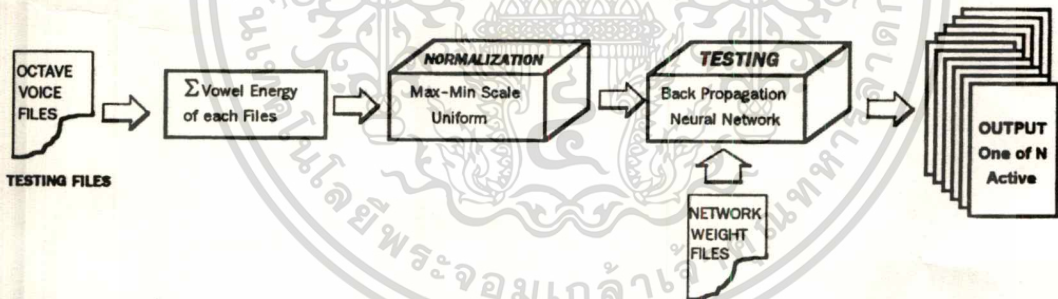
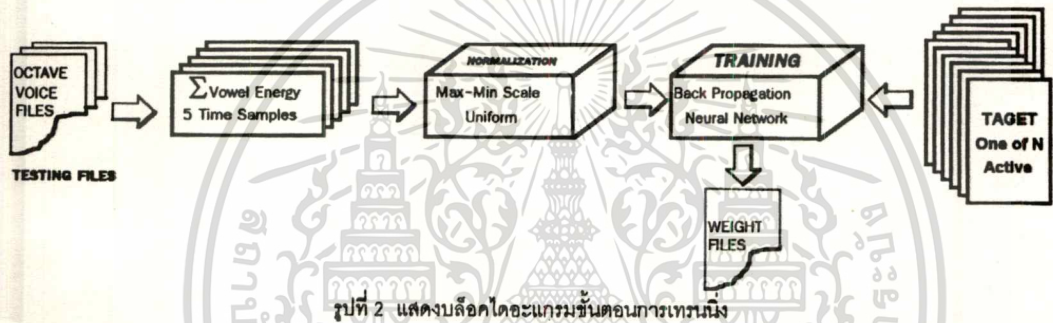
1. บทนำ

มนุษย์สามารถแยกคุณลักษณะของเสียงได้จากปริมาณความดังของเสียง (Volume), จำนวนของคลื่นต่อหน่วยเวลาที่ผ่านจุดอ้างอิง (Pitch) และลักษณะเสียงทางความถี่ (Tone) ซึ่งมีคุณสมบัติที่ซับซ้อนกว่า Volume และ Pitch การได้ยินเสียงของมนุษย์เกิดจากการแปลงการถ่ายทอดคลื่นสะท้อนจากอากาศผ่านช่องหู เข้าสู่อวัยวะรูปเปลือกหอยทากในหูส่วนใน (The snail-shaped cochlea) ที่มีช่องเหลวเป็นตัวกลาง สร้างคลื่นขึ้นบนเยื่อ ซึ่งมีอยู่ตลอดความยาวของ Cochlea ในเยื่อจะมีลักษณะเป็นเซลล์ขนเล็ก ๆ จำนวนนับพันแผ่อยู่ทั่ว เมื่อชนเหล่านี้ ถูกของเหลวทำให้เปลี่ยนลักษณะไป ทำให้เกิดเป็นสัญญาณไฟฟ้า(Electrical impulses)กระตุ้นส่งเป็นทอด ๆ ผ่านเส้นประสาทเสียง (Auditory nerve) เข้าสู่สมอง ความสามารถรู้จำ Pitch ของมนุษย์มีพื้นฐานจากเซลล์ที่ได้รับการกระตุ้นจากความถี่ต่ำที่ปรากฏที่ยอดของ Cochlea ในขณะที่ความถี่สูงจะกระตุ้นเซลล์ที่ตำแหน่งฐานของ Cochlea จากข้อความข้างต้น พอสรุปได้ว่าภายในCochlea จะมีเซลล์ขนเล็ก ๆ ที่มีคุณลักษณะในการรับรู้ความถี่ที่แตกต่างกัน วางตัวเรียงกัน ตั้งแต่ยอดถึงฐานของ Cochlea[6]

จากการศึกษาวิจัยที่รับรู้เสียงของมนุษย์ ทำให้ได้แนวทางการนำไปสร้างอุปกรณ์ที่สามารถแปลงสัญญาณการสะท้อนของเสียง เลียนแบบการทำงานของเซลล์ขนใน Cochlea ได้ และมีทฤษฎีนิวรัลเน็ตเวิร์ค[3][4][5]ที่เลียนแบบการทำงานของสมองไว้รองรับ ฉะนั้น ปัญหาการสร้างระบบการรู้จำเสียงน่าจะอยู่ที่การสรรหาคุณลักษณะสำคัญเฉพาะตัวของข้อมูลและวิธีการลำเลียงข้อมูลสู่นิวรัลเน็ตเวิร์ค



รูปที่ 1 แสดงขั้นตอนการปรับข้อมูลให้พร้อมนำไปเทรน หรือ ทดสอบ



2. ระบบการรู้จำเสียง

ระบบการรู้จำเสียงที่ทดลองสร้างขึ้น มีลักษณะดังรูปที่ 1, 2 และ 3 มีการทำงานเป็นแบบ Off-Line การทำงานของระบบแบ่งเป็น 2 ขั้นตอน คือ การเทรนนิ่ง และทดสอบ ดังนั้นจึงต้องเตรียมข้อมูล ไว้สองส่วนเช่นกัน การเตรียมข้อมูลขั้นต้นเป็นไปตามรูปที่ 1 คือแปลงสัญญาณเสียงตัวอย่างที่ได้จากไมโครโฟนให้เป็นสัญญาณดิจิทัล บันทึกเป็นไฟล์เสียง จากนั้นจึงแปลงข้อมูลให้อยู่ในรูปสเปกโตรแกรมความถี่-เวลา-พลังงาน ด้วยการแปลงแบบฟูเรียร์อย่างรวดเร็ว ดังรูปที่ 1ข. แล้วเปลี่ยนข้อมูลในแกนความถี่ให้เป็นข้อมูลในแนวแกนความถี่ฮาร์โมนิค⁽¹⁾ (รูปที่ 1ค. แสดงสเปกโตรแกรมที่แกนความถี่เปลี่ยนเป็นแกนความถี่ฮาร์โมนิค) การแปลงแกนความถี่ทำ เพื่อปรับสเปกโตรแกรมให้เหมาะสมสำหรับการวิเคราะห์เสียงเนื่องจาก

คุณลักษณะของเสียงที่สำคัญอยู่ในช่วงแรกของแถบความถี่และความสามารถในการรู้จำ pitch ของมนุษย์ สำหรับความถี่ต่ำจะดีกว่าความถี่สูง[6] ข้อมูลจากแถบความถี่ช่วงต่ำนี้จึงน่าสนใจนำมาใช้เป็นข้อมูลมาก และ 'pitch' เป็นความถี่พื้นฐานที่กำหนดระดับความสูงต่ำของเสียง เนื่องจากว่า ภาษาไทยเป็นภาษาที่เรียกว่า tonal language คือมีระดับโทนเสียง สามัญ เอก โท ตรี และจัตวา ซึ่งแต่ละโทนเสียงก็จะทำให้ ความหมายที่แตกต่างออกไปดังนั้น pitch จึงเป็นสิ่งสำคัญในการแยกแยะคำในภาษาไทย[12] เมื่อได้ข้อมูล เจึงความถี่ฮาร์โมนิค-เวลา-พลังงานแล้ว จะถูกเก็บสะสมไว้เป็นไฟล์ แล้วแบ่งส่วนเป็นข้อมูลมาตรฐาน เพื่อใช้เทรนนิ่ง และข้อมูลอีกส่วนใช้สำหรับทดสอบข้อมูลนี้ จะถูกนำมาผ่านขั้นตอนการ Normalization เสีย

เพื่อให้มีข้อมูลที่สอดคล้องกันสำหรับการฝึกสอน เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก่อนเพื่อลดขนาด, ความซับซ้อน และทำให้ลักษณะของข้อมูลเหมาะสมกับ Network ที่จะเทรนหรือทดสอบ หลังจากค่าผิดพลาดกำลังสองของเอพาทูที่ได้ออกเป็นเป้าหมาย(Taget)ที่ต้องการลดลงถึงจุดกำหนด, ค่า Weight ที่ได้จากการเทรนนิ่งจะถูกเก็บไว้เป็น Factor เพื่อนำมาคำนวณกับข้อมูลทดสอบ และให้ผลลัพธ์ที่ Output ที่สามารถแบ่งแยกชนิดของข้อมูลที่น่ามาทดสอบได้สอดคล้องกับข้อมูลทดสอบ ซึ่งเป็นตัว บอกถึงความสามารถจำของเน็ตเวิร์คได้

3.การแยกพยางค์ออกเพื่อพิจารณา

เนื่องจากการทดลองกำหนดเพื่อศึกษาวิธีการเตรียมข้อมูลและความสามารถในการจำของนิวรัลเน็ตเวิร์คเพื่อให้จำเสียงนับ "หนึ่งถึงสิบ" ซึ่งแต่ละพยางค์มีลักษณะเป็นคำโดดและมีช่วงว่างระหว่างคำเด่นชัด จึงสามารถแยกพยางค์ออกได้ไม่ยากนัก โดยสมการเป็นดังนี้

$$Eav(t) = \frac{\sum_{f=1}^M E_f}{M}$$

| $t = 1 \text{ to Maximum time Step, } \dots\dots\dots(1)$

และ

$Tref = K(Eav_{max} - Eav_{min}) + Eav_{min}$ (2)

ข้อมูลแต่ละพยางค์จะถูกแยกออกมาเมื่อ

$Tref < Eav(t)$

เมื่อ $Eav(t)$ คือ ค่าพลังงานเฉลี่ยตามแนวแกนความถี่ฮาร์โมนิคของช่วงเวลาStepที่ t

E_f คือค่าระดับพลังงานของแต่ละความถี่ตามแนวแกนความถี่ ณ.ความถี่ใดๆ ที่ช่วงเวลา t

f คือ ความถี่ที่กำลังสนใจในแกนความถี่ฮาร์โมนิค โดยเริ่มตั้งแต่ 1 ถึง 128

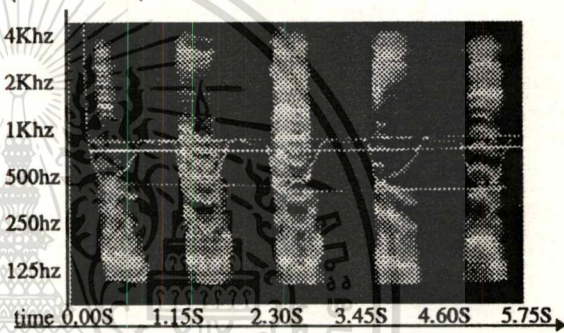
M คือ จำนวนช่วงความถี่ (Step) ที่แบ่งในแกนความถี่ฮาร์โมนิค(การทดลองให้ M=128)

t คือ เวลาแต่ละช่วง ตั้งแต่ 1 ถึงช่วงเวลาสิ้นสุดการแปลงข้อมูลด้วย FFT ความกว้างของผลการเปลี่ยนข้อมูลจากโดเมนของเวลามาเป็นความถี่ของชุดข้อมูลแต่ละชุด เมื่อนำมาเรียงกันจะเป็นความกว้างของสเปกโตรแกรมสามมิติ โดยความกว้างจะแทนเวลาที่ผ่านไป

Tref คือค่าเทรชโฮลที่กำหนดขึ้นเพื่อแยกช่วงพยางค์ออกไปพิจารณา

K เป็นเปอร์เซนต์ระหว่างค่าระดับพลังงานเฉลี่ยสูงสุด (Eav_{max}) และต่ำสุด (Eav_{min}) จากการทดลองค่าที่ใช้คือ 0.20 (20%)

จะสังเกตผลจากตัวอย่างสเปกโตรแกรมได้ในรูปที่ 1 ข, ค และภาพขยายในรูปที่ 4 โดยเส้นขวางบนแทนค่าเฉลี่ยต่ำสุด,เส้นขวางล่างแทนค่าเฉลี่ยสูงสุดซึ่งปรับตัวตามค่าที่สูงที่สุด, เส้นกลางคือค่า เทรชโฮล จะตัดกับกราฟพลังงานเฉลี่ยจุดตัดกราฟคือจุดพิจารณาการแยกพยางค์



รูปที่ 4 แสดงการแยกพยางค์ ออกจากสเปกโตรแกรม ความถี่ฮาร์โมนิค-เวลา-พลังงาน ของเสียงนับ "หนึ่ง" ถึง "ห้า"

4.การปรับข้อมูลให้เหมาะสมกับเน็ตเวิร์ค

เสียง เสียงเดียวกันที่เปล่งออกมาจากผู้พูดคนเดียว มักมีคุณสมบัติไม่เหมือนกันเสียทีเดียว ลักษณะการออกเสียงสามารถบ่งบอกถึงอารมณ์ของผู้พูดได้ ความยาวของการออกเสียงของคำคำเดียวกันไม่เท่ากัน แต่จากการทดลอง พบว่าใกล้เคียงกัน และแตกต่างกันกรณีต่างผู้พูดกัน เสียงพูดคำเดียวกันของผู้พูดแต่ละคนจะมีเอกลักษณ์เฉพาะบุคคลและมีกราฟของพลังงานเฉลี่ยแตกต่างกันโดยสิ้นเชิง แต่ด้วยวิเคราะห์ข้อมูลลักษณะอื่นพบว่ามีความร่วมเหมือนปะปนอยู่ด้วยบ้าง ในรายงาน ฉบับนี้ จะนำเสนอการตรวจสอบเสียงจากเอกลักษณ์ของคำ จากรูปของพลังงานเฉลี่ยในแกนเวลาของความถี่ฮาร์โมนิคของข้อมูลเสียงมาเป็นข้อกำหนดความแตกต่าง ลักษณะการรู้จำแบบนี้ เป็นแบบแผนซึ่งมีหลายวิธีที่สามารถใช้ปรับระยะเวลาการแปลงเสียง ของข้อมูลของเสียงใด ๆ ให้มีช่วงเวลาเท่ากันเป็นกรอบมาตรฐาน เช่น ปรับเวลาใด ๆ ให้เป็นช่วงเวลามาตรฐานค่าหนึ่ง แล้วคำนวณค่าระดับพลังงานในแกนความถี่ใหม่ ที่สอดคล้องกับข้อมูลเดิม อาจใช้ Least Square หรือวิธีอื่น ๆ แต่วิธีที่เลือกทำการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใช้วิธีหาค่าพลังงานเฉลี่ยของแต่ละความถี่ในแกนความถี่ฮาร์โมนิกเฉพาะในช่วงเวลาที่เปลี่ยนแปลง จะได้กราฟแสดงระดับพลังงานเฉลี่ยของแต่ละความถี่ทุกความถี่ของแกนความถี่ฮาร์โมนิก และไม่ขึ้นกับระยะเวลาการเปลี่ยนแปลง โดยใช้สมการที่ (3)

$$E_{av}(f) = \frac{\sum_{t=0}^N E_t}{N} \quad | \quad f = 1 \text{ to } M \quad (M=128) \dots\dots(3)$$

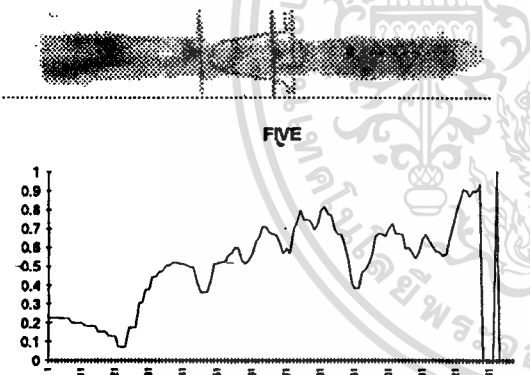
เมื่อ $E_{av}(f)$ คือ ค่าระดับพลังงานเฉลี่ยของความถี่ฮาร์โมนิกที่ f ของช่วงเวลาเปลี่ยนแปลงที่กำลังสนใจ

f คือ ช่วงความถี่ที่สนใจในแกนความถี่ฮาร์โมนิก โดยเริ่มตั้งแต่ 1 ถึง 128

E_t คือ ค่าระดับพลังงาน ณ เวลาใด ๆ ของช่วงที่เปลี่ยนแปลงของแกนความถี่ฮาร์โมนิกที่ f

M คือ จำนวนช่วงความถี่ (Step) ที่แบ่งในแกนความถี่ฮาร์โมนิก (การทดลองให้ $M=128$)

N คือจำนวนช่วงเวลาที่เปลี่ยนแปลงของค่าหนึ่งค่า



รูปที่ 5 แสดงกราฟระดับค่าพลังงานเฉลี่ยจากสเปกโตรแกรมเสียงห้าที่หาได้จากสมการที่ 3

จากสมการที่ 3. จะพบว่า N ที่ไปหาร ทำให้ได้ค่าเฉลี่ยเพียงค่าเดียวสำหรับ 1 ช่วงความถี่ ในแกนความถี่ฮาร์โมนิก จะได้ข้อมูลเพียง M ข้อมูล สำหรับเสียง 1 เสียง หรือค่า 1 ค่า,จากนี้จะได้คุณลักษณะพลังงานเฉลี่ยของข้อมูล 1 พยางค์ แต่พบว่าข้อมูลค่าสูงสุด-ต่ำสุด มีความแตกต่างกันมากหากนำข้อมูลไปเทรนทันที จะทำให้เน็ตเวิร์กใช้เวลาเทรนนิ่งมากเกินไป หรืออาจเกิดการเรียนรู้ไม่สู่สู่การแก้ปัญหาการรู้จำได้ จึงต้องตัดข้อมูลบางส่วนที่ไม่ใช่ออกไป โดยใช้วิธี Max-Min Scaling ทำให้ข้อมูลอยู่ใน Range พอดี เป็นUniform โดยใช้สมการที่ 4 ดังนี้.

$$X_s = \left[\frac{X_d - d_{min}}{d_{max} - d_{min}} \right] * [SCALE_{max} - SCALE_{min}] + SCALE_{min} \dots\dots\dots(4)$$

โดย X_s คือ ค่าที่ได้จากการทำ max-min scaled ของแต่ละจุดข้อมูล

X_d คือ ตัวข้อมูลดิบที่ต้องการนำมาคำนวณ max-min scaled

d_{max} คือ ค่าสูงสุดของชุดข้อมูลดิบ

d_{min} คือ ค่าต่ำสุดของชุดข้อมูลดิบ

$SCALE_{max}, SCALE_{min}$ เป็นค่าที่ใช้กำหนดช่วงสเกลที่ต้องการ Transfer ไปของ Max-Min Scaling เพื่อปรับข้อมูลดิบให้ได้ output อยู่ในช่วงที่ต้องการ ไปยัง สเกลสูงสุด,ต่ำสุด ในกรณีที่กำหนดสเกลสูงสุด = 1, ค่าต่ำสุด = 0 จะได้

$$X_s = \frac{X_d - d_{min}}{d_{max} - d_{min}} \dots\dots\dots(5)$$

หลังจากทำข้อมูลให้ลดลงและเป็น Unit value แล้วก็ได้ข้อมูลของค่าแต่ละค่าที่พร้อมนำไปเข้ากระบวนการเทรนนิ่งและทดสอบกับเน็ตเวิร์ก

5. อุปกรณ์และวิธีการทดลอง

วัตถุประสงค์การทดลองต้องการฝึกฝนและทดสอบความสามารถรู้จำเสียงเป็นคำไทยนับ "หนึ่ง" ถึง "สิบ" ของเน็ตเวิร์กที่มีความสามารถรู้จำเสียงเฉพาะบุคคลเพียงใด อุปกรณ์ที่ใช้ทดลองคือเครื่องไมโครคอมพิวเตอร์ขนาดประมวลผล 486DX4-100 ที่มีการ์ดเสียงรุ่น Sound Blaster 16, Neurolink library บนภาษาซี ของบริษัท NeuroDynamX และ Microsoft Excel ช่วยวิเคราะห์และนำเสนอข้อมูลที่ได้จากการทดลอง การทดลองเตรียมข้อมูลขั้นต้นใช้วิธีการของคุณธันวา[1] เริ่มจากการบันทึกข้อมูลเสียงตัวอย่าง โดยใช้ฮาร์ดดิสก์ 10KHz ขนาดข้อมูล 8 บิต ผ่านทางการ์ดเสียงเก็บเป็นไฟล์ตัวอย่างเสียงแล้วนำแต่ละไฟล์มาคำนวณเปลี่ยนข้อมูลจากเชิงเวลาให้เป็นข้อมูลเชิงความถี่ โดยตัดแบ่งข้อมูลเป็นชุดๆ ชุดละ 1024 ข้อมูลในลักษณะซ้อนทับกันโดยกำหนดระยะเวลาเลื่อนไว้ที่ 256 ข้อมูล เพื่อนำไปคำนวณด้วยวิธีการแปลงฟูเรียร์อย่างรวดเร็ว(Fast Fourier Transform) การคำนวณจะทำทีละชุดจนหมดข้อมูลจะได้สเปกโตรแกรมความถี่-เวลาพลังงาน เนื่องจากกำหนดขนาดของหน้าต่างเท่ากับ 1024 จุดข้อมูลจึง

ได้outputในแกนความถี่256ค่าแทนช่วงความถี่ตั้งแต่ 0 ถึง 4 KHz แล้วปรับเปลี่ยนสเกลแกนความถี่เดิมให้เป็นสเกลแกนความถี่ฮาร์โมนิค เพื่อให้สอดคล้องกับลักษณะการได้ยินของมนุษย์ การปรับเปลี่ยนสเกลดังกล่าวทำให้สามารถลดจำนวนข้อมูลให้เหลือข้อมูลเอาต์พุตเพียง 128 ข้อมูลต่อชุดข้อมูลอินพุต ซึ่งจะนำไปกำหนดและป้อนให้กับInput layer ของเน็ตเวิร์ก ก่อนจะป้อนให้กับเน็ตเวิร์ก จะต้องนำข้อมูลไปเลือกช่วงพิจารณาและปรับข้อมูลตามหัวข้อที่ 3และ4 เสียก่อน ขั้นตอนการดำเนินการเป็นดังนี้คือ

5.1 บันทึกเสียงนับ'หนึ่ง'ถึง'สิบ' จากผู้พูดคนเดียว 5 ตัวอย่าง เพื่อใช้ในขั้นตอนการเทรนเน็ตเวิร์ก และ 'หนึ่ง'ถึง'ห้า' กับ 'หก'ถึง'สิบ' อย่างละ 3 ตัวอย่าง เพื่อใช้ในขั้นตอนการทดสอบ การบันทึกข้อมูลทดสอบแต่ละครั้ง จะเว้นระยะห่างกันไม่ต่ำกว่า 3วินาที เพื่อเลี่ยงการจำลักษณะการออกเสียงของผู้พูดเมื่อออกเสียงต่อเนื่องซ้ำกัน

5.2 ทรานส์ฟอร์มข้อมูลเสียงไปเป็นข้อมูลสเปกโตรแกรม ความถี่ฮาร์โมนิค-เวลา-พลังงาน แล้วแยกพยางค์ออก

5.3 นำทุกพยางค์ไปผ่านขั้นตอนการในหัวข้อที่ 4 จะได้ชุดข้อมูลที่พร้อมจะใช้ ฝึกฝนและทดสอบ

5.4 จัดกลุ่มเสียงพยางค์เดียวกันเข้าด้วยกันจะได้ 10 กลุ่มเสียง กลุ่มละ 5ข้อมูลเสียง แล้วหาค่าเฉลี่ยของแต่ละกลุ่มเสียงทั้ง5 รวมกับข้อมูลกลุ่มเสียงเดิม เป็น6ข้อมูลเสียง 10กลุ่มเสียง จะได้ทั้งหมด 60ชุดข้อมูล

5.5จัดข้อมูลเพื่อการเทรนนิ่งโดยกำหนดTarget ในลักษณะ One-of-N active กำหนดค่า paramater ให้กับ Back Propagation Neural Network: BPNN, ดังนี้

input layer=128 input, output layer=10 output คือ 'หนึ่ง' ถึง 'สิบ'ในลักษณะ One of N active, จำนวนชั้นของเน็ตเวิร์ก =2Layerโดยกำหนด Hidden Layer เป็น Layer0 จำนวน=128 neural, Output Layer เป็น Layer1จำนวน=10neural, ในขั้นตอนการเทรนนิ่งใช้อัตราการเรียนรู้= 0.100, Transfer Function min=0.000, Max: 1.000, Gain:1.000 (Transfer Function ที่ใช้เป็นแบบNonlinear functionปกติใช้ Sigmoid Functiongสำหรับปรับปรุงผลรวม(NET)ของนิวรัล แล้วปรับให้สามารถกำหนด Rangeของเอาต์พุตได้ก่อนส่งออกเป็น output (o(x)) ,Transfer Function มีสมการดังนี้

$$O(x) = \frac{1}{1 + e^{-Gx}}$$

และ

$$f(x) = \text{Sigmoid}(x) = \min + (\max - \min) \cdot o(x)$$

โดย G เป็น Gain ของ Transfer Function, minเป็น Transfer Function Minimum, maxเป็นTransfer Function Maximum,

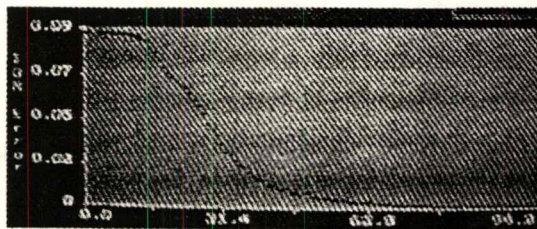
e เป็นฐานของ natural log และ x เป็นผลรวมของ Weight input) จากนั้นจึงเริ่มขั้นตอนการเทรนเน็ตเวิร์ก ขั้นตอนจะสิ้นสุดเมื่อค่า Square error ลดลงจนได้ค่าที่น่าพอใจ ค่าเวจท์จะถูกบันทึกไว้ใช้ในขั้นตอนการทดสอบ

5.6 เข้าสู่ขั้นตอนการทดสอบ โดยนำข้อมูลทดสอบที่หาได้จากข้อ 5.3 มาป้อนแทนข้อมูลที่ใช้เทรนนิ่ง ที่ Input ของเน็ตเวิร์ก ดังรายละเอียดในรูปที่ 3 และป้อนค่าเวจท์ที่หาได้จากการเทรนนิ่ง กลับสู่เน็ตเวิร์ก จะทำให้ได้ Output ลักษณะแบบเดียวกับ Target ในขั้นตอน เทรนนิ่ง, ระดับของ Output จะเปลี่ยนแปลงไปตามข้อมูลทดสอบที่คลาดเคลื่อนไปจากข้อมูลที่ใช้เทรน หากผลลัพธ์ผิดพลาดสามารถนำเฉพาะข้อมูลที่ผิดพลาดกลับไปเทรนร่วมกับข้อมูลเก่า จะทำให้เน็ตเวิร์กจำลักษณะข้อมูลที่ไม่เคยพบมากขึ้น

6. ผลการทดลองและสรุป

ผลการทดลองแบ่งเป็นสองส่วนคือ ผลการเทรนนิ่ง และผลการทดสอบ

6.1 ผลการเทรนนิ่ง การเทรนนิ่งของ BPNN ที่เวลาประมาณ 3ชั่วโมง ค่า Training average square error ลดลงเป็น 2.62×10^{-4} ค่า Maximum Training average square error = 8.464×10^{-3} , minimum=0, ที่รอบการเทรนนิ่ง 1570 รอบของข้อมูลจำนวน 60ชุด การลดลงของค่า Training average square error แสดงกราฟที่ 1



รูปกราฟที่ 1 แสดงค่า Training average square error เมื่อการ เทรนนิ่งผ่านไป หลังจากเสร็จสิ้นขั้นตอนการเทรนนิ่ง แล้วทดลองทดสอบด้วย ข้อมูลเดิมที่ใช้เทรนนิ่ง สามารถแยกแยะได้ถูกต้องทั้งหมดมีความผิดพลาดอยู่ในช่วงตั้งแต่ 0 ถึง 0.8% โดยมีค่าเฉลี่ยอยู่ที่ 0.0262% ถือได้ว่าเข้าเป้าหมายเชื่อถือ แต่เมื่อทดสอบด้วยข้อมูล

เสียงเดิมที่เน็ตเวิร์คไม่รู้จัก ค่าความผิดพลาดจะเพิ่มขึ้นตาม ลักษณะการออกเสียงที่แตกต่างไปอาจ ถึงขั้นแยกลักษณะเสียง ไม่ออก ถ้าคุณภาพเสียงที่เปล่งออกมาผิดไปมาก

6.2 ผลการทดสอบ เมื่อนำข้อมูลเสียงนับ “หนึ่ง”ถึง “ห้า” และ “หก”ถึง “สิบ”อย่างละ 3 ชุด การแยกแยะเสียงใช้วิธี การตรวจเลือกค่า Maximum Magnitude ของเอ๊าท์พุท ทั้งสิบของ เน็ตเวิร์ค พบว่าเน็ตเวิร์คมีความสามารถแยกแต่ละเสียงถูกต้องทั้ง หมด 4ชุด, แยกเสียงผิดชุดละ 1 เสียง 2ชุด คือเสียง “เจ็ด” และ “สิบ” โดยเสียง “เจ็ด”เน็ตเวิร์ค ให้ Maximum Magnitude เป็น สิบ และ เสียง “สิบ”เน็ตเวิร์ค ให้ Maximum Magnitude เป็น สี่ และเมื่อทดสอบนับ “หนึ่ง” ถึง “แปด”ให้เร็วขึ้นประมาณ 1.5 เท่า เน็ตเวิร์คก็ยังแยกได้ถูกต้องถึง 87.5 เปอร์เซ็นต์ ดังตารางที่ 1

INOUT	OUT:1	OUT:2	OUT:3	OUT:4	OUT:5	OUT:6	OUT:7	OUT:8	OUT:9	OUT:10
หนึ่ง	0.507	0	0.006	0.107	0.003	0	0	0	0	0.005
สอง	0	0.995	0	0.14	0	0.002	0	0.001	0.004	0
สาม	0	0	0.955	0	0.05	0	0	0.293	0.01	0.447
สี่	0.002	0	0	0.752	0	0	0.001	0.001	0	0.055
ห้า	0	0	0.036	0	0.817	0	0	0.202	0.056	0.259
หก	0	0.247	0	0.001	0	0.705	0.007	0.002	0.001	0
เจ็ด	0	0	0	0.017	0.001	0	0.011	0.043	0	0.118
แปด	0	0	0.078	0	0.003	0	0	0.995	0	0.759

ตารางที่ 1 แสดงผลการตรวจสอบเสียงที่เปล่งเร็วกว่าปกติ 1.5 เท่า ความผิดพลาดของการตรวจสอบแยกเสียง มิได้ เกิดจากความผิดพลาดของเน็ตเวิร์ค หากแต่เกิดจากความ แตกต่างของเสียงทดสอบที่ต่างไปจากข้อมูลที่ใช้เทรน การแก้ ปัญหาทำได้สองวิธีคือ 1. ผู้ออกเสียงพยายามออกเสียงให้ เหมือนเดิมและ 2. ทำการรีเทรนนี้ โดยใช้เฉพาะข้อมูลส่วนที่ ผิดพลาดไปผ่านขบวนการหลังจากเสร็จสิ้นการรีเทรนนี้ เน็ตเวิร์คจะรู้จักเสียงที่เคยรู้จำผิดพลาด กล่าวได้ว่า “นิวรัลเน็ตเวิร์คเป็นเน็ตเวิร์คที่มีความสามารถพัฒนา ความสามารถจำได้โดยการรีเทรนนี้” เวลาในการประมวล ยังคงเท่าเดิม แม้จะมีการรีเทรนนี้หลาย ๆ ครั้ง เพราะ จำนวนเวกที่ที่ใช้ยังคงเท่าเดิมคือเท่ากับจำนวนอินพุท เน็ตเวิร์คซึ่งมีจำนวนคงที่ตามที่กำหนด ข้อมูลเวกที่เท่านี้ที่ ถูกปรับเปลี่ยนไป ข้อที่ระวังคือความรู้จำอาจสับสนและล่ม สลายโดยสิ้นเชิง หากข้อมูลที่น่าไปรีเทรนนี้ ถูกกำหนด Taget ผิดไป, ยังมีข้อมูลอื่น ๆอีกที่แฝงเร้น อยู่ในเสียงที่ เปล่งออกมา เช่น อารมณ์, ความรู้สึก, ความแตกต่างเล็ก ๆ น้อย ๆ, ลักษณะคำ, การแยกแยะคำพยางค์เสียง ยังต้อง ใช้วิธีการพิเศษโดยเฉพาะเพื่อดึงลักษณะเด่นออกจากชุด

รูปภาพที่ 3 แสดงค่าพลังงานเฉลี่ยจากสเปกโตรแกรมเสียง “สาม”

ข้อมูล ให้กับนิวรัลเน็ตเวิร์ค ถ้าหากสามารถหาลักษณะ เ่นของข้อมูลวิธีอื่น ๆ มาพิจารณา ร่วมกับวิธีการที่นำเสนอ จะทำให้อัตราการแยกดีขึ้น แต่อาจมีผลต่อความยืดหยุ่นของ ระบบ ซึ่งเป็นทางเลือกตามวัตถุประสงค์ในการทำวิจัยต่อไป

เอกสารอ้างอิง

- [1.]ธันวา ศรีประโมง. การวิเคราะห์เสียงพูดภาษาไทยใน แกนความถี่ฮาร์โมนิค วิทยานิพนธ์ปริญญาโทบริหาร การศึกษาเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- [2.] ผศ.ดร.อภิชาติ ตั้งทางธรรม. การเปลี่ยนความเร็วของ เสียงพูด การประชุมวิชาการทางวิศวกรรมไฟฟ้าครั้งที่17, 2537
- [3.] เศรษฐพล ลีนปราชญา, กิตติ ไพฑูรย์วัฒนกิจ การใช นิวรัลเน็ตเวิร์คเพื่อการจดจำแพตเทิร์น.,45-59. J.Natl. Res. Council Thailand. 1993. 25 (1)
- [4.] R.P. Lippmann, 'An introduction to computing with neural networks', IEEE ASSP Mag., vol4, no 2. Apr 1987.
- [5.] James A. Freeman, David M. Skapura, 'Neural Network Algorithms, Applications, and Programming Techniques', Addison-Wesley Publishing Company, Inc. 1991.
- [6.] GROlier ELECTRONIC PUBLISHING, INC. , "EAR", 1995 Multimedia Encyclopedia.
- [7.] Dynamind Developer User's Guide version 4.0.1, NeuroDynamix, Inc. POBox 323 Boulder, Colorado 80306.

INOUT	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8	OUT9	OUT10
หนึ่ง	0.992	0.022	0.028	0.006	0	0.012	0	0.005	0	0
สอง	0.009	0.982	0	0.021	0	0.259	0	0.001	0.097	0
สาม	0.574	0.061	0.999	0.006	0.003	0	0	0.016	0.045	0.001
สี่	0.783	0.007	0.001	0.958	0	0.001	0	0	0	0
ห้า	0.002	0	0.063	0	0.509	0	0	0.005	0.834	0

ตารางที่ 2 แสดงผลการตรวจสอบเสียง “หนึ่ง”ถึง “ห้า” แยกถูกต้องทั้งหมด

INOUT	OUT1	OUT2	OUT3	OUT4	OUT5	OUT6	OUT7	OUT8	OUT9	OUT10
หนึ่ง	0	0	0	0.004	0.003	0.713	0.22	0	0.587	0
สอง	0	0	0	0.001	0.004	0	0.514	0.039	0	0.108
สาม	0	0	0.003	0	0.008	0	0	0.876	0.005	0.252
สี่	0	0.015	0.035	0	0.027	0	0	0.03	0.537	0
ห้า	0	0	0	0.005	0.006	0	0.123	0.016	0	0.904

ตารางที่ 3 แสดงผลการตรวจสอบเสียง “หก”ถึง “สิบ” แยกถูกต้องทั้งหมด

