การพัฒนาระบบการรู้จำเสียง

Speech Recognition System Development

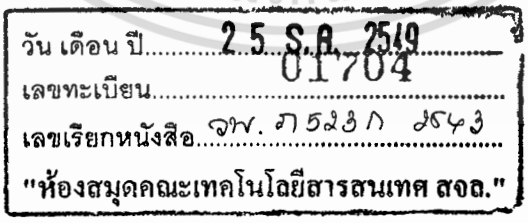
โดย

นางสาวภิญญ กำเนิดหล่ม

รหัส 41067191

อาจารย์ที่ปรึกษา

ดร. อาริต ธรรมโน



รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการศึกษากรณีพิเศษ

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ภาคเรียนที่ 1 ปีการศึกษา 2543

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบการรู้จำเสียง
นักศึกษา	นางสาวภิญญู กำเนิดหล่ม
อาจารย์ที่ปรึกษา	ดร.อาริต ธรรมโน
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2543

บทคัดย่อ

ระบบการรู้จำเสียง (Speech Recognition System) เป็นระบบที่พัฒนาขึ้นมาเพื่อเพิ่มประสิทธิภาพของคอมพิวเตอร์ในการรู้จำเสียงให้สูงขึ้น โดยระบบการรู้จำเสียงใช้เทคนิคของโครงข่ายประสาทเทียม (Artificial Neural Network) ซึ่งเป็นสถาปัตยกรรมหนึ่งที่มีโครงสร้างเลียนแบบโครงสร้างสมองมนุษย์และรูปแบบของโครงข่ายประสาทเทียมที่จะนำมาใช้ในการรู้จำ คือ Adaptive Resonance Theory (ART1) จากการพัฒนาระบบดังกล่าว คาดว่าจะได้โปรแกรมโครงข่ายประสาทเทียมชนิด ART1 เพื่อนำไปใช้ในการรู้จำเสียงมนุษย์ ทั้งนี้ เพื่อนำไปประยุกต์ใช้ในองค์กรต่อไป.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Speech Recognition System Development
Student	Miss Phinyoo Kamnerdlom
Advisor	Arit Thammano, Ph.D.
Level of Study	Mater of Science in Information Technology
Major	Information Science
Academic Year	2000

ABSTRACT

Speech Recognition System will be developed in order to increase computers' efficiency in recognizing human voice. This system employs Artificial Neural Network Technologies – one of the architectures which emulates human brain structure. An Artificial Neural Network used in this research in order to recognize speech will be Adaptive Resonance Theory (ART1). Elaborating this Speech Recognition System, will yield an artificial neural network program – type ART1. This program will be expected to have ability to recognize human voice and to adopt in any relevant organization.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบพระคุณ ดร.อาริต ธรรมโน อาจารย์ที่ปรึกษาที่กรุณาให้คำปรึกษา แนะนำ และให้ความช่วยเหลือด้วยดีตลอดมา ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ให้ความรู้และคำแนะนำในการศึกษา ขอขอบพระคุณกรรมสรรพากรที่ได้ให้โอกาสในการศึกษาในหลักสูตรนี้แก่ข้าพเจ้า ขอขอบคุณหัวหน้างานเพื่อนร่วมงานและเพื่อนร่วมรุ่นทุกคนที่ได้ให้ความช่วยเหลือและให้กำลังใจ ขอขอบคุณคุณสามารถ เหลี่ยวเจริญ ที่ได้ให้คำปรึกษาและคำแนะนำตั้งแต่ก่อนเข้ารับการศึกษาจนจบโครงงานนี้ และขอขอบคุณคุณชัย วุฒิวิวัฒน์ชัย ที่ให้คำปรึกษาและคำแนะนำในโครงงานนี้

สุดท้ายขอกราบขอบพระคุณบิดา มารดา พี่ชายและพี่สาวที่ได้เลี้ยงดูอบรมสั่งสอนคอยดูแลให้การสนับสนุนและให้กำลังใจข้าพเจ้าด้วยความรักและความเข้าใจตลอดมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป	VII
บทที่	
1. บทนำ	1
1.1 ความเป็นมา	1
1.2 วัตถุประสงค์	2
1.3 เป้าหมายและขอบเขต	2
1.4 ขั้นตอนและวิธีการดำเนินการ	2
1.5 เครื่องมือที่ใช้ในการทดลอง	3
1.6 ตารางเวลาในการพัฒนาระบบ	3
1.7 ประโยชน์ที่คาดว่าจะได้รับ	4
2. ทฤษฎีที่เกี่ยวข้อง	5
2.1 ระบบการรู้จำเสียงของมนุษย์	5
2.2 ระบบการส่งเสียงของมนุษย์	6
2.3 การรู้จำเสียง (Speech Recognition)	9
2.4 ทฤษฎีเซลล์ประสาททางชีววิทยาเบื้องต้น (Biological Neurons)	10
2.5 โครงข่ายประสาทเทียม (Neural Network)	12
3. Adaptive Resonance Theory 1 (ART1)	21
4. การทดลองและผลการทดลองในการทำวิจัย	32
4.1 ขั้นตอนการจัดเตรียมข้อมูล	29
4.2 ขั้นตอนการเรียนรู้ (ฝึกสอน)	45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

4.3 ขั้นตอนการทดสอบ 47
5. สรุปผลการวิจัยและข้อเสนอแนะ 50
5.1 ผลการทดลอง 50
5.2 สรุปผลการทดลอง 50
5.3 ข้อเสนอแนะ 51
บรรณานุกรม 52
ประวัติผู้เขียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

1.1 ตารางเวลาในการพัฒนาระบบ 6

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่

2.1	ระบบส่งเสียงของมนุษย์	6
2.2	ระบบการรู้จำเสียง	9
2.3	โครงสร้างเซลประสาทของมนุษย์	11
2.4	แบบจำลองทางตรรกของโครงข่ายประสาทเทียม	12
2.5	แสดงลักษณะของ Binary Sigmoid	14
2.6	แสดงลักษณะของ Bipolar Sigmoid	14
2.7	แสดงไดอะแกรมโครงข่ายประสาทเทียมพื้นฐาน	15
2.8	โครงข่ายประสาทเทียมชนิดเลเยอร์เดียว	16
2.9	โครงข่ายประสาทเทียมชนิดมัลติเลเยอร์	17
2.10	โครงข่ายประสาทเทียมชนิดคอมเพททิทิฟว์	17
3.1	ระบบการรู้จำเสียงของสถาปัตยกรรม ART1	22
3.2	ตัวอย่าง Pattern matching ของ ART1 โดยใช้ค่าวัดความคล้ายกันเท่ากับ 0.6	26
4.1	ตัวอย่างแฟ้มข้อมูลเสียงกลุ่มตัวอย่าง	32
4.2	แสดงลักษณะของเสียงที่ถูกกำหนดจุดเริ่มต้นและสิ้นสุดแล้ว	33
4.3	แสดงสัญญาณเสียงที่ปรับขนาดความยาวเท่ากับ 100%	33
4.4	ลักษณะของเสียที่ทำการวิเคราะห์ความถี่	34
4.5	แสดง Slope ของเสียงพูด	35
4.6	แสดง Slope ของเสียงพูด	36
4.7	แสดง Slope ของเสียงพูด	37
4.8	แสดง Slope ของเสียงพูด	38
4.9	แสดง Slope ของเสียงพูด	39
4.10	แสดง Slope ของเสียงพูด	40
4.11	แสดง Slope ของเสียงพูด	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

รูปที่

4.12	แสดง Slope ของเสียงพูด	42
4.13	แสดง Slope ของเสียงพูด	43
4.14	แสดง Slope ของเสียงพูด	44

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมา

ในการติดต่อสื่อสารของมนุษย์ เสียงพูดถือได้ว่าเป็นพื้นฐานของการติดต่อสื่อสารระหว่างกันซึ่งคนแต่ละคนจะมีเสียงพูดที่เป็นเอกลักษณ์ซึ่งสามารถระบุได้ว่าเสียงพูดนี้เป็นเสียงของใคร และในปัจจุบันได้มีการวิจัยเพื่อเพิ่มประสิทธิภาพให้คอมพิวเตอร์รู้จำเสียงเลียนแบบมนุษย์ เพื่อที่จะนำเอาเทคโนโลยีในการรู้จำเสียงมาประยุกต์ใช้ในงานด้านต่าง ๆ เช่น ในด้านระบบสื่อสารโทรคมนาคม, การป้อนข้อมูลด้วยเสียง, การรักษาความปลอดภัยของข้อมูล, การควบคุมอุปกรณ์ไฟฟ้าด้วยเสียง, เครื่องแปลภาษา, เครื่องให้บริหารให้ข้อมูลโดยใช้เสียง, เครื่องแปลงเสียงเป็นหนังสือ หรือแม้แต่การบริการของระบบธนาคาร เป็นต้น โดยมีจุดประสงค์หลักเพื่อให้คอมพิวเตอร์สามารถสื่อสารกับมนุษย์โดยใช้เสียงได้ ในระบบการรู้จำเสียงนั้น จะต้องมีการตรวจสอบว่าเสียงที่รับเข้ามาในระบบนั้นเป็นเสียงคำใด โดยเมื่อได้รับสัญญาณเสียงมาแล้วจะทำการแยกคุณลักษณะของเสียงนั้น แล้วนำมาเปรียบเทียบข้อมูลที่ได้เก็บไว้เป็นข้อมูลอ้างอิงในฐานข้อมูลที่ได้จากขั้นตอนการฝึกฝนก่อนหน้านั้น เมื่อได้เปรียบเทียบว่าสัญญาณเสียงที่เข้ามาตรงหรือใกล้เคียงกันหรือไม่แล้ว หลังจากนั้นจะทำการตัดสินใจว่าเป็นคำใด โดยทั่วไปการรู้จำเสียงสามารถแบ่งแยกชนิดการรู้จำได้เป็น 3 แบบ คือ การแบ่งตามการขึ้นอยู่กับผู้พูด, แบ่งตามระดับของเสียงที่รู้จำ และแบ่งตามหน่วยของเสียงที่ใช้เป็นตัวทาบเทียบ ซึ่งโครงงานนี้ใช้ชนิดการแบ่งตามระดับของเสียงที่รู้จำโดยเป็นการรู้จำเสียงคำเดียว

ปัจจุบันได้มีการวิจัยการรู้จำเสียงอย่างแพร่หลาย โดยมีรูปแบบและวิธีการในการรู้จำหลายวิธี เช่น Dynamic Time Warping (DTW), Hidden Markov Model(HMM), และวิธีหนึ่งที่นิยมใช้ คือ การรู้จำเสียงโดยใช้แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) ซึ่งเป็นสถาปัตยกรรมหนึ่งที่มีโครงสร้างเลียนแบบโครงสร้างของสมองมนุษย์ โครงสร้างของโครงข่ายประสาทเทียมมีแบบจำลองหลายแบบ เช่น Back Propagation, Perceptron, Adaptive Resonance 1 (ART1) เป็นต้น โครงงานนี้ใช้แบบจำลอง Adaptive Resonance Theory 1 (ART1) ซึ่งเป็นการรู้จำเสียงที่รับค่าอินพุทเป็นเลขฐานสอง (Binary) โดยมีลักษณะการทำงาน คือ จะแปลงสัญญาณเสียงที่รับเข้ามาเป็นสัญญาณอนาลอกให้เป็นสัญญาณดิจิตอล คือ อยู่ในรูปของเลขฐานสอง หลังจากนั้นจะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการแยกลักษณะสำคัญของเสียงแล้วนำไปเปรียบเทียบกับข้อมูลอ้างอิงในฐานข้อมูล แล้วทำการตัดสินใจว่าเป็นเสียงของใคร

1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษาและพัฒนาระบบการรู้จำเสียงโดยใช้โครงข่ายประสาทเทียม

1.2.2 พัฒนาระบบการรู้จำเสียงโดยใช้ Adaptive Resonance Theory 1

1.2.3 ศึกษาผลการนำโครงข่ายประสาทเทียมชนิด ART1 มาใช้ในการรู้จำเสียง

1.2.4 เพื่อศึกษาหาแนวทางในการรู้จำเสียงที่มีประสิทธิภาพ

1.3 เป้าหมายและขอบเขต

1.3.1 สร้างระบบการรู้จำเสียงโดยใช้ Adaptive Resonance Theory 1

1.3.2 สามารถรู้จำคำพูดได้ตรงตามชุดคำที่ได้กำหนดไว้

1.3.3 มีความแม่นยำในการรู้จำในอัตราที่ยอมรับได้

1.3.4 ระบบนี้เป็นการรู้จำเสียงที่เป็นคำเดียวโดด ๆ

1.4 ขั้นตอนและวิธีการดำเนินการ

1.4.1 ศึกษาคุณลักษณะของเสียงพูดและแบบจำลองการเปล่งเสียงพูด

1.4.2 ศึกษาค้นคว้าข้อมูลเกี่ยวกับการประมวลผลสัญญาณ (Signal Processing)

1.4.3 ศึกษาค้นคว้าข้อมูลเกี่ยวกับขั้นตอนการรู้จำเสียง

1.4.4 ศึกษาค้นคว้าข้อมูลเกี่ยวกับโครงข่ายประสาทเทียม

1.4.5 ศึกษาอัลกอริทึ่มของ Adaptive Resonance Theory 1

1.4.6 เก็บข้อมูลเสียงพูดจากกลุ่มตัวอย่าง

- กลุ่มตัวอย่างเสียงพูดเพื่อฝึกฝน (Training)
- กลุ่มตัวอย่างเสียงพูดเพื่อทดสอบ (Testing)

1.4.7 วิเคราะห์และพัฒนาโปรแกรม

1.4.8 ทดสอบอัตราความถูกต้องแม่นยำในการรู้จำเสียงพูด

1.4.9 ปรับปรุงแก้ไขโปรแกรม เพื่อเพิ่มอัตราความถูกต้องในการรู้จำ

1.4.10 สรุปและจัดทำเอกสารประกอบโครงงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 เครื่องมือที่ใช้ในการทดลอง

1.5.1 โครงงานนี้พัฒนาบนเครื่องไมโครคอมพิวเตอร์

- IBM – PC 300 GL
- Harddisk 2.1 GB
- Memory 32 MB
- Card LAN on Board
- Disk Drive 3.5 “
- Mouse, Keyboard, Monitor 17”

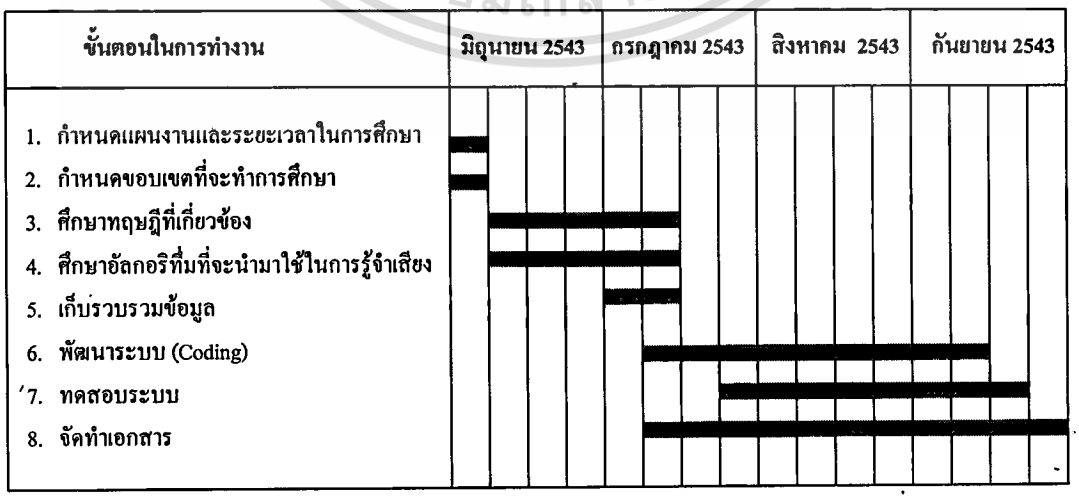
1.5.2 อุปกรณ์เพิ่มเติมในการทดลอง

- การ์ดเสียง (Sound Card)
- ไมโครโฟน
- ลำโพง

1.5.3 ซอร์ฟแวร์ที่ใช้

- Windows 98
- Microsoft office 97
- Cool Edit 2000
- Borland C++ Version 5.02 , Borland C++ Builder

1.6 ตารางเวลาในการพัฒนาระบบ



ตารางที่ 1.1 ตารางเวลาในการพัฒนาระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 ได้โปรแกรมโครงข่ายประสาทเทียมชนิด ART1 เพื่อนำไปใช้ในการรู้จำเสียง

1.7.2 ได้ทราบถึงองค์ประกอบข้อจำกัดและผลการใช้โปรแกรมโครงข่ายประสาทเทียมชนิด ART1

1.7.3 ได้ทราบถึงแนวทางและรูปแบบที่เหมาะสมในการพัฒนาระบบรู้จำเสียง

1.7.4 เพื่อหาแนวทางในการเพิ่มประสิทธิภาพในการรู้จำเสียงให้กับคอมพิวเตอร์

1.7.5 เป็นการพัฒนาเทคโนโลยีด้านคอมพิวเตอร์ให้สามารถทำงานเลียนแบบมนุษย์ได้อย่างมีประสิทธิภาพ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 ระบบการรู้จำเสียงของมนุษย์

ระบบการรู้จำเสียงของมนุษย์ มีความละเอียดอ่อนและซับซ้อนมาก มีการเรียนรู้และปรับปรุงความสามารถไปพร้อมกับการพัฒนาของร่างกายตามอายุ ระดับการรู้จำเสียงจะดำเนินไปตามขั้นตอน จนกระทั่งมีความชำนาญในการฟัง ประกอบกับมีโครงสร้างที่มีประสิทธิภาพสอดคล้องกับระบบการได้ยินทำให้สามารถแยกแยะคำและเสียงที่ซับซ้อนได้อย่างดี ความชำนาญในการฟังและการรู้จำเสียงของมนุษย์เกิดจากการเรียนรู้ ฝึกฝน และมีการพัฒนาอยู่ตลอดเวลา จากการศึกษาพฤติกรรมพบว่าการวิเคราะห์และการรู้จำเสียงของมนุษย์มีลักษณะและความสามารถในการฟังและการรู้จำเสียงหลายอย่าง ดังนี้

2.1.1 การตีความหมายของคำที่ได้ยินขึ้นกับสถานการณ์, เวลา, สถานที่, ความรู้สึกและเรื่องที่กำลังสนใจ

2.1.2 สามารถรับรู้ความหมายเดียวกันได้ถึงจะมีความแตกต่างกันทางด้าน ความดัง, ความถี่, พิท และความยาวของเสียงที่เปล่งออกมา

2.1.3 ขณะที่สนใจจะวิเคราะห์เสียงใดเสียงหนึ่ง ความสามารถวิเคราะห์เสียงอื่น รอบข้างจะด้อยลง

2.1.4 สามารถเลือกฟังเฉพาะเสียงที่ต้องการฟังเพื่อตีความหมาย

2.1.5 สามารถวิเคราะห์ได้ทั้งเสียงภาษาของบุคคลทั่วไปและบุคคลเฉพาะ

2.1.6 สามารถรับรู้ทิศทางแหล่งกำเนิดเสียง

2.1.7 สามารถตีความเสียงได้โดยไม่ขึ้นต่อชนิดของแหล่งกำเนิดเสียงพูด

2.1.8 รับรู้อารมณ์, ความรู้สึกและน้ำเสียงของแหล่งกำเนิดเสียงพูด

2.1.9 การรับรู้มีความถูกต้องมากขึ้นเมื่อคู่สนทนาได้เห็นสายตาและลักษณะท่าทางประกอบการพูด

2.1.10 สามารถแยกแยะคุณสมบัติของเสียงที่มีลักษณะใกล้เคียงกันได้

2.1.11 มีการเรียนรู้ตลอดเวลาและมีการปรับตัวเข้ากับสิ่งแวดล้อมได้อย่างรวดเร็ว

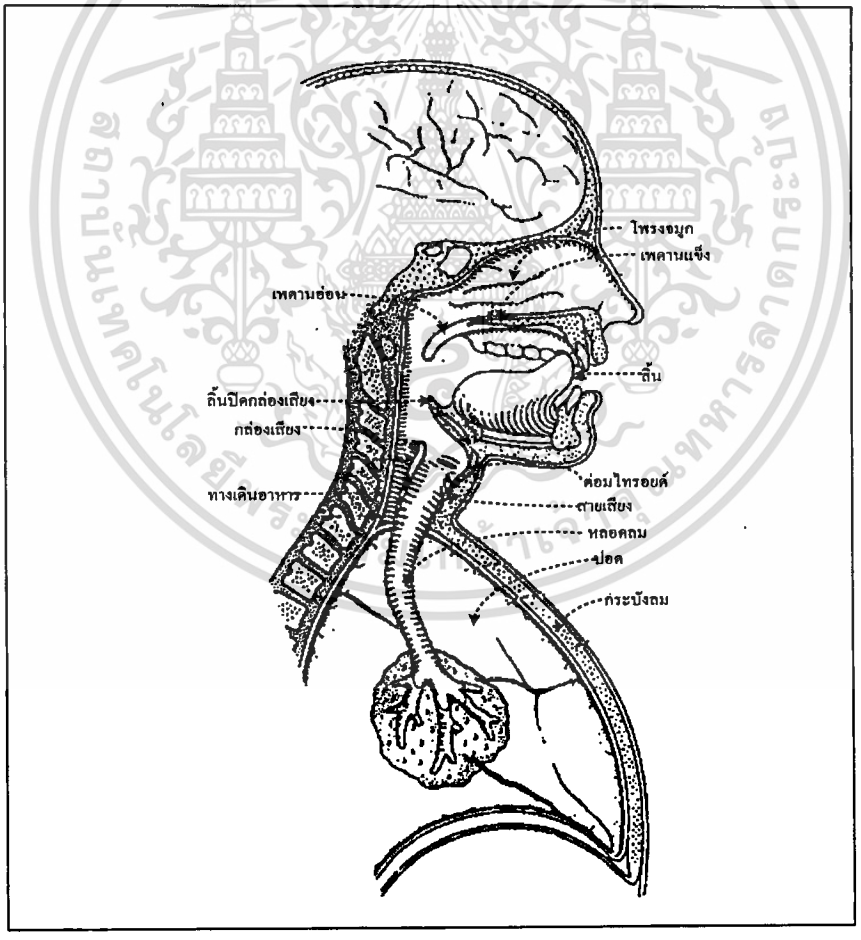
2.1.12 สามารถปรับระดับความดังของการรับฟังให้อยู่ในระดับที่สมองสามารถนำไปวิเคราะห์ได้อย่างอัตโนมัติโดยไม่คำนึงถึงแหล่งกำเนิดเสียงว่าจะดังหรือค่อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.13 มีการทำงานสอดคล้องประสานกันกับจิตใจที่สามารถควบคุมการวิเคราะห์จากการได้ยินได้

2.2 ระบบส่งเสียงของมนุษย์

การทำให้เกิดเสียงเป็นหน้าที่หนึ่งของระบบหายใจ การออกเสียงหรือการพูดของมนุษย์แต่ละครั้งจะต้องมีการทำงานร่วมกันของอวัยวะต่าง ๆ ของร่างกาย ซึ่งประกอบไปด้วย ปอด, กระบังลม, หลอดลม, เส้นเสียง, กล่องเสียง, ช่องปาก, หลอดอาหาร, ขากรรไกร, เพดานอ่อน, เพดานแข็ง, โพรงจมูก, รูจมูก ลิ้น, ฟัน และริมฝีปาก เป็นต้น ซึ่งอวัยวะต่าง ๆ เหล่านี้จะเป็นส่วนสำคัญในกระบวนการผลิตเสียง ของมนุษย์ ดังรูปที่ 2.1



รูปที่ 2.1 ระบบส่งเสียงของมนุษย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.1 จะแสดงถึงระบบส่งเสียงของมนุษย์ ดังนี้

2.2.1 ปอดและกระบังลม ทำหน้าที่ในการหายใจ และเป็นต้นกำเนิดการไหลของอากาศในกระบวนการผลิตเสียง

2.2.2 หลอดลมทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียง กล่องเสียงเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

2.2.3 สายเสียง หรือ เส้นเสียง เป็นอวัยวะสำคัญที่ทำให้เกิดเสียง มีลักษณะเป็นกล้ามเนื้อ 2 แผ่นปิดขวางอยู่บริเวณปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างเป็นทางผ่านให้ลมเข้าถึงปอดและออกจากปอดได้เรียกว่า ช่องระหว่างเส้นเสียงเส้นเสียงทั้ง 2 เส้นสามารถดึงออกห่างจากกันหรือดึงเข้าหากันได้

2.2.4 กล่องเสียง เป็นอวัยวะพิเศษที่ทำหน้าที่เป็นทางเดินอากาศเวลาหายใจ และเป็นตัวผลิตพัลส์ (pulse) ของอากาศขณะเปล่งเสียง ซึ่งประกอบ ด้วยเส้นเสียง และช่องสายเสียง

2.2.5 เพดานอ่อน เป็นกระดูกอ่อนที่อยู่ต่อจากเพดานแข็ง สามารถขยับขึ้นลงได้ ขณะที่หายใจเพดานอ่อนและลิ้นไก่ จะลดระดับลงมาปิดช่องให้ลมออกไปทางจมูก และถ้าหากออกเสียงนาสิกเพดานอ่อนจะลดระดับลงมาทำให้ลมออกทางช่องจมูก

2.2.6 เพดานแข็ง เป็นส่วนที่มีลักษณะโค้งเป็นกระดูกแข็ง หรือเรียกอีกอย่างหนึ่งว่าเพดานปาก

2.2.7 ช่องปากและส่วนของหลอดอาหารตอนต้น อยู่ต่อจากกล่องเสียงอาจเรียกว่าอวัยวะกำทอนเสียง (Vocal Tract) ทำหน้าที่กำทอนเสียง โดยให้กำทอนทั้งเสียงที่เกิดจากกล่องเสียงและเสียงที่เกิดภายในช่องปาก ขนาดขึ้นอยู่กับตำแหน่งของลิ้น ริมฝีปาก ขากรรไกร และเพดานอ่อน และเปลี่ยนแปลงไปตามการออกเสียง

2.2.8 โพรงจมูกเริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสอง ทำหน้าที่กำทอนเสียงรวมกับช่องปาก เมื่อมีการเปล่งเสียงที่ออกทางจมูก

2.2.9 ลิ้น ในเวลาที่มีการออกเสียงพูดลิ้นจะเป็นอวัยวะที่เคลื่อนไหวมากที่สุดทำให้เสียงที่พูดออกมาแตกต่างกันไป

2.2.10 ฟัน เป็นอวัยวะที่เป็นปัจจัยในการทำให้เกิดเสียงแตกต่างกันไปได้หลายแบบ

2.2.11 ริมฝีปาก เป็นอวัยวะที่ทำให้เสียงแตกต่างกันได้มาก เนื่องจากเคลื่อนไหวได้มาก และเสียงที่ออกมาจะขึ้นอยู่กับลักษณะของริมฝีปากด้วย

จากระบบส่งเสียงของมนุษย์จะเห็นได้ว่าเสียงที่ใช้ในภาษาพูดนั้น จะมีลักษณะที่สำคัญบางประการร่วมกัน เรียกว่าเป็นลักษณะร่วมของเสียงพูด ลักษณะดังกล่าวนี้มีหลายประการ คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก. ความก้อง หรือ ไม่ก้องของเสียง เสียงก้อง หรือ เสียงโฆษะ (Voice) คือ เสียงที่เกิดในขณะที่เส้นเสียงเกิดการตึงตัวหรือเรียกว่าเส้นเสียงปิด เมื่อมีแรงดันให้อากาศไหลผ่านกล่องเสียงในขณะที่เส้นเสียงปิดจะเกิดการสั่นสะบัดของเส้นเสียง เป็นผลให้สัญญาณเสียงที่ได้(Speech Waveform) มีลักษณะเป็นคาบ (Quasi-periodic) ซึ่งสามารถเรียกความถี่ในการปิด – เปิดของเส้นเสียงนี้ว่า "ความถี่มูลฐาน" (Fundamental Frequency : F_0) ตัวอย่างของเสียงก้องได้แก่ เสียงสระต่าง ๆ และเสียงพยัญชนะ เช่น บ ด ที่เกิดจากการเปล่งเสียงออกทางปาก หรือ เสียงพยัญชนะ ม น ง ที่เกิดจากการเปล่งเสียงออกทางจมูก

เสียงไม่ก้อง หรือ เสียงอโฆษะ (Unvoice หรือ Voiceless) คือ เสียงที่เกิดในขณะที่เส้นเสียงคลายจากการจึงหรือเรียกว่าเส้นเสียงเปิด เมื่อมีแรงดันให้อากาศไหลผ่านกล่องเสียงในขณะที่เส้นเสียงเปิด อากาศที่ไหลผ่านอย่างรวดเร็วจะเกิดการไหลวนและปั่นป่วนทำให้เกิดเสียงที่มีลักษณะเป็นเสียงของสัญญาณรบกวน (Noise) ซึ่งไม่เป็นคาบ ตัวอย่างของเสียงไม่ก้องได้แก่ เสียงพยัญชนะ ฟ ซ ส ฯลฯ หรือเกิดจากการสร้างแรงดันอากาศหลังตำแหน่งปิดกั้นของช่องทางเดินเสียง และเมื่อการปิดกั้นนี้ถูกเปิดออกอากาศจะถูกปล่อยออกมาอย่างทันที ทันใดเกิดเป็นเสียงที่เรียกว่า เสียงระเบิด (Plosive Sound) เช่น การเปล่งเสียงเริ่มแรกของพยัญชนะต้นของคำต่าง ๆ

ข. ความยาวของเสียง (Length) หมายถึง การที่เสียงใดเสียงหนึ่งเปล่งออกมาได้นานเท่าใด เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระ เสียงพยัญชนะนาสิก หรือเสียงพยัญชนะเสียดแทรกในภาษาไทยมีเพียงเสียงสระเท่านั้นที่เป็น เสียงพูดที่มีความยาว – สั้น เช่น อะ อิ อุ เป็นเสียงสั้น อา อี อู เป็นเสียงยาว เป็นต้น

ค. ระดับเสียงสูง–ต่ำ (Pitch) เสียงพูดจะมีระดับสูงหรือระดับต่ำนั้น อยู่ที่ความถี่ของเสียง (Fundamental frequency) ถ้าความถี่ต่ำระดับเสียงก็จะต่ำ อวัยวะส่วนที่ทำให้เสียงมีระดับ สูง–ต่ำ คือ เส้นเสียง ดังนั้นระดับเสียง สูง – ต่ำ คือ อัตราการสั่นสะบัดของเส้นเสียงนั่นเอง ความดัง (Loundness) ความดังของเสียงขึ้นอยู่กับปริมาณของลมที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่ง ๆ

ง. การลงน้ำหนัก (Stress) หมายถึง การออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นที่อยู่ข้างเคียง

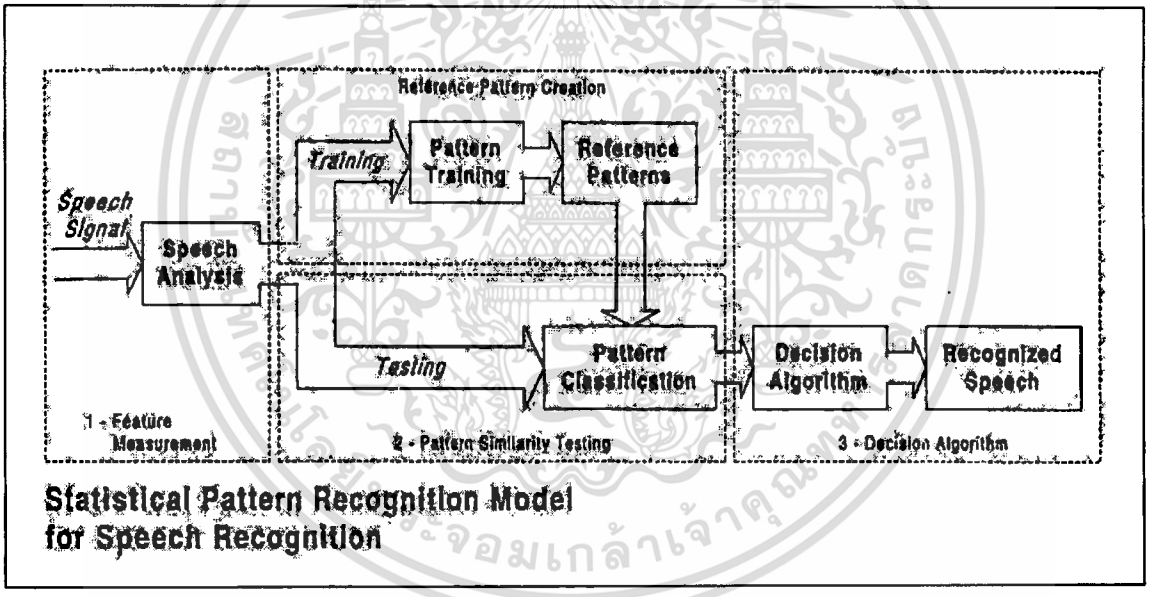
จ. ช่วงต่อของเสียง (Juncture) หมายถึง ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งอีกเสียงหนึ่งซึ่งเรียงกันมาเป็นลำดับ เสียงที่ประกอบกันเข้าเป็นพยางค์จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (Close Juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์หรือคนละคำ จะมีช่วง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

"ห่าง" จนสังเกตได้ชัด (Open Juncture) ดังนั้น ช่วงต่อของเสียงโดยเฉพาะช่วงต่อห่างจะมีความสำคัญมากในการแบ่งคำในภาษา

2.3 การรู้จำเสียง (Speech Recognition)

ในระบบการรู้จำเสียง จะเริ่มต้นจากจะมีการฝึก (Training) เพื่อให้ระบบจดจำรูปแบบอ้างอิง (Reference Patterns) เพื่อเก็บเป็นฐานข้อมูลในการเปรียบเทียบกับเสียงพูดที่ยังไม่ทราบรูปแบบ และเมื่อเปรียบเทียบกับรูปแบบอ้างอิงแล้วจะทำการตัดสินใจเลือกรูปแบบที่ใกล้เคียงมากที่สุดกับเสียงพูดที่นำมาเปรียบเทียบแล้วระบุว่าเป็นเสียงของใคร ขั้นตอนในระบบการรู้จำเสียงสามารถแสดงได้ ดังรูปที่ 2.2



รูปที่ 2.2 ระบบการรู้จำเสียง

จากรูปที่ 2.2 จะเห็นได้ว่าการรู้จำเสียงแบ่งเป็น 3 ขั้นตอน ดังนี้

- Feature Measurement เป็นขั้นตอนการวิเคราะห์และหาลักษณะสำคัญ โดยเริ่มต้นจากการรับสัญญาณเสียงเข้ามาแล้วทำการวิเคราะห์เพื่อแยกและเก็บลักษณะสำคัญจากเสียงพูดที่ยังไม่ทราบรูปแบบ
- Pattern Classification เป็นขั้นตอนการจำแนกรูปแบบของเสียง เมื่อได้ลักษณะของเสียงแล้วขั้นตอนนี้จะนำไปเปรียบเทียบกับรูปแบบของคำแต่ละคำที่ได้เก็บไว้ในฐานข้อมูล เพื่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หารูปแบบที่ใกล้เคียงมากที่สุดกับเสียงที่พูดเข้ามาใหม่

- Decision Algorithm เป็นขั้นตอนการตัดสินใจ หลังจากเปรียบเทียบว่าข้อมูลเสียงที่เข้ามานั้นใกล้เคียงกับรูปแบบอ้างอิงใดมากที่สุด ถ้ามีระบบจะตัดสินใจว่าเป็นเสียงของใครทันที แต่ถ้าไม่มีหรือใกล้เคียงระบบจะไม่ตัดสินว่าเป็นเสียงของใครแต่จะเรียนรู้และเก็บรูปแบบใหม่นั้น

2.3.1 การแบ่งแยกชนิดของการรู้จำเสียง

2.3.1.1 แบ่งตามการขึ้นกับผู้พูด ได้ 2 ชนิด ดังนี้

ก. ขึ้นกับผู้พูด (Speaker Dependent)

ข. ไม่ขึ้นกับผู้พูด

2.3.1.2 แบ่งตามระดับของเสียงที่รู้จำ ได้ 2 ชนิด ดังนี้

ก. การรู้จำคำเดี่ยว (Isolated Word)

ข. การรู้จำคำพูดต่อเนื่อง (Continuous Speech)

- คำต่อเนื่อง (Connected Word)
- พูดเป็นประโยค (Conversational Speech)

2.3.1.3 แบ่งตามหน่วยของเสียงที่ใช้เป็นตัวทาบเทียบ (Reference Template)

ก. ใช้คำเป็นตัวทาบเทียบ (Word Based)

ข. ใช้หน่วยเสียงเป็นตัวทาบเทียบ (Phoneme Based)

ค. ใช้หน่วยที่อยู่ระหว่างคำกับหน่วยเสียง เป็นตัวทาบเทียบ

2.3.2 ประเภทรูปแบบของการรู้จำได้

2.3.2.1 การเข้าคู่ต้นแบบ (Template Matching)

2.3.2.2 ระบบตามกฎเกณฑ์ (Rule-Based System)

2.3.2.3 ระบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model : HMM)

2.3.2.4 โครงข่ายประสาทเทียม (Artificial Neural Network : ANN)

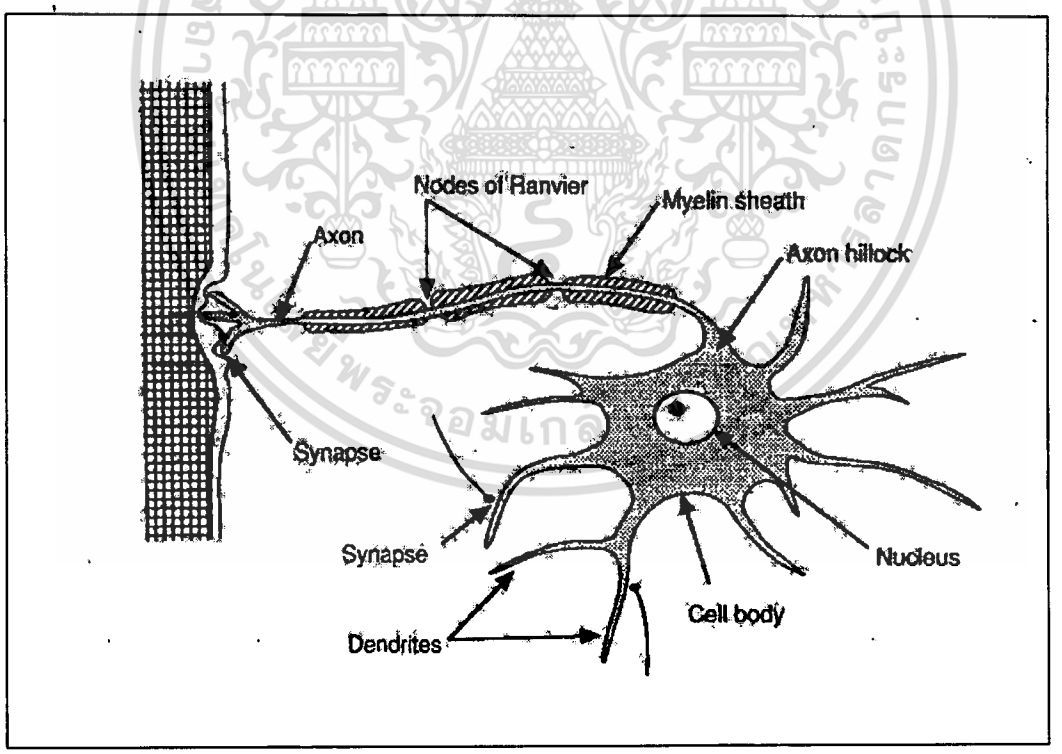
2.4 ทฤษฎีเซลล์ประสาททางชีววิทยาเบื้องต้น (Biological Neurons)

สมองมนุษย์ประกอบด้วยเซลประสาทขนาดเล็กประมาณ 2×10^{10} เซล เชื่อมโยงกันเป็นโครงข่าย ทำหน้าที่วิเคราะห์และสร้างสัญญาณกระตุ้นทางประสาท (Nurve Impulse) เพื่อควบคุมการทำงานต่าง ๆ ของร่างกาย โครงสร้างสำคัญของเซลประสาทประกอบด้วย ตัวเซลหรือโซม่า (Cell body/Soma) แอ็กซอน (Axon) และเดนไดรต์ (Dendrite) ตัวเซลมีนิวเคลียสอยู่ตรงกลางทำหน้าที่ประมวลผลสัญญาณกระตุ้นทางไฟฟ้า (Electrical Impulse) ที่ส่งเข้ามาทางเดนไดรต์ หากการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประมวลผลภายในตัวเซลต้องส่งสัญญาณกระตุ้นไปยังเซลอื่น ตัวเซลจะส่งสัญญาณกระตุ้นทางไฟฟ้ามีค่าระดับแรงดันประมาณ 50 มิลลิโวลท์ที่ช่วงเวลาประมาณ 2 มิลลิวินาทีผ่านส่วนของเซลที่เรียกว่า แอ็กซอน การส่งสัญญาณเอาท์พุทจากแอ็กซอนเพื่อเป็นสัญญาณอินพุทของอีกเซลหนึ่งเป็นกระบวนการทางเคมีที่เกิดขึ้นในส่วนของเซลที่เรียกว่าช่องว่างไซแน่พส์ (Synaptic gap)

ในการประมวลผลของนิวเคลียสว่าจะต้องส่งสัญญาณกระตุ้นออกจากตัวเซลหรือไม่นั้น พิจารณาจากผลรวมของสัญญาณอินพุทที่ผ่านเข้ามาทางเดนไดรต์ว่ามีค่าถึงระดับแรงดันกระตุ้น (Thershold Voltage) หรือไม่ หากมากว่าตัวเซลจะส่งสัญญาณกระตุ้นผ่านแอ็กซอนออกไป แต่ถ้าน้อยกว่าก็จะไม่มีการส่งสัญญาณกระตุ้นออกจากตัวเซล ลักษณะการส่งสัญญาณออกจากตัวเซลดังกล่าวจะเห็นว่าเซลประสาทมีฟังก์ชั่นการทำงานแบบ 2 สถานะ เช่นเดียวกันกับสถานะเปิดหรือปิด (on/off) ของสัญญาณไบนารีในเครื่องดิจิตอลคอมพิวเตอร์ โดยโครงสร้างเซลประสาทของมนุษย์แสดงได้ดังรูปที่ 2.3



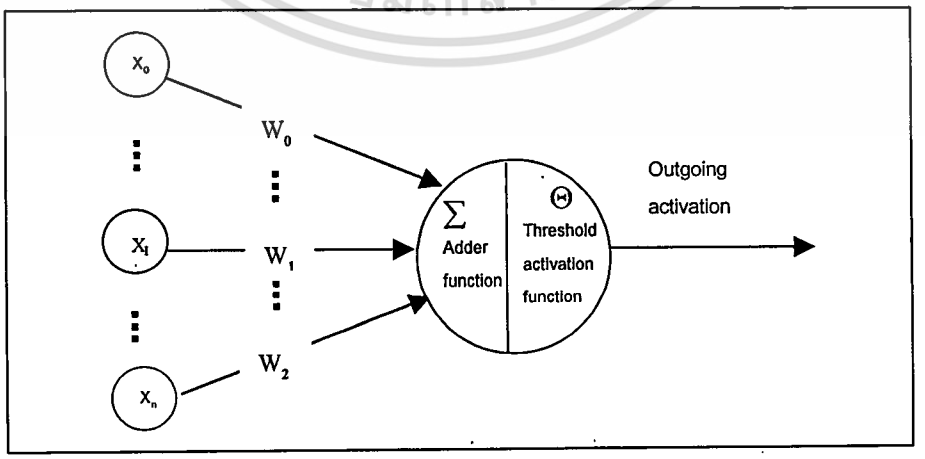
รูปที่ 2.3 โครงสร้างเซลประสาทของมนุษย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม เป็นสถาปัตยกรรมหนึ่งที่มีโครงสร้างที่เลียนแบบโครงสร้างของสมองมนุษย์ที่มีลักษณะเป็นเซลประสาทที่เชื่อมต่อกันแบบขนานกันจำนวนมาก ทำให้มีความสามารถในการคำนวณได้รวดเร็วและจำนวนมากพร้อมกัน และสมองมนุษย์มีวิธีการแก้ไขปัญหาที่เกิดจากการเรียนรู้ในอดีตและนำมาวิเคราะห์เพื่อแก้ไขปัญหาที่เข้ามาใหม่ เหตุผลข้างต้น จึงส่งผลให้มีการสร้างโครงข่ายประสาทเทียมเลียนแบบสมองมนุษย์ โดยที่โครงข่ายประสาทเทียมจะประกอบไปด้วยหน่วยประมวลผลจำนวนมากเชื่อมต่อกันเรียกว่า Neurons และมีส่วนที่เชื่อมต่อกันระหว่าง Neurons เรียกว่า Interconnection ซึ่งส่วนนี้จะเป็นส่วนที่เก็บความรู้ ที่จะใช้ในการแก้ไขปัญหาของโครงข่าย และความรู้ที่ใช้ในการแก้ปัญหานั้นได้มาจากการเรียนรู้จากตัวอย่าง ซึ่งเลียนแบบมาจากการแก้ไขปัญหาจากประสบการณ์ที่ผ่านมาของมนุษย์

โครงข่ายประสาทเทียมมีข้อดี คือมีความสามารถในการแยกแยะรูปแบบที่จะรู้จำได้ดีกว่าวิธีอื่น เนื่องจากมีคุณสมบัติ 3 ประการ คือ ประการที่หนึ่งการทนต่อความผิดพลาด (Fault tolerance) หมายถึง ระบบประมวลผลจะไม่หยุดทำงานกลางครันถึงแม้ว่าหน่วยประมวลผลบางหน่วยจะได้รับความเสียหายจนไม่อาจทำได้ ประการที่สอง คุณสมบัติ Generalization หมายถึง โครงข่ายจะยังให้ผลลัพธ์ที่สมเหตุสมผลแม้ว่าจะมีสัญญาณรบกวนหรือข้อมูลขาดความสมบูรณ์หรือข้อมูลไม่เคยเรียนรู้มาก่อนก็ตาม ประการที่สาม คุณสมบัติ Adaptability หมายถึง โครงข่ายสามารถเรียนรู้สภาพแวดล้อมที่เกิดขึ้นใหม่ได้รวดเร็วขึ้นสามารถวิเคราะห์หรือสร้างคำตอบสำหรับปัญหาที่มีลักษณะที่เหมือนหรือใกล้เคียงกับตัวอย่างที่เรียนรู้มาโดยที่ปัญหานั้นไม่เคยถูกนำมาสอนให้โครงข่ายเรียนรู้มาก่อน แบบจำลองทางตรรกะของโครงข่ายประสาทเทียมแสดงได้ ดังรูปที่ 2.4



รูปที่ 2.4 แบบจำลองทางตรรกของโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$x_0 \ldots, x_i \ldots, x_n$	สัญญาณอินพุทของโครงข่ายประสาทเทียม เทียบได้กับสัญญาณกระตุ้นที่ส่งเข้ามาทางแอ็กซอน
$w_0 \ldots, w_i \ldots, w_n$	ค่าน้ำหนัก (Weight) ของแต่ละบิตเทียบได้กับกระบวนการทางเคมีที่เกิดขึ้นในช่องว่างไซแนพส์เพื่อปรับระดับสัญญาณกระตุ้นก่อนส่งไปยังเดนไดรต์
$\sum$ Adder function	ฟังก์ชั่นรวมค่าสัญญาณอินพุทเทียบได้กับฟังก์ชั่นภายในตัวเซล
Thershold Activation	ฟังก์ชั่นพิจารณาผลรวมระดับแรงดันกระตุ้นที่เข้ามาในโครงข่ายเทียบได้กับกระบวนการตัดสินใจว่าจะส่งสัญญาณกระตุ้นออกจากตัวเซลหรือไม่
Outgoing Activation	สัญญาณเอาท์พุทของโครงข่าย เทียบได้กับสัญญาณกระตุ้นที่ส่งออกมาจากตัวเซลของประสาท

จากรูปจะเห็นได้ว่านิวรอน 1 โหนดจะมีค่าอินพุทหลายค่าแต่มีค่าเอาท์พุทเพียงค่าเดียวโดยเอาท์พุทในแต่ละนิวรอนเกิดจากผลรวมของผลคูณระหว่างอินพุทและตัวเลขน้ำหนักของอินพุทนั้นและเอาท์พุทที่ได้จากฟังก์ชั่นกระตุ้น เป็นค่าที่อยู่ในช่วง 0 ถึง 1 ถ้ามีค่าเป็น 1 หมายถึงสถานะการกระตุ้นของเซลประสาท ถ้ามีค่าเป็น 0 หมายถึงสถานะไม่กระตุ้นของเซลประสาท

ฟังก์ชั่นกระตุ้น (Activation Function : F()) คือ ฟังก์ชั่นคำนวณคำนวณสัญญาณเอาท์พุทของนิวรอนโดยพิจรณาจากระดับผลรวมของสัญญาณอินพุทที่ส่งเข้ามายังนิวรอนแต่ละหน่วยสำหรับอินพุทนิวรอน โดยฟังก์ชั่นกระตุ้นมีรูปแบบ ดังนี้

- Step Function เป็นฟังก์ชั่นที่ใช้ในการแปลงค่า Net Input ให้อยู่ในรูปไบนารี (Binary) คือ 0 และ 1 หรือ ไบโพลล่า (Bipolar) คือ 1 และ –1 ฟังก์ชั่นนี้จะใช้ค่าเทรสโฮลด์ (Threshold : θ) ในการกำหนดการแปลงค่า เรียกว่า Threshold Function โดยมีสมการดังนี้

สมการ Threshold Function

$$F(net) = 1 \quad \text{if net} \geq \theta$$
$$0 \quad \text{if net} \leq \theta$$

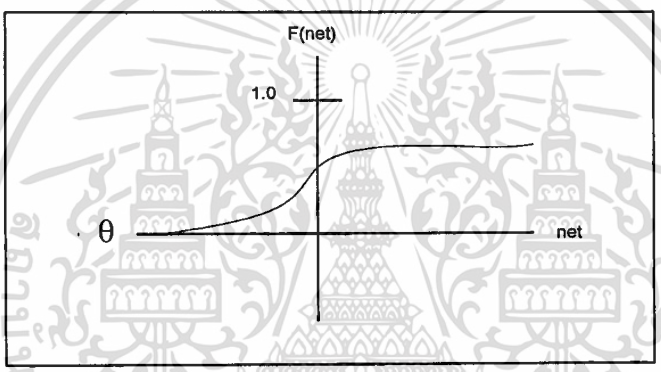
net : ค่า Net Input ที่ของเซลประสาทเทียม

θ : ค่าเทรสโฮลด์ที่ใช้ในการแปลงค่าของเซลประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

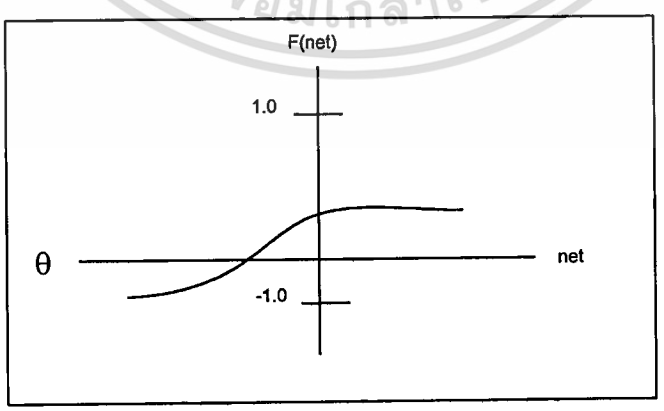
- Sigmoid Function (*S*-shaped curves) เป็นฟังก์ชั่นที่นิยมมากและฟังก์ชั่นนี้ให้ความละเอียดในการพิจาณาค่าเอาท์พุทสูง เพราะใช้ความสัมพันธ์ระหว่างสัญญาณอินพุทและค่าอนุพันธ์ของสัญญาณเอาท์พุท ซึ่งจะใช้ในการแปลงค่า Net Input ให้อยู่ในช่วง 0 ถึง 1 เรียกว่า Binary Sigmoid ฟังก์ชั่น หรืออยู่ในช่วง –1 ถึง 1 เรียกว่า Bipolar Sigmoid ซึ่งเป็นฟังก์ชั่นที่นำมาใช้ในการแปลงค่าเป็น Sigmoid Function สามารถแสดงด้วย Logistic Function และ Hyperbolic Tangent Function ตามรูปที่ 2.5 และรูปที่ 2.6

สมการ Logistic Function : $F(net) = 1/1(1 + e^{-net})$



รูปที่ 2.5 แสดงลักษณะของ Binary Sigmoid

สมการ Hyperbolic Tangent Function : $F(net) = (e^{net} - e^{-net}) / (e^{net} + e^{-net})$

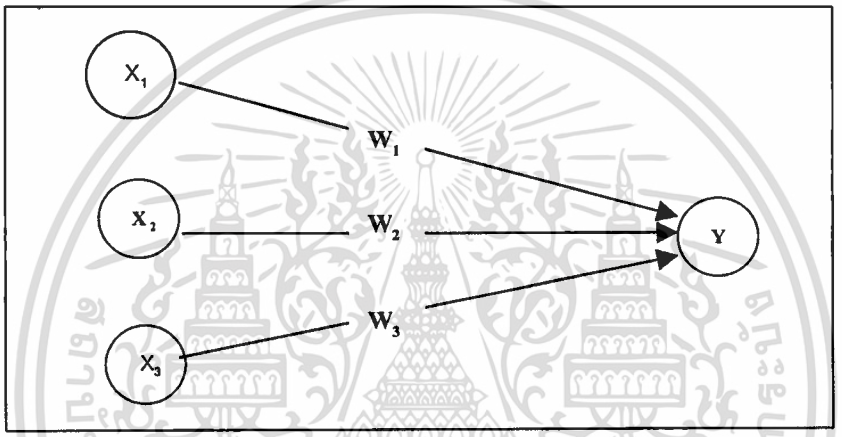


รูปที่ 2.6 แสดงลักษณะของ Bipolar Sigmoid

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.1 หลักการพื้นฐานของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมเป็นระบบการประมวลผลข้อมูลที่พัฒนาขึ้นโดยอาศัยหลักการทำงานพื้นฐานของโครงข่ายประสาททางชีวภาพที่กล่าวมาข้างต้น ภายในโครงข่ายประกอบด้วยหน่วยประมวลผลย่อยเรียกว่า นิวรอน (neuron) นิวรอนแต่ละหน่วยเชื่อมโยงกันแบบส่งสัญญาณโดยตรง (Direct Communication) แต่ละสายการเชื่อมโยง (Connection Link) มีค่าน้ำหนักเป็นตัวปรับระดับสัญญาณก่อนที่จะส่งเข้ามายังวงจรบวก ดังรูปที่ 2.7



รูปที่ 2.7 แสดงไดอะแกรมโครงข่ายประสาทเทียมพื้นฐาน

จากรูปที่ 2.7 แสดงไดอะแกรมโครงข่ายประสาทเทียมพื้นฐานประกอบด้วยนิวรอน x_1 x_2 x_3 และ Y โดยมี x_1 x_2 x_3 เป็นสัญญาณอินพุทของโครงข่าย w_1 w_2 w_3 เป็นค่าน้ำหนักเชื่อมอยู่ระหว่างนิวรอน x_1 x_2 x_3 ไปยังนิวรอน Y ตามลำดับ ผลรวมของสัญญาณอินพุท (y_in) ที่เข้ามายังนิวรอน Y เท่ากับ $y_in = w_1x_1 + w_2x_2 + w_3x_3$

หลังจากนั้นนิวรอน Y จะนำค่า y_in คำนวณค่าสัญญาณเอาท์พุทของโครงข่าย จากลักษณะพื้นฐานดังกล่าว สามารถสรุปขั้นตอนประมวลผลของโครงข่ายประสาทเทียมเป็นข้อ ๆ ดังนี้

2.5.1.1 การประมวลผลข้อมูลจะเกิดขึ้นในทุกนิวรอนที่ประกอบกันเป็นโครงข่าย

2.5.1.2 การส่งสัญญาณระหว่างนิวรอนกระทำผ่านสายการเชื่อมโยง

2.5.1.3 ในแต่ละสายการเชื่อมโยงมีน้ำหนักเป็นตัวคูณกับสัญญาณอินพุทของโครงข่าย

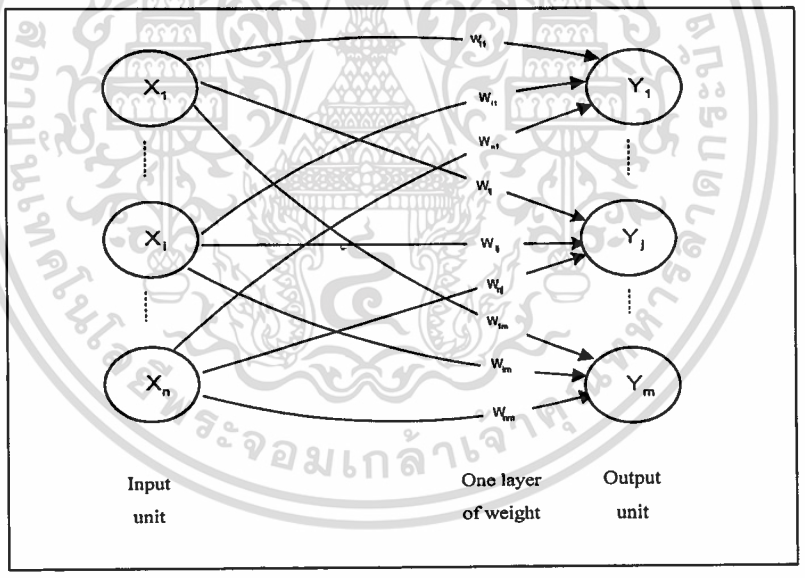
2.5.1.4 นิวรอนทุกหน่วยในโครงข่ายจะใช้ฟังก์ชั่นคำนวณค่าสัญญาณเอาท์พุทเดียวกัน

2.5.2 สถาปัตยกรรมโครงข่ายประสาทเทียม (Neural Network Architecture)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบ่งตามลักษณะโครงข่ายที่ประกอบกันเป็นโครงข่ายประสาทเทียมได้ 3 ประเภท คือ โครงข่ายประสาทเทียมแบบเลเยอร์เดียว (Single Layer), โครงข่ายประสาทเทียมแบบมัลติเลเยอร์(MultiLayer) และโครงข่ายประสาทเทียมแบบคอมเพททิทีฟว์เลเยอร์ (Competitive Layer) ซึ่งแต่ละแบบก็จะมีสถาปัตยกรรมที่แตกต่างกัน ดังนี้

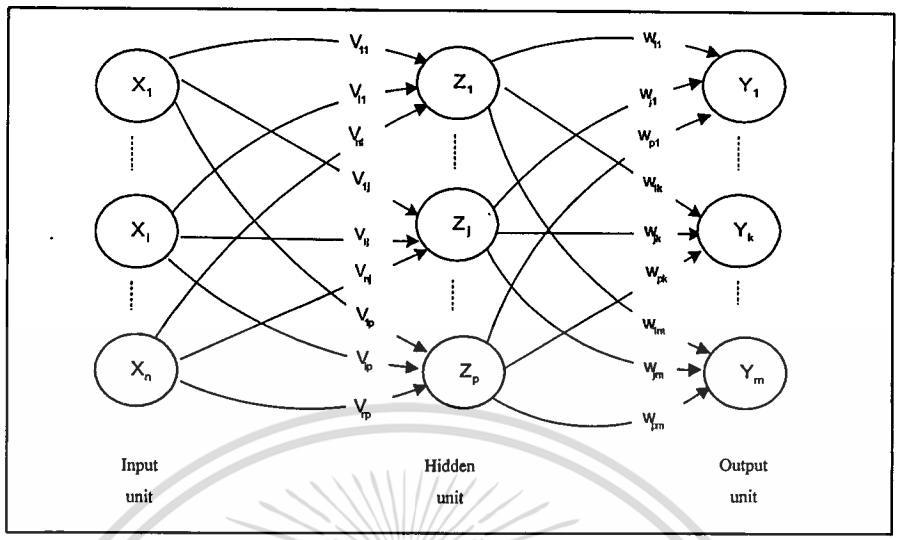
2.5.2.1 โครงข่ายประสาทเทียมแบบเลเยอร์เดียว (Single Layer) สถาปัตยกรรมโครงข่ายประสาทเทียมชนิดนี้ประกอบด้วย Input Neural เชื่อมต่อกับ Output Neural โดยมีค่าน้ำหนักเป็นตัวปรับระดับสัญญาณ Input ในหนึ่งเลเยอร์สามารถมีจำนวนนิวรอนได้มากกว่า 1 นิวรอน และภายในนิวรอนเดียวกันจะไม่เชื่อมต่อถึงกัน นิวรอนที่มีเลเยอร์สูงกว่าจะรับค่ามาจากเอาท์พุทของเลเยอร์ที่อยู่ต่ำกว่าเพื่อเป็นอินพุทของตัวเอง ส่วนค่าสัญญาณที่คำนวณได้จาก Output Unit จะเป็นคำตอบของโครงข่าย แสดงได้ ดังรูปที่ 2.8



รูปที่ 2.8 โครงข่ายประสาทเทียมชนิดเลเยอร์เดียว

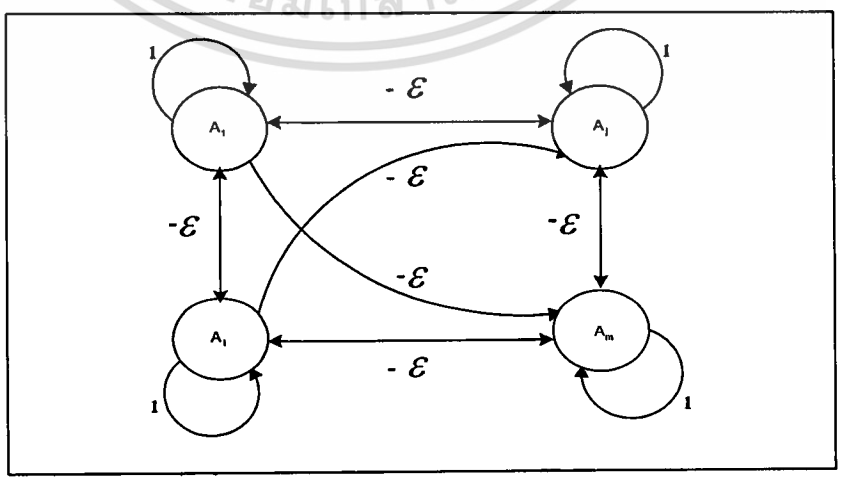
2.5.2.2 โครงข่ายประสาทเทียมชนิดมัลติเลเยอร์(MultiLayer) สถาปัตยกรรมโครงข่ายประสาทเทียมชนิดมัลติเลเยอร์ จะมีค่าน้ำหนักมากกว่า 1 เลเยอร์ โดยมี Hidden Neural เป็นตัวเชื่อมระหว่าง Input และ Output Neural ซึ่งโครงข่ายแบบมัลติเลเยอร์นี้ สามารถใช้วิเคราะห์ปัญหาที่ซับซ้อนได้ดีกว่าแบบเลเยอร์เดียวแต่จะมีกระบวนการที่ยุ่งยากกว่าดังรูปที่ 2.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.9 โครงข่ายประสาทเทียมชนิดมัลติเลเยอร์

2.5.2.3 โครงข่ายประสาทเทียมชนิดคอมเพททิทีฟว์เลเยอร์ (Competitive Layer) สถาปัตยกรรมโครงข่ายประสาทเทียมชนิดนี้ จะเป็นการเชื่อมโยงหลาย ๆ โครงข่ายเข้าด้วยกัน ซึ่งโครงข่ายย่อยที่ประกอบกันเป็นโครงข่ายใหญ่ จะเชื่อมต่อกันด้วยน้ำหนักซึ่งเป็นค่าคงที่ และแต่ละโครงข่ายย่อยก็จะมีสถาปัตยกรรมภายในของตัวเองไม่จำเป็นต้องเหมือนกันและเมื่อโครงข่ายย่อยประมวลผลภายในแล้วจะส่งสัญญาน Output ไปเป็นสัญญาณ Input ของโครงข่ายอื่นผ่านค่าน้ำหนัก จนกระทั่งเหลือเพียงโครงข่ายเดียวที่ให้ค่าสัญญาณ Output ไม่เท่ากับ 0 ดังรูปที่ 2.10



รูปที่ 2.10 โครงข่ายประสาทเทียมชนิดคอมเพททิทิฟว์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.3 การเรียนรู้ของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม เป็นโครงข่ายที่ฝึกให้เรียนรู้จำตัวอย่าง เพื่อนำไปใช้ในการแก้ไขปัญหาต่าง ๆ และแบบจำลองโครงข่ายประสาทเทียมแต่ละแบบจะมีขั้นตอนในการฝึกให้เรียนรู้แตกต่างกันไป ทั้งนี้สามารถแบ่งประเภทของการรู้จำได้ ดังนี้

2.5.3.1 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Leaning) เป็นการสอนโครงข่ายโดยการบันทึกข้อมูลอินพุทเข้าไปอย่างต่อเนื่องเพียงอย่างเดียว โดยไม่มีการส่งค่าเอาท์พุทเป้าหมายให้กับอินพุทแต่ละตัว การปรับค่าน้ำหนักใช้ข้อมูลที่นำมาสอนเป็นตัวปรับค่า โดยค่าน้ำหนักจะปรับตามกลุ่มที่อินพุทมีรูปแบบคล้าย ๆ กัน กระบวนการเรียนรู้ของโครงข่ายแบบนี้นิยมใช้กับงานง่าย ๆ ประเภทการเปรียบเทียบเอกลักษณ์ , รูปแบบที่สัมพันธ์กันระหว่างอินพุทและเอาท์พุท ที่ถูกกำหนดโดยโครงข่าย ไม่สามารถแก้ปัญหาที่เคร่งครัดได้

2.5.3.2 การเรียนรู้แบบมีผู้สอน (Supervised Learning) กระบวนในการเรียนรู้ของโครงโครงข่ายประสาทเทียม จะเริ่มจากก่อนที่จะนำโครงข่ายประสาทเทียมไปใช้งานเพื่อการรู้จำ ต้องมีกระบวนการเรียนรู้ให้กับโครงข่ายประสาทเทียมก่อนว่าจะรู้จำเสียงใด กระบวนการเรียนรู้นี้ คือ โครงข่ายประสาทเทียมจะได้รับข้อมูล Input – Output ตัวอย่าง แล้วให้โครงข่ายประสาทเทียมทำการปรับค่าน้ำหนักให้เหมาะสมกับข้อมูลที่ได้รับมา ซึ่งเป็นการสอนให้โครงข่ายทราบว่า ถ้า Input มีลักษณะเช่นนี้ Output ที่ถูกต้องควรมีลักษณะอย่างไร

โครงข่ายประสาทเทียมจะดำเนินกระบวนการเรียนรู้(ฝึกฝน) เป็นแบบวงรอบ ประกอบด้วย 3 ขั้นตอน คือ

ก. กระจายข้อมูลตัวอย่างเข้าสู่โครงข่าย

ข. คำนวณค่า Error Information Term

ค. การปรับค่าน้ำหนักภายในโครงข่าย

จุดสำคัญของกระบวนการในเรียนรู้ คือ การปรับค่าน้ำหนักภายในโครงข่ายประสาทเทียมแต่ละชั้น เพื่อรู้จำรูปแบบที่แตกต่างกันของข้อมูลเป้าหมาย โครงข่ายประสาทเทียม เป็นระบบโครงข่ายที่นำมาประยุกต์ใช้ในการรู้จำเสียงพูดซึ่งจะช่วยให้ระบบสามารถเรียนรู้และปรับปรุงตัวเองได้ เมื่อมีรูปแบบใหม่เข้ามา

2.5.4 แบบจำลองโครงข่ายประสาทเทียม

แบบจำลองโครงข่ายประสาทเทียมมี ดังนี้

2.5.4.1 Perceptron สำหรับ อ่านตัวพิมพ์ดีด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.4.2 ADALINE สำหรับ Adaptive Ewualizer

2.5.4.3 MDALINE สำหรับ Adaptive Ewualizer

2.5.4.4 Avalanche สำหรับ เข้าใจคำพูด

2.5.4.5 Cerebellatron สำหรับ ควบคุมการเคลื่อนไหวของแขนหุ่นยนต์

2.5.4.6 Back Propagation สำหรับ เลียนแบบฟังก์ชั่นทางคณิตศาสตร์ อ่านตัวเขียน, คำพูด และการทำนาย

2.5.4.7 Brain-State-in-a-box สำหรับ สกัดข้อมูลบางอย่างจากฐานข้อมูล

2.5.4.8 Adaptive Resonance สำหรับ Pattern Recognition

2.5.4.9 Self-Organizing Map สำหรับ จำลองลักษณะพื้นผิวของวัตถุ

2.5.4.10 Hopfield สำหรับ ทำนายส่วนของข้อมูลที่หายไป

2.5.4.11 Bidirectional Associative Memory สำหรับ Content-addressable Associative Memory

2.5.4.12 Boltzmann และ Cauchy Machine สำหรับ Pattern Recognition

2.5.4.13 propagation สำหรับ อัดข้อมูลของภาพให้น้อยลง, Table-lookup

2.5.5 ประวัติการพัฒนาโครงข่ายประสาทเทียม

การพัฒนาโครงข่ายประสาทเทียมเกิดขึ้นเมื่อมีการเสนอแบบจำลองทางตรรกะของโครงข่ายประสาทเทียมเป็นครั้งแรกในปี ค.ศ. 1943 โดย Warren McCulloch และ Walter Pitts แนวคิดสำคัญ คือ โครงข่ายจะมีการเชื่อมโยงจากชั้นอินพุทไปยังชั้นเอาท์พุทโดยผ่านน้ำหนัก (Weight) ซึ่งเป็นค่าน้ำหนักที่กำหนดไว้ตายตัว การคำนวณค่าสัญญาณเอาท์พุทโดยพิจารณาจากระดับกระตุ้น กล่าวคือ นิวรอนจะให้สัญญาณเอาท์พุทก็ต่อเมื่อผลรวมของสัญญาณอินพุทมีค่ามากกว่าระดับกระตุ้นซึ่งแนวคิดนี้ยังคงเป็นส่วนสำคัญในการพัฒนาโครงข่ายประสาทเทียมมาจนปัจจุบันนี้

ในปี ค.ศ. 1949 Dinal Hebb จากคณะฟิสิกส์มหาวิทยาลัยแมคกริลได้สร้างกฎการเรียนรู้ของโครงข่ายประสาทเทียมขึ้นเป็นครั้งแรก ทฤษฎีของ Hebb กล่าวว่าถ้านิวรอน 2 หน่วยทำงานพร้อมกันค่าน้ำหนักระหว่างสายการเชื่อมโยงของทั้งสองนิวรอนต้องลดลง ซึ่งกฎของทฤษฎีนี้คือขั้นตอนของกระบวนการปรับแต่งค่าน้ำหนักเพื่อให้เกิดค่าเอาท์พุทที่ต้องการ

ในช่วงปี ค.ศ. 1950 - 1960 ถือว่าเป็นช่วงที่มีการพัฒนาโครงข่ายประสาทเทียมกันอย่างกว้างขวาง เมื่อ John van Neumann บิดาแห่งวงการคอมพิวเตอร์ยุคใหม่ได้นำเอาแบบจำลองของสมองที่เสนอโดยนักวิจัยที่ทำงานเกี่ยวกับโครงข่ายประสาทเทียมไม่ว่าจะเป็น Warren McCulloch

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และคนอื่น ๆ มาเป็นแนวทางในการพัฒนาเครื่องคอมพิวเตอร์ยุคต่อมา จนในปี ค.ศ. 1956 ได้มีการสร้างแบบจำลองโดยใช้คอมพิวเตอร์ขึ้นสำเร็จ

ในปี ค.ศ. 1962 Frank Rosenblatt ได้เสนอทฤษฎีการเรียนรู้ของโครงข่ายที่เรียกว่า เพอร์เซพตรอน (Perceptrons) เป็นสถาปัตยกรรมที่ชั้นอินพุทต่อผ่านน้ำหนักไปยังชั้นเอาท์พุท การเรียนรู้ใช้การปรับค่าน้ำหนักภายในโครงข่ายจนกระทั่งได้ค่าสัญญาณเอาท์พุทตรงกับสัญญาณเอาท์พุทตัวอย่าง ต่อมา Bernard Widrow และ Marcian Hoff ได้พัฒนากฎการเรียนรู้ซึ่งใกล้เคียงกับการเรียนรู้ของเพอร์เซพตรอนและนำกฎนี้ไปใช้กับระบบหรือรูปแบบที่เรียกว่าอะดาไลน์ (ADALINE) และ Widrow ได้พัฒนารูปแบบมาดาไลน์ (MADALINE) ซึ่งเป็นรูปแบบที่พัฒนาจากอะดาไลน์

ในปี ค.ศ. 1972 Teuvo Kohonen ได้พัฒนาการเรียนรู้ของโครงข่ายด้วยการแก้ไขค่าน้ำหนักแบบแมตริกซ์ โดยอาศัยพื้นฐานจากทฤษฎีของ Hebb และในปีเดียวกันเขาก็ได้เสนอทฤษฎีการเรียนรู้ของโครงข่ายประสาทเทียมแบบจัดตัวเอง (Self-organizing) โดยการจัดรวมนิวรอนเป็นกลุ่ม เพื่อใช้ในการวิเคราะห์เสียงพูด โดยที่สถาปัตยกรรมของโครงข่ายประกอบด้วยชั้นอินพุทและชั้นเอาท์พุท การเชื่อมโยงที่ชั้นอินพุทไปยังนิวรอนที่ชั้นเอาท์พุทจะต่อผ่านน้ำหนัก กระบวนการเรียนรู้ของโครงข่ายนี้สามารถจัดกลุ่มข้อมูลอินพุทได้ด้วยตัวเอง เรียกการเรียนรู้แบบนี้ว่า การเรียนรู้แบบแข่งขัน (Competitive Learning) และต่อมา James Aderson ก็ได้พัฒนาโครงข่ายแบบ Brain-State-in-a-box เพื่อใช้ในการวินิจฉัยโรคและเรียนรู้สูตรคูณต่าง ๆ

ในปี ค.ศ. 1985 – 1990 Carpenter และ Grossberg ได้เสนอทฤษฎี Adaptive resonance สำหรับวิเคราะห์ข้อมูลแบบไบนารี่ (ART1) และข้อมูลอินพุทแบบต่อเนื่อง (ART2) David Parker ได้เสนอทฤษฎีการเรียนรู้แบบแพร่กลับ ทฤษฎีการเรียนรู้ของเขาใช้การนำค่าผิดพลาดของสัญญาณเอาท์พุทที่คำนวณได้กับสัญญาณเอาท์พุทตัวอย่างกลับมาเป็นค่าแก้ไขน้ำหนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

Adaptive Resonance Theory 1 (ART1)

3.1 Adaptive Resonance Theory 1 (ART1)

โครงข่าย ART1 พัฒนาขึ้นโดย Carpenter และ Grossberg (1987) และเป็นการเรียนรู้แบบแข่งขัน เป็นการย้อนกลับระหว่าง Competitive Layer กับ Input Layer ซึ่งมีความสามารถในการเรียนรู้ข้อมูลใหม่โดยไม่ลืมข้อมูลเก่า โดยในโครงข่าย ART1 ข้อมูลจะถูกส่งกลับและแพร่ออกไประหว่าง Layer จนกระทั่งวัดความคล้ายกันได้สำเร็จแต่ถ้าวัดความคล้ายกันไม่สำเร็จโครงข่ายจะเรียนรู้รูปแบบใหม่

3.1.1 ส่วนประกอบของโครงข่าย ART1

โครงข่าย ART1 เป็นโครงข่ายประสาทเทียมที่มี 2 Layer การเชื่อมต่อของเลเยอร์ ผ่าน Top-down และ Bottom-up สัญญาณที่ส่งผ่าน F_1 Layer โดยใช้ค่าน้ำหนักแบบ Bottom-up ไปยัง F_2 Layer เรียกว่า short-term memory (STM) และ Long-term memory (LTM) เป็นน้ำหนักที่เชื่อมต่อระหว่าง Layer และค่าน้ำหนักนี้เป็นข้อมูลสำคัญในการจดจำซึ่งเป็นส่วนหนึ่งของโครงข่าย

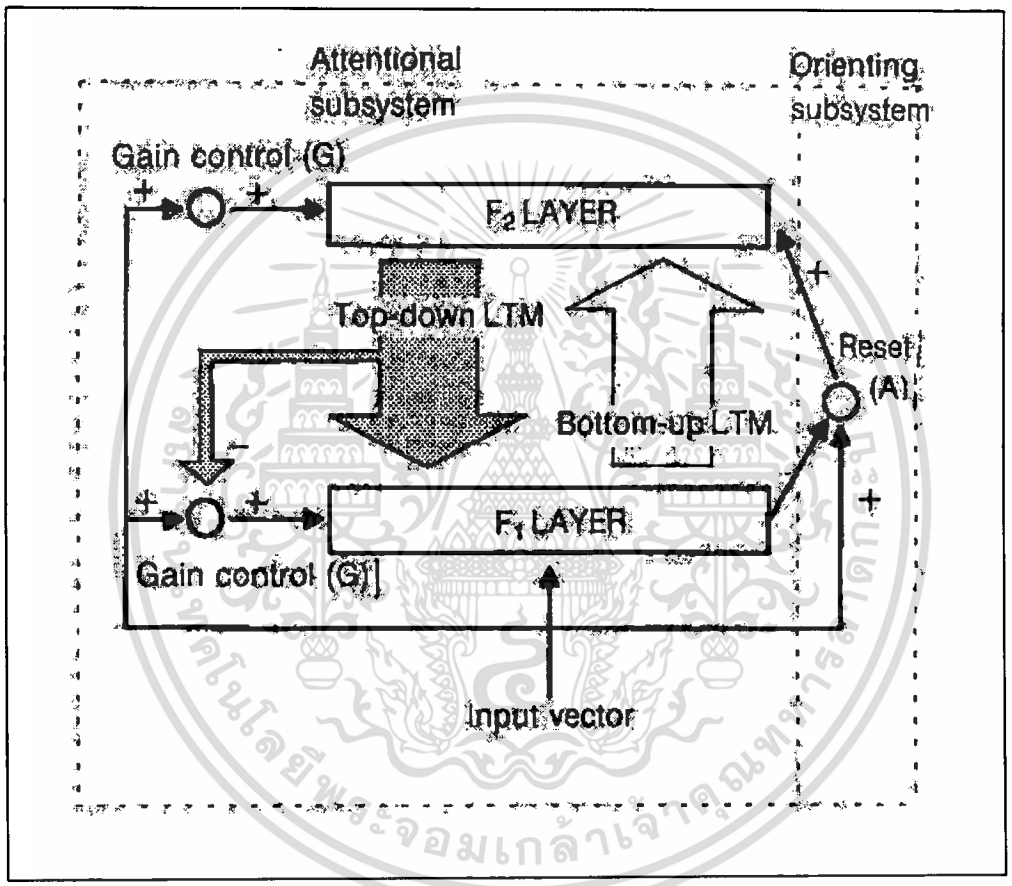
ระบบ ART1 ประกอบด้วย 2 ระบบย่อย คือ Attentional Supsystem จะโต้ตอบภายใน Pattern และสร้างการเรียนรู้แบบ Top-down LTM และ Bottom-up LTM ของ Pattern นั้น แต่ไม่สามารถสร้าง Pattern ใหม่ได้ และ Orienting Subsystem มีไว้เพื่อกำหนดรูปแบบการเรียนรู้และนำไปแสดงใน LTM หรือสร้างรูปแบบใหม่ที่ยังไม่เคยมีใน LTM โดย ART1 Model ประกอบไปด้วย

- เลเยอร์ของนิวรอน F_1 Layer เป็น อินพุทเลเยอร์หรือเลเยอร์ในการเปรียบเทียบ (Comparison Layer)
- โหนดของแต่ละเลเยอร์ เช่น Gain Control Unit
- เลเยอร์ของนิวรอน F_2 Layer เป็นเอาท์พุทเลเยอร์หรือเลเยอร์การรู้จำ (Recognition Layer)
- โหนด เช่น Reset Unit
- Bottom-up เป็นค่าน้ำหนักในการเชื่อมต่อจาก F_1 Layer ไป F_2 Layer
- Top-down เป็นค่าน้ำหนักในการเชื่อมต่อจาก F_2 Layer ไป F_1 Layer
- Inhibitory การเชื่อมต่อ (เป็น Negative Weight) จาก F_2 Layer ไป Gain Control

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Excitatory การเชื่อมต่อ (เป็น Positive Weight) จาก Gain Control ไปเลเยอร์
- Inhibitory การเชื่อมต่อจาก F_1 Layer ไป Reset Unit
- Excitatory เป็นการเชื่อมต่อจาก Reset Node ไป F_2 Layer

จากส่วนประกอบของ ART1 ดังกล่าวสามารถแสดงได้ ดังรูปที่ 3.1



รูปที่ 3.1 ระบบการรู้จำเสียงของสถาปัตยกรรม ART1

3.1.2 ลักษณะเฉพาะของ ART1

ลักษะหนึ่งของ ART1 Model คือ กฎ Two-third (Two-third rules) เป็นสิ่งที่จะกำหนดกิจกรรมของนิวรอนใน F_1 Layer โดยเข้ามา 3 ทาง คือ 1. มีอินพุทจากข้างนอก 2. เป็นอินพุทที่มาจากเอาท์พุทของ Gain Control และ 3. เป็นอินพุทที่มาจากเอาท์พุทของ F_2 Layer นิวรอน และ F_1 Layer จะไม่ทำงานจนกว่าจะมีอินพุทจาก 2 ใน 3 อินพุทดังกล่าว ลักษณะที่สอง มีค่าในการวัดความคล้ายกัน(Vigilance parameter) เป็นค่าที่ใช้ในการกำหนดกิจกรรมในการส่วนของ Reset Unit

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 เงื่อนไขในการคำนวณของ ART1

Z_{ij} คือ ค่าน้ำหนักในการเชื่อมต่อจากนิวรอนที่ I จาก F_1 Layer ไปนิวรอนที่ j ใน F_2 Layer

Z_{ji} คือ ค่าน้ำหนักในการเชื่อมต่อจากนิวรอนที่ j ใน F_2 Layer ไปนิวรอนที่ I ใน F_1 Layer

X_i เอาท์พุทของนิวรอนที่ i ใน F_1 Layer

Y_j เอาท์พุทของนิวรอนที่ j ใน F_2 Layer

z_i อินพุทของของนิวรอนที่ i ใน F_1 Layer จาก ใน F_2 Layer

ρ ค่าที่ใช้วัดความคล้าย (Vigilance Parameter) และมีค่าไม่มากกว่า 1

N_1 จำนวนของนิวรอนใน F_1 Layer

N_2 จำนวนของนิวรอนใน F_2 Layer

$|S^*|$ ผลรวมส่วนประกอบของอินพุทเวคเตอร์

$|I|$ ผลรวมเอาท์พุทของนิวรอนใน F_1 Layer

A, C, D เป็นพารามิเตอร์ที่มีค่า มากกว่าหรือเท่ากับ 0

L เป็นพารามิเตอร์ที่มีค่ามากกว่า 1

B เป็นพารามิเตอร์ที่มีค่าน้อยกว่า D + 1

3.1.4 อัลกอริทึมในการคำนวณของ ART1 (Algorithm for ART1 Calculations)

โครงข่ายแบบ ART1 รับค่า Input เป็นเลขฐานสอง (binary) ข้อดีของกระบวนการเรียนรู้ของโครงข่ายแบบ ART1 คือใช้เวลาในการเรียนรู้น้อย Freeman and Skapura(1991) ได้พัฒนาขั้นตอนการเรียนรู้ของ ART1 ดังนี้

Step 1 กำหนดให้ N_1 และ N_2 เป็นจำนวนของ Node ใน F_1 และ F_2 Layer

Step 2 A, B, C, D และ L เป็นค่าคงที่ ρ เป็นค่าที่ใช้ในการวัดความคล้ายกันในกระบวนการ matching

$$A, C, D >= 0$$

$$\text{Max } \{D, 1\} < B < D + 1$$

$$L > 1$$

$$0 < \rho <= 1$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 3 กำหนดค่าน้ำหนักใน Long-term memory

for $i = 1, \ldots, N_1$ และ $j = 1, \ldots, N_2$

Top-down weights : $Z_{ji} > B\text{-}1 \ / D$

Bottom-up weights: $0 < Z_{ij} < L \ / \ L - 1 \ + \ N_1$

Step 4 ปรับ input vector (I) ที่ F_1 Layer และคำนวณ X ผ่าน F_1 Layer

for $i = 1, \ldots, N_1$

$$X_i = I_i \ / \ 1 + A(I_i + B) + C$$

Step 5 คำนวณค่า Output vector (S) จาก F_1 Layer

for $i = 1, \ldots, N_2$

$S_i = 1$ if $X_i > 0$

$S_i = 0$ if $X_i <= 0$

Step 6 การแพร่ Pattern S ผ่าน Bottom-up LTM ไป F_2 Layer และ จะได้ Pattern Y ผ่าน F_2

for $j = 1, \ldots, N_2$

$$Y_j = \sum_{i=1}^{N_i} S_i Z_{ij}$$

Step 7 F_2 PEs ใช้การเรียนรู้แบบแข่งขัน โดยจะแข่งขันจนกว่าจะได้ Neuron ที่ชนะ และ Pattern U เป็น Output ที่เกิดจากการแข่งขัน

for $j = 1, \ldots, N_2$

$U_j = 1$ if $Y_j = \max \{Y\}$

$U_j = 0$ otherwise

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 8 Pattern U แพร่ย้อนกลับผ่าน Top-down LTM ไป F_1 Layer และ Pattern V เกิดจาก Pattern U จาก F_2 Layer

for $i = 1, \ldots, N_1$

$$V_i = \sum_{i=1}^{N} U_j Z_{ji}$$

Step 9 เริ่มคำนวณใหม่โดยใช้ X* จากทุก Input ของ F_1 Layer

for $i = 1, \ldots, N_1$

$$X^*_i = I_i + DV_i - B / 1 + A(I_i + DV_i) + C$$

Step 10 คำนวณ Output pattern ของ F_1 Layer ใหม่ โดย ใช้สมการเหมือนกันกับ Step 5 แต่ใช้ Pattern X* แทน X

Step 11 |P| เป็นจำนวนของสมาชิกที่ไม่ใช่ 0 ใน Pattern P , และค่า M เป็นค่าที่ใช้วัดความคล้ายกันระหว่าง Input Pattern I และ Pattern S คำนวณโดยใช้สูตร ดังนี้

$$M = |S^*| / |I|$$

Step 12 ถ้าค่า M น้อยกว่าค่า ρ ค่า J ใน F_2 Layer จะ ไม่เปลี่ยนแปลง และจะไปทำที่ Step 4 แต่ถ้าไม่ใช่จะทำการ Update Node J ใน F_2 Layer โดยทำตามขั้นตอนดังนี้

Step 13 ปรับค่าน้ำหนักของ Bottom-up weights

for $I = 1, \ldots, N_1$

$$Z_{ij} = I_i (L / L - 1 + |S|)$$

Step 14 ปรับค่าน้ำหนักของ Top-down

for $I = 1, \ldots, N_1$

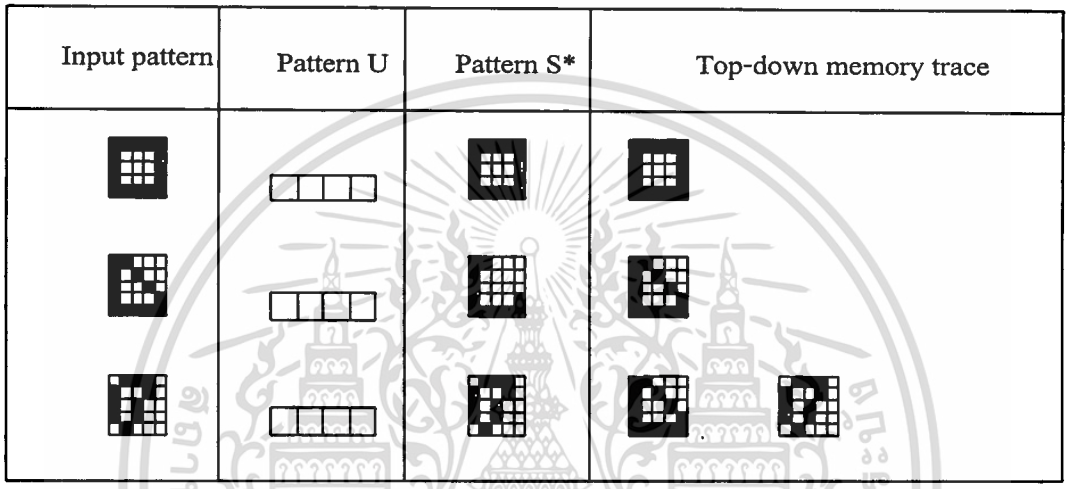
$$Z_{ji} = I_i$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 15 ลบทุก Node ใน F_2 Layer, รับค่า Input ใหม่เข้าสู่โครงข่าย แล้วไปทำ Step 4

3.1.5 ตัวอย่างการเรียนรู้ของ ART1

ตัวอย่างการเรียนรู้ของโครงข่าย ART1 จาก Input Pattern ตามรูปที่ 3.2



รูปที่ 3.2 ตัวอย่างของ Pattern matching ของ ATR1 โดยใช้ค่าวัดความคล้ายกันเท่ากับ 0.6

ตัวอย่าง การเรียนรู้ของ ART1 Pattern 1

Step 1 กำหนดค่า $N_1 = 25$, $N_2 = 4$

Step 2 กำหนดค่า $A = 1, B = 2, C = 1, D = 1.5, L = 2$

Step 3 Top-down weight $Z_{ji} = 0.75$

Bottom-up weights $Z_{ij} < 2 / 2 - 1 + 25 = 0.077$

Step 4 I = [1111110001100011000111111]

X = [0.2 0.2 0.2 0.2 0.2

0.2 0 0 0 0.2

0.2 0 0 0 0.2

0.2 0 0 0 0.2

0.2 0.2 0.2 0.2 0.2]

Step 5 S = [1111110001100011000111111]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 6 $Y = [\ 1.232\ 1.232\ 1.232\ 1.232]$

Step 7 $U = [\ 1\ 0\ 0\ 0]$

Step 8 $V = [0.75\ 0.75\ 0.75\ 0.75\ 0.75$

$0.75\ \ 0\ \ 0\ \ 0\ \ 0.75$

$0.75\ \ 0\ \ 0\ \ 0\ \ 0.75$

$0.75\ \ 0\ \ 0\ \ 0\ \ 0.75$

$0.75\ 0.75\ 0.75\ 0.75\ 0.75]$

Step 9 $X^* = [0.111\ \ 0.111\ 0.111\ 0.111\ \ 0.111$

$0.111\ \ -0.14\ \ -0.14\ \ -0.14\ \ 0.111$

$0.111\ \ -0.14\ \ -0.14\ \ -0.14\ \ 0.111$

$0.111\ \ -0.14\ \ -0.14\ \ -0.14\ \ 0.111$

$0.111\ \ 0.111\ 0.111\ \ 0.111\ 0.111]$

Step 10 $S^* = [1111110001100011000111111]$

Step 11 $M = 16 / 16 = 1$

Step 12 $M = 1$, $\rho = 0.6$ เพราะฉะนั้น $M > \rho$ จะไปทำ Step 13 - 14

Step 13 $Z_{i1} = [0.118\ 0.118\ 0.118\ 0.118\ 0.118$

$0.118\ \ 0\ \ 0\ \ 0\ \ 0.118$

$0.118\ \ 0\ \ 0\ \ 0\ \ 0.118$

$0.118\ \ 0\ \ 0\ \ 0\ \ 0.118$

$0.118\ 0.118\ 0.118\ 0.118\ 0.118\ ]$

$Z_{i2} = [0.077\ 0.077\ 0.077\ 0.077\ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ 0.077\ 0.077\ 0.077\ 0.077]$

$Z_{i3} = [0.077\ 0.077\ 0.077\ 0.077\ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ \ 0\ \ 0\ \ 0\ \ 0.077$

$0.077\ 0.077\ 0.077\ 0.077\ 0.077]$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Z_{i4} = [0.077 0.077 0.077 0.077 0.077
0.077 0 0 0 0.077
0.077 0 0 0 0.077
0.077 0 0 0 0.077
0.077 0.077 0.077 0.077 0.077]

Step 14 Z_{1i} = [1111110001100011000111111]

Step 15 ลบทุก Node ใน F_2 Layer, รับค่า Input ใหม่เข้าสู่โครงข่าย แล้วไปทำ Step 4

ตัวอย่าง การเรียนรู้ของ ART1 Pattern 2

Step 1 กำหนดค่า $N_1 = 25$, $N_2 = 4$

Step 2 กำหนดค่า A = 1, B = 2, C = 1, D = 1.5 , L = 2

Step 3 Top-down weight $Z_{ji} = 0.75$

Bottom-up weights $Z_{ij} = 0.118$

Step 4 I = [1100010100100101000111111]

X = [0.2 0.2 0 0 0
0.2 0 0.2 0 0
0.2 0 0 0.2 0
0.2 0 0 0 0.2
0.2 0.2 0.2 0.2 0.2]

Step 5 S = [1100010100100101000111111]

Step 6 Y = [1.298 1.001 1.001 1.001]

Step 7 U = [1 0 0 0]

Step 8 V = [0.75 0.75 0 0 0
0.75 0 0.75 0 0
0.75 0 0 0.75 0
0.75 0 0 0 0.75
0.75 0.75 0.75 0.75 0.75]

Step 9 X* =[0.111 0.111 -0.14 -0.14 -0.14
0.111 -0.14 0.111 -0.14 -0.14
0.111 -0.14 -0.14 0.111 0.111

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

0.111 -0.14 -0.14 -0.14 0.111

0.111 0.111 0.111 0.111 0.111]

Step 10 S* = [1100010000100001000111111]

Step 11 M = 11 / 13 = 0.84

Step 12 M = 0.84 , ρ = 0.6 เพราะฉะนั้น M > ρ

Step 13 Z_{i1} = [0.143 0.143 0 0 0

0.143 0 0.143 0 0

0.143 0 0.143 0 0

0.143 0 0 0.143 0

0.143 0.143 0.143 0.143 0.143]

Z_{i2} = [0.077 0.077 0 0 0

0.077 0 0.077 0 0

0.077 0 0.077 0 0

0.077 0 0 0.077 0

0.077 0.077 0.077 0.077 0.077]

Z_{i3} = [0.077 0.077 0 0 0

0.077 0 0.077 0 0

0.077 0 0.077 0 0

0.077 0 0 0.077 0

0.077 0.077 0.077 0.077 0.077]

Z_{i4} = [0.077 0.077 0 0 0

0.077 0 0.077 0 0

0.077 0 0.077 0 0

0.077 0 0 0.077 0

0.077 0.077 0.077 0.077 0.077]

Step 14 Z_{1i} = [1100010100100101000111111]

Step 15 ลบทุก Node ใน F_2 Layer, รับค่า Input ใหม่เข้าสู่โครงข่าย แล้วไปทำ Step 4

ตัวอย่าง การเรียนรู้ของ ART1 Pattern 3

Step 1 กำหนดค่า N_1 = 25, N_2 = 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 2 กำหนดค่า A = 1, B = 2, C = 1, D = 1.5 , L = 2

Step 3 Top-down weight $Z_{ji} = 0.75$

Bottom-up weights $Z_{ij} = 0.143$

Step 4 I = [01110100101010010100010 00]

$$X = \begin{bmatrix} 0 & 0.2 & 0.2 & 0.2 & 0 \\ 0.2 & 0 & 0 & 0.2 & 0 \\ 0.2 & 0 & 0.2 & 0 & 0 \\ 0.2 & 0 & 0.2 & 0 & 0 \\ 0 & 0.2 & 0 & 0 & 0 \end{bmatrix}$$

Step 5 S = [01110100101010010100010 00]

Step 6 Y = [0.715 0.77 0.77 0.77]

Step 7 U = [0 1 0 0 0]

Step 8 $V = \begin{bmatrix} 0 & 0.75 & 0.75 & 0.75 & 0 \\ 0.75 & 0 & 0 & 0.75 & 0 \\ 0.75 & 0 & 0.75 & 0 & 0 \\ 0.75 & 0 & 0.75 & 0 & 0 \\ 0 & 0.75 & 0 & 0 & 0 \end{bmatrix}$

Step 9 $X^* = \begin{bmatrix} -0.14 & 0.111 & 0.111 & 0.111 & -0.14 \\ 0.111 & -0.14 & -0.14 & 0.111 & -0.14 \\ 0.111 & -0.14 & 0.111 & -0.14 & -0.14 \\ 0.111 & -0.14 & 0.111 & -0.14 & -0.14 \\ -0.14 & 0.111 & -0.14 & -0.14 & -0.14 \end{bmatrix}$

Step 10 S* = [01110100101010010100010 00]

Step 11 M = 10/10 = 1

Step 12 M = 1 , ρ = 0.6 เพราะฉะนั้น M > ρ

Step 13 $Z_{il} = \begin{bmatrix} 0 & 0.166 & 0.166 & 0.166 & 0 \\ 0.166 & 0 & 0 & 0.166 & 0.166 \\ 0.166 & 0 & 0.166 & 0 & 0 \\ 0.166 & 0 & 0.166 & 0 & 0 \\ 0 & 0.166 & 0 & 0 & 0 \end{bmatrix}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Z_{i2} = [0 0.077 0.077 0.077 0
0.077 0 0 0.077 0.077
0.077 0 0.077 0 0
0.077 0 0.077 0 0
0 0.077 0 0 0]

Z_{i3} = [0 0.077 0.077 0.077 0
0.077 0 0 0.077 0.077
0.077 0 0.077 0 0
0.077 0 0.077 0 0
0 0.077 0 0 0]

Z_{i4} = [0 0.077 0.077 0.077 0
0.077 0 0 0.077 0.077
0.077 0 0.077 0 0
0.077 0 0.077 0 0
0 0.077 0 0 0]

Step 14 Z_{1i} = [011101001010100101000 1000]

Step 15 ลบทุก Node ใน F_2 Layer, รับค่า Input ใหม่เข้าสู่โครงข่าย แล้วไปทำ Step 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลองในการทำวิจัย

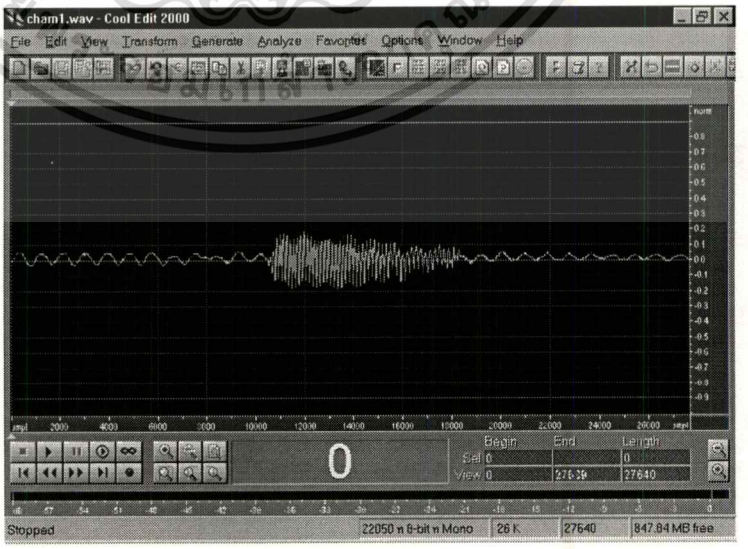
ขั้นตอนในการทำวิจัย

มีขั้นตอนในการทำวิจัยมี 3 ขั้นตอน คือ

4.1 ขั้นตอนการจัดเตรียมข้อมูล มีขั้นตอนการจัดเตรียม ดังนี้

4.1.1 เก็บข้อมูลกลุ่มเสียงตัวอย่างจากกลุ่มตัวอย่างเสียงจำนวน 15 คน โดยแบ่งเป็น 3 กลุ่ม คือ

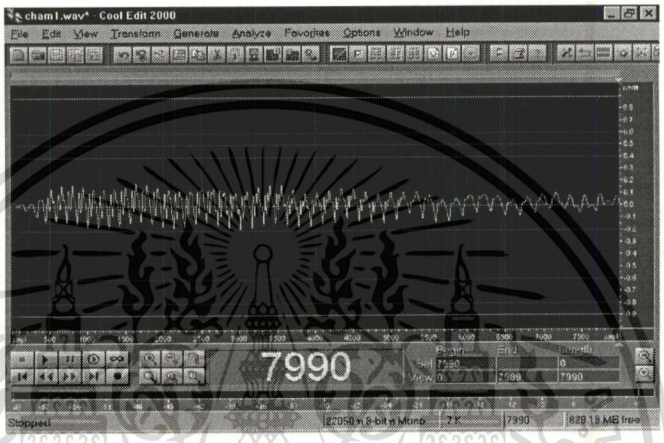
- กลุ่มที่ 1 จำนวน 10 คน จัดเป็นกลุ่มข้อมูลเสียงสำหรับการฝึกสอน
- กลุ่มที่ 2 จำนวน 5 คน จัดเป็นกลุ่มข้อมูลเสียงสำหรับทดสอบกลุ่มที่ 1
- กลุ่มที่ 3 จำนวน 5 คน จัดเป็นกลุ่มข้อมูลเสียงสำหรับทดสอบกลุ่มที่ 2
- ทำการบันทึกเสียงของทุกคนโดยให้แต่ละคนในกลุ่มพูดคำเดียวโดยพูดคำเดียวกัน 3 ครั้ง การทดลองในโครงงานนี้เป็นการรู้จำเสียงคำเดียว
- เก็บบันทึกเสียงไว้เป็นแฟ้มข้อมูล file.wave โดยพูด 1 คำ เก็บเป็น 1 แฟ้มข้อมูล ดังนั้นจะได้แฟ้มข้อมูลทั้งหมด 45 แฟ้ม และแฟ้มข้อมูลแต่ละแฟ้มจะมีลักษณะข้อมูล ดังรูปที่ 4.1



รูปที่ 4.1 ตัวอย่างแฟ้มข้อมูลเสียงกลุ่มตัวอย่าง

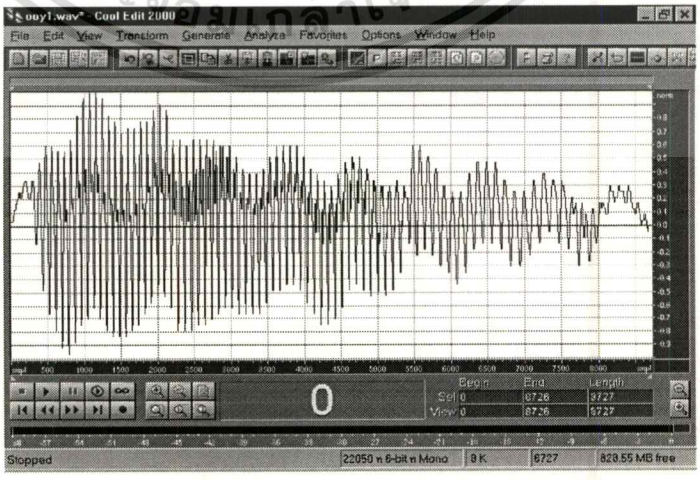
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 นำ file.wave ที่ได้จากการบันทึกเสียงมาทำการหาจุดสิ้นสุดของเสียงพูด โดยกำหนดจุดเริ่มต้นและจุดสิ้นสุดของคำในแต่ละแฟ้มข้อมูล โดยทำทั้งหมดทุกแฟ้มข้อมูลเพื่อแยกแยะเสียงพูด แล้วนำเอาเฉพาะส่วนที่เป็นเสียงพูดมาทำการวิเคราะห์ในขั้นตอนการรู้จำเสียงพูดของคำแต่ละคำ ซึ่งจะได้ลักษณะของเสียงก่อนกำหนดจุดเริ่มต้นและจุดสิ้นสุดมีลักษณะ ดังรูปที่ 4.2



รูปที่ 4.2 แสดงลักษณะของเสียงหลักจากกำหนดจุดเริ่มต้นและสิ้นสุดแล้ว

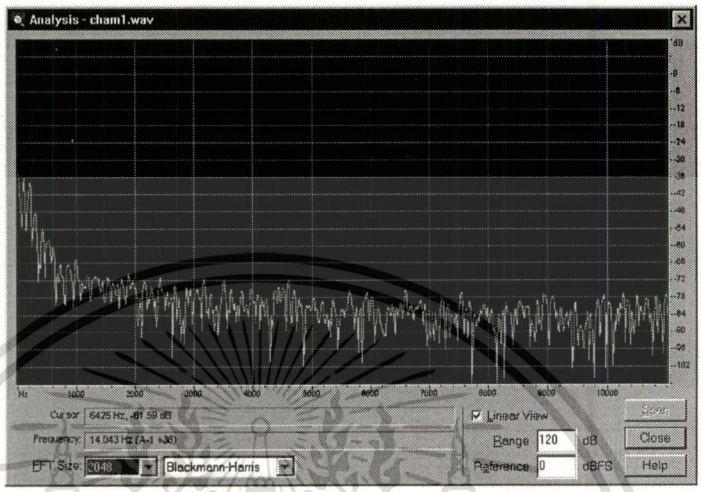
4.1.3 ทำการ Normalize เสียง เพื่อปรับขนาดความยาวของสัญญาณให้เหมาะสมเนื่องจากสัญญาณเสียงพูดที่ได้มีความยาวไม่เท่ากัน จึงกำหนดขนาดเพื่อใช้ในการหาลักษณะสำคัญและเปรียบเทียบสัญญาณเสียงโดยปรับให้มีขนาดเท่ากับ 100%



รูปที่ 4.3 แสดงสัญญาณเสียงที่ปรับขนาดความยาวเท่ากับ 100%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 วิเคราะห์ความถี่ของเสียงโดยมี FFT size 2048 แบบ Blackmann-Harrist โดยลักษณะเสียงที่ได้จะเป็นดังรูปที่ 4.4



รูปที่ 4.4 ลักษณะของเสียงที่ทำการวิเคราะห์ความถี่

4.1.5 นำค่าที่ได้จากการวิเคราะห์ความถี่ของเสียง แล้วทำการหาความถี่ตาม Slope ของเสียงแต่ละคน คนละ 3 แฟ้ม แล้วนำมาทำการหาค่าเฉลี่ยของแต่ละคน นำสัญญาณเสียงที่ได้จากการหาค่าเฉลี่ย มาทำการแบ่งสัญญาณเสียงเป็นส่วนย่อย (Frame) เป็น 4 ส่วน เมื่อได้ค่าเฉลี่ยแล้วจะมาทำการคำนวณหาค่าของแต่ละส่วนย่อย แล้วแปลงให้อยู่ในรูปอินพุทของ ART1 คือ อยู่ในรูปของเลขฐานสองโดยจะได้เลขฐานสองทั้งหมด 4 ชุด ชุดละ 7 Bit รวมทั้งหมด 28 Bit เก็บเป็นแฟ้มข้อมูลสำหรับฝึกสอน จำนวน 1 แฟ้ม แฟ้มข้อมูลสำหรับทดสอบกลุ่มที่ 1 จำนวน 1 แฟ้ม และแฟ้มข้อมูลสำหรับทดสอบกลุ่มที่ 2 จำนวน 1 แฟ้ม เพื่อจัดเตรียมเป็นข้อมูลอินพุทของกระบวนการฝึกสอนและการทดสอบของระบบ ในการหาค่าเฉลี่ยของเสียงแต่ละคนจะสามารถจัดทำกราฟเพื่อดู Slope ของเสียงได้ ดังรูปที่ 4.5 – 4.14

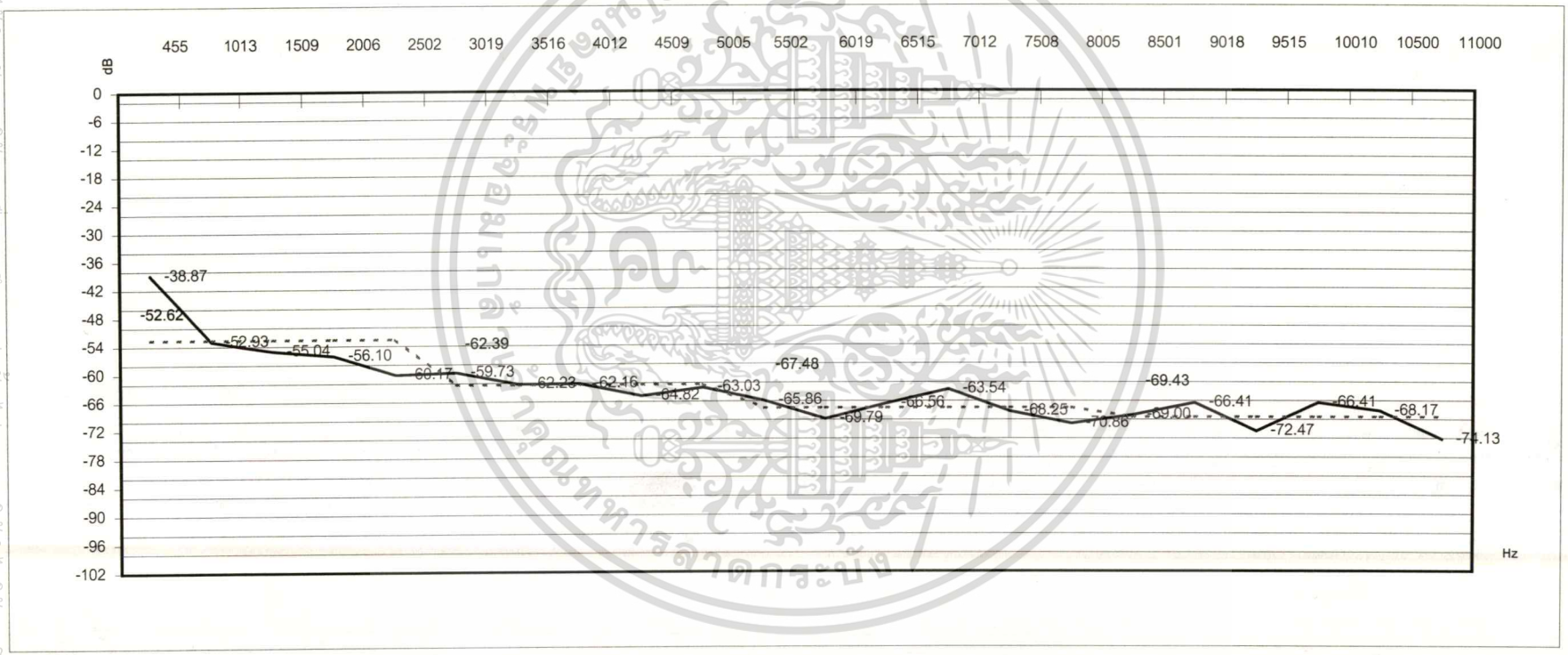
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-38.87	-52.93	-55.04	-56.10	-60.17	-59.73	-62.23	-62.16	-64.82	-63.03	-65.86	-69.79	-66.56	-63.54	-68.25	-70.86	-69.00	-66.41	-72.47	-66.41	-68.17	-74.13
-52.62	-52.62	-52.62	-52.62	-52.62	-62.39	-62.39	-62.39	-62.39	-62.39	-67.48	-67.48	-67.48	-67.48	-67.48	-67.48	-69.43	-69.43	-69.43	-69.43	-69.43	-69.43

Surashit

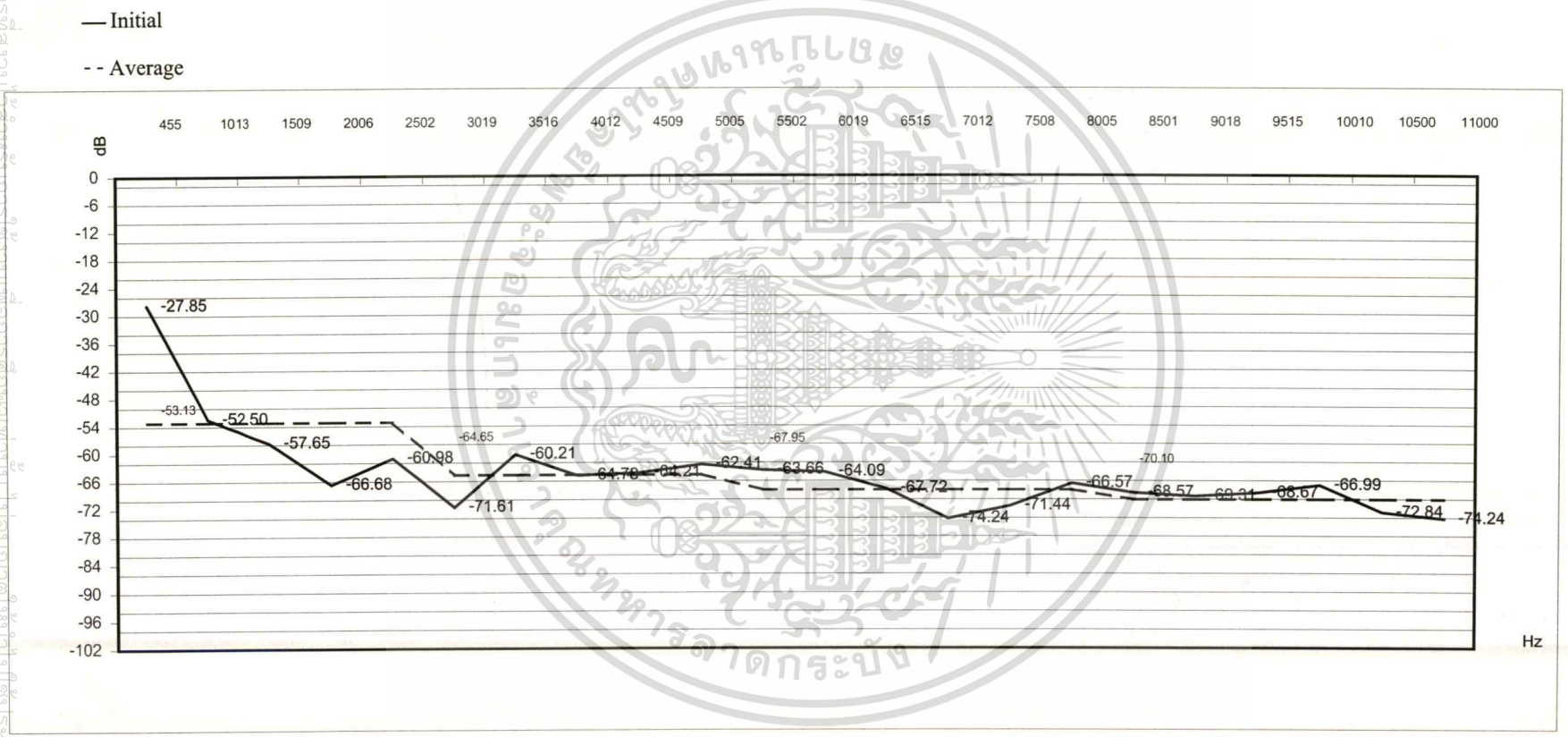
— Initial

- - Average



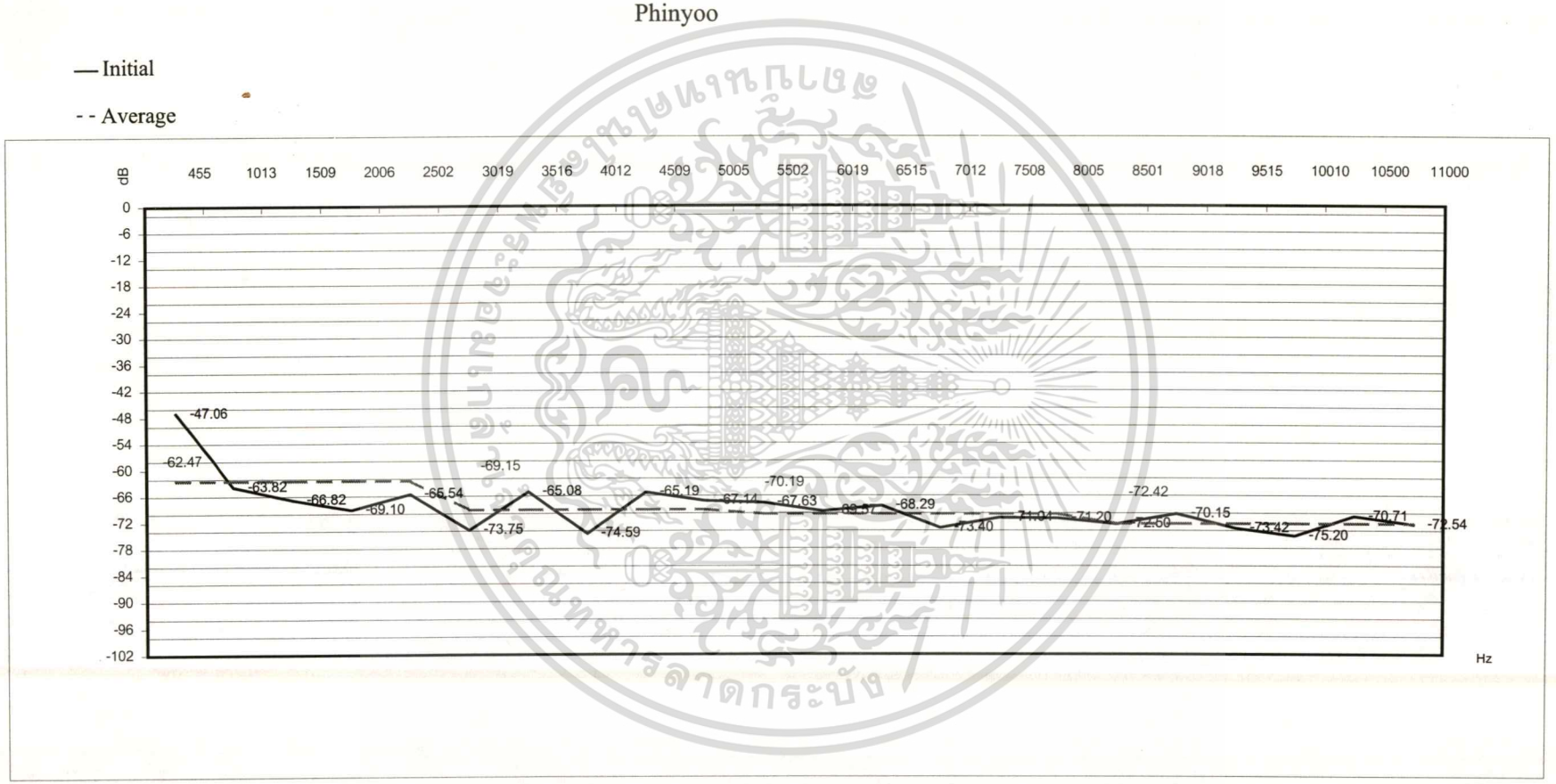
รูปที่ 4.5 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-27.85	-52.50	-57.65	-66.68	-60.98	-71.61	-60.21	-64.78	-64.21	-62.41	-63.66	-64.09	-67.72	-74.24	-71.44	-66.57	-68.57	-69.31	-68.67	-66.99	-72.84	-74.24
-53.13	-53.13	-53.13	-53.13	-53.13	-64.65	-64.65	-64.65	-64.65	-64.65	-67.95	-67.95	-67.95	-67.95	-67.95	-67.95	-70.10	-70.10	-70.10	-70.10	-70.10	-70.10



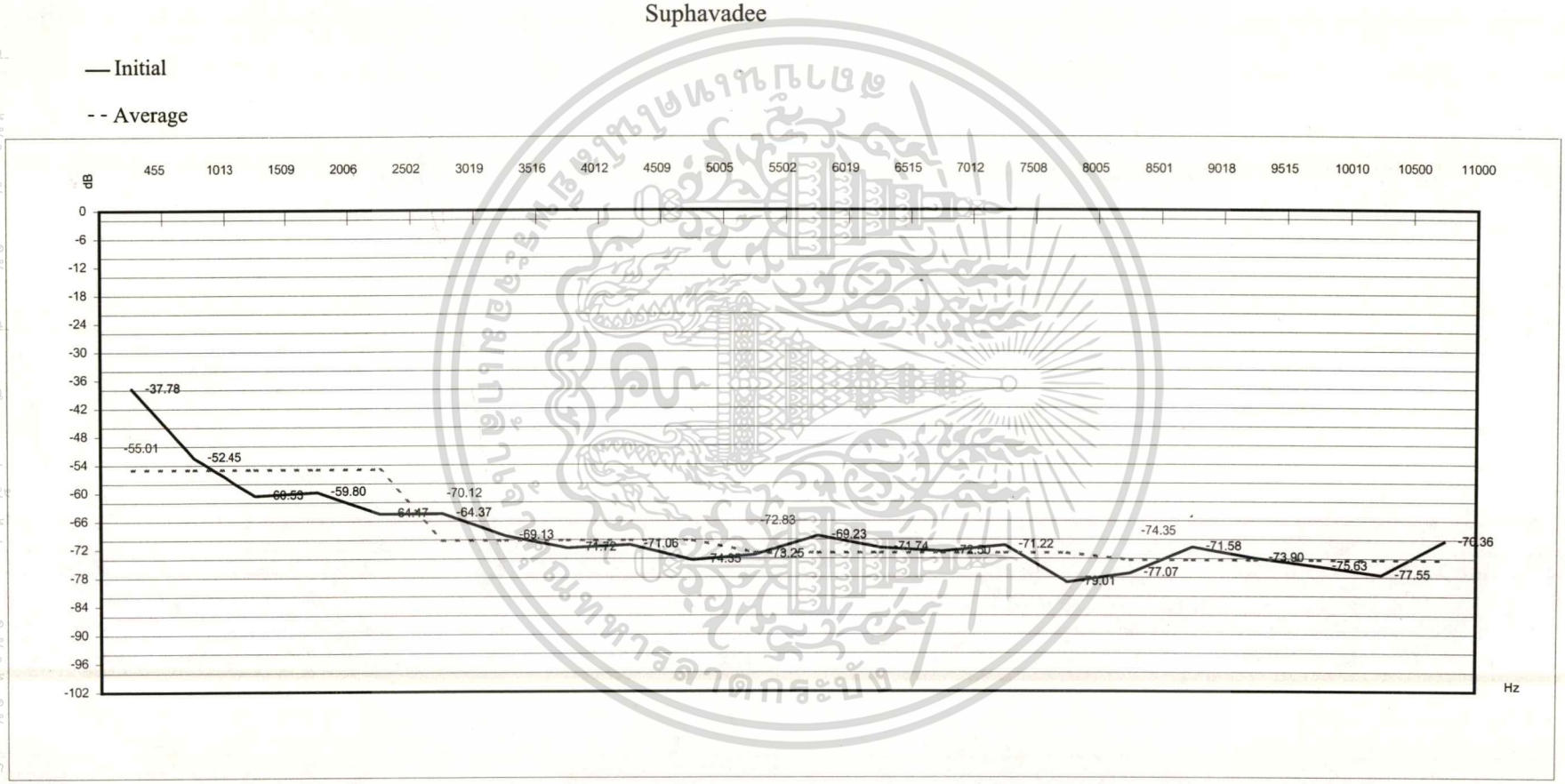
รูปที่ 4.6 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-47.06	-63.82	-66.82	-69.10	-65.54	-73.75	-65.08	-74.59	-65.19	-67.14	-67.63	-69.57	-68.29	-73.40	-71.04	-71.20	-72.50	-70.15	-73.42	-75.20	-70.71	-72.54
-62.47	-62.47	-62.47	-62.47	-62.47	-69.15	-69.15	-69.15	-69.15	-69.15	-70.19	-70.19	-70.19	-70.19	-70.19	-70.19	-72.42	-72.42	-72.42	-72.42	-72.42	-72.42



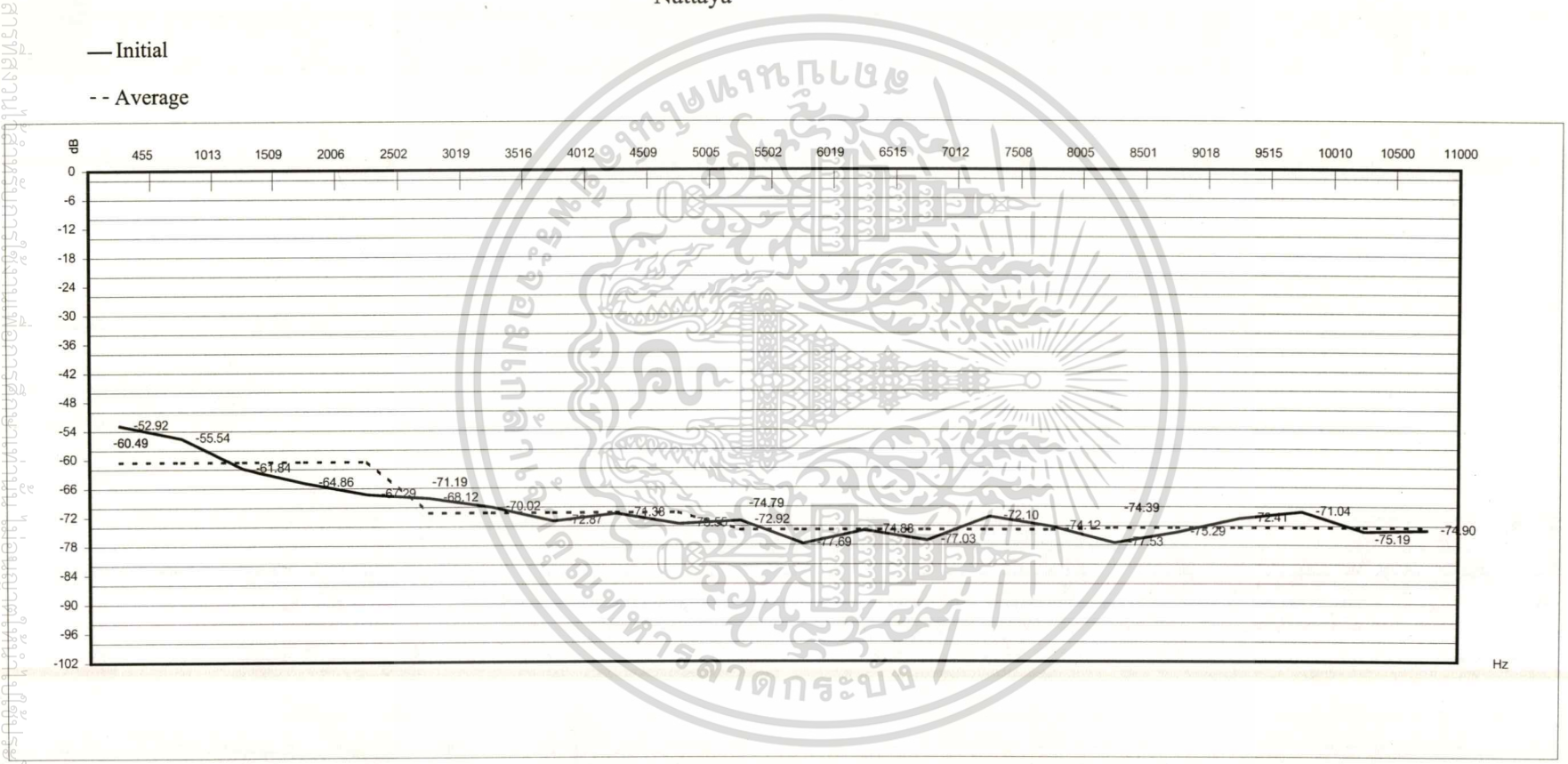
รูปที่ 4.7 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-37.78	-52.45	-60.53	-59.80	-64.47	-64.37	-69.13	-71.72	-71.06	-74.35	-73.25	-69.23	-71.74	-72.50	-71.22	-79.01	-77.07	-71.58	-73.90	-75.63	-77.55	-70.36
-55.01	-55.01	-55.01	-55.01	-55.01	-70.12	-70.12	-70.12	-70.12	-70.12	-72.83	-72.83	-72.83	-72.83	-72.83	-72.83	-74.35	-74.35	-74.35	-74.35	-74.35	-74.35



รูปที่ 4.8 แสดง Slope ของเสียงพูด

	455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
	-52.92	-55.54	-61.84	-64.86	-67.29	-68.12	-70.02	-72.87	-71.38	-73.55	-72.92	-77.69	-74.88	-77.03	-72.10	-74.12	-77.53	-75.29	-72.41	-71.04	-75.19	-74.90
	-60.49	-60.49	-60.49	-60.49	-60.49	-71.19	-71.19	-71.19	-71.19	-71.19	-74.79	-74.79	-74.79	-74.79	-74.79	-74.79	-74.39	-74.39	-74.39	-74.39	-74.39	-74.39



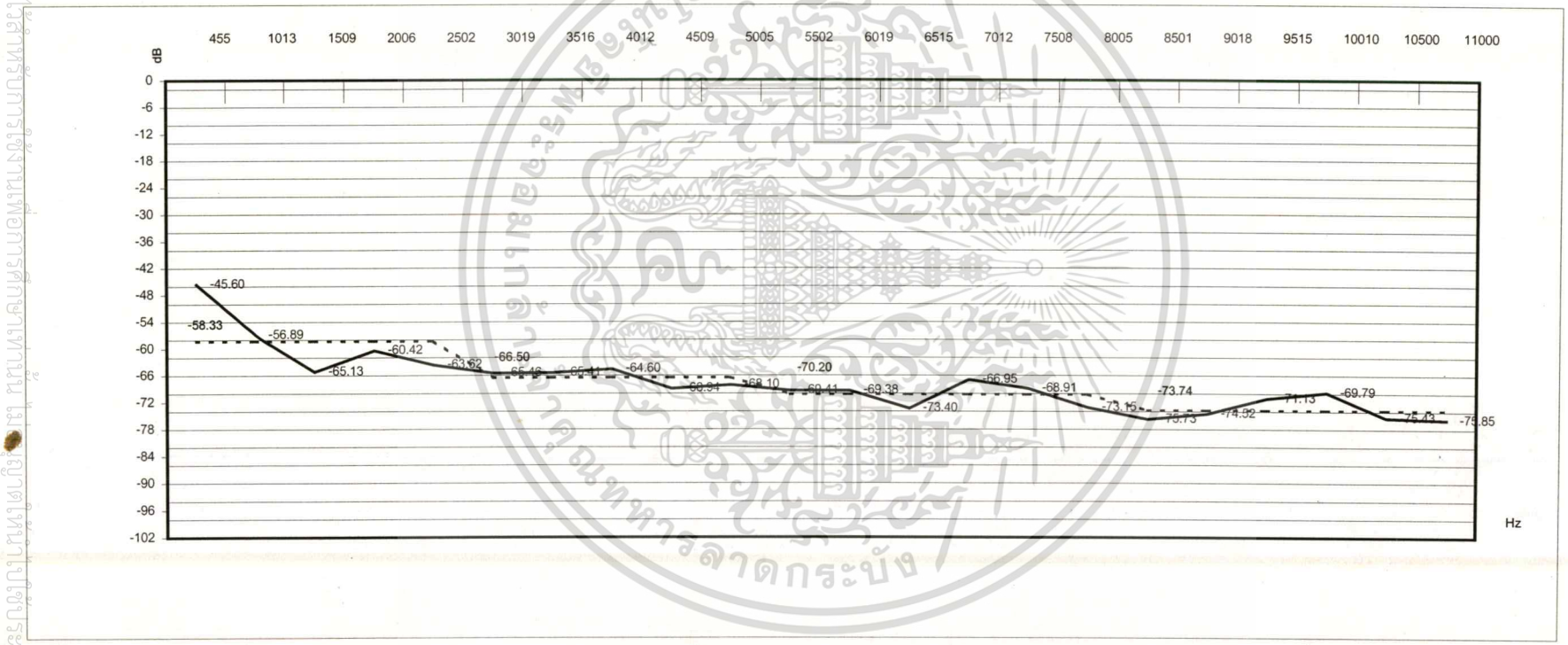
รูปที่ 4.9 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-45.60	-56.89	-65.13	-60.42	-63.62	-65.46	-65.41	-64.60	-68.94	-68.10	-69.41	-69.38	-73.40	-66.95	-68.91	-73.15	-75.73	-74.52	-71.13	-69.79	-75.43	-75.85
-58.33	-58.33	-58.33	-58.33	-58.33	-66.50	-66.50	-66.50	-66.50	-66.50	-70.20	-70.20	-70.20	-70.20	-70.20	-70.20	-73.74	-73.74	-73.74	-73.74	-73.74	-73.74

Siriporn

— Initial

- - Average



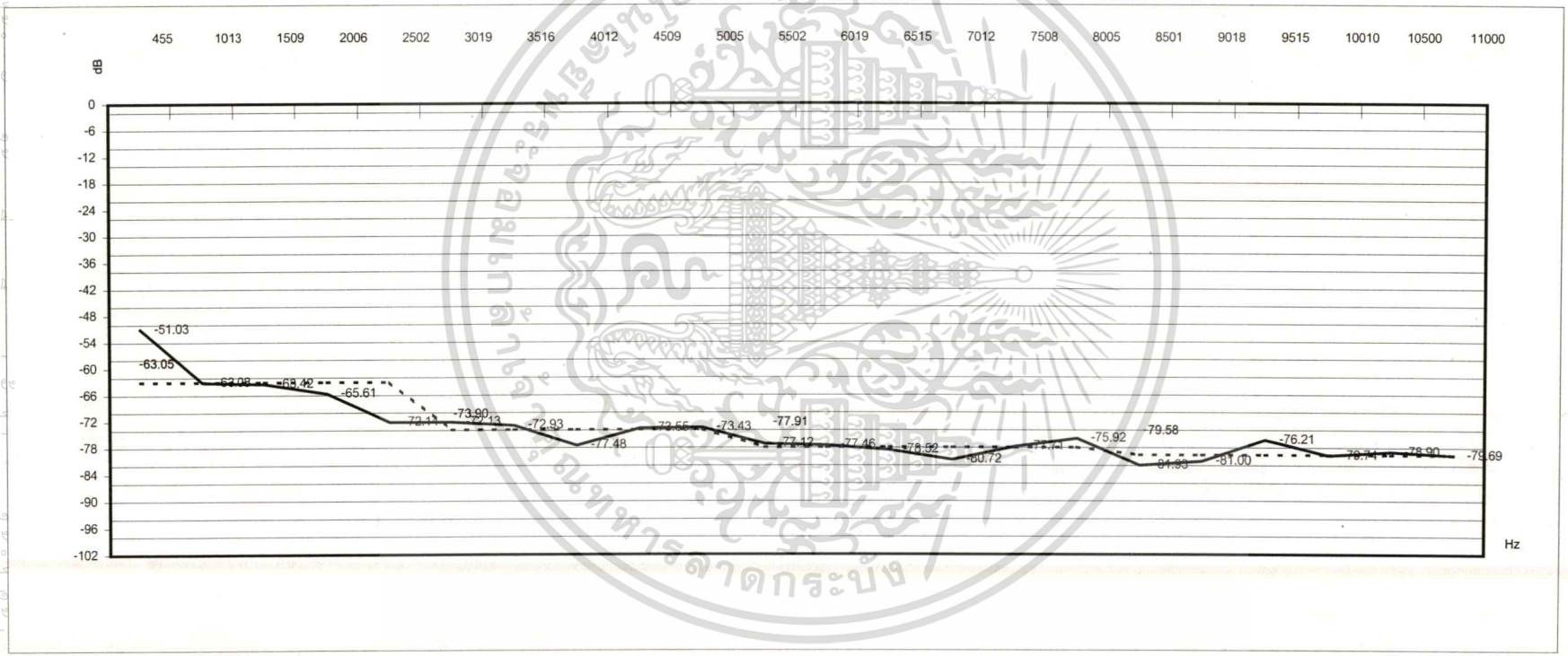
รูปที่ 4.10 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-51.03	-63.08	-63.42	-65.61	-72.14	-72.13	-72.93	-77.48	-73.55	-73.43	-77.12	-77.46	-78.52	-80.72	-77.71	-75.92	-81.93	-81.00	-76.21	-79.74	-78.90	-79.69
-63.05	-63.05	-63.05	-63.05	-63.05	-73.90	-73.90	-73.90	-73.90	-73.90	-77.91	-77.91	-77.91	-77.91	-77.91	-77.91	-79.58	-79.58	-79.58	-79.58	-79.58	-79.58

Phisal

— Initial

- - Average



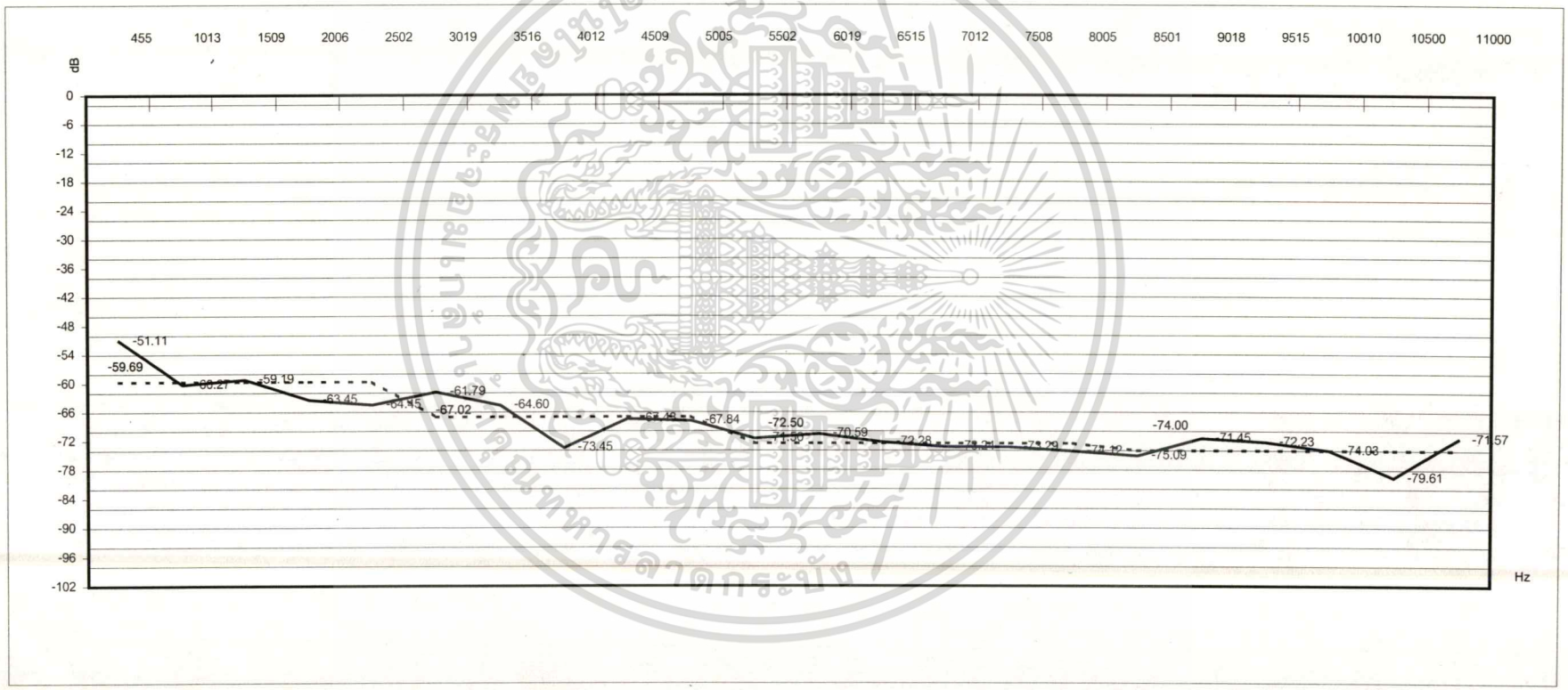
รูปที่ 4.11 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-51.11	-60.27	-59.19	-63.45	-64.45	-61.79	-64.60	-73.45	-67.43	-67.84	-71.50	-70.59	-72.28	-73.24	-73.29	-74.12	-75.09	-71.45	-72.23	-74.03	-79.61	-71.57
-59.69	-59.69	-59.69	-59.69	-59.69	-67.02	-67.02	-67.02	-67.02	-67.02	-72.50	-72.50	-72.50	-72.50	-72.50	-72.50	-74.00	-74.00	-74.00	-74.00	-74.00	-74.00

Kanyarat

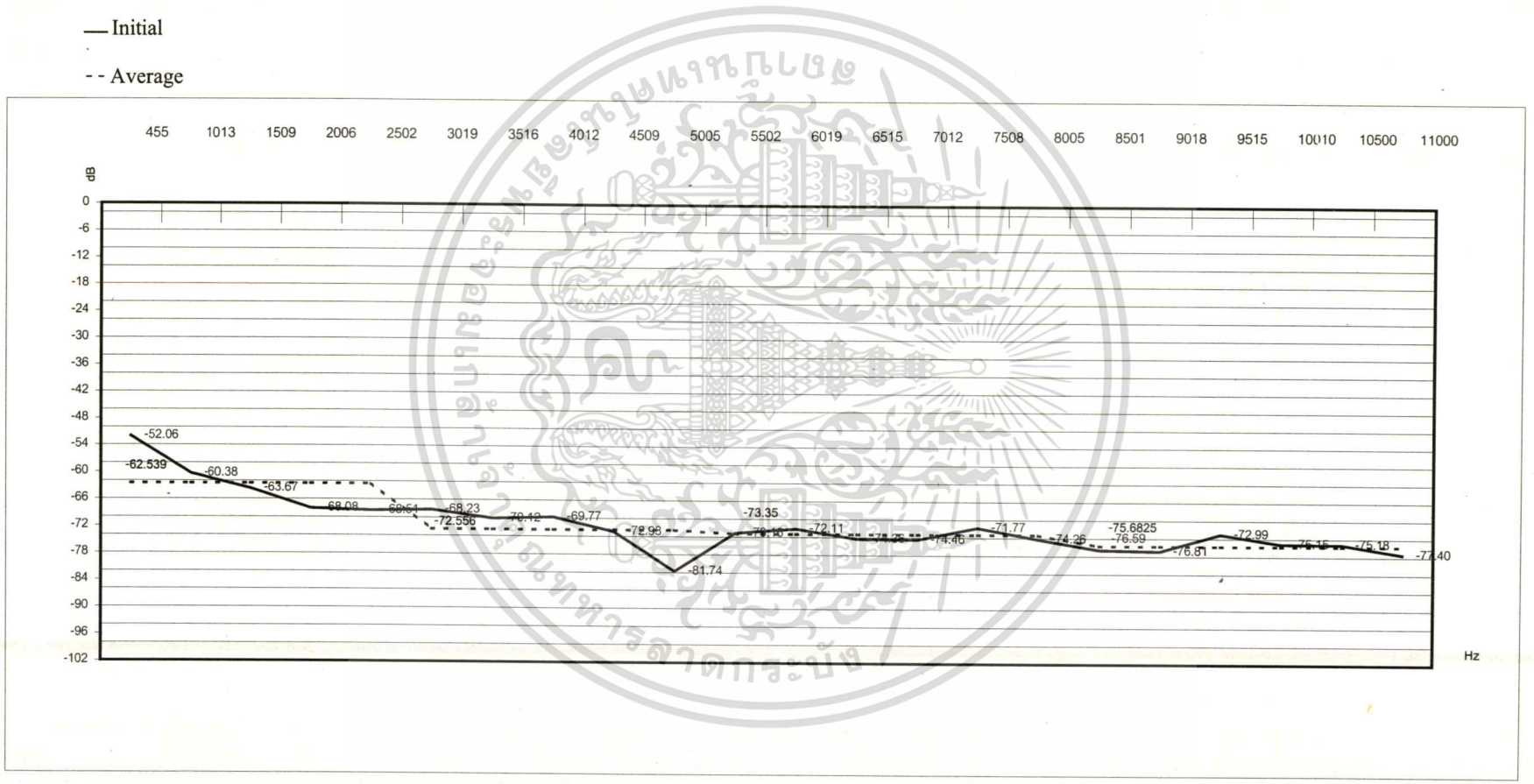
— Initial

- - Average



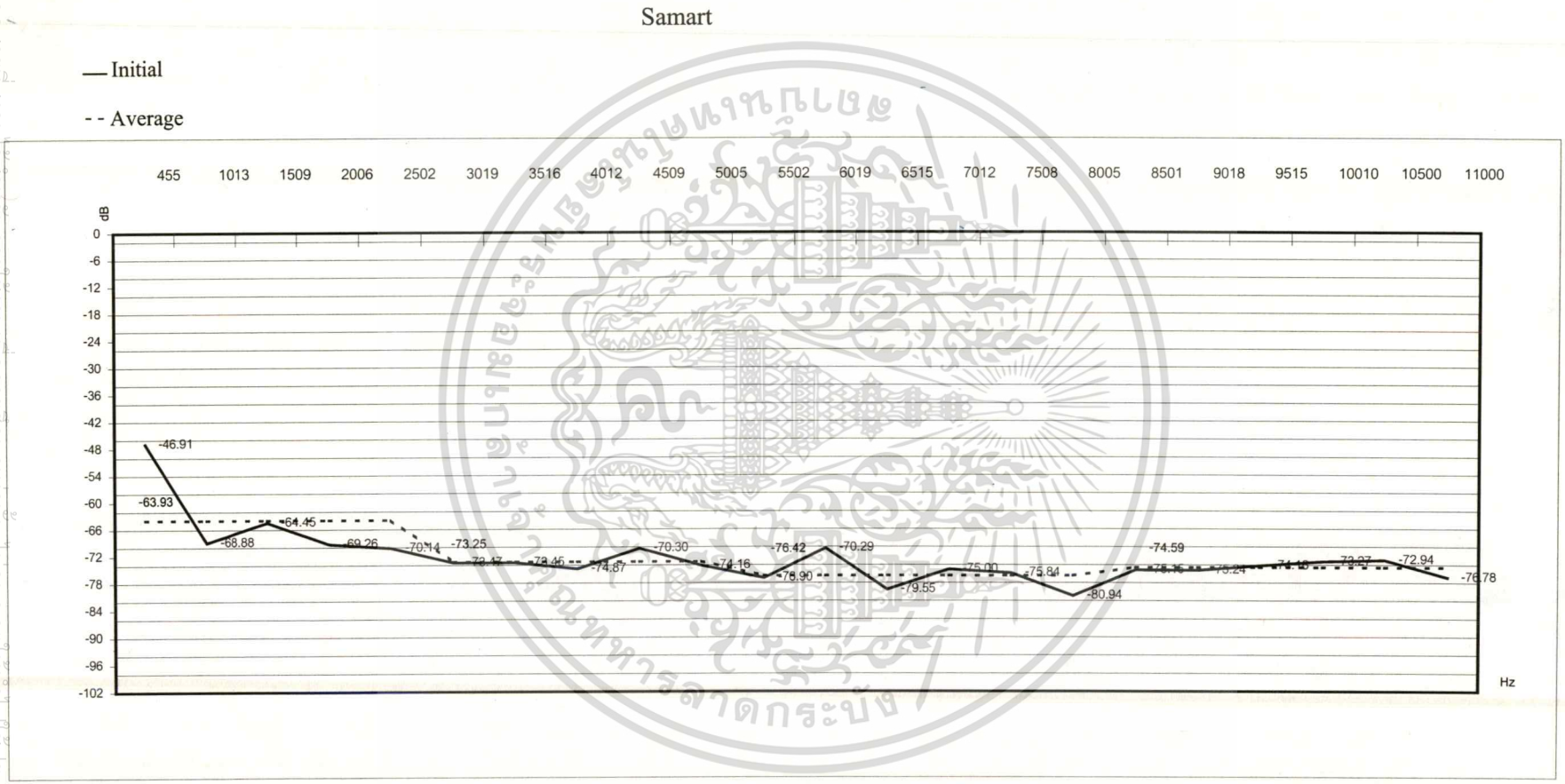
รูปที่ 4.12 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-52.06	-60.38	-63.67	-68.08	-68.51	-68.23	-70.12	-69.77	-72.93	-81.74	-73.16	-72.11	-74.36	-74.46	-71.77	-74.26	-76.59	-76.81	-72.99	-75.15	-75.18	-77.40
-62.54	-62.54	-62.54	-62.54	-62.54	-72.56	-72.56	-72.56	-72.56	-72.56	-73.35	-73.35	-73.35	-73.35	-73.35	-73.35	-75.68	-75.68	-75.68	-75.68	-75.68	-75.68



รูปที่ 4.13 แสดง Slope ของเสียงพูด

455	1013	1509	2006	2502	3019	3516	4012	4509	5005	5502	6019	6515	7012	7508	8005	8501	9018	9515	10010	10500	11000
-46.91	-68.88	-64.45	-69.26	-70.14	-73.47	-73.45	-74.87	-70.30	-74.16	-76.90	-70.29	-79.55	-75.00	-75.84	-80.94	-75.15	-75.24	-74.18	-73.27	-72.94	-76.78
-63.93	-63.93	-63.93	-63.93	-63.93	-73.25	-73.25	-73.25	-73.25	-73.25	-76.42	-76.42	-76.42	-76.42	-76.42	-76.42	-74.59	-74.59	-74.59	-74.59	-74.59	-74.59



รูปที่ 4.14 แสดง Slope ของเสียงพูด

4.2 ขั้นตอนการเรียนรู้ (ฝึกสอน)

ในขั้นตอนการเรียนรู้จะเป็นการเรียนรู้แบบมีผู้สอน โดยจะทำการส่งข้อมูลกลุ่มที่ 1 ซึ่งเป็นข้อมูลอินพุทของชุดฝึกสอนจากขั้นตอนการจัดเตรียมข้อมูล เข้าไปยังโครงข่าย ผ่านกระบวนการของ ART1 Algorithm กระบวนการจะทำการเรียนรู้และจำข้อมูลอินพุทไว้ โดยข้อมูลอินพุทในการฝึกฝนโครงข่าย จะมีข้อมูลของกลุ่มข้อมูลสำหรับฝึกสอน ดังนี้

Input 1 : 0 1 1 0 1 0 1 0 1 1 1 1 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 0

Input 2 : 0 1 1 0 1 0 1 1 0 0 0 0 0 1 1 0 0 0 1 0 0 1 0 0 0 1 1 0

Input 3 : 0 1 1 1 1 1 0 1 0 0 0 0 1 0 1 1 0 0 0 1 1 0 1 0 0 1 0 0 1

Input 4 : 0 1 1 0 1 1 1 1 0 0 0 1 1 0 1 0 0 1 0 0 1 1 0 0 1 0 1 0

Input 5 : 0 1 1 1 1 0 0 1 0 0 0 0 1 1 1 1 0 0 1 0 1 1 1 1 0 0 1 0 1 0

Input 6 : 0 1 1 1 0 1 0 1 0 0 0 0 0 1 1 1 0 0 0 1 1 0 1 0 0 1 0 1 0

Input 7 : 0 1 1 1 1 1 1 1 1 0 0 1 0 1 0 1 0 0 1 1 1 0 1 0 1 0 0 0 0

Input 8 : 0 1 1 0 1 1 1 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 0 1 0 0 0 1 0 1

Input 9 : 0 1 1 1 1 1 1 1 1 0 0 1 0 0 1 1 0 0 1 0 0 1 1 0 0 1 1 0 0

Input 10 : 1 0 0 0 0 0 0 1 0 0 1 0 0 1 1 0 0 1 1 0 1 1 0 0 1 0 1 1

ข้อมูลอินพุทจะมี 10 Record คือ 1 Record เป็นข้อมูลเสียงของคน 1 คน หลังจากที่ส่งข้อมูลอินพุทเข้าไปแล้วจะได้เอาท์พุทจากการทำงานของโครงข่ายและจะได้ผลลัพธ์ที่สามารถบอกได้ว่าโหนด ใดชนะ ดังนี้

โหนดชนะ Input 1 : 1000000000

โหนดชนะ Input 2 : 0100000000

โหนดชนะ Input 3 : 0010000000

โหนดชนะ Input 4 : 0001000000

โหนดชนะ Input 5 : 0000100000

โหนดชนะ Input 6 : 0000010000

โหนดชนะ Input 7 : 0000001000

โหนดชนะ Input 8 : 0000000100

โหนดชนะ Input 9 : 0000000010

โหนดชนะ Input 10 : 0000000001

จากผลลัพธ์ที่ได้เลข 1 หมายถึง โหนดที่ถูกฝึกสอนให้โครงข่ายเรียนรู้ และผลลัพธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่ได้จากการเรียนรู้ของโครงข่ายแล้ว สามารถบอกจำนวนรอบของการเรียนรู้ของโครงข่าย โดยการปรับค่าน้ำหนัก ดังนี้

1. 1000000000
2. 1000000000
3. 0100000000
4. 0100000000
5. 0010000000
6. 0100000000
7. 0010000000
8. 0001000000
9. 0001000000
10. 0000100000
11. 0010000000
12. 0000100000
13. 0000010000
14. 0000010000
15. 0000001000
16. 0100000000
17. 0010000000
18. 0000001000
19. 0000000100
20. 0100000000
21. 0001000000
22. 0000100000
23. 0000001000
24. 0000000100
25. 0000000010
26. 0000100000
27. 0000000010
28. 0000000001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่ 1 เป็นผลที่ได้จากการฝึกสอน Input 1 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 1 รอบ

ลำดับที่ 2 - 3 เป็นผลที่ได้จากการฝึกสอน Input 2 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 2 รอบ

ลำดับที่ 4 - 5 เป็นผลที่ได้จากการฝึกสอน Input 3 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 2 รอบ

ลำดับที่ 6 - 8 เป็นผลที่ได้จากการฝึกสอน Input 4 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 3 รอบ

ลำดับที่ 9 - 10 เป็นผลที่ได้จากการฝึกสอน Input 5 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 2 รอบ

ลำดับที่ 11 - 13 เป็นผลที่ได้จากการฝึกสอน Input 6 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 3 รอบ

ลำดับที่ 14 - 15 เป็นผลที่ได้จากการฝึกสอน Input 7 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 2 รอบ

ลำดับที่ 16 - 19 เป็นผลที่ได้จากการฝึกสอน Input 8 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 4 รอบ

ลำดับที่ 20 - 25 เป็นผลที่ได้จากการฝึกสอน Input 9 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 6 รอบ

ลำดับที่ 25 - 28 เป็นผลที่ได้จากการฝึกสอน Input 10 โครงข่ายทำการเรียนรู้โดยการปรับค่าน้ำหนัก จำนวน 3 รอบ

อินพุทที่ได้เรียนรู้ของโครงข่ายจะถูกเก็บไว้เป็นข้อมูลอ้างอิงต่อไป

4.3 ขั้นตอนการทดสอบ

ในขั้นตอนการทดสอบโครงข่าย จะทดสอบโดยส่งข้อมูลอินพุทเข้าไปในโครงข่ายผ่านกระบวนการ ART1 Algorithm เพื่อทดสอบว่าโครงข่ายสามารถจดจำและแยกแยะได้หรือไม่ว่าข้อมูลอินพุทที่รับเข้าไปใหม่นั้นตรงหรือใกล้เคียงมากที่สุดกับข้อมูลที่โครงข่ายได้เก็บไว้เป็นข้อมูลอ้างอิงจากขั้นตอนของการเรียนรู้ ในการทดลองนี้ใช้กลุ่มข้อมูลการทดสอบจำนวน 2 กลุ่ม กลุ่มทดสอบที่ 1 เป็นกลุ่มข้อมูลเสียงที่เก็บข้อมูลเสียงของกลุ่มตัวอย่างที่โครงข่ายเคยได้รับการฝึกสอนไปแล้ว โดยให้กลุ่มตัวอย่างพูดคำเดิม แล้วนำมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผ่านกระบวนการเตรียมข้อมูลนำมาเป็นอินพุทเพื่อทำการทดสอบโครงข่าย โดยใช้อัตราในการวัดค่าความคลายในการจำ 85% ผลที่ได้จากการทดสอบเป็นดังนี้

โหนดชนะ Input 1 : 1000000000

โหนดชนะ Input 2 : 0100000000

โหนดชนะ Input 3 : 0010000000

โหนดชนะ Input 4 : 00001 00000

โหนดชนะ Input 5 : 0000001000

จากเอาพุทที่ได้จะเห็นว่า โครงข่ายสามารถแยกแยะเสียงได้จำนวน 5 เสียงแต่จดจำเสียงของกลุ่มตัวอย่างเสียงได้เพียง 3 เสียง ส่วนอีก 2 เสียงระบบไม่สามารถบอกได้ว่าเป็นเสียงของใคร

กลุ่มทดสอบที่ 2 เป็นกลุ่มข้อมูลเสียงที่เก็บข้อมูลเสียงของกลุ่มตัวอย่างที่โครงข่ายไม่เคยได้รับการฝึกสอนมาก่อน โดยให้กลุ่มตัวอย่างพูดคำเดียวกันกับกลุ่มที่ 1 แล้วนำมาผ่านกระบวนการเตรียมข้อมูลนำมาเป็นอินพุทเพื่อทำการทดสอบโครงข่าย โดยใช้อัตราความแม่นยำในการจำ 85% โดยมีข้อมูลอินพุท ดังนี้

Input 1 : 0 1 1 0 0 1 1 0 1 1 1 1 1 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0

Input 2 : 0 1 0 1 1 0 0 0 1 1 1 1 0 0 1 0 0 0 0 1 0 1 0 0 1 1 1 0

Input 3 : 0 1 1 0 0 0 1 1 0 0 0 1 1 1 1 0 1 0 0 0 0 1 0 0 1 1 1 1

Input 4 : 0 1 1 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 0 1 1 1 1 1 0

Input 5 : 0 1 1 1 0 0 1 1 0 0 0 0 1 1 1 0 0 1 0 0 1 1 0 0 0 1 1 1

ข้อมูลอินพุทในการทดสอบจะมี 5 Record คือ 1 Record เป็นข้อมูลเสียงของคน 1 คน หลังจากที่ส่งข้อมูลอินพุทเข้าไปทำการทดสอบแล้วจะได้เอาท์พุทจากการทำงานของโครงข่ายและจะได้ผลลัพธ์ที่สามารถบอกได้ว่าโหนด ใดชนะ ดังนี้

โหนดชนะ Input 1 : 000000000010000

โหนดชนะ Input 2 : 000000000001000

โหนดชนะ Input 3 : 000000000000100

โหนดชนะ Input 4 : 000000000000010

โหนดชนะ Input 5 : 000000000000001

จากเอาพุทที่ได้จะเห็นว่าโหนดชนะ คือ โหนดที่ 11, 12, 13, 14, 15 ตามลำดับโดย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โหนดชนะจะไม่ไปซ้ำกับโหนดที่ 1 – 10 เนื่องจาก เป็นข้อมูลที่ไม่เคยถูกฝึกสอนให้กับโครงข่ายมาก่อน ดังนั้นโครงข่ายจึงสามารถเรียนรู้และจดจำข้อมูลใหม่เข้าไปได้ ทั้งหมด 5 เสียง เพื่อเก็บเป็นข้อมูลอ้างอิงต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้ เป็นการศึกษาเกี่ยวกับระบบการรู้จำเสียงคำโดด บนพื้นฐานของแบบจำลอง Adaptive Resonance Theory 1 (ART1) ระบบได้ถูกออกแบบให้ได้รับการฝึกสอนเพื่อหาต้นแบบอ้างอิงและเมื่อสิ้นสุดการฝึกสอนจะทำการทดสอบเพื่อหาอัตราความแม่นยำในการรู้จำ โดยส่งข้อมูลสำหรับทดสอบระบบและระบบก็สามารถให้เอาท์พุทได้สอดคล้องกับเป้าหมายและตัวอย่างเสียงที่เคยฝึกสอนไปแล้ว

5.1 ผลการทดลอง

จากการทดลองในบทที่ 4 เป็นการทดลองนำข้อมูลอินพุทเข้าสู่ระบบเพื่อให้ทำการเรียนรู้ในลักษณะของการเรียนรู้แบบมีผู้ฝึกสอนและทำการทดสอบด้วยชุดข้อมูลที่เตรียมไว้ ในการทดลองนี้เป็นการทดลองเพื่อหาอัตราการรู้จำคำโดดและลักษณะของข้อมูลเสียงที่นำมาใช้เป็นอินพุทจะเก็บมาจาก Slope เพียงอย่างเดียว ไม่มี Feature อื่น ซึ่งผลการทดลองมีดังนี้

5.1.1 การฝึกสอนโดยใช้ข้อมูลอินพุทของคน 10 คน เปล่งเสียง สามารถแยกความแตกต่างและเรียนรู้จำข้อมูลอินพุทได้ 10 คน โดยใช้ค่าที่ใช้วัดความคล้าย (ρ) 85%

5.1.2 ในการทดสอบโดยใช้ข้อมูลอินพุทจากกลุ่มทดสอบกลุ่มที่ 1 จำนวน 5 คน ระบบสามารถบอกว่ามีในระบบแล้ว 3 คน ส่วนอีก 2 คนไม่สามารถบอกได้ว่าเป็นใคร ดังนั้นอัตราความถูกต้องของการรู้จำในการทดสอบครั้งนี้ 60% โดยใช้ค่าที่ใช้วัดความคล้าย (ρ) 85%

5.1.3 ในการทดสอบโดยใช้ข้อมูลอินพุทจากกลุ่มทดสอบกลุ่มที่ 2 จำนวน 5 คน ระบบสามารถบอกได้ว่า ทั้ง 5 คนไม่มีในระบบและทำการเรียนรู้และจำข้อมูลใหม่เพิ่มเข้าไปในระบบ โดยใช้ค่าที่ใช้วัดความคล้าย (ρ) 85%

5.2 สรุปผลการทดลอง

จากการทดลอง โดยใช้แบบจำลองของ ART1 ในการรู้จำเสียง โดยเป็นวิธีการเรียนรู้แบบมีผู้สอน ซึ่งจากการเรียนรู้จำนี้ระบบจะจำจากการเรียนรู้จุดที่ซ้ำกันของ Pattern โดยมีการปรับค่าน้ำหนักไปเรื่อย ๆ เพื่อหารูปแบบที่ซ้ำกัน ถ้าหากมีรูปแบบที่ซ้ำกันจะถือว่าเป็นรูปแบบเดียวกัน แต่ถ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปรับค่าน้ำหนักแล้วไม่ซ้ำกับรูปแบบใด ระบบจะถือว่าเป็นรูปแบบใหม่และจะเรียนรู้จำรูปแบบใหม่นี้ โดยการรู้จำจะขึ้นอยู่กับค่าที่ใช้วัดความคล้าย (ρ) โดยให้ระบบวัดค่าความคล้ายจากจุดที่ซ้ำกัน ซึ่งจากการทดลองหลาย ๆ ครั้ง จะพบว่า ถ้ามีคนอยู่ 10 ระบบสามารถรู้จำและแยกความแตกต่างได้ 9 คน หรือ 10 คน ทั้งนี้ขึ้นอยู่กับการกำหนดค่าที่ใช้วัดความคล้าย (ρ) และความแตกต่างของเสียงพูดของคนในแต่ละครั้ง รวมทั้ง Feature ที่ใส่เข้าไปในข้อมูลเสียงด้วย

จากการทดลองการรู้จำ โดยใช้แบบจำลอง ART1 นั้น พบว่าระบบใช้เวลาในการเรียนรู้น้อย ซึ่งเมื่อเปรียบเทียบกับแบบจำลองอื่นแล้วจะเรียนรู้ได้เร็วกว่า และสามารถเรียนรู้รูปแบบใหม่ได้เอง แต่ยังมีข้อผิดพลาดในการแยกแยะเสียงที่มีลักษณะใกล้เคียงกันมากไม่ได้

5.3 ข้อเสนอแนะ

5.3.1 การทดลองในโครงงานนี้เป็นการเก็บลักษณะเสียงจาก Slope เพียงอย่างเดียว ซึ่งมีผลทำให้ลักษณะเด่นของเสียงพูดไม่ชัดเจน เกิดปัญหาในการแยกแยะรูปแบบของเสียง ดังนั้น ถ้าจะให้ได้ผลดี ควรจะต้องเพิ่มลักษณะของเสียงอย่างอื่นเข้าไปด้วย เช่น Time, Pitch, Frequency เป็นต้น

5.3.2 การทดลองในโครงงานนี้ มีการเก็บข้อมูลเสียงเพื่อฝึกสอนในการเรียนรู้จำนวนไม่มาก ซึ่งทำให้เกิดปัญหาในการทดสอบ ดังนั้นควรเก็บข้อมูลเพื่อการฝึกสอนจำนวนหลาย ๆ คน

5.3.3 ในการหาจุดเริ่มต้นและจุดสิ้นสุดของเสียงนั้น จะทำให้เกิดข้อผิดพลาดได้ เช่น ถ้าตัดจุดสิ้นสุดสั้นไป จะทำให้ลักษณะของเสียงผิดเพี้ยนไปด้วย ทำให้การวัดความคล้ายผิดพลาด

5.3.4 เพื่อให้การทดลองได้ผลดีควรใช้ค่าในการวัดความคล้าย (ρ) ที่แตกต่างกัน เพื่อหาอัตราการรู้จำที่เหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

ชัย วุฒิวิวัฒน์ชัย. 2540. "การรู้จำเสียงคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวิร์ค." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

จิตรลดา จารุมิศร์. 2542 "การออกแบบ แบบจำลองในการรู้จำเสียงวรรณยุกต์สำหรับภาษาไทยโดยใช้เทคนิคการควอนไตซ์พิทช์ และ Hidden Markov Modeling". วิทยานิพนธ์ศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

เสรี ปานซาง. 2540. "การรู้จำเสียงพูดคำไทยแบบไม่ขึ้นกับผู้พูดด้วยนิวรัลเน็ตเวอร์ค." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

สมรัฐ แถลงการณ์. 2540. "การวิเคราะห์ซอฟท์แวร์โครงขายประสาทเทียมเพื่อการรู้จำตัวอักษรและตัวเลข." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, มหาวิทยาลัยเชียงใหม่.

วิศรุต อาขุบุตร. 2539. "ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ." วิทยานิพนธ์วิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

A.Syrdal, R. Bennett, S. Greenspan. 1995. **Appied speech technology**. United States of America : CRC Press.

Cihan H.Dagli. 1994. **Artificial Neural Networks of Intelligent Manufacturing**. London : Chapman & Hall.

Gordon E. Pelton. 1992. **Voice Processing**. United States of America : McGraw-Hill, Inc.

Igor Aleksander. 1989. **Neural Computing Architecture**. London : North Oxford Academic Publishers LTd.

James A. Freeman, David M. Skapura. 1991. **Neural Networks Algorithms, Applications, and Programming Techniques**. United State of America: Addison-Wesley Publishing Company, Inc.

Valluru Rao & Hayagriva Rao. 1995. **C++ Neural Networks and Fuzzy Logic**. 2nd ed. United States of Amerrica: Henry Holt and Company, Inc.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นางสาวภิญญู กำเนิดหล่ม เกิดวันที่ 14 กันยายน 2511 ที่อำเภอภูกระดึง จังหวัดเลย สำเร็จการศึกษาปริญญาตรีครุศาสตร์บัณฑิต สาขาคอมพิวเตอร์ศึกษา จากวิทยาลัยครูมหาสารคาม ในปีการศึกษา 2535 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปี พ.ศ. 2542

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้