ปริญญานิพนธ์ปีการศึกษา 2536

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การประยุกต์ใช้งานโครงข่ายประสาทเทียม

ผู้จัดทำ

1. นาย ศักดิ์ชัย วิริยะคุณเกษม
2. นาย สมศักดิ์ บรรลือศรีศักดิ์

..... อาจารย์ที่ปรึกษา

(ดร. บุญธีร์ เครือตราชู)

การประยุกต์ใช้งานโครงข่ายประสาทเทียม

Neural Networks Application

โดย

1. นาย ศักดิ์ชัย วิริยะคุณเกษม 33100370
2. นาย สมศักดิ์ บรรลือศรีศักดิ์ 33100402

อาจารย์ที่ปรึกษา

ดร. บุญธีร์ เครือตราชู

บทคัดย่อ

การประยุกต์ใช้งานโครงข่ายประสาทเทียมเป็นการนำทฤษฎี และหลักการของโครงข่ายประสาทเทียม มาประยุกต์ใช้ในการจดจำเสียงพูด การทำงานของระบบจดจำเสียงพูดจะแบ่งออกเป็นสองส่วนงาน คือส่วนการเรียนรู้เสียงที่จะเริ่มต้นด้วยการเก็บตัวอย่างเสียงนำมาผ่านกระบวนการเพื่อแปลงให้สัญญาณที่มีฟังก์ชั่นในโดเมนเวลา ให้เป็นฟังก์ชันในโดเมนความถี่ และเวลาโดยใช้ FFT (Fast Fourier Transform) เพื่อเป็นข้อมูลในการเรียนรู้ของโครงข่ายประสาทเทียม ในงานที่สองจะเป็นการนำโครงข่ายประสาทเทียมที่ผ่านการเรียนรู้มาใช้แยกเสียง

Abstract

Using Neural Networks Application, it applies theories and principles of neural network to voice recognization. Voice recoginzation system devides into two parts. The former is voice learning which takes the example voice, by using Fast Fourier Transform, through time to frequency–time domain transformation process. By this reason, we will get voice learning data. The later is the learning neural network voice classification.

สารบัญ

			หน้า
บทที่ 1		บทนำ	1
บทที่ 2		ทฤษฎีและหลักการเบื้องต้น	4
	2.1	การจดจำเสียงพูด	4
	2.2	หลักการทำงานของการจดจำ	6
	2.3	การตรวจจับจุดสิ้นสุด	6
บทที่ 3		การแปลงฟูเรียร์	8
	3.1	หลักการเบื้องต้นในการแปลงฟูเรียร์	8
	3.2	การแปลงฟูเรียร์แบบเร็ว	9
	3.3	การแปลงกลับของฟูเรียร์	11
บทที่ 4		โครงข่ายประสาทเทียม	12
	4.1	ความรู้ทั่วไป	12
	4.2	แบบจำลองทางคณิตศาสตร์ของเซลประสาทเดียว	15
	4.3	เทคนิคการเรียนรู้แบบแพร่กลับ	16
บทที่ 5		ระบบการจดจำเสียง	21
	5.1	แนวทางการพัฒนา	25
	5.2	โมดูลที่ช่วยในการบันทึกเสียง	27
	5.3	โมดูลที่ช่วยในการแปลงฟาสท์ฟูเรียร์ทรานฟอร์ม	28
	5.4	โมดูลที่ช่วยในการหาพลังงานของเสียง	29
	5.5	โมดูลที่ช่วยแสดงผลทางกราฟฟิก	29

หน้า

บทที่ 5 ระบบการจดจำเสียง 21

5.6 โมดูลที่ช่วยในการพิมพ์ภาพออกเครื่องพิมพ์ 29

5.7 โปรแกรมเชื่อมต่อการใช้งานกับผู้ใช้ 30

5.8 อัลกอริทึมที่ช่วยในการตัดคำ 30

5.9 การเก็บข้อมูลตัวอย่าง 31

5.10 ขั้นตอนในการเรียนรู้ 32

บทที่ 6 ผลการทดลอง 33

6.1 ผลจากการทดสอบ 33

6.2 เกณฑ์การวัดประสิทธิภาพ 33

บทที่ 7 สรุปและวิจารณ์ 36

บทที่ 8 แนวทางการพัฒนาต่อไป 38

สารบัญรูปภาพ

		หน้า
รูปที่ 4.1	ส่วนประกอบของเซลประสาท	13
รูปที่ 4.2	แบบจำลองทางคณิตศาสตร์ของเซลประสาทเดียว	14
รูปที่ 4.3	One Layer Backpropagation Neural Network System	17
รูปที่ 4.4	การปรับตัวของ weight	20
รูปที่ 5	Voice Recognization System	22
รูปที่ 5.1	Error of word Seperator	30
รูปที่ 6.1	Error of word seperator	35
รูปที่ 6.2	Error of word seperator	35

สารบัญตาราง

หน้า

ตารางที่ 4.1 การพัฒนาของโครงข่ายประสาทเทียม 14

บทที่ 1
บทนำ

การทำงานในชีวิตประจำวันส่วนใหญ่ มักจะอาศัยความชำนาญของมนุษย์เพื่อให้งานเสร็จเรียบร้อย เนื่องจากมนุษย์สามารถติดต่อสื่อสารกันได้ สามารถแลกเปลี่ยนความรู้กันได้ นอกจากนี้มนุษย์ยังมีความสามารถในการเรียนรู้ สามารถเพิ่มพูนประสบการณ์ได้ แต่มนุษย์ก็ยังมีข้อเสียหลายประการ เช่น มนุษย์มีอายุที่จำกัด เราไม่สามารถที่จะเก็บรักษาผู้ชำนาญการด้านหนึ่งด้านใดไว้ได้ตลอดไป มนุษย์จะเกิดความเบื่อหน่ายที่ต้องทำงานเดิมบ่อยครั้งได้ง่าย เป็นต้น ดังนั้น ได้มีการคิดค้นประดิษฐ์อุปกรณ์ เพื่อให้สามารถทำงานแทนมนุษย์ได้ในด้านต่างๆ อุปกรณ์หนึ่งที่มีความสำคัญในปัจจุบัน และจะยังคงมีความสำคัญต่อไปในอนาคต คือ คอมพิวเตอร์ (Computer) คอมพิวเตอร์ได้เข้ามามีบทบาทต่อการดำเนินชีวิตของมนุษย์มากขึ้น คอมพิวเตอร์สามารถใช้งานได้ดี เนื่องจากมีข้อดีหลายอย่าง จึงมีการนำคอมพิวเตอร์มาทำงานแทนมนุษย์ในด้านต่างๆ แต่ในงานบางประเภทคอมพิวเตอร์ก็ไม่สามารถทำงานแทนมนุษย์ได้ จึงได้มีการพัฒนาความสามารถของคอมพิวเตอร์ต่อไป

ปัจจุบัน เครื่องคอมพิวเตอร์ถูกนำมาใช้กันอย่างแพร่หลาย เพื่อช่วยในการประมวลผลให้เป็นไปอย่างรวดเร็วและถูกต้อง ซึ่งการติดต่อระหว่างเครื่องคอมพิวเตอร์กับมนุษย์ โดยปกติเครื่องจะรับคำสั่งหรือข้อมูลทางแป้นพิมพ์ (keyboard) จะแสดงผลทางจอภาพ (terminal) หรือเครื่องพิมพ์ (printer) แป้นพิมพ์จะทำหน้าที่เปลี่ยนข้อมูลหรือ คำสั่งที่มนุษย์กดบนแป้นพิมพ์ให้เป็นภาษาสำหรับเครื่อง (0 กับ 1) แต่ในงานบางอย่างเราก็ไม่สะดวกในการที่จะกดแป้นพิมพ์

โดยปกติ วิธีสื่อสารที่สะดวกและรวดเร็ว คือการติดต่อกันด้วยเสียงพูด ทำให้มนุษย์มีความต้องการที่จะสามารถติดต่อกับเครื่อง (man-machine communication) โดยเสียงพูดซึ่งการติดต่อในระบบนี้จะต้องมีหน่วยรับข้อมูลที่สามารถรับเสียงมนุษย์ และสามารถแปลงเป็นสัญญาณเครื่อง (speech analyzer) และจะต้องมีส่วนผลลัพธ์ที่จะนำสัญญาณที่ได้ไปประมวลผล การสั่งงานด้วยเสียง สามารถนำมาใช้ประโยชน์ได้ดีในหลายด้าน ไม่ว่าจะเป็น การควบคุมเครื่องมืออุปกรณ์ในห้องผ่าตัด ของแพทย์ หรือการสั่งงานอุปกรณ์ต่างๆของผู้พิการ เครื่องคอมพิวเตอร์หรืออุปกรณ์เครื่องมือเหล่านั้น จะสามารถเข้าใจหรือจดจำคำพูดได้

อย่างไร คำถามนี้จึงเป็นหัวข้อศึกษาของงานวิจัยชิ้นนี้ ในการที่จะแก้ปัญหาโดยประยุกต์ใช้โครงข่ายประสาทเทียม (Neural Networks) ในการจดจำรูปแบบของเสียงพูด ทำให้เครื่องคอมพิวเตอร์สามารถเข้าใจเสียงพูดได้

โครงข่ายประสาทเทียมเกิดจาก ความพยายามที่จะสร้างเครื่องคำนวณที่สมบูรณ์แบบ ที่มีความสามารถเหมือนสมองของคน เครื่องคำนวณที่สมบูรณ์แบบต้องสามารถคิดเองได้ และโปรแกรมได้ด้วย สามารถเรียนรู้ได้เอง สะสมประสบการณ์ได้ โดยตอนเริ่มต้นเป็นการศึกษาพฤติกรรมของมนุษย์และโครงสร้างของสมองมนุษย์ จากนั้นได้เริ่มสร้างแบบจำลอง (model) เรียนแบบการทำงาน และได้รับการปรับปรุงจนกระทั่งได้แตกแขนงออกเป็นแบบจำลองย่อยๆหลายแบบ ซึ่งแบบที่ง่ายและนิยมใช้กันมากได้แก่ โมเดลแบบการแพร่กลับ (back propagation) ซึ่งเป็นแบบที่ถูกประยุกต์ใช้เพื่อการเรียนรู้และจดจำรูปแบบของเสียงพูดสำหรับโครงงานนี้

วัตถุประสงค์ของโครงงานการประยุกต์ใช้งานโครงข่ายประสาทเทียม โดยหลักจะเป็นการศึกษาทฤษฎีและหลักการของโครงข่ายประสาทเทียม และนำมาประยุกต์ใช้ในการจดจำรูปแบบของเสียงพูด เพื่อเป็นแนวทางในการพัฒนาไปสู่ระบบการรับข้อมูลด้วยเสียง หรือระบบสั่งงานด้วยเสียง ทำให้เราสามารถติดต่อสื่อสารกับคอมพิวเตอร์โดยใช้เสียงได้

การทำงานของระบบโครงข่ายประสาทเทียมประยุกต์เพื่อการจดจำเสียงพูด (Voice Recognization System) จะแบ่งออกเป็นสองส่วนงาน คือ ส่วนของการเรียนรู้เสียง และส่วนของการนำ Neural Networks ไปใช้งานแยกเสียง

ส่วนของการเรียนรู้เสียง

การทำงานเริ่มด้วยการเก็บตัวอย่างเสียง จากนั้นใช้ ฟาสท์ฟูเรียร์ทำการแปลงข้อมูลของเสียงจากโดเมนเวลา เป็นโดเมนความถี่ นำมาใช้เป็นข้อมูล ในการเรียนรู้ ของโครงข่ายประสาทเทียม ซึ่งจะได้ผลลัพธ์ มาเป็นแฟ้มข้อมูลน้ำหนัก (*.wgh file format) จากนั้นนำค่าน้ำหนักไปใช้งาน ในส่วนการนำโครงข่ายประสาทเทียมไปใช้แยกเสียง

ส่วนของการนำโครงข่ายประสาทเทียม ไปใช้ในการแยกเสียง

การทำงานเริ่มด้วยการเก็บตัวอย่างเสียง โดยใช้ซาวด์การ์ดเก็บเสียงในรูปแบบของจุดเวฟ (*.wav file format) จากนั้นใช้การแปลงฟาสท์ฟูเรียร์ทำการแปลงข้อมูลของเสียงจากโดเมนเวลา เป็นโดเมนความถี่ นำมาใช้เป็นอินพุท ในการตรวจจับเสียง ของโครงข่ายประสาทเทียม ซึ่งจะได้เอาท์พุท มาเป็นการบอกว่าเป็นเสียงของคำว่าอะไร

เนื้อหาโดยย่อของแต่ละบท

บทที่ 2 จะกล่าวถึงทฤษฎีและหลักการในเรื่องของการจดจำเสียง ความรู้ทั่วไปของการจดจำ การจดจำเสียงพูด หลักในการตัดและแยกพยางค์เพื่อนำมาใช้ในการจดจำเสียงพูด

บทที่ 3 จะกล่าวถึงทฤษฎีและหลักการในเรื่องของการแปลงฟาสท์ฟูเรียร์ ซึ่งนำมาใช้ในการแปลงข้อมูลที่อยู่ในโดเมนเวลามาเป็นข้อมูลในฟังก์ชันของโดเมนเวลากับความถี่ เพื่อใช้ในการเตรียมข้อมูล

บทที่ 4 จะกล่าวถึงทฤษฎีและหลักการในเรื่องของโครงข่ายประสาทเทียม กล่าวถึงหลักการของโครงข่ายประสาทเทียม โมเดลต่างๆของโครงข่ายประสาทเทียม

บทที่ 5 จะกล่าวถึงการทำงานของระบบการจดจำเสียงพูดด้วยโครงข่ายประสาทเทียม โครงสร้างและส่วนต่างๆ ของระบบการจดจำเสียงพูด

บทที่ 6 กล่าวถึงการทดลอง และผลการทดลอง ประสิทธิภาพของระบบที่ได้

บทที่ 7 กล่าวถึงบทสรุปและวิจารณ์ สรุปผลการทำงานและวิจารณ์การทำงานของระบบที่สร้างขึ้น นอกจากนี้ยังกล่าวถึงปัญหาที่เกิดขึ้นระหว่างการทำงาน

บทที่ 8 กล่าวถึงแนวทางการพัฒนาต่อไป การพัฒนาระบบให้มีประสิทธิภาพเพิ่มมากขึ้น การนำระบบไปประยุกต์ใช้

บทที่ 2
ทฤษฎีและหลักการเบื้องต้น

ทฤษฎีและหลักการที่ใช้ในโครงงานการประยุกต์ใช้โครงข่ายประสาทเทียมในการจดจำเสียง จะมีทฤษฎีที่ใช้หลักๆ ดังต่อไปนี้

1. ทฤษฎีและหลักการในเรื่องของการจดจำเสียง ซึ่งจะกล่าวถึงความรู้ทั่วไปของการจดจำ การจดจำเสียงพูด หลักในการตัดและแยกพยางค์เพื่อนำมาใช้ในการจดจำเสียงพูด
2. ทฤษฎีและหลักการในเรื่องของการแปลงฟาสท์ฟูเรียร์ ซึ่งนำมาใช้ในการแปลงข้อมูลที่อยู่ในโดเมนเวลามาเป็นข้อมูลในฟังก์ชันของโดเมนเวลากับความถี่ เพื่อใช้ในการเตรียมข้อมูล ซึ่งจะได้กล่าวต่อไปใน บทที่ 3
3. ทฤษฎีและหลักการในเรื่องของโครงข่ายประสาทเทียม กล่าวถึงหลักการของโครงข่ายประสาทเทียม ซึ่งจะได้พบในบทที่ 4

2.1 การจดจำเสียงพูด(Speech recognition)

2.1.1 ประเภทของการจำเสียงพูด

การจำเสียงพูดโดยทั่วไปนิยมแบ่งออกเป็น 4 กลุ่ม ดังต่อไปนี้

1. การจดจำแบบแยกคำ (Isolated-word recognition) เป็นการจำโดยต้องมีช่วงหยุดของแต่ละคำ
2. การจำเฉพาะคำ (Word spotting) ตรวจจับคำเฉพาะใดๆที่ปรากฏในประโยคนั้น
3. การจำแบบต่อเนื่อง (Connected speech recognition) จดจำคำพูดได้ทั้งประโยคโดยไม่ต้องมีช่วงหยุดระหว่างคำ
4. การเข้าใจเสียงพูด (Speech understanding) เป็นแบบสมบูรณ์ของการจำแบบต่อเนื่อง โดยสร้างแหล่งข้อมูลของการออกเสียงและหลักของภาษา จุดประสงค์หลักก็คือ การพูดที่ไม่ต้องถูกต้องมากนักแต่ให้เข้าใจความหมายของประโยคนั้นได้ ในทางทฤษฎี

ระบบเข้าใจเสียงพูดจะตัดทอนเอาเฉพาะความหมายของประโยค โดยไม่เข้มงวดเรื่องไวยากรณ์

การจำแบบแยกคำเป็นจุดเริ่มต้นของการพัฒนาการจำเสียงพูดจน เป็นการจำเฉพาะคำ และการจำแบบต่อเนื่องตามลำดับ ช่วงหยุดของแต่ละคำพูดทำให้การจดจำคำง่ายขึ้น เพราะสามารถแยกแยะจุดสิ้นสุดของแต่ละคำ (จุดเริ่มต้นและจุดสิ้นสุดของคำนั้น) ทำให้ผลที่เกิดจากความไม่ชัดเจนของการพูดลดน้อยลงไปได้มาก การจำแบบแยกคำขณะที่ออกเสียงจะต้องระมัดระวังมาก เพราะต้องการช่วงหยุดระหว่างการพูดแต่ละคำ ทำให้การพูดขาดความไพเราะของภาษาไป ไม่เหมือนกับวิธีการอื่นๆ ซึ่งสามารถพูดเป็นธรรมชาติได้มากกว่า และใช้ความระมัดระวังในการพูดน้อยกว่า

2.1.2 หลักการจดจำโดยใช้ลักษณะของเสียงพูด ตามการจำแนกแบบแยกคำ การจำแนกแบบแยกคำมีหลักการจดจำ โดยใช้ลักษณะของสัญญาณเสียงพูดดัง ต่อไปนี้

1. ขนาด หรือกำลังต่อเวลา
2. อัตราตัดผ่านศูนย์ (zero-crossing rate)
3. สมดุลกลุ่มความถี่รวม (gross spectrum balance) เป็นการเปรียบเทียบพลังงานของกลุ่มความถี่สูงกับกลุ่มพลังงานความถี่ต่ำ
4. ความถี่สำคัญอาจมาในรูปของ
   ก. ความถี่จากการแปลง DFT
   ข. F1,F2,F3
   ค. พารามิเตอร์ LPC หรือสัมประสิทธิ์ PARCOR
   ง. เอาท์พุทของฟิลเตอร์แบงก์ (filter-bank) วิธีนี้เป็นที่นิยมกันที่สุด เพราะราคาถูกและความเร็วสูง

ขนาดนอกจากเป็นตัวบอกจุดสิ้นสุดของแต่ละคำ และยังเป็นตัวชี้ความแตกต่างของพยัญชนะ และสระอีกด้วย

2.2 หลักการทำงานของการจดจำ

การทำงานของการจดจำโดยทั่วไปมีอยู่ 2 ขั้นตอน ดังนี้

1. การเรียนรู้ (Learning) เป็นการเก็บรวบรวมข้อมูลตัวอย่าง เพื่อนำมาให้ระบบปรับไลบรารี่ของรูปแบบอ้างอิง
2. การจดจำ เป็นการใช้ไลบรารี่ที่สมบูรณ์แบบเปรียบเทียบกับอินพุทที่ไม่รู้

ในการเรียนรู้ ข้อมูลสำหรับแต่ละคลาสจะถูกป้อนเข้าระบบ ระบบจะทำการสร้างรูปแบบอ้างอิงหรือต้นแบบของแต่ละคลาส โดยการเฉลี่ยรูปแบบจำนวนมากแล้วนำมารวมกันเป็นไลบรารี่โดยแยกออกเป็นคลาสต่างๆ ในการจดจำระบบจะทำการคำนวณรูปแบบของลักษณะเด่นของอินพุทที่ไม่รู้ และทำการตรวจสอบกับคลาสต่างๆเพื่อหาคลาสที่มีรูปแบบอ้างอิงใกล้เคียงกับลักษณะเด่นที่ได้มากที่สุด

การออกแบบระบบการจดจำ มีปัญหาที่สำคัญอยู่ 2 ประการ คือ

1. การเลือกลักษณะเด่นและการคำนวณค่า
2. ทางเลือกของกฎการตัดสินใจ(decision rule)

ปัญหาแรกที่มีความสำคัญมาก และยากมากกว่า เราจึงควรเริ่มต้นที่ปัญหาที่สองก่อน เพราะมีแนวความคิดหลายๆอย่าง ซึ่งจะช่วยในการอธิบายในหัวข้อการเลือกลักษณะเด่น

2.3 การตรวจจับจุดสิ้นสุด (Endpoint detection)

ปัญหาพื้นฐานของระบบจดจำเสียงพูด คือ เสียงที่เปล่งออกมาในขณะตรวจสอบนั้นอาจจะไม่เหมือนกันตอนที่เรียนรู้เสมอไป เสียงที่เปล่งออกมา 2 ครั้ง จะมีระยะเวลาที่ต่างกันและระยะระหว่างการออกเสียงที่ไม่แน่นอน หมายความว่าคุณสมบัติที่ขึ้นกับเวลา จะทำให้การเปรียบเทียบไม่สามารถทำได้ เพราะคำอ้างอิงกับคำที่มาเปรียบเทียบเวลาจะต่างกัน ในกรณีเช่นนี้ ถึงแม้ว่า จะเปรียบเทียบกับคำที่ไม่รู้กับรูปแบบที่ถูกต้องของตัวมันเองแล้วก็ตาม แต่ก็อาจจะให้ผลแตกต่างได้พอๆกับที่เปรียบเทียบกับรูปแบบอ้างอิงของคำอื่น ซึ่งจะเป็นปัญหา ถ้าคำที่ถูกเปรียบเทียบ เป็นทั้งคำต้นแบบ แทนที่จะเปรียบเทียบบางส่วนของคำ โดย

การเปรียบเทียบส่วนต่อส่วน ความจริงแม้ในระหว่างเสียงสองเสียงที่ใช้ในการเรียนรู้ รายละเอียดของเวลายังต่างกันเลย เมื่อเก็บข้อมูลเสียงจากหลายๆคน จึงต้องทำการเฉลี่ยในระหว่างการเรียนรู้ก่อนนำไปเปรียบเทียบ ในขั้นตอนการจดจำ ในหลายๆกรณี ความแม่นยำของการปรับ จะขึ้นอยู่กับความแม่นยำของการแสดงจุดสิ้นสุดของคำ ความผิดพลาดในการจดจำจุดเริ่มต้นและจุดสิ้นสุดของคำจะทำให้เกิดความยุ่งยากขึ้น โดยความผิดพลาดของจุดสิ้นสุดจะเกิดขึ้น ในคำที่เริ่มต้นและลงท้ายด้วยพยางค์ที่ออกเสียงเบา ผู้พูดบางคนอาจจบประโยคด้วยเสียงสั้น พยางค์เสียงสั้นนี้จะทำให้เกิดความผิดพลาดในการเปรียบเทียบคำ คุณสมบัติหลักของจุดสิ้นสุด คือพลังงานดังนั้น วิธีง่ายๆที่ใช้ในการตรวจจับจุดสิ้นสุด ก็คือการเปรียบเทียบพลังงานของสัญญาณกับค่าหรือระดับที่ตั้งไว้ และจุดสิ้นสุดจากตำแหน่งของคำที่มีพลังงานต่ำกว่าระดับที่ตั้งไว้ ค่าระดับที่ตั้งไว้นี้ต้องมีความเข้มระดับหนึ่ง ถ้าความแรงของสัญญาณเสียงพูดอ่อน อาจจะต้องตั้งระดับต่ำๆ

บทที่ 3
การแปลงฟูเรียร์ (Fourier Transform)

3.1 หลักการเบื้องต้นในการแปลงฟูเรียร์

ให้ f(x) เป็น ฟังก์ชัน(function) ของจำนวนจริงแบบต่อเนื่อง การแปลงฟูเรียร์ของ f(x) จะเขียนได้ดังสมการ

$$F(u) = \int_{-\infty}^{\infty} f(x)\exp(-j2\pi ux)dx$$

เมื่อ j = $\sqrt{-1}$ และให้ inverse Fourier transform ของ F(u) แทนด้วยสมการ

$$f(x) = \int_{-\infty}^{\infty} F(u)\exp(-j2\pi ux)du$$

function f(x) เองจะเป็นจำนวนจริงแต่ function F(u) อาจเป็นจำนวนเชิงซ้อนก็ได้ ซึ่งจะได้

$$F(u) = R(u) + jI(u)$$

เมื่อ R(u) เป็นส่วนจริง และ I(u) เป็นส่วนจินตภาพ จะได้

$$|F(u)| = \left[R^2(u) + I^2(u)\right]^{1/2}$$

เราเรียก |F(u)| ว่า แอมพลิจูด(Amplitude) หรือ ฟูเรียร์สเปกตรัม(Fourier spectrum) ของ f(x) และให้ พาวเวอร์สเปกตรัม (Power Spectrum) เป็น

$$P(u) = |F(u)^2|$$

เราเรียก u ว่าเป็น ค่าความถี่(frequency variable)

เราสามารถแปลงสมการ ฟูเรียร์ แบบ ต่อเนื่อง ให้เป็น แบบ ไม่ต่อเนื่อง (discrete) ได้ดังนี้

$$F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)\exp(-j2\pi ux/N)$$

เมื่อ u = 0,1,2,..,N-1 และ

$$f(x) = \frac{1}{N}\sum_{x=0}^{N-1} F(u)\exp(j2\pi ux/N)$$

เมื่อ x = 0,1,2,..,N-1

เช่นเดียวกันในกรณีของสองตัวแปรจะได้สมการคู่ต่อไปนี้

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[-j2\pi\left(ux/M + vy/N\right)\right]$$

เมื่อ u = 0,1,2,..,M-1,v = 0,1,2,..,N-1 และ

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[j2\pi\left(ux/M + vy/N\right)\right]$$

เมื่อ x = 0,1,2,..,M-1,y = 0,1,2,..,N-1

3.2 การแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform)

ใน fourier transform ความซับซ้อนของการคำนวณนั้นจะเป็นสัดส่วนกับ N^2 ซึ่งจะต้องใช้เวลามากเมื่อค่า N มากขึ้น จึงมีการพยายามลดความซับซ้อน ในการคำนวณให้เหลือเป็นลำดับของ $N\log_2 N$ โดยเรียกวิธีนี้ว่า ฟาสท์ฟูเรียร์ทรานฟอร์ม (FFT) ซึ่งทำให้มีความเร็วสูงขึ้นมาก ยกตัวอย่างการแปลง fourier เมื่อมีค่า N เป็น 8192 ถ้าใช้วิธีธรรมดาจะเสียเวลามากกว่า ฟาสท์ฟูเรียร์ทรานฟอร์ม ถึง 600 เท่า

อัลกอลิทึม (Algorithm)

ฟาสต์ฟูเรียร์ทรานฟอร์มอัลกอลิทึม (FFT algorithm) จะเป็น อัลกอลิทึม ประเภท ซักเซสทีฟอัลกอลิทึม (successive doubling method) โดยจากการแปลง ฟูเรียร์ ธรรมดา

$$F(u) = \sum_{x=0}^{N-1} f(x)\, W_N^{ux}$$

เมื่อ

$$W_N = \exp(-j2\pi / N)$$

033171

สมมุติให้ N อยู่ในรูป

$$N = 2^n$$

เมื่อ n เป็นจำนวนเต็มบวกใดๆ ดังนั้น N สามารถเขียนได้ในรูป

$$N = 2M$$

เมื่อ M เป็นจำนวนเต็มบวกเช่นกัน นำไปแทนในสมการแรกจะได้

$$F(u) = \frac{1}{2M}\sum_{x=0}^{2M-1} f(x) W_{2M}^{ux}$$

$$= \frac{1}{2}\left[\frac{1}{M}\sum_{x=0}^{M-1} f(2x) W_{2M}^{u(2x)} + \frac{1}{M}\sum_{x=0}^{M-1} f(2x+1) W_{2M}^{u(2x)}\right]$$

เนื่องจาก $W_{2m}^{2ux} = W_{M}^{ux}$ จะได้

$$F(u) = \frac{1}{2}\left[\frac{1}{M}\sum_{x=0}^{M-1} f(2x) W_{M}^{ux} + \frac{1}{M}\sum_{x=0}^{M-1} f(2x+1) W_{M}^{ux} W_{2M}^{u}\right]$$

กำหนดให้

$$F_{even}(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(2x) W_{M}^{ux}$$

เมื่อ u = 0,1,2,..,M-1 และ

$$F_{odd}(u) = \frac{1}{M}\sum_{x=0}^{M-1} f(2x+1) W_{M}^{ux}$$

เมื่อ u = 0,1,2,..,M-1 จะได้

$$F(u) = \frac{1}{2}\left[F_{even}(u) + F_{odd}(u) W_{2M}^{u}\right]$$

เพราะว่า $W_M = W_M, W_{2M} = -W_{2M}$ จะได้

$$F(u+M) = \frac{1}{2}\left[F_{even}(u) - F_{odd}(u) W_{2M}^{u}\right]$$

จากสมการที่ได้ จะเห็นว่าเราต้องการหา ฟูเรียร์ ของจุดสองจุด เราจะเริ่มจากการหาจุดแบบ1 จุด 2 ครั้ง แต่ว่าจากครั้งแรก F(u) เราสามารถหาค่า F_{even} และค่า F_{odd} ได้ ดัง

นั้นในการหาจุดที่ 2 F(u+M) ก็จะนำค่าที่คำนวณได้แล้วมาบวกกัน ในการหาสี่จุดก็เช่นกัน จะเริ่มจากการหา 2 จุด 2 ที่ส่วนหลังก็ไม่ต้องนำมาคำนวณอีก

3.3 การแปลงกลับของฟูเรียร์ (inverse fourier transform)

ในทำนองเดียวกัน การหา inverse fourier transform ก็สามารถใช้ algorithm ของ FFT ได้โดยจัดให้อยู่ในรูป

$$\frac{1}{N} f^*(x) = \frac{1}{N} \sum_{u=0}^{N-1} F^*(u) \exp(-j2\pi ux / N)$$

เมื่อ f*(x) หมายถึง conjugate ของ f(x) ซึ่งจะเห็นว่าสมการข้างบนอยู่ในรูปเดียวกับสมการที่แปลง fourier transform ธรรมดา ทำให้สามารถใช้ algorithm เดิมได้

บทที่ 4
โครงข่ายประสาทเทียม

4.1 ความรู้ทั่วไป

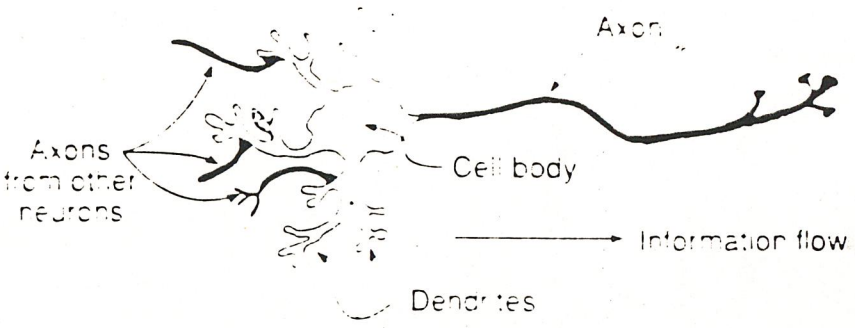
ความพยายามที่จะสร้างเครื่องคำนวณที่สมบูรณ์แบบ ที่มีความสามารถเหมือนสมองคนได้มีมานานนับศตวรรษแล้ว เครื่องคำนวณที่สมบูรณ์แบบต้องสามารถคิดเองได้ และโปรแกรมได้ด้วย นอกจากนั้นมันยังต้องสามารถที่จะหยั่งรู้เหตุการณ์ล่วงหน้าได้หรือผันได้ด้วย เครื่องคำนวณที่เรารู้จักกันดี คือ เครื่องคอมพิวเตอร์ หลักการทำงานของเครื่องคอมพิวเตอร์ที่ใช้กันในปัจจุบันตั้งอยู่บนพื้นฐานของคณิตศาตร์แบบบูลลีน ซึ่งไม่ได้มีส่วนเกี่ยวข้องกับหลักการทำงานเซลประสาทในสมองของคนเราแม้แต่น้อย ถึงแม้ว่าหลักการทำงานของคอมพิวเตอร์จะไม่ได้อิงการทำงานของเซลประสาท แต่มันก็สามารถทำให้คอมพิวเตอร์มีประสิทธิภาพในด้านการทำงานต่างๆที่เราต้องการ

ปัญหาที่สำคัญอย่างหนึ่งในการสร้างคอมพิวเตอร์ที่มีประสิทธิภาพ คือ ทำอย่างไรจึงจะทำให้คอมพิวเตอร์มีความเร็วสูงในการแก้ปัญหาที่ป้อนให้ เราจะเห็นได้ว่าการทำให้คอมพิวเตอร์มีความเร็วสูงมากๆ ทำได้สองวิธีคือ ใช้ชิ้นส่วนของสารกึ่งตัวนำที่มีความเร็วในการเปลี่ยนสภาวะทางไฟฟ้าสูง หรือใช้การแก้ปัญหาแบบขนาน นั่นคือแทนที่จะใช้หน่วยคำนวณอันเดียววิ่งโปรแกรมตั้งแต่ต้นจนจบ แต่จะใช้หน่วยคำนวณหลายๆหน่วย วิ่งโปรแกรมนี้โดยแบ่งโปรแกรมไปยังหน่วยเหล่านี้

ในการประชุมทางโครงข่ายประเภทเทียมที่รัฐ Maryland, USA เมื่อเดือนมิถุนายน 2535 H,Szen จากสถาบันวิจัยของกองทัพเรืออเมริกัน รายงานว่าการเพิ่มหน่วยคำนวณมากขึ้นไม่ได้ทำให้ความเร็วการแก้ปัญหาของคอมพิวเตอร์เพิ่มขึ้นโดยแท้จริง ถ้าจำนวนของหน่วยคำนวณเกินจำนวนหนึ่ง เหตุผลการเนื่องจากขณะที่หน่วยคำนวณทำงานพร้อมกัน มันจะต้องมีการติดต่อสื่อสารและส่งสัญญาณข้อมูลระหว่างกัน จะทำให้เกิดสภาวะ 'คอขวด' ซึ่งจะลดความเร็วและประสิทธิภาพของการแก้ปัญหาลงอย่างมาก การที่จะแก้ปัญหาการเกิดสภาวะแบบคอขวด ทำได้โดยการใช้เทคในโลยีใหม่คือ โครงข่ายประสาทเทียม (Neural Networks) การศึกษาและวิจัยโครงข่ายประสาทเทียมเริ่มแรกก็เพื่อที่จะเข้าใจถึงการทำงานของเซล

ประสาทในสมอง ในทางวิศวกรรมเราสนใจแค่แบบจำลองของเซลประสาทในเชิงคณิตศาสตร์และการประยุกต์ของมันเท่านั้น

ในธรรมชาติเซลประสาทของสิ่งมีชีวิตมีหลายประเภทแล้วแต่หน้าที่ของมัน เซลประสาทในตัวของคนเรา ก็เช่นกันมีหลายประเภทตามตำแหน่งและหน้าที่ของมัน เช่น เซลประสาทของกล้ามเนื้อ เซลประสาทในสมอง เซลประสาทที่ลิ้นและจมูกเป็นต้น โครงสร้างและส่วนประกอบของเซลประสาทโดยทั่วไป แสดงได้ดังรูปที่ 4.1



รูปที่ 4.1

เซลประสาทประกอบด้วยส่วนประกอบใหญ่ๆ สี่ส่วนคือ ตัวเซลประสาทซึ่งมีนิวเคลียส อยู่ตรงกลาง รอบๆตัว เซลประสาทมีสิ่งที่ยื่นออกไปเพื่อส่งและรับสัญญาณจากเซลปรระสาทอื่น สิ่งดังกล่าวเรียกว่า แอกซอน (axon) ที่ปลายกิ่งจะแตกออกเป็นก้านย่อยๆ เรียกว่า เดนไดรท์ (dendrite) รอยต่อระหว่างก้านของเซลประสาทที่ต่างกัน เรียกว่า ไซแนป (synapse) ซึ่งสามารถเปลี่ยนค่าความต้านทานได้ตามสัญญาณที่ส่งระหว่างกันของเซลประสาท การส่งสัญญาณระหว่างเซลประสาท ทำโดยการถ่ายเทสารประกอบของโซเดียมและโพแทสเซียม Hodgkin และ Huxley ซึ่งได้รับรางวัลโนเบลทางชีววิทยาได้ค้นพบว่า การไหลของสารประกอบโซเดียมและโพแทสเซียมของเซลประสาทของปลาหมึกได้ทำให้เกิดความต่างศักย์ระหว่างสารที่อยู่ในแอกซอน และสารที่อยู่นอกแอกซอน ค่าความต่างศักย์จะอยู่ระหว่าง 50 ถึง 70 มิลลิโวลต์ จากผลการศึกษาดังกล่าว ทำให้เราสามารถจำลองการทำงานเซลประสาทโดยอาศัยวงจรอิเลคทรอนิคได้

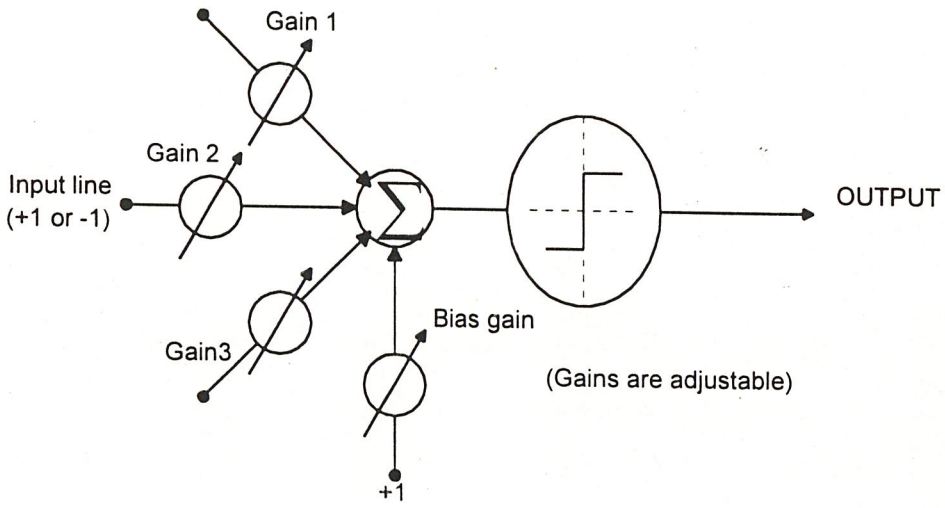
แบบจำลองต่างๆของโครงข่ายประสาทเทียม ที่ถูกพัฒนาขึ้นมาตั้งแต่ปี ค.ศ. 1957 จะสังเกตได้ว่าแนวคิดของโครงข่ายประสาทเทียมได้มีมานานกว่า 30 ปีแล้ว และถ้าสังเกตจากตารางที่ 4.1 จะเห็นได้ว่ามีช่วงเวลาที่การพัฒนาด้านนี้ลดน้อยลง โดยเฉพาะอย่างยิ่งช่วงปี ค.ศ. 1970 - 1984 เหตุผลเนื่องจาก การพัฒนาทางด้านอิเล็กทรอนิกส์ ที่จะมารับเทคโนโลยีของโครงข่ายประสาทเทียม ยังไม่มีประสิทธิภาพมากพอ ความก้าวหน้าทางเทคโนโลยีวงจรรวม (VLSI Very Large Scale Integration) ยังอยู่ในช่วงเริ่มต้นหัดเดินเท่านั้น ข้อมูลที่ใช้งานแบบจำลองทั้งหมดเป็นแบบจำลองที่ถูกพัฒนาขึ้นก่อนปี ค.ศ. 1986 และหลังจากปี ค.ศ. 1986 เป็นต้นไป มีการพัฒนาและขยายแนวคิดของโครงข่ายประสาทเทียมเป็นจำนวนมาก

ชื่อของแบบจำลอง	ผู้คิดค้น	ปีที่พัฒนา	การใช้งาน
1. Perceptron	F.Rosenblatt	1957	อ่านตัวพิมพ์ดีด
2. Madaline	B.Widrow	1960-62	adaptive equalizer
3. Avalanche	S.Grossberg	1967	เข้าใจคำพูด
4. Cerebellatron	D.Mar, J.Albus and A.Pellioonez	1969-82	ควบคุมการเคลื่อนไหวของแขนหุ่นยนต์
5. Back Propagation	P.Werbos,D.Parker, and D.Rumelhart	1974-85	เลียนแบบฟังก์ชันทางคณิตศาสตร์,อ่านตัวเขียน,คำพูด,ทำนายหุ้น
6. Brain-State-in-a-box	J.Anderson	1977	สกัดข้อมูลบางอย่างจากฐานข้อมูล
7. Neocognitron	K.Funkushima	1978-84	อ่านตัวเลขและตัวอักษรเขียน
8. Adaptive Resonance	G.Carpenter and S.Grossberg	1978-86	pattern recognition
9. Self-Organizing Map	T.Kohonen	1989	จำลองลักษณะพื้นผิวของวัตถุ เช่น ปีกของเครื่องบิน
10. Hopfield	J.Hopfield	1982	ทำนายส่วนของข้อมูลที่หายไป เช่น บางส่วนของหน้าคน และทำ optimization สำหรับปัญหาพวก combination
11. Bidirectional Associative Memory	B.Kosko	1985	content-addressable associative memory
12. Boltzman and Cauchy machine	J.Hinton,T Sejnowsky, and H.Szu	1985-86	pattern recognition
13. Counterpropagation	R.Hecht-Nielsen	1986	อัดข้อมูลของภาพให้น้อยลง,table-lookup

ตารางที่ 4.1

4.2 แบบจำลองทางคณิตศาสตร์ของเซลประสาทเดียว

จากโครงสร้างแบบง่ายของเซลประสาทที่แสดงในรูปที่ 4.1 ในปี ค.ศ. 1943 McCulloch และ Pitts ได้เสนอแบบจำลองทางคณิตศาสตร์ของเซลประสาทดังกล่าว โครงสร้างของแบบจำลองนี้แสดงดังรูปที่ 4.2



รูปที่ 4.2

สัญญาณที่ส่งออกมาจากเซลประสาทในรูปที่ 4.2 คือ n สามารถเขียนสมการคณิตศาสตร์ของ n ที่เวลา t+1 ได้ดังนี้

$$n_i(t+1) = f\left(\sum_j \omega_{ji} n_j(t) - T_i\right)$$

เมื่อ ω_{ji} คือค่าความต้านทานของไซแนปที่ j ของเซลประสาทตัวที่ i, ค่า $n_j(t)$ คือ สัญญาณที่เซลประสาทตัวที่ j ส่งออกมาที่เวลา t, T_i คือค่าเปรียบเทียบ ซึ่งเรียกว่า threshold ส่วน f( ) คือฟังก์ชัน f ถูกนิยามดังต่อไปนี้

f(x) = { 1 ถ้า x >= 0

0 ถ้า x < 0

ค่าของ x ในที่นี้ก็คือ $\left(\sum_j \omega_{ji} n_i(t) - T_i\right)$ นั่นเอง

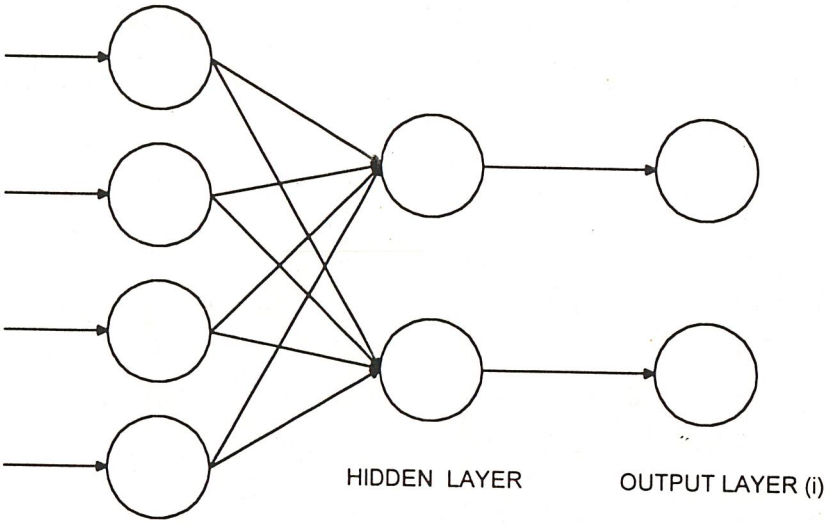
เซลประสาทเรียนรู้โดยการเปลี่ยนค่าของ ω_{ji} ดังนั้นสิ่งที่สำคัญมากที่สุดในการสร้างแบบจำลองของเซลประสาทนอกจากแบบจำลองเชิงคณิตศาสตร์แล้ว เราจำเป็นจะต้องกำหนดวิธีการที่จะเปลี่ยนค่าของ ω_{ji} ตามปัญหาต่างๆ ได้ด้วย เพื่อให้เซลประสาทส่งสัญญาณออกมาตามต้องการจากสมการข้างต้น

4.3 เทคนิคการเรียนรู้แบบแพร่กลับ (Back propagation)

เทคนิคของการเรียนรู้ของโครงข่ายประสาทเทียมมีอยู่หลายวิธีด้วยกัน วิธีที่นิยมและแพร่หลายมากที่สุด คือ วิธีการเรียนรู้หรือการเปลี่ยนค่าของ ω_{ji} โดยใช้วิธีที่เรียกว่า การแพร่กลับ (Back Propagation) การแพร่กลับพัฒนามาจากสมการพื้นฐานของการเรียนรู้แบบ McCullock และ Pitts โดยการพิจารณาฟังก์ชันของความคลาดเคลื่อนระหว่างค่าของเป้าหมายที่ต้องการ และค่าของสัญญาณที่ส่งออกมาจากตัว neuron เอง นอกจากนั้นวิธีการเรียนรู้โดยวิธีแพร่กลับยังพิจารณาค่าสัญญาณที่ส่งออกมาจากตัว neuron ในขอบข่ายของค่าจริงด้วย (ไม่ใช่แค่ 0 กับ 1 หรือ +1 กับ -1 เท่านั้น)

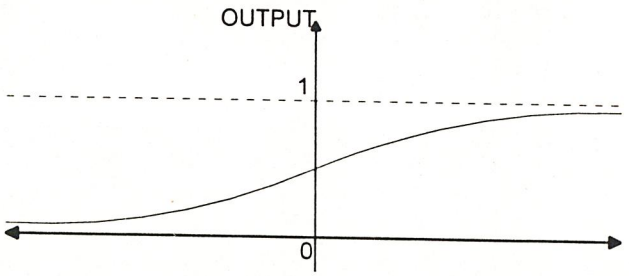
โครงข่ายประสาทเทียมที่มีการเรียนรู้โดยวิธีแพร่กลับนี้จะประกอบไปด้วย กลุ่มของเซลประสาทจัดตัวกันเป็นชั้นๆ ดังตัวอย่างที่แสดงในรูปต่อไปนี้

ONE LAYER BACKPROPAGATION
NEURAL NETWORK SYSTEM



กำหนดให้ p = จำนวน pattern
i = ขนาด output
j = ขนาด input

ที่ input layer จะเป็นการผ่าน data เข้ามาแล้วส่งไปยัง hidden layer ซึ่งใน hidden layer จะทำการหาผลรวมของแต่ละ input คูณด้วย weight จากนั้นจะนำผลรวมที่ได้ไปผ่าน sigmoid function ก่อนที่จะส่งไปยัง layer ถัดไป ลักษณะของ sigmoid function เป็นดังรูป



$$OUTPUT = \frac{1}{(1+\exp(-NET))}$$

$$NET = \sum_{i} I_{pj} \cdot W_{ij}$$

แต่ละขาของสัญญาณเข้าจะมีค่าของความต้านทาน หรือ ค่าของสัญญาณออกของแต่ละเซลคำนวณได้จากสมการต่อไปนี้

$$f(\alpha) = \frac{1}{1+e^{-(\alpha-\theta)}}$$

เมื่อ e คือ ค่าฐานของ log ธรรมชาติ ซึ่งมีค่าประมาณ 2.7182818

α คือ $\sum \omega_i X_i$

θ คือค่าคงที่

1. การกำหนดค่าความต้านทาน ω_i และ θ ของแต่ละเซลอย่างสุ่ม

2. ป้อนเวคเตอร์ $(X_1,X_2,...,X_n)$ และสัญญาณที่ต้องการให้ปรากฏที่ชั้นนอกสุด $(d_1,d_2,...,d_m)$

3. จากเวคเตอร์ $(X_1,X_2,...,X_n)$ คำนวณหาค่าสัญญาณที่ส่งออกจากเซลแต่ละเซลที่ชั้นนอกสุด โดยไล่ไปทีละชั้น จากชั้นที่ 1 การคำนวณทำโดยใช้สูตร $f(\alpha)$ ข้างต้น ให้ $(Y_1,Y_2,...,Y_n)$ เป็นค่าขอสัญญาณที่ได้จาการคำนวณเวคเตอร์ $(d_1,d_2,...,d_m)$ จะแตกต่างจากเวคเตอร์ $(Y_1,Y_2,...,Y_m)$

4. ปรับค่าของความต้านทานของแต่ละเซล โดยเริ่มจากชั้นนอกสุดไล่ไปจนถึงชั้นที่ 1 ค่าความต้านทานใหม่ปรับได้โดยสมการต่อไปนี้

$$\omega_{ij}(t+1) = \omega_{ij}(t) + \eta \delta_j x'_j$$

เมื่อ ω_{ij} คือ ความต้านทานระหว่างเซลประสาท i และ j ดังรูปที่ 4.3

η คือ ค่าคงที่ที่เรากำหนดเอง ค่าของ η เรียกว่าอัตตราการเรียนรู้โดยทั่วไป η จะมีค่าน้อยกว่า 1

j หาได้โดย

ก. ถ้าเซล j เป็นเซลชั้นนอกสุด ค่าของ j จะเท่ากับ

$$\delta_j = y_j(1-y_j)(d_j - y_j)$$

ข. ถ้าเซล j เป็นเซลในชั้นที่ซ่อนอยู่ ค่าของ จะเท่ากับ

$$\delta_j = x'_j(1-x'_j)\sum_k \delta_k \omega_{jk}$$

เมื่อ $x'j$ คือ ค่าสัญญาณขาเข้าของเซล

jk คือ จำนวนเซลทั้งหมดบนชั้นที่อยู่ถัดจากชั้นของเซล j ขึ้นไป

5. ทำซ้ำข้อ 2 ถึง 5 จนกว่าค่าความแตกต่าง E จะมีน้อยกว่า ΔE กำหนดดังนี้

$$E = \frac{1}{2}\sum_{i}(di - yi)^2$$

หลังจากเราสอนโครงข่ายประสาทให้เรียนรู้แล้ว ต่อไปจะนำมันมาใช้งานโดยการป้อนเวคเตอร์เข้าไป แล้วคำนวณหาค่าสัญญาณที่ออกจากแต่ละเซลประสาทโดยใช้สมการข้างต้น คือ

$$f(\alpha) = \frac{1}{1 + e^{-(\alpha - \theta)}}$$

โดยไล่ไปทีละเซลในแต่ละชั้นของโครงข่ายจากสมการข้างต้นเราสามารถหาการปรับตัวของค่าน้ำหนักได้ดังนี้

การปรับตัวของ weight

Least Mean Square Error (SLM)

กำหนดให้ p = จำนวน pattern i = ขนาด output

j = ขนาด input I = Input

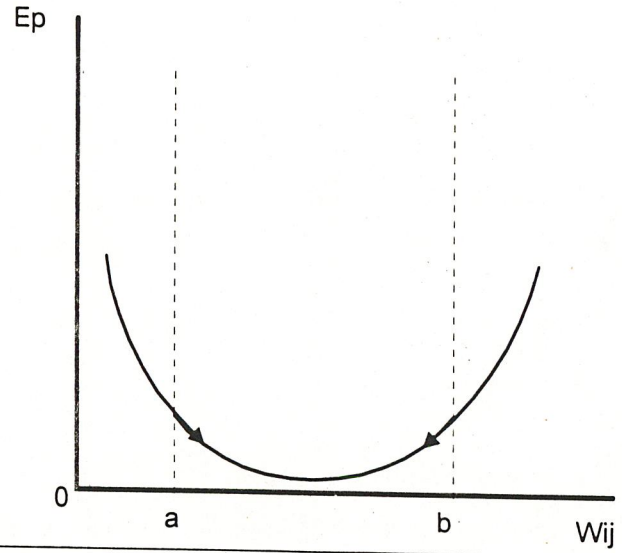
O = Output T = Target

E = Error

จะได้ว่า $E = \sum_{p} Ep$

$= \sum_{p} (Tpi-Opi)^2$

การปรับตัวของ weight จะเขียนได้ว่า $Wij(new) = Wij(old) + Wij$



จากรูปเราจะเห็นได้ว่า $\Delta Wij = -K\frac{\partial Ep}{\partial Wij}$

$= -K\frac{\partial (Tpi-Opi)^2}{\partial Opi} \cdot \frac{\partial Opi}{\partial NET} \cdot \frac{\partial NET}{\partial Wij}$

$= -K(-2Tpi+2Opi) \cdot Opi(1-Opi) \cdot Ipj$

$= 2K(Opi-Tpi)Opi(1-Opi)Ipj$

$= \varepsilon(Opi-Tpi)Opi(1-Opi)Ipj$

จะได้ว่า ε ค่าสัมประสิทธิ์ของการ learnning

บทที่ 5
ระบบการจดจำเสียง

ในขั้นแรกของการพัฒนาเริ่มต้นด้วยการศึกษาวิเคราะห์แนวทางต่างๆ ว่ามีวิธีใดบ้างที่ใช้กันอยู่ และมีประสิทธิภาพแค่ไหน จากนั้นเลือกแนวทางที่ทำโดยเปรียบเทียบข้อดีข้อเสียต่างๆ จากการทดลองจากนั้นก็เลือกแนวทางที่เห็นว่าเหมาะสมกับโครงงานนี้

โดยสรุปแล้วขั้นตอนในการดำเนินงานของโครงการ มีดังแผนภาพ Voice recognization system ในหน้าต่อไปนี้

จากแผนภาพ เราพบว่าระบบจดจำเสียงพูดแบ่งออกเป็น 2 ส่วน

- ส่วนการเก็บและเรียนรู้เสียง
- ส่วนการนำระบบจดจำเสียงไปใช้งาน

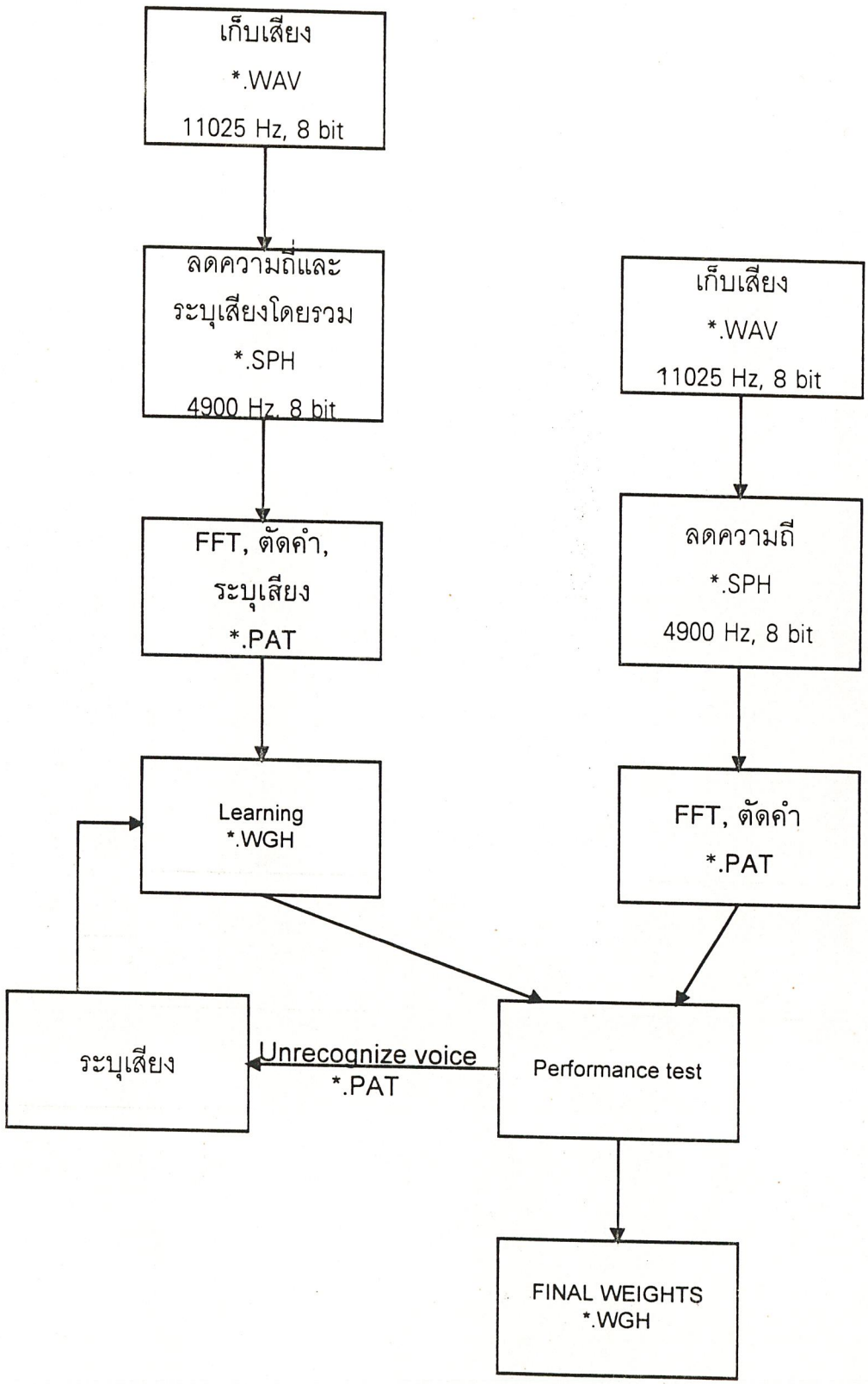
ส่วนการเก็บและเรียนรู้เสียง ประกอบด้วย

- การเก็บตัวอย่างเสียง
- การแปลงรูปแบบไฟล์ให้มีขนาดความถี่การสุ่มที่ลดลงและการระบุเสียง
- การแปลงเสียงจากโดเมนเวลาเป็นโดเมนความถี่
- การตัดคำและระบุเสียง
- การเรียนรู้โดยโครงข่ายประสาทเทียม
- การทดสอบและปรับปรุงประสิทธิภาพ

ส่วนการนำระบบจดจำเสียงไปใช้งาน ประกอบด้วย

- การเก็บตัวอย่างเสียง
- การแปลงรูปแบบไฟล์ให้มีขนาดความถี่ที่ลดลง
- การตัดคำ
- การบ่งบอกเสียงโดยโครงข่ายประสาทเทียม

VOICE RECOGNIZATION SYSTEM



การเก็บตัวอย่างเสียง

เป็นการออกจากโปรแกรม VRECPRJ.EXE มาเรียกโปรแกรม WREC.EXE ซึ่งเป็นทูลของซาวด์การ์ด เก็บตัวอย่างเสียง ในรูปแบบไฟล์เวฟฟอร์แมท ใช้อัตราความถี่ในการสุ่ม (sampling rate) เท่ากับ 11,025 Hz ใช้การสุ่มแบบ 8 บิท

การแปลงรูปแบบไฟล์ให้มีขนาดความถี่ในการสุ่มที่ลดลงและการระบุเสียง

เป็นการแปลงขนาดความถี่ในการสุ่ม จาก 11,025 Hz เป็น 4,900 Hz โดยการแปลงรูปแบบของไฟล์ และทำการเก็บข้อมูลว่าไฟล์ดังกล่าวเก็บข้อมูลคำพูดว่าอะไรไว้ที่ต้นไฟล์ในรูปของสตริงค์ การทำงานเรียกใช้รูทีน CHANGE_SCALE(char *fname,char *foname, char *speech) จากไฟล์ VRECPRJ.C มีอินพุทเป็นไฟล์เวฟฟอร์แมท (*.wav) เอาท์พุทเป็นไฟล์เอสพีเอชฟอร์แมท (*.sph) รูทีน change_scale มีพารามิเตอร์ดังนี้

char *fname : input file *.wav

char *foname : output file *.sph

char *speech : string ที่บ่งบอกว่าเป็นคำพูดอะไร เช่น '12345' ซึ่งจะเก็บไว้ที่ต้นไฟล์ ของไฟล์เอสพีเอชฟอร์แมท

การแปลงเสียงจากโดเมนเวลาเป็นโดเมนความถี่

เป็นการนำฟาสท์ฟูเรียร์ทรานฟอร์มมาประยุกต์ใช้ ในการแปลงเสียงจากโดเมนเวลาเป็นโดเมนความถี่ เพื่อใช้ในระบบการตัดคำต่อไป ในส่วนนี้จะเรียกใช้รูทีน FFT(void)

การตัดคำและระบุเสียง

ในส่วนนี้จะเป็นการตัดคำและบ่งบอกว่าคำที่ตัดได้เป็นเสียงของคำว่าอะไร การระบุเสียงจะ

เป็นการรับอินพุททางแป้นพิมพ์จากผู้ใช้ ที่บอกว่าเสียงของคำที่ตัดได้เป็นคำว่าอะไร ข้อมูลของการระบุจะกำหนดไว้ที่ต้นไฟล์ การทำงานในส่วนนี้จะมีอินพุทเป็นไฟล์แบบเอสพีเอชฟอร์แมท เอาท์พุทเป็นไฟล์ในรูปแบบพีเอทีฟอร์แมท (*.pat) ส่วนนี้จะเรียกใช้รูทีน FIND_NEW_WORD(int *startP,int *stopP,int mode) มีพารามิเตอร์ต่างๆ ดังนี้

int *startP : เป็นจุดเริ่มต้นของคำ

int *stopP : เป็นจุดสิ้นสุดของคำ

int mode : mode 0 : Show output in graphic mode to screen

1 : Not show

การเรียนรู้โดยโครงข่ายประสาทเทียม

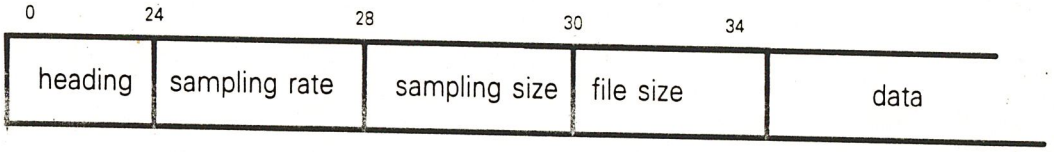
เป็นการนำไฟล์รูปแบบของพีเอทีฟอร์แมท ซึ่งเป็นไฟล์ที่เก็บข้อมูลส่วนหัวที่บอกว่าเป็นเสียงของคำว่าอะไร และแพทเทิร์นอินพุทที่มีขนาด 64x40 มาเป็นอินพุทของโครงข่ายประสาทเทียม โดยเรียกใช้ LEARNING.C สำหรับโครงข่ายประสาทเทียมที่ใช้เป็นแบบแพร่กลับแบบเลเยอร์เดียว (One-layer Back Propagation Neural Networks) มีอินพุทเป็นแพทเทิร์นขนาด 64x40 เอาท์พุทขนาด 8 บิท การทำงานของส่วนนี้จะได้น้ำหนัก (weight) เป็นไฟล์รูปแบบดับเบิลยูจีเอชฟอร์แมท (*.wgh)

การทดสอบและปรับปรุงประสิทธิภาพ

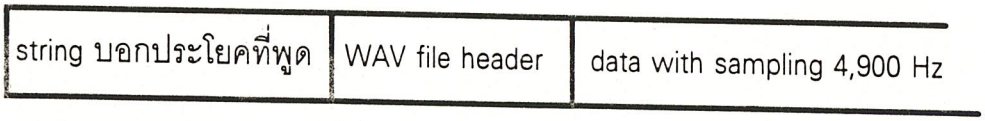
การทดสอบและปรับปรุงประสิทธิภาพเชื่อมต่อระหว่างส่วนการเรียนรู้เสียงและส่วนการนำโครงข่ายประสาทเทียมไปใช้ในการจดจำเสียง เป็นการนำข้อมูลในรูปแบบพีเอทีฟอร์แมท ที่ได้มาจากส่วนของการนำโครงข่ายประสาทเทียมไปใช้ในการจดจำเสียง นำมาทดสอบโดยโครงข่ายประสาทเทียมที่ใช้น้ำหนักที่ได้จากส่วนการเรียนรู้ จากการทดสอบนี้หากผลที่ได้ปรากฏว่าโครงข่ายประสาทเทียมจดจำได้ไม่ถูกต้อง จะนำตัวอย่างนั้นไปใช้ในการเรียนรู้ใหม่ เพื่อปรับปรุงค่าน้ำหนักให้ดียิ่งขึ้น ส่วนนี้จะเรียกใช้รูทีน PROVE(void)

สำหรับรูปแบบของไฟล์แต่ละแบบจะเป็น ดังรูปต่อไปนี้

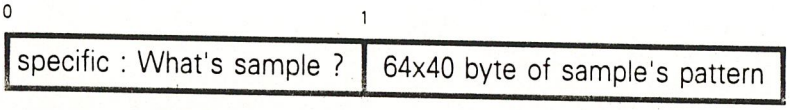
WAV file format



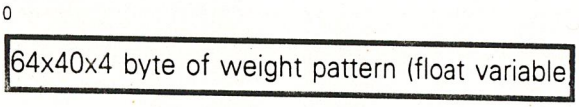
SPH file format



PAT file format



WGH file format



จำนวน 8 ไฟล์ สำหรับเอาท์พุทจำนวน 8 โหนด

5.1 แนวทางการพัฒนามีรายละเอียดคร่าว ๆดังนี้

5.1.1 ศึกษาทฤษฎีและแนวทางต่างๆที่จะใช้ ซึ่งได้แก่

1) ศึกษาหาวิธีที่จะดึงเอาลักษณะสำคัญของเสียง ไปเป็นข้อมูลเข้าไปยังส่วนของระบบการจดจำ พบว่าสิ่งที่เหมาะ ได้แก่ การหาส่วนประกอบของความถี่ของเสียง โดยใช้วิธี ฟาสท์ฟูเรียร์ทรานฟอร์ม และใช้ค่าพลังงานของเสียงเป็นส่วนที่ช่วยในการแบ่งคำ จากการศึกษาพบว่ายังมีอีกวิธีหนึ่งที่นิยมใช้เช่นกันคือ LPC แต่จากการศึกษาพบว่าถึงแม้จำนวนพารามิเตอร์ที่ได้จะมีจำนวนน้อยกว่าฟาสท์ฟูเรียร์ และใช้เวลาในการคำนวณน้อยกว่าฟาสท์ฟูเรียร์ก็ตาม แต่จะพบว่าค่าของพารามิเตอร์ที่ได้นั้นไม่ค่อยสื่อความหมายได้

เหมือนฟาสท์ฟูเรียร์ เพราะฉนั้นฟาสท์ฟูเรียร์จึงเหมาะสมในการที่จะใช้วิเคราะห์ในระดับลึกต่อไป

2) ศึกษาระบบโครงข่ายเทียม (Neural Network) แบบต่างๆ ซึ่งจากการศึกษาพบว่าแบบที่เหมาะสมที่สุดคือโครงข่ายประสาทเทียม แบบแพร่กลับ (Backpropagation Neutal Network) ซึ่งจากการทดสอบพบว่าแบบ วันเลเยอร์โครงข่ายประสาทเทียม (One Layer Neural Network) จะเร็วกว่าแบบ ทูเลเยอร์โครงข่ายประสาทเทียม (Two Layer Neural Network) เพราะฉะนั้นในโครงงานนี้จึงเลือกใช้วันเลเยอร์เป็นหลัก

5.1.2 เขียนโปรแกรม และ โปรแกรมช่วย (tools) ต่างๆ ที่จำเป็นและอำนวยความสะดวกต่อการทำงาน ต่อๆไป เช่น

1. โมดูลช่วยในการควบคุมซาวด์การ์ด ในการเก็บเสียง
2. โมดูลฟาสท์ฟูเรียร์ทรานฟอร์ม
3. โมดูลหาระดับพลังงานของเสียงในแต่ละช่วง
4. โมดูลนำเสนอข้อมูลในรูปแบบกราฟฟิก
5. โมดูลพิมพ์ภาพออกเครื่องพิมพ์
6. โมดูลเชื่อมต่อการใช้งานกับผู้ใช้
7. โมดูลที่ช่วยในการตัดคำ

5.1.3 หลังจากนั้นทำการเลือก ตัวแปรที่มีผลต่อระบบ (Parameter) โดยการนำข้อมูลตัวอย่างมาทดสอบและเลือกค่าที่เหมาะสมกับระบบต่อไป ซึ่งได้แก่

1. ความถี่ในการแซมปลิ้ง (sampling rate) : 11025 Hz
2. ขนาดของการแซมปลิ้ง (sampling size) : 8 bits
3. ขนาดของวินโดว์ที่จะเอาเข้าไปในฟาสท์ฟูเรียร์ :128
4. ขนาดของการเลื่อนวินโดว์ : 64
5. ขนาดของวินโดว์ที่จะนำไปเข้า โครงข่ายประสาทเทียม :64x40

5.1.4 เก็บข้อมูลตัวอย่างของเสียงมาทำการวิเคราะห์เพื่อดึงลักษณะจุดเด่นของเสียงเพื่อที่จะนำมาเป็นข้อมูลป้อนเข้าสู่ระบบการเรียนรู้

5.1.5 นำข้อมูลมาทำการเรียนรู้โดยวิธีต่างๆ

5.1.6 ทดสอบประสิทธิภาพในวิธีของแต่ละแบบ จากนั้นนำข้อมูลที่ได้ไปทำการวิเคราะห์เพื่อนำมาเป็นข้อมูลในการแก้ไขโครงสร้างต่อไป

5.1.7 นำผลลัพท์ที่ได้ไปใช้ในงานต่างๆ

5.2 โมดูลที่ช่วยในการบันทึกเสียง

ใช้วิธีการเรียกโปรแกรมภายนอกจากโปรแกรมเดิม (shell) ไปเรียกโปรแกรมที่ใช้ในการอัดเสียง ซึ่งมี 2 แบบให้เลือกคือ

1) เวฟฟอร์แมท (*.WAV)

2) วอคฟอร์แมท (*.VOC)

ซึ่งเวฟฟอร์แมทจะเป็นการเก็บข้อมูลโดยตรง แต่วอคฟอร์แมทจะเป็นการย่อข้อมูลก่อนการเก็บ ซึ่งในโครงงานนี้เลือกใช้ เวฟฟอร์แมท เนื่องจากสามารถเขียนโปรแกรมจัดการโดยตรงได้โดยง่าย และผู้ใช้สามารถระบุลักษณะการเก็บเสียงได้โดยการใส่พารามิเตอร์ลงไปต่อหลังจากชื่อไฟล์ได้ดังนี้ โดยพารามิเตอร์ที่ใส่ไปได้ทั้งหมดสามารถดูได้จาก เฮลพ์ (help) ซึ่งจะปรากฏข้อความดังนี้

```
C:\SB16> WREC /?
Creative Disk Wave Record  Version 1.03
Copyright (c) Creative Labs, Inc., 1992. All rights reserved.
Copyright (c) Creative Technology Ltd, 1992. All rights reserved.

Usage : WREC filename [/B:xx] [/R:xx] [/A:xx] [/S:xx] [/M:xx] [/T:xx]
        [/C:xx] [/Q] [/X="command line"]

/B:xx    set buffer size. Range of xx = 2 - 32
```

/R:xx set number of bits per sample. xx = 8 or 16

/A:xx set recording source. xx = MIC [+CD] [+LINE] [+FM]

/S:xx set sampling rate. xx = 11025, 22050 or 44100 Hz

/C:xx set compression format. xx = ALAW, MULAW or CTADPCM

if /C is used, /R is redundant.

/M:xx set recording mode. xx = MONO or STEREO

/T:xx set time limit for recording. Range of xx = 1 - 65535

/Q set to quiet mode

/X="command line" execute the command line directly where

command line is a set of executable statement.

C:\SB16>

พารามิเตอร์ที่ใช้ในโครงงานนี้มีดังนี้

C:\SB16\>WREC TEMP.WAV /B:32 /R:8 /A:MIC /S:11025 /M:MONO /T:4 /Q

5.3 โมดูลที่ช่วยในการแปลงฟาสท์ฟูเรียร์ทรานฟอร์ม

ค่าของอินพุทที่ใช้ในโดเมนเวลาในโครงงานนี้คือ 128 และ วินโดว์แต่ละอันจะขยับไปคราวละ 64 ตำแหน่ง

เนื่องจากเครื่องที่ใช้ในการทำโครงงานนี้เป็น 386-20 และฟังก์ชันเกี่ยวกับฟาสท์ฟูเรียร์ทรานฟอร์มต้องใช้คำสั่ง ไซน์ (sin) และคอส (cos) ซึ่งทั้งสองฟังก์ชันนี้ใช้เวลาในการคำนวณเป็นอันมาก ในโครงงานนี้จึงได้ทำการหลีกเลี่ยงการใช้ การคำนวณแบบโฟลท และ ฟังก์ชัน ไซน์และคอส โดยที่เราจะสังเกตุเห็นว่าค่าสูงสุดที่ได้จาก ฟูเรียร์ทรานฟอร์ม คือ ช่องที่ 0 ซึ่งค่านี้จะเป็นผลรวมทั้งหมดของ ข้อมูลบนโดเมนเวลา ซึ่งจะมีค่าสูงสุดที่เป็นไปได้คือ 128.0 จากข้อมูลเข้า 128 ตัว และคิดเผื่อทศนิยมไว้ 4 หลัก เพราะฉะนั้นจะเห็นได้ว่า ตัวแปรอินทีเจอร์ มีขนาดไม่เพียงพอที่จะรองรับค่าได้ ในโครงงานนี้จึงเลือกใช้ตัวแปร ลอง (LONG) ในการคำนวณแทนแบบโฟลท และใช้ตัวแปรอาเรย์(array) แบบ อินทีเจอร์ (integer) เนื่องจากเรากำหนดทศนิยมไว้ 4 หลัก ซึ่งตัวแปลอินทีเจอร์สามารถเก็บได้ ผลของการใช้ตัวแปร ลอง แทน แบบ โฟลท ทำให้โปรแกรมฟาสท์ฟูเรียร์ทรานฟอร์มมีความรวดเร็วขึ้นกว่าเดิมประมาณ

2-3 เท่า แต่จะต้องเสียหน่วยความจำไปบางส่วนกับการเก็บค่าของฟังก์ชันไซน์ ส่วนฟังก์ชันคอส ก็สามารถใช้ตารางร่วมกับฟังก์ชันไซน์ได้

เมื่อเราได้ข้อมูลความถี่แล้วเราจะเห็นได้ว่าค่าแมคนิจูด (Magnitude) จะสมมาตรซ้ายและขวา เพราะฉะนั้นข้อมูลที่จะนำไปใช้ต่อไปจะใช้เพียงครึ่งเดียว เพราะฉะนั้นกราฟที่ได้จะมีขนาดตามแกนความถี่เท่ากับ 128/2 = 64

5.4 โมดูลที่ช่วยในการหาพลังงานของเสียง

เราสามารถหาค่าพลังงานของเสียงในวินโดว์ นั้นโดยหารากของผลรวมของยกกำลังสองของข้อมูลเข้าในโดเมนเวลา แต่เรายังมีวิธีที่จะหาได้ง่ายและรวดเร็วกว่าวิธีนี้ก็คือ การหาผลรวมของค่า แมคนิจูด. ในโดเมนความถี่แต่ว่าค่าที่ได้ไม่ใช่ค่าพลังงานแต่จะแปรผันโดยตรงกับค่าพลังงาน ซึ่งกราฟที่ได้สามารถนำไปใช้ในการแบ่งแยกคำได้

5.5 โมดูลที่ช่วยในการแสดงผลทางกราฟฟิก

เนื่องจากในโครงงานนี้จำเป็นต้องใช้การแสดงผลกราฟฟิกในโหมดสูงๆเพื่อที่จะนำมาใช้ในการวิเคราะห์ได้สะดวกและถูกต้อง ซึ่งในโครงงานนี้เลือกใช้โหมด 640x480 จุด 256 สี และปรับค่าสีให้เป็นแบบ เกรสเกล (gray scale) ซึ่งในโหมดนี้ในแต่ละการ์ดจะมีวิธีใช้ไม่เหมือนกัน ซึ่งในโครงงานนี้ใช้การ์ดแสดงผล TSENG LAB ET 3000 ซึ่งโปรแกรมส่วนนี้อยู่ในไฟล์ชื่อ GRAPHIC4.CCC ซึ่งจะรวมฟังชันก์พื้นฐานทางกราฟฟิก ที่สำคัญๆไว้ทั้งหมด

5.6 โมดูลที่ช่วยในการพิมพ์ภาพออกเครื่องพิมพ์

ในโครงงานนี้จำเป็นต้องพิมพ์ภาพกราฟฟิคออกมาเนื่องจากเราไม่สามารถเปรียบเทียบข้อมูลบนหน้าจอกราฟฟิคได้โดยสะดวกจึงได้ทำการเขียนโปรแกรมช่วยในการพิมพ์ภาพออกสู่ เลเซอร์พรินเตอร์ ที่ความละเอียด 300 dpi เนื่องจากระดับของสีที่ออกมาจากเครื่องพิมพ์มีแค่ 2 ระดับแต่ภาพบนจอมีขนาดถึง 256 ระดับ จึงต้องหาวิธีในการแสดงผลที่ใกล้เคียงกับ 256 มากที่สุดโดยให้ใช้ภาพขนาด 8x8 จุดบนเลเซอร์พรินเตอร์มาแสดงแทน 1 จุดบนจอ

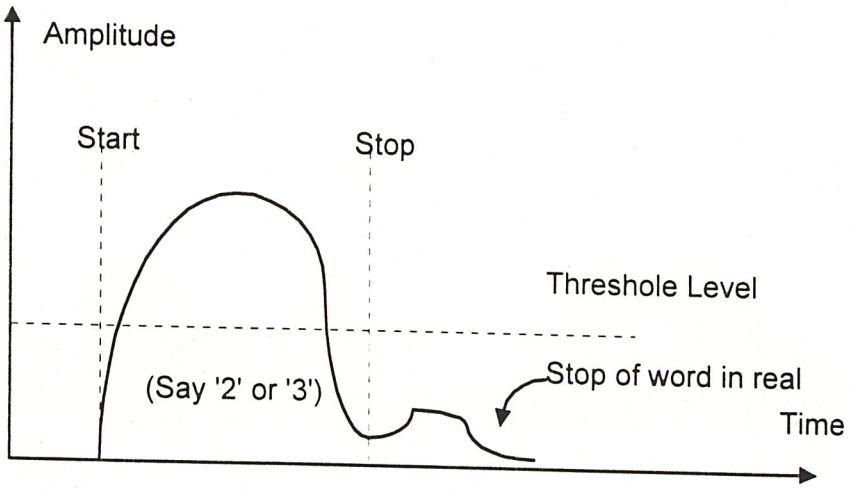
โดยลักษณะการวางจุดในบล็อกให้มีลักษณะกระจายแบบสุ่ม (random) ซึ่งโปรแกรมส่วนนี้อยู่ในไฟล์ชื่อ LASER2.CCC

5.7 โปรแกรมเชื่อมต่อการใช้งานกับผู้ใช้

เนื่องจากการทำงานของโปรแกรมนี้ส่วนใหญ่จะอยู่ในโหมดกราฟฟิก จึงจำเป็นที่จะต้องเขียนโมดูลที่ช่วยในการแสดงผล โดยใช้ ฟอนต์ของ โปรแกรม CW ในไฟล์ normal.fon และ ส่วนการนำข้อมูลเข้าก็จำเป็นที่ต้องเขียน ไลน์อีดิตเตอร์ (Line Editer) ขึ้นมาใช้งานเอง ส่วนของ การใช้งานของโปรแกรมโดยส่วนใหญ่จะเป็นแบบ ฮ็อทคีย์ (Hot Key)

5.8 อัลกอริทึมที่ช่วยในการตัดคำ

ในตอนแรกของโครงงานนี้ได้ใช้วิธีตั้งระดับในการตรวจเช็คค่าพลังงาน ปรากฏว่าใช้ได้ดีกับตัวเลขทั้งหมดโดยส่วนใหญ่แต่จะมีปัญหากับตัวเลข 2 กับ 3 ซึ่งจะมีลักษณะดังรูปที่ 5.1



Picture 5.1 : Error of word sepperator

ซึ่งจะทำให้เสียงในตอนท้าย ซึ่งเป็นเสียงในลำคอหายไป จึงได้นำวิธี สเตทเมนต์แมชชีนเข้ามาช่วยแก้ปัญหานี้ซึ่งทั้งหมดมีขั้นตอนคร่าวๆดังนี้

1. ใช้วิธีตรวจเทรชโฮล และ ความกว้างของสัญญาณ มาตรวจจับจุดเริ่มต้นของคำ
2. เมื่อตรวจจับจุดเริ่มต้นคำได้แล้วก็จะเป็นการตรวจหาจุดสิ้นสุดของคำโดยการสแกน (scan) หาค่าที่น้อยที่สุดจากระยะเริ่มต้นคำไปยังท้ายคำ
3. จากนั้นให้หาจุดเริ่มต้นของคำใหม่ที่อยู่ถัดไป
4. สแกนจากจุดจบของคำแรกไปจนถึงจุดเริ่มต้นของคำใหม่โดยใช้สแตทเมนท์แมชชีน (Statement Machine) ดังต่อไปนี้
   - เริ่มต้นด้วยให้ state=0
   - state=1 เมื่อค่ามากกว่า เทรชโฮลที่ตั้งไว้
   - state=2 เมื่อค่าต่ำกว่า เทรชโฮล และ ค่าสเตทเดิมเป็น 1
   - ถ้า state=2 แล้วให้หาจุดจบของคำใหม่โดยให้ลบค่าเก่าทิ้งเสียแต่ถ้า state<> 2 ให้ยึดถือค่าเก่าไว้

5.9 การเก็บข้อมูลตัวอย่าง

ก่อนที่เราจะบันทึกเก็บข้อมูลลงไปควรที่จะคำนึงถึง สิ่งต่างๆที่จะมีผลกระทบกับข้อมูลที่จะจัดเก็บ เพื่อให้ข้อมูลที่ได้มีลักษณะการกระจายครอบคลุมทุกกรณี ซึ่งได้แก่ ความดัง,ความถี่,ความเร็ว ในการพูด และ จากการทดลองครั้งแรกพบว่าข้อมูลตัวอย่างควรมีการพูดสลับกันไปแบบสุ่ม เนื่องจากในครั้งแรกๆของการทดลอง ได้ใช้เสียงของตัวเลขที่ต่อเนื่องที่เพิ่มขึ้นสองชุด 12345,67890 ซึ่งจากการวัดประสิทธิภาพพบว่า การพูดเรียงกันของตัวเลขในแบบอื่นๆจะทำให้เกิดข้อผิดพลาดได้โดยง่าย เช่น ในขั้นตอนของการเรียนรู้ เราป้อนคำพูดเลข 5 ที่ตามหลังจาก เลข 4 เท่านั้น จากการทดสอบประสิทธิภาพเราพบว่าเสียงเลข 5 จะผิดพลาดได้ง่ายถ้าพูดตามหลังอื่นๆ เช่นจาก เสียงเลข 9 ไปยังเสียงเลข 5 เป็นต้น ที่เป็นเช่นนี้ก็เพราะว่า การพูดตัวเลขหลายตัวต่อเนื่องกัน เสียงแต่ละเสียงจะมีผลต่อกันและกัน ทำให้ ขั้นตอนการรวบรวมข้อมูลต้องคำนึงถึงปัญหาลักษณะนี้ที่จะเกิดด้วย โดยการให้พูดเรียงสลับกันไปแบบแรนดอม (random) ข้อสังเกตุที่พบในการเรียนรู้ครั้งแรกๆอีกอย่างก็คือ การที่ผู้พูดพูดตัวเลขหลายตัวยาวต่อเนื่อง จะทำได้ตัวเลขท้ายๆมีข้อผิพลาดมากกว่าตัวเลขที่พูดในตอนแรกๆ ทำลักษณะการเก็บขอ้มูลในครั้งต่อๆมาต้องคำนึงถึงปัญหาที่จะเกิดนี้ด้วย โดยสรุปแล้วการเก็บข้อมูลเสียงที่พิจารณาแล้วว่าเหมาะสมในโครงการนี้มีดังนี้

1. ใช้ ความดัง 3 ระดับ ได้แก่ เสียงดัง,เสียงดังปานกลาง,เสียงค่อย

2. ใช้ ความเร็ว 3 ระดับ ได้แก่ พูดเร็ว,พูดเร็วปานกลาง,พูดช้า
3. ใช้ ความถี่ 3 ระดับ ได้แก่ เสียงสูงกว่าปกติ,เสียงธรรมดา,เสียงต่ำๆ
4. กลุ่มของตัวเลขมีขนาดกลุ่มละ 7 ตัว
5. ตัวเลขที่จะได้แต่ละตัวเกิดจากการ สุ่ม(random) ขึ้นมา

ปัญหาที่พบอีกอย่างหนึ่งตัวอย่างของเสียงที่จัดเก็บมีเป็นจำนวนมาก ทำให้ต้องมีการออกแบบการจัดเก็บเพื่อสะดวกแก่การใช้งานซึ่งมีลักษณะขั้นตอนดังนี้

1. ให้ผู้ที่จะอัดเสียงทำการป้อนชื่อไฟล์ที่ต้องการเก็บข้อมูล
2. สุ่มตัวเลขขึ้นมา 7 ตัวแล้วแสดงบนหน้าจอ พร้อมทั้งสุ่มคุณสมบัติที่ผู้พูดต้องพูดตาม ซึ่งได้แก่ ความดัง,ความถี่,ความเร็ว
3. ทำการอัดเสียง และบันทึกเสียงลงไฟล์ โดยนำตัวเลขที่เป็นครั้งที่ของการบันทึกเสียงมาต่อกับชื่อไฟล์ที่ผู้ใช้ระบุในตอนแรกๆ เพื่อความสะดวดในการค้นหาในภายหลัง ซึ่ง รูปแบบที่ใช้คือ เวฟฟอร์แมท (*.WAV)
4. ทำซ้ำข้อ 2 และ 3 ไปเรื่อยๆ จนได้ข้อมูลเสียงที่ครบทุกกรณี
5. ให้ผู้ใช้ระบุชื่อไฟล์ที่ต้องการเก็บความถี่เสียง
6. นำข้อมูลเสียงที่ได้ทั้งหมดมาทำการคำนวนหาส่วนประกอบของความถี่ของเสียง (FFT) และค่าพลังงานของเสียงแล้วแสดงบนหน้าจอกราฟฟิก
7. จากนั้นใช้การตัดคำโดยนำข้อมูลจากค่าพลังงานของเสียงมาพิจารณา
8. ให้ผู้ใช้ระบุว่าเสียงที่ตัดออกมาได้นั้นเป็นเสียงอะไรแล้วทำการเก็บเฉพาะความถี่เสียงตรงนั้นและความหมายของเสียงที่ผู้ใช้ระบุบันทึกลงไฟล์ควบคู่กันไปต่อกันไปเรื่อยๆ ในไฟล์ที่ผู้ใช้ระบุในข้อ 5
9. ทำข้อ 6, 7 และ 8 ไปเรื่อยๆ จนครบ

5.10 ขั้นตอนในการเรียนรู้

เริ่มต้นด้วยการสุ่มค่าน้ำหนักขึ้นมาจากนั้นทำการ แบ่งแยกโปรแกรมให้สามารถคำนวณบนเครื่องได้หลายเครื่องพร้อมกัน เนื่องจากการออกแบบได้ใช้จำนวน โหนด 8 โหนด จึงสามารถแบ่งโปรแกรมได้สูงสุดเป็น 8 โปรแกรม เมื่อเรียนรู้เสร็จแล้วก็ทำการทดสอบประสิทธิภาพในการจำ ถ้าข้อมูลใดจำไม่ได้หรือจำผิด ก็ให้นำข้อมูลเหล่านั้นมาทำการเรียนรู้ใหม่ ทำเช่นนี้ไปเรื่อยๆ ซึ่งจะเป็นการที่ทำให้ระบบมีประสิทธิภาพเพิ่มมากขึ้น

บทที่ 6
ผลการทดลอง

6.1 ผลจากการทดสอบ

ในเทอมแรกได้นำข้อมูลทั้งหมดมาคำนวณบนเครื่อง HP 370 จำนวน 2 เครื่อง พร้อมๆกัน ปรากฏว่าใช้เวลาคำนวณ 11.5 ชั่วโมง หรือ 23 ชั่วโมงบนเครื่องเดียว ปรากฏว่าจำได้ประมาณ 94 % และใช้เวลาในการตรวจสอบเสียง ประมาณ 1 นาที ต่อข้อมูล 1 วินาที

ในภาคการศึกษาที่ 2 ได้ทำการปรับปรุงขั้นตอนในการตรวจสอบเสียง และทำการแบ่งโปรแกรมออกเป็น 4 โปรแกรม เพื่อที่จะนำไปคำนวณบนเครื่อง 486-33 จำนวน 4 เครื่อง พร้อมกัน ปรากฏว่าใช้เวลาประมาณ 4 ชั่วโมง ต่อเครื่อง หรือใช้เวลาในการคำนวณประมาณ 4x4 = 16 ชั่วโมงบนเครื่องๆเดียว และใช้จำนวนครั้งในการปรับค่าน้ำหนักประมาณ 140 ครั้ง ต่อ 1 โหนด ซึ่งในครั้งที่สองนี้ได้ปรับปรุงการวัดประสิทธิภาพขึ้นมาใหม่ให้มีมาตรฐานขึ้นกว่าเดิม ซึ่งผลการวัดจำได้ประมาณ 95.6 % และความเร็วในการคำนวณประมาณ 0.58 วินาที ต่อข้อมูล 1 วินาที (ไม่นับเวลาในการ sampling) บนเครื่อง 486 DX2

6.2 เกณฑ์การวัดประสิทธิภาพในโครงงานนี้

ในการวัดประสิทธิภาพในตอนแรกเป็นการทดลองพูดไปเรื่อยๆ แตกต่างกันออกไปซึ่งจะทำให้มาตราฐานการวัดไม่ดีเท่าที่ควร ฉะนั้นในเทอมต่อมา จึงได้ปรับปรุงวิธีการวัดให้มีมาตราฐานขึ้น โดยการทดลองพูดคราวละ 7 ตัวเลข จำนวน 378 ครั้ง โดยลักษณะการพูดในแต่ละครั้งจะแตกต่างกันออกไปตามลักษณะ 3 อย่างดังนี้

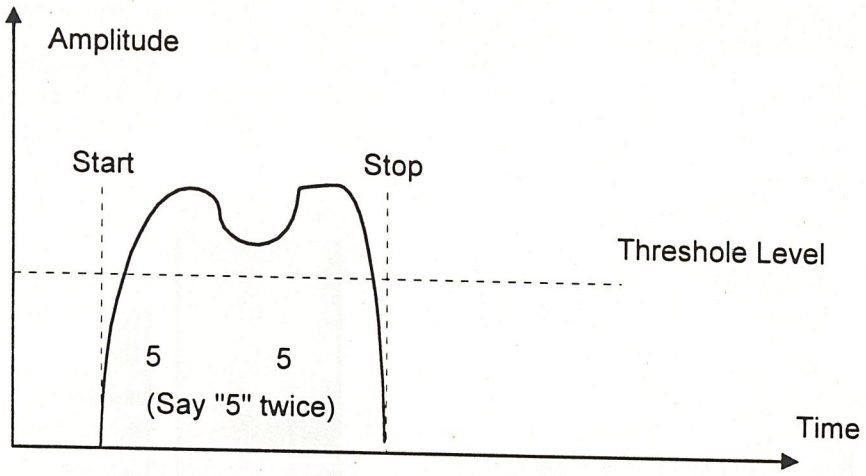
1. ความดังของเสียง มี 3 ระดับคือ เสียงดัง, ปานกลาง และเสียงค่อย
2. ความเร็ว มี 3 ระดับคือ เร็ว, ปานกลาง และช้า
3. ความถี่ของเสียง มี 3 ระดับคือ สูง, กลาง และต่ำ

รวมเป็นทั้งหมด 3*3*3 = 27 แบบ โดยแต่ละแบบให้พูดตัวเลข 2 ชุด

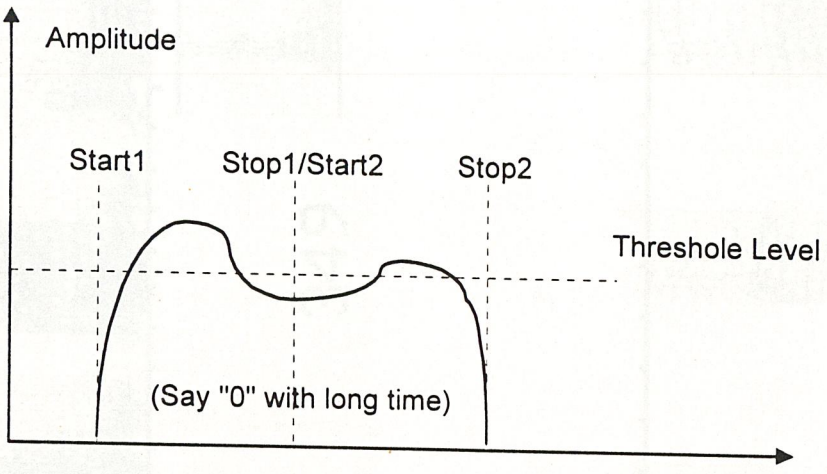
เพราะฉะนั้น เสียงที่ใช้ทดสอบทั้งหมดจึงเป็น 27 * 2 * 7 = 378 ตัวเลข
ผลการทดสอบพบว่ามี error ประมาณ 5 % ซึ่งเกิดจาก 2 กรณีใหญ่ ๆ คือ

1. เกิดจากการผิดพลาดจากการตัดคำ ประมาณ 1.8 % ซึ่งมักจะเกิดจากการพูดที่ต่อเนื่อง และเร็ว หรือพูดยาวมาก ๆ ซึ่งการพูดต่อเนื่อง และ เร็วๆ จะทำให้คำสองคำเชื่อมติดต่อกันเหมือนเป็นคำๆเดียวกันดังรูปที่ 6.1 และ การพูดยาวๆ ของคำบางคำเช่นเลข 0 จะทำให้ดูเหมือนเป็นคำสองคำดังรูปที่ 6.2

2. เกิดจากความผิดพลาดจากโครงข่ายประสาทเทียม ประมาณ 2.8 %



Picture 6.1 : Error of word seperator



Picture 6.2 : Error of word seperator

บทที่ 7
สรุปและวิจารณ์

ปัญหาที่พบ

1. เนื่องจากว่าหน่วยความจำสำรอง (buffer memory) ของเลเซอร์พรินเตอร์มีค่อนข้างจำกัดจึงทำให้ไม่สามารถที่พิมพ์ออกมาได้หมด

2. ต้องใช้หน่วยความจำเป็นจำนวนมากในการเก็บข้อมูลทำให้ไม่สามารถ ดีบัก (debug) โปรแกรม บน คอมไพเลอร์ได้โดยตรงได้ ทำให้การพัฒนาโปรแกรมเป็นไปด้วยความยากลำบาก

3. เนื่องจากเราไม่มีคู่มือการใช้งานของ การ์ดเสียง (sound card) ที่ละเอียดเพียงพอทำให้เราไม่สามารถที่จะเขียนโปรแกรมเจาะไปที่ การ์ดได้โดยตรง

4. เนื่องจากโครงงานนี้จำเป็นที่จะต้องใช้ กราฟฟิกในโหมดสูงๆ จึงจำเป็นที่จะต้องสร้างโปรแกรมกราฟฟิก ตลอดจนฟังก์ชันทุกอย่างขึ้นมาเองทำให้ต้องเสียเวลาไปกับส่วนนี้มาก

5. ไม่สามารถใช้โปรแกรมที่มีอยู่โดยทั่วๆไปมาจับภาพบนหน้าจอได้เลยแม้แต่โปรแกรมวินโดวก็ไม่สามารถที่จะจับได้ จึงต้องมีการเขียนโปรแกรมในส่วนของการพิมพ์ภาพออกสู่เลเซอร์ปรินเตอร์เองซึ่งก็ใช้ได้ดีระดับหนึ่งเท่านั้น

สรุป

จากกราฟความถี่จะทำให้เราเห็นส่วนประกอบ4อย่างหลักๆได้แก่

1) เสียงพยัญชนะ ซึ่งจะอยู่บริเวณเริ่มต้นของคำ
2) เสียงสระ จะปรากฏอยู่บริเวณช่วงตรงกลางของเสียง
3) เสียงวรรณยุกต์ จะเป็นลักษณะความเป็นไปของเสียง อาจจะเรียกได้ว่าความโค้งนั่นเองซึ่งอยู่บริเวณตรงกลางของเสียงเช่นกัน
4) เสียงตัวสะกด จะปรากฏอยู่ตอนท้ายของคำในช่วงสั้นๆ

จากองค์ประกอบทั้ง 4 อย่างจะเห็นได้ว่า เสียงตัวสะกดตรวจจับได้ยากที่สุด โดยเฉพาะเสียงกัก เช่นเสียงคำว่า 'ปก' และ 'ปด' จากความถี่ที่ได้แทบจะไม่เห็นความแตกต่างเลย หรือในเสียงตัวอักษรภาษาอังกฤษตัว เอ็ม (m) และ เอ็น (n) ก็จะแทบไม่เห็นความแตกต่างเลยทีเดียว ซึ่งนับว่าเป็นความโชคดีที่กลุ่มเสียงที่เลือกใช้ในโครงงานนี้เป็นเสียงเลข 0-9 ซึ่งทั้งสิบตัวนี้ไม่มีคุณสมบัติดังที่กล่าวมา วิธีแก้ไขเมื่อพบปัญหาเช่นนี้ก็คือ อาจใช้วิธีตรวจจับลักษณะของเสียงด้วยวิธีอื่นๆมาร่วมพิจารณาด้วย เช่นวิธี LPC เป็นต้น

จากผลการทดลองสรุปได้ว่า โครงข่ายประสาทเทียมมีความเหมาะสมกับการจำเสียงได้ดี โดยความเร็วที่เหมาะสมในการพูดโดยเฉลี่ยจะอยู่ในช่วง 0.3 - 0.5 วินาที ต่อหนึ่งคำ โดยที่มีโอกาสที่จะผิดพลาดอยู่ที่ประมาณ 4.6% จะเห็นได้ว่า การที่จะให้ error ลดลงได้นั้นมี 2 วิธี คือ ปรับปรุง algolithm ในการตัดคำให้ดีขึ้น หรือปรับปรุง ค่าน้ำหนัก ของ โครงข่ายประสาทเทียมให้ดีขึ้น โดยการนำเสียงที่จำผิดพลาดมาทำการ เรียนรู้ใหม่เรื่อย ๆ

จากเวลาที่ใช้ในการคำนวณประมาณ 0.58 วินาทีต่อข้อมูล 1 วินาที แสดงว่า ระบบนี้สามารถนำไปใช้ในงานที่ต้องใช้ความต่อเนื่องได้ แต่ยังไม่ถึงขนาดที่จะสามารถโต้ตอบได้ทันที ทันใด(real time) โดยเวลาที่เสียไปส่วนใหญ่จะไปอยู่ที่การแปลง ฟาสท์ฟูเรียร์ทรานฟอร์มมากกว่าตรงส่วนของโครงข่ายประสาทเทียม

ในเรื่องของการวัดประสิทธิภาพของระบบ เป็นการยากที่จะระบุมาตฐานในการวัดลงไปให้แน่ชัดลงไปว่าควรจะต้องเป็นอย่างไรเพราะฉะนั้นประสิทธิภาพของระบบจึงเป็นเพียงเครื่องชี้แสดงให้เห็นการพัฒนาของระบบนั้นๆเท่านั้น จึงไม่สามารถเอาตัวเลขนี้ไปเปรียบเทียบกับระบบอื่นๆได้ นอกเสียจากว่าจะมีระบบการวัดที่เหมือนกันทุกประการ

บทที่ 8
แนวทางการพัฒนาต่อไป

จะเห็นได้ว่าการที่จะเอาไปใช้งานได้จริงนั้นเนื่องจากงานที่จะนำไปใช้ส่วนใหญ่จะต้องเป็นงานที่โต้ตอบฉับพลัน (real time) เช่น การนำไปใช้ควบคุมอุปกรณ์ผ่าตัดในวงการแพทย์ด้วยเสียง จึงจำเป็นที่จะต้องพัฒนาอุปกรณ์ทาง ฮาร์ดแวร์ (Hardware) ที่มีประสิทธิภาพสูงยิ่งๆขึ้นไป อาจพัฒนาได้โดยการสร้างฮาร์ดแวร์ที่สามารถเก็บตัวอย่างเสียงได้รวดเร็วขึ้น หรือการสร้างระบบจดจำเสียงนี้ในรูปแบบของฮาร์ดแวร์ ซึ่งเป็นฮาร์ดแวร์มีทั้งส่วนเก็บตัวอย่างเสียง (แปลงสัญญาณเสียงจากอนาล็อกเป็นดิจิตอล,แปลงรูปแบบข้อมูลจากโดเมนเวลาเป็นฟังก์ชันโดเมนเวลา และความถี่,การตัดแยกพยางค์) และส่วนการจัดแยกเสียงว่าเป็นเสียงของคำว่าอะไร

แต่อย่างไรก็ตามระบบโครงข่ายประสาทเทียมนี้ยังไม่เหมาะสมกับงานที่เสี่ยงอันตรายเพราะถึงแม้ว่าค่าผิดพลาดที่ได้จากการทดสอบน้อยมากก็ตาม แต่ก็ไม่มีอะไรที่จะมารับประกันความผิดพลาดที่อาจจะเกิดขึ้นได้

โดยสรุปแล้วจากโครงงานนี้จะเห็นได้ว่า โครงข่ายประสาทเทียมมีความเหมาะสมในที่จะใช้ในการจดจำเสียง จึงสมควรที่จะพัฒนาระบบโครงข่ายประสาทเทียมให้มีโครงสร้างที่ดียิ่งขึ้นต่อไป

กิติกรรมประกาศ

ในเวลาที่ผ่านมาคณะผู้จัดทำสามารถทำปริญญานิพนธ์ฉบับนี้สำเร็จลงได้ ก็เนื่องจาก การได้รับความร่วมมือ และความช่วยเหลือจากบุคคลหลายฝ่าย ซึ่งมีส่วนช่วยให้การวิจัยโครงงานในครั้งนี้สำเร็จลุล่วงไปด้วยดี

ทางคณะผู้จัดทำ ขอขอบพระคุณ ดร.บุญธีร์ เครือตราชู ซึ่งเป็นอาจารย์ที่ปรึกษาที่ได้ให้คำแนะนำ ความรู้ต่างๆ ที่เป็นประโยชน์ในการศึกษา และวิจัยโครงงานนี้ และให้ความอนุเคราะห์ ในเรื่องเครื่องคอมพิวเตอร์ที่นำไปแสดงให้ผู้ใช้ดู และขอขอบคุณสำหรับความเอาใจใส่ดูแลโครงงานนี้จึงสำเร็จ

ขอขอบคุณอาจารย์อภิเนตร อูนากูล อาจารย์ วัชระ ฉัตรวิริยะ พี่ปู และไพโรจน์ ที่ได้ให้คำแนะนำ และความช่วยเหลือเป็นอย่างดีต่อผู้จัดทำจนโครงงานนี้สำเร็จ

ขอขอบพระคุณบิดา-มารดา ที่ให้ความดูแลเอาใจใส่ และเพื่อนๆพี่ๆที่ให้ความร่วมมือเป็นอย่างดีต่อคณะผู้จัดทำจนโครงงานนี้สำเร็จ

คณะผู้จัดทำ

หนังสืออ้างอิง

1. D.O. Shaughnessy.," Speech Communication ", Addison Wesley Publishing Company, pp.204-241,1987.

2. O.E. Brigham., " The Fast Fourier Transform and its applications ",Prentice-Hall international. Inc., 1988

3. Paul M. Embree and Bruce Kimble, " C Language Algorithms for Digital Signal Processing",Prentice-Hall International,Inc. , Englewood Cliffs, N.J.,1991

4.ศ.ดร. วัลลภ สุระกำพลธร, " การประมวลผลสัญญาณเชิงตัวเลข(Digital Signal Processing) ",บริษัทไดนาพรินท์ จำกัด,สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,กรุงเทพฯ,2533