บทที่ 4

การทดลองและการวิเคราะห์ผลการทดลอง

ในขั้นตอนการทดลองมีการเก็บข้อมูลเพื่อใช้ในการทดสอบประสิทธิภาพของระบบและการกำหนดตัวแปร (Parameter) ที่ใช้ในการฝึกโครงข่ายประสาทเทียม ซึ่งมีรายละเอียดดังต่อไปนี้

4.1 การทดลอง

4.1.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองได้จากการเก็บตัวอย่างเสียงพูดคำโดดภาษาไทยของผู้พูดจำนวน 10 คน ในสภาพแวดล้อมที่ไม่มีเสียงรบกวน (Signal to Noise Ratio > 10 dB) โดยผู้พูดแต่ละคนทำการบันทึกเสียง 20 ครั้ง ในแต่ละครั้งมีจำนวนคำพูด 7 คำคือ “เลี้ยวซ้าย” “เลี้ยวขวา” “เดินหน้า” “ถอยหลัง” “เร็วขึ้น” “ช้าลง” และ “หยุด” หลังจากนั้นแบ่งข้อมูลเสียงพูดออกเป็น 2 กลุ่ม กลุ่มแรก คือ กลุ่มตัวอย่าง (Training set) ที่ใช้ในการฝึกโครงข่าย เป็นเสียงคำพูด 10 ครั้ง จำนวน 700 คำ (10 คน x 10 ครั้ง x 7 คำ) โดยเสียงพูดในกลุ่มนี้จะถูกดึงค่าลักษณะสำคัญแล้วปรับจำนวนให้เท่ากัน (ช่วงต้นและช่วงท้าย) ก่อนจะบันทึกไว้ เพื่อใช้เป็นฐานข้อมูลสำหรับฝึกโครงข่ายทั้ง 2 โครงข่าย และกลุ่มที่สอง คือ กลุ่มทดสอบ เป็นเสียงคำพูดอีก 10 ครั้งที่เหลือจำนวน 700 คำ ใช้ในการทดสอบระบบรู้จำแบบขึ้นกับผู้พูด (Speaker-dependent)

นอกจากนั้นยังมีจากการเก็บตัวอย่างเสียงพูดของผู้พูดอีก 5 คน โดยผู้พูดในกลุ่มนี้จะทำการบันทึกเสียง 10 ครั้ง จำนวน 350 คำ เพื่อใช้ในการทดสอบระบบรู้จำแบบไม่ขึ้นกับผู้พูด (speaker-independent)

4.1.2 การฝึกโครงข่ายประสาทเทียมโดยใช้โปรแกรม MATLAB

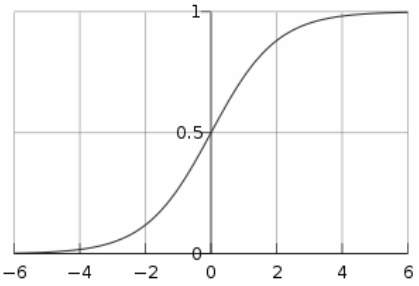
เริ่มจากการอ่านข้อมูลที่ใช้ในการฝึกโครงข่ายแต่ละโครงข่ายซึ่งมีจำนวน 125 x 700 ค่า (จำนวนโหนดในชั้นอินพุท x จำนวนคำในกลุ่มตัวอย่าง) จากนั้นสร้างโครงข่ายประสาทเทียมแบบ Feed-forward แล้วนำข้อมูลทั้งหมดป้อนเป็นข้อมูลอินพุทของโครงข่ายซึ่งมีชั้นซ่อนตัว 3 ชั้น

โดยในแต่ละชั้น ข้อมูลจะถูกส่งผ่านฟังก์ชันกระตุ้นแบบซิกมอยด์ (สมการที่ 4.1) จนถึงชั้นเอาท์พุต เพื่อหาค่าผิดพลาดระหว่างข้อมูลในชั้นเอาท์พุตกับค่าที่ต้องการ (ค่าที่ต้องการ คือ ค่าที่กำหนดให้เป็นตัวแทนของรูปแบบข้อมูลที่ป้อนเข้ามา) โดยเลือกใช้ฟังก์ชัน trainscg (Scaled conjugate gradient algorithm) ในการฝึกฝนเพื่อปรับค่าถ่วงน้ำหนักของโครงข่าย เมื่อค่าเอาท์พุตจริงของโครงข่ายเข้าใกล้ค่าที่ต้องการ (ค่าผิดพลาดลดลงต่ำกว่าค่าที่กำหนดไว้) จะหยุดการฝึกแล้วบันทึกค่าถ่วงน้ำหนัก (Weight & bias) ของโครงข่ายเก็บไว้

ตารางที่ 4.1 ค่าเอาท์พุตที่เป็นตัวแทนของคำที่ต้องการรู้จำ

คำที่ต้องการรู้จำ	ค่าที่ต้องการในชั้นเอาท์พุต
เลี้ยวซ้าย	[1; 0; 0; 0; 0; 0; 0]
เลี้ยวขวา	[0; 1; 0; 0; 0; 0; 0]
เดินหน้า	[0; 0; 1; 0; 0; 0; 0]
ถอยหลัง	[0; 0; 0; 1; 0; 0; 0]
เร็วขึ้น	[0; 0; 0; 0; 1; 0; 0]
ช้าลง	[0; 0; 0; 0; 0; 1; 0]
หยุด	[0; 0; 0; 0; 0; 0; 1]

$$a = \frac{1}{1+e^{-n}} \tag{4.1}$$



รูปที่ 4.1 ฟังก์ชัน Logistic Sigmoid

ตัวแปรที่ใช้ในการฝึกโครงข่ายประสาทเทียม คือ ค่าความผิดพลาดต่ำสุดที่ต้องการ (Goal) โดยปกติจะเป็นค่าความผิดพลาดเฉลี่ยของโหนดในชั้นเอาท์พุต (Mean squared error: mse) และจำนวนรอบ (Epochs) ในการฝึกโครงข่าย ส่วนค่า Parameter อื่นๆ จะใช้ค่าเริ่มต้นของ Neural Networks Toolbox

ในการฝึกโครงข่ายจะกำหนดให้หยุดการฝึกเมื่อค่า mse น้อยกว่า 0.00001 ในการทดลองครั้งที่ 1, น้อยกว่า 0.0001 ในการทดลองครั้งที่ 2 และน้อยกว่า 0.001 ในการทดลองครั้งที่ 3 โดยใช้จำนวนรอบในการฝึกไม่เกิน 250 รอบ

4.1.3 การทดสอบประสิทธิภาพการรู้จำ

การทดสอบประสิทธิภาพการรู้จำ เป็นการใช้ค่าลักษณะสำคัญช่วงต้นและช่วงท้ายของเสียงพูดในกลุ่มทดสอบที่ปรับจำนวนแล้ว (125 ค่า จำนวน 2 ชุด) ป้อนให้ชั้นอินพุทของโครงข่าย โดยนำค่าถ่วงน้ำหนักที่ได้จากขั้นตอนการฝึกมาใช้คำนวณหาค่าผลลัพธ์ของระบบ เนื่องจากจำนวนโหนดในชั้นเอาท์พุตของโครงข่ายเท่ากับจำนวนคำที่ต้องการรู้จำ แต่จะมีเพียง 1 โหนดที่แสดงค่าของรูปแบบข้อมูลที่ป้อนเข้ามา ดังนั้นการแยกแยะเสียงพูดจึงใช้วิธีการตรวจเลือกค่าสูงสุดของโหนดในชั้นเอาท์พุต ซึ่งแต่ละโหนดจะแทนคำพูดแต่ละคำ โดยที่ค่าสูงสุดนั้นต้องมากกว่าค่าการตัดสินใจ (กำหนดให้เท่ากับ 0.01) จึงจะเป็นการรู้จำเสียงพูดที่ถูกต้อง

ในการทดสอบการรู้จำในช่วงแรกจะทดสอบกับระบบบนคอมพิวเตอร์ที่สร้างโดยใช้โปรแกรม MATLAB เปรียบเทียบกับระบบบนอุปกรณ์ FPGA ซึ่งค่าผลลัพธ์ที่ได้จะใกล้เคียงกัน เนื่องจากระบบมีการทำงานที่เหมือนกัน ดังนั้นเพื่อความรวดเร็วในการทดสอบประสิทธิภาพการรู้จำของระบบบนอุปกรณ์ FPGA จึงใช้ผลลัพธ์ที่ได้จากระบบบนคอมพิวเตอร์แทน

4.2 ผลการทดลอง

4.2.1 เวลาในการทำงานของระบบรู้จำเสียงพูดบนอุปกรณ์ FPGA

ระบบบนอุปกรณ์ FPGA ใช้ MicroBlaze เป็นหน่วยประมวลผลข้อมูล ซึ่งเวลาโดยประมาณที่ใช้ในการประมวลผลแต่ละขั้นตอน แสดงในตารางที่ 4.2

ตารางที่ 4.2 เวลาในการประมวลผลของระบบบนอุปกรณ์ FPGA

กระบวนการ	เวลา (วินาที)
การเน้นล่วงหน้า (Preemphasis)	0.0331
การเพิ่มค่าความแตกต่างของเฟรมรอบข้าง	0.0437
การดึงค่าลักษณะสำคัญ (Feature extraction) และค่าพลังงานเสียง (Energy)	0.4047
การหาค่าความแตกต่างของสัมประสิทธิ์ในแต่ละเฟรม (Euclidean distance)	0.0008
การตรวจหาขอบเขตของคำ (Endpoint Detection)	0.0002
การปรับจำนวนค่าลักษณะสำคัญให้เท่ากัน	0.0001
การรู้จำรูปแบบด้วยโครงข่ายประสาทเทียม (Neural Networks)	0.0100

4.2.2 ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูด

ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูด คือ ความถูกต้องในการรู้จำเสียงคำพูดในกลุ่มทดสอบจำนวน 700 คำ ของผู้พูด 10 คน ที่เป็นผู้บันทึกเสียงในกลุ่มตัวอย่าง ได้ผลการทดลองดังนี้

ผลการรู้จำครั้งที่ 1 รู้จำถูกต้อง 695 เสียงพูด คิดเป็นร้อยละ 99.29 (ตารางที่ 4.3)

ผลการรู้จำครั้งที่ 2 รู้จำถูกต้อง 699 เสียงพูด คิดเป็นร้อยละ 99.86 (ตารางที่ 4.5)

ผลการรู้จำครั้งที่ 3 รู้จำถูกต้อง 698 เสียงพูด คิดเป็นร้อยละ 99.71 (ตารางที่ 4.7)

4.2.3 ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูด

ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูด คือ ความถูกต้องในการรู้จำเสียงคำพูดจำนวน 350 คำ ของผู้พูด 5 คน ที่ไม่ใช่ผู้บันทึกเสียงในกลุ่มตัวอย่าง ได้ผลการทดลองดังนี้

ผลการรู้จำครั้งที่ 1 รู้จำถูกต้อง 305 เสียงพูด คิดเป็นร้อยละ 87.14 (ตารางที่ 4.4)

ผลการรู้จำครั้งที่ 2 รู้จำถูกต้อง 317 เสียงพูด คิดเป็นร้อยละ 90.57 (ตารางที่ 4.6)

ผลการรู้จำครั้งที่ 3 รู้จำถูกต้อง 319 เสียงพูด คิดเป็นร้อยละ 91.14 (ตารางที่ 4.8)

ตารางที่ 4.3 ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.00001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	10	10	10	10	10	10	70
2	9	10	10	10	10	10	10	69
3	10	9	10	10	10	10	10	69
4	10	10	10	10	10	10	10	70
5	10	10	10	10	10	10	10	70
6	8	10	10	9	10	10	10	67
7	10	10	10	10	10	10	10	70
8	10	10	10	10	10	10	10	70
9	10	10	10	10	10	10	10	70
10	10	10	10	10	10	10	10	70

ตารางที่ 4.4 ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.00001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	7	10	7	8	10	8	60
2	9	10	10	10	10	10	10	69
3	4	9	6	2	10	10	10	51
4	10	10	9	10	10	10	10	69
5	7	8	9	8	10	5	9	56

ตารางที่ 4.5 ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.0001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	10	10	10	10	10	10	70
2	10	10	10	10	10	10	10	70
3	10	10	10	10	10	10	10	70
4	10	10	10	10	10	10	10	70
5	10	10	9	10	10	10	10	69
6	10	10	10	10	10	10	10	70
7	10	10	10	10	10	10	10	70
8	10	10	10	10	10	10	10	70
9	10	10	10	10	10	10	10	70
10	10	10	10	10	10	10	10	70

ตารางที่ 4.6 ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.0001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	5	9	8	8	10	9	59
2	9	10	10	10	10	10	10	69
3	8	10	7	9	10	10	10	64
4	10	10	9	9	10	10	10	68
5	7	7	10	9	10	5	9	57

ตารางที่ 4.7 ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	10	10	10	10	10	10	70
2	10	9	10	10	10	10	10	69
3	10	10	10	10	10	10	10	70
4	10	10	10	10	10	10	10	70
5	10	10	10	10	10	10	10	70
6	9	10	10	10	10	10	10	69
7	10	10	10	10	10	10	10	70
8	10	10	10	10	10	10	10	70
9	10	10	10	10	10	10	10	70
10	10	10	10	10	10	10	10	70

ตารางที่ 4.8 ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูด เมื่อกำหนดให้
หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.001

ผู้พูด	ความถูกต้อง							รวม
	เลี้ยวซ้าย	เลี้ยวขวา	เดินหน้า	ถอยหลัง	เร็วขึ้น	ช้าลง	หยุด	
1	10	8	9	10	9	10	10	66
2	10	10	10	10	10	10	10	70
3	5	10	7	9	10	10	10	61
4	10	10	10	9	10	10	10	69
5	7	9	8	6	10	4	9	53

4.2.4 ประสิทธิภาพการรู้จำโดยเฉลี่ย

เนื่องจากประสิทธิภาพการรู้จำของระบบเปลี่ยนแปลงตามค่าถ่วงน้ำหนักที่ได้จากการฝึกโครงข่าย ดังนั้นจึงทำการทดลองหาประสิทธิภาพการรู้จำทั้งแบบขึ้นกับผู้พูดและไม่ขึ้นกับผู้พูด 3 ครั้ง เพื่อหาค่าเฉลี่ยของอัตราการรู้จำที่ค่าการตัดสินใจเท่ากับ 0.01 ซึ่งได้ผลการทดลองดังแสดงในตารางที่ 4.9 และตารางที่ 4.10

ตารางที่ 4.9 ประสิทธิภาพโดยเฉลี่ยของการรู้จำแบบขึ้นกับผู้พูด

ครั้งที่	อัตราการรู้จำแบบขึ้นกับผู้พูด (%)		
	mse สำหรับการฝึกโครงข่าย		
	0.00001	0.0001	0.001
1	99.14	99.57	99.57
2	99.28	99.28	99.85
3	99	99.14	99.28
เฉลี่ย	99.14	99.33	**99.57**

ตารางที่ 4.10 ประสิทธิภาพโดยเฉลี่ยของการรู้จำแบบไม่ขึ้นกับผู้พูด

ครั้งที่	อัตราการรู้จำแบบไม่ขึ้นกับผู้พูด (%)		
	mse สำหรับการฝึกโครงข่าย		
	0.00001	0.0001	0.001
1	85.42	90.28	91.42
2	87.14	92.57	91.14
3	86.85	89.42	93.42
เฉลี่ย	86.47	90.76	**91.99**

4.2.5 ประสิทธิภาพการรู้จำเมื่อปรับค่าการตัดสินใจ

การทดสอบประสิทธิภาพการรู้จำทั้งแบบขึ้นกับผู้พูดและไม่ขึ้นกับผู้พูดที่ผ่านมาได้กำหนดให้ค่าสูงสุดของโหนดในชั้นเอาท์พุตจะต้องมากกว่าค่าการตัดสินใจ (0.01) จึงเป็นผลการรู้จำที่ถูกต้อง ในบางครั้งค่าสูงสุดที่น้อยกว่าค่าการตัดสินใจอาจเป็นคำที่รู้จำได้ถูกต้องแต่ระบบจะรวมคำนั้นเป็นการรู้จำที่ผิดพลาด ดังนั้นในการทดลองนี้จะแยกคำที่มีค่าสูงสุดน้อยกว่าค่าการตัดสินใจซึ่งเป็นคำที่ระบบรู้จำไม่ได้ออกแล้วหาเฉพาะคำที่ระบบรู้จำผิดพลาด จากนั้นทำการเปรียบเทียบประสิทธิภาพการรู้จำเมื่อกำหนดค่าการตัดสินใจ 3 ค่าคือ 0.01, 0.1 และ 0.5 ซึ่งได้ผลการทดลองดังแสดงในตารางที่ 4.11 และตารางที่ 4.12

ตารางที่ 4.11 ประสิทธิภาพการรู้จำแบบขึ้นกับผู้พูดเมื่อมีการปรับค่าการตัดสินใจ

mse	0.00001			0.0001			0.001		
ค่าการตัดสินใจ	0.01	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
คำที่รู้จำไม่ได้	12	21	42	5	26	48	0	14	50
คำที่รู้จำผิดพลาด	0	0	0	2	0	0	1	0	0
คำที่รู้จำถูกต้อง	688	679	658	693	674	652	699	686	650
อัตราการรู้จำแบบขึ้นกับผู้พูด (%)	98.3	97	94	99	96.3	93.1	99.9	98	92.9

ตารางที่ 4.12 ประสิทธิภาพการรู้จำแบบไม่ขึ้นกับผู้พูดเมื่อมีการปรับค่าการตัดสินใจ

mse	0.00001			0.0001			0.001		
ค่าการตัดสินใจ	0.01	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
คำที่รู้จำไม่ได้	38	73	96	28	65	97	6	47	108
คำที่รู้จำผิดพลาด	6	2	0	13	4	2	26	12	2
คำที่รู้จำถูกต้อง	306	275	254	309	281	251	318	291	240
อัตราการรู้จำแบบไม่ขึ้นกับผู้พูด (%)	87.4	78.6	72.6	88.3	80.3	71.7	90.9	83.1	68.6

4.2.6 ประสิทธิภาพการปฏิเสธคำที่ไม่ได้กำหนดไว้

ในการทดสอบประสิทธิภาพการรู้จำ ค่าสูงสุดของโหนดในชั้นเอาท์พุตที่มากกว่าค่าการตัดสินใจ คือ คำที่ต้องการรู้จำ แต่ในกรณีที่ผู้พูดไม่ได้พูดคำที่กำหนดไว้ ค่าสูงสุดของโหนดในชั้นเอาท์พุตควรจะน้อยกว่าค่าการตัดสินใจและระบบจะต้องปฏิเสธคำนั้น ในการทดลองได้ทำการบันทึกเสียงของผู้วิจัยจำนวน 10 ครั้ง ในแต่ละครั้ง เป็นเสียงคำพูดที่ใกล้เคียงกับคำที่ต้องการรู้จำ 6 คำ คือ "เลี้ยวหน้า" "เร็วขวา" "เดินซ้าย" "ถอยลง" "เลี้ยวขึ้น" และ "ซ้าหลัง" จากนั้นทำการทดสอบประสิทธิภาพการปฏิเสธคำที่ไม่ได้กำหนดไว้ จากคำพูดทั้งหมด 60 คำ โดยเปรียบเทียบค่าการตัดสินใจ 3 ค่า คือ 0.01, 0.1 และ 0.5 ซึ่งได้ผลการทดลองดังแสดงในตารางที่ 4.13

ตารางที่ 4.13 ประสิทธิภาพการปฏิเสธคำที่ไม่ได้กำหนดไว้

mse	0.00001			0.0001			0.001		
ค่าการตัดสินใจ	0.01	0.1	0.5	0.01	0.1	0.5	0.01	0.1	0.5
คำที่ปฏิเสธได้	16	22	29	11	18	24	2	11	16
ประสิทธิภาพ	26.6	36.6	48.3	18.3	30	40	3.3	18.3	26.6

4.3 วิเคราะห์ผลการทดลอง

จากผลการทดลองในตารางที่ 4.9 และ 4.10 พบว่าระบบมีประสิทธิภาพการรู้จำโดยเฉลี่ยสูงสุดเมื่อกำหนดให้หยุดการฝึกโครงข่ายประสาทเทียมที่ค่า mse น้อยกว่า 0.001 เนื่องจากการฝึกโครงข่ายที่กำหนดให้ค่า mse ต่ำ (0.00001) จะทำให้โครงข่ายปรับค่าถ่วงน้ำหนักให้เหมาะสมกับรูปแบบที่ใกล้เคียงกับกลุ่มตัวอย่างเท่านั้น ทำให้เสียงพูดคำเดียวกันที่มีรูปแบบแตกต่างจากเสียงพูดในกลุ่มตัวอย่างมีโอกาสที่จะถูกปฏิเสธได้ เมื่อค่าผลลัพธ์สูงสุดน้อยกว่าค่าการตัดสินใจ (0.01) ซึ่งทำให้ประสิทธิภาพการรู้จำลดลง ในขณะที่ค่าถ่วงน้ำหนักที่ได้จากการเรียนรู้อย่างพอเหมาะ แม้จะทำให้ค่าผิดพลาดของผลลัพธ์เพิ่มขึ้น (0.001) และอาจทำให้เกิดการรู้จำผิดพลาด แต่ก็ช่วยให้รูปแบบเสียงพูดที่แตกต่างจากกลุ่มตัวอย่างไม่ถูกปฏิเสธไปมากนัก ดังนั้นประสิทธิภาพการรู้จำโดยเฉลี่ยจึงมีค่าสูง

จากผลการทดลองในตารางที่ 4.11 และ 4.12 พบว่า การเพิ่มค่าการตัดสินใจมีผลทำให้อัตราการรู้จำลดลง เนื่องจากเป็นการกำหนดให้ระบบไม่ยอมรับคำที่มีค่าสูงสุดของโหนดในชั้นเอาท์พุตน้อยกว่าค่าการตัดสินใจ ทำให้เสียงพูดบางคำที่รู้จำได้ถูกต้องอาจถูกระบบปฏิเสธไปด้วย แต่ในขณะเดียวกันก็สามารถลดจำนวนคำที่ระบบรู้จำผิดพลาดลงได้

เมื่อพิจารณาการรู้จำแบบขึ้นกับผู้พูดพบว่า ประสิทธิภาพการรู้จำค่อนข้างสูง คือ มากกว่าร้อยละ 90 ในทุกการทดลอง ทั้งนี้อาจเป็นผลมาจากการที่ผู้พูดแต่ละคนบันทึกเสียงในช่วงเวลาที่ต่อเนื่องกัน ทำให้ข้อมูลที่ใช้ฝึกโครงข่ายและข้อมูลที่ใช้ทดสอบมีความแตกต่างกันไม่มากนัก ทำให้มีจำนวนคำที่รู้จำผิดพลาดน้อย สำหรับค่าการตัดสินใจที่เหมาะกับการรู้จำแบบขึ้นกับผู้พูด คือ 0.01 เนื่องจากสามารถปฏิเสธคำที่ระบบรู้จำผิดพลาดได้ส่วนหนึ่งและไม่ทำให้อัตราการรู้จำลดลงมากนัก

สำหรับการรู้จำแบบไม่ขึ้นกับผู้พูดพบว่า ประสิทธิภาพการรู้จำค่อนข้างดี คือ ร้อยละ 90.9 เมื่อกำหนดให้หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.001 และกำหนดให้ค่าการตัดสินใจเท่ากับ 0.01 แต่จะเห็นว่าจำนวนคำที่รู้จำผิดพลาดค่อนข้างมาก เมื่อเทียบกับประสิทธิภาพการรู้จำเมื่อกำหนดให้หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.00001 ที่ค่าการตัดสินใจค่าเดียวกัน ซึ่งมีอัตราการรู้จำใกล้เคียงกัน คือ ร้อยละ 87.4 แต่จำนวนคำที่รู้จำผิดพลาดน้อยกว่า สาเหตุมาจากค่าถ่วงน้ำหนักที่ได้จากการฝึกโครงข่ายที่กำหนดค่าผิดพลาดไว้ต่ำ (0.00001) มีโอกาสทำให้ระบบได้ผลลัพธ์ที่แม่นยำมากกว่าการฝึกโครงข่ายที่กำหนดค่าผิดพลาดไว้สูง (0.001) นอกจากนั้นผลการปฏิเสธคำที่ไม่ได้กำหนดไว้ ในตารางที่ 4.13 ยังชี้ให้เห็นว่าการฝึกโครงข่ายให้มีค่าผิดพลาดต่ำ สามารถทำให้ระบบปฏิเสธคำผิดได้ดีกว่าการฝึกโครงข่ายที่กำหนดค่าผิดพลาดไว้สูง โดยที่ค่าการตัดสินใจก็มีผลต่ออัตราการปฏิเสธคำผิดด้วยเช่นกัน

ในการสร้างค่าถ่วงน้ำหนักให้กับระบบรู้จำเสียงพูดบนอุปกรณ์ FPGA ได้กำหนดให้หยุดการฝึกโครงข่ายที่ค่า mse น้อยกว่า 0.00001 และกำหนดค่าการตัดสินใจเท่ากับ 0.01 เพราะระบบสามารถปฏิเสธการรู้จำที่ผิดพลาดได้และเหลือจำนวนคำที่รู้จำผิดไม่มากนัก โดยมีอัตราการรู้จำไม่ต่ำเกินไป