บทที่ 4

การทดลองและการวิเคราะห์ผลการทดลอง

4.1 ข้อมูลที่ใช้ในการทดสอบ

ในการทดสอบการรู้จำเสียงพูดนี้จะแบ่งกลุ่มเสียงพูดออกเป็น 2 กลุ่มดังนี้คือ

1 เสียงพูดกลุ่มที่ 1 จะประกอบด้วยเสียงพูดตัวเลขภาษาไทยจำนวน 10 คำคือ 0-9

2 เสียงพูดกลุ่มที่ 2 จะประกอบด้วยเสียงพูดตัวเลขภาษาไทยสองพยางค์ 6 คำคือ 11,12,13, 17,20,30 และคำสามพยางค์ 6 คำคือ 21,23,27,32,78,93 รวมทั้งสิ้น 12 คำ โดยเปล่งเสียงเป็นเสียงต่อเนื่องไม่มีการเว้นช่วงเงียบระหว่างพยางค์

ในการเก็บเสียงพูดเพื่อนำมาใช้ในการทดลองนี้ จะทำการบันทึกจากผู้พูดจำนวน 42 คน ผู้พูดทั้งหมดมีเพศชายอายุอยู่ในช่วง 17-25 ปี โดยแบ่งเป็น 2 กลุ่มดังนี้คือ

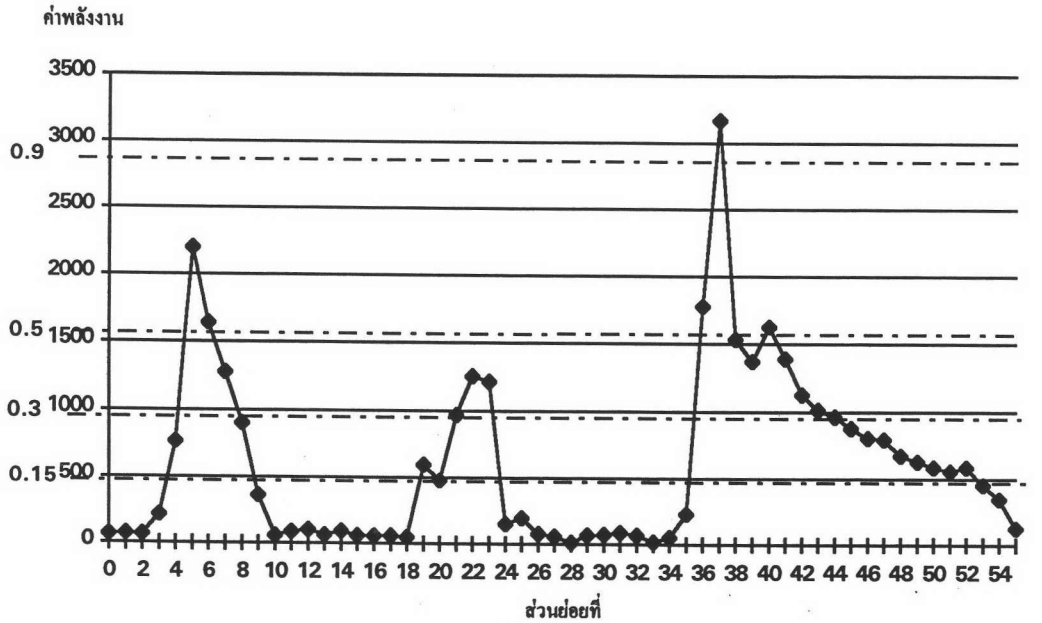
1 กลุ่ม A บันทึกเสียงคำละ 3 ครั้ง จำนวนคน 30 คนโดยแบ่งออกเป็น 2 กลุ่มคือ กลุ่ม A1 ซึ่งเป็นเสียงที่บันทึก 2 ครั้งแรกและกลุ่ม A2 ซึ่งเป็นเสียงที่บันทึกครั้งที่สาม เสียงในกลุ่ม A1 นำไปใช้ในการฝึกนิวรอลเน็ตเวิร์กส่วนเสียงในกลุ่ม A2 ใช้ในการทดสอบเบื้องต้น

2 กลุ่ม B บันทึกเสียงคำละ 3 ครั้ง จำนวนคน 12 คน เสียงพูดกลุ่มนี้ใช้ในการทดสอบเพื่อวัดประสิทธิภาพ ผู้พูดที่อยู่ในกลุ่มนี้ไม่ได้อยู่ในกลุ่มที่ใช้ฝึกนิวรอลเน็ตเวิร์ก ดังนั้นผลการรู้จำของเสียงในกลุ่มนี้จึงถือว่าเป็นผลการรู้จำที่ไม่ขึ้นกับผู้พูด

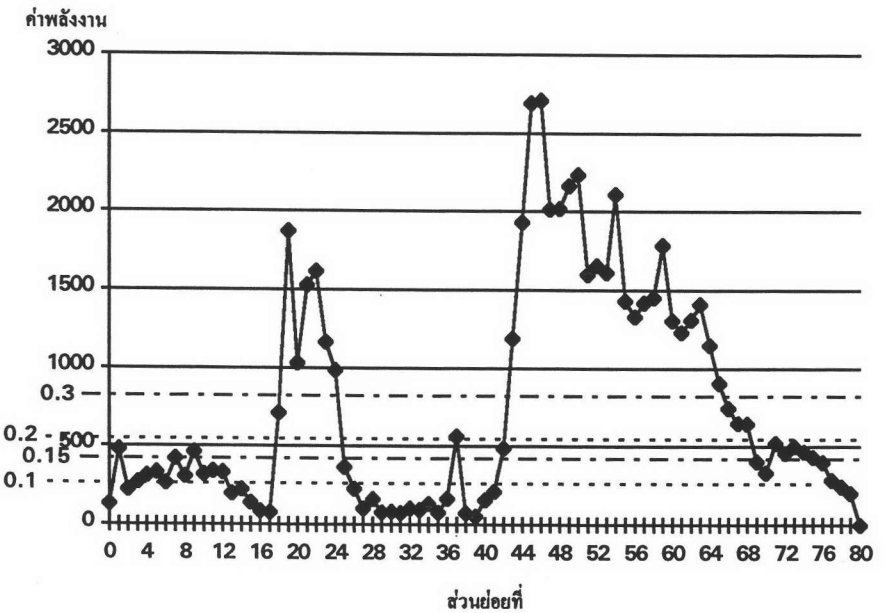
4.2 การวิเคราะห์หาค่าพารามิเตอร์ที่เหมาะสม

ค่าพารามิเตอร์ในการตัดหัวท้ายคำ

ค่าพารามิเตอร์ a, b, m และ n มีค่าเป็น 0.15 เท่า, 0.3 เท่า, 3 ส่วนและ 2 ส่วนตามลำดับ เนื่องจากเสียงพูดของคำ 2 หรือ 3 พยางค์ มีระดับความดังของแต่ละพยางค์ไม่เท่ากัน จึงเลือกใช้ค่าระดับพลังงานทั้งสองคือ a และ b เป็นค่า 0.15 เท่าและ 0.3 เท่า เพราะถ้าใช้ระดับพลังงานมากเกินไปจะทำให้ไม่สามารถตรวจพบพยางค์ที่มีระดับความดังต่ำได้ ตัวอย่างระดับพลังงานของเสียงพูด 3 พยางค์แสดงในรูปที่ 4.1 จากรูปพบว่าถ้าใช้ค่าระดับพลังงานเป็น 0.5 เท่าและ 0.9 เท่าทำให้เกิดการนับพยางค์ผิดพลาดในพยางค์ที่สอง ค่าระดับพลังงานสูง ๆ เช่น 0.5 เท่าและ 0.9 เท่าเหมาะสมเมื่อใช้กับการนับพยางค์ของคำโดดแต่ไม่เหมาะสมใช้กับคำหลายพยางค์ แต่ถ้าใช้ค่าระดับพลังงานต่ำเกินไปเช่น 0.1 เท่าและ 0.2 เท่าอาจทำให้เกิดการนับพยางค์เกินดังตัวอย่างที่แสดงในรูปที่ 4.2



รูปที่ 4.1 ค่าระดับพลังงานของเสียง "เจ็ดสิบแปด"



รูปที่ 4.2 ค่าระดับพลังงานของเสียง "สิบสอง"

ค่าระดับพลังงานอ้างอิงทั้งสองคือค่า a และ b เป็นค่าที่ได้จากการทดลอง โดยเลือกจากการทดสอบเบื้องต้นกับเสียงพูดทั้งคำพยางค์เดียวและหลายพยางค์จำนวน 880 เสียง ในขั้นแรกทดสอบหาค่าระดับพลังงานอ้างอิง b โดยกำหนดให้ค่าระดับพลังงานอ้างอิงทั้งสองมีค่าเท่ากัน

ทำให้เสมือนมีระดับพลังงานอ้างอิงเพียงระดับเดียวและกำหนดให้จำนวนส่วนย่อยของระดับพลังงานทั้งสองมีค่าเท่ากับหนึ่งเพื่อจำกัดขอบเขตการทดลอง จากนั้นทำการปรับค่าระดับพลังงานอ้างอิงเพื่อหาค่าระดับพลังงานที่ทำให้การวิเคราะห์จำนวนพยางค์ผิดพลาดน้อยที่สุด ผลการทดลองดังแสดงในตารางที่ 4.1 พบว่าค่าระดับพลังงานอ้างอิง b ที่เหมาะสมคือ 0.3 เท่า

ตารางที่ 4.1 ผลการทดลองหาค่า b ที่เหมาะสมเมื่อกำหนดให้ $a = b$ และ $m = n = 1$ ส่วน

ค่า b	วิเคราะห์คำ 1 พยางค์ผิด (คำ)	วิเคราะห์คำ 2,3 พยางค์ผิด (คำ)	รวมจำนวนคำที่วิเคราะห์ผิด	ผิดร้อยละ
0.05	77	316	393	44.65
0.10	203	46	249	28.29
0.15	172	3	175	19.88
0.20	116	0	116	13.18
0.25	101	0	101	11.47
0.30	98	0	98	11.13
0.35	104	3	107	12.15
0.40	106	5	111	12.61
0.45	123	8	131	14.88
0.50	145	11	156	17.72
0.60	208	28	236	26.81
0.70	233	69	302	34.31
0.80	250	125	375	42.61
0.90	241	200	441	50.11

จากนั้นทดสอบหาค่าระดับพลังงานอ้างอิง a โดยให้ค่าระดับพลังงานอ้างอิง b มีค่าเท่ากับ 0.3 เท่าซึ่งเป็นค่าที่เหมาะสมจากผลการทดลองในตารางที่ 4.1 และกำหนดให้จำนวนส่วนย่อยของระดับพลังงานทั้งสองมีค่าเท่ากับหนึ่งเพื่อจำกัดขอบเขตการทดลอง ผลการทดลองดังแสดงในตารางที่ 4.2 พบว่าค่าระดับพลังงานอ้างอิง a ที่เหมาะสมคือ 0.15 เท่า และจากผลการทดลองแสดงให้เห็นว่าการใช้ค่าระดับพลังงานอ้างอิงสองระดับช่วยปรับปรุงการนับจำนวนพยางค์และการตัดคำให้ดีขึ้นจากการใช้ระดับพลังงานอ้างอิงเพียงระดับเดียว

ตารางที่ 4.2 ผลการทดลองหาค่า a ที่เหมาะสมเมื่อกำหนดให้ $b = 0.3$ และ $m = n = 1$ ส่วน

ค่า a	วิเคราะห์คำ 1 พยางค์ผิด (คำ)	วิเคราะห์คำ 2,3 พยางค์ผิด (คำ)	รวมจำนวนคำที่วิเคราะห์ผิด	ผิดร้อยละ
0.00	0	480	480	54.54
0.05	1	342	343	38.97
0.10	7	69	76	8.63
0.15	16	6	22	2.50
0.20	26	0	26	2.95
0.25	46	0	46	5.22
0.30	98	0	98	11.13

จำนวนส่วนย่อยทั้งสองคือค่า m และ n เป็นค่าที่ได้จากการทดลอง โดยเลือกจากการทดสอบเบื้องต้นกับเสียงพูดทั้งคำพยางค์เดียวและหลายพยางค์จำนวน 880 เสียง ในขั้นแรกทดสอบหาจำนวนส่วนย่อยของระดับพลังงานต่ำโดยกำหนดให้จำนวนส่วนย่อยของระดับพลังงานสูงเท่ากับศูนย์เพื่อจำกัดขอบเขตการทดลอง ผลการทดลองดังแสดงในตารางที่ 4.3 พบว่าจำนวนส่วนย่อยของระดับพลังงานต่ำคือค่า m ที่เหมาะสมคือ 3 ส่วนย่อย

ตารางที่ 4.3 ผลการทดลองหาค่า m ที่เหมาะสมเมื่อกำหนดให้ $n = 0$

ค่า m	วิเคราะห์จำนวนพยางค์ผิด (คำ)	ผิดร้อยละ
1	22	2.50
2	20	2.27
3	16	1.81
4	28	3.18
5	60	6.81

จากนั้นทดสอบหาจำนวนส่วนย่อยของระดับพลังงานสูงโดยกำหนดให้จำนวนส่วนย่อยของระดับพลังงานต่ำเท่ากับ 3 ผลการทดลองดังแสดงในตารางที่ 4.4 พบว่าจำนวนส่วนย่อยของระดับพลังงานสูงคือค่า n ที่เหมาะสมคือ 2 ส่วนย่อย

ตารางที่ 4.4 ผลการทดลองหาค่า n ที่เหมาะสมเมื่อกำหนดให้ $m = 3$

ค่า n	วิเคราะห์จำนวนพยางค์ผิด (คำ)	ผิดร้อยละ
1	16	1.81
2	13	1.47
3	28	3.18
4	81	9.20

ค่าพารามิเตอร์ทั้งสี่ค่าที่ได้จากการทดลองสามารถแยกคำพยางค์เดียวออกจากคำหลายพยางค์ได้อย่างถูกต้อง เมื่อทดสอบกับเสียงพูดทั้งหมดที่ใช้ในการทดลอง พบว่าสามารถแบ่งแยกคำพยางค์เดียวได้ถูกต้อง 98.977% ดังแสดงรายละเอียดในตารางที่ 4.5

ตารางที่ 4.5 ความถูกต้องของโปรแกรมนับจำนวนพยางค์

	วิเคราะห์เป็นคำพยางค์เดียว	วิเคราะห์เป็นคำหลายพยางค์	แบ่งคำถูกต้อง
คำพยางค์เดียว	1742	18	98.977%
คำหลายพยางค์	42	1386	97.058%
รวมทั้งหมด	-	-	98.118%

ส่วนของคำที่ได้จากการวิเคราะห์ด้วยค่าระดับพลังงานคือระยะจากจุด a_1 ถึง b_1 จากนั้นเลื่อนส่วนหัวและส่วนท้ายคำเพื่อเก็บรายละเอียดของเสียงพูด โดยเลื่อนเป็นสัดส่วน 6/26 และ 1/26 ของจำนวนจุดสัญญาณเสียงพูดจากพยางค์แรกถึงพยางค์สุดท้าย เหตุผลที่เลื่อนส่วนหัวเป็นระยะมากกว่าระยะเลื่อนส่วนท้ายคำ คือ ส่วนหัวของเสียงพูดมักจะมีสัญญาณที่มีแอมพลิจูดต่ำเป็นเวลานานกว่าส่วนท้ายคำ ซึ่งเสียงของตัวเลขที่มีสัญญาณที่มีแอมพลิจูดต่ำในส่วนหัวนานคือเสียงที่มีเสียง “ส” ในส่วนต้นคำ คือเสียง“0”,“2”,“3”,“4” ซึ่งเสียงเหล่านี้มีสัญญาณในส่วนต้นคำเป็นเวลาประมาณ 1/4 ของคำ (ดังรูปคลื่นในภาคผนวก ก) จึงเลื่อนส่วนหัวเป็นสัดส่วน 6/26 ซึ่งมีค่าใกล้เคียง 1/4 การใช้ค่า 6/26 ไม่ใช้ค่า 1/4 เป็นเพราะต้องคำนวณค่า 1/26 ในการเลื่อนส่วนท้ายคำอยู่แล้วจึงนำค่าที่ได้ในตัวแปรมาใช้คำนวณต่อ สำหรับส่วนท้ายมีสัญญาณที่มีแอมพลิจูดต่ำเป็นเวลาสั้น ๆ จึงเลื่อนส่วนท้ายเป็นสัดส่วนเพียง 1/26 ค่า 1/26 เป็นระยะมากที่สุดที่ต้องการในการเลื่อนส่วนท้ายคำให้ครอบคลุมรายละเอียดของเสียง จากการทดลองกับตัวอย่างเสียงพูดจำนวน 100 เสียง

(รายละเอียดแสดงในภาคผนวก ข) เสียงพูดในระยะจากจุด a_2 ถึง b_2 ถูกใช้เป็นสัญญาณเสียงพูดที่จะทำการวิเคราะห์ต่อไป

ค่าพารามิเตอร์ในการประมาณพันธะเชิงเส้น

เวลาในแต่ละส่วนย่อยมีค่าเท่ากับ 20 ms (เสาวลักษณ์ อารีย์พงศา,2538) เพราะฉะนั้นพารามิเตอร์ N คือจำนวนจุดข้อมูลใน 1 ส่วนย่อยมีค่าเท่ากับ 20ms x 8kHz = 160 จุดข้อมูล

การเหลื่อมของส่วนย่อยมีค่าเท่ากับ 1/4 ส่วนย่อย (เสาวลักษณ์ อารีย์พงศา,2538) ดังนั้นค่า M คือจำนวนจุดข้อมูลใน 1 ส่วนย่อยที่ไม่เหลื่อมกับข้อมูลในส่วนย่อยอื่น มีค่าเท่ากับ 120 จุดข้อมูล

จากค่าทั้งสองจะได้ค่า L คือจำนวนส่วนย่อยในสัญญาณเสียงพูด 1 คำ มีค่าเท่ากับ 33 ส่วนย่อย

ค่า p หรืออันดับของค่าสัมประสิทธิ์ LPC มีค่าเท่ากับ 10 เพราะค่าอันดับ 10 เป็นค่าที่เหมาะสมสำหรับการวิเคราะห์สัญญาณเสียงที่สุ่มสัญญาณด้วยความถี่ 8 kHz (O'Shaughnessy, 1988) และได้มีการทดสอบว่าค่าอันดับ 10 เพียงพอสำหรับใช้ในการรู้จำเสียงโดย เสาวลักษณ์ อารีย์พงศา (2538)

ค่าพารามิเตอร์สำหรับนิวรอลเน็ตเวิร์ก

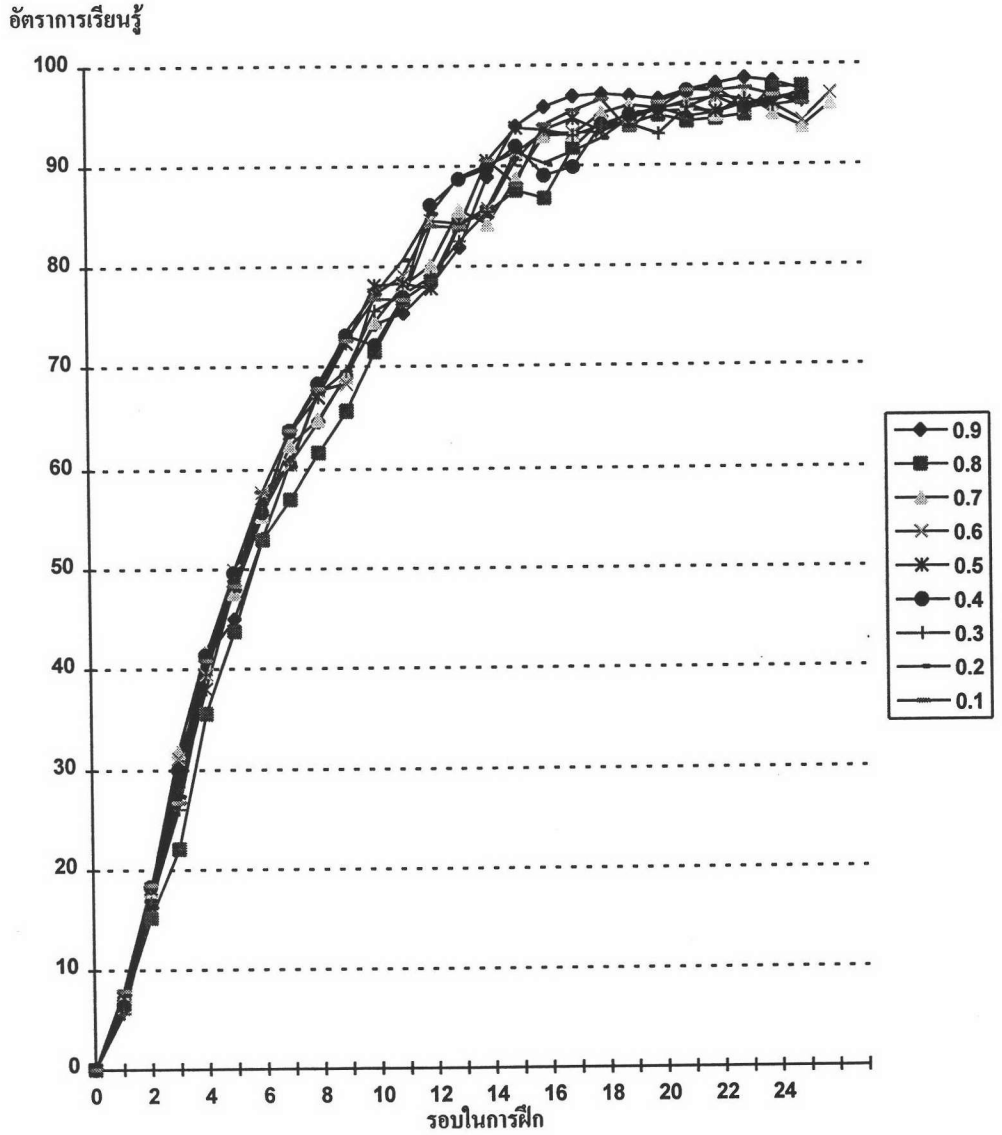
เนื่องจากค่า learning rate และ momentum เป็นค่าที่ขึ้นกับคุณลักษณะของพื้นผิวความผิดพลาด จึงต้องทำการทดลองเพื่อหาค่าที่เหมาะสม และเนื่องจากค่าทั้งสองมีผลกับเวลาที่ใช้ในการฝึกนิวรอลเน็ตเวิร์กจึงทำการทดลองเพื่อหาค่าที่ทำให้ใช้เวลาในการฝึกน้อยที่สุด จากสมการที่ 3.10 พบว่าค่า momentum กำหนดสัดส่วนของค่า learning rate ในการปรับค่าน้ำหนักการเชื่อมต่อด้วย จึงทำการทดลองโดยกำหนดให้ค่า learning rate มีค่าคงที่เท่ากับ 1 เพื่อจำกัดขอบเขตการทดลอง อีกเหตุผลหนึ่งที่เลือกใช้ค่านี้คือถ้า momentum มีค่าเท่ากับ 0 ค่า learning rate เท่ากับ 1 จะทำให้สัดส่วนการปรับค่าน้ำหนักการเชื่อมต่อมีค่ามากและใช้เวลาในการฝึกน้อย

การทดลองเพื่อหาค่า momentum จะปรับค่า momentum หลาย ๆ ค่า โดยกำหนดให้ค่า learning rate มีค่าคงที่เท่ากับ 1 เพื่อหาค่า momentum ที่ทำให้ใช้เวลาในการฝึกน้อยที่สุด จากผลการทดลองที่ได้ในตารางที่ 4.6 พบว่าการปรับค่า momentum มีผลกับเวลาที่ใช้ในการฝึกน้อยมาก จึงนำอัตราการเรียนรู้ในแต่ละรอบของการฝึกมาช่วยพิจารณาดังแสดงในรูปที่ 4.3

I17380546

ตารางที่ 4.6 ค่า momentum และเวลาในการฝึกที่ค่า learning rate = 1

ค่า momentum	เวลาที่ใช้ในการฝึก(นาที)
0.9	22
0.8	23
0.7	23
0.6	22
0.5	22
0.4	22
0.3	23
0.2	22
0.1	22



รูปที่ 4.3 อัตราการเรียนรู้ในแต่ละรอบของการฝึก

จากรูปที่ 4.3 พบว่าค่า momentum ที่เหมาะสมคือ 0.9 เพราะให้อัตราการเรียนรู้ที่สูงกว่าค่าอื่น และในช่วงท้ายของการฝึกให้อัตราการเรียนรู้ที่ค่อนข้างสม่ำเสมอ

จำนวนโหนดในระดับข้อมูลเข้ากำหนดโดยขนาดของข้อมูลอินพุตที่ใช้แทนเสียงพูด 1 คำ มีจำนวนโหนด 330 โหนดซึ่งตรงกับจำนวนส่วนย่อยในเสียงพูด 1 คำ คือ 33 ส่วนย่อย และ 1 ส่วนย่อยประกอบด้วยสัมประสิทธิ์ LPC 10 ค่า

จำนวนโหนดในระดับข้อมูลออกกำหนดโดยจำนวนกลุ่มข้อมูลที่ต้องการแบ่งแยก มีจำนวนโหนด 10 โหนดสำหรับนิวรอลเน็ตเวิร์กชุดแรก และมีจำนวนโหนด 12 โหนดสำหรับนิวรอลเน็ตเวิร์กชุดที่สอง

การทดลองเพื่อหาจำนวนโหนดในระดับซ่อนตัว เนื่องจากจำนวนโหนดในระดับซ่อนตัวกำหนดความสามารถในการเรียนรู้ของนิวรอลเน็ตเวิร์ก จึงออกแบบการทดลองโดยฝึกนิวรอลเน็ตเวิร์กที่มีจำนวนโหนดในระดับซ่อนตัวหลาย ๆ ค่า และดูผลอัตราความถูกต้องในการรู้จำเสียงของเสียงพูดในกลุ่ม B โดยที่กำหนดให้ค่า learning rate และ momentum มีค่าคงที่เท่ากับ 1 และ 0.9 ตามลำดับ เพื่อหาจำนวนโหนดที่ไม่มากเกินไปและมีความสามารถในการรู้จำเสียงได้ จากผลการทดลองที่ได้ในตารางที่ 4.7 พบว่าอัตราความถูกต้องในการรู้จำเสียงมีแนวโน้มที่สูงขึ้นเมื่อเพิ่มจำนวนโหนดในระดับซ่อนตัวขึ้น และจำนวนโหนดที่เหมาะสมคือ 50 โหนดเพราะใช้หน่วยความจำไม่มากเกินไป และให้ผลอัตราความถูกต้องในการรู้จำเสียงค่อนข้างสูง

ตารางที่ 4.7 ความสัมพันธ์ของจำนวนโหนดในระดับซ่อนตัวและอัตราการรู้จำ

จำนวนโหนดในระดับซ่อนตัว	อัตราการรู้จำเสียงโดยไม่ขึ้นกับผู้พูด(%)
30	87.5
40	85.6
50	88.9
60	90.6
70	90.6
80	91.7
90	91.7

4.3 ผลการทดลอง

4.3.1 การรู้จำเสียงหนึ่งพยางค์

นำค่าพารามิเตอร์ที่เหมาะสมจากผลการทดลองในหัวข้อ 4.2 มาใช้ในการทดลองเพื่อวัดประสิทธิภาพของนิวรอลเน็ตเวิร์กในการรู้จำเสียงพูดกลุ่มที่ 1 โดยใช้ข้อมูลในกลุ่ม A1 ฝึกนิวรอลเน็ตเวิร์ก 6 ครั้ง โดยใช้ผู้พูดจำนวน 5,10,15,20,25 และ 30 คน จากนั้นนำนิวรอลเน็ตเวิร์กที่ได้รับการฝึกแล้วมาทดสอบเพื่อวัดอัตราการรู้จำโดยแบ่งการทดสอบเป็น 3 กลุ่ม คือ

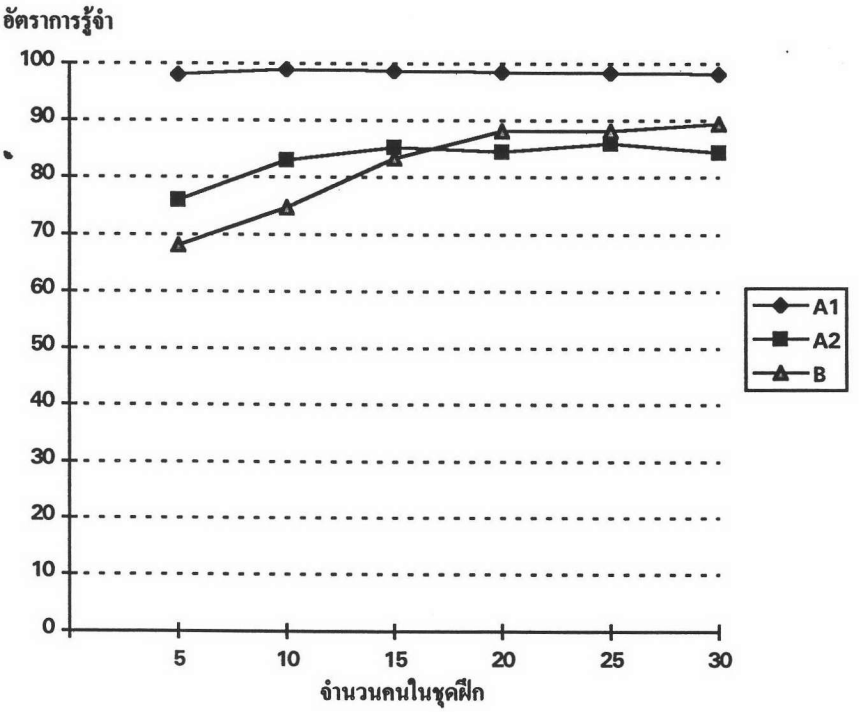
1 ทดสอบกับกลุ่ม A1 ตามจำนวนผู้พูดที่ใช้ในการฝึกนิวรอลเน็ตเวิร์ก

2 ทดสอบกับกลุ่ม A2 ตามจำนวนผู้พูดในกลุ่ม A1 ที่ใช้ในการฝึกนิวรอลเน็ตเวิร์ก

3 ทดสอบกับกลุ่ม B เพื่อวัดอัตราการรู้จำโดยไม่ขึ้นกับผู้พูด

ตารางที่ 4.8 อัตราการรู้จำของเสียงพูดกลุ่มที่ 1 โดยทดสอบตามจำนวนผู้พูดที่ใช้ในการฝึก

กลุ่มทดสอบ	ความถูกต้อง(%)					
	จำนวนผู้พูดที่ใช้ในการฝึกนิวรอลเน็ตเวิร์ก					
	5 คน	10 คน	15 คน	20 คน	25 คน	30 คน
A1	98	99	98.7	98.5	98.4	98.2
A2	76	83	85.3	84.5	86	84.3
B	68.1	74.7	83.3	88.1	88.1	89.4



รูปที่ 4.4 อัตราการรู้จำเทียบกับจำนวนคนในชุดฝึก

จากผลการทดลองในตารางที่ 4.8 แสดงให้เห็นว่านิวรอลเน็ตเวิร์กมีความสามารถในการเรียนรู้สูงมากเพราะให้อัตราการรู้จำที่สูงเมื่อทดสอบกับเสียงในกลุ่ม A1 และอัตราการรู้จำค่อนข้างคงที่ถึงแม้ว่าจะเพิ่มความซับซ้อนของข้อมูลโดยการเพิ่มจำนวนคนในชุดฝึก ส่วนอัตราการรู้จำเมื่อทดสอบกับเสียงในกลุ่ม A2 และ B ซึ่งไม่ใช่กลุ่มของเสียงที่ใช้ในการฝึกจะมีค่าต่ำเมื่อจำนวนคนในชุดฝึกน้อย และจะมีค่าสูงขึ้นเมื่อเพิ่มจำนวนคนในชุดฝึก เพราะการเพิ่มจำนวนคนในชุดฝึกจะ

ทำให้นิวรอลเน็ตเวิร์กมีความรู้เกี่ยวกับลักษณะของเสียงแต่ละคำมากขึ้น เมื่อเทียบอัตราการรู้จำระหว่างกลุ่ม A2 และ B จะเห็นว่าในช่วงแรกอัตราการรู้จำในกลุ่ม A2 ซึ่งเป็นเสียงของผู้พูดคนเดิมในชุดฝึกจะมีค่าสูงกว่าอัตราการรู้จำในกลุ่ม B เพราะเสียงพูดของคนเดียวกันถึงแม้จะไม่เหมือนกันแต่ก็มีความคล้ายคลึงกันมากกว่าเสียงพูดของคนอื่น แต่เมื่อเพิ่มจำนวนคนในชุดฝึกขึ้นมาก ๆ อัตราการรู้จำในกลุ่ม A2 กลับมีค่าต่ำกว่าอัตราการรู้จำในกลุ่ม B สิ่งนี้เป็นผลที่ขัดแย้งกับผลที่ควรจะเป็นในทางทฤษฎี ทั้งนี้อาจเป็นผลมาจากผู้ทดลองพยายามเลือกเสียงพูดที่มีสัญญาณรบกวนน้อยและมีรูปคลื่นสม่ำเสมอเพื่อใช้เป็นเสียงที่ใช้ในการฝึก เสียงที่เหลือของผู้พูดในชุดฝึกจะถูกใช้ในกลุ่มทดสอบ A2 ทำให้เสียงในกลุ่ม A2 เป็นเสียงที่มีคุณภาพไม่ดีนัก ในทางกลับกันเสียงที่ใช้ในกลุ่มทดสอบ B ประกอบด้วยเสียงจากการบันทึกทั้ง 3 ครั้งของบุคคลอีกกลุ่มหนึ่งโดยไม่มีการเลือกเสียงดังนั้นเสียงที่อยู่ในกลุ่ม B จะมีคุณภาพโดยเฉลี่ยดีกว่าเสียงในกลุ่ม A2

เมื่อนำผลการทดลองในตารางที่ 4.8 มาดูแนวโน้มดังแสดงในรูปที่ 4.4 พบว่าเมื่อเพิ่มจำนวนคนที่ใช้ในการฝึกจะทำให้อัตราการรู้จำโดยไม่ขึ้นกับผู้พูดเพิ่มขึ้น โดยมีแนวโน้มของการเพิ่มในอัตราที่ลดลง แสดงว่าเมื่อเพิ่มคนในชุดฝึกไปเรื่อย ๆ อัตราการรู้จำน่าจะเพิ่มขึ้นถึงค่าคงที่ค่าหนึ่ง

ตารางที่ 4.9-4.11 แสดงผลการรู้จำเสียงพูดในกลุ่มทดสอบ A1,A2 และ B แบ่งตามเสียงศูนย์ถึงเก้าเมื่อฝึกนิวรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน จากตารางทั้งสามพบว่าเสียงของตัวเลขที่ทำให้การรู้จำผิดพลาดมากคือเสียง “0”,“2”,“3” เสียงทั้งสามคำนี้ต่างก็มีเสียง “ส” ในส่วนต้นคำซึ่งจัดอยู่ในประเภทเสียงเสียดแทรก (fricative sound) คำทั้งสามมักจะผิดพลาดเป็นเสียงที่ขึ้นต้นด้วย “ส” ทั้งสิ้นและไม่ค่อยผิดเป็นคำที่ไม่ขึ้นต้นด้วย “ส” คือ “1”,“5”,“6”,“7”,“8”,“9” แต่ก็มีสิ่งที่น่าสังเกตคือเสียง “4” ก็ขึ้นต้นด้วย “ส” แต่ไม่ค่อยผิดกับคำทั้งสามทั้งนี้น่าจะเกิดจากเสียง “4” มีรูปคลื่นที่แตกต่างจากคำทั้งสาม และการที่เสียง “2” มีความผิดพลาดสูงเป็นเพราะเสียง “2” มีรูปคลื่นที่ก้ำกึ่งระหว่างเสียง “0” และ “3” ดังแสดงในภาคผนวก ก

ตารางที่ 4.9 ผลการรู้จำเสียงพูดกลุ่ม A1 จำนวน 30 คนพูดคำละ 2 ครั้ง เมื่อฝึกนิวรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน

คำในกลุ่มทดสอบ(A1)	คำที่วิเคราะห์ได้									
	0	1	2	3	4	5	6	7	8	9
0	59	1	0	0	0	0	0	0	0	0
1	0	60	0	0	0	0	0	0	0	0
2	1	0	56	2	0	0	0	1	0	0
3	0	0	0	59	0	0	0	0	1	0
4	0	1	0	0	58	0	1	0	0	0
5	1	0	0	0	0	58	0	0	1	0
6	0	0	0	0	0	0	60	0	0	0
7	0	0	0	0	0	0	0	60	0	0
8	0	0	0	0	0	0	0	0	60	0
9	0	0	0	1	0	0	0	0	0	59

ตารางที่ 4.10 ผลการรู้จำเสียงพูดกลุ่ม A2 จำนวน 30 คนพูดคำละ 1 ครั้ง เมื่อฝึกนิวรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน

คำในกลุ่มทดสอบ(A2)	คำที่วิเคราะห์ได้									
	0	1	2	3	4	5	6	7	8	9
0	25	1	3	0	0	1	0	0	0	0
1	0	27	0	0	2	0	1	0	0	0
2	5	0	17	4	0	1	1	0	0	2
3	1	1	2	24	0	1	1	0	0	0
4	0	1	0	0	26	0	2	1	0	0
5	0	0	1	1	1	25	0	0	1	1
6	0	0	0	0	0	0	30	0	0	0
7	0	0	0	0	1	1	2	26	0	0
8	0	0	0	1	0	1	2	0	24	2
9	0	0	0	0	0	1	0	0	0	29

ตารางที่ 4.11 ผลการรู้จำเสียงพูดกลุ่ม B จำนวน 12 คนพูดคำละ 3 ครั้ง เมื่อฝึกนิวรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน

คำในกลุ่มทดสอบ(B)	คำที่วิเคราะห์ได้									
	0	1	2	3	4	5	6	7	8	9
0	32	1	1	1	0	0	1	0	0	0
1	0	36	0	0	0	0	0	0	0	0
2	5	0	27	3	0	0	0	0	0	1
3	2	0	1	33	0	0	0	0	0	0
4	0	0	0	0	36	0	0	0	0	0
5	1	2	0	5	0	28	0	0	0	0
6	0	0	0	0	1	0	34	0	0	1
7	0	0	0	0	0	0	1	34	1	0
8	0	0	0	2	0	2	0	0	31	1
9	0	0	2	1	0	2	0	0	0	31

ตารางที่ 4.12 แสดงผลการรู้จำเสียงพูดในกลุ่มทดสอบ A1 และ A2 จำแนกตามรายบุคคล เมื่อฝึกนิวรอลเน็ตเวิร์กโดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน และตารางที่ 4.13 แสดงผลการรู้จำเสียงพูดในกลุ่มทดสอบ B จากตารางทั้งสองพบว่าผลการรู้จำเสียงพูดของแต่ละบุคคลในกลุ่ม B โดยเฉลี่ยมีค่าสูงกว่าผลในกลุ่ม A2 เพราะเสียงที่อยู่ในกลุ่ม B จะมีคุณภาพโดยเฉลี่ยดีกว่าเสียงในกลุ่ม A2 ซึ่งเกิดจากการเลือกเสียงของผู้ทดลองตามที่ได้กล่าวไปแล้ว

ตารางที่ 4.12 ผลการรู้จำเสียงพูดของแต่ละบุคคลในกลุ่ม A1,A2

ลำดับคนในกลุ่มฝึก	กลุ่ม A1 วิเคราะห์ถูก(คำ) จากจำนวน 20 คำ	กลุ่ม A2 วิเคราะห์ถูก(คำ) จากจำนวน 10 คำ
1	20	8
2	20	9
3	20	9
4	20	6
5	19	10
6	20	9
7	20	7
8	20	9
9	20	9
10	19	9
11	20	9
12	18	7
13	20	9
14	20	10
15	20	10
16	19	7
17	20	10
18	20	7
19	19	8
20	19	10
21	19	7
22	20	10
23	20	10
24	20	8
25	20	6
26	20	8
27	19	8
28	20	8
29	20	7
30	18	9

ตารางที่ 4.13 ผลการรู้จำเสียงพูดของแต่ละบุคคลในกลุ่ม B โดยพูด 10 คำ คำละ 3 ครั้ง

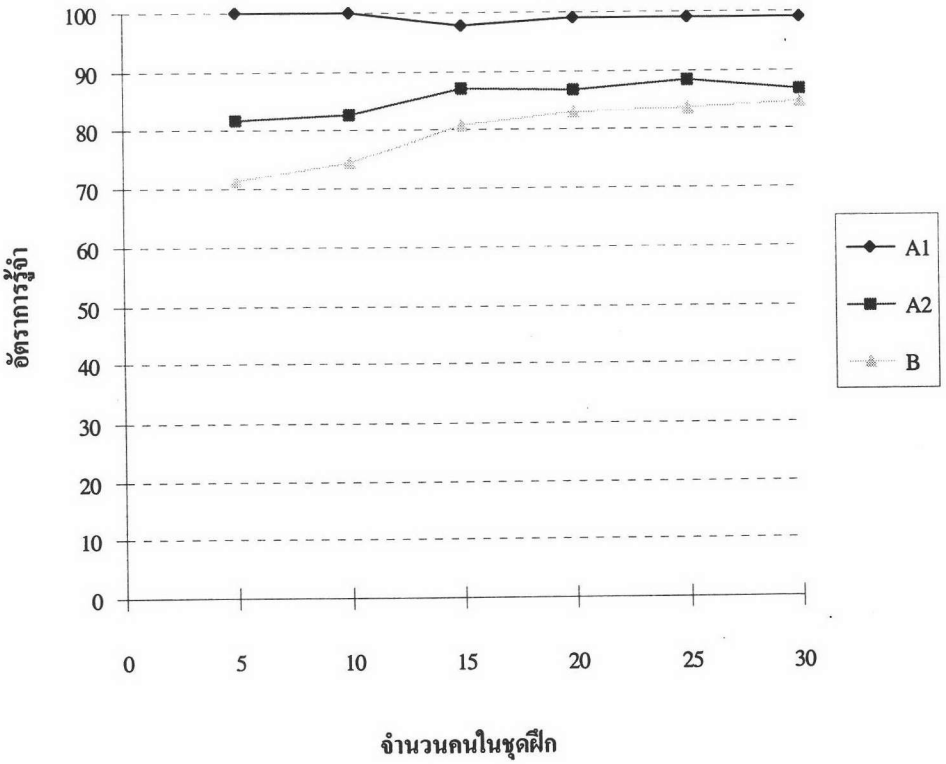
ลำดับคนในกลุ่มทดสอบ	กลุ่ม B วิเคราะห์ถูก(คำ) จากจำนวน 30 คำ
1	26
2	27
3	24
4	27
5	26
6	27
7	26
8	30
9	26
10	28
11	29
12	26

4.3.2 การรู้จำเสียงสองและสามพยางค์

การทดลองเพื่อวัดประสิทธิภาพของนิวรอลเน็ตเวิร์กในการรู้จำเสียงพูดกลุ่มที่ 2 ซึ่งเป็นเสียงของคำสองและสามพยางค์ โดยทำการฝึกนิวรอลเน็ตเวิร์กด้วยข้อมูลที่บันทึกจากคนในกลุ่มเดียวกับกลุ่มคนในการทดสอบกับคำพยางค์เดียว และใช้ค่าพารามิเตอร์เดิมที่ใช้ในการทดสอบกับคำพยางค์เดียวเพื่อจำกัดขอบเขตของการทดลองให้อยู่ในสภาพใกล้เคียงกัน ผลการทดลองที่ได้แสดงในตารางที่ 4.14 และแสดงแนวโน้มในรูปที่ 4.5

ตารางที่ 4.14 อัตราการรู้จำของเสียงพูดกลุ่มที่ 2 โดยทดสอบตามจำนวนผู้พูดที่ใช้ในการฝึก

กลุ่มทดสอบ	ความถูกต้อง(%)					
	จำนวนผู้พูดที่ใช้ในการฝึกนิวรอลเน็ตเวิร์ก					
	5 คน	10 คน	15 คน	20 คน	25 คน	30 คน
A1	100	100	97.8	99	99.2	99.2
A2	81.7	82.5	87.2	86.7	88.3	86.7
B	71.3	74.5	80.8	82.9	83.6	84.7



รูปที่ 4.5 อัตราการรู้จำของเสียงพูดกลุ่มที่ 2 เทียบกับจำนวนคนในชุดฝึก

เมื่อนำผลการทดลองในตารางที่ 4.14 พบว่ามีแนวโน้มเช่นเดียวกับตารางที่ 4.8 คือ เมื่อเพิ่มจำนวนคนที่ใช้ในการฝึกจะทำให้อัตราการรู้จำเพิ่มขึ้น โดยมีแนวโน้มของการเพิ่มในอัตราที่ลดลง และอัตราการรู้จำของกลุ่ม A1 อยู่ในระดับที่สูงและค่อนข้างคงที่แสดงถึงประสิทธิภาพในการเรียนรู้ของนิวรอลเน็ตเวิร์ก

ตารางที่ 4.15-4.17 แสดงผลการรู้จำเสียงพูดในกลุ่มทดสอบ A1,A2 และ B แบ่งตามเสียงในกลุ่มที่ 2 ซึ่งเป็นเสียงของคำสองและสามพยางค์ เมื่อฝึกนิวรอลเน็ตเวิร์กโดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน

ตารางที่ 4.15 ผลการรู้จำเสียงพูดกลุ่ม A1 จำนวน 30 คนพูดคำละ 2 ครั้ง เมื่อฝึกนิวรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน

คำในกลุ่ม	คำที่วิเคราะห์ได้											
ทดสอบ(A1)	11	12	13	17	20	30	21	23	27	32	78	93
11	58	0	0	1	0	1	0	0	0	0	0	0
12	0	60	0	0	0	0	0	0	0	0	0	0
13	0	0	60	0	0	0	0	0	0	0	0	0
17	0	0	0	60	0	0	0	0	0	0	0	0
20	0	0	1	0	58	0	0	0	0	0	0	1
30	0	0	0	0	0	60	0	0	0	0	0	0
21	0	0	0	0	0	0	59	0	0	0	0	1
23	0	0	0	0	0	0	0	60	0	0	0	0
27	0	0	0	0	0	0	0	0	60	0	0	0
32	0	0	0	0	0	0	0	0	0	60	0	0
78	0	0	0	0	0	0	0	0	0	0	59	1
93	0	0	0	0	0	0	0	0	0	0	0	60

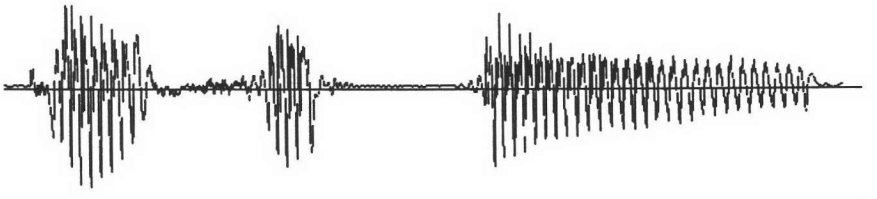
ตารางที่ 4.16 ผลการรู้จำเสียงพูดกลุ่ม A2 จำนวน 30 คนพูดคำละ 1 ครั้ง

คำในกลุ่ม	คำที่วิเคราะห์ได้											
ทดสอบ(A2)	11	12	13	17	20	30	21	23	27	32	78	93
11	26	0	0	3	0	1	0	0	0	0	0	0
12	0	26	1	0	0	0	0	0	0	3	0	0
13	1	4	23	0	0	0	0	0	0	1	0	1
17	5	0	0	24	0	0	0	0	1	0	0	0
20	0	0	0	0	27	0	1	1	0	0	0	1
30	0	0	1	1	0	28	0	0	0	0	0	0
21	0	0	2	0	4	0	21	1	2	0	0	0
23	0	0	0	0	1	0	0	27	0	0	1	1
27	0	0	0	0	0	0	3	1	25	0	0	1
32	0	0	0	0	0	0	0	0	0	30	0	0
78	0	0	0	0	0	0	1	0	0	0	29	0
93	0	0	1	0	0	1	0	1	0	1	0	26

ตารางที่ 4.17 ผลการรู้จำเสียงพูดกลุ่ม B จำนวน 12 คนพูดคำละ 3 ครั้ง

คำในกลุ่ม	คำที่วิเคราะห์ได้											
ทดสอบ B	11	12	13	17	20	30	21	23	27	32	78	93
11	28	0	0	5	0	0	0	0	0	0	1	2
12	0	30	4	0	0	1	0	0	0	0	0	1
13	0	0	30	1	0	1	0	2	0	1	0	1
17	6	0	0	26	1	1	1	0	0	0	0	1
20	0	0	0	0	36	0	0	0	0	0	0	0
30	0	0	0	0	0	36	0	0	0	0	0	0
21	1	0	1	0	2	0	19	0	12	0	0	1
23	0	0	1	0	0	0	0	31	3	0	0	1
27	0	0	0	0	0	0	4	0	32	0	0	0
32	0	0	0	0	0	0	0	0	0	35	0	1
78	0	0	0	0	0	0	3	2	0	1	30	0
93	0	0	1	1	0	0	0	1	0	0	0	33

จากผลการรู้จำในตารางที่ 4.15-4.17 พบว่าเสียงของตัวเลขที่ผิดพลาดมาก มักจะเป็นเสียงที่ใกล้เคียงกันเช่นเสียง “11” กับ “17”,เสียง “12” กับ “13”,เสียง “21” กับ “27” แต่มีคำที่เสียงต่างกันแต่ผิดมากเช่น เสียง “78” กับ “21” และ “23” ซึ่งเกิดจากเสียง “เจ็ด” คล้ายกับเสียง “ยี่” โดยเฉพาะอย่างยิ่งถ้าแอมพลิจูดของเสียงต่ำจะยิ่งคล้ายกันมากเพราะจะไม่ค่อยมีรายละเอียดของเสียง ตัวอย่างรูปคลื่นของเสียง “78” แสดงในรูปที่ 4.6 และรูปคลื่นของเสียง “23” แสดงในรูปที่ 4.7 แอมพลิจูดของเสียง “23” ในรูปที่ 4.7 ต่ำจึงทำให้ตัดส่วนท้ายทิ้งมากเกินไปจนคล้ายกับเสียง “แปด”



รูปที่ 4.6 เสียงเจ็ดสิบแปดที่ผิดเป็นเสียงยี่สิบเอ็ด

รูปที่ 4.7 เสียงยี่สิบสามที่ผิดเป็นเสียงเจ็ดสิบแปด

บทที่ 5
สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

1. โปรแกรมตัดหัวท้ายคำและนับจำนวนพยางค์สามารถทำงานได้อย่างถูกต้อง ถึงแม้ว่าแอมพลิจูดของสัญญาณเสียงไม่เท่ากัน แต่อาจตัดหัวท้ายคำผิดพลาดเมื่อแอมพลิจูดมีขนาดเล็กมาก สำหรับการนับจำนวนพยางค์ใช้เพื่อแยกนิวรอลเน็ตเวิร์กที่ใช้ในการรู้จำ สามารถแบ่งคำพยางค์เดียวออกจากคำหลายพยางค์ได้ถูกต้อง 98.1 %

2. นิวรอลเน็ตเวิร์กมีความสามารถในการเรียนรู้สูงมากโดยสังเกตได้จากอัตราการรู้จำของข้อมูลในกลุ่มที่ใช้ฝึกมีค่าสูงมากและค่อนข้างคงที่ ถึงแม้ว่าจะเพิ่มจำนวนคนในชุดฝึกมากขึ้น

3. อัตราการรู้จำแบบไม่ขึ้นกับผู้พูดมีค่าเพิ่มขึ้นเมื่อเพิ่มจำนวนคนในชุดฝึกขึ้น

4. จาก 2. และ 3. แสดงให้เห็นว่านิวรอลเน็ตเวิร์กเหมาะสมที่จะใช้ในการรู้จำเสียงพูดเมื่อมีตัวอย่างเสียงพูดที่ใช้ในการฝึกมากพอทำให้นิวรอลเน็ตเวิร์กรู้จักลักษณะทั่ว ๆ ไปของคำที่ต้องการรู้จำ จากการทดลองพบว่านิวรอลเน็ตเวิร์กสามารถประยุกต์ใช้ในการรู้จำเสียงตัวเลขภาษาไทยแบบคำโดดได้อย่างเหมาะสม โดยมีความถูกต้อง 89.4 % เมื่อฝึกนิวรอลเน็ตเวิร์กด้วยคน 30 คน

5. เมื่อทดสอบนิวรอลเน็ตเวิร์กกับเสียงตัวเลข 2 และ 3 พยางค์ พบว่าอัตราการรู้จำแบบไม่ขึ้นกับผู้พูดมีค่าเท่ากับ 84.7 % ซึ่งมีค่าต่ำกว่าผลการรู้จำคำโดด ทั้งนี้เป็นผลมาจากจำนวนคำศัพท์ที่เพิ่มขึ้นและการตัดคำผิดพลาดทำให้นิวรอลเน็ตเวิร์กตัดสินใจผิด

5.2 ปัญหา

1. เวลาที่ใช้ในการบันทึกเสียงพูดนานมากเพราะต้องการใช้เสียงของบุคคลจำนวนมาก

2. เสียงที่ได้จากการบันทึกมักมีเสียงรบกวน เพราะไม่ได้ทำการบันทึกเสียงในห้องเก็บเสียง เสียงรบกวนเหล่านี้มีผลทำให้การตัดหัวท้ายคำผิดพลาด

3. สำหรับโปรแกรมในส่วนนอร์แมลไลซ์ ได้มีการทดลองใช้วิธีนอร์แมลไลซ์โดยใช้การเปลี่ยนอัตราการชักตัวอย่าง ซึ่งควรให้ผลดีกว่าวิธีอื่น แต่จากการทดลองพบว่าวิธีนี้ใช้เวลาและหน่วยความจำในการคำนวณมาก และมีสัญญาณรบกวนขนาดเล็ก ๆ จำนวนมากเพิ่มเข้ามาในสัญญาณเสียง ทำให้การฝึกนิวรอลเน็ตเวิร์กไม่สามารถลู่เข้าสู่ระดับความผิดพลาดที่ต้องการได้ สัญญาณรบกวนเหล่านี้น่าจะเกิดจากความผิดพลาดสะสมในการคำนวณขณะนำสัญญาณไปผ่านวงจรกรองแบบผ่านต่ำ

4. ข้อมูลทั้งหมดที่ใช้ฝึกนิวรอลเน็ตเวิร์กมีขนาดใหญ่มาก ทำให้ต้องจัดรูปแบบการเก็บเป็นแฟ้มข้อมูล และอ่านข้อมูลที่ต้องการสู่หน่วยความจำทีละคู่ วิธีนี้ทำให้เวลาที่ใช้ในการฝึกนาน

ขึ้นเพราะต้องอ่านข้อมูลจากแฟ้มข้อมูลบ่อย ๆ อย่างไรก็ดีอาจใช้การสร้างไดรฟ์เสมือนด้วยหน่วยความจำเพื่อแก้ไขปัญหานี้ได้

5.3 ข้อเสนอแนะ

1. ควรเปลี่ยนขนาดข้อมูลในการบันทึกเสียงจาก 8 บิตเป็น 16 บิต เพราะจะทำให้อัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ทำให้ลดผลกระทบที่เกิดจากสัญญาณรบกวนลงได้

2. น่าจะมีการสร้างข้อมูลเสียงพูดมาตรฐานเก็บไว้ เพื่อลดเวลาที่ใช้ในการเก็บตัวอย่าง และการที่มีข้อมูลเสียงพูดมาตรฐานจะทำให้สามารถเปรียบเทียบผลการทดลองหลาย ๆ การทดลองได้ว่าวิธีใดมีประสิทธิภาพมากกว่า

3. การตัดหัวท้ายคำที่ใช้เป็นการตัดส่วนหัวและท้ายคำทิ้งเท่านั้น ไม่ได้ตัดช่วงเงียบที่อยู่ระหว่างพยางค์ อาจพัฒนาการตัดคำให้ดีขึ้นจนสามารถตัดเป็นพยางค์หรือหน่วยเสียง ซึ่งจะนำไปสู่การรู้จำคำต่อเนื่องได้

4. การนอร์แมลไลซ์เสียงวิธีใดก็ตาม ต่างทำให้เกิดความเพี้ยนทั้งสิ้น ถ้ามีทางแก้ไขได้ควรเลือกวิธีที่ไม่ต้องมีการนอร์แมลไลซ์สัญญาณเสียง

5. ปัญหาเรื่องหน่วยความจำ สามารถแก้ไขได้อีกวิธีโดยการเขียนโปรแกรมให้สามารถอ้างหน่วยความจำเกิน 1 เมกะไบต์ได้