บทที่ 4
การทดลอง และการวิเคราะห์ผลการทดลอง

4.1 ข้อมูล และการเก็บข้อมูล

งานวิจัยนี้จะทำการรู้จำเสียงคำไทยหลายพยางค์ ประกอบด้วยคำศัพท์ 4 ชุดดังนี้

ชุดที่ 1 เสียงคำ 1 พยางค์ 20 ตัว (ไม่ใช่เสียงตัวเลข)

ชุดที่ 2 เสียงคำ 2 พยางค์ 20 ตัว

ชุดที่ 3 เสียงคำ 3 พยางค์ 20 ตัว

ชุดที่ 4 เสียงตัวเลข 10 ตัว (0 ถึง 9)

คำศัพท์ที่ใช้เป็นคำที่พูดทั่วไปในชีวิตประจำวัน เช่น คำกริยา ชื่อวัน ชื่อสัตว์ และสิ่งของ เป็นต้น รายละเอียดคำศัพท์แต่ละชุดพร้อมทั้งรูปเสียง (Phonetic) แสดงในภาคผนวก ก การบันทึกเสียงมีข้อกำหนด และรายละเอียดดังนี้

1) เก็บข้อมูลจากผู้พูดที่ใช้ภาษาไทยได้ปกติ อายุในช่วง 18-25 ปี

2) เก็บข้อมูลจากผู้พูดจำนวน 60 คน แต่ละคนพูดคำศัพท์ละ 2 ครั้ง

3) บันทึกเสียงด้วยเครื่องคอมพิวเตอร์ 80486DX2-66 พร้อมด้วยการ์ดเสียง Sound Blaster 16 ของบริษัท Creative Technology และไมโครโฟน Philips Uni-directional Microphone รุ่น SBC 465 บันทึกเสียงโดยใช้โปรแกรม Creative WaveStudio for Windows version 2.0

นำข้อมูลที่ได้ทั้งหมดมาแบ่งเป็น 3 ชุด ได้แก่

ชุดฝึกฝน (Training Set) ประกอบด้วยเสียงพูดครั้งแรกของผู้พูด 50 คน

ชุดทดสอบที่ 1 (Dependent Set) ประกอบด้วยเสียงพูดครั้งที่ 2 ของผู้พูดในชุดฝึกฝน

ชุดทดสอบที่ 2 (Independent Set) ประกอบด้วยเสียงพูดครั้งหนึ่งของผู้พูดอีก 10 คน

โดยใช้ชุดฝึกฝนในการฝึกฝนนิวรอลเน็ตเวอร์ก และใช้ชุดฝึกฝน และชุดทดสอบทั้งสองในการทดสอบการรู้จำของระบบ การทดสอบด้วยชุดทดสอบที่ 2 จะได้ผลการรู้จำแบบไม่ขึ้นกับผู้พูด

4.2 การทดลองเพื่อวิเคราะห์หาค่าพารามิเตอร์ที่เหมาะสม

4.2.1 พารามิเตอร์ในการตัดหัวท้ายคำ

ขั้นตอนของโปรแกรมในการตัดหัวท้ายคำแสดงในรูปที่ ข.1 ในการคำนวณ จะกำหนดให้หนึ่งส่วนย่อยในการคำนวณในขั้นตอนนี้ ประกอบด้วยจุดข้อมูลจำนวน 100 จุด เมื่อคำนวณค่า $E(j)$ สำหรับส่วนย่อยที่ j ใดๆ พบว่าสัญญาณเสียงช่วงที่เงียบจะมีค่า $E(j)$ ไม่ถึง 5 เปอร์เซ็นต์ของค่า $E(j)$ สูงสุด (E_{max}) ที่เกิดขึ้นในคำนั้นๆ อย่างแน่นอน จึงกำหนด

$$E_{threshold} = 0.05E_{max} \tag{4.1}$$

จากการพิจารณาค่าพลังงานของสัญญาณในช่วงที่เปล่งเสียง ดังตัวอย่างกราฟในรูป 3.3 จะเห็นว่า ถ้ามีส่วนย่อยที่มีค่า $E(j)$ สูงกว่า $E_{threshold}$ ติดกันเกิน 2 ส่วนย่อย ก็จะถือได้ว่าเป็นส่วนของเสียงพูดได้แล้ว แต่เพื่อความถูกต้องแน่นอนของการหาจุดเริ่มต้น และจุดท้ายคำ จะกำหนดให้ต้องมีส่วนย่อยที่มี $E(j)$ สูงกว่า $E_{threshold}$ ติดกันเกิน 5 ส่วนย่อย (Frame Duration = 5) จึงจะนับเป็นช่วงที่เปล่งเสียง ส่วนจำนวนจุดข้อมูลที่ต้องบวกเผื่อที่ต้น และท้ายคำ จะต้องเก็บข้อมูลจากชุดฝึกฝนดังนี้

ใช้โปรแกรมตัดหัวท้ายคำตัวอย่าง ทุกเสียงคำศัพท์จากผู้พูดในชุดฝึกฝนจำนวน 5 คน โดยไม่มีการบวกเผื่อทั้งหัว และท้ายคำ (Front Add = Back Add = 0) จะได้จุดเริ่มต้นคำ และจุดท้ายคำที่ได้จากการโปรแกรมคือ จุด a_1 และ b_1 ตามลำดับ หลังจากนั้นพิจารณาจุดตัดที่ได้ในรูปคลื่นของสัญญาณเสียงว่าควรบวกเผื่อที่ต้นคำ และท้ายคำจากจุด a_1 และ b_1 ไปอีกเท่าไร จึงจะครอบคลุมสัญญาณช่วงที่เปล่งเสียงแน่นอน จุดหัว และท้ายคำใหม่ที่ได้จากการบวกเผื่อคือ จุด a_2 และ b_2 ตามลำดับ ดังตัวอย่างที่แสดงในรูปที่ 3.3 หลังจากนั้น จะหาค่าที่เหมาะสมเพียงค่าเดียวของจำนวนจุดข้อมูลที่ควรบวกเพิ่มสำหรับแต่ละคำศัพท์ ผลการทดลองแสดงไว้ในภาคผนวก ค เพื่อให้การตัดหัวท้ายคำครอบคลุมช่วงสัญญาณที่มีการเปล่งเสียงทุกคำศัพท์ จึงเลือกบวกเผื่อต้นคำ 2,000 จุดข้อมูล และท้ายคำ 2,500 จุดข้อมูล ค่าที่บวกเพิ่มนี้ ถึงแม้ว่าจะมีค่ามากเกินที่ควรจะเป็นสำหรับบางคำศัพท์ แต่การบวกเผื่อจะกระทำทุกๆ สัญญาณเสียง จึงทำให้มีผลกระทบต่อการรู้จำไม่มาก

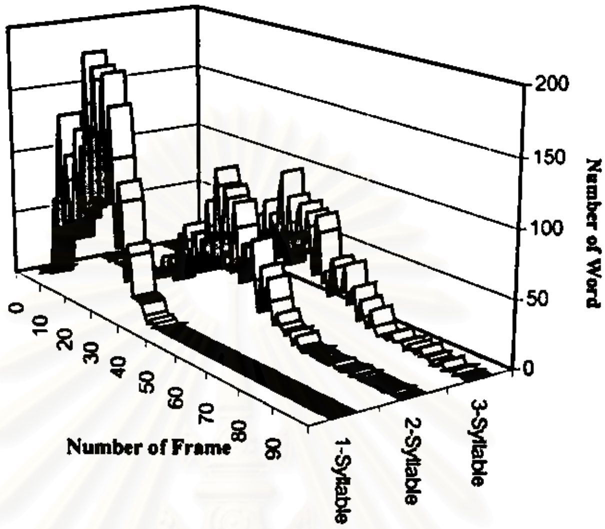
4.2.2 พารามิเตอร์ในการสกัดค่าลักษณะเด่น

ทำการแบ่งสัญญาณเสียงทั้งคำออกเป็นส่วนย่อย โดยมีความยาวส่วนย่อยละ 20 มิลลิวินาที (Furui, 1989) ซึ่งจะได้จำนวนจุดข้อมูลในหนึ่งส่วนย่อยเท่ากับ 220 จุด (20 มิลลิวินาที × 11 กิโลเฮิรตซ์) โดยมีการเหลื่อมกัน 1 ใน 4 ของส่วนย่อย (Furui, 1989) คือเหลื่อมกัน 55 จุดข้อมูล ใช้สัมประสิทธิ์ LPC จำนวน 10 อันดับ (Pornsukchantra, 1996) สำหรับสัญญาณเสียงแต่ละส่วนย่อย ซึ่งเพียงพอสำหรับใช้ในการรู้จำเสียง

4.2.3 พารามิเตอร์ในการนอร์มอลไลซ์ทางเวลา

เมื่อใช้ส่วนย่อยซึ่งประกอบด้วย 220 จุดข้อมูล โดยเหลื่อมกัน 55 จุดข้อมูล จะได้จำนวนส่วนย่อย L สำหรับเสียงคำแต่ละคำขึ้นอยู่กับจำนวนจุดข้อมูลทั้งหมดของคำนั้นๆ เสียงคำ

ที่มีจำนวนพยางค์ต่างกันจะถูกนอร์มอลไลซ์ทางเวลาให้มีจำนวนจุดข้อมูลต่างกัน ทำให้ได้จำนวนส่วนย่อยต่างกันด้วย รูปที่ 4.1 แสดงการกระจายของเสียงคำหลังจากผ่านการตัดหัวท้ายคำ โดยไม่มีการบวกเผื่อหัวท้ายคำ ซึ่งจะมีจำนวนส่วนย่อยต่างๆ กัน



รูปที่ 4.1 กราฟการกระจายของเสียงคำที่มีจำนวนส่วนย่อยต่างๆ กัน

การเลือกจำนวนส่วนย่อยที่เหมาะสมสำหรับเสียงคำแต่ละชุด (1, 2 และ 3 พยางค์) นอกจากพิจารณาจากกราฟในรูปที่ 4.1 แล้ว จากการทดลองยังพบว่าจำนวนส่วนย่อยต่างๆ กันมีผลต่ออัตราการรู้จำ และเวลาในการประมวลผลค่อนข้างมาก จึงจะทำการทดลองดังนี้

1) การทดลองเพื่อหาจำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 1 พยางค์ จะเลือกคำศัพท์จากชุดเสียงตัวเลข และชุดเสียงคำ 1 พยางค์มาจำนวนหนึ่ง ในที่นี้เลือกกลุ่มคำศัพท์ที่ 6 ดังแสดงในตารางที่ ก.5 ซึ่งมีจำนวนทั้งหมด 7 คำศัพท์ โดยใช้เสียงพูดจากผู้พูด 30 คนในชุดฝึกฝนเพื่อฝึกฝนนิวรอลเน็ตเวอร์ก ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC ไม่ต้องแปลงเป็นค่าสมาชิกภาพแบบฟัซซี และใช้เสียงจากผู้พูด 10 คนในชุดทดสอบที่ 2 ในการทดสอบการรู้จำ กำหนดค่าพารามิเตอร์ $E_t = 0.01$ $\varepsilon = 0.1$ และ $\alpha = 0.9$ และเปลี่ยนแปลงจำนวนจุดในการนอร์มอลไลซ์ทางเวลา ซึ่งก็คือการเปลี่ยนแปลงจำนวนส่วนย่อย โดยพิจารณาจำนวนส่วนย่อยในช่วงที่เป็นไปได้ในกราฟรูปที่ 4.1 รวมทั้งทดลองเปลี่ยนแปลงจำนวนโหนดในระดับชั้นซ่อนตัวซึ่งอาจจะเปลี่ยนแปลงไปเมื่อเปลี่ยนแปลงจำนวนโหนดในระดับชั้นข้อมูลเข้า ผลการทดลองแสดงได้ดังตารางที่ 4.1

จากตารางที่ 4.1 จำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 1 พยางค์เท่ากับ 46 ส่วนย่อยต่อคำ และจะใช้ค่านี้กับเสียงตัวเลขด้วย เนื่องจากเป็นคำ 1 พยางค์เช่นกัน

ตารางที่ 4.1 อัตราการรู้จำที่ได้จากชุดทดสอบที่ 2
เพื่อหาจำนวนส่วนย่อยที่เหมาะสมสำหรับเสียงคำ 1 พยางค์

จำนวนโหนดในระดับชั้นซ่อนตัว	อัตราการรู้จำ (%)									
	จำนวนส่วนย่อย (Number of Frame)									
	30	33	37	40	43	46	49	52	55	58
100	70.1	73.3	74.5	75.0	75.9	77.9	76.7	73.5	73.9	75.3
150	71.6	73.9	75.3	76.1	76.9	78.1	76.4	73.6	74.4	73.9
200	73.0	73.9	75.4	76.4	76.1	78.4	76.6	73.6	74.0	74.6

2) การทดลองเพื่อหาจำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 2 และ 3 พยางค์ จะทำการทดลองทำนองเดียวกับการหาจำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 1 พยางค์ แต่ใช้กลุ่มคำศัพท์ที่ 11 ซึ่งมีจำนวน 10 คำศัพท์ และกลุ่มคำศัพท์ที่ 15 ซึ่งมีจำนวน 8 คำศัพท์ ดังแสดงในตารางที่ ก.5 ได้ผลการทดลองสำหรับคำ 2 และ 3 พยางค์ ดังตารางที่ 4.2 และ 4.3 ตามลำดับ

จากตารางที่ 4.2 จำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 2 พยางค์เท่ากับ 49 ส่วนย่อยต่อคำ และจากตารางที่ 4.3 จำนวนส่วนย่อยที่เหมาะสมสำหรับคำ 3 พยางค์เท่ากับ 52 ส่วนย่อยต่อคำ

ตารางที่ 4.2 อัตราการรู้จำที่ได้จากชุดทดสอบที่ 2
เพื่อหาจำนวนส่วนย่อยที่เหมาะสมสำหรับเสียงคำ 2 พยางค์

จำนวนโหนดในระดับชั้นซ่อนตัว	อัตราการรู้จำ (%)									
	จำนวนส่วนย่อย (Number of Frame)									
	46	49	52	55	58	61	64	67	70	73
100	80.2	81.0	78.8	72.6	74.2	76.6	73.8	74.4	75.0	75.8
150	79.8	80.4	76.0	75.8	73.8	74.4	73.2	74.0	74.0	75.2
200	80.0	82.2	77.0	74.8	74.6	77.2	74.0	74.8	74.8	78.4

ตารางที่ 4.3 อัตราการรู้จำที่ได้จากชุดทดสอบที่ 2
เพื่อหาจำนวนส่วนย่อยที่เหมาะสมสำหรับเสียงคำ 3 พยางค์

จำนวนโหนดในระดับชั้นซ่อนตัว	อัตราการรู้จำ (%)									
	จำนวนส่วนย่อย (Number of Frame)									
	46	49	52	55	58	61	64	67	70	73
100	70.8	66.8	73.3	70.5	67.0	66.2	71.6	69.4	66.8	63.0
150	69.8	70.0	72.5	71.0	66.8	66.2	70.4	66.0	65.8	61.6
200	70.3	70.3	71.5	71.3	65.8	66.6	70.8	67.2	65.4	62.0

4.2.4 จำนวนโหนดที่เหมาะสมในระดับชั้นซ่อนตัว

จำนวนโหนดในระดับชั้นซ่อนตัว ขึ้นอยู่กับจำนวนโหนดในระดับชั้นข้อมูลเข้าและระดับชั้นข้อมูลออกเป็นหลัก จากผลจำนวนส่วนย่อยที่จะใช้สำหรับคำศัพท์ชุดต่างๆ ในหัวข้อที่ 4.2.3 ทำให้นิวรอลเน็ตเวอร์กย่อยมีจำนวนโหนดในระดับชั้นข้อมูลเข้าต่างๆ กันคือ ในกรณีข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC ของคำ 1, 2 และ 3 พยางค์ และกรณีข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีซึ่งจะมีจำนวนข้อมูลเข้าเพิ่มขึ้น 3 เท่า ดังนั้นจะมีจำนวนโหนดในระดับชั้นข้อมูลเข้าได้ 6 ค่าคือ { 460, 490, 520, 1380, 1470, 1560 } ส่วนจำนวนโหนดในระดับชั้นข้อมูลออกขึ้นอยู่กับจำนวนคำศัพท์ที่จะรู้จำด้วยนิวรอลเน็ตเวอร์กย่อยนั้นๆ ซึ่งมี 8 ค่าคือ { 2, 3, 5, 7, 8, 10, 20, 30 }

การทดลองเพื่อหาจำนวนโหนดในระดับชั้นซ่อนตัวที่เหมาะสมในกรณีต่างๆ ดังนี้

1) ใช้กลุ่มคำศัพท์ในการทดลอง ที่ใช้จำนวนโหนดในระดับชั้นข้อมูลออกเท่ากัน แต่เปลี่ยนแปลงจำนวนโหนดในระดับชั้นข้อมูลเข้าคือ กลุ่มคำศัพท์ที่ 10 กลุ่มคำศัพท์ที่ 14 และกลุ่มคำศัพท์ที่ 16 ดังแสดงในตารางที่ ก.5 ทั้งสามกลุ่มจะใช้จำนวนโหนดในระดับชั้นข้อมูลออกเท่ากันคือ 5 โหนด แต่มีจำนวนโหนดในระดับชั้นข้อมูลเข้าแตกต่างกัน โดยใช้ผู้พูดจำนวน 30 คนในชุดฝึกฝน เพื่อฝึกฝนนิวรอลเน็ตเวอร์ก และทดสอบด้วยชุดทดสอบที่ 2 กำหนดพารามิเตอร์อื่นๆ ให้คงที่คือ $E_t = 0.01$ $\varepsilon = 0.1$ และ $\alpha = 0.9$ ผลการทดลองดังตารางที่ 4.4

2) ใช้กลุ่มคำศัพท์ในการทดลอง ที่ใช้จำนวนโหนดในระดับชั้นข้อมูลเข้าเท่ากัน แต่เปลี่ยนแปลงจำนวนโหนดในระดับชั้นข้อมูลออกคือ กลุ่มคำศัพท์ที่ 1, 2, 3, 6, 7, 9 และ 10 ดังแสดงในตารางที่ ก.5 ซึ่งใช้จำนวนโหนดในระดับชั้นข้อมูลเข้าในกรณีข้อมูลแบบสัมประสิทธิ์ LPC เท่ากับ 460 โหนด แต่ใช้จำนวนโหนดในระดับชั้นข้อมูลออกตั้งแต่ 3 ถึง 30 โหนด ทำการทดลองทำนองเดียวกับขั้นตอนที่ 1) ผลการทดลองแสดงในตารางที่ 4.5

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.4 อัตราการรู้จำที่ได้จากชุดทดสอบที่ 2
สำหรับกลุ่มคำศัพท์ที่ใช้จำนวนโหนดในระดับชั้นข้อมูลออกเท่ากันคือ 5 โหนด

กลุ่มคำศัพท์ที่	ชนิดของข้อมูลเข้า	จำนวนโหนดเข้า	จำนวนโหนดออก	จำนวนโหนดในระดับชั้นซ่อนตัว							
				30	50	70	100	120	150	170	200
10	LPC	460	5	58.7	62.7	64.0	60.0	61.3	62.7	62.7	64.0
14	LPC	490	5	61.3	62.7	69.3	66.0	65.3	66.0	67.3	69.0
16	LPC	520	5	69.3	73.3	78.7	77.3	72.0	72.7	70.7	74.7
10	Fuzzy	1380	5	58.7	62.0	65.3	66.0	64.0	62.7	60.7	61.0
14	Fuzzy	1470	5	65.3	69.3	71.3	75.3	72.0	72.0	68.7	69.0
16	Fuzzy	1560	5	61.3	70.7	72.0	72.7	72.7	70.3	70.0	71.3

ตารางที่ 4.5 อัตราการรู้จำที่ได้จากชุดทดสอบที่ 2
สำหรับกลุ่มคำศัพท์ที่ใช้จำนวนโหนดในระดับชั้นข้อมูลเข้าเท่ากันคือ 460 โหนด

กลุ่มคำศัพท์ที่	ชนิดของข้อมูลเข้า	จำนวนโหนดเข้า	จำนวนโหนดออก	จำนวนโหนดในระดับชั้นซ่อนตัว							
				30	50	70	100	120	150	170	200
9	LPC	460	3	90.0	93.3	92.3	90.3	93.3	92.0	94.0	92.3
10	LPC	460	5	78.7	82.7	84.0	80.0	81.3	82.7	82.7	83.0
6	LPC	460	7	78.0	79.0	80.6	82.4	81.0	80.4	77.9	78.0
7	LPC	460	8	77.5	80.3	82.8	83.8	83.3	82.5	79.8	70.0
1	LPC	460	10	72.0	74.0	76.0	74.0	74.0	73.0	74.0	74.0
2	LPC	460	20	56.0	62.5	61.0	65.0	63.5	63.0	63.0	64.0
3	LPC	460	30	52.0	57.8	58.3	58.4	60.1	60.9	58.7	58.7

จะเห็นได้ว่า จำนวนโหนดในระดับชั้นซ่อนตัวมีผลมาจากจำนวนโหนดในระดับชั้นข้อมูลออกมากกว่าในระดับชั้นข้อมูลเข้า ดังแสดงให้เห็นในตารางที่ 4.4 เมื่อจำนวนโหนดในระดับชั้นข้อมูลเข้าเปลี่ยนแปลงไปจาก 460 เป็น 490 หรือ 520 โหนด จำนวนโหนดในระดับชั้น

ซ่อนตัวที่เหมาะสมยังคงมีค่าเท่าเดิมคือ 70 โหนด และจะเปลี่ยนไปเป็น 100 โหนดเมื่อจำนวนโหนดในระดับชั้นข้อมูลเข้าเพิ่มขึ้นสูงมากคือ 1380, 1470 หรือ 1560 โหนด แต่ถ้าพิจารณาในตารางที่ 4.5 เมื่อจำนวนโหนดในระดับชั้นข้อมูลออกเปลี่ยนแปลงไปจาก 3 หรือ 5 เป็น 7, 8 หรือ 10 จำนวนโหนดในระดับชั้นซ่อนตัวที่เหมาะสมจะเปลี่ยนแปลงจาก 50 เป็น 70 โหนด และจะเพิ่มสูงขึ้นไปอีกเอจำนวนโหนดในระดับชั้นข้อมูลออกเพิ่มขึ้นเป็น 20 และ 30 โหนด เมื่อพิจารณาดังนี้ จะเลือกใช้จำนวนโหนดในระดับชั้นซ่อนตัวดังในตารางที่ 4.6

ตารางที่ 4.6 จำนวนโหนดในระดับชั้นซ่อนตัวที่เหมาะสมสำหรับแต่ละกลุ่มคำศัพท์

จำนวนโหนดใน	จำนวนโหนดในระดับชั้นข้อมูลออก							
ระดับชั้นข้อมูลเข้า	2	3	5	7	8	10	20	30
460, 490, 520	50	50	70	70	70	70	100	150
1380, 1470, 1560	70	70	100	100	100	100	150	200

4.2.5 พารามิเตอร์ในกระบวนการเกี่ยวกับฟัซซี

1) ค่า i_{max} และ i_{min} สำหรับกำหนดขอบเขตของฟังก์ชันสมาชิกภาพแบบฟัซซี

หลังจากทำการสกัดค่าลักษณะเด่นแบบ LPC แล้วปรากฏว่าค่าสัมประสิทธิ์ LPC สูงสุด และต่ำสุดที่เกิดขึ้นมีค่าเท่ากับ 2.342 และ –2.376 ตามลำดับ ดังนั้นจะเลือกใช้ i_{max} และ i_{min} เท่ากับ 2.5 และ –2.5 ตามลำดับ

2) ค่า *fdenom* ซึ่งควบคุมการซ่อนทับในฟังก์ชันสมาชิกภาพแบบฟัซซีชนิด Pi

จากงานวิจัยของ Pal (Pal and Mitra, 1992) ได้เสนอไว้ว่าค่า *fdenom* ที่เหมาะสมคือ 0.8 แต่ชนิดของข้อมูล และคำศัพท์ที่แตกต่างกันอาจทำให้ต้องใช้ค่า *fdenom* เปลี่ยนแปลงไปได้ เนื่องจากค่านี้เป็นตัวกำหนดการซ้อนทับของฟังก์ชัน Pi มีผลต่อการกำหนดความกำกวม (Fuzziness) ของข้อมูลเข้าโดยตรง ดังนั้นจะทำการทดลองเพื่อหาค่า *fdenom* ที่เหมาะสม โดยฝึกฝนนิวรอลเน็ตเวอร์กด้วยกลุ่มคำศัพท์ที่ 12 ซึ่งประกอบด้วยค่าสัมประสิทธิ์ LPC ค่อนข้างครอบคลุมคือ มีค่าสัมประสิทธิ์ LPC สูงสุด และต่ำสุดคือ 2.332 และ –2.376 ใช้ผู้พูดจากกลุ่มฝึกฝนจำนวน 10 คน และทดสอบด้วยชุดทดสอบที่ 2 โดยใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi ที่เปลี่ยนแปลงค่า *fdenom* ส่วนค่าอื่นๆ คงที่คือ $E_t = 0.01$ $\varepsilon = 0.1$ และ $\alpha = 0.9$ ผลการทดลองแสดงดังตารางที่ 4.7

ตารางที่ 4.7 ผลอัตราการรู้จำจากชุดทดสอบที่ 2 สำหรับหาค่า *fdenom* ที่เหมาะสม

ค่า *fdenom*	อัตราการรู้จำ (%)
0.5	69.3
0.8	76.7
1.0	77.3
1.2	70.0

จากตารางที่ 4.7 ค่า *fdenom* ที่เหมาะสมคือ 1.0 ซึ่งจะส่งผลให้กราฟของฟังก์ชัน Pi มีการซ้อนทับกันน้อยกว่าค่า 0.8 นั่นคือข้อมูลเข้าจะมีความกำกวมลดลงจากการใช้ค่า 0.8 สาเหตุที่เป็นเช่นนี้อาจจะเพราะว่าข้อมูลเข้าแบบสัมประสิทธิ์ LPC ขนาด 10 อันดับ ทำให้จำนวนข้อมูลเข้ามาก เมื่อเทียบกับงานวิจัยของ Pal (Pal and Mitra, 1992) ซึ่งใช้ค่าความถี่ฟอร์แมนท์ (Formant) ลำดับที่ 1, 2 และ 3 เพียง 3 ค่า และเมื่อข้อมูลจำนวนมาก การแปลงค่าเป็นค่าสมาชิกภาพแบบฟัซซีก็ไม่ควรจะให้มีความกำกวมมากนัก

3) ค่า f_d และ f_e ซึ่งควบคุมความกำกวมของข้อมูลออกที่ต้องการ

ค่า f_d และ f_e ในสมการที่ 2.22 จะใช้ในกรณีข้อมูลออกที่ต้องการเป็นแบบค่าสมาชิกภาพของแต่ละคำศัพท์ ซึ่งเป็นค่าที่ควบคุมความกำกวมของข้อมูลออกที่ต้องการ งานวิจัยของ Pal (Pal and Mitra, 1992) เสนอค่าที่เหมาะสมคือ 5 และ 1 ตามลำดับ ในขณะที่งานวิจัยของ Reyes-Garcia และ Bandler เสนอค่าที่เหมาะสมในกรณีที่ใช้ระดับของลักษณะทางภาษา (Linguistic Properties) 5 ระดับคือ 6 และ 3 ตามลำดับ และในกรณีที่ใช้ระดับของลักษณะทางภาษา 7 ระดับคือ 4 และ 3 ตามลำดับ ในที่นี้จะทำการทดลองเพื่อหาค่า f_d และ f_e ที่เหมาะสมโดยทำการฝึกฝนนิวรอลเน็ตเวอร์กสำหรับกลุ่มคำศัพท์ที่ 12 ดังแสดงในตารางที่ ก.5 ด้วยเสียงผู้พูด 10 คนจากชุดฝึกฝน และทำการทดสอบด้วยเสียงชุดที่ใช้ฝึกฝน และเสียงผู้พูดอีก 10 คนจากชุดทดสอบที่ 2 ข้อมูลเข้าที่ใช้เป็นค่าสมาชิกภาพแบบฟัซซีชนิดสี่เหลี่ยมคางหมู ข้อมูลออกที่ต้องการหาได้ตามแผนภาพกระบวนการในรูปที่ ข.3 โดยเปลี่ยนแปลงค่า f_d และ f_e และคงค่าอื่นๆ คือ $E_t = 0.000001$ $\varepsilon = 0.1$ และ $\alpha = 0.9$ ผลการทดลองแสดงในตารางที่ 4.8

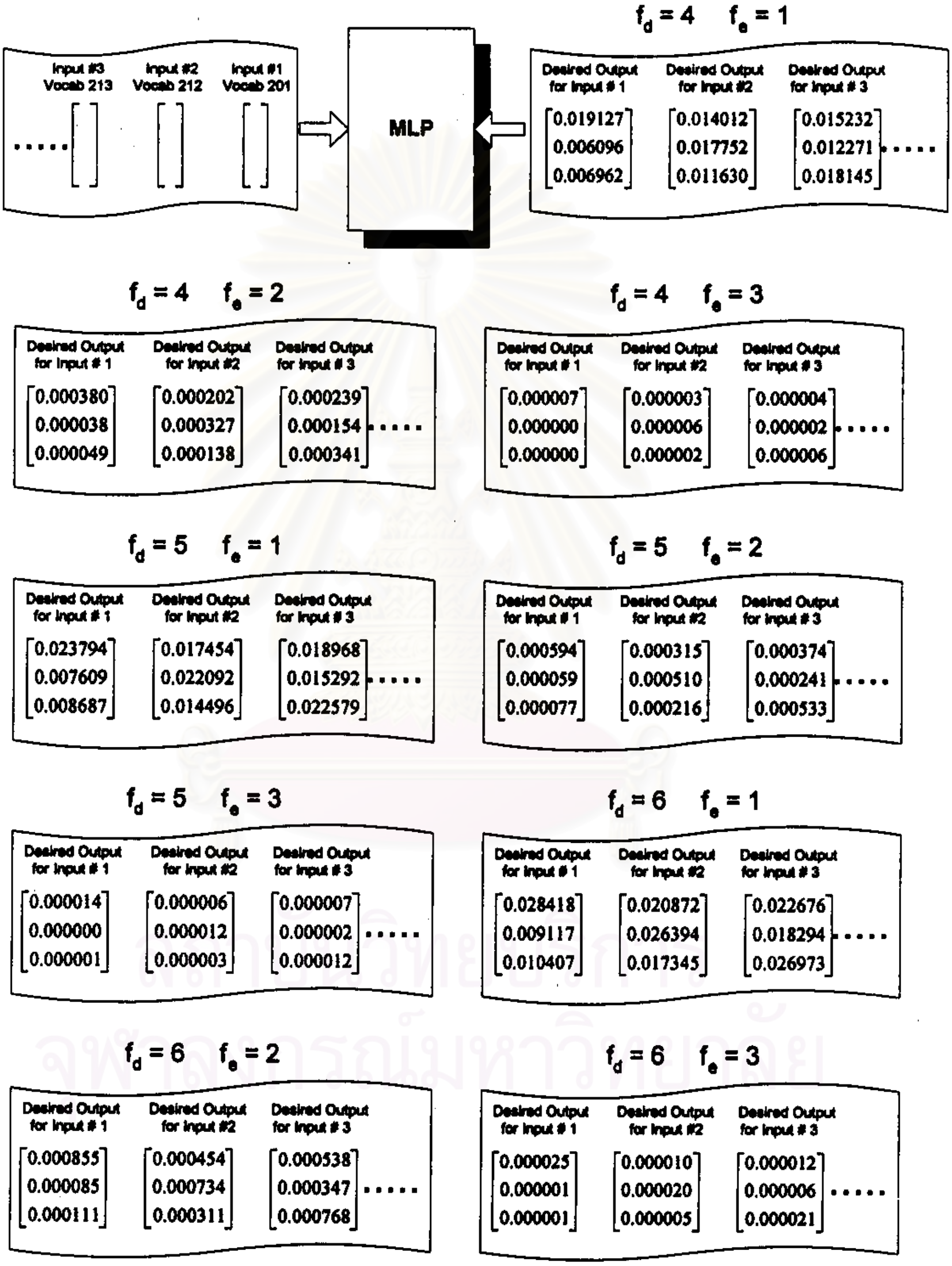
จากตารางที่ 4.8 ค่า f_d และ f_e ที่เหมาะสมคือ 5 และ 1 ตามลำดับ รูปที่ 4.2 แสดงตัวอย่างเวกเตอร์ข้อมูลออกที่ต้องการซึ่งใช้ในการทดลองนี้ เวกเตอร์ข้อมูลออกที่ต้องการซึ่งมีมิติเท่ากับจำนวนคำศัพท์ในกลุ่มคือ 3 คำศัพท์ จะประกอบด้วยค่าสมาชิกภาพของแต่ละคำศัพท์ $\mu_k(i)$ โดยจะมีค่าสูงสุด ณ ตำแหน่งโหนดที่สอดคล้องกับคำที่เข้ามาฝึกฝนขณะนั้น เมื่อพิจารณา

ค่าข้อมูลออกในแต่ละกรณีของ f_d และ f_e จะเห็นว่า เมื่อ $f_e = 1$ เช่นในกรณีที่ $f_d = 4$ และ $f_e = 1$ ค่าข้อมูลออกของเวกเตอร์ข้อมูลออกซึ่งสอดคล้องกับคำฝึกฝนคำแรก (คำศัพท์หมายเลข 201) จะมีค่าสูงสุดที่โหนดแรกคือ 0.019127 ส่วนโหนดอื่นๆ จะมีค่าต่ำกว่ามากคือ 0.006096 และ 0.006962 ในกรณีนี้ การใช้ค่า $E_t = 0.000001$ จะเพียงพอสำหรับการฝึกฝนนิวรอลเน็ตเวอร์ก แต่ในกรณีที่ $f_e \neq 1$ เช่นในกรณีที่ $f_d = 4$ และ $f_e = 2$ หรือ 3 ค่าข้อมูลออกที่ต้องการจะมีค่าต่ำมาก และโดยเฉพาะในกรณีที่ $f_e = 3$ นอกจากค่าต่ำมากแล้ว ยังมีค่าใกล้เคียงกันมากด้วย ลักษณะนี้ทำให้การใช้ $E_t = 0.000001$ ไม่ต่ำพอที่จะฝึกฝนนิวรอลเน็ตเวอร์กให้รู้จำได้ถูกต้องได้ ผลก็คือ อัตราการรู้จำจะตกลงอย่างเห็นได้ชัดดังในตารางที่ 4.8 อย่างไรก็ตาม ถ้าหากใช้ค่า E_t ต่ำกว่านี้ลงไปอีก จะสามารถฝึกฝนให้รู้จำได้ถูกต้องมากขึ้น แต่ก็ต้องการเวลาในการฝึกฝนมากขึ้นตามไปด้วย ซึ่งไม่มีความจำเป็น นอกจากนี้ เมื่อพิจารณากรณีที่ f_d เปลี่ยนแปลง แต่ $f_e = 1$ คงที่ จะเห็นว่า กรณีที่ $f_d = 5$ และ $f_e = 1$ ค่าข้อมูลออก ณ ตำแหน่งที่สอดคล้องกับคำที่เข้ามาฝึกฝน จะมีค่าสูงเด่นชัดกว่ากรณีอื่นๆ จึงสมควรที่จะได้อัตราการรู้จำมากที่สุด

ตารางที่ 4.8 ผลอัตราการรู้จำจากชุดทดสอบที่ 2
สำหรับหาค่า f_d และ f_e ที่เหมาะสม

ค่า f_d	ค่า f_e	อัตราการรู้จำ (%)	
		ชุดฝึกฝน	ชุดทดสอบที่ 2
4	1	93.3	54.2
4	2	33.3	33.3
4	3	33.3	33.3
5	1	93.3	63.3
5	2	26.7	33.3
5	3	33.3	33.3
6	1	91.7	48.3
6	2	30.0	40.0
6	3	33.3	33.3

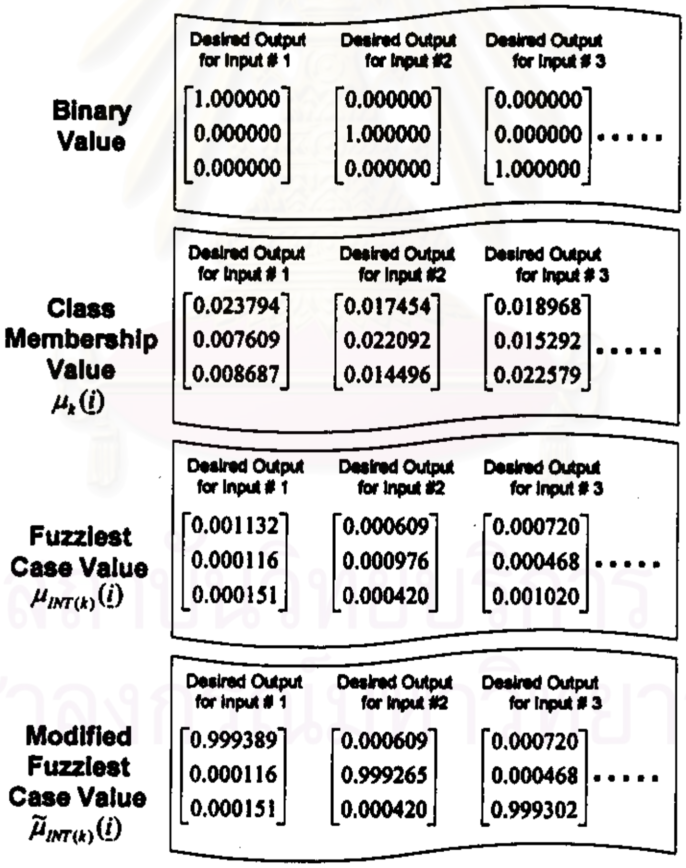
รูปที่ 4.2 ตัวอย่างเวกเตอร์ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ ในกรณี f_d และ f_e ต่างๆ กัน



4.3 การวิเคราะห์ค่าข้อมูลออกที่ต้องการในกรณีกำกวมที่สุด (Fuzziest Case)

ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ $\mu_k(i)$ ในบางกรณี อาจเกิดความกำกวมขึ้น ดังเช่นในรูปที่ 4.2 กรณี $f_d = 5$ และ $f_e = 1$ เวกเตอร์ข้อมูลออกที่ตรงกับคำฝึกฝนคำที่ 2 (คำศัพท์หมายเลข 212) และคำที่ 3 (คำศัพท์หมายเลข 213) ประกอบด้วยค่า $\mu_k(i)$ ที่ใกล้เคียงกันเกินไป เมื่อเทียบกับเวกเตอร์ข้อมูลออกที่ตรงกับคำฝึกฝนคำแรก (คำศัพท์หมายเลข 201) กรณีนี้เรียกว่าเกิดความกำกวมที่สุด (Fuzziest Case) ซึ่ง Pal (Pal and Mitra, 1992) ได้เสนอวิธีการแก้ไขไว้ โดยปรับค่า $\mu_k(i)$ ใหม่เป็นค่า $\mu_{INT(k)}(i)$ ตามสมการที่ 2.23 เมื่อนำมาใช้กับข้อมูลออกในกรณี $f_d = 5$ และ $f_e = 1$ ในรูปที่ 4.2 จะได้ผลของเวกเตอร์ข้อมูลออกในกรณีกำกวมที่สุดดังรูปที่ 4.3

รูปที่ 4.3 ตัวอย่างเวกเตอร์ข้อมูลออกที่ต้องการแบบต่างๆ



จะเห็นว่า $\mu_{INT(k)}(i)$ มีค่าต่ำมากและไม่ช่วยให้ค่าในแต่ละเวกเตอร์แยกแยะกันเด่นชัดขึ้นเลย สาเหตุเนื่องจากค่าข้อมูลออก $\mu_k(i)$ ที่ได้จากสมการที่ 2.22 จะมีค่าอยู่ในช่วง [0, 1] และสมการที่ 2.23 สร้างไว้สำหรับค่า $\mu_k(i)$ ที่กำกวม ซึ่งเป็นที่อยู่ในช่วงกึ่งกลางระหว่างค่า [0, 1] ดังที่ได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เห็นในสมการที่ 2.23 คือเมื่อค่า $\mu_k(i)$ มีค่าสูงกว่า 0.5 จะถูกปรับให้มีค่าสูงขึ้น แต่ถ้าค่า $\mu_k(i)$ มีค่าต่ำกว่า 0.5 จะถูกปรับให้มีค่าต่ำลง ดังนั้นจะได้ค่าข้อมูลออกใหม่แยกแยะกันเด่นชัดยิ่งขึ้น แต่สำหรับ $\mu_k(i)$ ที่ได้ในการทดลองของงานวิจัยนี้ มีค่าต่ำมากคืออยู่ในช่วง 0.1 ถึง 0.0001 ดังนั้นหากใช้สมการที่ 2.23 ค่าทั้งหมดจะถูกปรับให้ต่ำลงไปอีก ส่งผลให้ข้อมูลออกไม่ได้ถูกแยกแยะให้เด่นชัดขึ้น

อาศัยหลักการของสมการที่ 2.23 จึงได้ปรับเปลี่ยนสมการให้เหมาะสมกับข้อมูลในงานวิจัยนี้ โดยหาค่าเฉลี่ยของค่า $\mu_k(i)$ ทั้งหมดในชุดข้อมูลที่ใช้ฝึกฝน μ_{av} และจะใช้ค่านี้แทนค่า 0.5 ในสมการที่ 2.23 จะได้เป็นสมการที่ 4.2

$$\tilde{\mu}_{INT(k)}(i) = \begin{cases} 2[\mu_k(i)]^2 & , 0 \le \mu_k(i) \le \mu_{av} \\ 1-2[1-\mu_k(i)]^2 & , otherwise \end{cases} \tag{4.2}$$

เมื่อใช้สมการที่ 4.2 มาปรับเปลี่ยนค่าข้อมูลออกที่ต้องการจะได้ค่าดังแสดงในรูปที่ 4.2 ซึ่งจะเห็นว่ามีค่าใกล้เคียงกับกรณีข้อมูลออกแบบเลขฐานสอง (Binary) มาก ดังนั้นเพื่อลดระยะเวลาในการคำนวณ ในงานวิจัยนี้จึงได้เสนอการทดลองที่ใช้ข้อมูลออกที่ต้องการเพียง 2 แบบเท่านั้นคือแบบค่าสมาชิกภาพของแต่ละคำศัพท์ $\mu_k(i)$ และแบบเลขฐานสอง เท่านั้น

4.4 ผลการทดลองการรู้จำและการวิเคราะห์

นำค่าพารามิเตอร์ต่างๆ ที่ได้จากหัวข้อ 4.2 มาใช้ในการทดลองเพื่อวัดประสิทธิภาพของระบบรู้จำสำหรับกลุ่มคำศัพท์ต่างๆ ในตารางที่ ก.5

4.4.1 การรู้จำเสียงตัวเลข (กลุ่มคำศัพท์ที่ 1)

ในกลุ่มคำศัพท์ชุดนี้ จะทำการทดลองเพื่อเปรียบเทียบวิธีการหลายๆ วิธี และจะเป็นแนวทางในการคัดเลือกวิธีการที่ดีที่สุด เพื่อนำไปใช้ในการรู้จำกลุ่มคำศัพท์อื่นๆ โดยมีรายละเอียดของการทดลอง และค่าพารามิเตอร์ที่ใช้ในการทดลองดังนี้

1) ทำการฝึกฝนนิวรอลเน็ตเวอร์กด้วยเสียงของผู้พูดในชุดฝึกฝนจำนวน 10, 20, 30, 40 และ 50 คน เพื่อวิเคราะห์แนวโน้มของอัตราการรู้จำเมื่อเพิ่มจำนวนผู้พูดในการฝึกฝน

2) ทดสอบการรู้จำด้วยชุดทดสอบ 3 ชุดคือ ชุดฝึกฝนที่ใช้ในการฝึกฝน ชุดทดสอบที่ 1 (ชุดทดสอบแบบขึ้นกับผู้พูด) และชุดทดสอบที่ 2 (ชุดทดสอบแบบไม่ขึ้นกับผู้พูด)

3) ใช้ข้อมูลเข้า 2 แบบคือ แบบค่าสัมประสิทธิ์ LPC และแบบค่าสมาชิกภาพแบบฟัซซีโดยเปลี่ยนแปลงชนิดของค่าสมาชิกภาพแบบฟัซซี 3 ชนิดคือ ชนิดสี่เหลี่ยมคางหมู (Tp) ชนิดสามเหลี่ยม (Tg) และชนิด Pi (Pi)

4) เปลี่ยนแปลงค่าข้อมูลออกที่ต้องการ 2 แบบคือ แบบเลขฐานสอง และแบบค่าสมาชิกภาพของแต่ละคำศัพท์

5) พารามิเตอร์อื่นๆ ได้แก่ $\varepsilon = 0.1, \alpha = 0.9, fdenom = 1.0, f_d = 5, f_e = 1$, $E_t = 0.01$ สำหรับข้อมูลออกแบบเลขฐานสอง และ 0.000001 สำหรับข้อมูลออกแบบค่าสมาชิกภาพของแต่ละคำศัพท์ จำนวนโหนดในระดับชั้นข้อมูลเข้าเท่ากับ 460 และ 1380 ในกรณีข้อมูลเข้าแบบสัมประสิทธิ์ LPC และแบบค่าสมาชิกภาพแบบฟัซซี ตามลำดับ จำนวนโหนดในระดับชั้นซ่อนตัวเท่ากับ 70 และ 100 ในกรณีข้อมูลเข้าแบบสัมประสิทธิ์ LPC และแบบค่าสมาชิกภาพแบบฟัซซี ตามลำดับ จำนวนโหนดในระดับชั้นข้อมูลออกเท่ากับ 10 (เท่ากับจำนวนคำศัพท์ในกลุ่มนี้)

การทดลองแรกจะทำการเปรียบเทียบผลอัตราการรู้จำในกรณีที่ได้ตั้งสมมติฐานไว้ในหัวข้อที่ 3.4 คือ ในงานวิจัยนี้ การใช้ข้อมูลออกที่ต้องการแบบเลขฐานสอง น่าจะให้อัตราการรู้จำสูงกว่าข้อมูลออกแบบค่าสมาชิกภาพของแต่ละคำศัพท์ โดยการทดลองจะใช้ข้อมูลเข้าแบบค่าสมาชิกภาพแบบฟัซซีชนิดสี่เหลี่ยมคางหมู เนื่องจากใช้เวลาในการฝึกฝนน้อยที่สุดเมื่อเทียบกับกับชนิดอื่นๆ ผลการทดลองในกรณีใช้ข้อมูลออกแบบเลขฐานสอง และแบบค่าสมาชิกภาพของแต่ละคำศัพท์ แสดงในตารางที่ 4.9 และ 4.10 ซึ่งแสดงเป็นกราฟได้ดังรูปที่ 4.4(ก) และ 4.4(ข) ตามลำดับ

จากผลที่แสดงในกราฟรูปที่ 4.4 เป็นการยืนยันว่า การใช้ข้อมูลออกแบบเลขฐานสองจะให้อัตราการรู้จำสูงกว่ามาก ทั้งยังใช้เวลาในการฝึกฝนน้อยกว่ามากด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

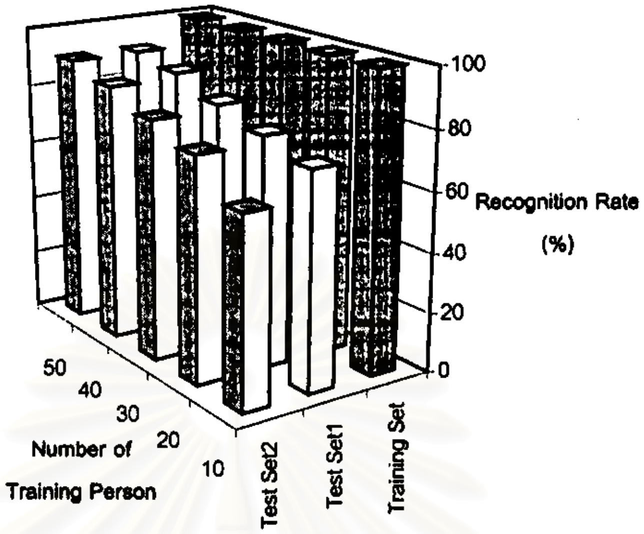
ตารางที่ 4.9 อัตราการรู้จำในกรณีใช้ข้อมูลออกที่ต้องการแบบเลขฐานสอง

จำนวนผู้พูดที่ใช้ฝึกฝน	อัตราการรู้จำ (%)		
	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
10	100	72.7	63.7
20	99.8	77.5	75.3
30	99.7	82.3	80.7
40	99.5	88.0	86.5
50	99.5	91.0	90.8

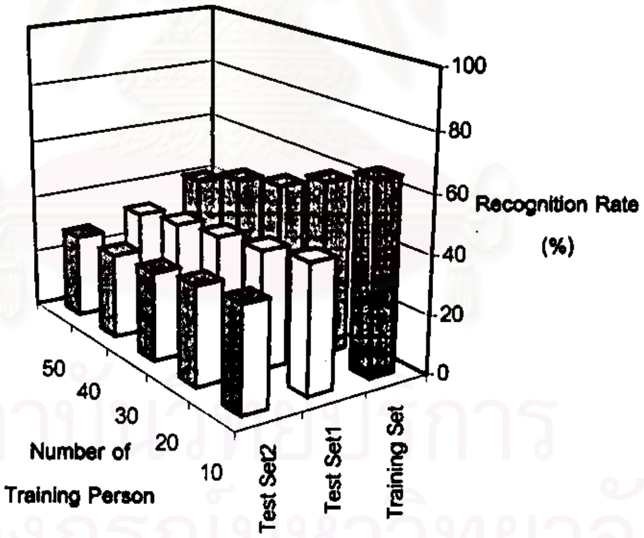
ตารางที่ 4.10 อัตราการรู้จำในกรณีใช้ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์

จำนวนผู้พูดที่ใช้ฝึกฝน	อัตราการรู้จำ (%)		
	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
10	66.0	44.0	36.0
20	59.0	40.5	34.0
30	51.5	38.0	30.5
40	47.5	35.5	29.0
50	41.0	32.5	28.5

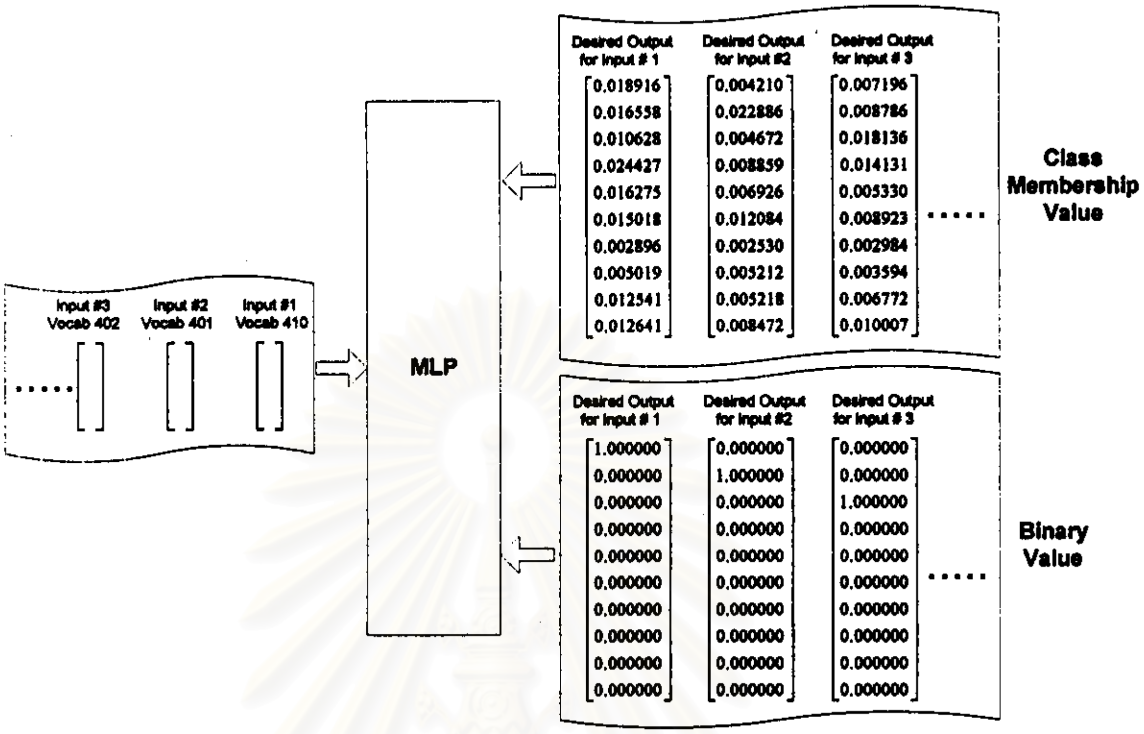
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.4(ก) กราฟแสดงอัตราการรู้จำเมื่อใช้ข้อมูลออกที่ต้องการแบบเลขฐานสอง



รูปที่ 4.4(ข) กราฟแสดงอัตราการรู้จำเมื่อใช้ข้อมูลออกที่ต้องการ
แบบค่าสมาชิกภาพของแต่ละคำศัพท์



รูปที่ 4.5 ตัวอย่างเวกเตอร์ข้อมูลออกที่ต้องการ
เพื่อจากวิเคราะห์ผลการทดลองในกราฟรูปที่ 4.4

เมื่อพิจารณารูปที่ 4.5 ซึ่งเป็นตัวอย่างของเวกเตอร์ข้อมูลออกที่ต้องการแบบเลขฐานสอง และแบบค่าสมาชิกของแต่ละคำศัพท์ ที่ใช้ในการทดลองนี้ จะวิเคราะห์ได้ถึงสาเหตุที่ได้ผลการทดลองดังกราฟรูปที่ 4.4 ดังนี้

1) การทดลองนี้ กำหนดให้คำศัพท์หมายเลข 410, 401, ..., 409 สอดคล้องกับโหนดที่ 1, 2, ..., 10 ในระดับชั้นข้อมูลออก ดังนั้นเวกเตอร์ข้อมูลออกที่ต้องการควรจะต้องมีค่าสูงสุด ณ ตำแหน่งที่สอดคล้องกับคำที่เข้ามาฝึกฝนขณะนั้น แต่เมื่อพิจารณารูปที่ 4.5 เวกเตอร์ข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ ซึ่งตรงกับคำฝึกฝนคำแรก (คำศัพท์หมายเลข 410) ควรจะต้องมีค่าที่โหนดแรกสูงที่สุด แต่ปรากฏว่ากลับมีค่าที่ตำแหน่งโหนดที่ 4 สูงที่สุด สาเหตุที่เป็นเช่นนี้เนื่องจาก เสียงพูดคำศัพท์หมายเลข 410 (เสียงเลขศูนย์) ของผู้พูดคนนี้ อาจจะค่อนข้างแตกต่างกับเสียงของผู้อื่น คือมีลักษณะใกล้เคียงกับเสียงคำศัพท์หมายเลข 403 (เสียงเลขสาม) มากกว่า รูปที่ 4.6 แสดงตัวอย่างรูปคลื่นของเสียงคำศัพท์หมายเลข 410 ที่ผิดพลาดไปคล้ายกับเสียงคำศัพท์หมายเลข 403

Vocab 410

Vocab 403

รูปที่ 4.6 ตัวอย่างรูปคลื่นของคำศัพท์หมายเลข 410 ซึ่งไปคล้ายกับคำศัพท์หมายเลข 403

ลักษณะที่เกิดขึ้นเช่นนี้ ทำให้มีการฝึกฝนให้จำแบบผิดๆ คือเป็นการฝึกฝนให้นิวรอลเน็ตเวอร์ก รู้จำคำศัพท์หมายเลข 410 นี้ เป็นคำศัพท์หมายเลข 403 เสมอ ดังนั้นเมื่อทดสอบการรู้จำย่อมได้ผลการรู้จำผิดอยู่เสมอ ดังผลการทดลองในตารางที่ 4.10 ในกรณีที่ทดสอบการรู้จำด้วยชุดที่ใช้ฝึกฝนจำนวน 10 คน ได้ผลอัตราการรู้จำเพียง 66 เปอร์เซ็นต์ และลักษณะเช่นนี้จะเกิดมากขึ้นได้ เมื่อเพิ่มจำนวนผู้พูดที่ใช้ฝึกฝน เนื่องจากเป็นการเพิ่มความแปรปรวนของข้อมูล ส่งผลให้ระยะห่าง z_k และค่า $\mu_k(i)$ ในสมการที่ 2.21 และ 2.22 ไม่ถูกต้องตามที่ควรจะเป็น ดังนั้นยิ่งเพิ่มจำนวนผู้พูดในการฝึกฝน จะยิ่งเป็นการลดความสามารถในการรู้จำของนิวรอลเน็ตเวอร์ก

2) เมื่อพิจารณาข้อมูลออกที่ต้องการแบบค่าสมาชิกภาพของแต่ละคำศัพท์ในรูปที่ 4.5 แม้ว่าเวกเตอร์ข้อมูลออกที่ต้องการส่วนใหญ่จะมีค่าข้อมูลออกที่ต้องการ สอดคล้องกับคำที่ฝึกฝนขณะนั้น แต่ยังเกิดกรณีกำกวมที่สุด (Fuzziest Case) ขึ้นได้อีกคือ มีค่าใกล้เคียงกัน ไม่แยกแยะกันเด่นชัดเท่าที่ควร ผลคือหลังจากการฝึกฝน นิวรอลเน็ตเวอร์กจะให้ค่าข้อมูลออกในขณะทดสอบกำกวม หรือผิดพลาดไปด้วย อีกทั้งการฝึกฝนนิวรอลเน็ตเวอร์กทำได้ยาก ใช้เวลานานกว่าการใช้ข้อมูลออกที่ต้องการแบบเลขฐานสองมาก ในบางครั้งอาจจะกำกวมเกินไปจนไม่สามารถฝึกฝนจนสำเร็จได้ วิธีการแก้ไขอย่างหนึ่งคือ การใช้สมการที่ 4.2 มาปรับค่าข้อมูลออกที่ต้องการใหม่ แต่ผลของค่าใหม่ที่ได้แทบจะเท่ากับกรณีที่ใช้ข้อมูลออกแบบเลขฐานสอง ดังที่แสดงให้เห็นในหัวข้อ 4.3

3) เมื่อใช้ข้อมูลออกแบบเลขฐานสอง แม้ว่าคำศัพท์ที่เข้ามาฝึกฝนจะมีลักษณะแตกต่างไปจากคำศัพท์เดียวกันที่ได้จากผู้พูดส่วนใหญ่ แต่การฝึกฝนก็ยังคงพยายามปรับค่าน้ำหนัก

การเชื่อมต่อเพื่อให้ได้ค่าข้อมูลออกสอดคล้องกับคำศัพท์นี้อยู่เสมอ ผลคือทำให้ได้อัตราการรู้จำสูงสุดถึง 90.8 เปอร์เซ็นต์ เมื่อใช้ผู้พูดจำนวน 50 คนในการฝึกฝน

การทดลองต่อไป จะทำการเปรียบเทียบผลอัตราการรู้จำ ในกรณีที่ใช้ข้อมูลเข้าแบบค่าสัมประสิทธิ์ LPC และค่าสมาชิกภาพแบบฟัซซีทั้ง 3 ชนิดคือ สี่เหลี่ยมคางหมู (Tp) สามเหลี่ยม (Tg) และ Pi โดยใช้ข้อมูลออกที่ต้องการแบบเลขฐานสองเท่านั้น ผลการทดลองในกรณีข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิดสี่เหลี่ยมคางหมูคือตารางที่ 4.9 ที่ผ่านมา ส่วนในกรณีข้อมูลเข้าแบบเลขฐานสอง ค่าสมาชิกภาพแบบฟัซซีชนิดสามเหลี่ยม และ Pi แสดงในตารางที่ 4.11, 4.12 และ 4.13 ตามลำดับ

ผลการทดลองนำมาแสดงเป็นกราฟ โดยแยกเปรียบเทียบสำหรับชุดทดสอบที่ 1 และชุดทดสอบที่ 2 ได้ดังรูปที่ 4.7(ก) และ 4.7(ข) ตามลำดับ

ตารางที่ 4.11 อัตราการรู้จำในกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC

จำนวนผู้พูดที่ใช้ฝึกฝน	อัตราการรู้จำ (%)		
	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
10	100	72.3	56.0
20	100	76.7	68.0
30	100	78.6	73.0
40	100	83.8	80.3
50	100	87.0	86.0

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

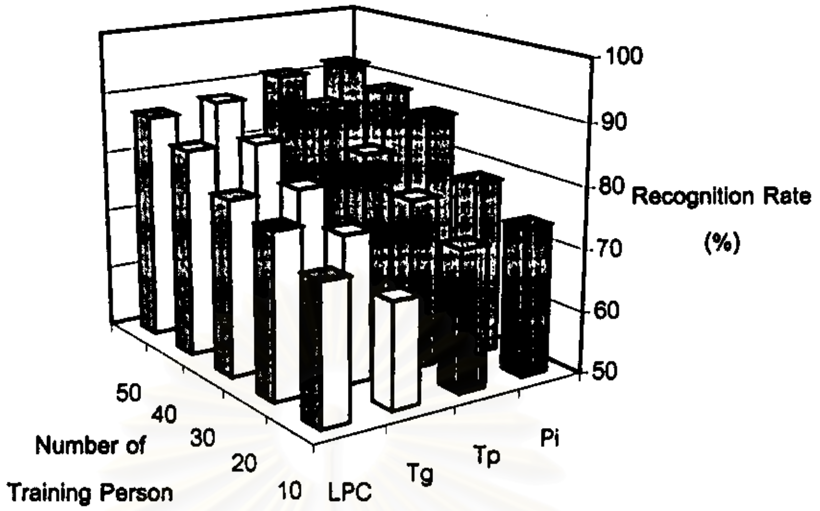
ตารางที่ 4.12 อัตราการรู้จำในกรณีที่ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิดสามเหลี่ยม (Tg)

จำนวนผู้พูดที่ใช้ฝึกฝน	อัตราการรู้จำ (%)		
	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
10	100	67.0	62.0
20	100	74.0	70.0
30	99.7	78.5	75.5
40	99.5	83.3	83.0
50	99.3	88.0	87.7

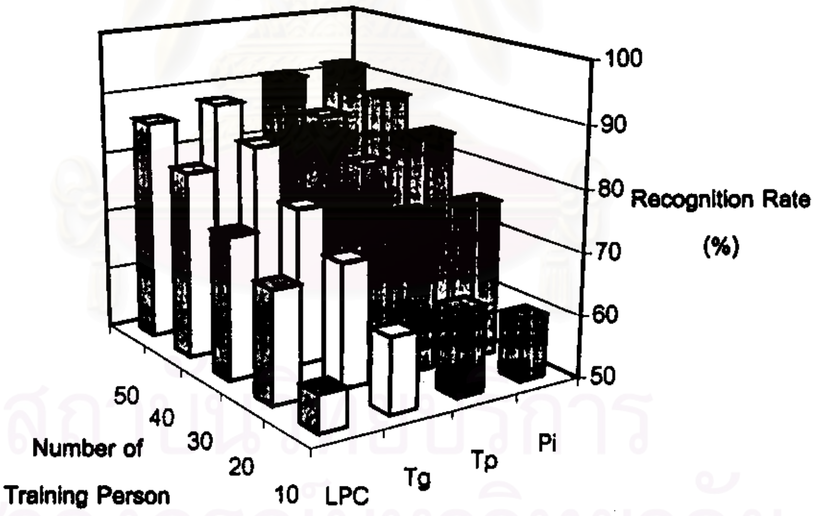
ตารางที่ 4.13 อัตราการรู้จำในกรณีที่ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi

จำนวนผู้พูดที่ใช้ฝึกฝน	อัตราการรู้จำ (%)		
	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
10	100	74.0	60.0
20	99.8	78.5	75.3
30	100	87.0	84.0
40	99.5	89.0	88.3
50	99.5	91.7	91.5

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.7(ก) กราฟเปรียบเทียบอัตราการรู้จำจากชุดทดสอบที่ 1 (ขึ้นต่อผู้พูด)

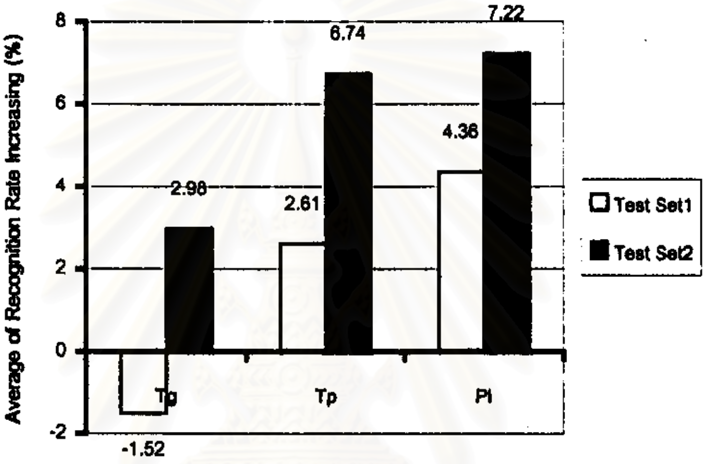


สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.7(ข) กราฟเปรียบเทียบอัตราการรู้จำจากชุดทดสอบที่ 2 (ไม่ขึ้นต่อผู้พูด)

จากผลการทดลองมีจุดที่น่าสนใจ ซึ่งจะวิเคราะห์ได้ดังนี้

1) ผลอัตราการรู้จำจากชุดทดสอบที่ 1 ซึ่งเป็นชุดที่ขึ้นต่อผู้พูด จะสูงกว่าผลอัตราการรู้จำจากชุดทดสอบที่ 2 ซึ่งไม่ขึ้นต่อผู้พูด โดยจะมีอัตราการรู้จำเพิ่มขึ้นเมื่อเพิ่มจำนวนผู้พูดในชุดฝึกฝน โดยเฉพาะอย่างยิ่ง อัตราการรู้จำจากชุดทดสอบที่ 2 จะเพิ่มขึ้นอย่างมากเมื่อเพิ่มผู้พูดในชุดฝึกฝนจาก 10 คนเป็น 20 คน อัตราการรู้จำจากชุดทดสอบที่ 2 สูงที่สุดถึง 91.5 เปอร์เซ็นต์ ในกรณีใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi และใช้ผู้พูดถึง 50 คนในการฝึกฝน



รูปที่ 4.8 กราฟแสดงผลอัตราการรู้จำที่เพิ่มขึ้นเฉลี่ย เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิดต่างๆ เทียบกับเมื่อใช้ข้อมูลเข้าแบบค่าสัมประสิทธิ์ LPC

2) การใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี แทนข้อมูลเข้าแบบค่าสัมประสิทธิ์ LPC จะช่วยเพิ่มอัตราการรู้จำของระบบได้อย่างเด่นชัด ทั้งนี้ขึ้นอยู่กับชนิดของฟังก์ชันสมาชิกภาพแบบฟัซซี อัตราการรู้จำเมื่อใช้ข้อมูลเข้าแบบค่าสมาชิกภาพแบบฟัซซีแต่ละชนิด เมื่อเทียบกับใช้ข้อมูลแบบค่าสัมประสิทธิ์ LPC จะเพิ่มขึ้นโดยเฉลี่ยดังแสดงในกราฟรูปที่ 4.8 จะเห็นว่าฟังก์ชันสมาชิกภาพแบบฟัซซีชนิด Pi จะให้อัตราการรู้จำเพิ่มขึ้นมากที่สุด ส่วนชนิดสี่เหลี่ยมคางหมู และชนิดสามเหลี่ยมจะลดลงมาตามลำดับ สาเหตุที่เป็นเช่นนี้ เนื่องจากธรรมชาติของเสียงพูดจะมีลักษณะแปรเปลี่ยนอยู่เสมอ และเมื่อสกัดค่าลักษณะเด่นแบบสัมประสิทธิ์ LPC ก็จะมีค่าแปรเปลี่ยนอยู่เสมอด้วย และเพื่อให้ได้ค่าลักษณะเด่นแบบฟัซซีที่ครอบคลุมการแปรเปลี่ยน ซึ่งทำให้เกิดความกำกวมเหล่านี้ ฟังก์ชันสมาชิกภาพแบบฟัซซีที่ใช้ก็ควรจะให้ค่าสมาชิกภาพแบบฟัซซีเปลี่ยนแปลงไปอยู่เสมอ เมื่อค่าสัมประสิทธิ์ LPC มีการเปลี่ยนแปลง ซึ่งลักษณะเช่นนี้เกิดขึ้นได้เมื่อใช้ฟังก์ชันชนิด Pi และชนิดสามเหลี่ยม แต่จะไม่เกิดขึ้นในบางช่วงของฟังก์ชันชนิดสี่เหลี่ยม

คางหมู อย่างไรก็ตาม ค่าสมาชิกภาพแบบฟัซซีที่ได้ก็ไม่ควรเปลี่ยนแปลงมากเกินไป เมื่อค่าสัมประสิทธิ์ LPC เปลี่ยนแปลงไปเล็กน้อย ซึ่งฟังก์ชันชนิด Pi เหมาะสมที่สุด ส่วนฟังก์ชันชนิดสามเหลี่ยมจะให้ค่าสมาชิกภาพแบบฟัซซีเปลี่ยนแปลงมากเกินไป ผลคือข้อมูลใหม่ที่ได้จะมีความกำกวมมากเกินไป ทำให้ในบางกรณี อัตราการรู้จำที่ได้จะต่ำกว่าข้อมูลเข้าแบบสัมประสิทธิ์ LPC ได้ ดังเช่นจากรูปที่ 4.8 ในกรณีอัตราการรู้จำจากชุดทดสอบที่ 1

ตารางที่ 4.14, 4.15, 4.16 และ 4.17 แสดงผลการรู้จำจากชุดทดสอบที่ 2 ในกรณีที่ใช้ผู้พูดจำนวน 50 คนในการฝึกฝน โดยใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC และค่าสมาชิกภาพแบบฟัซซีชนิดสามเหลี่ยม สี่เหลี่ยมคางหมู และ Pi ตามลำดับ

จากตารางที่ 4.14 จะพบว่า คำในชุดทดสอบที่ระบบรู้จำผิดพลาดมากคือ เสียงเลขสอง และเสียงเลขสาม โดยผิดเป็นเสียงเลขสาม เลขสอง และเลขศูนย์ เนื่องจากเสียงคำดังกล่าวมีความกำกวมของเสียงพยัญชนะต้นซึ่งเป็นตัวเดียวกันคือ “ส” ทั้งยังมีเสียงวรรณยุกต์จัตวาเหมือนกัน นอกจากนี้ยังมีเสียงเลขเจ็ดที่ผิดมาก โดยผิดเป็นเสียงเลขหก และเลขหนึ่ง เนื่องจากเป็นความกำกวมของเสียงที่มีสระเสียงสั้น ทำให้พูดให้ชัดเจนได้ยากกว่าสระเสียงยาว แต่เมื่อพิจารณาตารางที่ 4.15-4.17 การใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีโดยเฉพาะชนิดสี่เหลี่ยมคางหมู และชนิด Pi จะเป็นการเพิ่มความละเอียดในการพิจารณาเพื่อตัดสินใจ ซึ่งจะช่วยแก้ปัญหาตรงจุดดังกล่าวได้ รูปที่ 4.9(ก) แสดงตัวอย่างเสียงเลขสองที่วิเคราะห์ผิดเป็นเลขสาม และรูปที่ 4.9(ข) แสดงตัวอย่างเสียงเลขเจ็ดที่วิเคราะห์ผิดเป็นเลขหก แต่ทั้งสองรูป จะวิเคราะห์ได้ถูกต้องเมื่อเปลี่ยนมาใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.14 ตัวอย่างผลการรู้จำเสียงจำนวน 100 เสียงจากชุดทดสอบที่ 2 เมื่อใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC โดยใช้ผู้พูดจำนวน 50 คนในการฝึกฝน

คำในชุดทดสอบที่ 2	คำศัพท์ที่วิเคราะห์ได้									
	410	401	402	403	404	405	406	407	408	409
410	9	0	1	0	0	0	0	0	0	0
401	0	10	0	0	0	0	0	0	0	0
402	2	0	6	2	0	0	0	0	0	0
403	0	0	1	8	0	1	0	0	0	0
404	1	0	0	0	9	0	0	0	0	0
405	0	0	0	0	0	9	0	0	0	1
406	0	0	0	0	1	0	9	0	0	0
407	0	1	0	0	0	0	1	8	0	0
408	0	0	0	0	0	1	0	0	9	0
409	0	0	0	1	0	0	0	0	0	9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.15 ตัวอย่างผลการรู้จำเสียงจำนวน 100 เสียงจากชุดทดสอบที่ 2 เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิดสามเหลี่ยม โดยใช้ผู้พูดจำนวน 50 คนในการฝึกฝน

คำในชุดทดสอบที่ 2	คำศัพท์ที่วิเคราะห์ได้									
	410	401	402	403	404	405	406	407	408	409
410	9	0	0	1	0	0	0	0	0	0
401	0	9	0	0	0	0	1	0	0	0
402	1	0	7	1	0	0	0	0	0	1
403	0	0	1	9	0	0	0	0	0	0
404	0	0	1	0	9	0	0	0	0	0
405	0	1	0	0	0	9	0	0	0	0
406	0	0	0	0	0	0	10	0	0	0
407	0	1	0	0	0	0	0	9	0	0
408	0	0	0	1	0	1	0	0	8	0
409	0	0	0	0	0	1	0	0	0	9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.16 ตัวอย่างผลการรู้จำเสียงจำนวน 100 เสียงจากชุดทดสอบที่ 2 เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิดสี่เหลี่ยมคางหมู โดยใช้ผู้พูดจำนวน 50 คนในการฝึกฝน

คำในชุดทดสอบที่ 2	คำศัพท์ที่วิเคราะห์ได้									
	410	401	402	403	404	405	406	407	408	409
410	9	0	1	0	0	0	0	0	0	0
401	0	9	0	0	0	0	1	0	0	0
402	0	0	9	1	0	0	0	0	0	0
403	1	0	0	9	0	0	0	0	0	0
404	0	0	0	1	9	0	0	0	0	0
405	0	0	0	0	0	9	0	0	0	1
406	0	1	0	0	0	0	9	0	0	0
407	0	0	0	0	1	0	0	9	0	0
408	0	0	0	1	0	0	0	0	9	0
409	0	0	0	1	0	0	0	0	0	9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.17 ตัวอย่างผลการรู้จำเสียงจำนวน 100 เสียงจากชุดทดสอบที่ 2 เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi โดยใช้ผู้พูดจำนวน 50 คนในการฝึกฝน

คำในชุดทดสอบที่ 2	คำศัพท์ที่วิเคราะห์ได้									
	410	401	402	403	404	405	406	407	408	409
410	9	0	1	0	0	0	0	0	0	0
401	0	10	0	0	0	0	0	0	0	0
402	0	0	9	1	0	0	0	0	0	0
403	0	0	1	9	0	0	0	0	0	0
404	0	0	0	0	10	0	0	0	0	0
405	0	0	0	1	0	9	0	0	0	0
406	0	0	0	0	1	0	9	0	0	0
407	0	0	0	0	0	0	1	9	0	0
408	0	0	0	0	0	1	0	0	9	0
409	0	0	0	0	0	1	0	0	0	9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.9(ก) ตัวอย่างรูปคลื่นของเสียงตัวเลขสอง ที่วิเคราะห์ผิดเป็นเลขสาม
แต่วิเคราะห์ได้ถูกต้องเมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี

รูปที่ 4.9(ข) ตัวอย่างรูปคลื่นของเสียงตัวเลขเจ็ด ที่วิเคราะห์ผิดเป็นเลขหก
แต่วิเคราะห์ได้ถูกต้องเมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี

3) จากตารางที่ 4.11-4.13 การใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี จะทำให้ผลอัตราการรู้จำจากชุดฝึกฝนลดลงเล็กน้อย กล่าวได้ว่านิวรอลเน็ตเวอร์กมีความสามารถในการฝึกฝนข้อมูลเข้าประเภทนี้ต่ำลง เนื่องจากข้อมูลเข้าที่เป็นค่าสมาชิกภาพแบบฟัซซีมีความละเอียดมากกว่าข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC มาก แต่ในทางกลับกันจะเป็นการเพิ่มความสามารถในการพิจารณารายละเอียดในการรู้จำชุดทดสอบอื่นๆ อย่างไรก็ตาม การฝึกฝนนิวรอลเน็ตเวอร์กโดยใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีจะต้องใช้เวลานานกว่าข้อมูลเข้าแบบค่าสัมประสิทธิ์ LPC มาก โดยที่กรณีใช้ฟังก์ชันสมาชิกภาพชนิดสามเหลี่ยมจะใช้เวลานานที่สุด ส่วนชนิด Pi จะใช้เวลาน้อยกว่ามาก และชนิดสี่เหลี่ยมคางหมูจะใช้เวลาน้อยที่สุด สาเหตุที่เป็นเช่นนี้เพราะกรณีใช้ฟังก์ชันสมาชิกภาพชนิดสามเหลี่ยม และ Pi จะทำให้ข้อมูลเข้ามีค่าไม่เท่ากับ 0 หรือ 1 แทบทุกค่า ในขณะที่ชนิดสี่เหลี่ยมคางหมูมีโอกาสเกิดขึ้นมาก ผลคือทำให้การคำนวณเพื่อฝึกฝนง่ายกว่า และชนิดสามเหลี่ยมจะใช้เวลานานมาก เนื่องจากมีความกำกวมสูงที่สุดดังที่ได้กล่าวมาแล้ว

4.4.2 การรู้จำเสียงคำศัพท์กลุ่มอื่นๆ

จากผลการวิเคราะห์ในหัวข้อที่ 4.2, 4.3 และ 4.4.1 วิธีการที่เหมาะสมที่สุดในการทดลองเพื่อรู้จำเสียงคำศัพท์กลุ่มอื่นๆ คือ ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi และข้อมูลออกที่ต้องการแบบเลขฐานสอง โดยใช้ผู้พูดจำนวน 50 คนในการฝึกฝน ส่วนพารามิเตอร์อื่นๆ คือ $\varepsilon = 0.1, \alpha = 0.9, fdenom = 1.0$, $E_t = 0.01$ และจำนวนโหนดในระดับชั้นต่างๆ ขึ้นอยู่กับกลุ่มคำศัพท์แต่ละกลุ่มดังที่กล่าวมาแล้ว

ดังที่ได้กล่าวมาแล้วในหัวข้อที่ 3.6 เพื่อเพิ่มประสิทธิภาพของระบบรู้จำ จะนำคำศัพท์ทั้ง 70 คำมาแบ่งกลุ่มเบื้องต้นด้วยวิธีการตรวจสอบจำนวนพยางค์ ซึ่งได้กลุ่มคำศัพท์ย่อย 3 กลุ่มคือ กลุ่มคำศัพท์ที่ 3, 4 และ 5 ดังแสดงในตารางที่ ก.5 ดังนั้นจะทำการทดลองเพื่อรู้จำกลุ่มคำศัพท์ดังกล่าว ยกเว้นการทดลองสำหรับกลุ่มคำศัพท์ที่ 3 กรณีที่ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซี จะไม่ทำการทดลอง เพราะใช้เวลาในการฝึกฝนนานเกินไป นอกจากนี้จะทำการทดลองเพื่อรู้จำกลุ่มคำศัพท์ที่ 2 เพื่อใช้เปรียบเทียบกรณีกลุ่มคำศัพท์หนึ่งพยางค์ที่ไม่ใช่ตัวเลข ผลการทดลองเปรียบเทียบกับกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC แสดงในตารางที่ 4.18

ตารางที่ 4.18 ผลอัตราการรู้จำสำหรับกลุ่มคำศัพท์ที่ 2-5 เมื่อใช้ผู้พูด 50 คนในการฝึกฝน

กลุ่มคำศัพท์ที่	จำนวนคำศัพท์ในกลุ่ม	อัตราการรู้จำ (%)					
		ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC			ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi		
		ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
2	20	100	82.9	79.0	98.4	85.9	83.2
3	30	99.1	78.8	72.5	N.A.		
4	20	100	87.2	84.3	99.2	90.1	88.0
5	20	100	87.6	80.2	98.9	89.7	87.5
เฉลี่ย	70	99.7	83.5	78.3	98.8	88.6	86.2

จากผลการทดลอง โดยอาศัยสมการที่ 3.12 และ 3.13 จะได้อัตราการรู้จำเฉลี่ยของระบบกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC สำหรับชุดทดสอบที่ 1 และชุดทดสอบที่ 2 เท่า

กับ 83.5 และ 78.3 เปอร์เซ็นต์ ตามลำดับ ส่วนกรณีที่ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi จะได้เท่ากับ 88.6 และ 86.2 เปอร์เซ็นต์ ตามลำดับ นั่นคือได้อัตราการรู้จำเฉลี่ยสำหรับชุดทดสอบที่ 1 และชุดทดสอบที่ 2 เพิ่มขึ้น 5.1 และ 7.9 เปอร์เซ็นต์ ตามลำดับ ซึ่งนับว่าใกล้เคียงกับกรณีกลุ่มคำศัพท์ที่ 1 ซึ่งเป็นตัวเลขอย่างเดียว

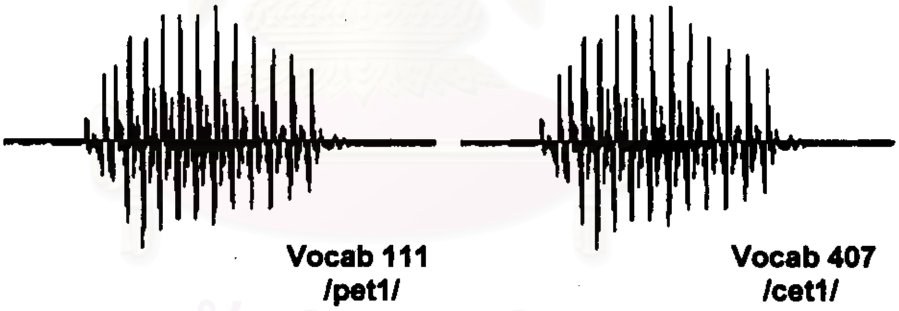
เมื่อพิจารณาผลการทดลองสำหรับกลุ่มคำศัพท์ที่ 2 และ 3 ในตารางที่ 4.18 และผลการทดลองสำหรับกลุ่มคำศัพท์ที่ 1 ในตารางที่ 4.11-4.13 จะพบว่าเมื่อเพิ่มจำนวนคำศัพท์ในการรู้จำจาก 10 เป็น 20 และ 30 คำศัพท์ สำหรับนิวรอลเน็ตเวอร์กหนึ่งๆ อัตราการรู้จำจะลดลงอย่างมาก นอกจากนี้ผลอัตราการรู้จำสำหรับกลุ่มคำศัพท์ที่มีจำนวนคำศัพท์ที่จะรู้จำเท่ากันยังขึ้นอยู่กับคำศัพท์ในกลุ่มนั้นๆ ด้วย ดังเช่นผลการทดลองสำหรับกลุ่มคำศัพท์ที่ 2, 4 และ 5 ในตารางที่ 4.18 จะเห็นว่ากลุ่มคำศัพท์ที่ 4 และ 5 ซึ่งประกอบด้วยคำสองพยางค์ และสามพยางค์ตามลำดับ จะมีอัตราการรู้จำสูงกว่ากรณีกลุ่มคำศัพท์ที่ 2 ซึ่งประกอบด้วยคำหนึ่งพยางค์ เนื่องจากคำหลายพยางค์ ย่อมมีโอกาสที่จะมีลักษณะแตกต่างกันมากกว่าคำหนึ่งพยางค์ แต่อย่างไรก็ตาม การพูดคำสามพยางค์ ผู้พูดแต่ละคนสามารถพูดให้แตกต่างกันได้มากกว่าคำสองพยางค์ ดังนั้น ผลการทดลองสำหรับกลุ่มคำศัพท์ที่ 4 ซึ่งประกอบด้วยคำสองพยางค์จึงให้อัตราการรู้จำสูงที่สุด

เพื่อเพิ่มประสิทธิภาพของการรู้จำ จะทำการแบ่งกลุ่มคำศัพท์ให้เล็กลงไปอีก โดยอาศัยวิธีการตรวจสอบเสียงวรรณยุกต์ของพยางค์แรกของคำศัพท์ ซึ่งจะแบ่งได้เป็นกลุ่มคำศัพท์ที่ 6-19 ดังที่แสดงในตารางที่ ก.5 และทำการทดลองใช้นิวรอลเน็ตเวอร์กย่อย เพื่อรู้จำคำศัพท์ในกลุ่มคำศัพท์แต่ละกลุ่มทำนองเดียวกับการทดลองสำหรับกลุ่มคำศัพท์ที่ 2-5 ยกเว้นกลุ่มคำศัพท์ที่ 17 และ 19 ซึ่งประกอบด้วยคำศัพท์เพียงคำเดียว ผลการทดลองทุกกลุ่มคำศัพท์แสดงในตารางที่ 4.19 โดยที่ค่าเฉลี่ยในตารางที่ 4.19 คิดเฉพาะกลุ่มที่ 6-19 ซึ่งเป็นการแบ่งกลุ่มย่อยด้วยจำนวนพยางค์และระดับเสียงของพยางค์แรก

จากผลการทดลอง โดยอาศัยสมการที่ 3.13 อัตราการรู้จำเฉลี่ยสำหรับชุดทดสอบที่ 1 และชุดทดสอบที่ 2 ในกรณีที่ใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC เท่ากับ 91.1 และ 89.9 เปอร์เซ็นต์ ตามลำดับ ส่วนในกรณีที่ใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด Pi จะได้เท่ากับ 94.4 และ 93.3 เปอร์เซ็นต์ ตามลำดับ นั่นคืออัตราการรู้จำเฉลี่ยสำหรับชุดทดสอบที่ 1 และชุดทดสอบที่ 2 เพิ่มขึ้น 3.3 และ 3.4 เปอร์เซ็นต์ ตามลำดับ เป็นที่น่าสังเกตว่าอัตราการรู้จำเฉลี่ยที่เพิ่มขึ้นนี้ ลดลงเมื่อเทียบกับการทดลองที่ผ่านมา เนื่องมาจากการแบ่งกลุ่มคำศัพท์ให้เล็กลง จะทำให้อัตราการรู้จำเมื่อใช้ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC จะเพิ่มขึ้นมาก และเมื่ออัตราการรู้จำสูงอยู่แล้ว เมื่อใช้ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีจะช่วยเพิ่มอัตราการรู้จำได้อีกไม่มากนัก

เมื่อพิจารณาผลการทดลอง กลุ่มคำศัพท์ส่วนใหญ่จะมีอัตราการรู้จำสำหรับชุดทดสอบที่ 1 ต่ำกว่าชุดทดสอบที่ 2 ทั้งที่ชุดทดสอบที่ 2 เป็นแบบไม่ขึ้นต่อผู้พูด สาเหตุเนื่องมาจากในการฝึกฝนเพื่อรู้จำโดยใช้ผู้พูด 50 คน จะใช้ชุดทดสอบที่ 1 ประกอบด้วยเสียงของผู้พูดกลุ่มเดียวกับที่ใช้ในชุดฝึกฝนทั้ง 50 คน แต่ชุดทดสอบที่ 2 จะใช้เสียงจากผู้พูดจำนวน 10 คน เมื่อจำนวนคนในชุดทดสอบที่ 1 มีจำนวนมากๆ อาจจะเกิดความแปรปรวนได้มากกว่าชุดทดสอบที่ 2 จึงทำให้อัตราการรู้จำต่ำกว่าได้ นอกจากนี้ยังขึ้นอยู่กับการพูดของแต่ละบุคคลในขณะทำการบันทึกเสียงแต่ละครั้ง

ในบางกรณีของการทดลอง สำหรับกลุ่มคำศัพท์ที่มีจำนวนคำศัพท์ใกล้เคียงกัน เช่น การทดลองสำหรับกลุ่มคำศัพท์ที่ 8 ดังแสดงในตารางที่ 4.19 และการทดลองสำหรับกลุ่มคำศัพท์ที่ 1 ดังแสดงในตารางที่ 4.13 แม้ว่าจำนวนคำศัพท์ในกลุ่มคำศัพท์ที่ 1 จะมีจำนวนมากกว่ากลุ่มคำศัพท์ที่ 8 แต่อัตราการรู้จำสำหรับชุดทดสอบที่ 2 กลับมีค่าใกล้เคียงกัน ทั้งนี้ขึ้นอยู่กับคำศัพท์ในกลุ่มมีความคล้ายคลึงกันเพียงใด ดังเช่นรูปที่ 4.10 แสดงตัวอย่างรูปคลื่นเสียงเลขเจ็ด (คำศัพท์หมายเลข 407) และเสียงคำว่าเป็ด (คำศัพท์หมายเลข 111) ซึ่งมีความคล้ายกันมาก ทำให้ระบบวิเคราะห์ผิดอยู่เสมอ ส่วนในกลุ่มคำศัพท์ที่ 1 ไม่มีคำศัพท์ที่ใกล้เคียงกันขนาดนี้ จึงให้ผลการรู้จำสูงกว่าได้



รูปที่ 4.10 ตัวอย่างรูปคลื่นของคำศัพท์ของคำศัพท์หมายเลข 111
ซึ่งใกล้เคียงกับคำศัพท์หมายเลข 407 และอยู่ในกลุ่มคำศัพท์เดียวกัน

ตารางที่ 4.19 ผลอัตราการรู้จำสำหรับทุกกลุ่มคำศัพท์
ยกเว้นกลุ่มคำศัพท์ที่ 17 และ 19 เมื่อใช้ผู้พูด 50 คนในการฝึกฝน

กลุ่มคำศัพท์ที่	จำนวนคำศัพท์ในกลุ่ม	อัตราการรู้จำ (%)					
		ข้อมูลเข้าเป็นค่าสัมประสิทธิ์ LPC			ข้อมูลเข้าเป็นค่าสมาชิกภาพแบบฟัซซีชนิด PI		
		ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2	ชุดฝึกฝน	ชุดทดสอบที่ 1	ชุดทดสอบที่ 2
1	10	100	87.0	86.0	99.5	91.0	90.8
2	20	100	82.9	79.0	98.4	85.9	83.2
3	30	99.1	78.8	72.5	N.A.		
4	20	100	87.2	84.3	99.2	90.1	88.0
5	20	100	87.6	80.2	98.9	89.7	87.5
6	7	100	87.1	88.1	99.7	92.1	92.5
7	8	100	87.5	84.3	99.7	93.0	91.7
8	7	100	88.8	86.3	99.8	92.4	89.8
9	3	100	95.3	92.2	100	96.1	94.2
10	5	100	87.1	94.0	99.9	93.1	95.5
11	10	100	91.5	89.7	99.5	94.3	93.0
12	3	100	95.6	96.7	100	96.5	97.2
13	2	100	98.0	100	100	100	100
14	5	100	92.9	94.7	99.9	95.0	96.3
15	8	100	90.5	81.3	99.6	93.4	88.1
16	5	100	95.6	100	100	98.4	100
18	5	100	92.8	87.3	100	94.6	90.0
เฉลี่ย	70	100	91.1	89.9	99.8	94.4	93.3