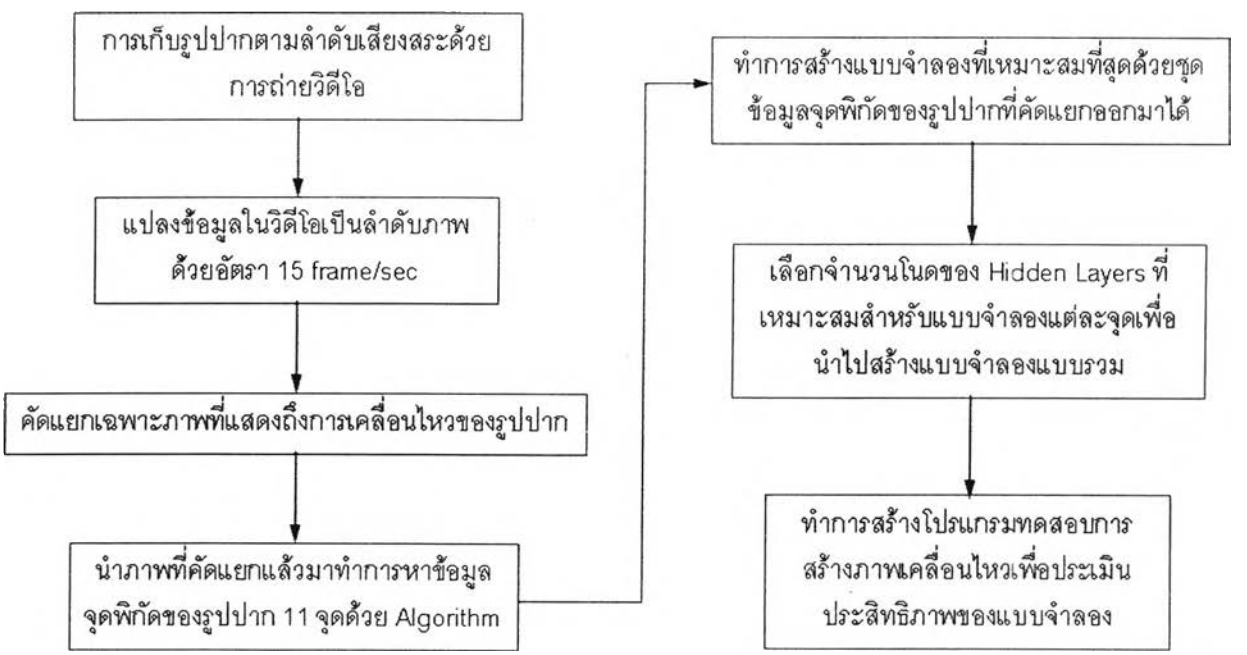
บทที่ 3

การสร้างแบบจำลองการเคลื่อนไหวของรูปปาก

ในบทนี้ จะกล่าวถึงกระบวนการในการสร้างแบบจำลองการเคลื่อนไหวของรูปปากซึ่งจะประกอบไปด้วย ขั้นตอนการเก็บข้อมูล การสร้างแบบจำลองด้วย Backpropagation Neural Network วิธีการสร้างภาพเคลื่อนไหวจากแบบจำลองและการประเมินประสิทธิภาพของแบบจำลอง



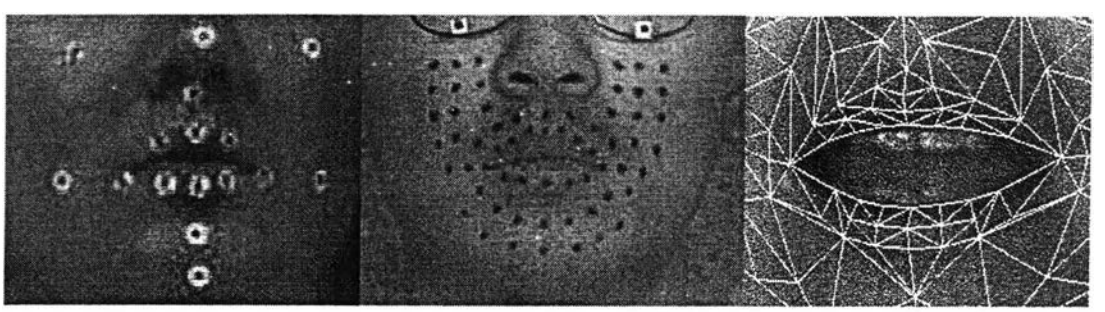
รูปที่ 3.1 แผนผังขั้นตอนการดำเนินการวิจัย

3.1 การเก็บข้อมูล

ในขั้นตอนการเก็บข้อมูลนี้จะประกอบไปด้วยขั้นตอนย่อยต่อไปนี้ การกำหนดจุดรอบปากเพื่อทำการเก็บข้อมูล การถ่ายภาพวิดีโอเพื่อบันทึกการเคลื่อนไหวของรูปปาก การคัดเลือกภาพที่จะนำมาเป็นตัวแทนการเคลื่อนไหวของรูปปาก และการระบุจุดพิกัดรอบปากด้วย Algorithm

3.1.1 การกำหนดจุดพิกัดรอบปากเพื่อทำการเก็บข้อมูล

ในงานวิจัยของ Gutierrez,R.,Osuna. และคณะ[8] ได้กำหนดจุดพิกัดรอบปากไว้ 8 – 10 จุด (รูปที่ 3.2(ก)) ในขณะที่งานวิจัยของ Hong, Pengyu., Wen, Zhen., และ Huang, S., Thomas. [4] ใช้จำนวนจุดพิกัดรอบปากถึง 14 จุด (รูปที่ 3.2(ข)) ซึ่งความละเอียดของการเคลื่อนไหวของรูปปากจากงานวิจัยทั้ง 2 ไม่ได้แตกต่างกันมากนัก เราจึงกำหนดจุดพิกัดรอบริมฝีปากโดยพิจารณาจุดที่แสดงการเคลื่อนไหวอย่างชัดเจนโดยอาศัยจุดพิกัดจากงานวิจัยที่ได้กล่าวมาทั้ง 2 ชิ้น และงานวิจัยของ Kalberer A.G. และ Luc Van Gool [2] ที่ใช้ Triangular mesh model ในการเก็บรูปแบบการเคลื่อนไหวของรูปปาก (รูปที่ 3.2(ค)) โดยอาศัยรูปแบบของจุดพิกัดที่มีตำแหน่งใกล้เคียงกันจากงานวิจัยทั้ง 3 ชิ้น ทำให้เราได้จุดพิกัดสำหรับใช้ในงานวิจัย 11 จุดที่สามารถครอบคลุมการเคลื่อนไหวของรูปปากได้ ตามที่แสดงในรูปที่ 3.3 ที่แสดงจุดพิกัดของรูปปากที่ใช้ในงานวิจัยของเรา



(ก) (ข) (ค)

รูปที่ 3.2 จุดพิกัดที่ถูกนำมาใช้ในการสร้างภาพเคลื่อนไหวในงานวิจัยชิ้นอื่น

รูปที่ 3.3 จุดพิกัดของรูปปากทั้ง 11 จุดที่นำมาใช้ในการวิจัย

3.1.2 การถ่ายภาพวิดีโอเพื่อบันทึกการเคลื่อนไหวของรูปปาก

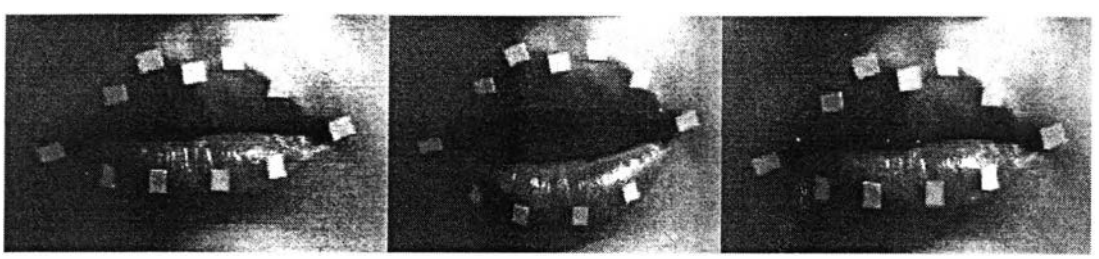
งานวิจัยของเราต้องการเก็บรูปแบบการเคลื่อนไหวเพื่อนำมาสร้างแบบจำลอง ดังนั้นจึงเลือกที่จะเก็บรูปแบบการเคลื่อนไหวของรูปปากในขณะพูดด้วยการถ่ายวิดีโอของตัวอย่างด้วยอัตราความถี่ 15 frames/sec และกำหนดคำที่นำมาใช้ในการทดลองเป็นเสียงสระเดี่ยว 16 เสียง ดังที่แสดงในตารางที่ 3.1 โดยคัดมาเฉพาะสระเสียงเดี่ยวตามหลักภาษาไทยและการออกเสียงภาษาไทยและเป็นการพูดคำเดี่ยวไม่ต่อเนื่อง ทำการถ่ายวิดีโอขณะพูดของตัวอย่าง 5 คน คนละ 5 ครั้ง

ตารางที่ 3.1 ค่า Index ของเสียงสระเดี่ยวที่ใช้ในการวิจัยทั้ง 16 เสียง

Index no.	สัญลักษณ์เสียง	เสียงสระ
1	/a/	อะ
2	/aa/	อา
3	/i/	อิ
4	/ii/	อี
5	/v/	อึ
6	/vv/	อือ
7	/u/	อุ
8	/uu/	อู
9	/e/	เอะ
10	/ee/	เอ
11	/x/	แอะ
12	/xx/	แอ
13	/@/	เอาะ
14	/@@/	ออ
15	/ai/	ไอ
16	/ai/	ใอ

3.1.3 การคัดเลือกภาพที่จะนำมาเป็นตัวแทนการเคลื่อนไหวของรูปปาก

ในงานวิจัยของ Gutierrez,R.,Osuna. และคณะ[8] ได้แสดงถึงทิศทางการเคลื่อนไหวของจุดพิกัดรอบริมฝีปากในลักษณะที่ใกล้เคียงกับเส้นตรง ซึ่งทำให้สามารถคาดคะเนการเคลื่อนไหวของจุดพิกัดได้ง่ายขึ้น และไม่จำเป็นต้องนำเอาภาพการเคลื่อนไหวของจุดพิกัดมาทำการพิจารณาทุก frame จึงได้ทำการคัดเลือกเฉพาะภาพที่แสดงการเคลื่อนไหวของจุดพิกัดอย่างชัดเจนมาเพียงแค่ 3 ภาพ ต่อการพูด 1 คำ คือ ภาพก่อนที่จุดพิกัดรอบริมฝีปากจะมีการเคลื่อนไหว ภาพที่จุดพิกัดรอบริมฝีปากมีการเคลื่อนไหวมากที่สุด และภาพที่จุดพิกัดกลับมาที่ตำแหน่งเริ่มต้นและหยุดการเคลื่อนไหวซึ่งแสดงถึงการสิ้นสุดการพูด ดังแสดงในรูปที่ 3.4 และได้ทำการกำหนดค่า index ลำดับของการเคลื่อนไหวสำหรับนำไปใช้ในการสร้างแบบจำลองเป็น 1 ,2 และ 3 ตามลำดับ ตามตารางที่ 3.2



(ก) (ข) (ค)

รูปที่ 3.4 ภาพในช่วงการเคลื่อนไหวที่เลือกเก็บมาเป็นข้อมูล (ก) ภาพก่อนที่จุดพิกัดรอบริมฝีปากจะทำการเคลื่อนไหว (ข) ภาพที่จุดพิกัดรอบริมฝีปากมีการเคลื่อนไหวมากที่สุด (ค) ภาพที่จุดพิกัดรอบริมฝีปากเคลื่อนไหวกลับมายังตำแหน่งเริ่มต้นและหยุดการเคลื่อนไหว

ตารางที่ 3.2 ค่า index ของลำดับการเคลื่อนไหวและลักษณะของภาพ

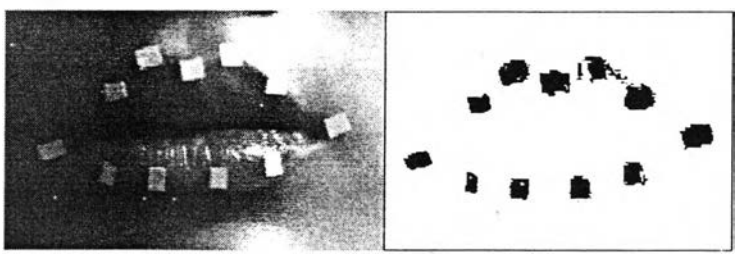
ค่า Index	ลักษณะของภาพ
1	ภาพก่อนที่จุดพิกัดรอบริมฝีปากจะทำการเคลื่อนไหว
2	ภาพที่จุดพิกัดรอบริมฝีปากมีการเคลื่อนไหวมากที่สุด
3	ภาพที่จุดพิกัดรอบริมฝีปากเคลื่อนไหวกลับมายังตำแหน่งเริ่มต้นและหยุดการเคลื่อนไหว

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3.1.4 การระบุจุดพิกัดรอบปากด้วย Algorithm

เมื่อทำการคัดเลือกภาพที่จะนำมาเป็นตัวแทนการเคลื่อนไหวของรูปปากได้แล้ว ในขั้นตอนนี้จะเป็นการระบุจุดพิกัดรอบริมฝีปากออกมาเป็นชุดตัวเลขใน Coordinate – XY จำนวน 11 จุด เพื่อนำไปใช้สร้างแบบจำลองต่อไป ซึ่งการระบุจุดพิกัดนี้จะดำเนินการด้วยชุด Algorithm สำหรับการดึงจุดพิกัดของภาพที่สร้างขึ้นมาเพื่อใช้งานโดยเฉพาะด้วยโปรแกรม Matlab ขั้นตอนของการระบุจุดพิกัดรอบริมฝีปากเป็นตัวเลขมีดังต่อไปนี้

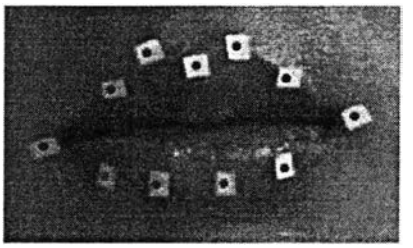
ขั้นตอนที่ 1 ทำการคัดแยกจุดพิกัดบนภาพทั้งหมดที่สอดคล้องกับเงื่อนไขว่า ค่าของสีพิกเซลในพิกัดนั้นมีค่า R, G และ B มากกว่า 200 ขึ้นไปแล้วทำการบันทึกค่าพิกัดของพิกเซลนั้นเพื่อนำไปสู่กระบวนการขั้นถัดไป จุดพิกัดที่ได้จากขั้นตอนนี้จะมีลักษณะดังรูปที่ 3.5



รูปที่ 3.5 ภาพจุดพิกัดรอบริมฝีปากที่ทำการแยกเอาเฉพาะกลุ่มของจุดพิกัดออกมาแล้ว

ขั้นตอนที่ 2 นำข้อมูลที่บันทึกจากในขั้นตอนที่ 1 มาแยกตามกลุ่มข้อมูลเป็น 11 กลุ่ม โดยมีหลักการว่า ข้อมูลที่อยู่ในจุดพิกเซลติดกัน เป็นข้อมูลกลุ่มเดียวกัน จึงกำหนดเงื่อนไขว่า ในการหาข้อมูลของกลุ่มที่ n ถ้าจุดที่ j มีระยะห่างจากจุดที่ i น้อยกว่าหรือเท่ากับ $\sqrt{2}$ แล้ว ให้ข้อมูลของจุดที่ j ย้ายไปอยู่ในกลุ่ม n ไม่เช่นนั้นก็ให้เก็บค่านั้นไว้เพื่อทำการคัดเข้าสู่กลุ่มถัดไป โดยจะทำการตรวจสอบข้อมูลของจุดทั้งหมดที่บันทึกไว้จากขั้นตอนที่ 1 เทียบกับข้อมูลที่ถูกจัดเข้ากลุ่มไปแล้ว จากนั้นจึงทำการบันทึกข้อมูลที่ทำการคัดแยกกลุ่มแล้วและจำนวนของข้อมูลในแต่ละกลุ่มเพื่อนำไปสู่ขั้นตอนที่ 3

ขั้นตอนที่ 3 นำข้อมูลที่ทำการคัดแยกกลุ่มแล้วและจำนวนของข้อมูลมาทำการหาค่าตัวแทนของกลุ่ม โดยพิจารณาเฉพาะกลุ่มข้อมูลที่มีจำนวนของข้อมูลตั้งแต่ 5 จุดขึ้นไป เพราะถือว่ากลุ่มข้อมูลที่มีจำนวนของข้อมูลน้อยกว่า 5 จุด เป็นค่าที่นอกเหนือกลุ่มหรือเป็น noise แล้วนำเอาค่าที่มีจำนวนข้อมูลตั้งแต่ 5 จุด มาหาค่าจุดกึ่งกลางของกลุ่มเพื่อเป็นตัวแทนของกลุ่ม



รูปที่ 3.6 ภาพที่ทำการบันทึกจุดพิกัดลงบนภาพโดยตรงเพื่อใช้เป็นชุดข้อมูลอ้างอิง

ขั้นตอนที่ 4 นำชุดข้อมูลที่ได้จากขั้นตอนที่ 3 มาเทียบกับชุดข้อมูลอ้างอิงโดยกำหนดเงื่อนไขว่า หากมีจุดพิกัดตัวแทนข้อมูลใดอยู่ใกล้กับจุดของชุดข้อมูลอ้างอิงมากที่สุด แต่ไม่เกินระยะ 15 หน่วย ให้ทำการบันทึกจุดพิกัดตัวแทนข้อมูลจุดนั้นตามลำดับของจุดพิกัดอ้างอิง กล่าวคือ ในจุดพิกัดอ้างอิงทั้ง 11 จุดมีการระบุค่าลำดับที่ของข้อมูลในลักษณะของ Array ดังนั้น หากมีค่าพิกัดที่สอดคล้องกับเงื่อนไขข้างต้น ก็ให้บันทึกค่าพิกัดนั้นลงใน Array ตามลำดับของจุดพิกัดอ้างอิง ซึ่งในขั้นตอนนี้จะสามารถทำการคัดแยกข้อมูลที่ถือว่าเป็น noise ที่เราไม่ต้องการออกไปได้อีก และทำการจัดลำดับของจุดพิกัดเพื่อบันทึกลง Array พร้อมกับตรวจสอบว่ามี จุดพิกัดในลำดับ Array ใดหายไปหรือไม่เพื่อทำการทดแทนขึ้นมาใหม่ในขั้นถัดไป

ขั้นตอนที่ 5 ค่า Array ทั้ง 11 ค่าจะถูกนำมาตรวจสอบว่า มีค่าใดที่หายไปหรือไม่ ซึ่งค่าที่หายไป จะถูกบันทึกใน Array ว่าเป็น (0,0) จากนั้นจึงทำการสร้างจุดพิกัดทดแทนโดยทำการคำนวณจากชุดข้อมูลอ้างอิงและชุดข้อมูลที่บันทึกใน Array ดังสมการต่อไปนี้

กำหนดให้ (x_i, y_i) เป็นจุดพิกัดที่ขาดหายไป

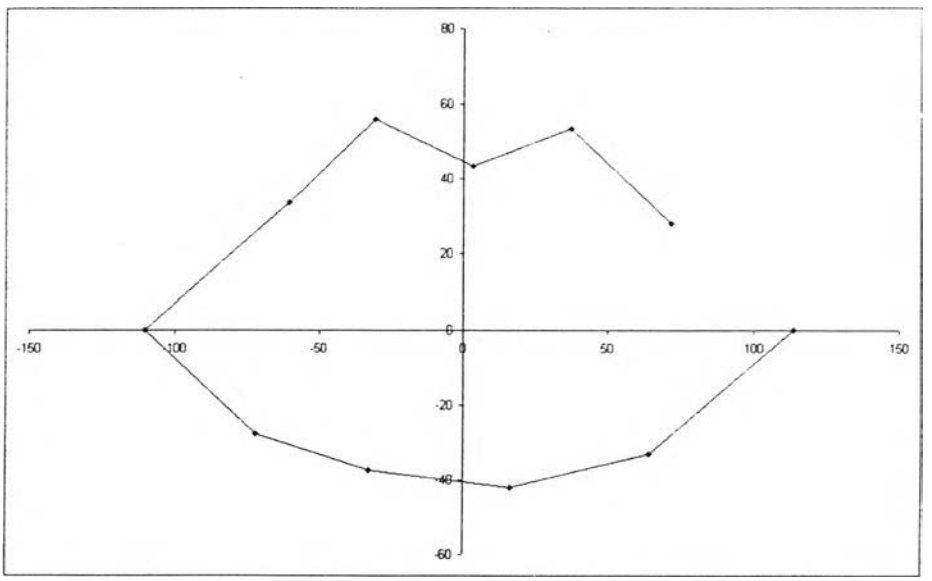
$$\vec{D} = \left\{ \sum_1^n (x_i - xr_i), \sum_1^n (y_i - yr_i) \right\} \tag{3.1}$$

$$(x_i, y_i) = (xr_i, yr_i) + \vec{D} \tag{3.2}$$

จากสมการที่ 3.1 ค่าของ $\vec{D}$ จะเป็นค่าเวคเตอร์ระยะห่างของจุดในกลุ่มข้อมูล x กับชุดข้อมูลอ้างอิง (xr_i, yr_i) และค่าเวคเตอร์ $\vec{D}$ จะนำไปใช้ปรับค่าของชุดข้อมูลอ้างอิงในลำดับของชุดข้อมูลที่ขาดหายไป เพื่อนำค่าข้อมูลในชุดอ้างอิงนั้นเข้าไปแทนที่จุดพิกัดในลำดับที่ขาดไปโดยประมาณ

ขั้นตอนที่ 5 ทำการปรับจุดพิกัดให้มีจุดกึ่งกลางที่ (0,0) และทำการนำค่าของจุดพิกัดลำดับที่ 1 และ 6 ที่มีค่าตามแนวแกน X มากที่สุด และน้อยที่สุด มาทำการหมุนจุดพิกัดทั้งหมดเพื่อแก้ภาพรวมของชุดจุดพิกัดที่เอียงไปเนื่องจากขั้นตอนการถ่ายภาพ

ซึ่งข้อมูลที่ได้จากขั้นตอนการจัดเก็บจุดพิกัดรอบริมฝีปากด้วย Algorithm ทั้งหมดจะมีลักษณะดังรูปที่ 3.7 และข้อมูลที่จัดเก็บได้จะถูกนำไปใช้ในการสร้างแบบจำลองด้วย Backpropagation Neural Network ต่อไป



รูปที่ 3.7 ข้อมูลจุดพิกัดที่ได้จากเก็บข้อมูลด้วย Algorithm

3.2 การสร้างแบบจำลองด้วย Backpropagation Neural Network

จากรูปที่ 3.7 จะเห็นได้ว่าชุดข้อมูลทั้งหมดของรูปปากรูปหนึ่งจะประกอบไปด้วยจุดพิกัดทั้งหมด 11 จุด และได้ทำการกำหนดค่า parameters ที่จะนำมาใช้ในการสร้างแบบจำลองตามลักษณะขององค์ประกอบของรูปปากทั้งหมด 6 ค่า และค่า index ของเสียงสระ และค่า index ของลำดับการเคลื่อนไหว รวม 8 ค่า ได้แก่

1. ค่าความสูงของปากทั้งหมด
2. ค่าความกว้างของปาก
3. ค่าความหนาของริมฝีปากบน
4. ค่าความหนาของริมฝีปากล่าง

5. ค่าความสูงของช่วงริมฝีปากส่วนใต้จมูก

6. ค่าความกว้างของช่วงริมฝีปากส่วนใต้จมูก

7. ค่า index ของเสียงสระ

8. ค่า index ของลำดับการเคลื่อนไหว

และได้กำหนดรูปแบบของการสร้างแบบจำลองโดย Backpropagation Neural Network ในลักษณะของ Multi-Model คือการสร้างแบบจำลองสำหรับทำการจำลองค่าของจุดพิกัดเฉพาะในลำดับของจุดพิกัดนั้นๆ เนื่องจากการสร้างแบบจำลองด้วย Neural Network จำเป็นต้องมีการระบุจำนวนของ Hidden Nodes ในชั้นของ Hidden Layer ซึ่งจำนวนของ Hidden Nodes จะมีผลอย่างมากความแม่นยำของแบบจำลองและสาเหตุที่จำนวนของ Hidden Nodes มีจำนวนไม่เท่ากัน เป็นเพราะลักษณะของข้อมูลที่มีการกระจายตัวไม่เท่ากัน ทำให้จำเป็นต้องมีจำนวนของ Hidden Nodes สำหรับการจำแนกข้อมูลที่ต่างกันด้วย ด้วยเหตุนี้จึงทำให้เราพิจารณาการสร้างแบบจำลองเป็นแบบ Multi-Model เพื่อเพิ่มความยืดหยุ่นสำหรับการพิจารณาค่า weight และค่า bias ที่จะนำมาใช้ในแบบจำลองแบบอิสระ

ด้วยเหตุนี้ลักษณะของแบบจำลองที่เราสร้างขึ้นจึงมีทั้งหมด 11 ชุด โดยที่แต่ละชุดจะมีการสร้าง Hidden Nodes ในจำนวนที่เป็นอิสระต่อกัน แต่ใช้ค่าของ Input Layer เดียวกันทั้งหมด และได้ทำการหาจำนวน Hidden Nodes ที่เหมาะสมที่สุดสำหรับแบบจำลองแต่ละชุดในช่วงตั้งแต่ 1-100 โดยใช้ Rprop เป็น Learning Function และใช้ Hyperbolic Tangent เป็น Activate Function

ในการสร้างแบบจำลองนั้นได้ทำการสร้างแบบจำลองทั้งหมด 3 ชุด ได้แก่

1. แบบจำลองที่สร้างจากตัวอย่างบุคคลเดียวชุดที่ 1

2. แบบจำลองที่สร้างจากตัวอย่างบุคคลเดียวชุดที่ 2

3. แบบจำลองที่สร้างจากตัวอย่างหลายบุคคล

ดังนั้นค่าของ Hidden Nodes ที่เหมาะสมสำหรับแบบจำลองทั้ง 3 ชุดจะต่างกัน ซึ่งจำนวนของ Hidden Nodes ที่เหมาะสมสำหรับแบบจำลองแต่ละชุดสามารถดูได้ในตารางที่ 3.3 และ 3.4

ตารางที่ 3.3 จำนวน Hidden Nodes ที่เหมาะสมสำหรับแบบ

จำลองที่สร้างจากตัวอย่างบุคคลเดียวชุดที่ 1 และ 2

ลำดับจุดพิกัด	จำนวน Hidden Nodes ของแบบจำลองชุดที่ 1
1	29
2	92
3	22
4	17
5	93
6	93
7	40
8	55
9	65
10	18
11	90

ลำดับจุดพิกัด	จำนวน Hidden Nodes ของแบบจำลองชุดที่ 2
1	32
2	62
3	36
4	76
5	41
6	54
7	83
8	46
9	11
10	20
11	13

ตารางที่ 3.3 จำนวน Hidden Nodes ที่เหมาะสมสำหรับแบบจำลองที่สร้างจากตัวอย่างหลายบุคคล

ลำดับจุดพิกัด	จำนวน Hidden Nodes
1	89
2	62
3	51
4	82
5	80
6	44
7	17
8	90
9	63
10	69
11	21

3.3 วิธีการสร้างภาพเคลื่อนไหวจากแบบจำลองและการประเมินประสิทธิภาพของแบบจำลอง

จากรูปที่ 3.8 จะเห็นได้ว่า แบบจำลองที่ทำการสร้างขึ้นมาได้นี้จะสามารถสร้างจุดพิกัดของรูปปากทั้ง 11 จุดขึ้นมาได้ด้วยค่า parameters ที่ทำการส่งเข้าสู่แบบจำลองทั้ง 8 ค่า ซึ่งเมื่อทำการส่งค่า parameters ที่เรียงตามค่า index ลำดับการเคลื่อนไหวของรูปปากลำดับที่ 1, 2 และ 3 แล้ว ก็จะได้ชุดจุดพิกัดการเคลื่อนไหวของรูปปาก 1 คำตามเสียงสระที่ส่งค่าเข้าไปด้วย parameter ค่า index ของเสียงสระ และเมื่อสามารถรู้จุดพิกัดของภาพที่ 1 ,2 และ 3 แล้ว จึงทำการส่งค่า Duration Time ระหว่างภาพที่ 1 กับ 2 และระหว่างภาพที่ 2 กับ 3 เพื่อทำการคำนวณจำนวน Frames ที่ต้องสร้างเพื่อหาช่วงการเคลื่อนไหวในแต่ละ Frames ซึ่งจะถือว่ารูปแบบการเคลื่อนไหวของจุดพิกัดรูปปากที่เราต้องการสร้างขึ้นมานั้นมีลักษณะเป็นเส้นตรง จึงทำการคำนวณค่าระยะห่างระหว่างจุดพิกัดของภาพที่ 1 และ 2 มาเก็บไว้ใน Array ที่ใช้เก็บค่าระยะห่างของจุดพิกัด ดังสมการที่ 3.3 แล้วจึงทำการสร้างจุดพิกัดสำหรับแต่ละเฟรมในช่วงระหว่างภาพที่ 1 กับ 2 และระหว่างภาพที่ 2 กับ 3 ตามแบบการสร้าง In-between หรือช่วงเฟรมระหว่างจุดเริ่มต้นและจุดสุดท้ายในการสร้าง animation

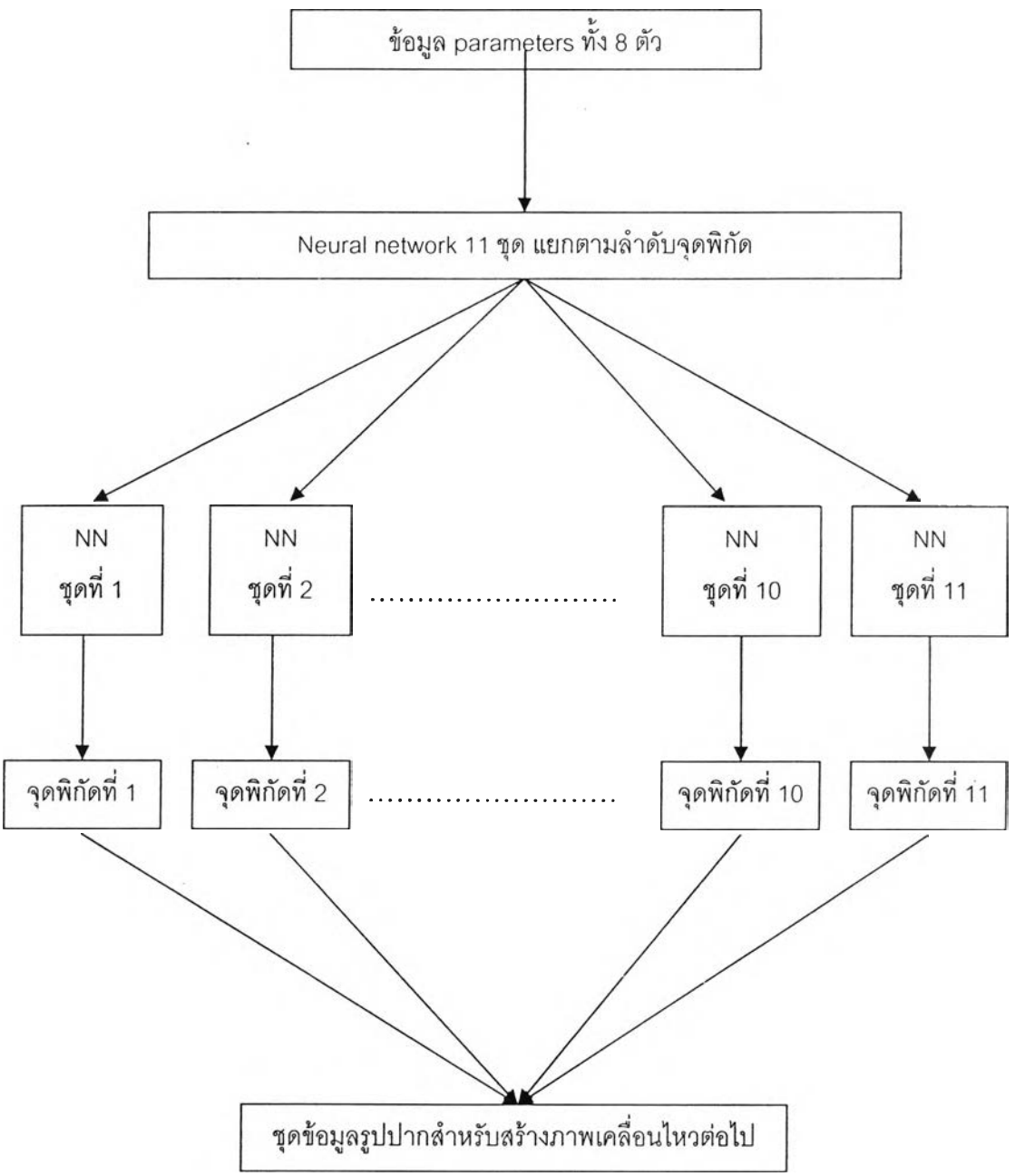
$$d1 = \left((x_1 - x_2), (y_1 - y_2) \right) \tag{3.3}$$

โดยที่ x_1 และ x_2 คือค่าตามแนวแกน X ของจุดพิกัดของภาพที่ 1 และ 2 และ y_1 และ y_2 คือค่าตามแนวแกน Y ของจุดพิกัดของภาพที่ 1 และ 2

การประเมินประสิทธิภาพของแบบจำลองใช้การเปรียบเทียบแบบจำลองทั้ง 3 ชุด ด้วยชุดข้อมูลสำหรับทดสอบ 2 ชุด คือชุดข้อมูล parameters ของตัวอย่างบุคคลที่ 1 และชุดข้อมูล parameters ของตัวอย่างบุคคลที่ 2 แล้วทำการเปรียบเทียบระหว่างจุดพิกัดที่สร้างจากแบบจำลอง กับจุดพิกัดจริงที่บันทึกไว้โดยทำการวัดการเปรียบเทียบด้วยค่า SSE (Sum Square Error) และค่า $Cos\,\theta$ ระหว่าง Array ของค่าระยะห่างของจุดพิกัดที่สร้างจากแบบจำลองและ Array ของค่าระยะห่างของจุดพิกัดที่ทำการบันทึกจากตัวอย่างผู้พูดจริงๆ

วิธีการคำนวณค่า SSE (Sum Square Error) แสดงถึงระยะห่างจากจุดพิกัดที่สร้างด้วยแบบจำลองและจุดพิกัดที่บันทึกจากผู้พูดจริงๆ ซึ่งบอกให้ทราบถึงความคลาดเคลื่อนของจุดพิกัดที่สร้างจากแบบจำลองซึ่งคำนวณได้ตามสมการที่ 3.4 โดยค่า x และ y เป็นค่าตาม

แนวแกน X และ Y ของจุดพิกัดที่สร้างจากแบบจำลอง และค่า xr และ yr เป็นค่าตามแนวแกน X และ Y ของจุดพิกัดที่ทำการบันทึกจากตัวอย่างผู้พูดจริงๆ

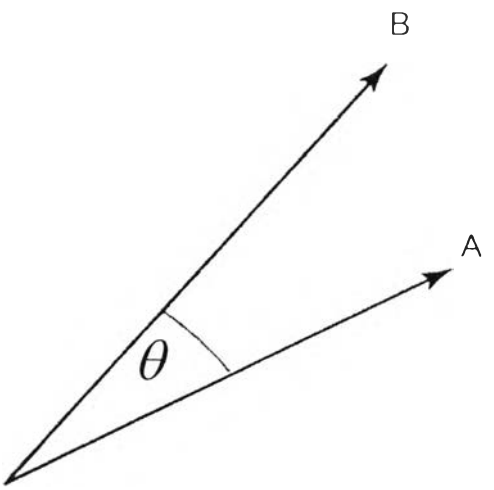


รูปที่ 3.8 แผนผังการทำงานของแบบจำลองแบบ Multi-Model

ที่สร้างด้วย Backpropagation Neural Network

$$SSE = \sum_{i=1}^{11} ((x_i - xr_i)^2 + (y_i - yr_i)^2) \quad (3.4)$$

สำหรับการคำนวณค่า $Cos\,\theta$ สามารถบอกให้เราทราบได้ถึงความสอดคล้องกันของทิศทางการเคลื่อนที่ของจุดพิกัดจากภาพที่ 1 ไปยังภาพที่ 2 และ จากภาพที่ 2 ไปยังภาพที่ 3 ได้ระหว่างจุดพิกัดที่สร้างจากแบบจำลองและจุดพิกัดที่บันทึกจากตัวอย่างผู้พูดจริงๆ



รูปที่ 3.9 ตัวอย่างทิศทางการเคลื่อนที่ของจุดพิกัดที่สร้างจาก

แบบจำลองและจุดพิกัดที่บันทึกจากตัวอย่างผู้พูด

วิธีการคำนวณค่า $Cos\,\theta$ ทำโดยการนำเอาค่า Vector ทิศทางการเคลื่อนที่ของจุดพิกัดแต่ละจุดที่ได้จากการสร้างด้วยแบบจำลองมาเปรียบเทียบกับค่า Vector ทิศทางการเคลื่อนที่ของจุดพิกัดแต่ละจุดที่ได้จากการบันทึกจากตัวอย่างผู้พูดจริง โดยพิจารณาจากค่า $Cos\,\theta$ ของ Vector ทั้ง 2 ที่ทำมุมกัน θ ดังในรูปที่ 3.9 ซึ่งสามารถหาค่าได้จากการคำนวณด้วยสมการ 3.5

$$Cos\,\theta = Cos\left(Arc\tan\left(\frac{\Delta y}{\Delta x}\right) - Arc\tan\left(\frac{\Delta yr}{\Delta xr}\right)\right) \quad (3.5)$$

โดยที่

$$\Delta y = y_2 - y_1 \quad (3.6)$$

$$\Delta x = x_2 - x_1 \tag{3.7}$$

$$\Delta yr = yr_2 - yr_1 \tag{3.8}$$

$$\Delta xr = xr_2 - xr_1 \tag{3.9}$$