

การรู้จำภาษามือไทยแบบต่อเนื่องโดยใช้แบบจำลองสวิตเตนมาร์คอฟโมเดล

CONTINUOUS THAI SIGN LANGUAGE RECOGNITION USING  
HIDDEN MARKOV MODELS

อรุณี จารวนาวัตน์  
ARUNEE JARUWANAWAT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษิตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2550

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การรู้จำภาษามือไทยแบบต่อเนื่องโดยใช้แบบจำลองฮิดเดนมาร์คอฟโมเดล

CONTINUOUS THAI SIGN LANGUAGE RECOGNITION USING  
HIDDEN MARKOV MODELS

อรุณี จารูวานวัฒน์

ARUNEE JARUWANAWAT

เลขหมู่.....  
เลขทะเบียน..... 75126  
วัน,เดือน,ปี..... 19 ต.ค. 2550

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2550

**CONTINUOUS THAI SIGN LANGUAGE RECOGNITION USING  
HIDDEN MARKOV MODELS**

**ARUNEE JARUWANAWAT**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENT FOR THE DEGREE OF  
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY  
SCHOOL OF GRADUATE STUDIES  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**2007**

**COPYRIGHT 2007**

**SCHOOL OF GRADUATE STUDIES**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

หัวข้อวิทยานิพนธ์	การรู้จำภาษามือไทยแบบต่อเนื่อง โดยใช้แบบจำลองฮิดเดน มาร์คอฟโมเดล
นักศึกษา	นางสาวอรุณี จารุณาววัฒน์
รหัสประจำตัว	45066025
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
พ.ศ.	2550
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.นพพร โชติคกำจร

### บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ นำเสนอวิธีการรู้จำภาษามือไทยแบบต่อเนื่อง โดยการใช้ฮิดเดนมาร์คอฟโมเดล ปัญหาหนึ่งที่พบในการรู้จำภาษามือไทยแบบต่อเนื่อง เกิดจากการเคลื่อนไหวของมือในช่วงการเปลี่ยนผ่านจากท่ามือหนึ่งไปสู่อีกท่ามือหนึ่ง ซึ่งท่ามือเคลื่อนไหวนี้เป็นท่ามือที่ไม่มีความหมายเรียกว่าท่าเคลื่อนไหวระหว่างมือ แต่มีความจำเป็นต้องมีการโมเดลท่าเคลื่อนไหวระหว่างมือ เพื่อให้ประสิทธิภาพในการรู้จำดีขึ้น ในงานวิจัยนี้ได้นำเสนอเทคนิคการสร้างโมเดลของท่าเคลื่อนไหวระหว่างมือด้วยวิธีการเชิงอัลกอริทึม เพื่อหลีกเลี่ยงความจำเป็นที่จะต้องจัดเก็บข้อมูลท่าเคลื่อนไหวระหว่างมือในการนำมาใช้เพื่อทำการสอนให้กับฮิดเดนมาร์คอฟโมเดล ซึ่งวิธีการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการเชิงอัลกอริทึมนี้จะถูกนำไปใช้ร่วมกับฮิดเดนมาร์คอฟโมเดลของข้อมูลในระดับคำ นอกจากจะลดความจำเป็นที่ต้องทำการสอนท่าเคลื่อนไหวระหว่างมือให้แก่ระบบแล้ว ยังเป็นการลดปริมาณพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดลที่ต้องจัดเก็บอยู่ในระบบ ซึ่งได้มีการทดสอบความถูกต้องในการรู้จำเปรียบเทียบกันระหว่างการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่ได้จากวิธีการเชิงอัลกอริทึม, การใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือและการไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือในการรู้จำ ผลความถูกต้องคือ 89.04%, 94.04% และ 62.69%

<b>Thesis Title</b>	Continuous Thai Sign Language Recognition using Hidden Markov Models
<b>Student</b>	Ms. Arunee Jaruwanawat
<b>Student ID.</b>	45066025
<b>Degree</b>	Master of Science
<b>Program</b>	Information Technology
<b>Year</b>	2007
<b>Thesis Advisor</b>	Assoc. Prof. Dr. Nopporn Chotikakamthorn

## **ABSTRACT**

This thesis presents the creation transitional gesture model, segmentation and recognition methods for Continuous Thai Sign Language Translation. One of the problems found in sign language translation is on segmenting a hand movement that is part of a transitional movement from one hand gesture to another. This transitional gesture conveys no meaning, but serves as a connecting period between two consecutive gestures. This research introduces algorithmic hand movement modeling technique to avoid the necessary of collecting hand movement data for HMM training process in word level. This technique also decreases the number of collected HMM's parameter. The experiment was performed to compare the accuracy of recognition process between algorithmic based modeling method, hand gesture collected based modeling method and the method without hand movement model. The experimental result is 89.04%, 94.04% and 62.69% respectively.

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างดี ส่วนหนึ่งก็เพราะได้รับความช่วยเหลือจากบุคคลเหล่านี้ ซึ่งผู้วิจัยขอแสดงความขอบคุณ มา ณ โอกาสนี้

ขอบคุณพระคุณครอบครัวจรรวนาวัดน์ ที่ได้ให้กำลังใจ ความเอาใจใส่ เป็นผู้ผลักดันให้ผู้วิจัยมีความพยายามที่จะทำงานวิจัยให้สำเร็จลุล่วง รวมทั้งให้ความช่วยเหลือทางด้านทุนทรัพย์ตลอดมา

ขอขอบพระคุณอาอนุชิต จรรวนาวัดน์ ที่คอยสนับสนุนและช่วยเหลือในทุกๆเรื่อง อีกทั้งยังได้ให้คำแนะนำด้วยดีเสมอมา

ขอขอบพระคุณ รศ. ดร. นพพร โชติกอำร อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำปรึกษาชี้แนะแนวทางในการทำวิจัย ให้ข้อมูลเพิ่มเติมที่เป็นประโยชน์ในการศึกษาการทำวิจัย รวมทั้งแนะนำเทคนิคต่างๆ ในการดำเนินงานวิจัยนี้ อีกทั้งยังให้การสนับสนุนอุปกรณ์และเอกสารที่เกี่ยวข้องกับงานวิจัยได้อย่างดีเยี่ยม จนทำให้งานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่กรุณาช่วยแนะนำงานวิจัยในส่วนต่างๆและวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณอาจารย์ทุกท่าน จากคณะเทคโนโลยีสารสนเทศ ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้ตลอดมาตั้งแต่เข้าเรียนในคณะนี้

ขอขอบคุณเจ้าหน้าที่สำนักวิจัยการสื่อสารและเทคโนโลยีสารสนเทศ (ReCCIT) โดยเฉพาะน้องคู่ที่ช่วยในการเก็บข้อมูลอีกด้วย และเจ้าหน้าที่คณะเทคโนโลยีสารสนเทศ โดยเฉพาะพี่จีที่ได้อำนวยความสะดวกต่างๆตั้งแต่ต้นจนจบการศึกษานี้

ขอขอบคุณสมาชิก MVLab ทุกๆท่าน ที่เป็นที่ปรึกษาและคอยให้กำลังใจเป็นอย่างดี

ขอขอบคุณภพ.ป๊อง, พี่เอ และคุณชายที่ให้คำแนะนำต่างๆ ทั้งเป็นแรงผลักดันและให้กำลังใจตลอดมา

ขอขอบคุณเพื่อนๆทุกคนที่ให้การสนับสนุน กำลังใจ และคำแนะนำต่างๆด้วยดีเสมอมา

อนึ่ง งานวิจัยที่นำเสนอในงานวิทยานิพนธ์ฉบับนี้นั้น ส่วนหนึ่งกระทำภายใต้ห้องปฏิบัติการสื่อประสมและเสมือน (Multimedia and Virtual Research Laboratory: MVLab) สำนักวิจัยการสื่อสารและเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอบแต่บุพการี บุรพคณาจารย์ และผู้มีพระคุณทุกท่าน

อรุณี จรรวนาวัดน์

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย .....	I
บทคัดย่อภาษาอังกฤษ .....	II
กิตติกรรมประกาศ.....	III
สารบัญ .....	IV
สารบัญตาราง .....	VII
สารบัญรูป .....	VIII
บทที่ 1 บทนำ .....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา .....	1
1.3 สมมุติฐานของการศึกษา .....	2
1.4 ขอบเขตของการดำเนินงานวิจัย.....	2
1.5 ขั้นตอนการศึกษา .....	2
1.6 ประโยชน์ที่คาดว่าจะได้รับ .....	3
1.7 โครงสร้างวิทยานิพนธ์ .....	3
บทที่ 2 งานวิจัยที่เกี่ยวข้อง .....	4
บทที่ 3 ทฤษฎีที่เกี่ยวข้องในงานวิจัย .....	8
3.1 ภาษามือไทย.....	8
3.1.1 ความหมายและคำจำกัดความของภาษามือไทย.....	8
3.1.2 องค์ประกอบของภาษามือไทย.....	9
3.2 อุปกรณ์ที่ใช้ในงานวิจัยภาษามือ.....	13
3.2.1 ประเภทของอุปกรณ์นำเข้าข้อมูล .....	13
3.2.2 เปรียบเทียบข้อดีและข้อเสียของอุปกรณ์นำเข้าข้อมูล.....	14
3.2.3 สรีระของมือ.....	15
3.2.4 ถุงมืออิเล็กทรอนิกส์.....	16
3.2.5 เครื่องมือตรวจวัดตำแหน่งแบบ 6 แกน .....	17
3.3 ทฤษฎีแบบจำลองฮิดเดนมาร์คอฟโมเดล .....	18

## สารบัญ (ต่อ)

	หน้า
3.3.1 ส่วนประกอบของแบบจำลองฮิดเดนมาร์คอฟโมเดล.....	18
3.3.2 ชนิดของแบบจำลองฮิดเดนมาร์คอฟโมเดล .....	18
3.3.3 ปัญหาพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟโมเดล .....	20
<b>บทที่ 4 การรู้จำท่ามือแบบต่อเนื่องในภาษามือไทย .....</b>	<b>27</b>
4.1 ระบบการแปลภาษามือไทย.....	27
4.1.1 กระบวนการสอนฮิดเดนมาร์คอฟโมเดล และการใช้งานฮิดเดนมาร์คอฟโมเดล..	29
4.1.2 การใช้ฮิดเดนมาร์คอฟในการรู้จำท่าในภาษามือแบบไม่ต่อเนื่อง .....	30
4.2 วิธีการรู้จำภาษามือไทยชนิดต่อเนื่อง.....	31
4.2.1 การเลือกจำนวนสเตทของฮิดเดนมาร์คอฟโมเดล .....	31
4.2.2 การสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือ .....	32
4.2.3 การรู้จำด้วย Viterbi algorithm.....	34
<b>บทที่ 5 อุปกรณ์และข้อมูลที่ใช้ในการทดลอง .....</b>	<b>36</b>
5.1 อุปกรณ์ที่ใช้ในการเก็บข้อมูล.....	36
5.2 ข้อมูลที่ใช้ในการทดลอง.....	36
5.2.1 สภาพแวดล้อมขณะเก็บข้อมูล.....	36
5.2.2 ท่ามือที่ใช้ในการทดลอง .....	36
<b>บทที่ 6 วิธีการทดลองและผลการทดลอง .....</b>	<b>42</b>
6.1 การทดลองเพื่อหาจำนวนสเตทของฮิดเดนมาร์คอฟโมเดล .....	42
6.2 การทดสอบการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ..	45
6.3 การทดสอบการรู้จำโดยการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่า มือ .....	47
6.4 การทดสอบการรู้จำโดยไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือ .....	50
<b>บทที่ 7 สรุปผลการทดลองและข้อเสนอแนะ .....</b>	<b>54</b>

## สารบัญ (ต่อ)

	หน้า
เอกสารอ้างอิง.....	56
ภาคผนวก.....	58
ภาคผนวก ก. บทความและผลงานวิจัยที่ได้รับการตีพิมพ์.....	59
ประวัติผู้เขียน.....	70

## สารบัญตาราง

ตารางที่	หน้า
5.1 แสดงข้อมูลท่ามือท่าเคลื่อนไหว.....	38
5.2 แสดงข้อมูลท่าหนึ่ง .....	39
5.3 แสดงประโยคทั้งหมดที่ใช้ในการทดลอง.....	39
6.1 ตารางแสดงการแบ่งกลุ่มที่ให้ค่าความถูกต้องสูงสุด .....	43
6.2 ตารางแสดงผลการทดลองเมื่อนำโมเดลท่าเคลื่อนไหวระหว่างมือด้วยการคำนวณมาใช้ในการ รู้จำประโยค .....	45
6.3 ตารางแสดงผลการทดลองเมื่อนำโมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือมาใช้ ในการรู้จำประโยค .....	48
6.4 ตารางแสดงผลการทดลองเมื่อไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือในการรู้จำประโยค.....	51

# สารบัญรูป

รูปที่	หน้า
2.1 โค้ดแแกรมของ OP pre-searching.....	5
3.1 ตัวอย่างรูปมือในภาษามือไทย.....	9
3.2 (a) ทำมือที่มีความหมายว่า “ฉัน” (b) ทำมือที่มีความหมายว่า “รู้”.....	9
3.3 ลักษณะการเคลื่อนไหว และทิศทางของการเคลื่อนไหวปกติ.....	9
3.4 สัญลักษณ์แทนการเคลื่อนไหว และทิศทางของการเคลื่อนไหวอย่างรวดเร็ว.....	10
3.5 สัญลักษณ์แทนการเคลื่อนที่จากจุดหนึ่งไปยังอีกจุดหนึ่งโดยก้าวสั้นๆ.....	10
3.6 สัญลักษณ์แทนการเคลื่อนที่ไปในทางเดียวหลายๆ ครั้ง.....	10
3.7 สัญลักษณ์แทนการเคลื่อนไหวที่ไป-กลับหลายๆ ครั้ง.....	10
3.8 สัญลักษณ์แทนการเคลื่อนไหวที่มือทั้งสองเคลื่อนสลับกัน.....	10
3.9 สัญลักษณ์แทนการหมุนข้อมือหรือส่วนของแขนระหว่างข้อศอกกับมือเป็นวงกลม.....	11
3.10 สัญลักษณ์แทนการเคลื่อนที่ไปสองทิศทางที่มีการสั้น.....	11
3.11 สัญลักษณ์แทนการแตะเบาๆ หลายครั้ง.....	11
3.12 สัญลักษณ์แทนการหมุนปลายแขน.....	11
3.13 สัญลักษณ์แทนการเคลื่อนไหวที่มีการบิดเข้าเล็กน้อย, การกระพริบ.....	12
3.14 สัญลักษณ์แทนปลายนิ้วคลายออกจากใต้นิ้วหัวแม่มือ.....	12
3.15 สัญลักษณ์แทนการกระดิกนิ้วในลักษณะท่าเล่นเปียโน การพิมพ์ดีด (A) กระดิก 2 นิ้ว (B) กระดิก 5 นิ้ว.....	12
3.16 สัญลักษณ์แทนสิ่งที่มีลักษณะละเอียดหรือเป็นผง (A) นิ้วมือทั้งสี่กับนิ้วหัวแม่มือหลายๆ ครั้ง (B) ใช้นิ้วชี้ นิ้วกลาง และนิ้วหัวแม่มือเท่านั้น.....	12
3.17 สัญลักษณ์แทนการเคลื่อนไหวที่เริ่มจากท่ากางนิ้ว.....	13
3.18 สัญลักษณ์แทนการใช้นิ้วหัวแม่มือสัมผัสนิ้วอื่นครั้งหนึ่ง.....	13
3.19 (a) ทำมือที่มีความหมายว่า “ฉัน” (b) ทำมือที่มีความหมายว่า “เธอ”.....	13
3.20 แสดงโครงสร้างทางกายวิภาคมือขวาของมนุษย์.....	15
3.21 อุปกรณ์ถุงมืออิเล็กทรอนิกส์.....	17
3.22 อุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกนรุ่น 3SPACE® ISOTRAK II™ พร้อมด้วยเซนเซอร์ 2 ตัว และ Transmitter.....	17
3.23 แบบจำลองชนิดต่างๆของแบบจำลองฮิดเดนมาร์คอฟโมเดล.....	19
3.24 กระบวนการไปยังหน้า.....	22

## สารบัญรูป (ต่อ)

รูปที่	หน้า
3.25 กระบวนการย้อนกลับ .....	23
3.26 ลำดับการคำนวณการเกิดค่าปรากฏรวมซึ่งจะอยู่ที่สเตต $i$ เวลา $t$ และอยู่ที่สเตต $j$ เวลา $t+1$ ....	25
4.1 การแบ่งแยกรูปแบบของการแสดงท่าทางในระบบภาษามือไทย .....	27
4.2 ระบบรู้จำภาษามือไทยชนิดต่อเนื่องอัตโนมัติ .....	28
4.3 กระบวนการสอนฮิดเดนมาร์คอฟโมเดล .....	29
4.4 ตัวอย่างโทโปโลยีของแต่ละฮิดเดนมาร์คอฟโมเดลแบบ 3 สเตตซึ่ง 1 โมเดลแทน 1 คำ.....	30
4.5 การใช้งานฮิดเดนมาร์คอฟโมเดล .....	31
4.6 ท่าเคลื่อนไหวระหว่างท่ามือซึ่งอยู่ระหว่างคำ A และ คำ B .....	32
4.7 แสดงการจำแนกของ $\bar{U}$ ออกเป็น 4 submatrice .....	34
4.8 แผนภาพแสดงการรู้จำด้วย Viterbi algorithm .....	35
6.1 แสดงการแบ่งกลุ่มของค่า $Q$ ด้วย Threshold .....	42
6.2 กราฟแสดงผลการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ.....	47
6.3 กราฟแสดงผลการรู้จำโดยการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือ .....	50
6.4 กราฟแสดงผลการรู้จำโดยไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือ.....	53

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันคนพิการทางหู และคนพิการทางเสียง สื่อสารกันโดยใช้ภาษามือซึ่งคนที่สื่อสารด้วยนั้นต้องรู้ภาษามือเช่นกัน จะทำให้คนที่ไม่รู้ภาษามือ จะสื่อสารกับคนพิการทางหูและทางเสียงได้ลำบาก เพราะไม่เข้าใจภาษามือ ปัจจุบันมีการวิจัยและพัฒนาเพื่อแก้ไขปัญหาคาดช่องว่างระหว่างคนพิการกับคนปกติในต่างประเทศมากมาย เช่น ในสหรัฐอเมริกา เป็นต้น ซึ่งการวิจัยเหล่านี้มีจุดประสงค์ที่จะทำให้คนพิการมีส่วนร่วมในสังคมและดำรงชีวิตเหมือนคนปกติ ในจำนวนคนพิการในแต่ละประเทศสามสี่ คนหูหนวกเป็นคนพิการที่มีความบกพร่องทางประสาทรับฟังเท่านั้น แต่สำหรับคนที่หูหนวกมาแต่ตั้งแต่กำเนิดจะไม่สามารถพูดเสียงที่ถูกต้องได้ เพราะคนหูหนวกจะไม่เคยได้รับฟังเสียงของคำแต่ละคำ ทำให้ไม่สามารถออกเสียงได้ถูกต้อง ซึ่งทำให้การสื่อสารระหว่างคนหูหนวกหรือคนพูดไม่ได้กับคนปกติยังมีความยากลำบากอยู่ กล่าวคือ ในเวลาที่คนหูหนวกคุยกับคนหูหนวกด้วยกันเองจะใช้ภาษามือ ซึ่งใช้ลักษณะท่าทางและตำแหน่งของมือสื่อความหมายและภาษามือก็ถูกนำมาใช้ในการสนทนาหรือติดต่อกับคนปกติด้วยเพราะคนหูหนวกไม่สามารถฟังเสียงของคนปกติได้ คนปกติจึงจำเป็นต้องรู้ภาษามือหรืออาจต้องคุยกับคนหูหนวกโดยผ่านล่ามภาษามือ ดังนั้นจึงได้มีระบบการจดจำภาษามือเกิดขึ้น ซึ่งในปัจจุบันมีการนำทฤษฎีต่างๆ มาใช้อย่างมากมาย เช่น Template matching, Neuron networks และ Hidden Markov Model และในงานวิจัยฉบับนี้ได้นำเสนอการพัฒนาสร้างโมเดลของการเคลื่อนไหวระหว่างมือด้วยการคำนวณ เพื่อให้การรู้จำภาษามือไทยชนิดต่อเนื่องมีความถูกต้องมากขึ้น

ผู้วิจัยจึงตระหนักในปัญหาที่เกิดขึ้น จึงริเริ่มทำงานวิจัยเกี่ยวกับการติดต่อสื่อสารในกลุ่มผู้พิการทางหู ซึ่งมีวัตถุประสงค์ดังนี้

1. เพื่อให้ผู้พิการทางหูติดต่อสื่อสารกับผู้อื่นได้
2. เพื่อให้ผู้พิการทางหูติดต่อกับคอมพิวเตอร์ได้

### 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

ความมุ่งหมายและวัตถุประสงค์ของการศึกษางานวิทยานิพนธ์นี้ เพื่อศึกษาและพัฒนาวิธีการรู้จำภาษามือไทยแบบต่อเนื่อง เพื่อใช้ในระบบการแปลภาษามือไทยอัตโนมัติ โดยนางงานวิจัยนี้ไปผสมผสานกับระบบเดิมในส่วนของโครงข่ายทำนอง และใช้ฮิดเดนมาร์คอฟโมเดลในการรู้จำท่าเคลื่อนไหวไม่ซ้ำ เพื่อใช้ในระบบการแปลภาษามือไทยอัตโนมัติ ซึ่งช่วยทำให้ระบบโดยรวม

สามารถรู้จำและแปลภาษามือไทยอัตโนมัติได้สมบูรณ์ยิ่งขึ้น สามารถสรุปวัตถุประสงค์ออกเป็นได้ดังนี้

1.2.1 เพื่อศึกษาและพัฒนาวิธีการรู้จำภาษามือไทยแบบต่อเนื่อง โดยใช้ฮิดเดนมาร์คอฟโมเดล

1.2.2 เพื่อพัฒนาวิธีสร้างโมเดลท่าเคลื่อนไหวระหว่างท่ามือ เพื่อเพิ่มประสิทธิภาพและความถูกต้องในการรู้จำ โดยไม่ต้องอาศัยการสอนท่าเคลื่อนไหวระหว่างมือให้แก่ระบบ

1.2.3 เพื่อศึกษาและพัฒนาในการลดเส้นทางการค้นหาด้วย Viterbi algorithm

### 1.3 สมมุติฐานของการศึกษา

ในระบบแปลภาษามือไทยที่เป็นประโยชน์นั้น หากมีการสร้างโมเดลของท่าเคลื่อนไหวระหว่างมือ (Transitional gesture) โดยการเฉลี่ยค่าได้ที่จากการสอนระหว่างสเตตสุดท้ายของโมเดลท่ามือที่มีความหมายก่อนหน้ากับสเตตแรกของโมเดลท่ามือที่มีความหมายถัดไป จะช่วยเพิ่มประสิทธิภาพในการรู้จำภาษามือแบบต่อเนื่องได้ อีกทั้งลดความซับซ้อนให้แก่ระบบและยังเป็นการลดขนาดฐานข้อมูลอีกด้วย

### 1.4 ขอบเขตของการดำเนินงานวิจัย

1.4.1 เป็นงานวิจัยที่เป็นแบบ Signer dependence คือ แบบที่ขึ้นกับผู้ทำท่ามือ โดยจะหนึ่งคนจะมีหนึ่งฐานข้อมูล และเป็นการทำท่ามือที่ต่อเนื่องโดยผู้ทำท่ามือต้องไม่มีความลังเลในการทำท่ามือ

1.4.2 การเริ่มทำท่ามือจะต้องเริ่มด้วยท่ามือที่กำหนดไว้

1.4.3 ท่ามือที่ใช้จะเลือกมาจากวิดิทัศน์ประกอบการเรียนการสอนภาษามือไทย และหนังสือภาษามือไทยเล่ม 1-6 [12] ซึ่งเลือกมาใช้เฉพาะคำที่ใช้บ่อยในชีวิตประจำวันประมาณ 100 คำ

1.4.4 ภาษามือที่ใช้จะไม่รวมการแสดงออกทางสีหน้า (Single Hand)

### 1.5 ขั้นตอนการศึกษา

1.5.1 กำหนดหัวข้อ เป้าหมาย จุดประสงค์ และขอบเขตของการทำวิทยานิพนธ์

1.5.2 ศึกษาทฤษฎี และหลักการพื้นฐานที่ใช้ในการวิจัย

1.5.3 ศึกษาปัญหา และวิเคราะห์วิธีการแก้ปัญหาที่เกิดขึ้น

1.5.4 พัฒนาโปรแกรมคอมพิวเตอร์ประกอบการทดลอง

1.5.5 ทำการทดลอง ปรับปรุง และสรุปผล

1.5.6 จัดทำเอกสารประกอบวิทยานิพนธ์

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

ในงานวิจัยนี้ได้นำเสนอการรู้จำภาษามือไทยแบบต่อเนื่องและการสร้างโมเดลของการเคลื่อนไหวก่อนหน้าท่ามือด้วยการคำนวณ ซึ่งเป็นการลดการจับเก็บการเคลื่อนไหวก่อนหน้าท่ามือลงในฐานข้อมูล การสร้างโมเดลของการเคลื่อนไหวก่อนหน้านั้นเป็นการสร้างขึ้นมาใช้ชั่วคราวในหน่วยความจำในขณะที่โปรแกรมทำงานเท่านั้น ไม่ได้เป็นการจัดเก็บถาวร นอกจากนี้ยังเป็นการลดความซับซ้อนและความยุ่งยากในการสอนการเคลื่อนไหวก่อนหน้าท่ามือให้แก่ระบบอีกด้วย และงานวิจัยนี้นำระบบงานวิจัยก่อนหน้าในส่วนของ การตรวจจับท่ามือมาผสมผสาน เพื่อไม่ต้องใช้ฮาร์ดแวร์ โมเดลอย่างเดียวกันทั้งในการรู้จำท่ามือและท่าเคลื่อนไหวก่อนหน้า ซึ่งเป็นการลดการคำนวณและความซับซ้อนอีกทางหนึ่ง และจากระบบงานวิจัยก่อนหน้ายังไม่สามารถแยกและรู้จำท่าเคลื่อนไหวก่อนหน้าได้ ซึ่งงานวิจัยนี้จะช่วยทำให้ระบบแปลภาษามือไทยแบบต่อเนื่องมีความสมบูรณ์ยิ่งขึ้น

## 1.7 โครงสร้างวิทยานิพนธ์

เพื่อความสะดวกในการศึกษาวิทยานิพนธ์ฉบับนี้ วิทยานิพนธ์ฉบับนี้จึงแบ่งเป็นส่วนที่สำคัญ ดังนี้

บทที่ 1 บทนำ กล่าวถึงความเป็นมาและความสำคัญของปัญหา ความมุ่งหมายและวัตถุประสงค์ของการศึกษา สมมุติฐานของการศึกษา ขอบเขตของการดำเนินงานวิจัย ขั้นตอนการดำเนินงานวิจัย และประโยชน์ที่คาดว่าจะได้รับ และ โครงสร้างวิทยานิพนธ์

บทที่ 2 งานวิจัยที่เกี่ยวข้อง กล่าวถึงงานวิจัยที่ผ่านมาในอดีตถึงปัจจุบัน

บทที่ 3 ทฤษฎีที่เกี่ยวข้องในงานวิจัย อธิบายถึงลักษณะของภาษามือไทย สรีระของมือ ข้อมูลอุปกรณ์ที่ใช้ในการวิจัย และทฤษฎีแบบจำลองฮิดเดนมาร์คอฟโมเดล

บทที่ 4 รู้จำท่าทางในภาษามือไทย การสร้างโมเดลท่าเคลื่อนไหวก่อนหน้าท่ามือ รวมถึง อัลกอริทึมที่ใช้ในงานวิจัย

บทที่ 5 อุปกรณ์และข้อมูลที่ใช้ในการทดลอง

บทที่ 6 วิธีการทดลองและผลการทดลองในการรู้จำท่าทางในงานวิจัย

บทที่ 7 สรุปผลการวิจัยและข้อเสนอแนะ

## บทที่ 2

### งานวิจัยที่เกี่ยวข้อง

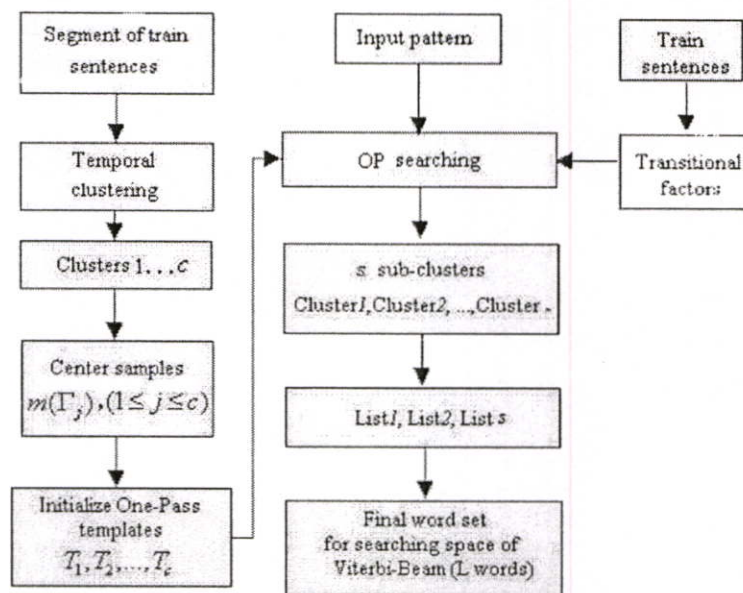
งานวิจัยที่เกี่ยวกับการแปลภาษามือมีหลายๆ งานวิจัยที่ผ่านมา ดังนี้

Gaolin Fang และ Wen Gao [1] ได้ใช้อุปกรณ์ถุงมืออิเล็กทรอนิกส์ร่วมกับอุปกรณ์ตรวจวัดตำแหน่ง 6 แกน ได้มีการรู้จำท่ามือต่อเนื่องที่เป็นแบบ Signer-Independent วิธีที่ใช้คือได้ทำการ pre-segment ด้วย Simple Recurrent Network (SRN) เพื่อที่จะตัดการเคลื่อนไหวระหว่างท่ามือ (Transition Movement) ออกก่อนที่จะนำข้อมูลมารู้จำด้วยฮิดเดนมาร์คอฟโมเดล ทำให้ไม่ต้องสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือ เพื่อที่จะทำให้ sequence สั้นลง ทำให้เวลาที่ใช้ Viterbi algorithm ในการค้นลำดับของคำที่ดีที่สุดก็จะสั้นลงด้วย ในการสร้างโมเดลนั้นสเตทของฮิดเดนมาร์คอฟโมเดลอยู่ระหว่าง 1-4 สเตท ซึ่งขึ้นอยู่กับความยาวของข้อมูล ในการทดลองมีการเก็บท่ามือจากผู้ทำท่ามือทั้งหมดจำนวน 7 คน ซึ่งกระทำท่าหนึ่ง 208 ท่าจำนวนท่าละ 3 ครั้ง สำหรับประโยคต่อเนื่องจำนวน 100 ประโยคจะประกอบไปด้วยท่า 208 ท่าดังกล่าว จากผลการทดลองนั้นทำให้ทราบได้ว่ารู้จำท่ามือจากผู้ใช้ 5 ใน 7 คน มีค่าเฉลี่ย 95.3% ปัญหาคือ ถ้า segment ผิดจะมีผลต่อเนื่องกันไป ซึ่งทำให้ความถูกต้องต่ำกว่าการใช้ฮิดเดนมาร์คอฟโมเดลเพียงอย่างเดียว

Wen Gao, Gaolin Fang, Debin Zhao และ Yiqiang Chen [2][3] ได้นำเสนอการพัฒนาวิธีการแก้ปัญหาคำศัพท์ที่มีจำนวนมากมายของท่าต่อเนื่อง ซึ่งเน้นการสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือด้วยการ Segment ในส่วนของการเคลื่อนไหวระหว่างท่ามือก่อน แล้วนำมา Clustering การเคลื่อนไหวระหว่างท่ามือด้วย Dynamic Time Warping (DTW) เมื่อได้ข้อมูลที่ Cluster แล้วก็นำมา train ให้แก่ระบบ แล้วใช้ Transition Models ที่ได้มาใช้ในการ segment ประโยคที่ต่อเนื่องให้เป็นท่าย่อยๆ ซึ่งได้ทำการทดลอง 25,565 ท่าจากท่าเดียว 5,113 ท่า โดยแต่ละท่าท่าละ 5 ครั้ง และเลือกข้อมูล 4 ชุดมาใช้ในกระบวนการสอนและอีกหนึ่งชุดที่เหลือใช้สำหรับการทดสอบ ซึ่งผลการทดลองมีความถูกต้อง 95.4% สำหรับท่าเดียว 5,113 ท่า และสำหรับท่าระหว่างมือจาก 1,500 ประโยค (เก็บข้อมูล 750 ประโยคเป็นจำนวน 2 ครั้ง) โดยข้อมูลชุดแรกใช้ในกระบวนการสอน และอีกชุดใช้สำหรับการทดสอบ ในหนึ่งประโยคมีค่าโดยเฉลี่ยคือ 6.6 คำ ซึ่งผลที่ได้คือมีความถูกต้อง 90.8% ซึ่งวิธีการนี้จะพบปัญหาคือ การ train ท่าเคลื่อนไหวระหว่างมือต้องใช้เวลา และต้อง train ให้ครอบคลุมจึงจะนำไปทำการ Cluster ได้

Jianjun Ye, Hongxun Yao และ Feng Jiang [4] ได้ใช้ฮิดเดนมาร์คอฟโมเดลร่วมกับ SVM (Support Vector Machines) รวมเป็น 2 ส่วนในการรู้จำบน Confusable set และได้ใช้ฮิดเดนมาร์คอฟโมเดลในการรู้จำท่าปกติ ซึ่งจะเป็นตัวกำหนดขอบเขตของท่าที่ไม่สามารถแยกได้ (Confusability) ในตอนแรก และใช้ SVM ในการแก้ปัญหาความไม่แน่นอนหลังจากที่ใช้ฮิดเดน

มาร์คอฟโมเดลในการรู้จำ ซึ่งในงานวิจัยนี้ได้พบว่าหากมีทำเป็นจำนวนมากๆ ก็จะมีบางทำที่มีความคล้ายคลึงกัน ซึ่งจะทำให้การแบ่งแยกมีคุณภาพลดลง จึงนำฮิดเดนมาร์คอฟโมเดลใช้ร่วมกับ SVM เพื่อให้การรู้จำมีประสิทธิภาพมากขึ้น หากมีข้อมูลของทำมือเข้ามาก็จะทำการรู้จำด้วยฮิดเดนมาร์คอฟโมเดล ถ้าข้อมูลไม่ได้อยู่ในกลุ่มของ Confusable set ก็จะพิจารณาว่าเป็นผลของการรู้จำได้เลย แต่ถ้าอยู่ในกลุ่มของ Confusable set ก็จะรู้จำทำมือจาก Confusable sign ใน Confusable set ซึ่งได้ทำการทดลองกับ 4,942 ทำ ซึ่งกระทำโดยผู้ทำทำจำนวน 6 คน (รวมเป็น 59,304 ทำ) โดยแบ่งข้อมูลจากผู้ทำทำคนละ 2 ชุด ซึ่งจาก 12 ชุด จะใช้ 1 ชุดสำหรับการทดสอบ และ 11 ชุดสำหรับกระบวนการสอน ซึ่งได้ผลการรู้จำ 89.40% ซึ่งในงานวิจัยนี้ยังไม่ได้นำมาใช้กับทำมือที่ต่อเนื่อง



รูปที่ 2.1 โค้ดแกรมของ OP pre-searching

Guilin Yao, Hongxun Yao, Xin Liu และ Feng Jiang [5] ได้นำเสนอถึงความเร็วในการรู้จำสำหรับจำนวนข้อมูลที่มีมากมาย ซึ่งจะมีการคิดเลขของเวลาเป็นระยะเวลานาน ดังนั้นจึงได้ใช้วิธีการ One-Pass (OP) ในการ pre-searching (แสดงดังรูปที่ 2.1) เพื่อค้นหาแต่ละ Sub-Cluster ที่เป็นไปได้เท่านั้น คือแบ่งประโยคให้เป็นส่วนย่อยด้วยการ Cluster ทำมือที่มีความคล้ายคลึงกันนั่นเอง ซึ่งได้ทำการหาถึงกลางของคำระหว่าง 2 คำแล้วทำการ cluster จะพบว่าความเร็วในการรู้จำและความถูกต้องนั้นมีความสัมพันธ์กันกับจำนวน Cluster ที่ได้จาก Temporal Clustering Algorithm และซึ่งการทดลองได้กระทำกับฐานข้อมูลขนาดใหญ่ โดยใช้ทำเดี่ยวจำนวน 4,942 ทำ โดยแต่ละทำทำจำนวน 4 ครั้ง จากผู้ทำทำ 2 คน ซึ่งจะใช้ข้อมูลทั้งหมดในกระบวนการสอน และทำต่อเนื่องจำนวน 543 ประโยค โดยกระทำประโยคละ 4 ครั้งเช่นกัน แล้วทำการเลือกข้อมูล 3 ชุดสำหรับการแบ่งกลุ่มและการ segment ด้วย Viterbi-beam และอีก 1 ชุดใช้ในการทดสอบ โดยแต่ละประโยค

ประกอบด้วยคำโดยเฉลี่ย 6.10 คำ ซึ่งผลการทดลองถ้ายิ่งแบ่งกลุ่มมากขึ้น ความเร็วในการรู้จำจะมีมากขึ้นแต่ความถูกต้องจะลดลงงานวิจัยนี้ได้พบว่ามีข้อจำกัดคือ ในการใช้ Center matching method บางการรู้จำจะผิดพลาดได้ และต้องไม่มีการเคลื่อนไหวระหว่างท่ามีอระหว่าง Center และจากการใช้ Transition factor ในการ Classify ของ OP-searching จะมีข้อดีคือ ทำให้การรู้จำแบบ real time ดีกว่าการใช้ Viterbi-beam algorithm เพียงอย่างเดียว เนื่องจาก error อาจเกิดขึ้นในขั้นตอนแรกคือ การ Classify ผิด Cluster

วุฒิชัย วิศาลคุณา [14] ได้นำเสนอการจดจำภาษามือไทยโดยใช้ฮิดเดนมาร์คอฟโมเดล โดยใช้ถุงมืออิเล็กทรอนิกส์ร่วมกับอุปกรณ์ตรวจวัดตำแหน่ง 6 แกนเป็นตัวรับข้อมูล แล้วนำข้อมูลไปผ่านขั้นตอนประมวลผลเบื้องต้น เพื่อทำการแบ่งข้อมูลออกเป็นชุด คือ รูปแบบของมือ ตำแหน่งของมือ การหมุนของมือ และการเคลื่อนที่ของมือ เพื่อทำการหาจุดเริ่มต้นและจุดสิ้นสุดของภาษามือท่านั้น ซึ่งจะเป็นการแบ่งแยกภาษามือที่เป็นท่าหนึ่งและท่าเคลื่อนไหว ซึ่งจะทำให้สามารถแบ่งแยกคำในการทำภาษามือที่ทำต่อเนื่องกันได้ แต่หากหาขอบเขตของภาษามือไม่ดี จะทำให้โมเดลภาษามือไทยที่ประมาณได้มีคุณภาพไม่ดี ในส่วนของฮิดเดนมาร์คอฟโมเดลนั้น ได้นำข้อมูลจากขั้นตอนการประมวลผลเบื้องต้นมาทำการสอนให้ระบบแล้วจึงทำการจดจำข้อมูลภาษามือชุดนั้น หลังจากทำการสอนให้ระบบจดจำภาษามือแล้ว เมื่อผู้ใช้ทำภาษามือลักษณะเดียวกับแบบที่ทำการสอนระบบจะสามารถบอกได้ว่าผู้ใช้ทำภาษามือคำใด หลังจากหาภาษามือที่เหมาะสมกับข้อมูลในขณะเวลาดังกล่าวได้แล้ว จึงนำข้อมูลที่ที่ได้มาเรียงกันเป็นประโยคที่ต้องการตามหลักไวยากรณ์และความหมาย โดยใช้หลักการทางการประมวลผลภาษามือธรรมชาติ ซึ่งงานวิจัยนี้ไม่ได้วิจัยในส่วนของการเคลื่อนไหวระหว่างท่ามือและเป็นแบบ Signer dependence

วรวิทย์ วีระพันธุ์ [6] ได้ทำการศึกษาท่ามือภาษาไทย ซึ่งท่ามือภาษาไทยนั้นแบ่งออกเป็น 2 ส่วน คือ ส่วนที่เป็นท่ามือนิ่ง (Static Gesture or Posture) และส่วนที่เป็นท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) ซึ่งปัญหาที่พบในการรู้จำท่ามือที่เป็นท่าต่อเนื่อง (Continuous Gesture) ก็คือ จะมีวิธีการอย่างไรที่จะสามารถแยกการเคลื่อนไหวของท่ามือ ซึ่งจะประกอบไปด้วยการเคลื่อนไหวของมือที่มีความหมาย (Meaningful Dynamic Gesture) และการเคลื่อนไหวของมือที่ไม่มีความหมายในตัวเอง (Transitional Gesture) นอกจากนั้นยังมีปัญหาเรื่องความเร็วในการทำท่ามือ, ตำแหน่งของท่ามือ และทิศทางองศาหันฝ่ามือ ซึ่งแต่ละคนจะมีการทำที่ไม่เหมือนกันหรือไม่เท่ากัน โดยในงานวิจัยนี้ยังได้นำเสนอวิธีการในการแก้ไขปัญหาเหล่านี้ โดยยึดหลักที่ว่าท่ามือภาษาไทยและท่ามือภาษาอื่น การเคลื่อนไหวของมือที่มีความหมายนั้น จะเป็นการเคลื่อนที่แบบซ้ำๆ (Periodic) ดังนั้นการแยกท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) ออกจากการเคลื่อนที่ของมือที่ไม่มีความหมายได้ โดยใช้ Fourier Analysis กับข้อมูลในแต่ละช่วงเวลา ซึ่งข้อมูลที่ได้โดยการใช้ถุงมืออิเล็กทรอนิกส์ที่เรียกว่า CyberGlove ในกรณีของท่ามือที่มีการเคลื่อนไหวของมือ (Dynamic Gesture) การประมาณค่าช่วงสัญญาณของท่ามือที่ได้จากการวิเคราะห์สามารถนำไปทำ Time Normalization ดังนั้น จึง

สามารถลดความผิดพลาดเนื่องจากการทำท่ามือที่มีความเร็วไม่เท่ากันของแต่ละคนได้ การทดลองจะใช้ข้อมูลที่เป็นท่ามือนิ่งและท่ามือที่มีการเคลื่อนไหวอย่างละ 100 ท่า และใช้วิธีการรู้จำที่เรียกว่า Bayesian Estimator ซึ่งผลการทดลองที่ได้ปรากฏว่ามีความถูกต้องในการรู้จำถึง 100% แต่วิธีการในงานวิจัยนี้ยังไม่สามารถรู้จำท่าเคลื่อนไหวแบบไม่ซ้ำได้

## บทที่ 3

# ทฤษฎีที่เกี่ยวข้องในงานวิจัย

### 3.1 ภาษามือไทย

#### 3.1.1 ความหมายและคำจำกัดความของภาษามือไทย

ด้วยเหตุที่คนหูหนวกไม่ได้ยินเสียงพูดเหมือนคนปกติ จึงไม่สามารถพูดได้ แต่สายตาของเขาปกติมองเห็นกิริยาอาการ ทำท่าต่างๆที่เคลื่อนไหวไปมาได้ ภาพต่างๆที่แลเห็นนั้นเป็นสื่อทำให้คนหูหนวกเรียนรู้ความหมาย แม้จะเข้าใจได้ไม่มากหรืออาจจะเข้าใจไม่ลึกซึ้งนัก แต่ก็เป็นส่วนหนึ่งที่มีอิทธิพลผลักดันให้คนหูหนวกพยายามใช้ท่าทาง ร่างกายและสีหน้า เพื่อแสดงความรู้สึกภายในของเขามืออยู่ให้คนอื่นเข้าใจความต้องการของเขาได้บ้าง ท่าทางที่แสดงนั้น เราจะสังเกตได้ว่าเป็นท่าทางที่เลียนแบบธรรมชาติมากที่สุด และจากท่าทางธรรมชาตินั้นเองได้มีการพัฒนาขึ้น โดยใช้มือทำท่าทางต่างๆเป็นส่วนใหญ่ ทำให้เกิดเป็นท่าทางใช้แทนความหมายในคำพูดของคนปกติได้ เราเรียกภาษาท่าทางที่ได้รับการพัฒนานั้นว่า “ภาษามือ” [7]

ภาษามือ คือ ภาษาที่ใช้มือ สีหน้า และกิริยาท่าทางประกอบในการสื่อความหมายและถ่ายทอดอารมณ์แทนการพูด ภาษามือของแต่ละชาติมีความแตกต่างกันเช่นเดียวกับภาษาพูด ซึ่งต่างกันตามขนบธรรมเนียมประเพณี วัฒนธรรมและลักษณะทางภูมิศาสตร์ เช่น ภาษามือจีน ภาษามืออเมริกัน และภาษามือไทย เป็นต้น ภาษามือเป็นภาษาที่นักการศึกษาทางด้านการศึกษาของคนหูหนวกตกลงยอมรับกันแล้วว่า เป็นภาษาหนึ่งสำหรับสื่อความหมายระหว่างคนหูหนวกกับคนหูหนวกด้วยกัน และระหว่างคนปกติกับคนหูหนวก ในภาษาอังกฤษเรียกว่า Sign Language หรือ Manual Communication

ภาษามือที่ใช้อยู่ในกลุ่มคนหูหนวกชาติหนึ่งๆ มักแบ่งออกเป็น 2 ประเภท ประเภทแรกคือ อย่างแรกคือ ภาษามือธรรมชาติ (Sign Language) ซึ่งคนหูหนวกเป็นผู้สร้างขึ้น และใช้ร่วมกันในแต่ละชุมชนหรือในแต่ละชาติ เช่น American Sign Language, British Sign Language, Swedish Sign Language เป็นต้น ซึ่งส่วนมากเป็นท่าเลียนแบบธรรมชาติ ที่จะช่วยคนหูหนวกให้มีพัฒนาการในภาษาประจำชาติเท่าเทียมกับคนปกติ ประเภทที่สองคือ ภาษามือประดิษฐ์ (Signed) คือ ภาษามือที่ครู ผู้ปกครอง หรือญาติมิตรของคนหูหนวกคิดขึ้นแทนภาษาพูดและภาษาเขียนประจำชาติ เพื่อให้มีคำใช้ให้เพียงพอในการศึกษา และการสื่อความหมาย โดยเฉพาะในเรื่องนามธรรม ภาษามือที่ประดิษฐ์ขึ้นนี้ บางทีเรียกว่า ภาษามือที่ใช้ในห้องเรียน หรือภาษามือที่ใช้ในการศึกษา ซึ่งเป็นภาษาที่ทำท่าคำแต่ละคำตามไวยากรณ์ภาษาพูด หรือภาษาเขียนของคนปกติ ภาษามือประดิษฐ์มักจะนำแบบสะกดนิ้วมือมาประสมด้วย

### 3.1.2 องค์ประกอบของภาษามือไทย

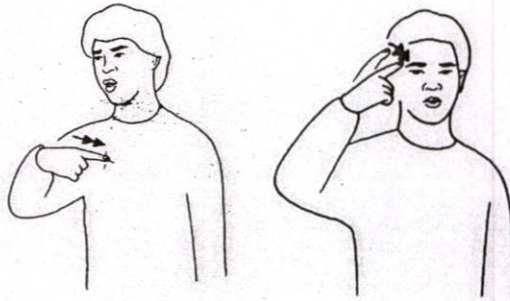
องค์ประกอบของภาษามือไทยประกอบไปด้วย 5 องค์ประกอบที่สำคัญ [8] คือ

3.1.2.1 รูปร่างมือ (handshape) คือ ลักษณะของมือต่างๆ เช่น การกำหมัด การแบมือ การชี้นิ้ว ในภาษามือไทยรูปร่างมือบางท่าอาจจะถือว่าเป็นรากศัพท์ เป็นต้น ดังรูปที่ 3.1



รูปที่ 3.1 ตัวอย่างรูปร่างมือในภาษามือไทย

3.1.2.2 ตำแหน่งของมือ (location) คือ การใช้ท่ามือโดยที่ตำแหน่งของมืออยู่ในตำแหน่งต่างๆ กันไป จะทำให้ความหมายที่แตกต่างกัน ถึงแม้ว่ารูปร่างมือจะเป็นท่าเดียวกันก็ตาม เช่น การใช้ นิ้วชี้ที่หน้าอก มีความหมายว่า “ฉัน” แต่ถ้าหากเปลี่ยนเป็นที่ขมับ มีความหมายว่า “รู้” เป็นต้น ดังรูปที่ 3.2

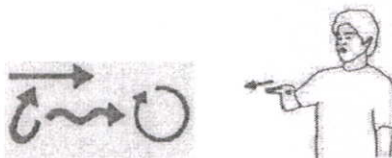


(a) (b)

รูปที่ 3.2 (a) ท่ามือที่มีความหมายว่า “ฉัน” (b) ท่ามือที่มีความหมายว่า “รู้”

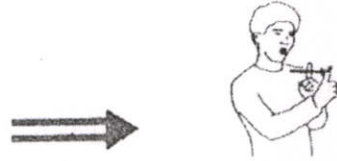
3.1.2.3 การเคลื่อนไหว (movement) คือ การเคลื่อนไหวร่างกาย, ใบหน้าและมือของผู้พูด ในขณะที่สื่อสาร เช่น การเคลื่อนมือขึ้น-ลง, การเอนลำตัวเข้า-ออก จากตัวผู้ฟัง, การพยักหน้า, การหมุนมือ เป็นต้น โดยการเคลื่อนไหวในภาษามือไทยแบ่งการเคลื่อนไหวเป็น 3 กลุ่ม คือ

1. จังหวะของการเคลื่อนไหวขั้นพื้นฐาน (Rhythms of basic movements)
  - ลูกศรธรรมดาแสดงถึงทิศทางและระยะทางของการเคลื่อนไหว โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.3



รูปที่ 3.3 ลักษณะการเคลื่อนไหว และทิศทางของการเคลื่อนไหวปกติ

- การเคลื่อนไหวอย่างรวดเร็ว โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.4



รูปที่ 3.4 สัญลักษณ์แทนการเคลื่อนไหว และทิศทางของการเคลื่อนไหวอย่างรวดเร็ว

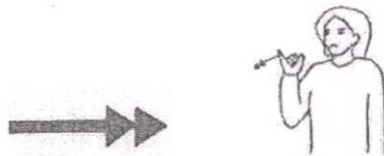
- การเคลื่อนที่จากจุดหนึ่ง ไปยังอีกจุดหนึ่งโดยก้าวสั้นๆ ในแต่ละการก้าวจะต้องมีการหยุดนิ่งก่อนที่จะก้าวไปยังจุดถัดไป โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.5



รูปที่ 3.5 สัญลักษณ์แทนการเคลื่อนที่จากจุดหนึ่ง ไปยังอีกจุดหนึ่ง โดยก้าวสั้นๆ

2. การเคลื่อนไหวที่ซ้ำๆ กัน (Repeated movements)

- การเคลื่อนที่ไปในทางเดียวหลายๆ ครั้ง โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.6



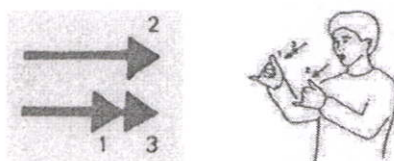
รูปที่ 3.6 สัญลักษณ์แทนการเคลื่อนที่ไปในทางเดียวหลายๆ ครั้ง

- การเคลื่อนไหวที่ไป-กลับหลายๆ ครั้ง โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.7



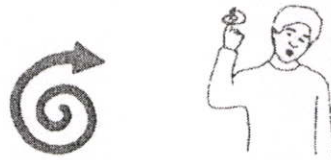
รูปที่ 3.7 สัญลักษณ์แทนการเคลื่อนไหวที่ไป-กลับหลายๆ ครั้ง

- การที่มือทั้งสองเคลื่อนสลับกันในทิศทางเดียวกัน (ซ้ายครั้ง ขวาครั้ง) โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.8



รูปที่ 3.8 สัญลักษณ์แทนการเคลื่อนไหวที่มือทั้งสองเคลื่อนสลับกัน

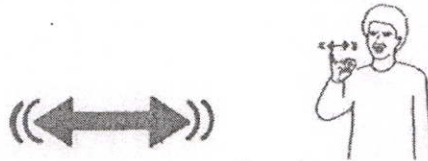
- การหมุนข้อมือหรือส่วนของแขนระหว่างข้อศอกกับมือเป็นวงกลมหลายครั้ง (เป็นรูปใกล้เคียงวงกลมซึ่งง่ายที่สุดสำหรับที่กล้ามเนื้อจะสร้างขึ้นมาได้) โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.9



รูปที่ 3.9 สัญลักษณ์แทนการหมุนข้อมือหรือส่วนของแขนระหว่างข้อศอกกับมือเป็นวงกลม

3. การเคลื่อนไหวที่มีการสั่น การขยับและการหมุนปลายแขน (Vibrating and nodding movements, and forearm rotation)

- การเคลื่อนที่ไปสองทิศทางที่มีการสั่น (สั่นและเร็วมาก) โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.10



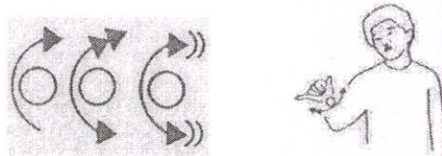
รูปที่ 3.10 สัญลักษณ์แทนการเคลื่อนที่ไปสองทิศทางที่มีการสั่น

- การตะเบาๆ หลายครั้ง โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.11



รูปที่ 3.11 สัญลักษณ์แทนการตะเบาๆ หลายครั้ง

- การหมุนปลายแขนโดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.12



รูปที่ 3.12 สัญลักษณ์แทนการหมุนปลายแขน

4. การงอนิ้วและการเปลี่ยนท่ามือ (Finger bending and changes of handshape)

- การเคลื่อนไหวที่มีการปิดเข้าเล็กน้อย นิ้วทุกนิ้วเกือบจะรวมเข้าด้วยกัน, การกระพริบ โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.13



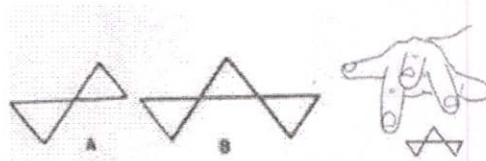
รูปที่ 3.13 สัญลักษณ์แทนการเคลื่อนไหวที่มีการปิดเข้าเล็กน้อย, การกระพริบ

- ปลายนิ้วคลายออกจากได้นิ้วหัวแม่มือ หมายถึงประกายไฟและกระแสไฟฟ้า โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.14



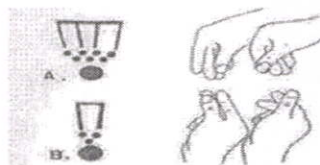
รูปที่ 3.14 สัญลักษณ์แทนปลายนิ้วคลายออกจากได้นิ้วหัวแม่มือ

- การกระดิกนิ้วในลักษณะท่าเล่นเปียโน การพิมพ์ดีด โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.15



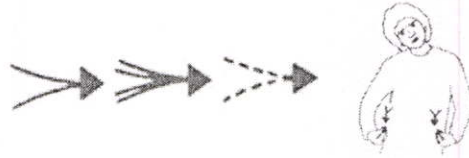
รูปที่ 3.15 สัญลักษณ์แทนการกระดิกนิ้วในลักษณะท่าเล่นเปียโน การพิมพ์ดีด (A) กระดิก 2 นิ้ว (B) กระดิก 5 นิ้ว

- สัญลักษณ์นี้แสดงถึงสิ่งที่มีลักษณะละเอียดหรือเป็นผง โดยให้นิ้วมือทั้งสี่ถูกกับนิ้วหัวแม่มือหลายครั้ง (สัญลักษณ์ที่สองใช้นิ้วชี้ นิ้วกลาง และนิ้วหัวแม่มือเท่านั้น) โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.16



รูปที่ 3.16 สัญลักษณ์แทนสิ่งที่มีลักษณะละเอียดหรือเป็นผง (A) นิ้วมือทั้งสี่ถูกกับนิ้วหัวแม่มือหลายครั้ง (B) ใช้นิ้วชี้ นิ้วกลาง และนิ้วหัวแม่มือเท่านั้น

- การเคลื่อนไหวที่เริ่มจากท่าทางนิ้วแล้วจบลง โดยการขยับปลายนิ้วทั้งหมดเข้าหากัน โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.17



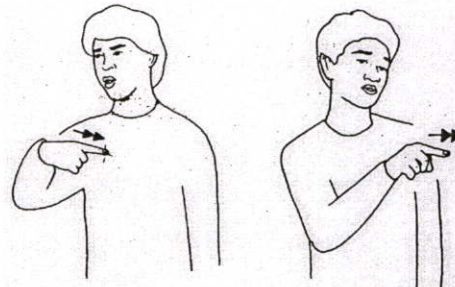
รูปที่ 3.17 สัญลักษณ์แทนการเคลื่อนไหวที่เริ่มจากท่าทางนิ้ว

- การใช้นิ้วหัวแม่มือสัมผัสนิ้วอื่นครั้งหนึ่ง โดยเริ่มจากนิ้วก้อยจนถึงนิ้วชี้ โดยใช้สัญลักษณ์แทนได้ดังรูปที่ 3.18



รูปที่ 3.18 สัญลักษณ์แทนการใช้นิ้วหัวแม่มือสัมผัสนิ้วอื่นครั้งหนึ่ง

3.1.2.4 การหันของฝ่ามือ (palm orientation) คือ การที่ฝ่ามือหันเข้าหาตัวหรือหันออกจากตัว เช่น คำว่า “ฉัน” กับคำว่า “คุณ” เป็นต้น ดังรูป 3.19 สังเกตว่ารูปมือเป็นรูปมือเดียวกันแต่ลักษณะการหันของมือนั้นต่างกันจึงทำให้มีความหมายที่ต่างกัน



(a)

(b)

รูปที่ 3.19 (a) ท่ามือที่มีความหมายว่า “ฉัน” (b) ท่ามือที่มีความหมายว่า “เธอ”

3.1.2.5 การแสดงออกทางสีหน้า (facial expression) คือ แววตาสีหน้าประกอบการทำท่าทาง เช่น การเลิกคิ้ว, การพองลมที่แก้ม, การทำสีหน้าตกใจ เป็นต้น ซึ่งถือว่าเป็นองค์ประกอบที่สำคัญในการแสดงท่ามือ เช่น ประโยคบอกเล่า แสดงสีหน้าปกติ, ประโยคคำถามแสดงสีหน้าบ่งบอกถึงความสงสัยโดยการยกคิ้วสูง หรือเป็นประโยคคำสั่งโดยการจ้องตาคู่สนทนา เป็นต้น

## 3.2 อุปกรณ์ที่ใช้ในงานวิจัยภาษามือ

### 3.2.1 ประเภทของอุปกรณ์นำเข้าข้อมูล

งานวิจัยภาษามือในปัจจุบันสามารถแบ่งได้ออกตามเทคโนโลยีอุปกรณ์การรับข้อมูล โดยสามารถแบ่งอุปกรณ์ได้ 2 ชนิด คือ

1.1.1.1 ถุงมืออิเล็กทรอนิกส์ โดยใช้ร่วมกับอุปกรณ์บอกตำแหน่งแบบ 6 แกน (Glove-based) งานวิจัยจะใช้อุปกรณ์เซนเซอร์ ใช้วัดความเปลี่ยนแปลงซึ่งเกิดขึ้นจากการบิดหรือองของข้อต่อต่างๆ จากสรีระของฝ่ามือ รวมถึงวัดตำแหน่งของมือโดยใช้การเปลี่ยนแปลงของตำแหน่งโดยจะแบ่งออกตามเทคโนโลยีว่าจะใช้ในลักษณะใด เช่น ใช้คลื่นแม่เหล็กไฟฟ้า คลื่นความถี่เหนือเสียง หรือใช้อินฟราเรด เป็นต้น

1.1.1.2 กล้องวิดีโอ (Vision-based) การวิจัยใช้อุปกรณ์วิดีโอในการเก็บข้อมูลเป็นรูปภาพของผู้ทำท่ามือ ในบางงานวิจัยก็มีการใช้กล้องวิดีโอมากกว่า 2 ตัวเพื่อให้ได้ข้อมูลมากขึ้น ในบางงานวิจัยใช้การใส่ถุงมือที่มีสีประกอบ เมื่อได้ข้อมูลมาแล้วจำเป็นต้องทำการ Preprocessing ก่อนเพื่อให้ได้ข้อมูลมือที่ต้องการและนำเอาข้อมูลมาประมวลผลเพื่อหาความหมายต่อไป

### 3.2.2 เปรียบเทียบข้อดีและข้อเสียของอุปกรณ์นำเข้าข้อมูล [9]

การเก็บข้อมูลโดยใช้อุปกรณ์ทั้งสองมีข้อดีและข้อเสียต่างๆ กันไปโดยเปรียบเทียบได้เป็น 9 หัวข้อ ดังต่อไปนี้

3.2.2.1 ราคา เมื่อเปรียบเทียบถึงราคาระหว่างอุปกรณ์ทั้งสองชนิดนี้ จะเห็นได้ว่าอุปกรณ์ที่เป็นถุงมืออิเล็กทรอนิกส์จะมีราคาที่สูงกว่า กล้องวิดีโอมาก แต่ก็จะได้คุณภาพของข้อมูลที่สูงเช่นเดียวกัน

3.2.2.2 ความสะดวกสบาย อุปกรณ์ที่เป็นถุงมือต้องมีการสวมใส่และยังต้องมีการเชื่อมต่อกับอุปกรณ์คอมพิวเตอร์ทำให้ถูกจำกัดการเคลื่อนไหว ซึ่งต่างจากการใช้กล้องวิดีโอ คือ สามารถใช้งานได้สะดวกกว่าและเคลื่อนไหวได้ง่ายกว่าเนื่องจากไม่จำเป็นต้องสวมใส่อุปกรณ์เพื่อใช้ในการวัดข้อมูล

3.2.2.3 ความต้องการพลังในการประมวลผล เมื่อเปรียบเทียบอัลกอริทึมที่ใช้ในการดึงเอาข้อมูลที่ต้องการจากถุงมือ และวิดีโอ สรุปได้ว่าการนำเอาข้อมูลจากถุงมือมีความต้องการพลังในการประมวลผลที่ต่ำกว่ามาก อีกทั้งข้อมูลที่ได้ยังเก็บได้ง่ายกว่าแบบที่ใช้กล้องวิดีโอ

3.2.2.4 สรีระของมือ ปัญหาจะเกิดขึ้นกับถุงมืออิเล็กทรอนิกส์เมื่อเซนเซอร์ไม่ตรงกับข้อมือจะทำให้ค่ามีการผิดพลาดบ้าง แต่ถ้าผู้ใช้มีขนาดของมือที่ต่างกันข้อมูลที่ได้จากถุงมืออิเล็กทรอนิกส์จะแตกต่างกัน

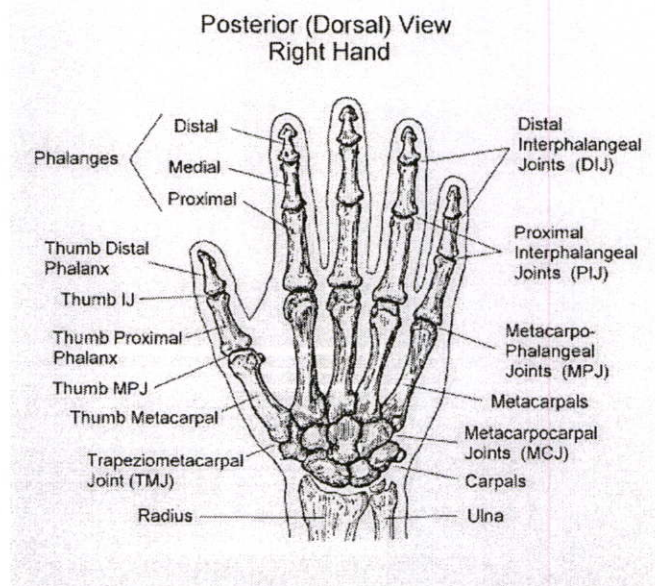
3.2.2.5 สัญญาณรบกวน (Noise) เมื่อเปรียบเทียบกันในเรื่องของสัญญาณรบกวนในถุงมืออิเล็กทรอนิกส์จำเป็นต้องมีอัลกอริทึมเพื่อมาลดทอนสัญญาณรบกวนและการสั่นของสัญญาณ ในถุงมืออิเล็กทรอนิกส์บางบริษัทได้จัดการกับปัญหานี้แล้ว ส่วนสัญญาณรบกวนที่เกิดขึ้นกับกล้อง

วิถีโอมิมากและแก้ไขได้ยากกว่า เช่น แสงสว่าง, ความแตกต่างของสี เป็นต้น แต่ถ้าหากมีการติดตั้งระบบได้ดี และมีความเข้มของแสงเพียงพอ ปัญหาเรื่องสัญญาณรบกวนดังกล่าวก็จะน้อยลง

3.2.2.6 ความละเอียด (Accuracy) ความละเอียดของข้อมูลเป็นความจำเป็นมากเนื่องจากถ้าได้ข้อมูลที่มีความละเอียดสูงในการรู้จำ ก็มีความเป็นไปได้สูงว่าจะมีความถูกต้องมากขึ้น เมื่อเปรียบเทียบข้อมูลที่ได้จากถุงมืออิเล็กทรอนิกส์กับกล้องวิดีโอแล้วจะเห็นได้ว่า ความถี่ของข้อมูลของถุงมืออิเล็กทรอนิกส์สูงกว่ามาก ถึงแม้ว่าจะมีความถี่ของข้อมูลเท่ากัน ก็ยังต้องคำนึงถึงความต้องการการประมวลผลอีกเนื่องจากการใช้กล้องวิดีโอก่อนที่จะนำเอาข้อมูลมาใช้ได้ต้องการการประมวลผลเบื้องต้นก่อนซึ่งอาจทำให้ข้อมูลที่สูญหายอย่างหายไป

### 3.2.3 สรีระของมือ

มือของมนุษย์เป็นส่วนที่มีความซับซ้อนที่สุดเนื่องจากเป็นอวัยวะที่มนุษย์ใช้งานมากที่สุด มือประกอบไปด้วยข้อต่อที่สำคัญดังรูปที่ 3.20 ประกอบไปด้วยกระดูกและเส้นเอ็นมากมาย รวมถึงเส้นประสาทที่มีความซับซ้อนสูง ข้อต่อต่างๆ ที่สำคัญในมือมีดังต่อไปนี้



รูปที่ 3.20 แสดงโครงสร้างทางกายวิภาคมือขวาของมนุษย์

3.2.3.1 MPJ (Metacarpo-Phalangeal Joint) เป็นข้อต่อที่เชื่อมระหว่างกระดูกนิ้วข้อล่าง (Proximal Phalanx) กับกระดูกฝ่ามือ (Metacarpal) เข้าด้วยกัน

3.2.3.2 PIJ (Proximal Interphalangeal Joint) เป็นข้อต่อที่อยู่ถัดจากข้อต่อ MPJ ขึ้นมาหนึ่งข้อหรืออาจเรียกได้ว่าเป็นส่วนที่เชื่อมระหว่างกระดูกที่ติดกับโคนนิ้วกับกระดูกนิ้วข้อกลางเข้าด้วยกัน

3.2.3.3 DIJ (Distal Interphalangeal Joint ) เป็นข้อต่อที่อยู่ถัดจากข้อต่อ PIJ ขึ้น ไปเพื่อเชื่อมระหว่างกระดูกนิ้วข้อกลางกับกระดูกปลายนิ้วเข้าด้วยกัน

3.2.3.4 TMJ (Trapeziometacarpal Joint ) เป็นข้อต่อมีอยู่ที่นิ้วหัวแม่มือถัดจากข้อต่อ MPJ ของนิ้วหัวแม่มือลงมา มีลักษณะเป็นแอ่งที่ฐานของนิ้วหัวแม่มือเข้าใกล้กับข้อมือ

### 3.2.4 ถุงมืออิเล็กทรอนิกส์ [10]

อุปกรณ์ถุงมืออิเล็กทรอนิกส์เป็นอุปกรณ์ถุงมือมีลักษณะดังรูปที่ 3.21 ที่มีตัวต้านทานปรับค่าได้ฝังในถุงมืออยู่ในตำแหน่งตามข้อต่อที่มีจุดเปลี่ยนบนมือและข้อมือนั้นได้ใช้ถุงมือของบริษัทอิมเมอร์ชัน โควออปเปอร์เรต (Immersion Corporate) ทั้งชายและหญิง ถุงมือสองข้างจะมีความแตกต่างกันเล็กน้อย ในถุงมือข้างขวาจะมีเซนเซอร์ติดอยู่ทั้งหมด 22 ชิ้น โดยจะอยู่ในที่ต่างๆ กันตามข้อต่อต่างๆ บนฝ่ามือ ค่าที่ได้จะเป็นค่าจากเซนเซอร์ ซึ่งในเซนเซอร์จะเป็นตัวต้านทานที่ปรับค่าได้ เมื่อมีการเปลี่ยนแปลงค่าที่ได้ก็จะเปลี่ยนแปลงไปด้วย เซนเซอร์ติดอยู่ในตำแหน่งต่างๆ ของมือ ตำแหน่งของตัวรับข้อมูลในถุงมืออิเล็กทรอนิกส์ที่ใช้ในงานวิจัยมีดังต่อไปนี้

นิ้วหัวแม่มือ	TMJ, MPJ และ IJ
นิ้วชี้	MPJ ,PIJ
นิ้วกลาง	MPJ ,PIJ
นิ้วนาง	MPJ ,PIJ
นิ้วก้อย	MPJ ,PIJ
ง่ามนิ้ว	หัวแม่มือและนิ้วชี้, นิ้วชี้และนิ้วกลาง, นิ้วกลางและนิ้วนาง, นิ้วนางและนิ้วก้อย
ฝ่ามือ	การงอของฝ่ามือระหว่างนิ้วหัวแม่มือกับนิ้วก้อย
ข้อมือ	บิดข้อมือไปข้างหน้า (Pitch) และ บิดข้อมือไปข้างๆ (Yaw)

ค่าที่ได้จากเซนเซอร์จะอยู่ระหว่าง 0-255 มีค่าขึ้นอยู่กับสรีระของผู้สวมใส่ถุงมือ เซนเซอร์ที่ถูกตัดทิ้งจากถุงมือ คือ ตำแหน่ง DIJ ของนิ้วชี้ นิ้วกลาง นิ้วนาง และนิ้วก้อย เนื่องจากในตำแหน่งดังกล่าวถ้าไม่มีการหยิบจับวัตถุ ค่าที่เกิดขึ้นสามารถคำนวณได้โดยใช้สูตรดังสมการที่ 3.1 [4]

$$DIJ = \frac{2}{3} PIJ \quad (3.1)$$

## สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง



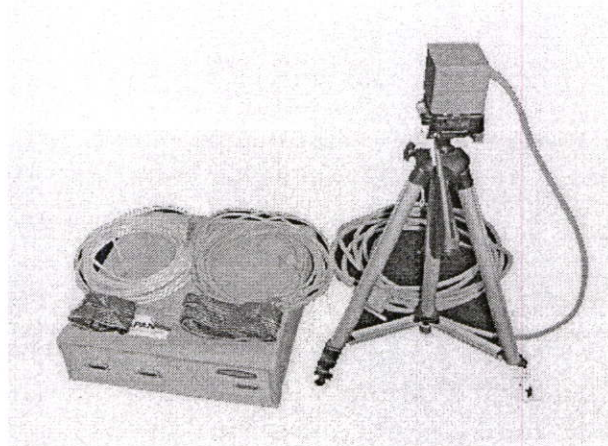
รูปที่ 3.21 อุปกรณ์ถุงมืออิเล็กทรอนิกส์

### 3.2.5 เครื่องมือตรวจวัดตำแหน่งแบบ 6 แกน [11]

อุปกรณ์เครื่องมือตรวจวัดตำแหน่งแบบ 6 แกน เป็นเครื่องมือที่อาศัยหลักการของสนามแม่เหล็กไฟฟ้าเพื่อตรวจวัดตำแหน่งของตัวรับข้อมูลว่าอยู่ในตำแหน่งใดๆ อุปกรณ์จะประกอบไปด้วยตัวส่งคลื่นแม่เหล็กไฟฟ้า และตัวรับคลื่นแม่เหล็กไฟฟ้า ดังรูปที่ 3.22 ในการใช้อุปกรณ์นี้ ตัวรับข้อมูลจะถูกติดกับข้อมือทั้งสองข้างรวมเข้ากับอุปกรณ์ถุงมืออิเล็กทรอนิกส์ อุปกรณ์นี้จะส่งข้อมูล 6 ค่า โดยแบ่งเป็นสองส่วน คือ

ค่าพิกัดในแนวแกน จะประกอบไปด้วยค่าตำแหน่งของตัวรับข้อมูล 3 ค่า ในแนวแกน X, Y และ Z โดยอ้างอิงกับแนวแกนของอุปกรณ์ตัวส่งคลื่นแม่เหล็กไฟฟ้า โดยค่าพิกัดในแนวแกนจะมีหน่วยวัดเป็นเซนติเมตร

ค่าองศาหมุนรอบตัวเองของเซนเซอร์นั้นจะประกอบไปด้วยค่าองศา 3 ค่าได้แก่ ค่าองศาหมุนในแนวระนาบ (yaw), ค่าองศาหมุนเบย (pitch) และค่าองศาหมุนของการม้วน (roll) โดยแนวแกนของเซนเซอร์จะเทียบกับทิศทางแนวแกนของอุปกรณ์ส่งคลื่นแม่เหล็กไฟฟ้า ในแต่ละค่าจะมีหน่วยเป็นเรเดียน



รูปที่ 3.22 อุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกนรุ่น 3SPACE® ISOTRAK II™ พร้อมด้วยเซนเซอร์ 2 ตัว และ Transmitter

### 3.3 ทฤษฎีแบบจำลองฮิดเดนมาร์คอฟโมเดล (Hidden Markov Model: HMM) [13]

แบบจำลองฮิดเดนมาร์คอฟโมเดลเป็นแบบจำลองทางสถิติ ซึ่งพัฒนาเพื่อแบ่งกลุ่มของอนุกรมทางเวลาหรือสัญญาณไม่คงที่ นั่นคือใช้สำหรับจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknown signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลองฮิดเดนมาร์คอฟโมเดลได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียงพูด (Speech Recognition) หรือแม้แต่การรู้จำท่าทาง (Gesture Recognition)

#### 3.3.1 ส่วนประกอบของแบบจำลองฮิดเดนมาร์คอฟโมเดล

แบบจำลองฮิดเดนมาร์คอฟโมเดลมีพารามิเตอร์ที่สำคัญ ที่ใช้ในการสร้างแบบจำลองในการรู้จำ โดยมีค่าพารามิเตอร์ต่างๆดังนี้

-  $T$  คือ ความยาวของลำดับข้อมูล ซึ่งมีขนาดความยาวของลำดับเท่ากับจำนวนเฟรมทั้งหมดของข้อมูลที่เกี่ยวข้องๆหนึ่ง ซึ่งจะใช้เป็นข้อมูลอินพุทในส่วนของ HMM ซึ่งเรียกแทนว่า “ลำดับของค่าปรากฏ” (Observation sequence)

-  $N$  คือ จำนวนสเตทในแบบจำลอง ถ้ากำหนดให้เซตของสเตทเป็น  $\{1, 2, \dots, N\}$  จะสามารถแทนสเตทที่เปลี่ยนไปตามเวลา  $t$  ด้วยเซตของ  $Q = \{q_1, q_2, \dots, q_N\}$

-  $M$  คือ จำนวนของค่าปรากฏที่สามารถเป็นไปได้ต่อหนึ่งสเตท แทนสัญลักษณ์ด้วย  $V = \{v_1, v_2, \dots, v_M\}$  ซึ่งจากการจัดระดับของการเปลี่ยนแปลงของความถี่ ( $\Delta F$ ) ออกเป็น 3 ระดับ จะได้เซตของค่าปรากฏที่สามารถเป็นไปได้ในแต่ละสเตทมีค่าเป็น  $V = \{-1, 0, 1\}$

- ค่าความน่าจะเป็นในการย้ายสเตท:  $A = \{a_{ij}\}$

โดย  $a_{ij}$  แทนการย้ายสเตทจาก  $i$  ไป  $j$

เมื่อ

$$a_{ij} = P [q_t = j \mid q_{t-1} = i] \quad ; 1 \leq i, j \leq N \quad (3.2)$$

- การกระจายความน่าจะเป็น ของค่าปรากฏที่สามารถเป็นไปได้ภายในสเตท:  $B = \{b_j(k)\}$

$$\text{โดยที่} \quad b_j(k) = P [v_k \text{ ที่เวลา } t \mid q_j \text{ ที่เวลา } t] \quad ; 1 \leq k \leq M \quad (3.3)$$

เป็นนิยามการกระจายสัญลักษณ์ในสเตท  $j$  เมื่อ  $j = 1, 2, \dots, N$

- ค่าความน่าจะเป็นของการเป็นสเตทเริ่มต้น:  $\pi = \{\pi_i\}$

$$\text{เมื่อ} \quad \pi_i = P [q_1 \text{ ที่เวลา } t=1] \quad ; 1 \leq i \leq N \quad (3.4)$$

จะเห็นว่าแบบจำลองฮิดเดนมาร์คอฟโมเดลต้องการพารามิเตอร์ของแบบจำลองคือ  $N, M$  และกลุ่มของความน่าจะเป็น  $A, B, \pi$  ดังนั้นในการแสดงเซตของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองอ้างอิง จะแทนด้วยสัญลักษณ์

$$\lambda = (A, B, \pi) \quad (3.5)$$

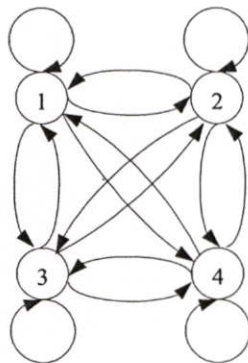
#### 3.3.2 ชนิดของแบบจำลองฮิดเดนมาร์คอฟโมเดล

แบ่งตามการย้ายสเตทของเมตริกซ์ A

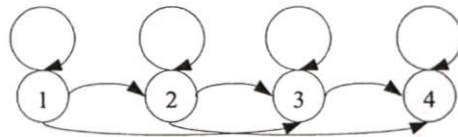
1. แบบ Ergodic Model หรือ Fully Connected Model

การย้ายสเตทสามารถย้ายไปยังทุกๆ สเตทของแบบจำลอง ดังรูปที่ 3.23 (a) เป็นตัวอย่างของแบบจำลองที่มี  $N = 4$  ซึ่งจากรูปนี้มีค่าของเมตริกซ์ A เป็น

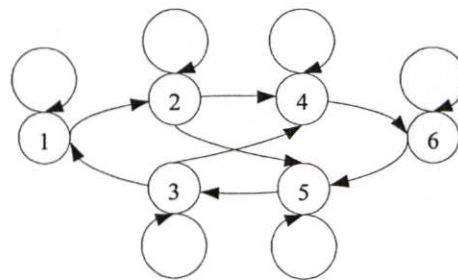
$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \tag{3.6}$$



(a)



(b)



(c)

รูปที่ 3.23 แบบจำลองชนิดต่างๆของแบบจำลองฮิดเดนมาร์คอฟโมเดล

2. แบบ Left-Right Model หรือ Bakis Model

การย้ายสเตทจะย้ายจากซ้ายไปขวา ซึ่งจะมีคุณสมบัติของสัมประสิทธิ์ในการย้ายสเตท

ดังนี้

$$a_{ij} = 0, j < i \tag{3.7}$$

จะไม่มีการย้ายสเตตไปยังสเตตที่ต่ำกว่าสเตตปัจจุบัน และนอกจากนี้ก็ยังมีความน่าจะเป็นของสเตตเริ่มต้นดังนี้

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (3.8)$$

ลำดับของสเตตจะต้องเริ่มที่สเตตที่ 1 เสมอ และ Left-Right Model นี้มักมีกฎบังคับการย้ายสเตต เพื่อไม่ให้มีการเปลี่ยนแปลงดัชนีของสเตตมากนัก กล่าวคือ

$$a_{ij} = 0, \quad j < i + \Delta i \quad (3.8)$$

ดังรูปที่ 3.23 (b) ค่าของ  $\Delta i = 2$  นั่นคือจะไม่มีการย้ายข้ามสเตตไปเกิน 2 สเตต และมีเมตริกซ์ในการย้ายสเตตเป็น

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.9)$$

จะเห็นว่าสเตตสุดท้าย สัมประสิทธิ์การย้ายสเตตจะเป็น

$$a_{NN} = 1 \quad (3.10)$$

$$a_{Nj} = 0, \quad i < N \quad (3.11)$$

แบบจำลองแบบนี้ จะเหมาะกับสัญญาณที่มีลักษณะเปลี่ยนแปลงตามเวลาอย่างต่อเนื่อง

### 3. แบบ Parallel Left-Right Model

เป็นแบบจำลองที่มีความยืดหยุ่นมากกว่าแบบที่ 2 แสดงได้ดังรูปที่ 3.23 (c)

#### 3.3.3 ปัญหาพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟโมเดล

ปัญหาของแบบจำลองฮิดเดนมาร์คอฟโมเดลมี 3 ข้อ ซึ่งต้องใช้อัลกอริทึมวิธีต่างๆ ในการคำนวณเพื่อแก้ปัญหา

**ปัญหาที่ 1** เมื่อมีลำดับของค่าปรากฏ  $O = \{O_1 O_2 O_3 \dots O_T\}$  และมีแบบจำลอง  $\lambda = (A, B, \pi)$  จะคำนวณหาความน่าจะเป็น  $P(O|\lambda)$  ของลำดับค่าปรากฏนั้นได้อย่างไร

**ปัญหาที่ 2** เมื่อมีลำดับของค่าปรากฏ  $O = \{O_1 O_2 O_3 \dots O_T\}$  และมีแบบจำลอง  $\lambda = (A, B, \pi)$  จะคำนวณหาลำดับสเตต  $q = \{q_1 q_2 q_3 \dots q_T\}$  ที่เหมาะสมกับลำดับค่าปรากฏนั้นได้อย่างไร

**ปัญหาที่ 3** เราจะปรับพารามิเตอร์ของแบบจำลอง  $\lambda = (A, B, \pi)$  เพื่อให้ได้ค่า  $P(O|\lambda)$  สูงสุดได้อย่างไร

การคำนวณเพื่อแก้ปัญหาของแบบจำลองฮิดเดนมาร์คอฟโมเดล

**การแก้ปัญหาที่ 1** เป็นการคำนวณหาว่าแบบจำลอง  $\lambda$  ใดๆ มีโอกาสจะให้ค่าลำดับเป็นไปตามลำดับของค่าปรากฏนั้น ด้วยค่าความน่าจะเป็นมากน้อยเท่าใด

การแก้ปัญหาสามารถทำได้โดยระบุสเตตให้กับลำดับของค่าปรากฏซึ่งยาว  $T$  (โดยที่ค่าปรากฏหนึ่งตัวมีความเป็นไปได้ที่จะอยู่ในสเตตได้  $N$  สเตต) ซึ่งสามารถเป็นไปได้ถึง  $N^T$  แบบ ซึ่งให้สเตตต่างๆ แทนด้วย

$$q = q_1 q_2 q_3 \dots q_T \quad (3.12)$$

เมื่อ  $q_1$  เป็นสเตตเริ่มต้นที่เวลา  $t = 1$  ความน่าจะเป็นของลำดับค่าปรากฏ  $O$  ที่กำหนดคือ

$$P(O | q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) \quad (3.13a)$$

ความน่าจะเป็นในการเกิดค่าปรากฏคือ

$$P(O | q, \lambda) = b_{q_1} O_1 \cdot b_{q_2} O_2 \cdot \dots \cdot b_{q_T} O_T \quad (3.13b)$$

และความน่าจะเป็นในการย้ายข้ามสเตต  $q$  จะเป็น

$$P(q | \lambda) = \pi_{q_1} \cdot a_{q_1 q_2} \cdot a_{q_2 q_3} \cdot \dots \cdot a_{q_{T-1} q_T} \quad (3.14)$$

ดังนั้นเมื่อนำความน่าจะเป็นของการเกิดค่าปรากฏ  $O$  และค่าความน่าจะเป็นในการย้ายสเตต  $q$  มารวมกัน ซึ่งนั่นก็คือความน่าจะเป็นที่  $O$  และ  $q$  จะเกิดขึ้นพร้อมกัน จะได้

$$\begin{aligned} P(O, q | \lambda) &= P(O | q, \lambda) P(q | \lambda) \\ &= (b_{q_1} O_1 \cdot b_{q_2} O_2 \cdot \dots \cdot b_{q_T} O_T) (\pi_{q_1} \cdot a_{q_1 q_2} \cdot a_{q_2 q_3} \cdot \dots \cdot a_{q_{T-1} q_T}) \end{aligned} \quad (3.15)$$

โดยที่ความน่าจะเป็นของ  $O$  ได้มาจากผลรวมของความน่าจะเป็นที่  $O$  และ  $q$  เกิดขึ้นพร้อมกัน โดยคิดจากทุกสเตต  $q$  ที่จะเป็นไปได้ดังนี้

$$P(O | \lambda) = \sum_{\text{all } q} P(O | q, \lambda) P(q | \lambda) \quad (3.16)$$

$$= \sum_{q_1 q_2 \dots q_T} \pi_{q_1} b_{q_1} (O_1) a_{q_1 q_2} b_{q_2} (O_2) \dots a_{q_{T-1} q_T} b_{q_T} O_T \quad (3.17)$$

ที่เวลาเริ่มต้น ( $t = 1$ ) เราจะอยู่ที่สเตต  $q_1$  ด้วยค่าความน่าจะเป็น  $\pi_{q_1}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏ  $O_1$  ที่สเตตนี้ด้วย  $b_{q_1} O_1$

ที่เวลาเพิ่มขึ้นจาก  $t$  ไปยัง  $t+1$  ( $t=2$ ) เราแทนการย้ายสเตตจากสเตต  $q_1$  ไปยัง  $q_2$  ด้วยค่าความน่าจะเป็น  $a_{q_1 q_2}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น  $O_2$  ด้วยค่าความน่าจะเป็น  $b_{q_2} O_2$

จนกระทั่ง ณ เวลา  $T$  เราแทนการย้ายสเตตจากสเตต  $q_{T-1}$  ไปยัง  $q_T$  ด้วยค่าความน่าจะเป็น  $a_{q_{T-1} q_T}$  และแทนค่าความน่าจะเป็นในการเกิดค่าปรากฏเป็น  $O_T$  ด้วยค่าความน่าจะเป็น  $b_{q_T} (O_T)$

จะเห็นว่าสมการนี้มีการคำนวณที่ยุ่งยากเนื่องจากการคูณกันเป็นจำนวนมากในรูปของลำดับ  $2T \cdot N^T$  ดังนั้นจึงมีการคิดหาวิธีมาช่วย ซึ่งแบ่งออกเป็น

1. กระบวนการ ไปข้างหน้า (Forward Procedure);  $\alpha_t(i)$  = Forward variable นิยาม

$$\alpha_t(i) = P(O_1 O_2 \dots O_T, q_t = i \mid \lambda) \quad (3.18)$$

คือความน่าจะเป็นของการเกิดลำดับของค่าปรากฏ  $O_1 O_2 \dots O_T$  และอยู่ที่สแตต  $q_t$  ณ เวลา  $t$  โดยมีแบบจำลองเป็น  $\lambda$  เราสามารถหา  $\alpha_t(i)$  ได้ดังนี้

1. การเริ่มต้น (Initialization)

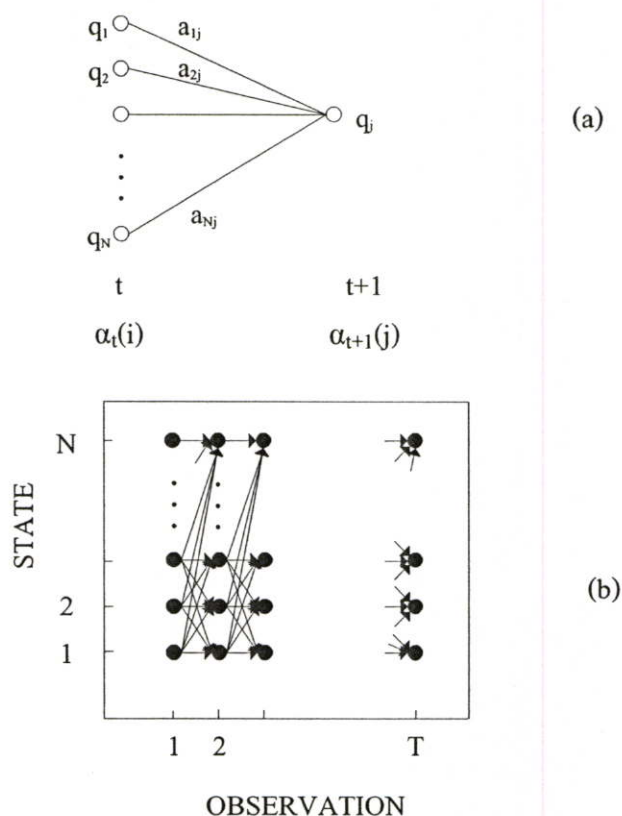
$$\alpha_t(i) = \pi_i b_i O_1 \quad ; 1 \leq i \leq N \quad (3.19)$$

เริ่มต้นด้วยการกำหนดความน่าจะเป็นไปข้างหน้าซึ่งเป็นความน่าจะเป็นร่วมของสแตต  $i$  และมีเหตุการณ์เริ่มต้นเป็น  $O_1$

2. การเหนี่ยวนำ (Induction)

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad ; \begin{array}{l} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{array} \quad (3.20)$$

หมายความว่าสแตต  $j$  ที่เวลา  $t+1$  สามารถมาได้จากสแตตก่อนหน้านี้ ซึ่งเป็นไปได้ถึง  $N$  สแตต (สแตต  $i$  ณ เวลา  $t$  โดยที่  $1 \leq i \leq N$ ) ดังรูป 3.24(a)



รูปที่ 3.24 กระบวนการไปข้างหน้า

จากรูปที่ 3.24 (b) แสดงให้เห็นว่าการคำนวณค่าความน่าจะเป็นแบบไปข้างหน้า (Forward probability) มีโครงสร้างการคำนวณคล้ายๆ ลักษณะของโครงผลึก และเนื่องจากมี

จำนวนสเตตเพียง  $N$  สเตต (แทนด้วยจำนวนโหนดในแต่ละช่วงเวลา  $t$  ใดๆในโครงผลึก) จำนวนลำดับสเตตจะถูกจัดเรียงลงในโหนดเหล่านี้ โดยในเวลา  $t=1$  จะทำการคำนวณค่าของ  $\alpha_t(i)$  ในทุกๆสเตต,  $1 \leq i \leq N$  และที่เวลา  $t = 2, 3, \dots, T$  จะทำการคำนวณค่าของ  $\alpha_t(j)$  ในทุกๆสเตต,  $1 \leq i \leq N$  โดยในแต่ละค่าจะทำการคำนวณมาจาก  $\alpha_{t-1}(i)$  จำนวน  $N$  ค่าก่อนหน้านี

3. การสิ้นสุด (Termination)

$$P(O|\lambda) = \sum_{i=1}^N \alpha_t(i) \quad ; 1 \leq i \leq N \tag{3.21}$$

เราสามารถหา  $P(O|\lambda)$  ได้จากผลรวมของ  $\alpha_t(i)$  จากทุกๆสเตต

2. กระบวนการย้อนกลับ (Backward Procedure);  $\beta_t(i) =$  Backward variable

นิยาม

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | i_t = q_i, \lambda) \tag{3.22}$$

คือ ความน่าจะเป็นของลำดับค่าปรากฏส่วนหลังจากเวลา  $t+1$  ไปจนจบ โดยกำหนดว่าต้องอยู่ที่สเตต  $i$  ที่เวลา  $t$  และมีแบบจำลองเป็น  $\lambda$  เราจะคำนวณหา  $\beta_t(i)$  ได้ดังนี้

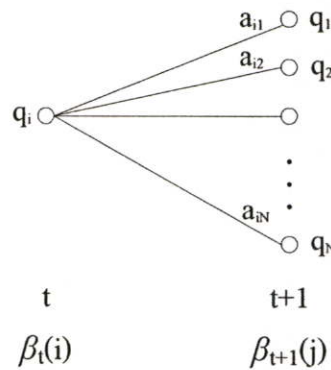
1. การเริ่มต้น (Initialization)

$$\beta_t(i) = 1 \quad ; 1 \leq i \leq N \tag{3.23}$$

2. การเหนี่ยวนำ (Induction)

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \tag{3.24}$$

เมื่อ  $t = T-1, T-2, \dots, 1 \quad ; 1 \leq i \leq N$



รูปที่ 3.25 กระบวนการย้อนกลับ

จากรูปที่ 3.25 เพื่อที่จะให้ค่าปรากฏอยู่ที่สเตต  $i$  ณ เวลา  $t$  โดยคาดคะเนจากลำดับค่าปรากฏจากเวลา  $t+1$  ซึ่งเราจะต้องพิจารณาจากสเตต  $j$  ที่เป็นไปได้ทั้งหมด โดยจะขึ้นอยู่กับค่า  $a_{ij}$  และ  $b_j(O_{t+1})$

การแก้ปัญหาที่ 2 ใช้วิธอร์บีอัลกอริทึม (Viterbi Algorithm) เพื่อที่จะหาลำดับสเปคที่ ดีที่สุด  $q = (q_1, q_2, q_3, \dots, q_T)$  ให้กับลำดับของค่าปรากฏ  $O = \{O_1, O_2, O_3, \dots, O_T\}$  ที่มีอยู่ โดยนิยามให้

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P [q_1, q_2, \dots, q_{t-1}, q_t = i, O_1, O_2, \dots, O_t | \lambda] \quad (3.25)$$

เมื่อ  $\delta_t(i)$  คือ ความน่าจะเป็นสูงสุด (highest probability) ของเส้นทาง (path) ซึ่งจะหาได้ จากค่าความน่าจะเป็นสูงสุด เมื่อเทียบกับสเปคทุกสเปคในการให้ค่าปรากฏเป็นไปตามค่าปรากฏที่ กำหนดให้ ที่ขณะเวลา  $t$  ใดๆ และจากการอาศัยคุณสมบัติของการเหนี่ยวนำจะได้

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}) \quad (3.26)$$

โดยกำหนดให้  $\psi_t(j)$  เป็นอาร์เรย์ที่เก็บตำแหน่งของสเปค ที่ให้ค่าความน่าจะเป็นสูงที่สุดที่ คำนวณได้ในแต่ละเวลา  $t$  และแต่ละลำดับ  $j$  ซึ่งจะสามารถหาลำดับสเปคที่ดีที่สุด โดยใช้ กระบวนการต่อไปนี้

#### 1. การเริ่มต้น (Initialization)

$$\delta_1(i) = \pi_i b_i O_1 \quad ; 1 \leq i \leq N \quad (3.27)$$

$$\psi_1(i) = 0 \quad (3.28)$$

#### 2. การย้อนกลับ (Recursion)

$$\delta_t(i) = [\max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] \cdot b_j(O_t) \quad ; \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq N \end{matrix} \quad (3.29)$$

$$\psi_t(i) = [\arg \max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] \quad ; \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq N \end{matrix} \quad (3.30)$$

#### 3. การสิ้นสุด (Termination)

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.31)$$

$$q_T = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.32)$$

#### 4. เส้นทางเดินย้อนกลับ (Backtracking)

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad ; t = T-1, T-2, \dots, 1 \quad (3.33)$$

การแก้ปัญหาที่ 3 จากที่กล่าวมาแล้วข้างต้นว่าแบบจำลองของท่าภาษามือไทยจะแทนด้วย ค่าพารามิเตอร์  $\lambda = (A, B, \pi)$  ดังนั้นเมื่อมีลำดับของค่าปรากฏจำนวนหนึ่ง เพื่อที่จะนำมาสร้าง แบบจำลองอ้างอิง จะต้องทำการคำนวณหาค่าพารามิเตอร์  $A, B, \pi$  ของแบบจำลอง ซึ่งจะอยู่ในรูป ของค่าความน่าจะเป็น โดยวิธีที่เลือกใช้ก็คือ วิธีของบาม-เวลล์ (Baum-Welch method) หรือเรียกอีก ชื่อหนึ่งว่า EM (Expectation-Maximization method) โดยมี

นิยาม 1. คือ

$$\gamma_t(i) = P (q_t = i | O, \lambda) \quad (3.34)$$

เมื่อ  $\gamma_t(i)$  คือ ค่าความน่าจะเป็นที่จะอยู่ที่สแตต  $i$  ที่ขณะเวลา  $t$  โดยให้ลำดับของค่าปรากฏด้วยโมเดล  $\lambda$  โดยที่กำหนดลำดับของค่าปรากฏให้ สามารถแสดงค่า  $\gamma_t(i)$  ได้ดังนี้

$$\begin{aligned}\gamma_t(i) &= P(q_t = i | O, \lambda) \\ &= \frac{P(O, q_t = i | \lambda)}{P(O | \lambda)} \\ &= \frac{P(O, q_t = i | \lambda)}{\sum_{i=1}^N P(O, q_t = i | \lambda)}\end{aligned}\quad (3.35)$$

เนื่องจาก  $P(q_t = i | O, \lambda)$  มีค่าเท่ากับ  $\alpha_t(i)\beta_t(i)$  ดังนั้นสามารถเขียน  $\gamma_t(i)$  ได้เป็น

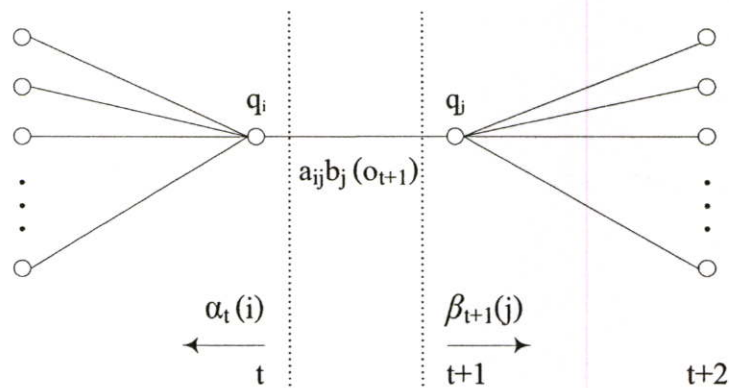
$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)}\quad (3.36)$$

โดย  $\alpha_t(i)$  เริ่มจาก  $O_1, O_2, \dots, O_t$  จนถึงสแตต  $i$  ที่เวลา  $t$

โดย  $\beta_t(i)$  เริ่มจาก  $O_{t+1}, O_{t+2}, \dots, O_T$  จนถึงสแตต  $q_t = i$  ที่เวลา  $t$

**นิยาม 2.**  $\epsilon_t(i,j) = P(q_t = i, q_{t+1} = j | O, \lambda)$  (3.37)

เมื่อ  $\epsilon_t(i,j)$  คือความน่าจะเป็นที่จะอยู่ที่สแตต  $i$  ที่เวลา  $t$  และสแตต  $j$  ที่เวลา  $t+1$  เมื่อกำหนดแบบจำลองและลำดับค่าปรากฏให้



**รูปที่ 3.26** ลำดับการคำนวณการเกิดค่าปรากฏร่วมซึ่งจะอยู่ที่สแตต  $i$  เวลา  $t$  และอยู่ที่สแตต  $j$  เวลา  $t+1$

จากรูปแสดงลำดับการคำนวณการเกิดค่าปรากฏร่วม ซึ่งระบบจะอยู่ในสแตต  $i$  ที่เวลา  $t$  และอยู่ที่สแตต  $j$  ที่เวลา  $t+1$  โดย  $\alpha_t(i)$  เริ่มจาก  $t=1$  ที่ค่าปรากฏแรกจนถึงสแตต  $q_i$  ที่เวลา  $t$  และ  $a_{ij}b_j(o_{t+1})$  เป็นการเปลี่ยนสแตตที่เวลา  $t$  ไปเป็น  $q_j$  ที่เวลา  $t+1$  และให้ค่าปรากฏเป็น  $O_{t+1}$

ซึ่งจากนิยามของตัวแปรไปข้างหน้า  $\alpha_t(i)$  และตัวแปรย้อนกลับ  $\beta_t(i)$  สามารถนำมาสัมพันธ์กับ  $\epsilon_t(i,j)$  ได้เป็น

$$\epsilon_t(i,j) = \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)}$$

$$\begin{aligned}
&= \frac{\alpha_i(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} \\
&= \frac{\alpha_i(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}
\end{aligned} \tag{3.38}$$

จากที่ได้นิยาม  $\gamma_t(i)$  แล้ว นำมาสัมพันธ์กับ  $\varepsilon_t(i,j)$  ได้เป็น

$$\gamma_t(i) = \sum_{j=1}^N \varepsilon_t(i,j) \tag{3.39}$$

$$\text{เมื่อ } \sum_{i=1}^{T-1} \gamma_t(i) = \text{จำนวนของการย้ายสแตทจากสแตท } i \text{ ในลำดับค่าปรากฏ } O \tag{3.40a}$$

$$\sum_{i=1}^{T-1} \varepsilon_t(i,j) = \text{จำนวนของการย้ายสแตทจากสแตท } i \text{ ไป } j \text{ ในลำดับค่าปรากฏ } O \tag{3.40b}$$

ดังนั้นสามารถคำนวณหาค่าของพารามิเตอร์ได้ดังนี้

$$\pi'_i = \text{จำนวนครั้งในการอยู่ที่สแตท } i \text{ ที่เวลา } t=1$$

$$\pi'_i = \gamma_1(i) \quad ; 1 \leq i \leq N \tag{3.41a}$$

$$a'_{ij} = \frac{\text{จำนวนครั้งที่คาดไว้ของการย้ายสแตทจาก } i \text{ ไป } j}{\text{จำนวนครั้งที่คาดว่าจะย้ายจากสแตท } i}$$

$$a'_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{3.41b}$$

$$b'_j(k) = \frac{\text{จำนวนครั้งที่คาดว่าจะอยู่ในสแตท } j \text{ และเกิดค่าปรากฏเป็น } V_k}{\text{จำนวนครั้งที่คาดว่าจะอยู่ที่สแตท } j}$$

$$b'_j(k) = \frac{\sum_{t=1, O_t=V_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \tag{3.41c}$$

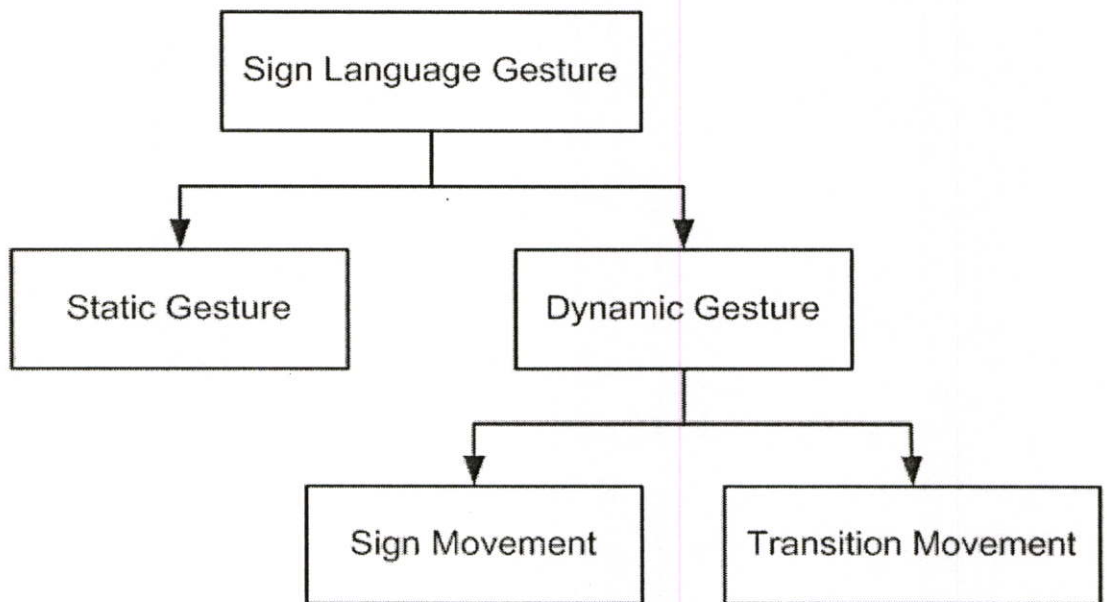
จากกระบวนการข้างต้นถ้าให้  $\lambda = (A, B, \pi)$  เป็นแบบจำลองปัจจุบัน และใช้  $\lambda$  นี้คำนวณในด้านขวาของสมการที่ (3.41a-c) และให้แบบจำลองที่ได้จากการคำนวณซ้ำเป็น  $\lambda' = (A', B', \pi')$  เป็นแบบจำลองที่ได้จากด้านซ้ายของสมการที่ (3.41a-c) ซึ่งจะได้จุดวิกฤติของฟังก์ชันความน่าจะเป็นในกรณีที่  $\lambda' = \lambda$  หรือถ้า  $\lambda'$  มีความน่าจะเป็นมากกว่าแบบจำลอง  $\lambda$  [ $P(O|\lambda') > P(O|\lambda)$ ] นั่นคือจะได้แบบจำลอง  $\lambda'$  ใหม่ ที่น่าจะทำให้เกิดลำดับของค่าปรากฏ  $O$  ที่ดีกว่า

## บทที่ 4

# การรู้จำท่ามือแบบต่อเนื่องในภาษามือไทย

### 4.1 ระบบการแปลภาษามือไทย

ภาษามือไทยจะประกอบไปด้วย รูปมือ, ตำแหน่งของมือ, การเคลื่อนไหว, การหันของฝ่ามือ และการแสดงออกทางสีหน้า ซึ่งสามารถแบ่งส่วนประกอบของลักษณะการแสดงท่ามือได้ดังต่อไปนี้



รูปที่ 4.1 การแบ่งแยกรูปแบบของการแสดงท่าทางในระบบภาษามือไทย

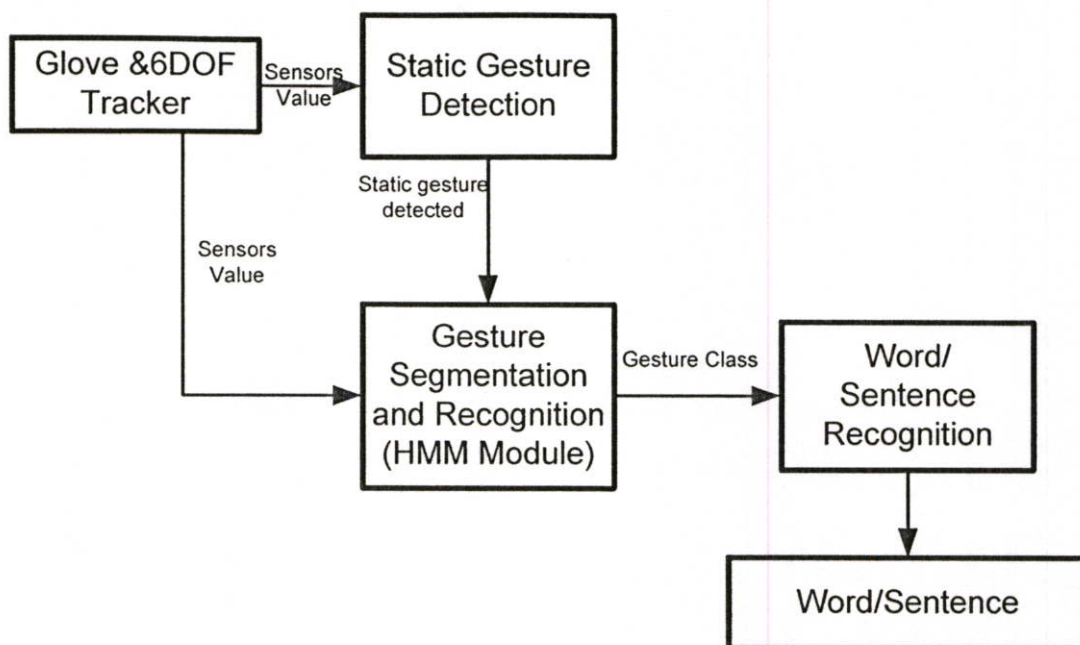
ทำนิ่ง (Static gesture, Posture) คือ การแสดงท่ามือที่ไม่มีการเคลื่อนไหวของมือและตำแหน่งต่างๆ

ท่าเคลื่อนไหว (Dynamic gesture) คือ การแสดงท่ามือที่มีการเคลื่อนไหวมือ หรือตำแหน่งของมือ ซึ่งสามารถแบ่งออกได้ คือ

- การเคลื่อนไหวที่มีความหมาย (Sign Movement) คือ การเคลื่อนไหวที่เป็นส่วนที่มีความหมายในภาษามือไทย

- การเคลื่อนไหวระหว่างท่ามือ (Transition Movement) คือ การเคลื่อนไหวในระหว่างการเปลี่ยนจากท่ามือหนึ่งไปสู่อีกท่าหนึ่ง

ซึ่งระบบการแปลภาษามืออัตโนมัติสามารถแสดงผังแผนภาพต่อไปนี้



รูปที่ 4.2 ระบบรู้จำภาษามือไทยชนิดต่อเนื่องอัตโนมัติ

ระบบการรู้จำภาษามือไทยต่อเนื่องอัตโนมัติ ดังรูปที่ 4.2 ประกอบไปด้วยส่วนสำคัญ 4 ส่วน คือ ส่วนแรกจะมีการรับข้อมูลจากถุงมืออิเล็กทรอนิกส์และอุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกน ส่วนที่สอง คือ การตรวจจับท่าหนึ่ง ซึ่งผลจากส่วนนี้จะถูกนำมาใช้เพื่อช่วยให้ HMM recognizer สามารถลดเส้นทางการหาเส้นด้วย Viterbi algorithm ได้ (เรียกว่าการ Pruning) ส่วนที่สาม คือ ส่วนของการรู้จำท่าทางด้วยฮิดเดนมาร์คอฟโมเดลและส่วนสุดท้าย คือ การรวมส่วนประกอบต่างๆ เข้าด้วยกันเป็นคำหรือประโยค ซึ่งในส่วนของโมดูล Static Gesture Detection นั้นเป็นการดำเนินงานจากงานวิจัยก่อนหน้า [6] โดยกระบวนการทำงานของระบบ สามารถแยกออกเป็น ส่วนการทำงานได้ดังต่อไปนี้

การรับข้อมูลจากถุงมืออิเล็กทรอนิกส์และอุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกน (Glove & 6DOF Tracker) เป็นการรับข้อมูลจากเซนเซอร์ต่างๆที่ติดไว้ที่ถุงมืออิเล็กทรอนิกส์ ซึ่งในงานวิจัยนี้ ใช้เซนเซอร์จากถุงมือจำนวน 18 เซนเซอร์และจากอุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกนจำนวน 6 เซนเซอร์

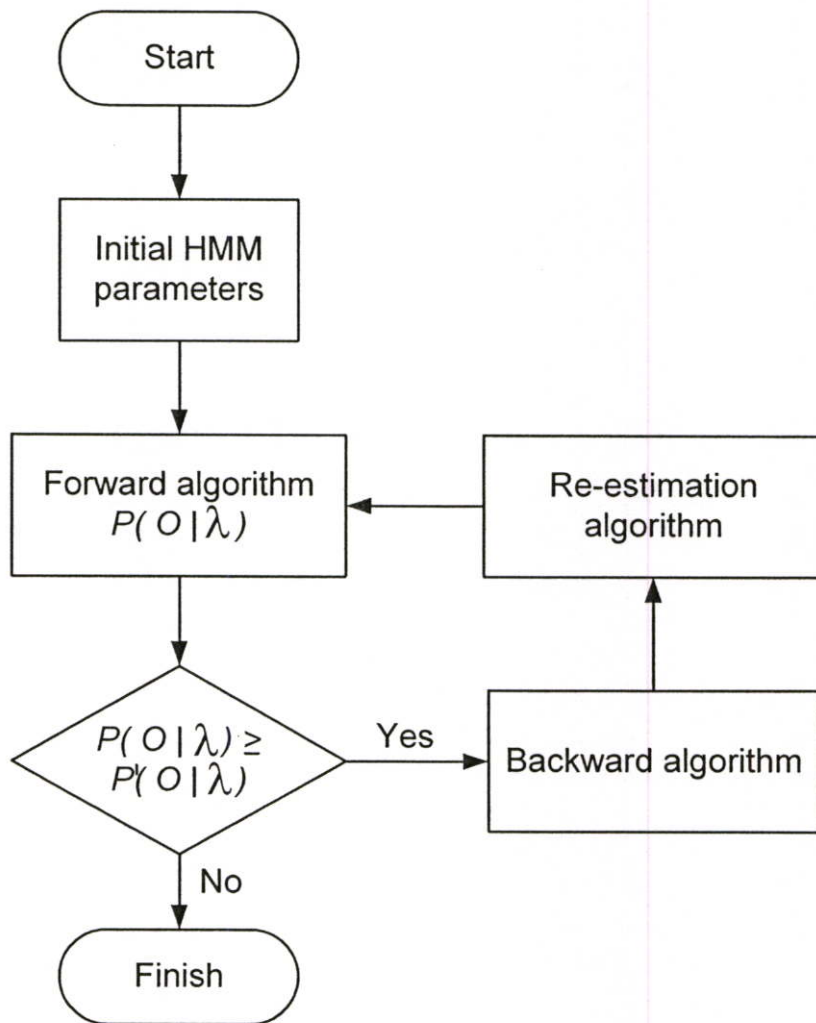
การแบ่งแยกท่าหนึ่งจากท่าเคลื่อนไหว (Static Gesture Detection) กระบวนการนี้จะทำหน้าที่แยกท่าหนึ่งออกจากท่าเคลื่อนไหว จะใช้วิธีการของ Bayesian Estimator เนื่องจากเป็นวิธีการที่ดีที่สุดในการเชิงสถิติสำหรับการรู้จำท่าหนึ่ง

การแบ่งแยกท่าเคลื่อนไหว (Gesture Segmentation and Recognition: HMM Module) เป็นกระบวนการรู้จำท่าเคลื่อนไหวที่มีความหมายออกจากการเคลื่อนไหวระหว่างท่ามือออกมาได้ โดยใช้ฮิดเดนมาร์คอฟโมเดล

การรวมส่วนประกอบต่างๆ เข้าด้วยกันเป็นประโยค (Word/Sentence Recognition) เป็นการนำผลที่ได้จากการรู้จำในส่วนต่างๆ มาประกอบกันเป็นประโยค

#### 4.1.1 กระบวนการสอนฮิดเดนมาร์คอฟโมเดล และการใช้งานฮิดเดนมาร์คอฟโมเดล

ในหัวข้อนี้จะนำเสนอถึงกระบวนการสอนฮิดเดนมาร์คอฟโมเดล และการใช้งานฮิดเดนมาร์คอฟโมเดล ซึ่งในการสอนฮิดเดนมาร์คอฟโมเดลให้รู้จำ หรือแบ่งแยกสัญญาณที่ไม่ทราบ ทำได้โดยการนำสัญญาณที่ไม่ทราบซึ่งอยู่ในรูปของลำดับของค่าที่ปรากฏ นำมาสอนให้แก่ฮิดเดนมาร์คอฟโมเดล โดยกระทำตามขั้นตอนดังรูปที่ 4.3 โดยมีขั้นตอนในการสอนดังนี้



รูปที่ 4.3 กระบวนการสอนฮิดเดนมาร์คอฟโมเดล

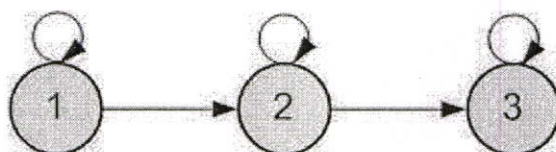
- เริ่มต้นด้วยการสุ่มค่าพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดล ได้แก่  $A, B, \pi$

- คำนวณหาค่าความน่าจะเป็น โดยใช้กระบวนการไปข้างหน้า ซึ่งจะได้ค่าความน่าจะเป็นของโมเดลออกมา
- พิจารณาค่าความน่าจะเป็นใหม่กับค่าความน่าจะเป็นก่อนหน้า จะทำการปรับค่าพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดล โดยใช้กระบวนการย้อนกลับ แล้วใช้บาม-เวลส์อัลกอริทึม (Baum-Welch algorithm) ในการปรับค่าพารามิเตอร์  $A, B, \pi$  ให้ได้ค่าที่เหมาะสม แล้วย้อนกลับไปทำกระบวนการไปข้างหน้าอีกครั้งเพื่อหาค่าความน่าจะเป็นของโมเดล แต่ถ้าค่าความน่าจะเป็นใหม่ที่ได้มีค่าลดลง หรือมีค่าไม่เปลี่ยนแปลงเป็นระยะเวลานาน จึงจะหยุดวงจรการคำนวณ แล้วเก็บพารามิเตอร์  $A, B, \pi$  โดยพารามิเตอร์เหล่านี้เป็นตัวแทนของลักษณะลำดับของค่าที่ปรากฏที่ไม่รู้จัก โดยวิธีการดังกล่าวแสดงดังรูปที่ 4.3

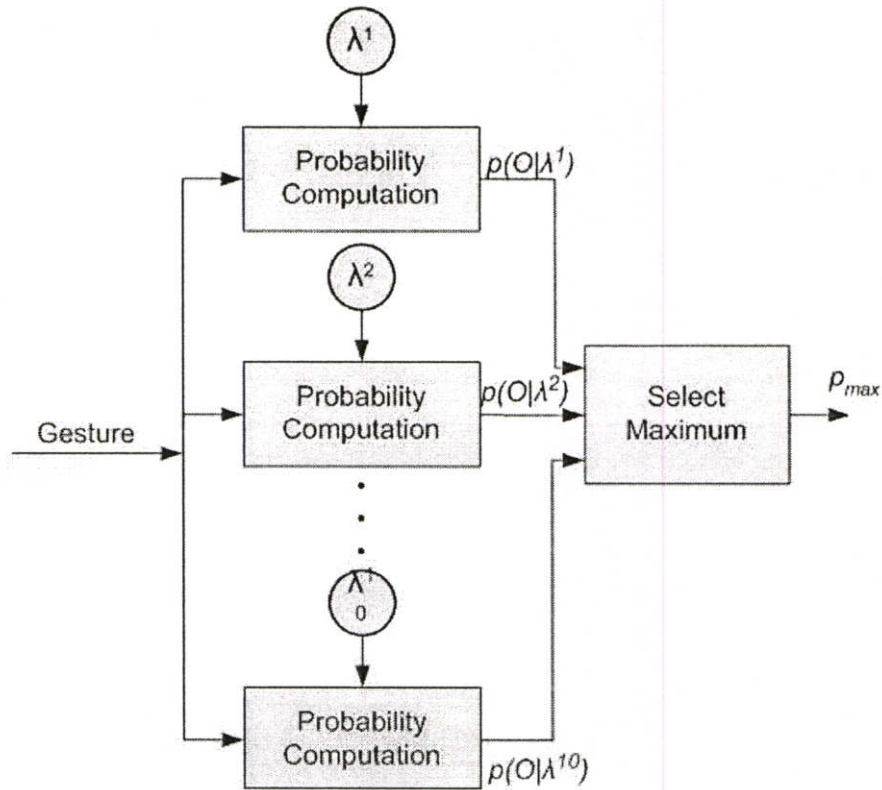
#### 4.1.2 การใช้ฮิดเดนมาร์คอฟในการรู้จำท่าในภาษามือแบบไม่ต่อเนื่อง

ในการใช้ฮิดเดนมาร์คอฟในการรู้จำท่าในภาษามือแบบไม่ต่อเนื่อง (Isolated sign) สามารถอธิบายได้ดังรูปที่ 4.5 ซึ่งจากรูปแต่ละบล็อกของ Probability Computation นั้นแทนฮิดเดนมาร์คอฟโมเดล ซึ่งแต่ละโมเดลถูกสร้างจากท่ามือ 1 ท่า สำหรับโมเดลที่  $i^{\text{th}}$  เราจะให้  $\lambda^i$  แทนโมเดลของฮิดเดนมาร์คอฟโมเดล ซึ่งมีพารามิเตอร์คือ  $A^i, B^i, \pi^i$  ซึ่ง  $A^i$  คือ เซตของ state transition probability,  $B^i$  คือเซตของ observation probability และ  $\pi^i$  คือเซตของ initial state probability โดยที่สเตตของการสอนฮิดเดนมาร์คอฟโมเดลจะมีการประมาณค่าพารามิเตอร์ดังกล่าว โดยใช้บาม-เวลส์อัลกอริทึม

ในกรณีของการจำแนก เซตของข้อมูลหรือที่เรียกว่า Observation data จะถูกคัดออกจากข้อมูลที่ได้รับมาทั้งหมดโดยถุงมืออิเล็กทรอนิกส์ ซึ่งจะเป็นข้อมูลเข้าในโมดูล Gesture Segmentation and Recognition (HMM Module) การประมาณค่าความน่าจะเป็นสูงสุดได้ใช้ Viterbi algorithm สำหรับแต่ละโมเดลที่ถูกเปรียบเทียบกับโมเดลอื่นๆ ค่าที่มีค่าสูงสุดแทนด้วย  $p_{\max}$  จะถูกเลือก



รูปที่ 4.4 ตัวอย่างโทโปโลยีของแต่ละฮิดเดนมาร์คอฟโมเดลแบบ 3 สเตตซึ่ง 1 โมเดลแทน 1 คำ



รูปที่ 4.5 การใช้งานฮิดเดนมาร์คอฟโมเดล

## 4.2 วิธีการรู้จำภาษาไทยชนิดต่อเนื่อง

### 4.2.1 การเลือกจำนวนสเตทของฮิดเดนมาร์คอฟโมเดล

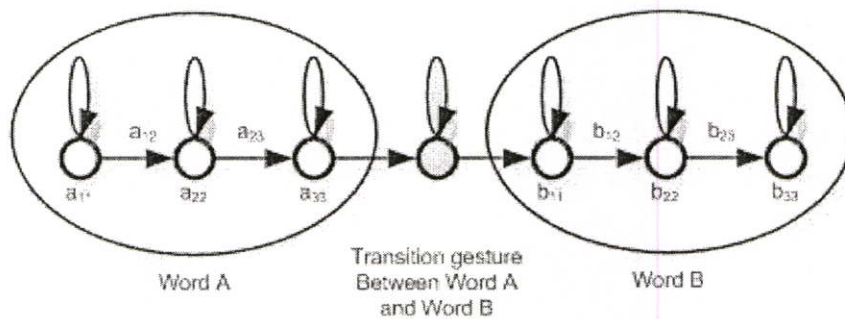
การเลือกจำนวนสเตทของฮิดเดนมาร์คอฟโมเดลนั้น จะพบปัญหา 2 กรณี คือ หากมีจำนวนสเตทน้อยเกินไป จะทำให้การรู้จำท่าที่มีความคล้ายคลึงกันเกิดความรู้จำผิดพลาดได้ และหากมีจำนวนสเตทมากเกินไป (Over fit) ก็จะทำให้การรู้จำท่าเดียวกันผิดพลาดได้ เนื่องจากการทำท่ามือในแต่ละครั้งในส่วนของการทำงานท่ามือ การหมุนมือ การเคลื่อนไหวของมืออาจจะไม่ได้เหมือนกันอย่างพอดีทุกครั้ง จากปัญหาดังกล่าวจึงได้นำการแบ่งข้อมูลเค-มีน (K-means Clustering) มาประยุกต์ใช้ในการเลือกจำนวน สเตท การแบ่งข้อมูลเค-มีนนั้นจะต้องกำหนดจำนวนกลุ่ม (K) ไว้ล่วงหน้า ซึ่งในงานวิจัยนี้ได้ทำการแบ่งกลุ่มข้อมูลในระดับคำ โดยแต่ละคำจะนำข้อมูลมาทำการเลือกการแบ่งกลุ่มออกเป็น 2, 3 และ 4 ทำการหาคู่อันดับของระหว่าง 2 กลุ่มที่มีค่าที่ใกล้ที่สุดหรือระยะห่างน้อยที่สุด (Euclidean distance) แล้วนำมาหาระยะทางระหว่างจุดสองจุด หลังจากนั้นหารด้วยผลบวกของค่า Standard Deviation ระหว่างจุดของจุดข้างต้น ซึ่งสามารถเขียนเป็นสมการได้ดังนี้

$$Q = \frac{d}{STD_1 + STD_2} \quad (4.1)$$

เมื่อได้ค่า  $Q$  ของแต่ละชุดของข้อมูลในการแบ่งกลุ่ม 2, 3 และ 4 กลุ่มแล้ว จึงนำค่าทั้งหมดมาทำการหา Threshold 2 ค่า เพื่อทำการแบ่งกลุ่มค่า  $Q$  ที่ได้นี้ออกเป็น 3 กลุ่ม หลังจากนั้นทำการทดสอบกับข้อมูลในระดับค่า เพื่อทดสอบว่ากลุ่มใดที่มีความถูกต้องมากที่สุด แล้วจึงนำกลุ่มที่มีความถูกต้องมากที่สุดนี้ ไปทำการทดสอบการรู้จำในระดับประโยคต่อไป

#### 4.2.2 การสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือ

เนื่องจากการเคลื่อนไหวระหว่างท่ามือนั้นเป็นท่าที่ไม่มี ความหมาย ซึ่งเป็นการเคลื่อนไหวของมือระหว่างท่าที่มีความหมาย 2 ท่า แสดงได้ดังรูปที่ 4.6 จากงานวิจัย [2] ต้องทำการเก็บข้อมูลของการเคลื่อนไหวระหว่างท่ามือลงในฐานข้อมูล แล้วทำการสอนให้กับระบบ ซึ่งทำให้ฐานข้อมูลมีขนาดใหญ่ขึ้น และคำนวณซับซ้อนขึ้นเนื่องจากต้องสอนให้กับระบบทุกๆท่าเคลื่อนไหวระหว่างมือ



รูปที่ 4.6 ท่าเคลื่อนไหวระหว่างท่ามือซึ่งอยู่ระหว่างคำ A และ คำ B

ซึ่งในงานวิจัยนี้จึงขอแนะนำเสนอ วิธีในการสร้างฮิดเดนมาร์คอฟโมเดล ของการเคลื่อนไหวระหว่างท่ามือ (Transition movement) ด้วยการคำนวณ โดยสร้างโมเดลแบบสเตตเดียว เพื่อการลดความจำเป็นในการที่ต้องเก็บข้อมูลลงในฐานข้อมูล ซึ่งทำให้ต้องเสียพื้นที่ในการจัดเก็บข้อมูล อีกทั้งการเก็บข้อมูลในแต่ละท่านั้นต้องใช้เวลาและต้องนำมาสอนให้แก่ระบบ หากโมเดลที่ได้นั้นยังไม่เหมาะสมก็ต้องทำการคำนวณใหม่ ซึ่งทำให้การคำนวณนั้นซับซ้อนมากขึ้น และการสร้างโมเดลจากการเก็บข้อมูลเข้าป้อนนั้น ข้อมูลที่สร้างจะอยู่ในฐานข้อมูลตลอดเวลา ซึ่งวิธีการทำงานวิจัยนี้ได้แนะนำนั้น จะทำการสร้างโมเดลของการเคลื่อนไหวระหว่างท่าด้วยคำนวณพารามิเตอร์จากสเตตสุดท้ายของท่าก่อนหน้าร่วมกับสเตตแรกของท่าถัดไป ซึ่งในขณะที่โปรแกรมทำงานก็จะทำการสร้างขึ้นเป็นการชั่วคราว เพื่อนำมาใช้ในการหาเส้นทางที่ดีที่สุดสำหรับการรู้จำภาษาไทยชนิดต่อเนื่องอัตโนมัติ และหลังจากนั้นระบบก็จะทำการลบทิ้งไป ทำให้ประหยัดพื้นที่ในการเก็บข้อมูลและหน่วยความจำ

ซึ่งการสร้างโมเดลนั้นประกอบด้วยพารามิเตอร์  $(\pi, A, \mu, U)$  โดย  $\pi$  คือ ค่าเริ่มต้นของการเปลี่ยน สเตทจากจุดแรก,  $A$  คือ เมตริกซ์ความน่าจะเป็นในการเปลี่ยนสเตท,  $\mu$  คือ เวกเตอร์เฉลี่ย และ  $U$  คือ เมตริกซ์ Covariance กำหนดให้  $(\pi, A, \mu, U)$  เป็นพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดลของคำก่อนหน้า,  $(\pi', A', \mu', U')$  เป็นพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดลของคำถัดไป และ  $(\bar{\pi}, \bar{A}, \bar{\mu}, \bar{U})$  เป็นพารามิเตอร์ของฮิดเดนมาร์คอฟโมเดลของการเคลื่อนไหวระหว่างมือที่ได้จากการคำนวณ ซึ่งในการสร้างโมเดลของท่าเคลื่อนไหวระหว่างท่ามือจะกำหนดค่าต่างๆดังต่อไปนี้

ค่าเริ่มต้นของการเปลี่ยนสเตทจะได้

$$\bar{\pi} = 1 \quad (4.2)$$

เนื่องจากโมเดลเป็นแบบสเตทเดียวดังนั้นค่าความน่าจะเป็นในการเปลี่ยนสเตทจะได้

$$\bar{A} = 1 \quad (4.3)$$

ค่าเวกเตอร์เฉลี่ยคำนวณจากค่าเวกเตอร์เฉลี่ยในสเตทสุดท้ายของคำก่อนหน้า บวกกับค่าเวกเตอร์เฉลี่ยในสเตทแรกของคำถัดไปแล้วหารด้วย 2 จะได้

$$\bar{\mu} = \frac{\mu_k + \mu'_1}{2} \quad (4.4)$$

$$\bar{U}_{11} = \frac{U_{k,11} + U'_{1,11}}{2} \quad (4.5)$$

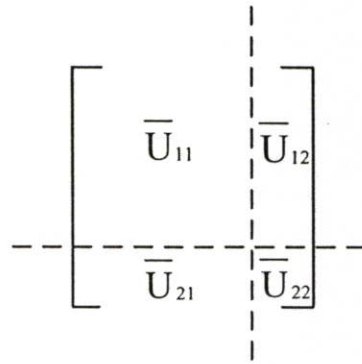
$$\bar{U}_{12} = \bar{U}_{21}^T = 0 \quad (4.6)$$

$$\bar{U}_{22} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_L^2 \end{bmatrix} \quad (4.7)$$

จากสมการที่ 4.5  $\bar{U}_{11}$  คือ  $M \times M$  top-left submatrix ของ  $\bar{U}$  ซึ่ง  $M$  คือจำนวนข้อมูลซึ่งได้จากจุดมีอิเลคตรอนิกส์ในที่นี้คือ 18 ในทำนองเดียวกัน  $U_{3,11}$  และ  $U'_{1,11}$  คือ  $M \times M$  top-left submatrix ของ  $U_3$  และ  $U'_1$  ตามลำดับ ในสมการที่ 4.6 และ 4.7  $\bar{U}_{12}$ ,  $\bar{U}_{21}$ , และ  $\bar{U}_{22}$  คือ submatrice ที่เหลือของ  $\bar{U}$  แสดงดังรูปที่ 4.7 ขนาด  $L$  ของ  $\bar{U}_{22}$  นั้นเท่ากับจำนวนข้อมูลซึ่งได้จากอุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกนซึ่งในที่นี้คือ 6 และค่าความแปรปรวนในสมการที่ 4.7 สามารถแสดงได้ดังสมการต่อไปนี้

$$\sigma_i^2 = \{|\mu_{3,M+i} - \mu'_{1,M+i}|/\alpha\}^2 \quad (4.8)$$

ซึ่ง  $\mu_{3,M+i}$  คือ ค่าเฉลี่ยที่อยู่ในโมเดลซึ่งเป็นสเตทสุดท้ายของคำแรก เช่นเดียวกับข้อมูลลำดับที่  $i^{\text{th}}$  ที่ได้จากเซนเซอร์ และ  $\mu'_{1,M+i}$  ก็เช่นเดียวกัน

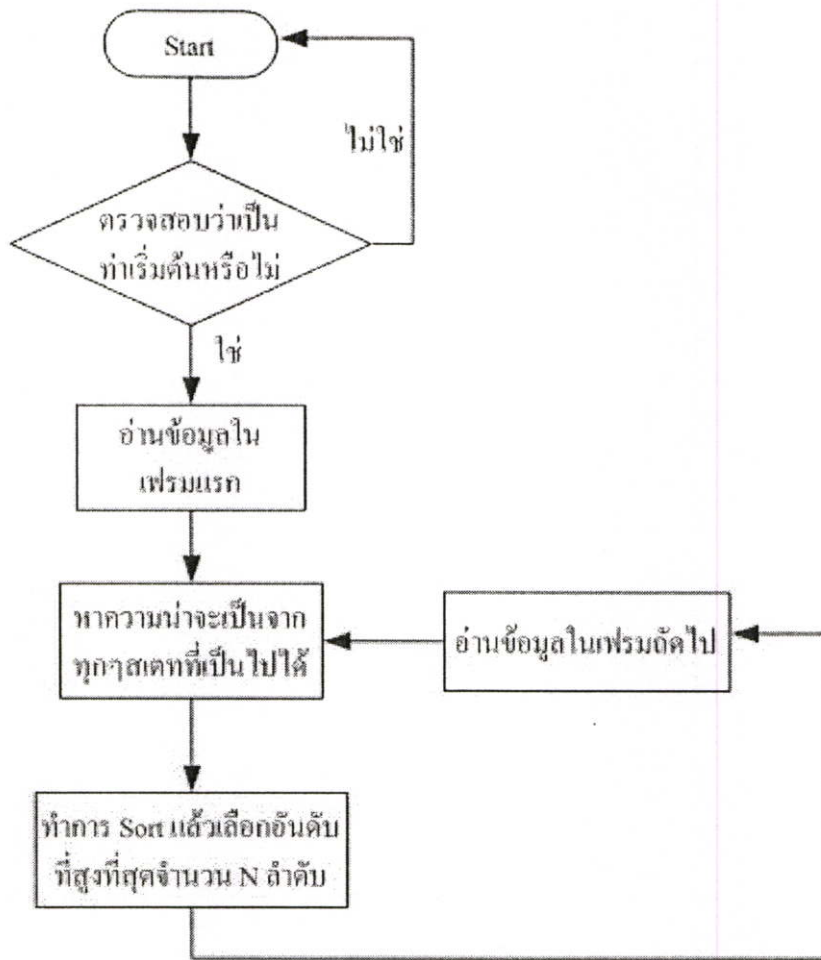


รูปที่ 4.7 แสดงการจำแนกของ  $\bar{U}$  ออกเป็น 4 submatrice

#### 4.2.3 การรู้จำด้วย Viterbi algorithm

การจดจำภาษามือ นั้น เป็นการหาลำดับของภาษามือที่เหมาะสมกับข้อมูลภาษามือ โดยใช้ข้อมูลโมเดลที่ได้จากการประมาณจากขั้นตอนการสอน ซึ่งจะทำการหาลำดับของภาษามือที่มีความน่าจะเป็นที่สูงที่สุด ในการหาลำดับของภาษามือที่น่าจะเป็นไปได้ จะทำการหาได้โดยพิจารณาลำดับที่เป็นไปได้ของทุกสแตทและทุกโมเดล ซึ่งเป็นการทำให้การประมวลผลในขั้นตอนนี้มีจำนวนมาก และจะทำให้กระบวนการในรู้จำนั้นช้า ดังนั้นเพื่อที่จะทำการลดปัญหาจำนวนในการค้นหาจึงใช้ Viterbi algorithm ซึ่งเป็นหลักการพื้นฐานช่วยในการเลือกเส้นทางที่ดีที่สุด ทำให้ไม่ต้องค้นหาในทุกๆเส้นทางได้ ซึ่งขั้นตอนนั้นเริ่มแรกจะทำการพิจารณาข้อมูลภาษามือที่เวลา  $t=1$  ว่ามีความน่าจะเป็นอยู่ในสแตทแรกของโมเดลภาษามือโมเดลใด ต่อไปจะทำการพิจารณาข้อมูลภาษามือที่เวลา  $t$  ต่อไป โดยคำนวณจนกระทั่งครบจำนวนข้อมูลภาษามือที่เวลา  $t_n$  โดยพิจารณาจากสแตทก่อนหน้า ซึ่งการทำแบบนี้จะมีลักษณะต่อเนื่องเป็นกราฟต้นไม้ (Tree graph) และเมื่อมาถึงเฟรมสุดท้าย ณ เวลา  $t_n$  ของการค้นหา จะทำการเลือกเส้นทางที่มีความน่าจะเป็นที่สูงที่สุด นอกจากนี้ Viterbi ยังลดเส้นทางด้วยการลบเส้นทางที่มีปลายทางเหมือนกันอีกด้วย โดยจะเลือกจากลำดับสแตทที่มีความน่าจะเป็นสูงที่สุด

จากรูปที่ 4.7 ได้แสดงการรู้จำด้วย Viterbi algorithm ซึ่งในงานวิจัยนี้ได้ทำการกำหนดจำนวนลำดับที่จะจัดเก็บลงในหน่วยความจำ โดยกำหนดให้  $N$  แทนจำนวนลำดับดังกล่าว เพื่อช่วยในการลดการจัดเก็บในหน่วยความจำ โดยทำการหาความน่าจะเป็นจากทุกๆสแตทที่เป็นไปได้ในแต่ละเฟรม แล้วมาทำการจัดเรียง (Sort) และเลือกอันดับที่สูงที่สุด  $N$  ลำดับ แล้วจึงทำการการอ่านข้อมูลในเฟรมถัดไป ทำเช่นนี้ไปจนหมดข้อมูล



รูปที่ 4.8 แผนภาพแสดงการรู้จำด้วย Viterbi algorithm

## บทที่ 5

# อุปกรณ์และข้อมูลที่ใช้ในการทดลอง

### 5.1 อุปกรณ์ที่ใช้ในการเก็บข้อมูล

อุปกรณ์ที่ใช้ในการเก็บข้อมูลประกอบด้วย

- เครื่องคอมพิวเตอร์ PC Intel Pentium IV 1.8 GHz
- ถุงมืออิเล็กทรอนิกส์ของบริษัท Immersion Corporate แบบมือขวา 22 เซนเซอร์
- อุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกน รุ่น 3SPACE® ISOTRAK II™ ของบริษัท Polhemus Inc. พร้อมเซนเซอร์ จำนวน 2 เซนเซอร์
- โปรแกรมจัดเก็บข้อมูลที่พัฒนาจากโปรแกรม Microsoft® Visual C++ 6.0 และใช้ไลบรารีของ Virtual Hand Library Version 2.4
- เก็บข้อมูลเข้าสู่ฐานข้อมูล Microsoft Access โดยผ่าน ODBC
- MatLab® Version 6.5 โดยดึงข้อมูลจากฐานข้อมูล และวิเคราะห์ข้อมูล

### 5.2 ข้อมูลที่ใช้ในการทดลอง

#### 5.2.1 สภาพแวดล้อมขณะเก็บข้อมูล

ผู้เก็บท่ามือมีทั้งหมด 2 คน ซึ่งจะสวมใส่อุปกรณ์ถุงมืออิเล็กทรอนิกส์เพียงข้างขวา ในขณะที่มีการเก็บข้อมูลจะมีผู้ควบคุมการทดลองเพื่อบอกถึงวิธีการแสดงท่ามือและการกำหนดการเริ่มต้นและสิ้นสุด รวมถึงการยกเลิกการเก็บเมื่อมีการทำท่ามือผิดพลาด และลักษณะการจัดวางของการทดลองจะกำหนดตำแหน่งของผู้เข้าเก็บข้อมูล และตัวส่งข้อมูลของอุปกรณ์ตรวจวัดตำแหน่งแบบ 6 แกน ให้อยู่ในตำแหน่งเดิมทุกครั้ง

#### 5.2.2 ท่ามือที่ใช้ในการทดลอง

ท่ามือที่ใช้ในการทดลองสามารถแบ่งออกได้เป็น 3 ประเภทดังต่อไปนี้

- ท่ามือประเภทคำ เป็นท่ามือที่เป็นคำในภาษามือไทยที่คัดเลือกมาแล้วจากหนังสือภาษามือไทย เล่มที่ 1-6 ประกอบด้วยท่ามือทั้งหมด 106 คำ ซึ่งประกอบไปด้วย ท่ามือนิ่งจำนวน 7 คำ ท่ามือเคลื่อนไหวจำนวน 99 คำ โดยเก็บข้อมูลท่าละ 40 ครั้ง ในการเก็บข้อมูลจะต้องเริ่มจากตำแหน่งเริ่มต้นในการเก็บข้อมูล คือ ตำแหน่งซิดแนบข้างลำตัวและกำมือ โดยแบ่งข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลที่เป็นท่าเคลื่อนไหวจำนวน 99 คำ ดังตารางที่ 5.1 และท่ามือนิ่งจำนวน 7 คำ ดังตารางที่ 5.2

หน้านี้ไม่มีในต้นฉบับ

ตารางที่ 5.1 แสดงข้อมูลท่ามือท่าเคลื่อนไหว

ลำดับ	ท่ามือ	ลำดับ	ท่ามือ	ลำดับ	ท่ามือ
1	สบายดี	34	เปรี้ยว	67	ไข่ดาว
2	ไม่สบายใจ	35	จี๊ด	68	ขนมจีน
3	ภาษามือ	36	สู้ม	69	เหงา
4	เรียน	37	มะม่วง	70	เมฆ
5	พบ	38	ลับประรด	71	ท้องฟ้า
6	ลืบ	39	ชมพู	72	ฟ้าแลบ
7	บ่าย	40	ทุเรียน	73	แม่น้ำ
8	ทั้งวันทั้งคืน	41	เซอริ	74	บุรุษไปรษณีย์
9	วัน	42	แดงโม	75	ชาวนา
10	วันจันทร์	43	อ้อย	76	ชอบ
11	วันศุกร์	44	แพะ	77	คิดถึง
12	สัปดาห์	45	ข้าง	78	ฝัน
13	วันนี้	46	ยีราฟ	79	ขี้ม
14	พรุ่งนี้	47	แมลงวัน	80	หิว
15	มะรืนนี้	48	หอยแครง	81	อิม
16	เมื่อวาน	49	ผักคะน้า	82	มีความสุข
17	เมื่อวานซืน	50	เส้นหมี่	83	เสียใจ
18	กรกฎาคม	51	หน้า	84	สงสัย
19	อะไร	52	เสื้อกล้าม	85	เบื่อ
20	ทำไม	53	เสื้อเชิต	86	เศร้า
21	เท่าไร	54	กางเกงขายาว	87	ยมนาสดิก
22	อย่างไร	55	ผ้าเช็ดหน้า	88	ว่ายน้
23	พ่อ	56	ผ้าถุง	89	ยิงธนู
24	พี่น้อง	57	ผ้าห่ม	90	ฟิวเจอร์ปาร์ครังสิต
25	ลุง	58	โต๊ะ	91	เบลเยี่ยม
26	หลาน	59	ขวด	92	เดนมาร์ก
27	สีดำ	60	เครื่องโทรสาร	93	ภาษาอังกฤษ
28	สีขาว	61	จักรยาน	94	กฎหมาย

ตารางที่ 5.1 (ต่อ)

ลำดับ	ท่ามือ	ลำดับ	ท่ามือ	ลำดับ	ท่ามือ
29	สีน้ำตาล	62	รถไฟฟ้า	95	บัญชี
30	สีชมพู	63	วัด	96	การออกแบบ
31	สีครีม	64	วิทยาลัย	97	เชิญ
32	หวาน	65	ไป	98	นั่ง
33	อโรย	66	มา	99	กบ

ตารางที่ 5.2 แสดงข้อมูลท่ามือนั่ง

ลำดับ	ท่ามือ
1	ท่าเริ่มต้น(ท่ามือแนบลำตัว)
2	ฉัน
3	คุณ
4	เขา
5	พักผ่อน
6	ใหญ่
7	กิน

- ประโยค ซึ่งประโยคที่ใช้ในการทดลองนั้นเป็นการผสมกันระหว่างคำที่เลือกจากตารางที่ 5.1 และ 5.2 ซึ่งในที่นี้มีทั้งหมด 52 ประโยคดังตารางที่ 5.3 โดยแต่ละประโยคประกอบไปด้วยจำนวนคำ 2-4 คำ โดยแต่ละประโยคนั้นทำการเก็บประโยคละ 20 ครั้ง

ตารางที่ 5.3 แสดงประโยคทั้งหมดที่ใช้ในการทดลอง

ลำดับ	ประโยค	คำที่ประกอบในประโยค
1	ช้างตัวใหญ่	ช้าง - ใหญ่
2	ส้มมีรสชาติเปรี้ยว	ส้ม - เปรี้ยว
3	กบกินแมลงวัน	กบ - กิน - แมลงวัน
4	ชมพู่มีสหวาน	ชมพู่ม - หวาน
5	ฉันทพบเขาที่เดินมารักเดือนกรกฎาคม	พบ - เขา - เดินมารัก - กรกฎาคม
6	เชิญนั่งสิ	เชิญ - นั่ง
7	พ่อสบายดีไหม	พ่อ - สบายดี

ตารางที่ 5.3 (ต่อ)

ลำดับ	ประโยค	คำที่ประกอบในประโยค
8	วันนี้วันจันทร์	วันนี้ - วันจันทร์
9	วันศุกร์ฉันพักก่อน	ฉัน - พักก่อน - วันศุกร์
10	ฉันมีพี่น้องสิบคน	พี่น้อง - ฉัน - สิบ
11	ฉันมีแพะสิบตัว	แพะ - ฉัน - สิบ
12	เรียนภาษาอังกฤษวันจันทร์	เรียน - ภาษาอังกฤษ - วันจันทร์
13	จะไปว่ายน้ำมะริณี	ว่ายน้ำ - มะริณี
14	ฉันเรียนภาษามือวันจันทร์ตอนบ่าย	เรียน - ภาษามือ - วันจันทร์ - บ่าย
15	ชอบวิหะอะไรระหว่างภาษาอังกฤษกับ กฎหมาย	ภาษาอังกฤษ - กฎหมาย - ชอบ - อะไร
16	ชอบวิหะอะไรระหว่างบัญชีกับการออกแบบ	บัญชี - การออกแบบ - ชอบ - อะไร
17	ฉันเรียนที่ประเทศเบลเยียม	เรียน - เบลเยียม
18	คุณหน้าขาวจ้ง	หน้า - ขาว
19	เมื่อวานกินมะม่วงและสับปะรด	กิน - มะม่วง - สับปะรด - เมื่อวานนี้
20	เมื่อวานขึ้นกินทุเรียนและเชอร์รี่	กิน - ทุเรียน - เชอร์รี่ - เมื่อวานขึ้น
21	ข้างกินอ้อย	ข้าง - กิน - อ้อย
22	พรุ่งนี้จะไปวัดอย่างไร	ไป - วัด - พรุ่งนี้ - อย่างไร
23	จะไปโดยรถไฟฟ้าหรือจักรยาน	ไป - รถไฟฟ้า - จักรยาน - อะไร
24	ฉันชอบยิมนาสติกและว่ายน้ำ	ฉัน - ชอบ - ยิมนาสติก - ว่ายน้ำ
25	สงสัยว่ายิงธนูอย่างไร	สงสัย - ยิงธนู - อย่างไร
26	ยี่ราฟกินผักคะน้า	ยี่ราฟ - กิน - ผักคะน้า
27	คุณไปศูนย์การค้าฟิวเจอร์ปาร์ครังสิตทำไม	ไป - ศูนย์การค้าฟิวเจอร์ปาร์ครังสิต - ทำไม
28	ลุงไม่สบายใจทั้งวันทั้งคืน	ลุง - ไม่สบายใจ - ทั้งวันทั้งคืน
29	เรียนภาษามือสัปดาห์ละเท่าไร	เรียน - ภาษามือ - สัปดาห์ - เท่าไร
30	หลานเรียนที่วิทยาลัย	หลาน - เรียน - วิทยาลัย
31	ฉันเบื่อกินไข่ดาว	เบื่อ - กิน - ไข่ดาว
32	ฉันเศร้าและเสียใจทั้งวันทั้งคืน	ฉัน - เศร้า - เสียใจ - ทั้งวันทั้งคืน
33	แดงโมลูกนี้หวาน	แดงโม - หวาน
34	คุณยิ้มหวาน	คุณ - ยิ้ม - หวาน

ตารางที่ 5.3 (ต่อ)

ลำดับ	ประโยค	คำที่ประกอบในประโยค
35	มะม่วงลูกนี้อร่อย	มะม่วง - อร่อย
36	ฉันกินเส้นหมี่และหอยแครง	กิน - เส้นหมี่ - หอยแครง
37	เครื่องโทรสารสีดำ	เครื่องโทรสาร - สีดำ
38	ผ้าห่มสีขาวหรือสีน้ำตาล	ผ้าห่ม - ขาว - น้ำตาล - อะไร
39	โต๊ะสีน้ำตาล	โต๊ะ - สีน้ำตาล
40	ผ้าถุงและผ้าเช็ดหน้าราคาเท่าไร	ผ้าถุง - ผ้าเช็ดหน้า - เท่าไร
41	คุณมาที่วัดอย่างไร	คุณ - มา - วัด - อย่างไร
42	พຽ່ງนี้ไปว่าน้ำที่แม่น้ำ	ว่าน้ำ - แม่น้ำ - พຽ່ງนี้
43	เมฆและท้องฟ้าสีขาว	เมฆ - ท้องฟ้า - สีขาว
44	ฉันกินขนมจีนจันอิม	กิน - ขนมจีน - อิม
45	ลูกเป็นบุรุษไปรษณีย์หรือชาวนา	ลูก - บุรุษไปรษณีย์ - ชาวนา - อะไร
46	ฉันคิดถึงหลานทั้งวันทั้งคืน	คิดถึง - หลาน - ทั้งวันทั้งคืน
47	ขวดน้ำสีชมพู	ขวดน้ำ - สีชมพู
48	เสื้อกล้ามสีขาวเสื้อเชิ้ตสีครีม	เสื้อกล้าม - สีขาว - เสื้อเชิ้ต - สีครีม
49	กางเกงขายาวราคาเท่าไร	กางเกงขายาว - เท่าไร
50	เมื่อวานฟ้าแลบ	ฟ้าแลบ - เมื่อวาน
51	มะม่วงลูกนี้จืด	มะม่วง - จืด
52	ฉันฝันว่ามีความสุข	ฉัน - ฝัน - มีความสุข

- การเคลื่อนไหวระหว่างมือ ซึ่งเป็นการเคลื่อนไหวระหว่างมือที่สัมพันธ์กับประโยคทั้ง 52 ประโยค เป็นจำนวน 178 ท่า

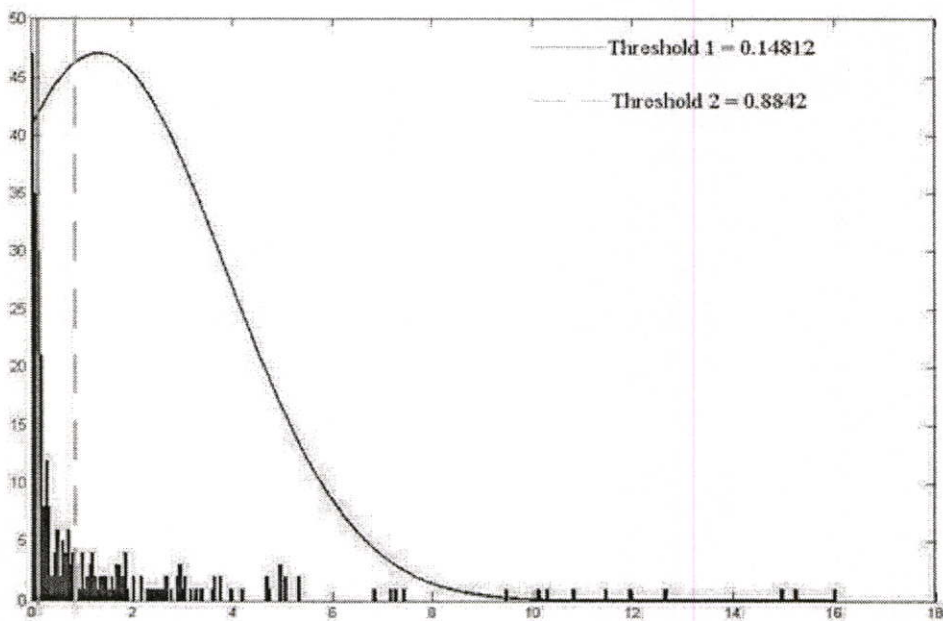
## บทที่ 6

### วิธีการทดลองและผลการทดลอง

ในการทดลองได้ทำการเลือกประโยคคั้งที่ได้กล่าวไปแล้วในบทที่ 5 โดยมีผู้จัดทำมือจำนวน 2 คนกระทำคนละจำนวน 52 ประโยค แต่ละประโยคประกอบไปด้วยคำ 2-4 คำต่อประโยค โดยเก็บประโยคละ 10 ครั้ง และสำหรับคำที่ประกอบอยู่แต่ละประโยคในที่นี่คือ 106 คำโดยแต่ละคำเก็บเป็นจำนวนคำละ 40 ครั้ง โดยแบ่งออกมาทำการสอนให้กับระบบเป็นจำนวน 30 และอีก 10 เอาไว้สำหรับทดสอบก่อนนำไปทดสอบกับการรู้จำทำมือต่อเนื่อง

#### 6.1 การทดลองเพื่อหาจำนวนสเตทของฮิดเดนมาร์คอฟโมเดล

ในการทดลองนี้ได้ทำการทดลองเพื่อหาค่า  $Q$  (จากหัวข้อ 4.1) โดยนำแต่ละคำจากตารางที่ 5.1 ทำการแบ่งกลุ่มเค-มิน โดยแบ่งข้อมูลออกเป็น 2, 3 และ 4 กลุ่ม แล้วทำการหาค่า  $Q$  หลังจากนั้นทำการหาค่า Threshold 2 คำ เพื่อแบ่งกลุ่มของค่า  $Q$  ออกเป็น 3 กลุ่ม ซึ่งได้ค่า Threshold คือ 0.14812 และ 0.8842



รูปที่ 6.1 แสดงการแบ่งกลุ่มของค่า  $Q$  ด้วย Threshold

จากรูปที่ 6.1 เป็นการแสดงการแบ่งกลุ่มของค่า  $Q$  ด้วย Threshold ซึ่งกลุ่มของค่า  $Q$  ที่มีค่าน้อยกว่า Threshold 1 จะถูกจัดเป็นกลุ่มที่ 1, ค่า  $Q$  ที่มีค่ามากกว่า Threshold 1 แต่น้อยกว่า Threshold 2

จะถูกจัดเป็นกลุ่มที่ 2 และค่าที่มากกว่า Threshold 2 ถูกจัดเป็นกลุ่มที่ 3 หลังจากนั้นนำทั้ง 3 กลุ่มมา ทำสอบกับการรู้จำระดับคำเพื่อหากลุ่มที่มีความถูกต้องมากที่สุด เพื่อหาจำนวนสเดทที่เหมาะสม ซึ่งได้ผลการทดลองในการแบ่งกลุ่มที่ให้ค่าความถูกต้องสูงสุดดังตารางต่อไปนี้

ตารางที่ 6.1 ตารางแสดงการแบ่งกลุ่มที่ให้ค่าความถูกต้องสูงสุด

ลำดับ	ทำมือ	จำนวนกลุ่ม	ลำดับ	ทำมือ	จำนวนกลุ่ม
1	สบายดี	2	51	หน้า	3
2	ไม่สบายใจ	3	52	เสื้อกล้าม	2
3	ภาษามือ	2	53	เสื้อเชิ้ต	2
4	เรียน	2	54	กางเกงขายาว	2
5	พบ	2	55	ผ้าเช็ดหน้า	4
6	สืบ	2	56	ผ้าถุง	3
7	บ่าย	2	57	ผ้าห่ม	2
8	ทั้งวันทั้งคืน	2	58	โต๊ะ	2
9	วัน	2	59	ขวด	4
10	วันจันทร์	2	60	เครื่องโทรสาร	2
11	วันศุกร์	2	61	จักรยาน	2
12	สัปดาห์	2	62	รถไฟฟ้า	2
13	วันนี้	3	63	วัด	2
14	พรุ่งนี้	2	64	วิทยาลัย	2
15	มะรืนนี้	2	65	ไป	3
16	เมื่อวาน	2	66	มา	3
17	เมื่อวานซืน	2	67	ไข่ดาว	4
18	กรกฎาคม	2	68	ขนมจีน	3
19	อะไร	2	69	เหงา	2
20	ทำไม	2	70	เมฆ	2
21	เท่าไร	2	71	ท้องฟ้า	2
22	อย่างไร	4	72	ฟ้าแลบ	2
23	พ่อ	2	73	แม่น้ำ	2
24	พี่น้อง	2	74	บุรุษไปรษณีย์	2
25	ลุง	2	75	ชานา	2

ตารางที่ 6.1 (ต่อ)

ลำดับ	ท่ามือ	จำนวนกลุ่ม	ลำดับ	ท่ามือ	จำนวนกลุ่ม
26	หลาน	2	76	ชอบ	2
27	สีคำ	2	77	คิดถึง	3
28	สีขาว	2	78	ฝัน	3
29	สีน้ำตาล	2	79	ยิ้ม	3
30	สีชมพู	2	80	หิว	2
31	สีครีม	2	81	อึ้ง	2
32	หวาน	2	82	มีความสุข	2
33	อรรอย	2	83	เสียใจ	2
34	เปรี้ยว	2	84	สงสัย	3
35	จืด	2	85	เบื่อ	3
36	ส้ม	2	86	เศร้า	4
37	มะม่วง	2	87	ขมनासतिग	4
38	สับปะรด	2	88	ว่ายน้	4
39	ชมพู	3	89	ยิงธนู	3
40	ทุเรียน	3	90	ฟิวเจอร์ปาร์ครังสิต	3
41	เชอร์รี่	2	91	เบลเยียม	3
42	แดงโม	2	92	เดนมาร์ก	4
43	อ้อย	4	93	ภาษาอังกฤษ	2
44	แพะ	4	94	กฎหมาย	2
45	ช้าง	4	95	บัญชี	2
46	ยีราฟ	2	96	การออกแบบ	2
47	แมลงวัน	3	97	เชิญ	2
48	หอยแครง	2	98	นั่ง	2
49	ผักคะน้า	3	99	กบ	2
50	เส้นหมี่	2			

ในการทดลองนี้ได้นำโมเดลที่เป็นจำนวนสเตทที่เหมาะสมไปเปรียบเทียบกับการสอนแบบ 3 สเตทในทุกๆโมเดลโดยการนำไปรู้จำในระดับคำ เมื่อนำกลุ่มของโมเดลซึ่งได้จากผลการทดลองข้างต้นไปรู้จำในระดับคำจะให้ความถูกต้อง 98.57% และนำกลุ่มของโมเดลโดยการสร้างฮิดเดน

มาร์คอฟโมเดลแบบ 3 สเตท จะได้รับความถูกต้อง 97.76% อนึ่ง โดยอาศัยการกำหนดจำนวนสเตทตามตารางที่ 6.1 ได้จำนวนสเตทจริงจำนวนเท่ากับ 87.76%

## 6.2 การทดสอบการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ

ในการทดลองนี้ได้ทำการทดลองเพื่อทดสอบประสิทธิภาพ ในการสร้างโมเดลของท่าเคลื่อนไหวระหว่างมือแล้วนำมาใช้ในการรู้จำภาษามือแบบต่อเนื่อง ซึ่งได้ผลการทดลองดังนี้

ตารางที่ 6.2 ตารางแสดงผลการทดลองเมื่อนำโมเดลท่าเคลื่อนไหวระหว่างมือด้วยการคำนวณมาใช้ในการรู้จำประโยค

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
1	ข้างตัวใหญ่	70	80
2	สัมผัสรสชาติเปรี้ยว	90	90
3	กบกินแมลงวัน	90	80
4	ชมพู่มีรสหวาน	90	90
5	ฉันพบเขาที่เดนมาร์กเดือนกรกฎาคม	90	90
6	เชิญนั่งสิ	90	90
7	พ่อสบายดีไหม	90	90
8	วันนี้วันจันทร์	90	90
9	วันศุกร์ฉันพักผ่อน	90	90
10	ฉันมีพี่น้องสิบคน	90	90
11	ฉันมีแพะสิบตัว	100	90
12	เรียนภาษาอังกฤษวันจันทร์	90	90
13	จะไปว่ายน้ำสระรีนนี่	100	90
14	ฉันเรียนภาษามือวันจันทร์ตอนบ่าย	100	90
15	ชอบวิชาอะไรระหว่างภาษาอังกฤษกับกฎหมาย	90	100
16	ชอบวิชาอะไรระหว่างบัญชีกับการออกแบบ	80	80
17	ฉันเรียนที่ประเทศเบลเยียม	100	90
18	คุณหน้าขาวจัง	90	90
19	เมื่อวานกินมะม่วงและสับปะรด	100	90

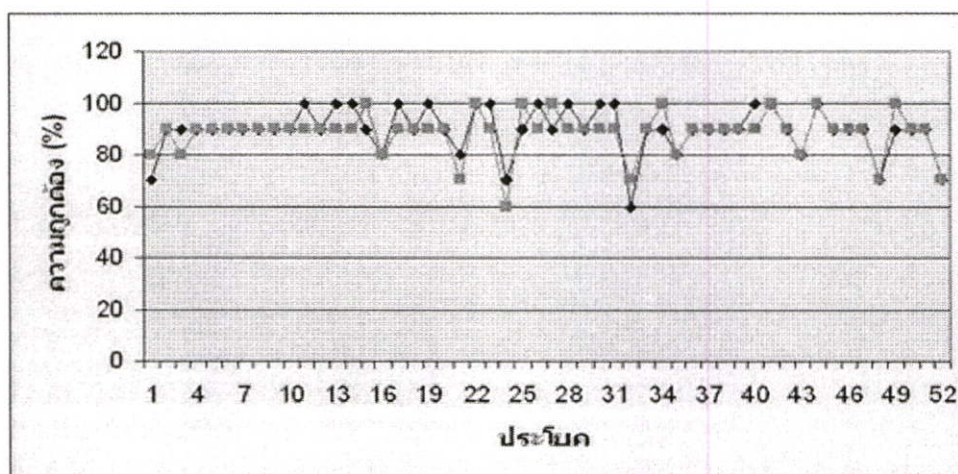
ตารางที่ 6.2 (ต่อ)

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
20	เมื่อวานฉันกินทุเรียนและเชอรี่	90	90
21	ข้างกินอ้อย	80	70
22	พรุ่งนี้จะไปวัดอย่างไร	100	100
23	จะไปโดยรถไฟฟ้าหรือจักรยาน	100	90
24	ฉันชอบยิมนาสติกและว่ายน้ำ	70	60
25	สงสัยว่ายิงธนูอย่างไร	90	100
26	ยี่ราฟกินผักคะน้า	100	90
27	คุณไปศูนย์การค้าฟิวเจอร์ปาร์ครังสิตทำไม	90	100
28	ลุงไม่สบายใจทั้งวันทั้งคืน	100	90
29	เรียนภาษามือสัปดาห์ละเท่าไร	90	90
30	หลานเรียนที่วิทยาลัย	100	90
31	ฉันเบื่อกินไข่ดาว	100	90
32	ฉันเศร้าและเสียใจทั้งวันทั้งคืน	60	70
33	แดงโมลูกนี้หวาน	90	90
34	คุณยิ้มหวาน	90	100
35	มะม่วงลูกนี้อร่อย	80	80
36	ฉันกินเส้นหมี่และหอยแครง	90	90
37	เครื่องโทรสารสีดำ	90	90
38	ผ้าห่มสีขาวหรือสีน้ำตาล	90	90
39	โต๊ะสีน้ำตาล	90	90
40	ผ้าถุงและผ้าเช็ดหน้าราคาเท่าไร	100	90
41	คุณมาที่วัดอย่างไร	100	100
42	พรุ่งนี้ไปว่ายน้ำที่แม่น้ำ	90	90
43	เมฆและท้องฟ้าสีขาว	80	80
44	ฉันกินขนมเงินจันอิม	100	100
45	ลุงเป็นบุรุษไปรษณีย์หรือชาวนา	90	90
46	ฉันคิดถึงหลานทั้งวันทั้งคืน	90	90

ตารางที่ 6.2 (ต่อ)

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
47	ขวดน้ำสีชมพู	90	90
48	เสื้อกล้ามสีขาวเสื้อเชิ้ตสีครีม	70	70
49	กางเกงขายาวราคาเท่าไร	90	100
50	เมื่อวานฟ้าแลบ	90	90
51	มะม่วงลูกนี้จี๊ด	90	90
52	ฉันฝันว่ามีความสุข	70	70

จากผลการทดลองในตารางข้างต้นสามารถสรุปเป็นกราฟได้ดังนี้



รูปที่ 6.2 กราฟแสดงผลการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ

จากการทดลองนี้การรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ ซึ่งได้ความถูกต้องในการรู้จำรวมทั้ง 2 ผู้ทำท่ามือเท่ากับ 89.04%

### 6.3 การทดสอบการรู้จำโดยใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือ

ในการทดลองนี้จะใช้โมเดลของท่าเคลื่อนไหวระหว่างมือที่ได้มีการเก็บข้อมูลของท่าเคลื่อนไหวระหว่างมือจำนวน 178 ท่า แล้วจึงนำมาสร้างโมเดลของท่าเคลื่อนไหวระหว่างมือ โดยกำหนดให้ 1 โมเดลมี 1 สเตท ซึ่งได้ผลการทดลองดังนี้

ตารางที่ 6.3 ตารางแสดงผลการทดลองเมื่อนำโมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือมาใช้ในการรู้จำประโยค

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
1	ช้างตัวใหญ่	90	90
2	สัมมิรสชาติเปรี้ยว	100	90
3	กบกินแมลงวัน	90	90
4	ชมพู่มีระหวาน	90	90
5	ฉันพบเขาที่เดนมาร์กเดือนกรกฎาคม	100	90
6	เชิญนั่งสิ	100	100
7	พ่อสบายดีไหม	100	100
8	วันนี้วันจันทร์	90	90
9	วันศุกร์ฉันพักผ่อน	90	90
10	ฉันมีพี่น้องสิบคน	100	90
11	ฉันมีแพะสิบตัว	100	100
12	เรียนภาษาอังกฤษวันจันทร์	90	90
13	จะไปว่ายน้ำมะริ่นี้	100	100
14	ฉันเรียนภาษามือวันจันทร์ตอนบ่าย	100	100
15	ขอวิชาอะไรระหว่างภาษาอังกฤษกับกฎหมาย	90	100
16	ขอวิชาอะไรระหว่างบัญชีกับการออกแบบ	90	90
17	ฉันเรียนที่ประเทศเบลเยียม	100	90
18	คุณหน้าขาวจัง	100	100
19	เมื่อวานกินมะม่วงและสับปะรด	100	90
20	เมื่อวานฉันกินทุเรียนและเชอร์รี่	90	90
21	ช้างกินอ้อย	90	90
22	พรุ่งนี้จะไปวัดอย่างไร	100	100

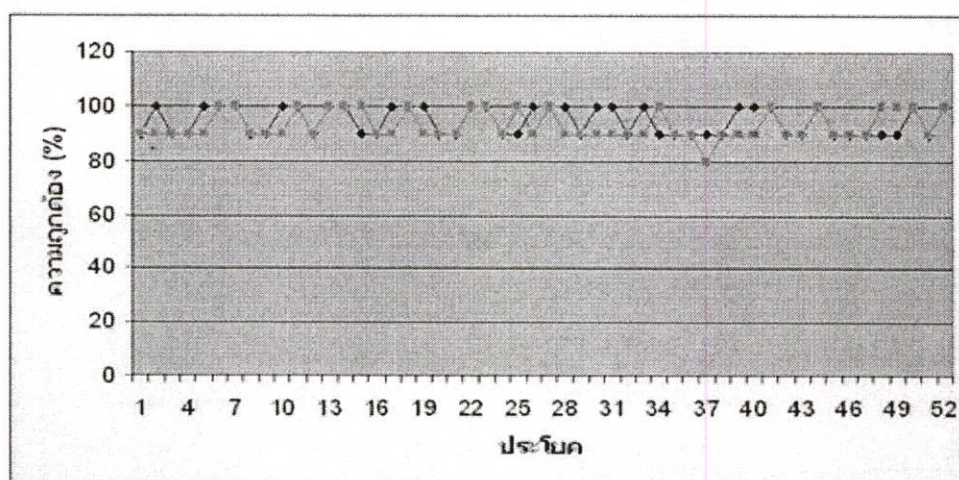
ตารางที่ 6.3 (ต่อ)

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
23	จะไปโดยรถไฟหรือจักรยาน	100	100
24	ฉันชอบยิมนาสติกและว่ายน้ำ	90	90
25	สงสัยว่ายิงธนูอย่างไร	90	100
26	ยี่ราฟกินผักคะน้า	100	90
27	คุณไปศูนย์การค้าฟิวเจอร์ปาร์ครังสิตทำไม	100	100
28	ลุงไม่สบายใจทั้งวันทั้งคืน	100	90
29	เรียนภาษามือสัปดาห์ละเท่าไร	90	90
30	หลานเรียนที่วิทยาลัย	100	90
31	ฉันเนื้อที่จะกินไข่ดาว	100	90
32	ฉันเศร้าและเสียใจทั้งวันทั้งคืน	90	90
33	แดงโมลูกนี้หวาน	100	90
34	คุณยิ้มหวาน	90	100
35	มะม่วงลูกนี้อร่อย	90	90
36	ฉันกินเส้นหมี่และหอยแครง	90	90
37	เครื่องโทรสารสีดำ	90	80
38	ผ้าห่มสีขาวหรือสีน้ำตาล	90	90
39	โต๊ะสีน้ำตาล	100	90
40	ผ้าถุงและผ้าเช็ดหน้าราคาเท่าไร	100	90
41	คุณมาจากที่ใดอย่างไร	100	100
42	พรุ่งนี้ไปว่ายน้ำที่แม่น้ำ	90	90
43	เมฆและท้องฟ้าสีขาว	90	90
44	ฉันกินขนมเงินจันอ้อม	100	100
45	ลุงเป็นบุรุษไปรษณีย์หรือชาวนา	90	90
46	ฉันคิดถึงหลานทั้งวันทั้งคืน	90	90
47	ขวดน้ำสีชมพู	90	90
48	เสื้อกล้ามสีขาวเสื้อเชิ้ตสีครีม	90	100
49	กางเกงขายาวราคาเท่าไร	90	100

ตารางที่ 6.3 (ต่อ)

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
50	เมื่อวานฟ้าแลบ	100	100
51	มะม่วงลูกนี้จี๊ด	90	90
52	ฉันฝันว่ามีความสุข	100	100

จากผลการทดลองในตารางข้างต้นสามารถสรุปเป็นกราฟได้ดังนี้



รูปที่ 6.3 กราฟแสดงผลการรู้จำโดยการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือ

จากการทดลองนี้การรู้จำโดยนำโมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือมาใช้ในการรู้จำประโยค ซึ่งได้ความถูกต้องในการรู้จำรวมทั้ง 2 ผู้ทำท่ามือเท่ากับ 94.04%

#### 6.4 การทดสอบการรู้จำโดยไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือ

ในการทดลองนี้จะเป็นการทดลองโดยไม่ใช้โมเดลของท่าเคลื่อนไหวระหว่างมือมาใช้ในการรู้จำ ซึ่งได้ผลการทดลองดังนี้

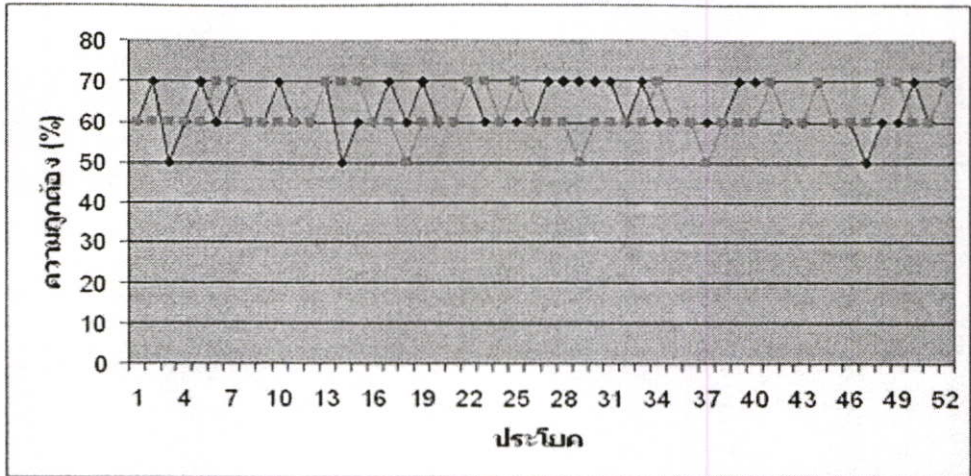
ตารางที่ 6.4 ตารางแสดงผลการทดลองเมื่อไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือในการรู้จำประโยค

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
1	ช้างตัวใหญ่	90	90
2	ส้มมีรสชาติเปรี้ยว	100	90
3	กบกินแมลงวัน	90	90
4	ชมพูมีรสหวาน	90	90
5	ฉันพบเขาที่เคนมาร์กเดือนกรกฎาคม	100	90
6	เชิญนั่งสิ	100	100
7	พ่อสบายดีไหม	100	100
8	วันนี้วันจันทร์	90	90
9	วันศุกร์ฉันพักผ่อน	90	90
10	ฉันมีพี่น้องสิบคน	100	90
11	ฉันมีแพะสิบตัว	100	100
12	เรียนภาษาอังกฤษวันจันทร์	90	90
13	จะไปว่ายน้ำมะริ่นี้	100	100
14	ฉันเรียนภาษาเมื่อวันจันทร์ตอนบ่าย	100	100
15	ขอวิชาอะไรระหว่างภาษาอังกฤษกับกฎหมาย	90	100
16	ขอวิชาอะไรระหว่างบัญชีกับการออกแบบ	90	90
17	ฉันเรียนที่ประเทศเบลเยียม	100	90
18	คุณหน้าขาวจัง	100	100
19	เมื่อวานกินมะม่วงและสับปะรด	100	90
20	เมื่อวานฉันกินทุเรียนและเชอร์รี่	90	90
21	ช้างกินอ้อย	90	90
22	พรุ่งนี้จะไปวัดอย่างไร	100	100
23	จะไปโดยรถไฟฟ้าหรือจักรยาน	100	100
24	ฉันชอบยิมนาสติกและว่ายน้ำ	90	90
25	สงสัยว่ายังธนูอย่างไร	90	100
26	ยี่ราฟกินผักคะน้า	100	90
27	คุณไปศูนย์การค้าฟิวเจอร์ปาร์คครั้งล่าสุดทำไม	100	100

ตารางที่ 6.4 (ต่อ)

ลำดับ	ประโยค	ความถูกต้องในการรู้จำ (%)	
		ผู้ทำท่า 1	ผู้ทำท่า 2
28	ลุงไม่สบายใจทั้งวันทั้งคืน	100	90
29	เรียนภาษามือสัปดาห์ละเท่าไร	90	90
30	หลานเรียนที่วิทยาลัย	100	90
31	ฉันทิ้งที่ตะกอนไข่ดาว	100	90
32	ฉันทเศร้าและเสียใจทั้งวันทั้งคืน	90	90
33	แต่งโมลูนี่หวาน	100	90
34	คุณยิ้มหวาน	90	100
35	มะม่วงลูกนี้อร่อย	90	90
36	ฉันทกินเส้นหมี่และหอยแครง	90	90
37	เครื่องโทรสารสีดำ	90	80
38	ผ้าห่มสีขาวหรือสีน้ำตาล	90	90
39	โต๊ะสีน้ำตาล	100	90
40	ผ้าถุงและผ้าเช็ดหน้าราคาเท่าไร	100	90
41	คุณมาที่วัดอย่างไร	100	100
42	พรงนี้ไปว่ายน้ำที่แม่น้ำ	90	90
43	เมฆและท้องฟ้าสีขาว	90	90
44	ฉันทกินขนมจีนจมนม	100	100
45	ลุงเป็นบุรุษไปรษณีย์หรือชาวนา	90	90
46	ฉันทคิดถึงหลานทั้งวันทั้งคืน	90	90
47	ขวดน้ำสีชมพู	90	90
48	เสื้อกล้ามสีขาวเสื้อเชิ้ตสีครีม	90	100
49	กางเกงขายาวราคาเท่าไร	90	100
50	เมื่อวานฟ้าแลบ	100	100
51	มะม่วงลูกนี้จัด	90	90
52	ฉันทฝันว่ามีความสุข	100	100

จากผลการทดลองในตารางข้างต้นสามารถสรุปเป็นกราฟได้ดังนี้



รูปที่ 6.4 กราฟแสดงผลการรู้จำโดยไม่ใช่โมเดลท่าเคลื่อนไหวระหว่างมือ

จากการทดลองนี้การรู้จำโดยไม่ใช่โมเดลท่าเคลื่อนไหวระหว่างมือ ซึ่งได้รับความถูกต้องในการรู้จำรวมทั้ง 2 ผู้ทำท่ามือเท่ากับ 62.69%

## บทที่ 7

### สรุปผลการทดลองและข้อเสนอแนะ

จากผลการทดลองในบทที่ 6 ในส่วนแรกจะเป็นการทดลองเพื่อหาจำนวนสแตกที่ดีที่สุด เนื่องจากข้อมูลที่ใช้ในการทดลองนั้นมีจำนวนมาก ทำให้ไม่สามารถแบ่งจำนวนสแตกทีละค่าได้ จึงต้องทำการหาวิธีเพื่อหาจำนวนสแตกที่เหมาะสมเพื่อนำมาสอนให้แก่ระบบ กล่าวคือจึงได้นำวิธีการแบ่งกลุ่มเค-มินมาประยุกต์ใช้ในการหาจำนวนสแตกที่เหมาะสม ซึ่งได้จำนวนสแตกที่เหมาะสมดังตารางที่ 6.1 เมื่อได้จำนวนสแตกที่เหมาะสมแล้วก็นำมาเปรียบเทียบกับการสอน โมเดลแบบ 3 สแตก โดยการนำไปรู้จำในระดับค่า ซึ่งเมื่อนำกลุ่มของโมเดลที่ได้จากการแบ่งกลุ่มเค-มินไปรู้จำในระดับค่าจะได้รับความถูกต้อง 98.57% และนำกลุ่มของโมเดลโดยการสร้างฮิดเดนมาร์คอฟโมเดลแบบ 3 สแตก จะได้รับความถูกต้อง 97.76% และนอกจากนี้การหาจำนวนสแตกที่เหมาะสมด้วยการแบ่งกลุ่มเค-มินนั้นมีอัตราเฉลี่ยเท่ากับ 2.42 ซึ่งน้อยกว่า 3 นั่นคือทำให้การคำนวณของระบบนั้นน้อยลงอีกด้วย แต่จากการแบ่งกลุ่มเค-มินนั้นจำนวนสแตกตรงจากความเป็นจริงคิดเป็น 87.76% และนอกจากนี้จะพบว่าค่า  $Q$  ที่ได้ส่วนใหญ่จะมีแนวโน้มเพิ่มขึ้นเมื่อจำนวนกลุ่มเค-มินเพิ่มขึ้น ซึ่งในการทดลองนี้ใช้เพียง 3 ค่า คือ 2, 3 และ 4 ดังนั้นในอนาคตอาจจะนำมาปรับปรุงด้วยการเพิ่มจำนวนกลุ่มที่จะแบ่งขึ้นไปอีก

เมื่อได้จำนวนสแตกที่เหมาะสมแล้ว จึงนำมาทดสอบการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณ ซึ่งได้รับความถูกต้องในการรู้จำรวมทั้ง 2 ผู้ทำท่ามือเท่ากับ 89.04%, เมื่อนำโมเดลมาทดสอบการรู้จำโดยการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือนั้น ได้รับความถูกต้องเท่ากับ 94.04% และเมื่อทดสอบโดยไม่ใช้โมเดลท่าเคลื่อนไหวระหว่างมือในการรู้จำประโยค ได้รับความถูกต้องเท่ากับ 62.69% ซึ่งสรุปได้ว่าการรู้จำโดยการสร้างโมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณนั้น มีความถูกต้องใกล้เคียงกันกับการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือ แต่การใช้โมเดลท่าเคลื่อนไหวระหว่างมือด้วยวิธีการคำนวณในการรู้จำมีประสิทธิภาพน้อยกว่าการใช้โมเดลท่าเคลื่อนไหวระหว่างมือที่สร้างจากการทำท่ามือ เนื่องจากการสร้างโมเดลของการเคลื่อนไหวระหว่างมือนั้นเป็นการเฉลี่ยในแนวเส้นตรง (Linear Interpolation) แต่สระระของคนเรานั้นบางครั้งไม่ได้มีการเคลื่อนไหวเป็นเส้นตรง ซึ่งในอนาคตอาจจะนำวิธีการแบบ Non-Linear Interpolation มาทำให้ระบบสมบูรณ์ยิ่งขึ้น

นอกจากนี้การรู้จำภาษาไทยแบบต่อเนื่องโดยการสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือด้วยการคำนวณ เป็นการลดการจับเก็บการเคลื่อนไหวระหว่างท่ามือลงในฐานข้อมูล อีกทั้งการสร้างโมเดลของการเคลื่อนไหวระหว่างท่ามือด้วยการคำนวณนั้นเป็นการสร้างขึ้นมาใช้ชั่วคราวในหน่วยความจำในขณะที่โปรแกรมทำงานเท่านั้น ไม่ได้เป็นการจัดเก็บถาวร นอกจากนี้ยังเป็นการลด

ความซับซ้อนและความยุ่งยากในการสอนการเคลื่อนไหวกว่าทำมือให้แก่ระบบอีกด้วย ซึ่งงานวิจัยนี้จะช่วยทำให้ระบบแปลภาษามือไทยแบบต่อเนื่องมีความสมบูรณ์ยิ่งขึ้น

## เอกสารอ้างอิง

- [1] Gaolin F. and Wen G., "A SRN/HMM system for signer-independent continuous sign language recognition", Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, 20-21 May 2002 Page(s): 297 -302
- [2] Wen G., Gaolin F., Debin Z. and Yiqiang C., "Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition", Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, 17-19 May 2004 Page(s):553 – 558
- [3] G. Fang, W. Gao and D. Zhao, "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models", Systems, Man and Cybernetics, Part A, IEEE Transactions on Volume 37, Issue 1, Jan. 2007 Page(s):1 – 9
- [4] Jianjun Y., Hongxun Y. and Feng J., "Based on HMM and SVM Multilayer Architecture Classifier for Chinese Sign Language Recognition with Large Vocabulary", Image and Graphics, 2004. Proceedings. Third International Conference on, 18-20 Dec. 2004 Page(s):377 - 380
- [5] Guilin Y., Hongxun Y., Xin L. and Feng J., "Real Time Large Vocabulary Continuous Sign Language Recognition Based on OP/Viterbi Algorithm", Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Volume 3, 20-24 Aug. 2006 Page(s):312 - 315
- [6] วรวิทย์ วีระพันธุ์. (2548). *วิธีการแยกและการรู้จำรูปแบบท่าทางสำหรับการแปลภาษามือไทยอัตโนมัติ*. วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต. ภาควิชาการเทคโนโลยีสารสนเทศบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- [7] องค์การศึกษาวิทยาศาสตร์และวัฒนธรรมแห่งสหประชาชาติ. *หนังสือภาษามือ* กรุงเทพมหานคร : มูลนิธิอนุเคราะห์คนหูหนวกในพระบรมราชินูปถัมภ์ 2526.
- [8] สมาคมคนหูหนวกแห่งประเทศไทย. *ปทานุกรมภาษามือไทย ฉบับปรับปรุงและขยายเพิ่มเติม*. ไทยวัฒนาพานิช, 1990.
- [9] J. LaViola., "A Survey of Hand Posture and Gesture Recognition Techniques and Technology", Technical Report CS-99-11, Brown University, Department of Computer Science, Providence RI, June, 1999.
- [10] Virtual Technologies, Inc. **VirtualHand® Software Library Reference Manual**. Aug. 5,

1998.

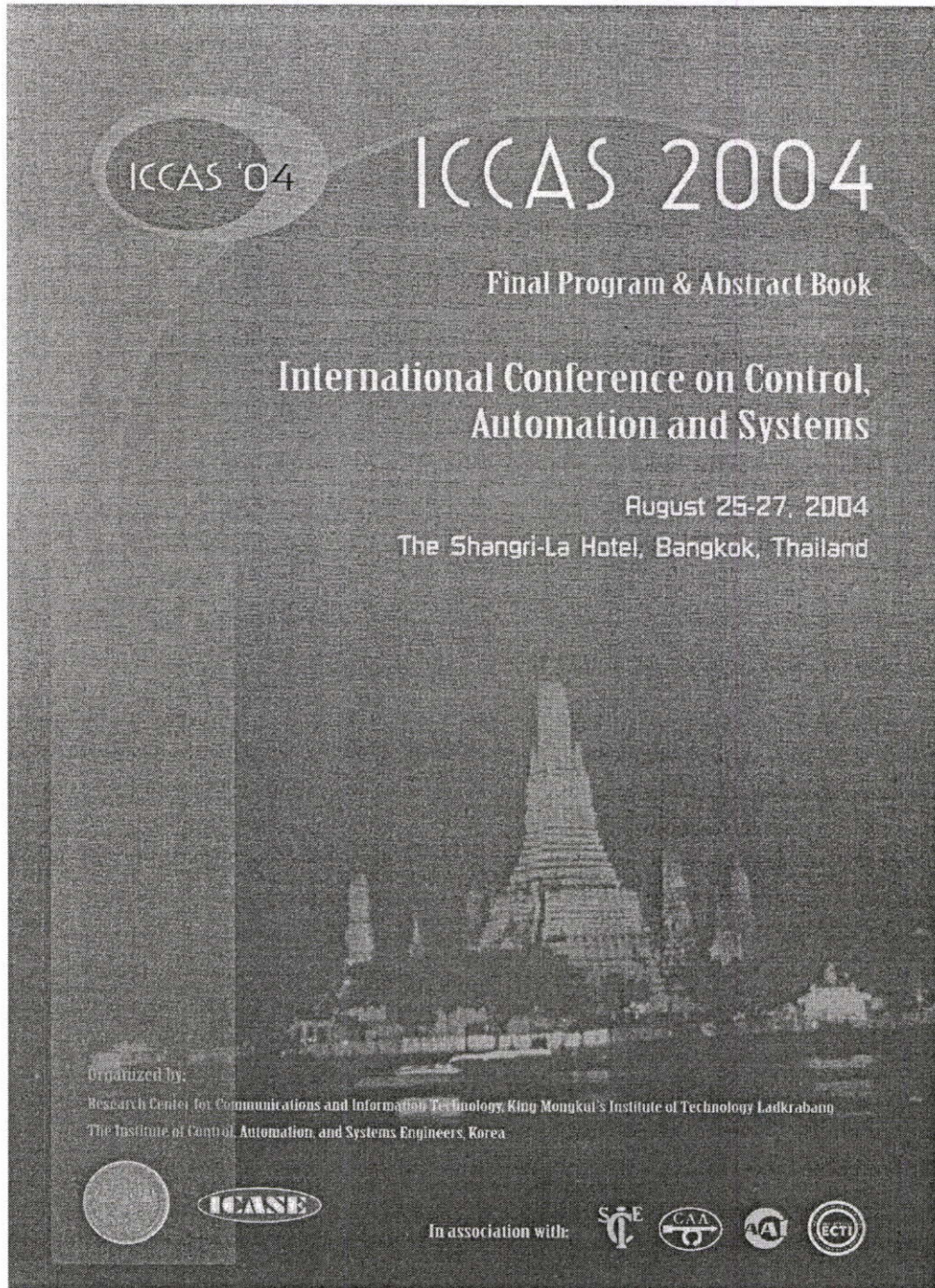
- [11] Polhemus, Inc. **3SPACE<sup>®</sup> User's Manual**. May 1993.
- [12] สมาคมคนหูหนวกแห่งประเทศไทย. **หนังสือภาษามือไทยเล่ม 1-6**. บพิธการพิมพ์, 2542.
- [13] L. Rabiner, "A Tutorial on Hidden Markov Models and Select Applications in speech recognition.", *Proceedings of IEEE*, Vol 77, pp. 257-286, No.2, February 1989.
- [14] วุฒิชัย วิศาลकुณา. (2548). **ระบบจดจำภาษามือไทยโดยใช้ฮิดเดนมาร์คอฟโมเดล**. วิทยานิพนธ์ วิศวกรรมศาสตรมหาบัณฑิต. สาขาวิชาวิศวกรรมเครื่องกล วิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณนรี.

ภาคผนวก

ภาคผนวก ก.

## บทความและผลงานวิจัยที่ได้รับการตีพิมพ์

1. Arunee Jaruwanawat and Nopporn Chotikakamthorn. "Hybrid HMM for Transitional Gesture Classification in Thai Sign Language" ICCAS 2004, pp. 60



### Hybrid HMM for Transitional Gesture Classification in Thai Sign Language Translation

Aruncce Jaruwanawat, Nopporn Chotikakamthorn, and Worawit Werapun

Faculty of Information Technology &  
 Research Center for Communications and Information Technology  
 King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand  
 (E-mail: s5066025@kmitl.ac.th)

**Abstract:** A human sign language is generally composed of both static and dynamic gestures. Each gesture is represented by a hand shape, its position, and hand movement (for a dynamic gesture). One of the problems found in automated sign language translation is on segmenting a hand movement that is part of a transitional movement from one hand gesture to another. This transitional gesture conveys no meaning, but serves as a connecting period between two consecutive gestures. Based on the observation that many dynamic gestures as appeared in Thai sign language dictionary are of quasi-periodic nature, a method was developed to differentiate between a (meaningful) dynamic gesture and a transitional movement. However, there are some meaningful dynamic gestures that are of non-periodic nature. Those gestures cannot be distinguished from a transitional movement by using the signal quasi-periodicity. This paper proposes a hybrid method using a combination of the periodicity-based gesture segmentation method with a HMM-based gesture classifier. The HMM classifier is used here to detect dynamic signs of non-periodic nature. Combined with the periodic-based gesture segmentation method, this hybrid scheme can be used to identify segments of a transitional movement. In addition, due to the use of quasi-periodic nature of many dynamic sign gestures, dimensionality of the HMM part of the proposed method is significantly reduced, resulting in computational saving as compared with a standard HMM-based method. Through experiment with real measurement, the proposed method's recognition performance is reported.

**Keywords:** Sign Language, Gesture Recognition, Hidden Markov Model

#### 1. INTRODUCTION

Sign languages are used in most parts of the world among community of deaf people. A sign language can also be used as a means for communication between a normal and a deaf persons. Research has been carried out as part of an attempt to automatically recognize hand shapes and movements as defined in some sign languages [1]. Although this automated sign language recognition research may find a relatively less commercial and practical value as compared with the speech recognition counterpart, the result should be beneficial to both deaf and normal people in many ways. Not just it can be used to simplify communication between the deaf and other people, result from automated sign recognition research can be applied to improve human-computer interaction efficiency through gesture type of input.

Regardless of the way the hand shape and movement are captured, current sign language translation techniques are differentiated by the recognition approach employed. Examples of feasible approaches include a simple template matching, use of a artificial neural network, and approach based on the Hidden Markov Model (HMM) [4]. Among these approaches, it appears that the HMM principle has received a relatively more attention due to its ability in handling gesture time variation. Another reason for its popularity is its level of success when applied to the speech recognition problem. Example of HMM-based sign language translation research is the work of Vogler and Metaxas [3]. In [3], HMM principle

was applied to recognize 53 signs drawn from the American sign language.

Although the HMM principle is an elegant approach for recognition of time-based patterns (such as speech and signs), the method requires high computational complexity when dealing with large patterns. The problem is remedied by the use of sub-word components (phoneme) in a speech recognition system. Similar idea was explored for automatic sign language recognition [2]. However, unlike speech, there is no standard or widely-accepted phoneme-like counterpart in a sign language. In addition, while continuous speech contains no easily-detected boundary points that can be used for word or sub-word segmentation, movement of a signer's hand while performing signs continuously contains pauses and turning points of the movement. It is thus arguable that by exploiting these natural segmental points, use of the HMM method may be reduced. Research works that exploit these features include those of [5]. In [5], such natural segmental points are exploited along with the (roughly) quasi-periodic nature of many dynamic signs as found in Thai sign language. However, certain dynamic signs in Thai sign language are non-periodic [5]. Therefore, those signs are distinguishable from a transitional movement by the method as described in [5].

This paper describes a hybrid Thai sign language translation method. The method is an extension of that as described in [5]. In particular, a HMM module is added to the system in [5] to deal with non-periodic dynamic signs. Use of such a hybrid scheme offers benefits of both the HMM-based

and non HMM-based methods. While computational complexity of this hybrid method is slightly higher as compared with that of [5], it can deal with all types of signs: static, periodic, or non-periodic.

The paper is organized as follows. Section 2 describes major modules in our Thai sign language translation system. The proposed hybrid scheme based on the HMM principle is then described in Section 3. Experimental results are reported in Section 4. Last, discussion and conclusion remarks are also included.

## 2. THAI SIGN LANGUAGE TRANSLATION SYSTEM

A typical structure of a hand gesture word/sentence in Thai sign language is shown in Fig. 1. From the figure, each sign word/sentence is a sequence of hand gestures. Those gestures are either static or dynamic. A static gesture is a hand of a particular shape without any movement. A dynamic gesture is a hand in motion. A dynamic gesture can be further categorized into either (approximately) quasi-periodic and non-periodic gestures. Among the non-periodic gestures, some of them correspond to transitional movement while some correspond to non-periodic signs. A transitional movement serves as a connection between two consecutive gesture/posture, and conveys no meaning.

Next, consider the system used in this study. Our sign language translation system employs a right-hand instrumented glove, used in combination with a magnetic 6-DOF tracker device. Such devices as described are used here to ensure that data collected and result as obtained are not affected by the accuracy of the acquisition equipment. The architecture of the recognition system is as shown in Fig. 2. The system is composed of 7 modules. The first module performs static gesture detection. If the data obtained from the system's sensors is classified as non-static, the sampled data is passed to the second module to detect hand movement turning points. When this is the case, it is first assumed that the data samples correspond to a moving hand contain segment(s) of transitional movement, or non-periodic gesture, or periodic gesture, or those combination. The outputs of the second module are the sample numbers where change in hand velocity occurs. These turning points are then passed to the 'periodic/non-periodic' gesture classification module. In [5], based on the observation that most meaningful dynamic gestures in Thai sign language can be approximated as quasi-periodic (i.e., those gestures consist of certain hand movement patterns, cyclically repeated), a method was developed to classify samples of hand movement into periodic and non-periodic segments. Most of the data corresponding to non-periodic segments belong to a transitional movement. Some words commonly used in Thai sign language are, however, represented by a non-periodic hand movement [7]. This special case must be handled by the HMM-based transitional/non-periodic gesture classifier. Results from Posture Classifier module, Periodic Gesture classifier module, and the HMM-based classifier module, are then fed to the Word/Sentence Recognition module to arrive at a final recognized word/sentence. In this paper, only the Periodic-gesture and HMM-based modules are discussed in detail.

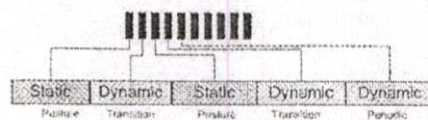


Fig. 1 Typical structure of a hand gesture sequence found in Thai sign language.

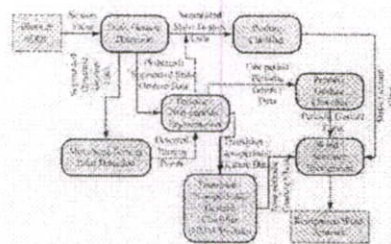


Fig. 2 Block diagram of the Thai sign language translation system

## 3. HYBRID TRANSITIONAL/NON-PERIODIC GESTURE CLASSIFICATION

### 3.1 Use of HMM for Non-periodic Gesture Classification

In Fig. 2, HMM is used to classify among non-periodic gesture classes, as well as to identify a transitional gesture. A detailed block diagram of this HMM-based module is shown in Fig. 3. From the figure, each block labeled as 'probability computation' represents a HMM model corresponding to each non-periodic gesture class. In Fig. 3 there are ten HMM models corresponding to ten non-periodic gesture classes. The structure of each HMM model is shown in Fig. 4.

For the  $i^{\text{th}}$  model, let  $A^i$  be the corresponding set of HMM parameters, which are  $A^i, B^i, \pi^i$ . Based on the notation as detailed in [6],  $A^i$  is the set of state transition probability,  $B^i$  is the set of observation probabilities, and  $\pi^i$  is the set of initial state probabilities. At the training stage, these parameters are estimated, a standard Baum-Welch algorithm [6] is applied here.

During the classification phase, the set of observation data as extracted from a sequence of electronic-glove data samples by the periodic/non-periodic segmentation module, is provided as an input to each of the HMM models. The estimated maximum probability of occurrence using the Viterbi algorithm for each model is compared against others. The one with the highest value, denoted by  $p_{\text{max}}$ , is chosen and compared against a predefined threshold. If  $p_{\text{max}}$  is greater than the threshold, its corresponding gesture class is associated with the input data. Otherwise, the input data is classified as a transitional gesture.

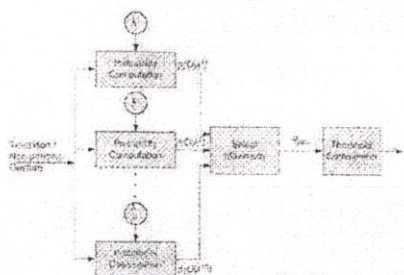


Fig. 3 Detailed block diagram of the non-periodic gesture classifier module.

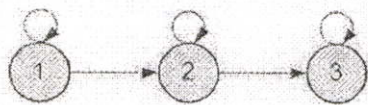


Fig. 4 Structure (topology) of each HMM model representing a non-periodic gesture class

### 3.2 Algorithms using hybrid non-periodic/periodic and HMM modules.

As described in [5], quasi-periodicity as found in many dynamic gestures is exploited to differentiate them from non-periodic and transitional gestures. A segment of hand gesture data samples is classified as periodic if its spectrum as obtained by Fourier analysis contains a spectral peak that exceeds a pre-defined threshold. This threshold of periodicity is denoted by  $T_p$ . By pre-classifying a gesture as periodic or non-periodic using this periodic/non-periodic gesture segmentation module, we reduce the need to perform transitional/non-periodic gesture classification, thus avoiding evaluation of the HMM models. By considering only the gesture classified as dynamic by the previous modules in Fig. 2, the following algorithm describes how the hybrid classifier, which is composed of the periodic/non-periodic segmentation module and the HMM-based non-periodic/transitional gesture classifier module, operates.

1. A gesture classified as dynamic by a previous module is first analyzed by the periodic/non-periodic segmentation module. Result of the analysis falls into one of the following cases

- The gesture is classified as a (vibrated) static gesture, if this is the case, the corresponding captured data is passed to the posture classifier module.
- The gesture is classified as a periodic gesture. This is the case when the spectral peak of the captured data is greater than  $T_p$ . If this is the case, the captured data is passed to the periodic gesture classifier module.

- The gesture is neither periodic nor static. When this is the case, the captured data is passed to the HMM-based non-periodic/transitional gesture classifier module, and go to the next step.

2. The HMM module is used to classify the captured data. The obtained maximum estimated probability,  $P_{max}$ , is compared against a predefined threshold. There are the same numbers of thresholds as the number of HMM models, as denoted by  $T_{a,k}$  for the one corresponds to the  $k^{th}$  HMM model. The threshold corresponds to the model which achieves maximum estimated probability is used for the comparison. Let's assume that  $P_{max}$  is due to the  $k^{th}$  model.

Then, if  $P_{max} > T_{a,k}$ , the gesture is classified as non-periodic and belongs to the  $k^{th}$  non-periodic gesture class. If this is not the case, go to the next step.

3. If  $P_{max} \leq T_{a,k}$ , the gesture is classified as ambiguous.

This ambiguity means that the gesture is either transitional or periodic. The ambiguity case must be resolved by the subsequent word/sentence recognition module. (Detail of how this can be achieved is, however, outside the scope of this paper.)

## 4. EXPERIMENTAL RESULT

Three experiments were carried out in this study. Details of the experiments are reported below.

### EXP#1

First we experimented with periodic/non-periodic segmentation by using the corresponding module as shown in Fig. 2. The experiment was carried out to find appropriate threshold  $T_p$ , as used in the algorithm of Section 3. A total of 880 gesture data sets were recorded by asking a signer to perform 10 different isolated signs drawn from the Thai sign language dictionary. Among 34 signs, they consist of 14 periodic signs, 10 non-periodic signs, and 10 static signs. The data sets corresponding to static signs also contain the same amount of transitional gesture segments. Each data set was put into the system in Fig. 2. The data sets classified as non-static were further classified by the periodic/non-periodic segmentation module. In this module, for each data set, the normalized peak magnitude,  $|c_{max}^p|$ , of its Discrete Fourier Transform (DFT) coefficients was computed (The normalization was performed by dividing the DFT spectral peak by the total sum of all DFT coefficient magnitudes). The result, categorized by the gesture types, is shown in Fig. 5.

From Fig. 5,  $T_p = 0.046$  was chosen empirically. By using this threshold value, it was found that 0.33% of the periodic gesture data sets have their  $|c_{max}^p|$  below the threshold, while all non-periodic and transitional gesture data sets were correctly classified as not periodic by using the threshold. This largely eliminates the need to perform HMM module evaluation for almost all periodic gestures. And because

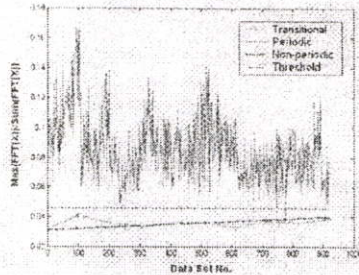


Fig. 5 The normalized peak magnitude for different gesture types, and the so-obtained empirical threshold  $T_{p,k}$ .

periodic gestures account for about 20% of all gestures found in the Thai sign language, by avoiding the need to perform more computing-intensive HMM module evaluation, this results in reduced system computational complexity.

#### EXP2

In the second experiment, the classification performance of the HMM module was evaluated. Ten commonly-used non-periodic gesture signs were taken from the Thai sign language dictionary. Each sign was performed by the same signer for forty times. The resulting 40 data sets for each sign were divided into 30 data sets used for HMM training, and the remaining 10 data sets for testing. All ten HMM models are based on a 3-state topology as shown in Fig. 4. The recognition rate for all ten gesture classes is shown in Table 1. From the table, it was found that all data sets were correctly classified. This perfect result is perhaps due to the limited number of gestures used in the experiment, however. Thus, more extensive experiment may be needed to confirm the result.

Table 1 Recognition rate for 10 non-periodic signs

Sign	Recognition Rate (%)
Turn left	100
Turn right	100
Beautiful	100
Italy	100
Nepal	100
Myanmar	100
Weather	100
Write	100
Fly	100
Walk on a rough surface	100

Because non-periodic signs account for a small percentage of all signs in the Thai sign language, evaluation of the HMM module to classify those non-periodic signs will not significantly increase the system's overall computational complexity.

#### EXP3

In this experiment, a total of 500 data sets drawn from all dynamic gesture types (periodic, non-periodic, and transitional) were used to obtain the empirical value of  $T_{p,k}$ . The 500 data sets consist of

- 420 data sets from 14 periodic signs
- 60 data sets from 6 transitional signs
- 400 data sets from 10 non-periodic signs.

Classification error was categorized into

- Missed classification: the rate at which the non-periodic gestures were incorrectly classified (as periodic or transitional gesture), by using the empirical thresholds as shown in Table 2.
- False classification: the rate at which either periodic or transitional gesture was classified, based on the empirical thresholds as shown in Table 2, as one of the non-periodic gestures.

Figs. 6-8 show the HMM probability values  $p(O|\lambda^k)$  for the third non-periodic sign ('beautiful') corresponding to periodic, non-periodic (excluding data sets belonging to the sign 'beautiful'), and transitional gesture data sets. The empirical value of  $T_{p,k}$  is also shown in all figures.

Table 2 Classification error rates, and empirical values of  $T_{p,k}$

Sign	Threshold	Missed Classification (%)	False Classification (%)
Turn left	$-7.13 \times 10^3$	0	0
Turn right	$-1.16 \times 10^4$	0	0
Beautiful	$-7.26 \times 10^3$	0	0
Italy	$-1.42 \times 10^3$	0	1.67%
Nepal	$-9.37 \times 10^3$	0	0
Myanmar	$-8.22 \times 10^3$	0	0
Weather	$-1.36 \times 10^4$	0	0
Write	$-9.34 \times 10^3$	0	0
Fly	$-1.36 \times 10^4$	0	0
Walk on a rough surface	$-1.98 \times 10^4$	0	0

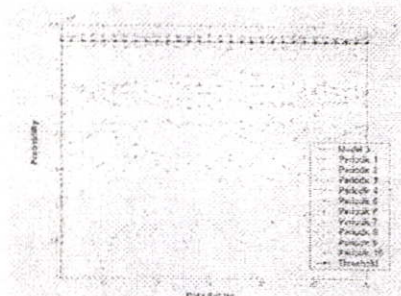


Fig. 6  $p(O|\lambda')$  corresponding to periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

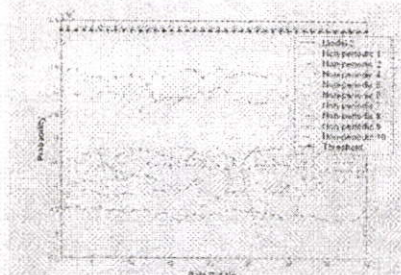


Fig. 7  $p(O|\lambda')$  corresponding to other non-periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

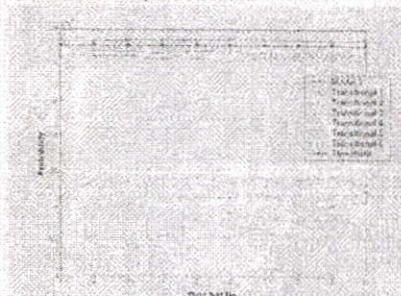


Fig. 8  $p(O|\lambda')$  corresponding to transitional data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

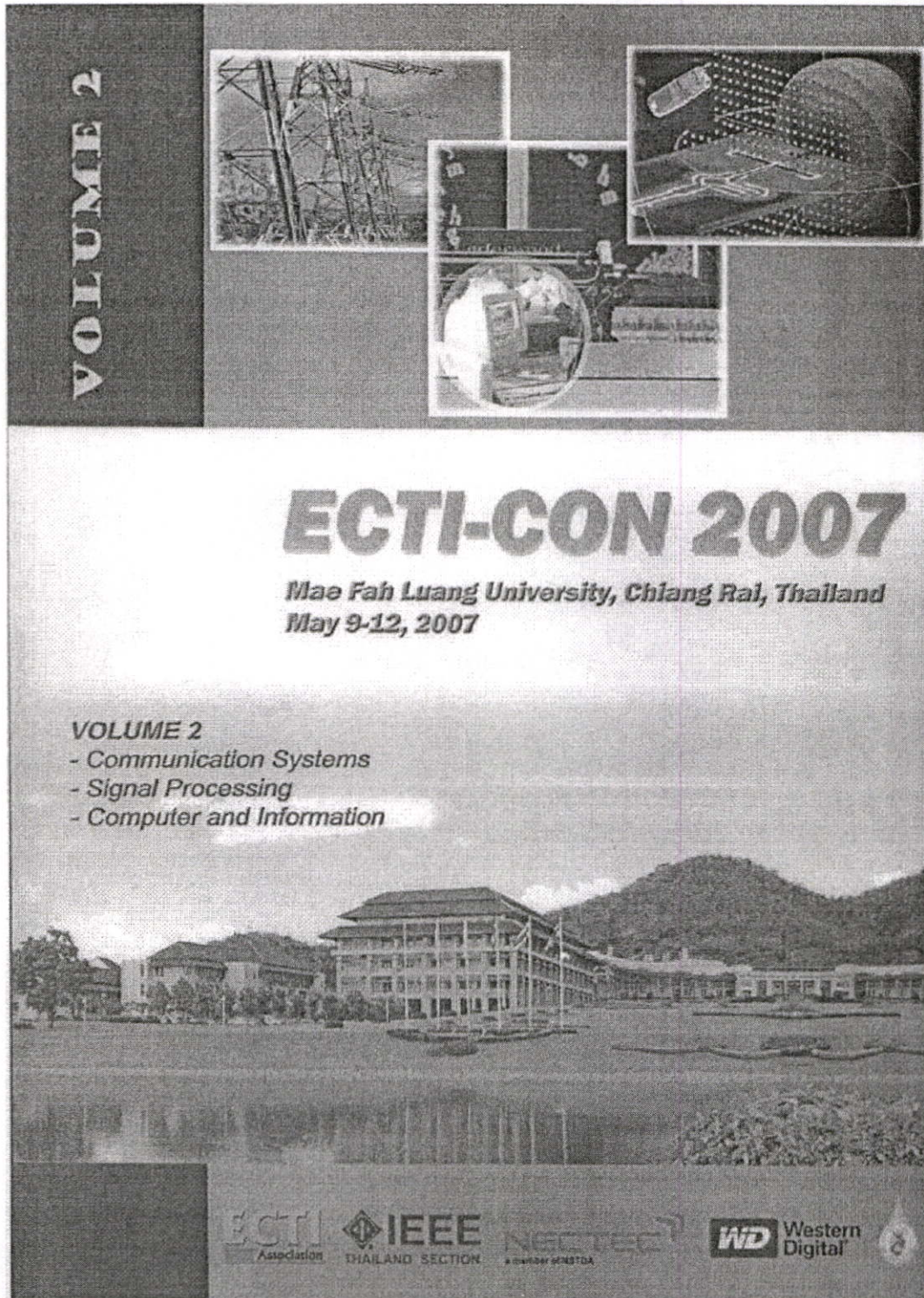
## 5. CONCLUSION

In this paper, a hybrid approach to automated Thai sign language translation has been proposed. The proposed method combines the HMM-based classifier with another two non-HMM-based classifiers developed in previous work. Because the HMM module is used for classifying a relatively small number of signs, the hybrid method retains the advantage of the non-HMM-based method in terms of computational saving. In addition, use of the HMM module here makes possible the transitional/non-periodic gesture classification. This classification of non-periodic gesture has been a major weakness in our previous Thai sign language translation system. As a result of this hybrid scheme, the system can now distinguish both periodic and non-periodic gestures from transitional movement. Future work includes more experiments based on a larger number of data sets.

## REFERENCES

- [1] Guolin F. and Wen G., "A SRNHMM system for signer-independent continuous sign language recognition", Proceedings, Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 20-21 May 2002, pp. 297-302.
- [2] Hermann Henz, Britta Bauer, and Karl-Friedrich Krüsis, "HMM-Based Continuous Sign Language Recognition Using Stochastic Grammars", In: Brattfort A. et al (Eds.): Gesture-Based Communication in Human Computer Interaction CW 99, 1999, pp. 185-196.
- [3] Vogler, C. and Metaxas, D., "ASL recognition based on a coupling between HMMs and 3D motion analysis", Sixth International Conference on Computer Vision, 4-7 Jan. 1998, pp. 363-369.
- [4] Wu J., Gao W., Song Y., Liu W. and Pang B. "A simple sign language recognition system based on data glove", Proceedings Fourth International Conference on Signal Processing (ICSP '98), Vol. 2, 12-16 Oct. 1998, pp. 1257-1260.
- [5] W. Werapan, and N. Chotikamthorn "Gesture Segmentation Technique for Thai Sign Language Translation", The Second International Symposium on Communications and Information Technology, Chonburi, Thailand, 2002, pp. 483-486.
- [6] Rabiner, L.R., "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, Vol. 77, Issue: 2, Feb. 1989, pp. 257-286.
- [7] The National Association of the Deaf in Thailand, *The Thai Sign Language Dictionary, Revised and Expanded Edition*, Thai Watana Panich, 1990.

2. Arunee Jaruwanawat and Nopporn Chotikakamthorn. "Continuous Thai Sign Language Recognition using HMM with Transition Models" ECTI Conference 2007 Volume 2, pp. 991-994



# Continuous Thai Sign Language Recognition using HMM with Transition Models

Arunce Jaruwawat and Nopporn Chotikakamthorn

Faculty of Information Technology, and  
Research Center for Communications and Information Technology  
King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand  
(E-mail: s5066025@kmitl.ac.th)

**Abstract**-This paper addresses the problem of continuous Thai sign language recognition. Sign language is typically composed of hand shapes and gestures formed by the movement of hands, arms and perhaps body also. One of the problems in performing continuous sign recognition is the existence of hand and arm movement in transition between two signs. This spontaneous transition movement conveys no meaning. However, not taking it into account results in poor recognition performance. This paper proposes the inclusion of algorithmic-based HMM transition movement models within a HMM-based sign language recognition system. A transition movement is represented by a single-state HMM model, of which its model parameters are obtained algorithmically based on the HMM parameters of the two signs involved in the transition. This avoids the need to collect extensive training data for modeling all possible transition movements in a conventional manner. In addition, to realize continuous sign language recognition, in this paper a pruning strategy for reducing the complexity of a Viterbi search is also described. Based on the experiment result, inclusion of such model within a HMM-based recognizer was found to offer improved recognition performance.

**Keywords**- Sign Language, Hidden Markov Model, Transition Movement

## I. INTRODUCTION

People with hearing and speech impairments communicate within their communities through the use of a sign language. Use of this specialized language, however, is limited to those who have acquired knowledge of the language. This largely prevents deaf people from communicating with those outside their communities. Also, while currently there exists a speech recognition system, allowing a computer user to input textual data through voice, there is not yet an equivalent product for deaf people. A sign language recognition system has been seen as an enabling technology to allow for natural human-computer interaction, as well as to assist in communications between those who understand the sign language as those who do not.

Previous works in continuous sign language recognition, with the emphasis on problems related to transition movement, include those of [1-6]. In [1], equipped with an electronic glove device and a 6-DOF tracking sensor, a method was proposed to segment continuous sentences such that transition movement can be removed before the segmented data being processed by a HMM-based recognizer. This 'hard' pre-segmentation strategy is, however, suboptimal as compared to a 'soft' segmentation inherent in a general HMM-based pattern recognition method, where segmentation and recognition are jointly performed to achieve optimal result. In [2, 4], transition movements are first extracted.

This extracted data is then used to perform data clustering using a Dynamic Time Warping (DTW) method. The so-obtained clusters are used to form HMM models for transition movements. The problem with this method based on conventional HMM model construction is its need for extensive training data to sufficiently cover possible transition movement patterns.

In [3], a large vocabulary sign language recognition is considered. Due to a large sign database, extensive Viterbi search is required. The solution suggested in [3] is based on a two-stage classification strategy. A sign gesture input is first classified into one of the clusters of similar signs. Then, a Viterbi search is performed on a small subset of HMM models corresponding to signs belonging to that cluster. Although efficient computation is obtained, this two-stage classification method is suboptimal for the search is performed over a subset of a set of all feasible solutions. It also results in poorer recognition rate as compared with a direct search over all feasible HMM models [1].

Most research works in sign language recognition deal with languages used by people of large population such as American and Chinese sign languages. However, it is known that a sign language, like its spoken language counterpart, is different from one region to another. For Thai sign language, early works include those of [5-6]. In [5, 6], a Thai sign language recognition system based on a statistical-based estimator was described. The method performs a pre-segmentation of input data from an electronic glove into static and dynamic gestures. Recognition is performed from the classified data using a Bayesian estimator. The method can perform recognition of signs with certain regularity in the movement patterns. However, it cannot deal with signs of irregular patterns, as well as transition movements. A solution to automated Thai sign language recognition based on HMM models was suggested in [5]. Use of HMM models allows signs of arbitrary movement patterns to be handled in a unified manner. However, the method in [5] was applied on a single-sentence basis. In [6] a method was developed to classify periodic gestures as found in Thai sign language and those with non-periodic nature, of which transitional gestures form a large part of them. However, the method in [6] can not differentiate between some non-periodic and yet meaningful gestures from those due to transition movements. No known work has been done on recognition of continuous Thai sign language containing both periodic and non-periodic gestures.

In this paper, a problem of continuous Thai sign language recognition is considered. In particular, the problem due to the existence of a transition movement is addressed. It is known that, by neglecting such spontaneous movement in modeling the recognition system, recognition performance is degraded [3]. The method described in this paper is based on a HMM model. Unlike those in [2, 4], this paper proposes an algorithmic-based method for construction of a transition model. This avoids the need for

collection and processing of extensive training data for transition movements. Also, it can be used to handle transition movement between any two arbitrary signs. To address complexity in performing Viterbi search on continuous signs, a pruning strategy based on the detection of static signs is also described.

This paper first introduces a general framework of a Thai sign language recognition system in Section 2. Then, a detailed discussion on modeling a transition movement is described in Section 3. A Viterbi search strategy is also given. Experimental result is provided in Section 4, with concluding remarks given in Section 5.

## II. THAI SIGN LANGUAGE RECOGNITION SYSTEM

Like other sign languages, Thai sign language is composed of hand shapes and gestures formed by the movement of hands, arms and body. Sometimes, facial expression also plays a crucial role in sign interpretation. Here, only part of the gestures formed by hand shapes and movement of hands and arms are considered. In this study (see Figure 1), a sign gesture is roughly classified into static (posture) and dynamic gestures. Dynamic gestures are further classified into sign movements and transition movements.

- Static gesture (posture) is a sign of which there is no hand or arm movement. Only the hand posture that defines the meaning of the sign.
- Dynamic gesture contains a hand/arm movement. It is further divided into
  - dynamic signs, which are meaningful hand movement, and
  - transition movements, which occur during the transition from one sign to another and convey not meaning.

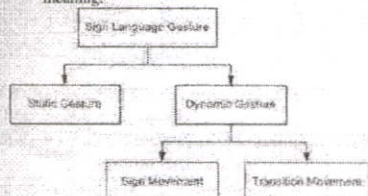


Figure 1. Classification of Thai sign gestures

An overall view of the Thai sign language recognition system, which contains the work described here as its part, is illustrated in Figure 2.

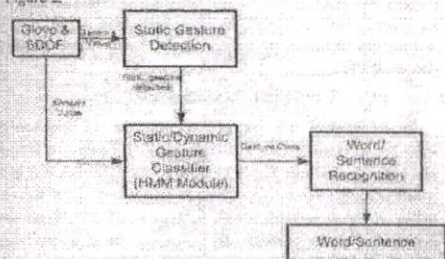


Figure 2. Block diagram of the continuous Thai sign language recognition system.

From the figure, the system is composed of four main parts. The first one is responsible for acquisition of hand shapes and movements. Here, an electronic glove with a 6-DOF electromagnetic sensor is used as an acquisition device. The second part is the HMM-based gesture recognizer. The third part is the static gesture detector. The result of the detection is used to assist the HMM recognizer in selecting feasible Viterbi paths (pruning). And the last part is the word/sentence recognizer. This part uses the result from the gesture recognizer, along with a sign language model, to produce the final recognized word/sentence. In this paper, the focus is mainly on the second and, to a lesser extent, the third parts of the system.

## III. CONTINUOUS THAI SIGN LANGUAGE RECOGNITION WITH ALGORITHMIC-BASED TRANSITION MODELS

### A. Modeling the transition movement

Here, an HMM model is adopted for the recognition of both static and dynamic signs. As pointed out by previous work [2, 4], transition movement should be incorporated in the model to improve recognition rate. Instead of directly training the model using data of transition movements, in this paper we propose that a transition movement is modeled using a single state HMM of which its model parameters are obtained algorithmically at run-time. Thus, given two signs (the last states of the word models A and B respectively) connected by a particular transition movement, the series of interconnected HMM models is described by Figure 3.

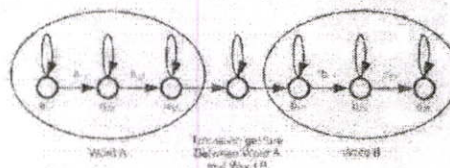


Figure 3. Series of HMM states, connected by a transition state, to model continuous signing.

In our study, a single-state HMM transition model is constructed at run-time based on HMM parameters corresponding to the last state of Word A and the first state of Word B. To explain the algorithm, some notations for HMM models are first introduced. Let  $(\pi, A, \mu, U)$  be an HMM model parameters corresponding to the first word (Word A in Figure 3). Here,  $\pi$  is the initial probability of being in the first state,  $A$  is a matrix of state transition probabilities. In addition,  $\mu_i$  is the mean vector of the  $i^{\text{th}}$  state outcome and  $U_i$  is the corresponding covariance matrix. Similarly,  $(\pi', A', \mu', U')$  are HMM parameters corresponding to the second word (Word B in Figure 3). In addition, we define  $(\pi'', A'', \mu'', U'')$  as model parameters corresponding to the single-state transition model. These parameters are obtained below.

First, because each transition model is designed for a particular pair of sign words, its initial probability can be set to  $1 - a_{11}$ , where  $a_{11}$  is the probability that the system, once reaching the last state in

Word A, remains on the same state. Next, because the model contains a single state,  $\bar{A}$  is scalar. It is set to  $1 - \pi'$ . The model parameters related to the outcome probability is obtained according to the following equations.

$$\bar{\mu} = \frac{\mu_s + \mu'_s}{2} \quad (1)$$

$$\bar{U}_{11} = \frac{U_{s,11} + U'_{s,11}}{2} \quad (2)$$

$$\bar{U}_{12} = \bar{U}'_{21} = 0 \quad (3)$$

$$\bar{U}_{22} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_N^2 \end{bmatrix} \quad (4)$$

From Eq. (2),  $\bar{U}_{11}$  is the  $M \times M$  top-left submatrix of  $\bar{U}$ , where  $M$  is the number of data points obtained from an electronic glove. Similarly,  $U_{s,11}$  and  $U'_{s,11}$  are the  $M \times M$  top-left submatrices of  $U_s$  and  $U'_s$  respectively. In Eqs (3, 4),  $\bar{U}_{12}$ ,  $\bar{U}_{21}$ , and  $\bar{U}_{22}$  are the remaining submatrices of  $\bar{U}$  (see Figure 4, for an illustration). The size  $N$  of  $\bar{U}_{22}$  is equal to the number of data points obtained from a 6-DOF electromagnetic location sensor ( $N = 6$ ). The variance parameters in Eq. (4) are obtained from the following equation.

$$\sigma_i^2 = \frac{(\mu_{s,2+i} - \mu'_{s,2+i})^2}{\alpha} \quad (5)$$

Here,  $\mu_{s,2+i}$  is the mean value at the third state in the first word model, corresponding to the  $i^{\text{th}}$  data point from the location sensor. In addition,  $\mu'_{s,2+i}$  should be interpreted similarly. From the same equation,  $\alpha$  is a scaling factor, arbitrarily chosen as 6.

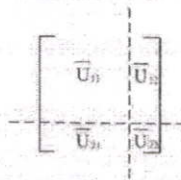


Figure 4. Decomposition of  $\bar{U}$  into four submatrices

**B. Viterbi search strategy**

Generally, pattern recognition by HMM model involves a search for a best path (i.e., a sequence of HMM states) using a Viterbi algorithm. However, for continuous sign language recognition, with medium to large vocabulary support, the search complexity and space is increased significantly. To maintain the search complexity and memory space requirement to a practically realizable level, certain search strategy must be applied to a standard Viterbi algorithm. Here, by observing that many words in Thai sign language are composed at least partly of static signs,

those static moments are detected and utilized (along with some heuristic rules) as part of the pruning strategy in an attempt to keep the number of active paths to an acceptable figure. The search strategy used here is summarized in Figure 5.

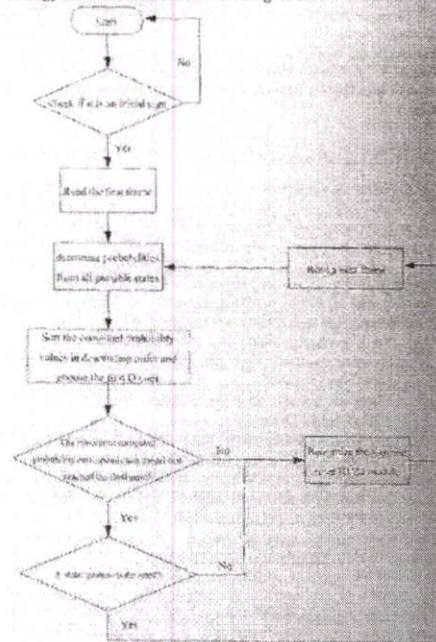


Figure 5. A flow chart describing a pruning strategy, to be used with a standard Viterbi algorithm

From Figure 5, the parameter  $Q$  is the maximum number of feasible Viterbi paths that can be kept at any time instance. Therefore, from time to time, rules must be enforced to select  $Q$  best paths from all candidates. Here, in addition to a simple accumulated probability sorting, it is proposed that a detection of static sign is used to significantly reduce infeasible paths. Detail of a static sign detector is not given here. It is, however, similar to that used in [6].

**IV. EXPERIMENTAL RESULT**

The experiment was carried out based on 13 sentences of everyday use. Each sentence is composed of 2-3 words on average and was selected from [7]. For each sentence, 20 data sets were collected using a right-hand electronic glove and a 6-DOF location sensor. A signer performed each sentence by starting in the pre-defined initialized position. In summary, 40 data sets were collected for each sign word. Of these, 30 sets were used for training and the remaining data sets were used for testing.

Two HMM-based sign language recognition techniques were applied to the same data sets. The first one employed the proposed transition models as part of the HMM models. The second one did not model the transition movement. For the latter model, each HMM sign model was simply connected to the next one. The results are shown in Tables 1 and 2. From the tables, it is seen that the proposed method offers noticeable performance improvement. It is noted, however, that this is achieved without additional work on training the transition movement.

Table 1. Recognition performance using the method with algorithmic-based transition models

Sentence	Word sequence	Accuracy (%)
1. I was born in Denmark.	Born-Denmark	100
2. I met him in Denmark.	Meet-Him-Denmark	80
3. Please, sit down.	Please-Sit	100
4. Your mom is fine.	Mother-Fine	95
5. Today is Monday.	Today-Monday	95
6. I take a rest on Friday.	Friday-Rest	80
7. I have ten relatives.	Relative-Ten	85
8. I have ten goats.	Goat-Ten	85
9. I study English on Monday.	Monday-Study-English	90
10. I'll go swimming in the day after tomorrow.	Swim-The day after tomorrow	100
11. Orange has a sour taste.	Orange-Sour	90
12. Elephant is big.	Elephant-Big	80
13. Frog eat fly.	Frog-Eat-Fly	90

Table 2. Recognition performance using the method without modeling transition movement.

Sentence	Word sequence	Accuracy (%)
1. I was born in Denmark.	Born-Denmark	75
2. I met him in Denmark.	Meet-Him-Denmark	70
3. Please, sit down.	Please-Sit	85
4. Your mom is fine.	Mother-Fine	80
5. Today is Monday.	Today-Monday	85
6. I take a rest on Friday.	Friday-Rest	70
7. I have ten relatives.	Relative-Ten	70
8. I have ten goats.	Goat-Ten	75
9. I study English on Monday.	Monday-Study-English	75
10. I'll go swimming in the day after tomorrow.	Swim-The day after tomorrow	80
11. Orange has a sour taste.	Orange-Sour	75
12. Elephant is big.	Elephant-Big	70
13. Frog eat fly.	Frog-Eat-Fly	75

## V. CONCLUSION

In this paper, the problem of modeling transition movements for Thai sign language recognition has been addressed. The transition models, based on algorithmic construction at run-time, have been described. It has been verified through the experiment that the use of the proposed transition model offers improved performance over the standard HMM-based method without transition movement modeling. The improvement is attained without the need for additional training work for modeling these spontaneous movements.

## REFERENCES

- [1] F. Gao and G. Wen, "A SRNMM system for signer-independent continuous sign language recognition", *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 20-21 May 2002, Page(s): 297 -302
- [2] G. Wen, F. Gao, Z. Dohu and C. Yikang, "Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition", *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 17-19 May 2004, Page(s):553 - 558
- [3] Y. Guolin, Y. Hongxin, L. Xin and J. Feng, "Real Time Large Vocabulary Continuous Sign Language Recognition Based on OP-Viterbi Algorithm", *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, Volume 3, 20-24 Aug. 2006, Page(s):312 - 315
- [4] G. Fang, W. Gao and D. Zhao, "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models", *Systems, Man and Cybernetics, Part A, IEEE Transactions on* Volume 37, Issue 1, Jan. 2007, Page(s):1 - 9
- [5] C. Stam and V. Wittichai, "Isolated Thai Sign Language Recognition System using Hidden Markov Models", *Proceedings of the 25th National Electrical Engineering Conference (EECON-25)*, Hadyai, Thailand, November 21-22, 2002.
- [6] Weerawat Werapan, "Gesture Segmentation and Recognition Techniques for Automated Thai Sign Language Translation", *Master of Science in Information Technology School of Graduate Studies, King Mongkut's Institute of Technology Ladkrabang, 2005.*
- [7] The National Association of the Deaf in Thailand, *The Thai Sign Language Dictionary. Revised and Expanded Edition*, Thai Wistana Patich, 1990.

## ประวัติผู้เขียน

ชื่อผู้เขียน

วัน เดือน ปีเกิด

วุฒิการศึกษาระดับปริญญาตรี

สถานที่สำเร็จการศึกษา

ปีที่สำเร็จการศึกษา

ประสบการณ์การทำงาน

นางสาวอรุณี จารუნาวัดน์

วันที่ 29 กรกฎาคม พ.ศ. 2523

วิทยาศาสตร์บัณฑิต สาขาวิทยาการคอมพิวเตอร์

มหาวิทยาลัยบูรพา

2545

พฤศจิกายน 2544 – เมษายน 2545 ตำแหน่งผู้ช่วย

นักวิจัย ประจำห้อง Multimedia and Virtual

Research Lab. สำนักวิจัยการสื่อสารและ

เทคโนโลยีสารสนเทศ

เมษายน – พฤษภาคม 2550 ตำแหน่งผู้ช่วยนักวิจัย

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระ

จอมเกล้าเจ้าคุณทหารลาดกระบัง