



ระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ  
EMG-based Thai Speech Recognition

ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ  
Acting Sub. Lt. Niyawadee Srisuwan

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา  
ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า  
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy in Electrical Engineering  
Prince of Songkla University

2559

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์



ระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ  
EMG-based Thai Speech Recognition

ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ  
Acting Sub. Lt. Niyawadee Srisuwan

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา  
ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า  
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy in Electrical Engineering  
Prince of Songkla University

2559

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์      ระบบบำรุงการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ  
ผู้เขียน              ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ  
สาขาวิชา            วิศวกรรมไฟฟ้า

---

**อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก**

**คณะกรรมการสอบ**

.....ประธานกรรมการ  
(รองศาสตราจารย์ ดร. ชูศักดิ์ ลิ้มสกุล)      (รองศาสตราจารย์ ดร. มนต์รี กาญจนเดชะ)

**อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม**

.....กรรมการ  
(รองศาสตราจารย์ ดร. ชูศักดิ์ ลิ้มสกุล)

.....  
(รองศาสตราจารย์ ดร. พรชัย พฤษภักทรานนท์)

.....กรรมการ  
(รองศาสตราจารย์ ดร. พรชัย พฤษภักทรานนท์)

.....กรรมการ  
(รองศาสตราจารย์ ดร. มิตรชัย จงเขียวชำนาญ)

.....กรรมการ  
(รองศาสตราจารย์ ดร. ชูชาติ ปิณฑวิรุจน์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้รับวิทยานิพนธ์ฉบับนี้  
เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

.....  
(รองศาสตราจารย์ ดร. วีระพล ศรีชนะ)

คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ .....

(รองศาสตราจารย์ ดร. ชูศักดิ์ ลิ้มสกุล)

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ลงชื่อ .....

(รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนท์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

ลงชื่อ .....

(ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ)

นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน  
และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ .....

(ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ)

นักศึกษา

ชื่อวิทยานิพนธ์	ระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ
ผู้เขียน	ว่าที่ร้อยตรีหญิง นິยวดี ศรีสุวรรณ
สาขาวิชา	วิศวกรรมไฟฟ้า
ปีการศึกษา	2559

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างต้นแบบของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งสามารถรู้จำคำพูดตัวเลขภาษาไทย 11 คำ และรู้จำวรรณยุกต์ 5 เสียง โดยได้ออกแบบตำแหน่งอิเล็กโทรดที่ใช้ในการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อของการพูด ออกแบบอุปกรณ์และวิธีการเก็บข้อมูลของสัญญาณ และออกแบบเทคนิคในการประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อ ในส่วนของการประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อมีการศึกษาและเปรียบเทียบเทคนิคต่างๆ ในการประมวลผล เพื่อเลือกเทคนิคที่เหมาะสมที่สุด โดยได้เปรียบเทียบเทคนิคการคัดเลือกลักษณะเด่น และเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด ซึ่งพบว่าเทคนิคการคัดเลือกลักษณะเด่นที่ให้ความแม่นยำในการจำแนกสูงสุด คือ เทคนิคการเลือกลักษณะเด่นแบบแรปเปอร์ซึ่งใช้ผลความแม่นยำในการจำแนกของตัวจำแนกเชิงเส้นแบบเบย์นอร์มัลเป็นตัวประเมินลักษณะเด่น และใช้เทคนิคการค้นหาเซตย่อยของลักษณะเด่นแบบไปข้างหน้าสำหรับการเปรียบเทียบประสิทธิภาพของตัวจำแนก พบว่าตัวจำแนกเชิงเส้นแบบเบย์นอร์มัลเป็นตัวจำแนกที่ให้ความแม่นยำในการจำแนกสูงสุดและใช้เวลาในการประมวลผลน้อย โดยความแม่นยำเฉลี่ยในการจำแนกสำหรับการจำแนกคำพูดหลัก 11 คำ คือ 92.45% และ 81.45% สำหรับการพูดแบบออกเสียง และ การพูดแบบไม่ออกเสียง ตามลำดับ สำหรับการจำแนกวรรณยุกต์ 5 เสียงได้รับความแม่นยำเฉลี่ย 75.28% สำหรับการพูดแบบออกเสียง ซึ่งความแม่นยำเฉลี่ยดังกล่าวคำนวณจากความแม่นยำของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร 10 คน หลังจากการออกแบบระบบในส่วนของระบบรู้จำ ได้สร้างแบบจำลองการทำงานของระบบโดยนำส่วนของการจำแนกคำพูดหลัก 11 คำ และการจำแนกวรรณยุกต์มารวมกัน และแปลงเอาต์พุตที่ได้จากระบบรู้จำไปเป็นเสียงพูด

**Thesis Title**           EMG-based Thai Speech Recognition  
**Author**                 Acting Sub. Lt. Niyawadee Srisuwan  
**Major Program**       Electrical Engineering  
**Academic Year**       2016

## ABSTRACT

The objective of this thesis is to create the prototype of Thai speech recognition and synthesis based on Electromyography signal (EMG) that can classify 11 Thai isolated words (11 numbers) and 5 Thai tones. We designed the suitable electrode positions of EMG capturing, the process of EMG capturing, and the technique of feature selection and learning algorithm. In order to get the best technique, we compared the performance of the feature selection methods and 8 types of learning algorithm (classifier). From our results, the best feature selection technique was the wrapper method that use an accuracy rate of the linear Bayes normal classifier as a feature evaluation criterion (W\_LBN) and the forward search (FW) for finding the best feature subset. For the classifier comparison, we found that the linear Bayes normal classifier (LBN) yielded the best result. For the 11 words classification, the best average accuracy rate was 92.45% and 81.45% for audible and silent speech respectively. For the 5 tones classification, we obtained a 75.28% average classification rate that were calculated from the accuracy of 10 subjects. Moreover, the demonstration of the working of this system was created by MATLAB. The 11 words and 5 tones classification model were employed to classify 11 words and 5 tones. The output from recognition part was converted to sound speech.

## กิตติกรรมประกาศ

สำหรับการดำเนินการวิจัยและจัดทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอขอบคุณรองศาสตราจารย์ ดร. ชูศักดิ์ ลิ้มสกุล ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และ รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนต์ ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ในการให้ความช่วยเหลือให้คำปรึกษาในการทำวิจัยในทุกๆ ด้าน ทั้งในส่วนของการทำงานวิจัยและการเขียนบทความเพื่อการประชุมวิชาการและตีพิมพ์ลงวารสาร

ขอขอบคุณประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. มนต์รี กาญจนเดชะ และกรรมการสอบวิทยานิพนธ์ทุกท่าน รองศาสตราจารย์ ดร. มิตรชัย จงเขี้ยว ชำนาญ รองศาสตราจารย์ ดร. ชูชาติ ปิณฑวิรุจน์ รองศาสตราจารย์ ดร. ชูศักดิ์ ลิ้มสกุล และ รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนต์ ที่ได้อุทิศเวลาในการเป็นกรรมการสอบวิทยานิพนธ์ ให้คำแนะนำ และตรวจทานวิทยานิพนธ์

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ที่ให้ความร่วมมือในการเป็นอาสาสมัครในการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อเพื่อใช้ในการวิจัย

ขอขอบคุณโครงการร่วมให้ทุนปริญญาเอกกาญจนาภิเษก (คปก.) ระหว่างสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) และมหาวิทยาลัยสงขลานครินทร์ (สัญญาเลขที่ PHD/0164/2552) ที่ให้การสนับสนุนด้านการเงินสำหรับการทำวิจัยนี้ทั้งหมด ทั้งการทำวิจัยในประเทศไทยและในประเทศเยอรมัน

ขอขอบคุณ Prof. Dr. Tanja Schultz และ เพื่อนๆ ที่ Cognitive Systems Lab (CSL) ประเทศเยอรมัน สำหรับการให้ความช่วยเหลือและให้คำปรึกษาในการทำงานวิจัยร่วมกับห้องทดลองที่ประเทศเยอรมัน

ขอขอบคุณอาจารย์และบุคลากรภาควิชาวิศวกรรมไฟฟ้าทุกคนที่ให้คำปรึกษาและให้ความช่วยเหลือในการทำงานวิจัยนี้

สุดท้ายนี้ขอขอบคุณ คุณพ่อ คุณแม่ และญาติพี่น้อง ที่คอยให้กำลังใจในการทำงานวิจัยนี้ให้สำเร็จลุล่วง



## สารบัญ

	หน้า
สารบัญ.....	(8)
รายการรูปภาพ.....	(11)
รายการตาราง.....	(19)
สารบัญคำศัพท์.....	(22)
สารบัญคำย่อ.....	(29)
<b>บทที่ 1 บทนำ.....</b>	<b>1</b>
1.1 ความเป็นมาของงานวิจัย.....	1
1.2 การตรวจเอกสาร.....	2
1.2.1 กระบวนการเก็บข้อมูลของสัญญาณ.....	4
1.2.2 กระบวนการเตรียมสัญญาณเบื้องต้น.....	13
1.2.3 กระบวนการสกัดลักษณะเด่นของสัญญาณ.....	16
1.2.4 กระบวนการลดมิติของข้อมูล.....	23
1.2.5 กระบวนการรู้จำระบบ.....	23
1.2.6 การประยุกต์ใช้งาน.....	26
1.2.7 สรุปการตรวจเอกสาร.....	27
1.3 วัตถุประสงค์.....	28
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	28
1.5 ขอบเขตงานวิจัย.....	29
<b>บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....</b>	<b>30</b>
2.1 ระบบรู้จำเสียงพูดอัตโนมัติ.....	30
2.2 ระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	31
2.3 กลไกการสร้างเสียงพูด.....	33
2.4 กล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูด.....	34
2.5 ระบบเสียงพูดในภาษาไทย.....	39
2.6 สัญญาณไฟฟ้ากล้ามเนื้อ.....	42

## สารบัญ (ต่อ)

	หน้า
2.7 การคำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล้ำมเนื้อ.....	43
2.7.1 ลักษณะเด่นที่วิเคราะห์ในโดเมนเวลา.....	44
2.7.2 ลักษณะเด่นที่วิเคราะห์ในโดเมนความถี่.....	50
2.8 การคัดเลือกลักษณะเด่น.....	50
2.9 ตัวจำแนกรูปแบบที่ใช้ในการจำแนกสัญญาณไฟฟ้ากล้ำมเนื้อ.....	59
2.10 การสังเคราะห์เสียงพูด.....	72
<b>บทที่ 3 วิธีการวิจัย.....</b>	<b>74</b>
3.1 การออกแบบรูปแบบการเก็บข้อมูลของสัญญาณ.....	74
3.1.1 การออกแบบตำแหน่งอิเล็กทรอนิกส์โทรด.....	75
3.1.2 การออกแบบอุปกรณ์และวิธีการเก็บข้อมูลของสัญญาณ.....	82
3.1.3 การออกแบบคำพูด จำนวนคำพูด วิธีการพูด จำนวนอาสาสมัคร และประเภทอาสาสมัครที่ใช้สำหรับการเก็บข้อมูลสัญญาณ.....	87
3.2 กระบวนการเตรียมสัญญาณเบื้องต้น.....	91
3.2.1 การกรองสัญญาณรบกวน.....	91
3.2.2 การตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ำมเนื้อ.....	93
3.3 การประมวลผลสัญญาณไฟฟ้ากล้ำมเนื้อเพื่อออกแบบระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อ.....	104
3.3.1 การออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด).....	104
3.3.2 การออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง).....	137
3.3.3 สรุปผลการออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด) และการออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง).....	164
<b>บทที่ 4 การปรับปรุงประสิทธิภาพของระบบรู้จำ.....</b>	<b>165</b>
4.1 การจำแนกคำพูดหลัก 11 คำ และวรรณยุกต์ 5 เสียง โดยใช้ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน (TDN) ซึ่งใช้ตัวจำแนก HMM ในการจำแนกระบบ.....	165
4.2 การเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อ.....	171

## สารบัญ (ต่อ)

	หน้า
4.3 การเปรียบเทียบตัวประเมินลักษณะเด่น 14 ชนิด สำหรับระบบรู้จำการพูดภาษาไทย โดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	180
4.4 การเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกเซตย่อยของลักษณะเด่นโดยผ่านกระบวนการค้นหา.....	189
4.5 การคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกตัวเลขไทย 11 คำ โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN.....	198
4.6 การคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW.....	203
4.7 สรุปผลการปรับปรุงระบบ.....	210
<b>บทที่ 5 การสร้างแบบจำลองการทำงานของระบบ.....</b>	<b>212</b>
<b>บทที่ 6 บทสรุปและข้อเสนอแนะ.....</b>	<b>218</b>
6.1 สรุปผลการวิจัย.....	218
6.2 ข้อจำกัดและแนวทางวิจัยในอนาคต.....	222
6.3 ข้อเสนอแนะ.....	224
บรรณานุกรม.....	226
ภาคผนวก.....	237
ภาคผนวก ก. สัทอักษรสากล.....	238
ประวัติผู้เขียน.....	239

## รายการรูปภาพ

รูปที่	หน้า	
1.1	แผนภูมิแสดงกลุ่มของงานวิจัยซึ่งทำการวิจัยเกี่ยวกับระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	3
1.2	แผนภูมิแสดงจำนวนบทความของงานวิจัยเกี่ยวกับระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	4
1.3	อิเล็กทรอนิกส์ชนิดต่าง ๆ ที่ถูกนำมาใช้ในงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	6
1.4	ตำแหน่งของอิเล็กทรอนิกส์ที่ใช้ในงานวิจัยจากการทบทวนวรรณกรรม (ก) ระบุกล้ามเนื้อโดยตรง (ข) ระบุตำแหน่งบนผิวหนัง.....	8
1.5	จำนวนช่องสัญญาณของการจับสัญญาณไฟฟ้ากล้ามเนื้อที่ถูกใช้ในงานวิจัยจากการทบทวนวรรณกรรม.....	11
1.6	จำนวนข้อมูลของอาสาสมัครในการจับสัญญาณ สำหรับงานวิจัยนี้จากการทบทวนวรรณกรรม.....	12
1.7	จำนวนคำพูดที่ใช้ในการรู้จำ จากการทบทวนวรรณกรรม ซึ่งมีทั้งการรู้จำคำพูดโดด คำพูดโดด และคำพูดต่อเนื่อง.....	12
1.8	ลักษณะเด่น 10 อันดับแรกที่ถูกนำมาทดสอบมากที่สุด จากการทบทวนวรรณกรรม.....	21
1.9	ตัวจำแนกที่ถูกใช้ในงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	25
2.1	โครงสร้างของระบบรู้จำเสียงพูดอัตโนมัติ.....	30
2.2	โครงสร้างของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ.....	33
2.3	อวัยวะในการสร้างเสียงพูด.....	33
2.4	กล้ามเนื้อที่เกี่ยวข้องสำหรับกลไกการออกเสียงพูด.....	39
2.5	กราฟแสดงรูปแบบการเปลี่ยนแปลงระดับเสียงซึ่งทำให้เกิดเสียงวรรณยุกต์ 5 เสียง.....	42
2.6	แสดงการแบ่งเฟรมของสัญญาณไฟฟ้ากล้ามเนื้อแต่ละสัญญาณ.....	48
2.7	การคำนวณลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน.....	49
2.8	กระบวนการคัดเลือกลักษณะเด่น.....	52

## รายการรูปภาพ (ต่อ)

รูปที่		หน้า
2.9	แสดงแผนภาพสถานะแบบไม่ต่อเนื่องของแบบจำลองมาร์คอฟพื้นฐาน.....	61
2.10	แสดงแผนภาพ HMM 3 สถานะ.....	61
2.11	แบบจำลองเซลล์ประสาทชั้นเดียว.....	63
2.12	การจำแนกโดยตัวจำแนก KNN.....	65
2.13	ระบบสร้างเสียงพูดโดยใช้การแปลงข้อความเป็นเสียงพูด.....	73
3.1	โครงสร้างหลักของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อในส่วน ของระบบรู้จำ.....	74
3.2	เครื่องขยายและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อยี่ห้อ Mobi6.....	75
3.3	อิเล็กโทรดแบบผิวหน้า Ag/AgCl ชนิดใช้แล้วทิ้งยี่ห้อ ARBO.....	75
3.4	ตำแหน่งของกล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูด กล้ามเนื้อที่อยู่ในกรอบ สี่เหลี่ยมเป็นกล้ามเนื้อที่ถูกคัดเลือก.....	77
3.5	สัญญาณไฟฟ้ากล้ามเนื้อของการพูดเสียงวรรณยุกต์ 5 เสียง "ดู, ดู, ดู, ดู, ดู" ซึ่งได้ จากกล้ามเนื้อบริเวณหน้า (Depressor Anguli Oris).....	77
3.6	กล้ามเนื้อบริเวณคอ (strap muscles) ซึ่งมีผลในการออกเสียงวรรณยุกต์.....	78
3.7	ตำแหน่งของอิเล็กโทรดที่ใช้สำหรับการทดลองเพื่อเลือกตำแหน่งที่เหมาะสม สำหรับการแยกเสียงวรรณยุกต์.....	79
3.8	สัญญาณไฟฟ้ากล้ามเนื้อจากกล้ามเนื้อบริเวณส่วนล่างของลำคอสำหรับการออก เสียงวรรณยุกต์แต่ละเสียง.....	80
3.9	ตำแหน่งอิเล็กโทรดที่ดีที่สุดจากการทดลอง เพื่อใช้ในการแยกเสียงวรรณยุกต์.....	80
3.10	สรุปตำแหน่งอิเล็กโทรด 5 ตำแหน่งที่ถูกคัดเลือก.....	82
3.11	รูปแบบการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ.....	84
3.12	ซ้าย: ตัวอย่างวิดีโอที่ใช้ขณะเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อขณะออก เสียงพูด ขวา: ลักษณะการกวดสวิตซ์ขณะการออกเสียงพูดเพื่อเก็บข้อมูล สัญญาณไฟฟ้ากล้ามเนื้อ.....	84

## รายการรูปภาพ (ต่อ)

รูปที่	หน้า
3.13 ภาพขณะทำการทดลองเพื่อเก็บข้อมูลของสัญญาณจากอาสาสมัคร 1) อิเล็กโทรดแบบผิวหน้า 2) สายนำสัญญาณ 3) Mobi6 4) หน้าจอสำหรับแสดงวิดีโอ 5) คอมพิวเตอร์ 6) วงจรสร้างสัญญาณการกวดสวิตช์ 7) สวิตช์แบบกดติดปลั๊กยัด.....	86
3.14 ลักษณะการบันทึกสัญญาณการออกเสียงพูด "ศุณย์" ถึง "สิบ" ที่ใช้ในการจำแนกคำพูดหลัก.....	88
3.15 ลักษณะการบันทึกสัญญาณการออกเสียงพูดคำพูดที่ใช้ในการจำแนกวรรณยุกต์.....	88
3.16 สัญญาณไฟฟ้ากล่อมเนื้อช่องที่ 1 สำหรับการพูด "ศุณย์" ถึง "สิบ" ภาพบน: สัญญาณดิบ ภาพล่าง: สัญญาณที่ผ่านการกรองโดยใช้ Butterworth filter และ Notch filter.....	92
3.17 สัญญาณไฟฟ้ากล่อมเนื้อช่องที่ 4 สำหรับการพูด "ศุณย์" ถึง "สิบ" ภาพบน: สัญญาณดิบ ภาพล่าง: สัญญาณที่ผ่านการกรองโดยใช้ Butterworth filter และ Notch filter.....	93
3.18 การตัดสัญญาณไฟฟ้ากล่อมเนื้อโดยใช้สัญญาณการกวดสวิตช์เป็นตัวกำหนดขอบเขต.....	94
3.19 การปรากฏของสัญญาณคลื่นไฟฟ้าหัวใจที่ไม่สามารถกำจัดได้หมด ซึ่งจะปรากฏเฉพาะสัญญาณช่องที่ 4 และ ช่องที่ 5.....	97
3.20 การปรากฏของสัญญาณไฟฟ้ากล่อมเนื้อที่ไม่ได้เกิดจากการพูดซึ่งอยู่นอกช่วงการกวดสวิตช์.....	98
3.21 แสดงตำแหน่งของจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อขณะพูดสำหรับแต่ละช่องสัญญาณ ซึ่งมีเวลาเริ่มต้นและสิ้นสุดไม่พร้อมกัน.....	99
3.22 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อ จากสัญญาณของคำว่า "สี่" ซึ่งผลการตรวจจับถูกต้อง.....	102
3.23 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อ ซึ่งผลของการตรวจจับเข้าเงื่อนไขของการตรวจจับผิดพลาด.....	103

## รายการรูปภาพ (ต่อ)

รูปที่	หน้า	
3.24	กราฟแสดงการเปรียบเทียบผลการประเมินโดยใช้ RES สำหรับการทดลองสามรูปแบบแรก คือ การใช้ข้อมูล 5 ช่องสัญญาณ 3 ช่องสัญญาณ และ 2 ช่องสัญญาณสุดท้าย สำหรับการพูดแบบออกเสียง.....	110
3.25	กราฟแสดงการเปรียบเทียบผลการประเมินโดยใช้ RES สำหรับการทดลองสามรูปแบบแรก คือ การใช้ข้อมูล 5 ช่องสัญญาณ 3 ช่องสัญญาณ และ 2 ช่องสัญญาณสุดท้าย สำหรับการพูดแบบไม่ออกเสียง.....	111
3.26	กราฟแสดงผลการประเมินโดยใช้ RES สำหรับการทดลองรูปแบบสุดท้าย การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 1 ช่องสัญญาณ (การพิจารณาแต่ละช่องสัญญาณ) สำหรับการพูดแบบออกเสียง...	112
3.27	กราฟแสดงผลการประเมินโดยใช้ RES สำหรับการทดลองรูปแบบสุดท้าย การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 1 ช่องสัญญาณ (การพิจารณาแต่ละช่องสัญญาณ) สำหรับการพูดแบบไม่ออกเสียง.....	113
3.28	กราฟเปรียบเทียบผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูดโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวนช่องสัญญาณที่แตกต่างกัน.....	118
3.29	กราฟเปรียบเทียบผลการประเมินความสามารถในการจำแนกกลุ่มของลักษณะเด่นสัญญาณไฟฟ้ากล่อมเนื้อ โดยใช้ RES เมื่อใช้จำนวนช่องสัญญาณที่แตกต่างกัน.....	118
3.30	กราฟเปรียบเทียบผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูด สำหรับการใช้อินพุตของตัวจำแนก ANN ที่แตกต่างกัน.....	121
3.31	แผนภาพแสดงขั้นตอนการประเมินและเลือกลักษณะเด่นโดยอ้างอิงข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน.....	125
3.32	กราฟแสดงผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อปรับจำนวนลักษณะเด่น 3 5 7 9 ตามลำดับ ซ้าย: รูปแบบที่หนึ่ง ขวา: รูปแบบที่สอง บน: ออกเสียง ล่าง: ไม่ออกเสียง.....	132

## รายการรูปภาพ (ต่อ)

รูปที่	หน้า	
3.33	แผนภูมิแท่งแสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ (แกนปฐมภูมิ) และ กราฟเส้นตรงแสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (แกนทุติยภูมิ) ซ้าย: รูปแบบที่หนึ่ง ขวา: รูปแบบที่สอง บน: ออกเสียงล่าง: ไม่ออกเสียง.....	133
3.34	แผนภูมิแท่งแสดงผลการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ระหว่างวิธีการพูดแบบออกเสียง และไม่ออกเสียง ก) รูปแบบที่หนึ่ง ข) รูปแบบที่สอง.....	134
3.35	แผนภูมิแท่งแสดงผลการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ คน เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ระหว่างการใช้ลักษณะเด่นจากการประเมินรูปแบบที่หนึ่ง (pattern 1) และรูปแบบที่สอง (pattern 2) ก) ออกเสียง ข) ไม่ออกเสียง .....	135
3.36	ผลการประเมินความสามารถในการตัดแยกกลุ่มวรรณยุกต์ 5 เสียงของลักษณะเด่น 64 ตัว โดยการใช้ตัวประเมิน RES.....	143
3.37	แผนภาพการกระจายของลักษณะเด่น (ซ้าย) RES สูงสุด (ขวา) RES ต่ำสุด.....	143
3.38	แผนภาพการกระจายแสดงการแยกแยะกลุ่มของวรรณยุกต์ 5 เสียง (ก) พิจารณาวรรณยุกต์ทั้ง 5 เสียง (ข) พิจารณาเฉพาะวรรณยุกต์สามัญ โท และ ตี (ค) พิจารณาเฉพาะวรรณยุกต์สามัญและโท.....	144
3.39	แผนภาพเทคนิคการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย.....	145
3.40	แผนภาพแสดงการแบ่งกลุ่มวรรณยุกต์ในแต่ละขั้นตอนของเทคนิคการจำแนกแบบสามขั้นตอน.....	148
3.41	ผังงานแสดงการทำงานของโปรแกรมการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย.....	149



## รายการรูปภาพ (ต่อ)

รูปที่	หน้า	
3.42	ผลความแม่นยำในการจำแนกในแต่ละโครงข่ายสำหรับเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย (ก) โครงข่ายที่ 1 (ข) โครงข่ายที่ 2 (ค) โครงข่ายที่ 3.....	152
3.43	ลักษณะเด่นของสัญญาณไฟฟ้ากล้ำมเนื้อจากการคำนวณทั้งสองรูปแบบ ทั้งหมด 124 ตัว ซึ่งถูกใช้ในการทดลองในหัวข้อ 3.3.2.3.....	156
3.44	กราฟแสดงผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง ซึ่งเปรียบเทียบความแม่นยำของข้อมูลสัญญาณไฟฟ้ากล้ำมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน.....	158
3.45	กราฟแกนปฐมภูมิ (ซ้าย) แสดงผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว กราฟแกนทุติยภูมิ (ขวา) แสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน.....	161
3.46	แผนภูมิแท่งแสดงการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกระหว่างรูปแบบการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอน.....	161
4.1	ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ (ก) การพูดออกเสียง (ข) การพูดไม่ออกเสียง.....	168
4.2	ผลความแม่นยำเฉลี่ยของลักษณะเด่น TDO โดยเปรียบเทียบความแม่นยำที่ได้ระหว่างการไม่ใช้ LDA ในการลดมิติของข้อมูล และการใช้ LDA ในการลดมิติของข้อมูล.....	169
4.3	ผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง.....	169
4.4	เปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ ระหว่างการใช้ตัวจำแนก 8 ชนิด (ก) การพูดแบบออกเสียง (ข) การพูดแบบไม่ออกเสียง.....	177

## รายการรูปภาพ (ต่อ)

รูปที่	หน้า	
4.5	เปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง ระหว่างการใช้ตัว จำแนก 8 ชนิด (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสาม ขั้นตอน.....	178
4.6	เวลาในการประมวลผล (elapsed time) สำหรับการจำแนกคำพูดหลัก 11 คำ ของตัวจำแนกแต่ละตัว.....	179
4.7	เวลาในการประมวลผล (elapsed time) สำหรับการใช้ตัวประเมินแต่ละชนิด.....	185
4.8	เปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด (C1-C14) เมื่อ ทดสอบด้วยตัวจำแนก 4 ชนิด (ก) FLDA (ข) LOGL (ค) LBN (ง) ANN (สำหรับ การพูดแบบออกเสียง).....	187
4.9	เปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด (C1-C14) เมื่อ ทดสอบด้วยตัวจำแนก 4 ชนิด (ก) FLDA (ข) LOGL (ค) LBN (ง) ANN (สำหรับ การพูดแบบไม่ออกเสียง).....	188
4.10	ผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณ S08 สำหรับ การพูดแบบออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหาลักษณะเด่น 3 วิธี โดยการ ใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN.....	194
4.11	ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้า กล้ามเนื้อทั้งหมด สำหรับการพูดแบบออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหา ลักษณะเด่น 3 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN.....	196
4.12	ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้า กล้ามเนื้อทั้งหมด สำหรับการพูดแบบไม่ออกเสียง ซึ่งเปรียบเทียบเทคนิคการ ค้นหาลักษณะเด่น 4 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN.....	197
4.13	ผลการประเมินและเลือกลักษณะเด่นโดยใช้เทคนิคการค้นหาแบบ FW โดยการใช้ตัว ประเมิน W_LBN สำหรับสัญญาณ S08.....	202

## รายการรูปภาพ (ต่อ)

รูปที่	หน้า	
4.14	ผลความแม่นยำเฉลี่ยในการจำแนกของสัญญาณไฟฟ้ากล้ำมเนื้อจากอาสาสมัคร ทั้ง 10 คน เมื่อเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว.....	202
4.15	ผลการประเมินและเลือกลักษณะเด่น (124 ตัว) โดยใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W_LBN และ W_LOGL สำหรับสัญญาณ S04.....	208
4.16	ผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้า กล้ำมเนื้อจากอาสาสมัครทั้ง 10 คน เมื่อเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว สำหรับการใช้ตัวประเมินและตัวจำแนก 2 ชนิด.....	209
4.17	ผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล้ำมเนื้อ ของอาสาสมัครแต่ละคน โดยพิจารณาจากการใช้จำนวนลักษณะเด่น 20 ตัว ซึ่ง คัดเลือกจากวิธีการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ W_LBN เป็นตัวประเมิน และจำแนกโดยใช้ LBN.....	209
5.1	รูปแบบการทำงานของแบบจำลองระบบรู้จำการพูดภาษาไทยโดยใช้ สัญญาณไฟฟ้ากล้ำมเนื้อ.....	213
5.2	แผนผังแสดงการทำงานของแบบจำลองระบบรู้จำการพูดภาษาไทยโดยใช้ สัญญาณไฟฟ้ากล้ำมเนื้อ.....	215
5.3	แสดงตัวอย่างผลการทำงานของโปรแกรมจำลองการทำงานของระบบรู้จำการพูด ภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อ.....	217
6.1	ตัวอย่างโครงสร้างรวมของระบบรู้จำเมื่อใช้ระบบจำแนกวรรณยุกต์เข้ามามีส่วน ในการจำแนกร่วมกับการจำแนกคำพูดหลัก.....	223

## รายการตาราง

ตารางที่	หน้า	
1.1	ลักษณะเด่นในโดเมนเวลาทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม.....	19
1.2	ลักษณะเด่นในโดเมนความถี่ทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม.....	20
1.3	ลักษณะเด่นในโดเมนเวลา-ความถี่ทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม.....	21
1.4	บทความที่มีการทดลองเพื่อเปรียบเทียบประสิทธิภาพของลักษณะเด่น.....	22
2.1	รายชื่อและหน้าที่ของกล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูดทั้ง 7 กลุ่ม.....	35
2.2	หน่วยเสียงพยัญชนะภาษาไทยซึ่งแทนด้วยสัญลักษณ์ของแต่ละหน่วยเสียงตาม IPA.....	40
2.3	หน่วยเสียงสระภาษาไทย ซึ่งแทนด้วยสัญลักษณ์ของแต่ละหน่วยเสียงตาม IPA.....	41
3.1	การจัดกลุ่มของมัดกล้ามเนื้อที่มีหน้าที่คล้ายกัน.....	81
3.2	ข้อมูลอาสาสมัคร.....	89
3.3	คำพูดทั้งหมดที่ถูกใช้สำหรับการเก็บข้อมูลของสัญญาณ.....	90
3.4	สรุปข้อมูลจำนวนอาสาสมัคร วิธีการพูด จำนวนการพูด และข้อมูลทั้งหมดสำหรับการเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง.....	91
3.5	ลักษณะเด่นที่ถูกใช้ในการทดลอง.....	105
3.6	เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ (รูปแบบที่หนึ่ง).....	114
3.7	เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ 3 ช่องสัญญาณแรก (รูปแบบที่สอง).....	114

## รายการตาราง (ต่อ)

ตารางที่	หน้า	
3.8	เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า กล้ามเนื้อ 2 ช่องสัญญาณสุดท้าย (รูปแบบที่สาม).....	115
3.9	เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า กล้ามเนื้อสำหรับแต่ละช่องสัญญาณ (รูปแบบที่สี่).....	116
3.10	ผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวน ลักษณะเด่นที่แตกต่างกัน และใช้จำนวนข้อมูลทั้ง 5 ช่องสัญญาณ.....	119
3.11	ผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวน ลักษณะเด่นที่แตกต่างกัน โดยพิจารณาลักษณะเด่นที่ดีที่สุดในแต่ละ ช่องสัญญาณ.....	120
3.12	ลักษณะเด่นที่ใช้ในการทดลอง.....	124
3.13	ลักษณะเด่นที่มีผลการประเมินดีที่สุด 5 อันดับแรก(การประเมินรูปแบบที่หนึ่ง) สำหรับสัญญาณไฟฟ้ามื้อของอาสาสมัครแต่ละคน จากสัญญาณ ช่องสัญญาณที่ 1 สำหรับการพูดแบบออกเสียง.....	126
3.14	การเรียงลำดับของลักษณะเด่นจากตารางที่ 3.13 โดยพิจารณาจากความถี่และ ค่าเฉลี่ย RES.....	127
3.15	ผลการประเมินและเลือกลักษณะเด่น จากการประเมินรูปแบบที่สอง.....	127
3.16	ผลการประเมินและเลือกลักษณะเด่น จากการประเมินรูปแบบที่หนึ่ง.....	128
3.17	ลักษณะเด่นรูปแบบที่หนึ่ง.....	138
3.18	ลักษณะเด่นรูปแบบที่สอง.....	139
3.19	ผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง ของลักษณะเด่นที่ดีที่สุด 5 อันดับแรกจากผลการประเมินด้วย RES.....	141
3.20	ลักษณะเด่นที่มีผลการประเมินด้วย RES ดีที่สุด 8 อันดับแรกในแต่ละ กลุ่ม.....	151
3.21	ผลความแม่นยำในการจำแนกวรรณยุกต์โดยใช้ตัวจำแนก ANN.....	151

## รายการตาราง (ต่อ)

ตารางที่	หน้า
3.22 ลักษณะเด่นที่ถูกเลือก สำหรับใช้ในรูปแบบการจำแนกแบบสามขั้นตอน.....	162
3.23 สรุปผลการออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด) และ การ ออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง).....	164
4.1 จำนวนมิติข้อมูลของเวกเตอร์ลักษณะเด่นในแต่ละเฟรม.....	167
4.2 เปรียบเทียบความแม่นยำเฉลี่ยระหว่างการใช้เทคนิคเดิมและเทคนิคใหม่.....	171
4.3 ตัวจำแนกทั้งหมดที่ใช้ในการทดลองนี้.....	173
4.4 เปรียบเทียบผลความแม่นยำเฉลี่ยระหว่างเทคนิคการแบ่งข้อมูลแบบสุ่มด้วยการ แบ่งร้อยละและเทคนิคการตรวจสอบไขว้.....	174
4.5 สรุปตัวจำแนกที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรก.....	176
4.6 ลักษณะเด่น 22 ชนิด.....	182
4.7 ตัวประเมินลักษณะเด่น 14 ชนิด.....	183
4.8 ตัวประเมินที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรกสำหรับตัวจำแนกแต่ละ ชนิด.....	186
4.9 เทคนิคการค้นหาลักษณะเด่น 3 วิธี.....	191
4.10 เปรียบเทียบเทคนิคการเลือกลักษณะเด่นแบบ FR และ FSS.....	195
4.11 เปรียบเทียบประสิทธิภาพของตัวจำแนก FLDA และ LBN.....	196
4.12 ลักษณะเด่นที่ถูกเลือกในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่นด้วยวิธี FW และผลการประเมินเซตย่อยของลักษณะเด่นในแต่ละรอบด้วยวิธีการใช้ตัว ประเมิน W_LBN.....	201
4.13 เวลาในการฝึกสอนและทดสอบระบบสำหรับการใช้ตัวจำแนกสองชนิด.....	206
4.14 ลักษณะเด่นที่ถูกเลือกในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่นด้วยวิธี FW และผลการประเมินเซตย่อยของลักษณะเด่นในแต่ละรอบด้วยวิธีการใช้ตัว ประเมิน W_LBN.....	207
4.15 สรุปผลการปรับปรุงระบบ (ความแม่นยำเฉลี่ยของสัญญาณไฟฟ้ากล่อมเนื้อทั้ง 10 คน).....	211
4.16 สรุปผลการปรับปรุงระบบ (ความแม่นยำสูงสุด).....	211

## สารบัญคำศัพท์

แบบจำลองเสียง	acoustic model
สัญญาณรบกวนล้อมรอบ	ambient noise
อิเล็กโทรดแบบจัดเรียง	array electrode
แบบจำลองคุณลักษณะการออกเสียงพูด	articulatory features/phonetic features
โครงข่ายประสาทเทียม	artificial neural network
ระบบรู้จำเสียงพูดอัตโนมัติ	automatic speech recognition
อัตโนมัติถดถอย	autoregressive
การค้นหาแบบไปข้างหลัง	backward search
ตัวกรองแถบความถี่ผ่าน	band pass filter
อิเล็กโทรดแบบแท่ง	bar electrode
กฎของเบย์	Bayes' theorem
การค้นหาแบบบีบ	beam search
เมตริกซ์โควาเรียนซ์ระหว่างคลาส	between-class covariance matrix
อิเล็กโทรดแบบสองขั้ว	bipolar electrode
การค้นหาแบบบรานซ์แอนด์บาวด์	branch and bound search
การจัดกลุ่มของคุณลักษณะการออกเสียงพูด	bundled phonetic features
บัตเตอร์เวิร์ท	Butterworth
อิเล็กโทรดแบบกระดุม	button electrode
สัมประสิทธิ์เซปสตรัล	cepstral coefficient
ความหนาแน่นอย่างมีเงื่อนไขของคลาส	class-conditional density
การจำแนก/การรู้จำ	classification/recognition
ตัวจำแนก	classifier
การค้นหาแบบสมบูรณ์	complete search
ตัวกรองแบบบริบท	contextual filter
คำพูดต่อเนื่อง	continuous speech
สหสัมพันธ์	correlation
เมตริกซ์โควาเรียนซ์	covariance matrix

### สารบัญคำศัพท์ (ต่อ)

สัญญาณรบกวนจากสัญญาณไฟฟ้ากล้ามเนื้อ	cross talk noise
จากกล้ามเนื้อที่อยู่ติดกัน	
การเก็บข้อมูลของสัญญาณ	data collection
กฎการตัดสินใจ	decision rule
ต้นไม้ตัดสินใจ	decision tree
สัมประสิทธิ์อินพุตอันดับหนึ่งของการแปลงแบบ	delta coefficient of short time Fourier
ฟูรีเยร์ช่วงสั้น	transform
ตัวกรองเดลต้า	delta filter
ฟังก์ชันการจำแนก	discriminant function
อิเล็กโทรดแบบใช้ครั้งเดียวทิ้ง	disposable electrode
ตัวแปลงลงแบบดิจิทัล	down digital converter
อิเล็กโทรดแบบแห้ง	dry electrode
สัญญาณคลื่นไฟฟ้าหัวใจ	Electrocardiography
อิเล็กโทรด	electrode
สัญญาณไฟฟ้ากล้ามเนื้อ	electromyography
เทคนิคฝังตัว/เทคนิคไฮบริด	embedded method/hybrid method
ระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ	EMG-based speech recognition
เอนโทรปี	entropy
ระยะทางแบบยูคลิด	Euclidean distance
เกณฑ์การประเมิน	evaluation criterion
อีวิดเอนซ์เฟกเตอร์	evidence factor
การค้นหาแบบเอกซ์โพเนนเชียล	exponential search
ลักษณะเด่น	feature
การสกัดลักษณะเด่นของสัญญาณ	feature extraction
การเรียงลำดับลักษณะเด่น	feature ranking
การคัดเลือกลักษณะเด่นของสัญญาณ	feature selection
การเลือกเซตย่อยของลักษณะเด่น	feature subset selection



### สารบัญคำศัพท์ (ต่อ)

เทคนิคฟิลเตอร์	filter method
ตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้นของพีชเซอร์	Fisher's linear discriminant analysis classifier
การค้นหาแบบไปข้างหน้า	forward search
ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน	frame-based time domain feature with stacking filter
ลักษณะเด่นในโดเมนความถี่	frequency domain feature
ความถี่มูลฐาน	fundamental frequency
อิเล็กโทรดแบบตาราง	grid electrode
แบบจำลองฮิดเดนมาร์คอฟ	Hidden Markov model
ตัวกรองความถี่สูงผ่าน	high pass filter
เกณฑ์ค่าเกินความรู้	information gain criterion
ค่าผลรวมของค่าสัมบูรณ์	integrated absolute value
เกณฑ์ระยะทางภายในกลุ่มและระหว่างกลุ่ม	inter-intra distance criterion
ข้อมูล	
สัทอักษรสากล	international phonetic alphabet
คำโดด	isolated word
เกณฑ์ค่าสหสัมพันธ์แบบเคนดอล	Kendall rank correlation criterion
การตรวจสอบแบบไขว้ K ครั้ง	K-fold cross validation
ตัวจำแนกเพื่อนบ้านใกล้ที่สุด K ตัว	K-nearest neighbor classifier
แบบจำลองภาษา	language model
ไลค์ลิฮูด	likelihood
ตัวจำแนกเชิงเส้นแบบเบย์นอร์มัล	linear Bayes normal classifier
การวิเคราะห์การจำแนกประเภทแบบเชิงเส้น	linear discriminant analysis
การทำนายการเข้ารหัสแบบเชิงเส้น	linear predictive coding,
ตัวจำแนกเชิงเส้นแบบลอจิสติก	logistic linear classifier
ตัวกรองความถี่ต่ำผ่าน	low pass filter

## สารบัญคำศัพท์ (ต่อ)

แมนนวล	manual
ค่าสูงสุดของความน่าจะเป็นหลัง	Maximum A Posterior
ค่าสูงสุดของไลค์ลิฮูด	maximum likelihood
ค่าเฉลี่ยของสัญญาณ	mean
ค่าเฉลี่ยของค่าสัมบูรณ์	mean absolute value
ค่าความชันของค่าเฉลี่ยค่าสัมบูรณ์	mean absolute value slope
ค่าเฉลี่ยของความถี่	mean frequency
ค่าเฉลี่ยขององค์ประกอบความถี่สูงของสัญญาณ	mean of high frequency part
ค่าเฉลี่ยขององค์ประกอบความถี่ต่ำของสัญญาณ	mean of low frequency part
ค่าความถี่กลาง	median frequency
สัมประสิทธิ์เซปสตรัมความถี่เมล	Mel-frequency cepstral coefficient
เกณฑ์ค่าต่ำสุดของระยะทางมาฮาลานอบิสแบบประมาณ	minimum of estimated Mahalanobis distance criterion
เกณฑ์ค่าต่ำสุดของระยะทางแบบยูคลิดกำลังสอง	minimum of squared Euclidean distance criterion
ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 1	modified mean absolute value type 1
ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 2	modified mean absolute value type 2
อิเล็กโทรดแบบขั้วเดียว	monopolar electrode
สัญญาณรบกวนการเคลื่อนที่	motion artifact noise
ตัวกรองค่าเฉลี่ยเคลื่อนที่	moving average filter
ตัวจำแนกค่าเฉลี่ยที่ใกล้ที่สุด	nearest mean classifier
การกระจายแบบปกติ	normal distribution
นอร์มัลไลซ์	normalize
ตัวกรองนอตช์	Notch filter
ความถี่ไนควิสต์	Nyquist frequency
การเพิ่มอัตราสุ่ม	over-sampling
เกณฑ์ค่าสหสัมพันธ์แบบเชิงเส้นของเพียร์สัน	Pearson's linear correlation criterion

### สารบัญคำศัพท์ (ต่อ)

การแบ่งข้อมูลแบบร้อยละ	percentage split /hand out method
การค้นหาแบบบวกแอลลบอาร์	plus L take away R search
ความน่าจะเป็นหลัง	posterior probability
ค่าเฉลี่ยขององค์ประกอบความถี่สูงของสัญญาณ	power of high frequency part
ค่ากำลังขององค์ประกอบความถี่ต่ำของสัญญาณ	power of low frequency part
การเตรียมสัญญาณเบื้องต้น	pre-processing
การวิเคราะห์องค์ประกอบหลัก	principal component analysis
ความน่าจะเป็นก่อนหน้า	prior probability
ตัวจำแนกควอดราติกเบย์นอร์มัล	quadratic Bayes normal classifier
ฟังก์ชันแบบควอดราติก	quadratic function
แรนดอมฟอร์เรส	random forest
การแปลงเวฟเลตไม่ต่อเนื่องแบบรีดันแดนด์	redundant discrete wavelet transform
อิเล็กโทรดชนิดนำกลับมาใช้ใหม่ได้	reusable electrode
ค่ารากเฉลี่ยกำลังสอง	root mean square
การชักตัวอย่าง	sampling
อัตราการชักตัวอย่าง	sampling rate
กราฟการกระจาย	scatter plot
การแบ่งส่วนของสัญญาณ	segmentation
ลำดับข้อมูล	sequence
การค้นหาตามลำดับ	sequential search
การแปลงแบบฟูรีเยร์ช่วงสั้น	short time Fourier transform
ค่าการเปลี่ยนแปลงความชัน	slope sign change
เกณฑ์ค่าสหสัมพันธ์แบบสเปียร์แมน	Spearman rank correlation criterion
ค่าสเปคตรัลโมเมนต์	spectral moment
ตัวกรองซ้อน	stacking filter
ความกว้างการซ้อน	stacking width

### สารบัญคำศัพท์ (ต่อ)

ค่าเบี่ยงเบนมาตรฐาน	standard deviation
เกณฑ์ผลรวมของระยะทางมาฮาลานอบิส ประมาณ	sum of estimated Mahalanobis distance criterion
เกณฑ์ผลรวมของระยะทางแบบยูคลิดกำลังสอง	sum of squared Euclidean distance criterion
ฟังก์ชันการรวม	summation function
ซัพพอร์ตเวกเตอร์แมชชีน	Support Vector Machine
อิเล็กโทรดแบบพื้นผิว	surface electrode
การทดสอบระบบ	testing
สถิติอันดับสาม	third order statistics
ค่าขีดแบ่งของสามเท่าของค่าเบี่ยงเบนมาตรฐาน	Three standard deviation threshold
ค่าขีดแบ่ง	threshold
ลักษณะเด่นในโดเมนเวลา	time domain feature
ลักษณะเด่นในโดเมนเวลาและความถี่	time-frequency domain feature
แบบจำลองเสียงวรรณยุกต์	tone model
การฝึกสอนระบบ	training
ฟังก์ชันการแปลง	transfer function
ตัวกรองเทรนด์	trend filter
ค่าความแปรปรวน	variance
ค่าความยาวรูปคลื่น	waveform length
การแปลงเวฟเลตพ็อกเก็ต	wavelet pocket transform
การแปลงเวฟเลต	wavelet transform
อิเล็กโทรดแบบเปียก	wet electrode
ค่าแอมพลิจูดของวิลลิสัน	Willison amplitude
เมตริกซ์การกระจายภายในคลาส	within-class scatter matrix
แบบจำลองคำพูดหลัก	word model
เทคนิคแรปปเปอร์	wrapper method

**สารบัญคำศัพท์ (ต่อ)**

ค่าจุดตัดศูนย์

zero crossing

ค่าจุดตัดผ่านศูนย์ขององค์ประกอบความถี่สูง  
ของสัญญาณ

zero crossing of high frequency part

## สารบัญคำย่อ

AF/PF	แบบจำลองคุณลักษณะการออกเสียงพูด
ANN	โครงข่ายประสาทเทียม
AR	อัตโนมัติถดถอย
ASR	ระบบรู้จำเสียงพูดอัตโนมัติ
BDPFs	การจัดกลุ่มของลักษณะการออกเสียงพูด
BW	การค้นหาแบบไปข้างหลัง
delta STFT	สัมประสิทธิ์อนุพันธ์อันดับหนึ่งของการแปลงแบบฟูริเยร์ช่วงสั้น
DT	ต้นไม้ตัดสินใจ
ECG	สัญญาณคลื่นไฟฟ้าหัวใจ
EM	เทคนิคฝังตัว/เทคนิคไฮบริด
EMG	สัญญาณไฟฟ้ากล้ามเนื้อ
F_CorrK	เกณฑ์ค่าสหสัมพันธ์แบบเคนดอลล์
F_CorrP	เกณฑ์ค่าสหสัมพันธ์แบบเชิงเส้นของเพียร์สัน
F_CorrS	เกณฑ์ค่าสหสัมพันธ์แบบสเปียร์แมน
F_Eucl-m	เกณฑ์ค่าต่ำสุดของระยะทางแบบยูคลิดกำลังสอง
F_Eucl-s	เกณฑ์ผลรวมของระยะทางแบบยูคลิดกำลังสอง
F_IG	เกณฑ์ค่าเกณฑ์ความถี่
F_in-in	เกณฑ์ระยะทางภายในกลุ่มและระหว่างกลุ่มข้อมูล
F_Maha-m	เกณฑ์ค่าต่ำสุดของระยะทางมาฮาลานอบิสแบบประมาณ
F_Maha-s	เกณฑ์ผลรวมของระยะทางมาฮาลานอบิสแบบประมาณ
FLDA	ตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้นของฟิชเชอร์
FM	เทคนิคฟิลเตอร์
FR	การเรียงลำดับลักษณะเด่น
FSS	การเลือกเซตย่อยของลักษณะเด่น
FW	การค้นหาแบบไปข้างหน้า
HMM	แบบจำลองฮิตเดนมาร์คอฟ
IAV	ค่าผลรวมของค่าสัมบูรณ์

### สารบัญคำย่อ (ต่อ)

IPA	สัทอักษรสากล
KNN	ตัวจำแนกเพื่อนบ้านใกล้ที่สุด K ตัว
LBN	ตัวจำแนกเชิงเส้นแบบเบย์นอร์มัล
LDA	การวิเคราะห์การจำแนกประเภทแบบเชิงเส้น
LOGL	ตัวจำแนกเชิงเส้นแบบลอจิสติก
LPC	การทำนายการเข้ารหัสแบบเชิงเส้น
LR	การค้นหาแบบบวกแอลบอาร์
MATLAB	โปรแกรมเมตแลบ
MAV	ค่าเฉลี่ยของค่าสัมบูรณ์
MAVS	ค่าความชันของค่าเฉลี่ยค่าสัมบูรณ์
MDF	ค่าความถี่กลาง
MFCC	สัมประสิทธิ์เซปสตรัมความถี่เมล
ML	ค่าสูงสุดของไลค์ลิตูด
MMAV1	ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 1
MMAV2	ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 2
MN	ค่าเฉลี่ยของสัญญาณ
MNF	ค่าเฉลี่ยของความถี่
MNH	ค่าเฉลี่ยขององค์ประกอบความถี่สูงของสัญญาณ
MNL	ค่าเฉลี่ยขององค์ประกอบความถี่ต่ำของสัญญาณ
NM	ตัวจำแนกค่าเฉลี่ยที่ใกล้ที่สุด
PCA	การวิเคราะห์องค์ประกอบหลัก
PWH	ค่าเฉลี่ยขององค์ประกอบความถี่สูงของสัญญาณ
PWL	ค่ากำลังขององค์ประกอบความถี่ต่ำของสัญญาณ
QBN	ตัวจำแนกควอดราติกเบย์นอร์มัล
RDWT	การแปลงเวฟเลตไม่ต่อเนื่องแบบรีดันแดนต์
RES/F_RES	เกณฑ์อัตราส่วนระหว่างระยะทางแบบยูคลิดและค่าเบี่ยงเบนมาตรฐาน (ดัชนีอาร์อีเอส)

### สารบัญคำย่อ (ต่อ)

RF	แรนคอมฟอร์เรส
RMS	ค่ารากเฉลี่ยกำลังสอง
RMS_1	ลักษณะเด่นชนิด RMS ของสัญญาณส่วนที่หนึ่ง
SM	ค่าสเปคตรัลโมเมนต์
SM1	ค่าสเปคตรัลโมเมนต์ลำดับที่ 1
SSC	ค่าการเปลี่ยนแปลงความชัน
STFT	การแปลงฟูริเยร์ช่วงสั้น
SVM	ซัพพอร์ตเวกเตอร์แมชชีน
TDn	ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน
TOS	สถิติอันดับสาม
VAR	ค่าความแปรปรวน
W_ANN	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกโครงข่ายประสาทเทียม
W_FLDA	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้นของฟิชเชอร์
W_LBN	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกเชิงเส้นแบบเบย์นอร์มัล
W_LOGL	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกเชิงเส้นแบบลอจิสติก
WAMP	ค่าแอมพลิจูดของวิลลิสัน
WL	ค่าความยาวรูปคลื่น
WM	เทคนิคแรปเปอร์
WPT	การแปลงเวฟเลตพอกเก็ต
WT	การแปลงเวฟเลต
ZC	ค่าจุดตัดศูนย์
ZCH	ค่าจุดตัดผ่านศูนย์ขององค์ประกอบความถี่สูงของสัญญาณ



## บทที่ 1

### บทนำ

#### 1.1 ความเป็นมาของงานวิจัย

ระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ (EMG based speech recognition) ถูกพัฒนาจากระบบรู้จำการพูดแบบเดิม (conventional automatic speech recognition) ซึ่งใช้สัญญาณเสียงพูดเป็นข้อมูลเข้าหรืออินพุตให้กับระบบ โดยระบบใหม่ที่ถูกพัฒนาขึ้นนี้จะใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่เกี่ยวข้องกับการออกเสียงพูดมาเป็นอินพุตให้กับระบบแทน หรืออาจจะใช้ทั้งสองสัญญาณร่วมกัน ซึ่งจากการทบทวนวรรณกรรมพบว่าสัญญาณไฟฟ้ากล้ามเนื้อดังกล่าวสามารถใช้เป็นตัวแทนของสัญญาณเสียงพูดได้

สาเหตุที่ระบบรู้จำการพูดแบบใหม่นี้ถูกพัฒนาขึ้นมา มีสาเหตุหลักอยู่ 4 ประการ ซึ่งในส่วนของเหตุผล 3 ประการแรกจะเป็นปัญหาที่เกิดขึ้นจากการใช้งานระบบรู้จำการพูดแบบเดิม ได้แก่ ปัญหาการใช้งานระบบขณะที่มีเสียงรบกวนอย่างมาก ซึ่งทำให้สัญญาณเสียงที่ป้อนเข้าสู่ระบบถูกรบกวนด้วยสัญญาณรบกวนดังกล่าว ดังนั้นอินพุตของระบบจะผิดเพี้ยนและทำให้ระบบจำแนกข้อมูลผิดพลาด ปัญหาที่สอง คือ เมื่อต้องการสื่อสารด้วยข้อมูลที่เป็นส่วนตัวหรือเป็นความลับ เช่น การประยุกต์ใช้งานทางด้านทหาร ธนาคาร เป็นต้น การใช้งานด้วยระบบรู้จำการพูดแบบเดิมจะทำให้ข้อมูลอาจถูกเปิดเผย ซึ่งจะส่งผลกระทบต่อความปลอดภัยของข้อมูลและผู้ใช้งานได้ สำหรับปัญหาสุดท้ายของการใช้งานระบบรู้จำการพูดแบบเดิม คือ ระบบจะไม่สามารถใช้งานได้เมื่อถูกใช้งานในสถานการณ์ที่การสื่อสารด้วยเสียงพูดเป็นไปได้ยาก เช่น ขณะที่ผู้พูดสวมชุดซึ่งมีหน้ากากปิดใบหน้า ได้แก่ ชุดดับเพลิง ชุดอวกาศ ชุดป้องกันสารเคมี เป็นต้น นอกจากนี้ระบบรู้จำการพูดแบบใหม่ถูกพัฒนาขึ้นเพื่อใช้เป็นเครื่องมือช่วยในการสื่อสารสำหรับผู้ป่วยหรือผู้สูงอายุที่มีปัญหาเกี่ยวกับการออกเสียงพูด เช่น การได้รับการผ่าตัดมะเร็งกล่องเสียง การได้รับความกระทบกระเทือนทางสมองจนระบบประสาทที่เกี่ยวข้องกับกลไกการพูดมีปัญหา เป็นต้น ซึ่งผู้ป่วยเหล่านี้ไม่สามารถพูดได้อย่างปกติเหมือนคนทั่วไป ถึงแม้ว่าจะมีอุปกรณ์หรือเทคนิคที่จะช่วยให้ผู้ป่วยสามารถพูดได้ เช่น วิธีการพูดโดยใช้อากาศจากกระเพาะอาหาร (esophageal speech) การใช้เครื่องช่วยพูดอิเล็กทรอนิกส์ (electrolarynx) เป็นต้น แต่อุปกรณ์และเทคนิคเหล่านี้ยังไม่มีประสิทธิภาพเพียงพอ

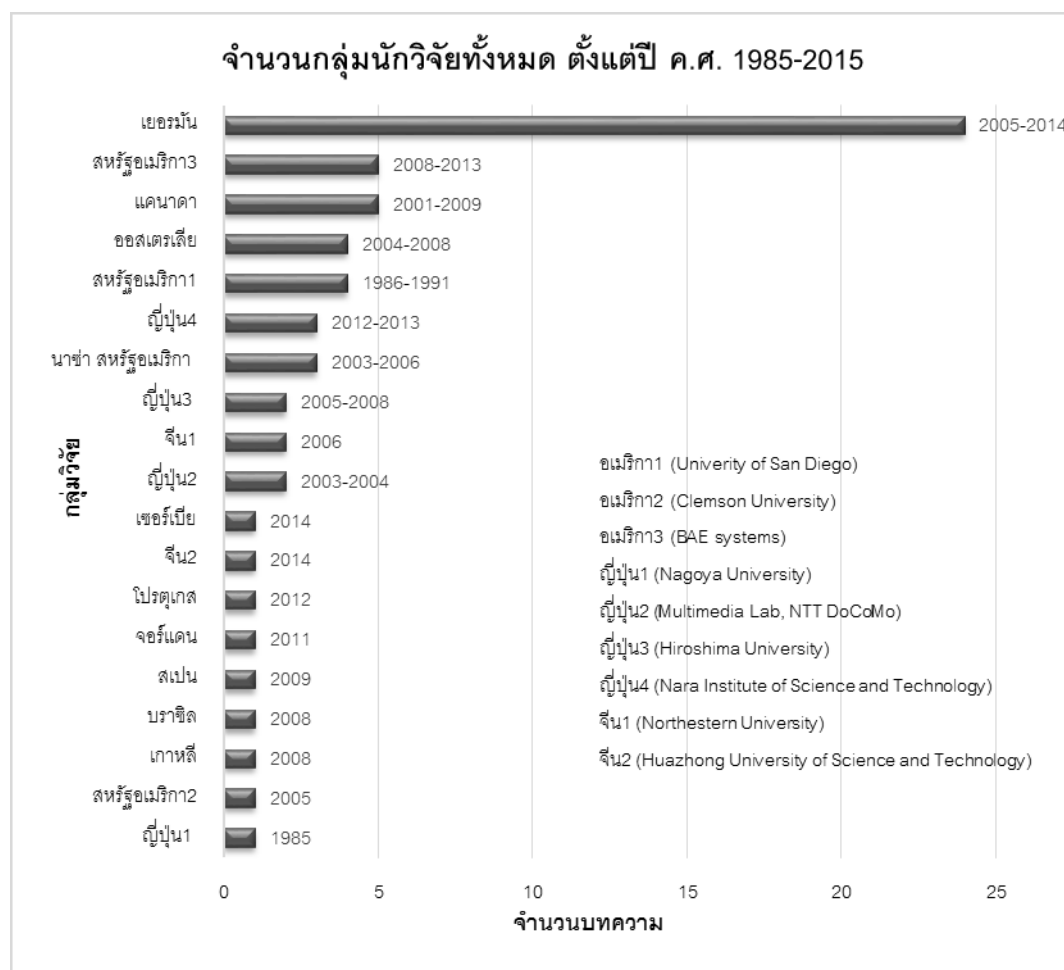
ระบบรู้จำการพูดแบบใหม่ที่ถูกพัฒนาขึ้นโดยหลากหลายกลุ่มวิจัยเพื่อแก้ไขปัญหาดังกล่าว อย่างไรก็ตามงานวิจัยเหล่านี้มีข้อจำกัดในด้านภาษาของผู้ใช้งาน กล่าวคือ ระบบถูกพัฒนาขึ้นสำหรับผู้ใช้งานสำหรับแต่ละภาษาโดยเฉพาะ ซึ่งความแตกต่างของภาษาที่หลากหลายมีผลทำให้ระบบรู้จำการพูดแบบใหม่ที่ถูกพัฒนาขึ้นในหลากหลายภาษา เช่นเดียวกับงานวิจัยนี้ ซึ่งจะพัฒนาระบบดังกล่าวให้สามารถใช้ได้กับผู้ที่ใช้ภาษาไทยในการสื่อสาร โดยเฉพาะอย่างยิ่งเพื่อสามารถนำไปใช้งานกับผู้ป่วยหรือผู้สูงอายุที่มีปัญหาตามที่กล่าวไว้แล้วข้างต้น ซึ่งอาจจะไม่สามารถพูดภาษาอื่นได้นอกจากภาษาไทย

## 1.2 การตรวจเอกสาร

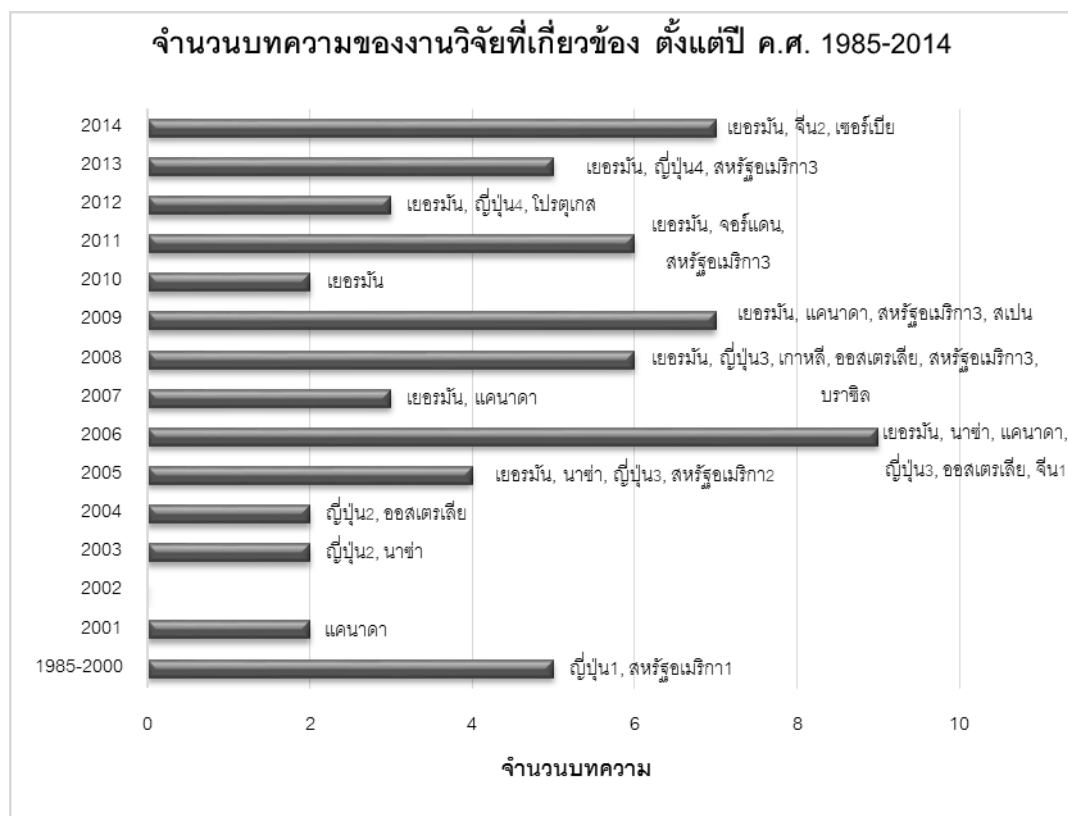
งานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อเริ่มขึ้นเมื่อปี ค.ศ. 1985 โดยกลุ่มนักวิจัยชาวญี่ปุ่น [1] ซึ่งพวกเขาได้สังเกตเห็นถึงปัญหาของการเพิ่มขึ้นเรื่อยๆ ของผู้ป่วยที่มีความบกพร่องของระบบการทำงานด้านการสร้างเสียงพูด เช่น ผู้ป่วยที่ต้องผ่าตัดมะเร็งกล่องเสียง ผู้ป่วยที่เป็นโรคอัมพาตเนื่องจากสมองถูกกระทบกระเทือน (cerebral palsy) ซึ่งทำให้ระบบประสาทส่วนการควบคุมการพูดมีปัญหา ถึงแม้ว่าผู้ป่วยเหล่านี้จะมีเทคนิคหรืออุปกรณ์ในการช่วยพูดเพื่อฟื้นฟูการพูดของผู้ป่วยให้สามารถสื่อสารกับผู้อื่นได้ เช่น เทคนิคการพูดโดยใช้ลมจากกระเพาะอาหาร การใช้อุปกรณ์ช่วยพูด เป็นต้น แต่วิธีการดังกล่าวยังมีข้อจำกัดในด้านคุณภาพของเสียงที่ใช้ในการสื่อสาร โดยเฉพาะเทคนิคการพูดโดยใช้ลมจากกระเพาะอาหาร ซึ่งจำเป็นต้องใช้เวลาในการฝึกฝนเป็นเวลานาน ทำให้ผู้ป่วยส่วนใหญ่เลือกใช้วิธีอื่นในการสื่อสารแทน เช่น การอ่านปาก การใช้ภาษามือหรือภาษาเขียน เป็นต้น ซึ่งวิธีการเหล่านี้ก็ยังคงยากต่อการสื่อสาร จากปัญหาดังกล่าวพวกเขาจึงได้พัฒนาอุปกรณ์การสร้างเสียงเทียมให้กับผู้ป่วยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ที่ได้จากการจับสัญญาณไฟฟ้ากล้ามเนื้อรอบปาก ซึ่งกล่าวได้ว่างานวิจัยนี้เป็นต้นแบบของแนวคิดการพัฒนากระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

ในเวลาต่อมางานวิจัยด้านนี้ได้รับการพัฒนาโดยหลากหลายกลุ่มนักวิจัย และหลากหลายภาษา [1]-[63] ซึ่งพบว่ามีกลุ่มนักวิจัย 19 กลุ่มจาก 12 ประเทศ (รูปที่ 1.1) โดยในช่วงเริ่มแรกของงานวิจัย (ค.ศ.1985-2000) จะมีเพียง 2 กลุ่ม คือ กลุ่มของญี่ปุ่น (Nagoya university) [1] และ สหรัฐอเมริกา (university of San Diego) [2]-[5] หลังจากนั้นก็มีกลุ่มนักวิจัยอื่นเพิ่มขึ้นมาอีกในปีต่อมา โดยปี ค.ศ.2006 และ ค.ศ.2008 มีจำนวนกลุ่มนักวิจัยมากที่สุด 6 กลุ่ม นอกจากนี้เมื่อพิจารณาจำนวนของบทความที่เกี่ยวข้องกับงานวิจัยนี้ตั้งแต่ปี ค.ศ. 1985-2014 ซึ่งแสดงดังรูป

ที่ 1.2 (ทั้งหมด 63 บทความ) แสดงให้เห็นว่างานวิจัยทางด้านนี้มีการพัฒนามาอย่างต่อเนื่อง โดยเฉพาะในปี ค.ศ. 2006 ปรากฏผลงานตีพิมพ์มากที่สุด 9 บทความจากกลุ่มนักวิจัย 6 กลุ่ม ซึ่งถือได้ว่าเป็นช่วงที่มีงานวิจัยทางด้านนี้เกิดขึ้นมากที่สุด โดยกลุ่มงานวิจัยที่เด่นที่สุด คือ กลุ่มเยอรมัน ซึ่งได้ทำวิจัยมาอย่างต่อเนื่องตั้งแต่ปี ค.ศ. 2005 ถึงปัจจุบัน และมีผลงานตีพิมพ์มากที่สุด 24 บทความ [6]-[29]



**รูปที่ 1.1** แผนภูมิแสดงกลุ่มของงานวิจัยซึ่งทำการวิจัยเกี่ยวกับระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ



**รูปที่ 1.2** แผนภูมิแสดงจำนวนบทความของงานวิจัยเกี่ยวกับระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

ในงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อของแต่ละกลุ่มนักวิจัย จะทำการปรับปรุงโครงสร้างของระบบในแต่ละส่วน ซึ่งประกอบไปด้วย กระบวนการเก็บข้อมูลของสัญญาณ (data collection) กระบวนการเตรียมสัญญาณเบื้องต้น (pre-processing) กระบวนการสกัดลักษณะเด่นของสัญญาณ (feature extraction) กระบวนการเลือกลักษณะเด่นของสัญญาณ (feature selection) และกระบวนการรู้จำระบบหรือการจำแนกระบบ (recognition or classification) ซึ่งจะทำให้วิเคราะห์โครงสร้างในแต่ละส่วนของระบบเพื่อเชื่อมโยงกับการทบทวนวรรณกรรม ดังนี้

### 1.2.1 กระบวนการเก็บข้อมูลของสัญญาณ

อินพุตของระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ คือ สัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากการตรวจจับสัญญาณดังกล่าวโดยใช้อิเล็กโทรด (electrode) จากตำแหน่งของ

กล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูดของมนุษย์ สำหรับบางงานวิจัย อินพุตของระบบจะใช้ทั้งสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงพูดเพื่อใช้ในการจำแนกระบบ [30]-[36] อย่างไรก็ตาม สำหรับบางงานวิจัยสัญญาณเสียงพูดจะถูกใช้เพื่อช่วยในการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อเท่านั้น [6]-[29]

### (ก) อิเล็กโทรด

สำหรับอิเล็กโทรดที่ใช้ในการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ ส่วนใหญ่จะใช้ อิเล็กโทรดแบบพื้นผิว (surface electrode) ชนิดซิลเวอร์ซิลเวอร์คลอไรด์ (Ag/AgCl) เนื่องจากมีความสะดวกในการใช้งาน มีประสิทธิภาพในการวัดสัญญาณ มีมวลน้อย ขนาดเล็ก และมีความนำเชื่อถือสูง ซึ่งอิเล็กโทรดชนิดนี้สามารถแบ่งออกเป็น อิเล็กโทรดแบบแห้ง (dry electrode) และอิเล็กโทรดแบบเปียก (wet electrode) ในส่วนของอิเล็กโทรดแบบแห้งจะอยู่ในรูปของอิเล็กโทรดแท่ง (bar electrode) สำหรับอิเล็กโทรดแบบเปียกจะอยู่ในรูปของอิเล็กโทรดแผ่นกลมที่มีลักษณะเป็นกระดุม (button electrode) ซึ่งมีทั้งแบบใช้ครั้งเดียวทิ้ง (disposable electrode) และชนิดที่นำกลับมาใช้ใหม่ได้ (reusable electrode) อย่างไรก็ตามงานวิจัยส่วนใหญ่นิยมใช้อิเล็กโทรดแบบเปียกมากกว่าอิเล็กโทรดแบบแห้ง เนื่องจากอิเล็กโทรดชนิดนี้มีเจลที่ช่วยลดความต้านทานที่เกิดขึ้นระหว่างผิวหนังกับอิเล็กโทรด มีน้ำหนักเบา และใช้งานง่าย สำหรับอิเล็กโทรดแบบแห้งซึ่งอยู่ในรูปของอิเล็กโทรดแท่งจะมีวงจรรีเลย์หรือทรานซิสเตอร์อยู่ภายใน ทำให้มีน้ำหนักมากใช้งานไม่สะดวก อย่างไรก็ตามอิเล็กโทรดชนิดนี้ก็นำมาใช้ในบางกลุ่มวิจัย เนื่องจากไม่ต้องมีการทาเจลและไม่มีกาวติดกับผิวหนังจึงเหมาะสมสำหรับการใช้งานบางประเภท เช่น งานวิจัย [37] ได้ใช้อิเล็กโทรดแบบแท่งติดอยู่กับอุปกรณ์แบบพกพา (รูปที่ 1.3(จ)) โดยขณะใช้งานอุปกรณ์นี้จะถูกสวมไปที่นิ้วโป้ง นิ้วชี้ และนิ้วกลางของผู้ใช้ เพื่อจับสัญญาณไฟฟ้ากล้ามเนื้อรอบปากขณะพูด ซึ่งอุปกรณ์นี้จะเชื่อมต่อกับโทรศัพท์มือถือ งานวิจัยนี้เป็นการประยุกต์ใช้งานเพื่อการติดต่อสื่อสารแบบเสียงผ่านโทรศัพท์มือถือในกรณีที่ผู้ใช้อยู่ในสถานที่ที่มีรบกวนอย่างมาก นอกจากนี้ยังมีการนำอิเล็กโทรดฝังเข้ากับหน้ากากออกซิเจนของนักบิน (รูปที่ 1.3(ง)) เพื่อจับสัญญาณไฟฟ้ากล้ามเนื้อของนักบินขณะพูดในสถานการณ์ที่มีเสียงรบกวนอย่างมาก [30] อิเล็กโทรดที่ใช้สำหรับงานวิจัยนี้ได้ถูกพัฒนาขึ้นมาเรื่อยๆ ซึ่งงานวิจัยล่าสุดมีการพัฒนาอิเล็กโทรดแบบจัดเรียง (array electrode) [23] และอิเล็กโทรดแบบตาราง (grid electrode) [38]-[40] เพื่อให้สามารถจับสัญญาณไฟฟ้ากล้ามเนื้อได้หลายช่องสัญญาณ ซึ่งสามารถเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อขนาดเล็กได้ อย่างไรก็ตามเมื่ออิเล็กโทรดแบบใหม่ถูกพัฒนาขึ้นมา จำเป็นจะต้องออกแบบเทคนิคต่างๆ เพื่อให้

เหมาะสมกับการใช้งานอิเล็กโทรดนั้นๆ ดังนั้นจะเห็นได้ว่าการเลือกใช้อิเล็กโทรดขึ้นอยู่กับหลายปัจจัย โดยเฉพาะอย่างยิ่งการเลือกใช้ให้เหมาะสมกับการนำไปใช้งานแต่ละประเภท



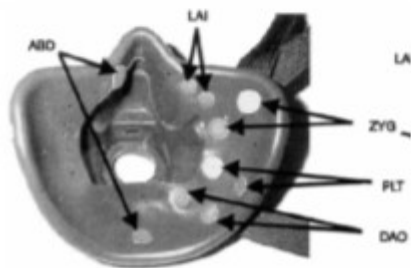
(ก)



(ข)



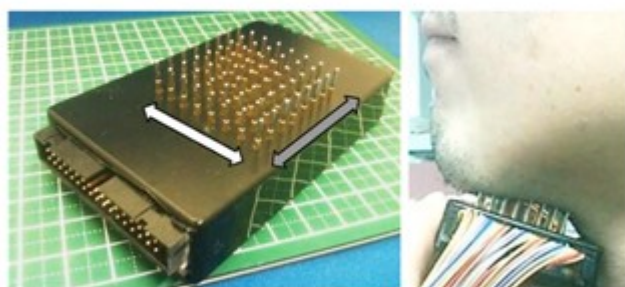
(ค)



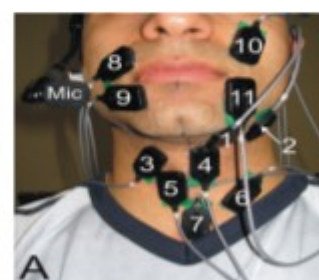
(ง)



(จ)



(ฉ)

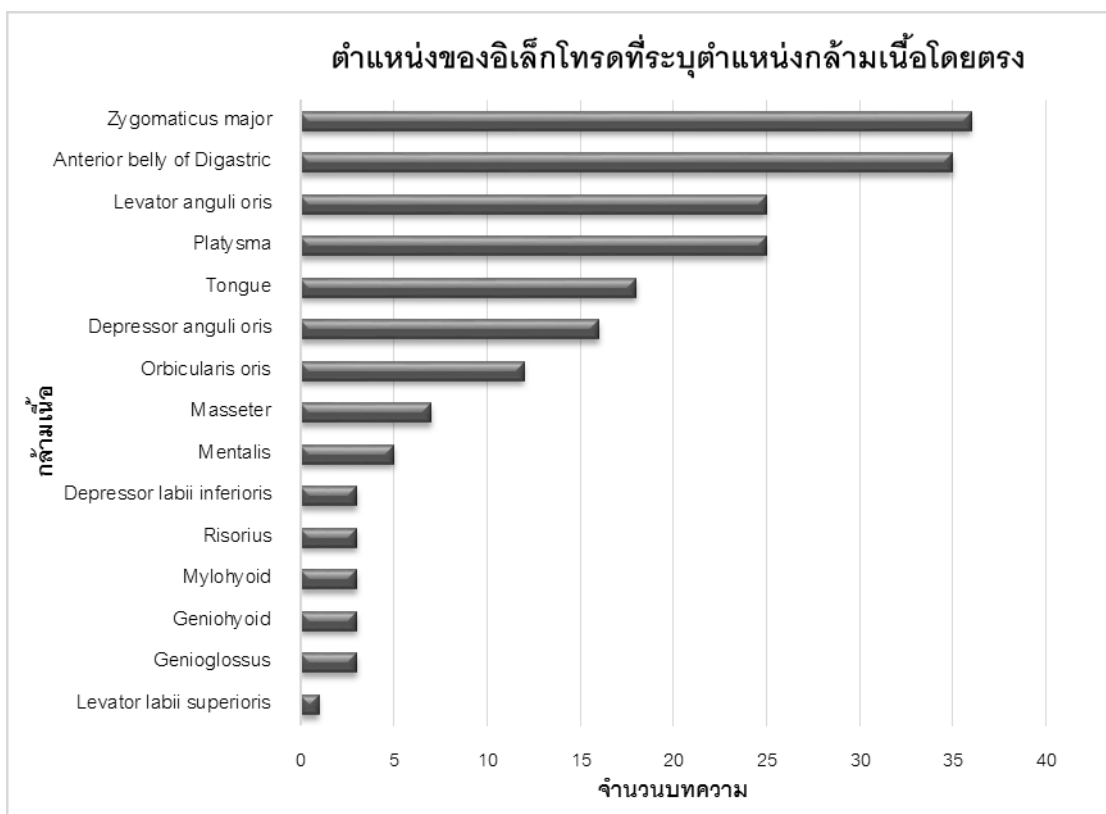


(ซ)

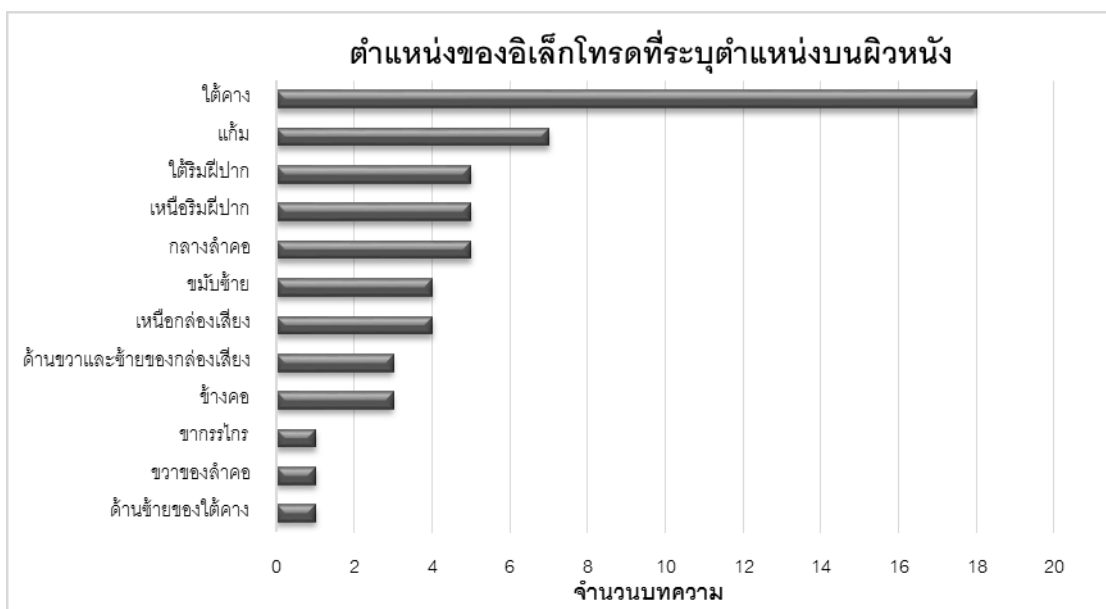
**รูปที่ 1.3** อิเล็กโทรดชนิดต่าง ๆ ที่ถูกนำมาใช้ในงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ (ก), (ข) คือ อิเล็กโทรดแบบกระดุมวงกลมซึ่งเป็นอิเล็กโทรดแบบเปียก โดยที่ (ก) คือ อิเล็กโทรดที่นำกลับมาใช้ใหม่ได้ [16] และ (ข) คือ อิเล็กโทรดที่ใช้ครั้งเดียวทิ้ง [33]; (ค) คือ อิเล็กโทรดแบบจัดเรียง [23]; (จ), (ซ) คือ อิเล็กโทรดแบบแท่ง ชนิดอิเล็กโทรดแห้ง [37], [41]; (ฉ) คือ อิเล็กโทรดแบบตาราง [39]

## (ข) ตำแหน่งของอวัยวะ

สำหรับตำแหน่งของอวัยวะที่ใช้ในการจับสัญญาณไฟฟ้ากล้ามเนื้อ จากการทบทวนวรรณกรรมพบว่า การระบุตำแหน่งดังกล่าวมีทั้งรูปแบบที่ระบุชื่อกล้ามเนื้อโดยตรง และระบุตำแหน่งบนผิวหนัง ในกรณีที่บางงานวิจัยระบุตำแหน่งอวัยวะโดยใช้ตำแหน่งของผิวหนังนั้น เนื่องจากกล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูดเป็นกล้ามเนื้อที่มีขนาดเล็กและมีการซ้อนทับกันระหว่างมัดกล้ามเนื้อแต่ละชนิด ทำให้ยากต่อการระบุมัดกล้ามเนื้อที่เฉพาะเจาะจง ซึ่งได้สรุปตำแหน่งอวัยวะที่ถูกใช้ในงานวิจัยทางด้านนี้ดังรูปที่ 1.4 สำหรับตำแหน่งของอวัยวะที่ระบุกล้ามเนื้อโดยตรง (รูปที่ 1.4 (ก)) พบว่ากล้ามเนื้อ Zygomaticus major และ Anterior belly of Digastric ถูกใช้มากที่สุด เมื่อพิจารณาหน้าที่หลักของกล้ามเนื้อ Zygomaticus major พบว่ามีหน้าที่ในการยกมุมปากขึ้นไปข้างบนและออกไปด้านข้าง ซึ่งช่วยในการออกเสียงเสียดแทรก (fricative) ที่เกิดขึ้นที่ริมฝีปากและฟัน (labiodentals) [64] (เช่น เสียง /f/) และช่วยในการออกเสียงสระ (เสียง /i/ , /e/) สำหรับกล้ามเนื้อ Anterior belly of Digastric ทำหน้าที่ในการเคลื่อนที่กระดูกขากรรไกรล่าง (jaw bone) ดึงกระดูกโคนลิ้น (hyoid bone) ขึ้นบนและไปข้างหน้า เพื่อช่วยนำลิ้นไปข้างหน้าและยกขึ้น กล้ามเนื้อนี้ช่วยในการออกเสียงที่เกิดขึ้นที่ฟันและปุ่มเหงือก (alveolar) ( เสียง /t/ , /d/ , /n/ , /r/ , /s/ , /l/) [64] และช่วยในการออกเสียงสระหน้าสูง (high front vowels) ทั้งสองมัดกล้ามเนื้อนี้ถูกนำมาใช้มากในงานวิจัยเนื่องจากเป็นกล้ามเนื้อที่ช่วยในการออกเสียงสระเป็นหลัก ซึ่งงานวิจัยในช่วงแรกๆ ส่วนใหญ่จะเน้นรู้จำเสียงสระ โดยเฉพาะงานวิจัยของญี่ปุ่นซึ่งจำแนกเสียงสระญี่ปุ่น ในภาษาญี่ปุ่นสระเป็นองค์ประกอบที่สำคัญมากดังนั้นจึงนิยมใช้กล้ามเนื้อดังกล่าวในการรู้จำ อย่างไรก็ตามมีการใช้กล้ามเนื้อชนิดอื่นอีกหลากหลาย เนื่องจากกล้ามเนื้อที่ช่วยในการออกเสียงพูดทั้งหมดเป็นกล้ามเนื้อที่มีขนาดเล็ก และซ้อนทับกันอยู่หลาย ๆ มัดกล้ามเนื้อ ซึ่งหน้าที่ของแต่ละกล้ามเนื้อก็มีความคล้ายคลึงกัน ดังนั้นการเลือกใช้กล้ามเนื้อในงานวิจัยนี้จึงมีความหลากหลาย สำหรับการระบุกล้ามเนื้อที่จะนำมาใช้งานโดยระบุตำแหน่งบนผิวหนัง (รูปที่ 1.4 (ข)) พบว่า ตำแหน่งใต้คาง (under the chin) ถูกเลือกใช้มากที่สุด ซึ่งเป็นตำแหน่งที่ครอบคลุมหลายมัดกล้ามเนื้อ เช่น กล้ามเนื้อ Anterior belly of Digastric กล้ามเนื้อการเคลื่อนที่ของลิ้น (tongue) และ กล้ามเนื้อ Platysma ซึ่งกล้ามเนื้อ Platysma นี้เป็นกล้ามเนื้อที่อยู่ชั้นนอกสุดหุ้มกล้ามเนื้อรอบคอทั้งหมด



(ก)



(ข)

**รูปที่ 1.4** ตำแหน่งของอวัยวะที่ถูกใช้ในงานวิจัยจากการทบทวนวรรณกรรม (ก) ระบุกล้ามเนื้อโดยตรง (ข) ระบุตำแหน่งบนผิวหนัง



### (ค) จำนวนช่องสัญญาณ

สำหรับจำนวนช่องสัญญาณที่ใช้ในการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อในงานวิจัยนี้มีการใช้จำนวนช่องสัญญาณตั้งแต่ 1 ช่องสัญญาณ ถึง 56 ช่องสัญญาณ แต่ในงานวิจัยที่ผ่านมาพบว่า การใช้ช่องสัญญาณมากกว่า 5 ช่องสัญญาณไม่มีผลในการปรับปรุงประสิทธิภาพของระบบ [6], [41] แต่การใช้ช่องสัญญาณน้อยกว่า 5 ช่องสัญญาณจะทำให้ประสิทธิภาพในการจำแนกลดลง อย่างไรก็ตามการใช้จำนวนช่องสัญญาณมากจะทำให้เกิดปัญหาความไม่สะดวกในการใช้งานเมื่อพิจารณาถึงการประยุกต์ใช้งานจริง เนื่องจากการติดอิเล็กโทรดหลายอันบนใบหน้าและลำคอทำให้การใช้งานไม่สะดวก ดังนั้นจึงมีบางงานวิจัยที่พยายามลดจำนวนอิเล็กโทรดและยังคงทำให้ประสิทธิภาพของระบบเท่าเดิม เช่น งานวิจัย [42] ใช้จำนวนช่องสัญญาณเพียง 1 ช่องสัญญาณ ซึ่งได้ปรับปรุงกระบวนการเตรียมสัญญาณเบื้องต้น โดยใช้เทคนิคการเพิ่มอัตราสุ่ม (over-sampling) และเทคนิคการแยกองค์ประกอบของเวฟเลตสี่ชั้น (4 layer wavelet decomposition) เพื่อลดสัญญาณรบกวน และใช้ตัวแปลงลงแบบดิจิตอล (down digital converter) สำหรับลดความถี่ในการสุ่มสัญญาณให้เท่ากับความถี่ไนควิสต์ (Nyquist frequency) โดยไม่สูญเสียข้อมูล และปรับปรุงกระบวนการสกัดลักษณะเด่นโดยใช้ลักษณะเด่นแบบอัตโนมัติถดถอย (autoregressive) ซึ่งเทคนิคดังกล่าวเหมาะสำหรับการใช้ช่องสัญญาณเพียง 1 ช่องสัญญาณ และออกแบบตัวจำแนกโดยใช้ตัวจำแนกโครงข่ายประสาทเทียม (artificial neural network, ANN) 2 โครงข่ายอนุกรมกันเพื่อจำแนกสระภาษาจีน 6 ตัว ซึ่งต่อมาก็ได้ขยายการทดลองโดยจำแนกตัวเลขจีนอีก 11 ตัว และได้ปรับปรุงการสกัดลักษณะเด่นโดยเพิ่มลักษณะเด่นแบบการแปลงเวฟเลต (wavelet transform) และ สัมประสิทธิ์เซปสตรัล (cepstral coefficient) และเปลี่ยนตัวจำแนกเป็นซัพพอร์ตเวกเตอร์แมชชีน (support vector machine, SVM) อย่างไรก็ตามผลการทดลองนี้ยังจำกัดอยู่ที่การจำแนกสระภาษาจีน 6 ตัว และตัวเลขจีน 11 ตัว ซึ่งยังไม่สามารถสรุปได้ว่าวิธีนี้สามารถใช้งานได้ดีเมื่อต้องการจำแนกคำพูดที่หลากหลายกว่านี้ และยังมีอีกหนึ่งงานวิจัยที่แก้ปัญหาที่คือ งานวิจัย [43] ได้ทำการลดจำนวนช่องสัญญาณเช่นเดียวกัน แต่แทนที่จะใช้สัญญาณไฟฟ้ากล้ามเนื้อซึ่งวัดจากจากความแตกต่างของสัญญาณไฟฟ้ากล้ามเนื้อจากอิเล็กโทรดแบบสองขั้วบนกล้ามเนื้อเดียว ก็ใช้สัญญาณไฟฟ้ากล้ามเนื้อจากความแตกต่างของสัญญาณไฟฟ้ากล้ามเนื้อระหว่างกล้ามเนื้อ 2 มัด จากอิเล็กโทรดแบบขั้วเดียวแทน และปรับปรุงกระบวนการจำแนกให้เหมาะกับการลดจำนวนอิเล็กโทรดดังกล่าว อย่างไรก็ตามก็ยังจำกัดอยู่ที่การจำแนกสระเพียง 6 ตัว

ถึงแม้ว่าการเพิ่มจำนวนช่องสัญญาณของอิเล็กทรอนิกส์โทรคมนาคมเกินไปจะมีข้อเสียดังที่ได้กล่าวไว้แล้วข้างต้น อย่างไรก็ตามมีงานวิจัยที่ออกแบบให้จับสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้จำนวนช่องสัญญาณมากถึง 35 และ 56 ช่อง [38]-[40], [23] โดยใช้อิเล็กทรอนิกส์โทรคมนาคมแบบใหม่ที่เรียกว่าอิเล็กทรอนิกส์โทรคมนาคมจัดเรียงและอิเล็กทรอนิกส์โทรคมนาคมแบบตาราง ซึ่งอิเล็กทรอนิกส์โทรคมนาคมชนิดนี้มีขนาดเล็กแต่สามารถตรวจจับสัญญาณได้หลายช่อง อิเล็กทรอนิกส์โทรคมนาคมชนิดนี้ออกแบบมาเพื่อให้สามารถจับสัญญาณของกล้ามเนื้อได้หลายจุดในแต่ละกล้ามเนื้อ ซึ่งสัญญาณจากช่องสัญญาณทั้งหมดไม่ได้ถูกนำมาใช้ทุกสัญญาณ จะมีการเลือกเฉพาะบางสัญญาณเท่านั้น ซึ่งเป็นข้อดีทำให้สามารถเลือกสัญญาณที่ดีที่สุดที่จะใช้ในกระบวนการจำแนกต่อไป

### (ง) จำนวนอาสาสมัครและประเภทของอาสาสมัครที่จะถูกจับสัญญาณ

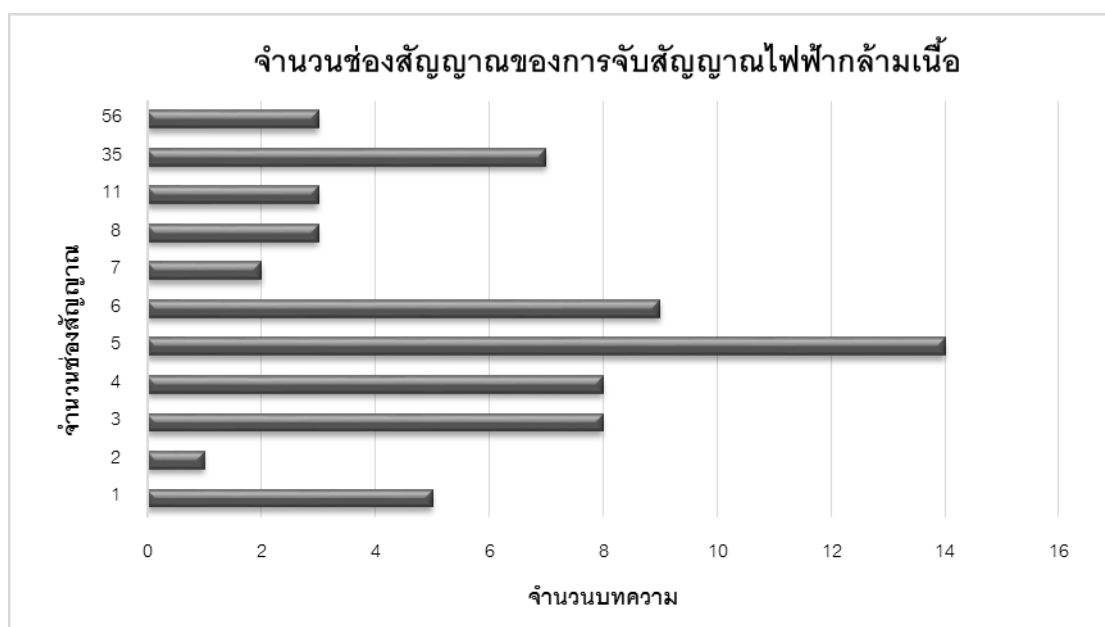
สำหรับงานวิจัยทางด้านนี้ พบว่าปัจจุบันมีเพิ่มจำนวนของอาสาสมัครในการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้อาสาสมัครจำนวนมาก ซึ่งมากที่สุดถึง 78 คน [13] โดยเป็นกลุ่มวิจัยของเยอรมันซึ่งมีการพัฒนางานวิจัยทางด้านนี้มากที่สุดในขณะนี้ ซึ่งได้ทำการทดลองโดยใช้คลังข้อมูลจำนวนมาก สำหรับงานวิจัยอื่นๆ ส่วนมากจะใช้จำนวนอาสาสมัครประมาณ 1-3 คน เนื่องจากอยู่ในช่วงเริ่มแรกของงานวิจัย สำหรับประเภทของอาสาสมัครส่วนใหญ่จะกำหนดอาสาสมัครที่มีสุขภาพดี มีบางงานวิจัยเท่านั้นที่ใช้อาสาสมัครที่เป็นผู้ป่วย ซึ่งเน้นศึกษาเพื่อประยุกต์ใช้งานทางการแพทย์โดยตรง เพื่อช่วยผู้ป่วยที่มีปัญหาบกพร่องทางการพูด

### (จ) คำพูดที่ใช้ในการฝึกสอนระบบ (training) และการรู้จำหรือจำแนกระบบ

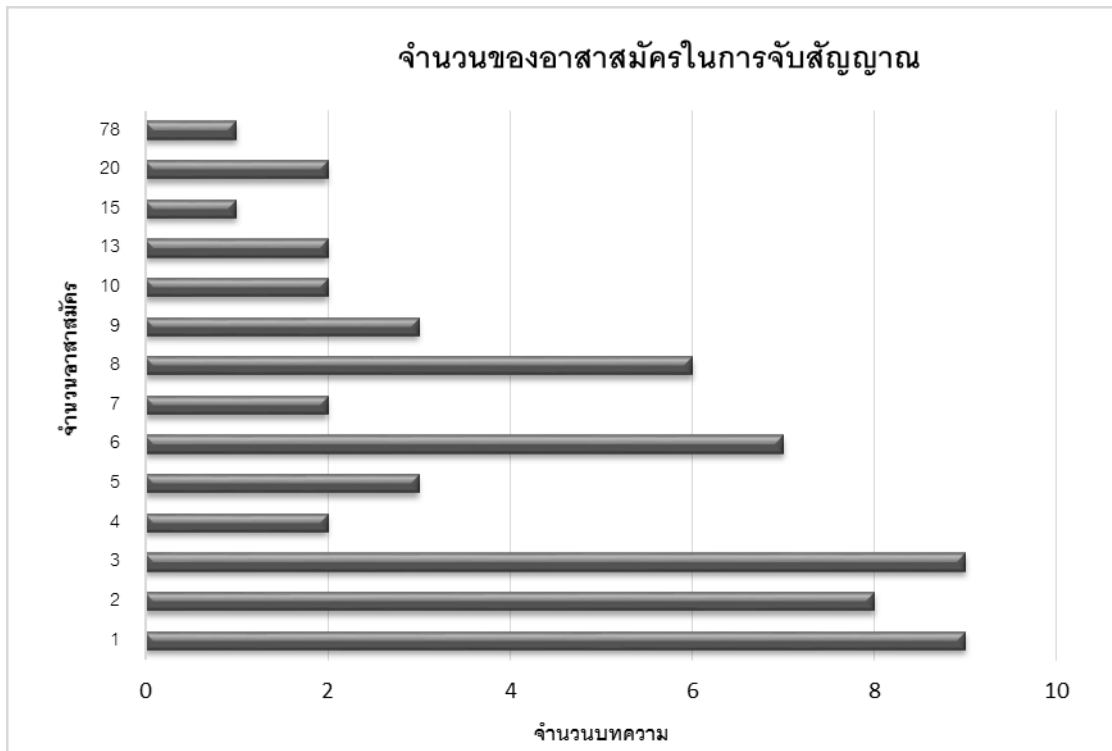
ในช่วงแรกๆ งานวิจัยส่วนใหญ่จะเริ่มต้นจากการรู้จำคำโดด (isolated word) ประมาณ 5-10 คำ โดยจะเป็นการรู้จำสระ ตัวเลขโดด คำสั่งสั้น ๆ ที่ใช้ในงานเฉพาะ ข้อดีของการรู้จำคำโดดคือ ง่ายในการตรวจจับช่วงเริ่มต้นและช่วงสิ้นสุดของคำพูด แต่ข้อเสียคือ ต้องสร้างแบบจำลองเสียง (acoustic model) ในการเรียนรู้จำนวนมาก ต่อมาจึงได้มีการพัฒนาการรู้จำพยางค์และหน่วยเสียงแทนคำโดด (หน่วยเสียงคือ ส่วนที่เล็กที่สุดของภาษาซึ่งเป็นองค์ประกอบย่อยของคำพูดแต่ละคำ) ข้อดีของการรู้จำหน่วยเสียงและพยางค์คือ ทำให้สร้างแบบจำลองเสียงน้อยลง และสามารถขยายคำที่จะรู้จำได้มากขึ้นจากแบบจำลองเสียงเดิมที่มีอยู่ นอกจากนี้การรู้จำหน่วยเสียงนั้นยังเป็นพื้นฐานในการพัฒนาการรู้จำแบบคำพูดต่อเนื่องได้ ซึ่งงานวิจัยในขณะนี้พบว่ามีอยู่สองกลุ่มวิจัย คือ กลุ่มเยอรมัน [9]-[29] และ สหรัฐอเมริกา 3 [44] ที่สามารถพัฒนาระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ โดยที่สามารถรู้จำคำพูดต่อเนื่องได้ และรู้จำคำ

ได้จำนวนมากถึง 2,012 คำ [18] และ 2,517 คำ [44] สำหรับกลุ่มงานวิจัยอื่นนั้น ยังคงจำกัดอยู่ที่การรู้จำคำโดด ซึ่งคำพูดที่ใช้ในการรู้จำมากที่สุด 65 คำ

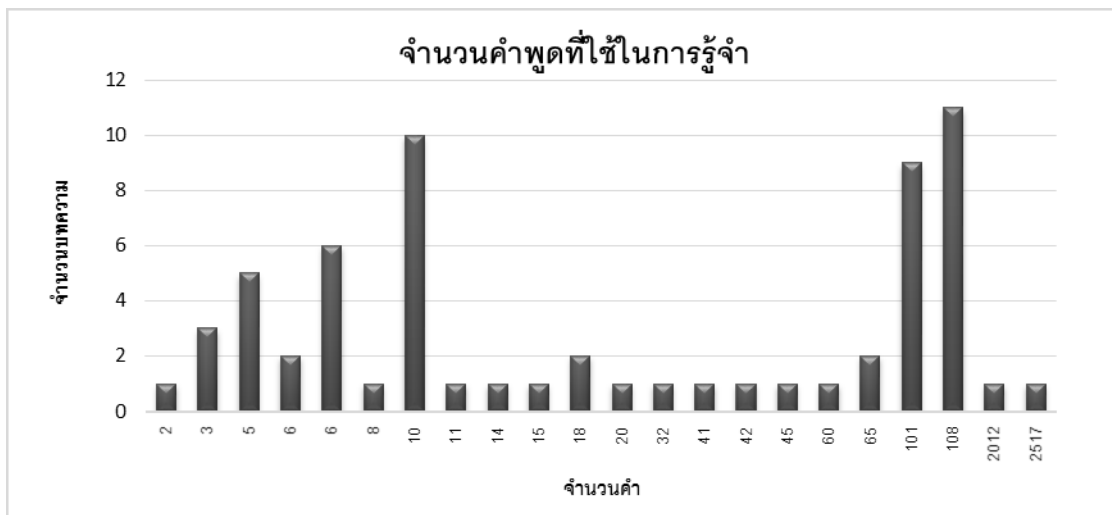
สำหรับภาษาที่ใช้ในการรู้จำส่วนใหญ่จะเป็นภาษาอังกฤษ เนื่องจากเป็นภาษาสากลที่ใช้กันทั่วโลก ส่วนภาษาอื่นๆ ที่มีการวิจัย ได้แก่ ภาษาจีน [42], [45]-[46] ภาษาญี่ปุ่น [1], [37]-[40], [43], [47]-[48] ภาษาเกาหลี [49] ภาษาโปรตุเกส [50]-[51] ภาษาสเปน [52] ภาษาอารบิก [53] และภาษาเซอร์เบีย [54] สำหรับภาษาที่แตกต่างกันมีส่วนทำให้ระบบรู้จำการพูดของแต่ละภาษาแตกต่างกันด้วย เพราะคุณลักษณะเฉพาะของแต่ละภาษาแตกต่างกันดังนั้นจึงมีการทำวิจัยหลากหลายภาษา ซึ่งส่วนใหญ่ก็นักวิจัยแต่ละกลุ่มก็ทำการวิจัยภาษาของตัวเองเพื่อประโยชน์ในการพัฒนาประเทศของตัวเอง



**รูปที่ 1.5** จำนวนช่องสัญญาณของการจับสัญญาณไฟฟ้ากล้ามเนื้อที่ถูกใช้ในงานวิจัยจากการทบทวนวรรณกรรม



**รูปที่ 1.6** จำนวนข้อมูลของอาสาสมัครในการจับสัญญาณ สำหรับงานวิจัยนี้จากการทบทวนวรรณกรรม



**รูปที่ 1.7** จำนวนคำพูดที่ใช้ในการรู้จำ จากการทบทวนวรรณกรรม ซึ่งมีทั้งการรู้จำคำพูดโดด คำพูดโดด และคำพูดต่อเนื่อง

## 1.2.2 กระบวนการเตรียมสัญญาณเบื้องต้น

ในส่วนกระบวนการเตรียมสัญญาณเบื้องต้นที่จะกล่าวถึงในที่นี้ คือ กระบวนการกรองสัญญาณรบกวน (filtering) และ กระบวนการแบ่งส่วนของสัญญาณ (segmentation)

### (ก) กระบวนการกรองสัญญาณรบกวน

เนื่องจากสัญญาณไฟฟ้ากล่อมเนื้อที่ได้จากกระบวนการเก็บข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อขณะพูดถูกรบกวนด้วยสัญญาณรบกวนหลากหลายชนิด ดังนั้นก่อนนำสัญญาณดังกล่าวไปประมวลผลในขั้นตอนต่อไป จำเป็นต้องผ่านกระบวนการกรองสัญญาณรบกวนก่อน ปกติสัญญาณไฟฟ้ากล่อมเนื้อที่มีองค์ประกอบความถี่หลักอยู่ที่ 50-150 เฮิร์ตซ์ [65] ซึ่งครอบคลุมช่วงความถี่ตั้งแต่ 20-500 เฮิร์ตซ์ สัญญาณรบกวนที่ปนมากับสัญญาณไฟฟ้ากล่อมเนื้อ ได้แก่ สัญญาณรบกวนล้อมรอบ (ambient noise) สัญญาณรบกวนการเคลื่อนที่ (motion artifacts noise) สัญญาณรบกวนจากสัญญาณไฟฟ้ากล่อมเนื้อจากกล้ามเนื้อที่อยู่ติดกัน (cross talk noise) สัญญาณรบกวนของสัญญาณทางชีวภาพอื่น เป็นต้น

สัญญาณรบกวนล้อมรอบ เป็นสัญญาณรบกวนที่มีความถี่ 50 เฮิร์ตซ์ หรือ 60 เฮิร์ตซ์ โดยปกติจะกำจัดสัญญาณรบกวนประเภทนี้ด้วยตัวกรองนอตช์ (Notch filter) แต่พบว่างานวิจัยบางกลุ่มโดยเฉพาะงานวิจัยของกลุ่มเยอรมันจะไม่กำจัดสัญญาณรบกวนประเภทนี้ เนื่องจากทำให้สูญเสียข้อมูลที่สำคัญของสัญญาณไฟฟ้ากล่อมเนื้อไป อย่างไรก็ตามมีหลายงานวิจัยที่กำจัดสัญญาณรบกวนประเภทนี้ เนื่องจากสัญญาณรบกวนดังกล่าวมีขนาดแอมพลิจูดสูงกว่าสัญญาณไฟฟ้ากล่อมเนื้อมาก ทำให้มีผลในการลดความแม่นยำในการจำแนก

สัญญาณรบกวนการเคลื่อนที่ เกิดจากการเคลื่อนที่ของสายนำสัญญาณที่ต่อระหว่างอิเล็กโทรดและเครื่องขยาย สัญญาณรบกวนประเภทนี้มีความถี่อยู่ในช่วง 0-20 เฮิร์ตซ์ การแก้ไขปัญหาของสัญญาณรบกวนประเภทนี้เริ่มต้นจากการแก้ปัญหาที่ต้นเหตุ กล่าวคือ การแปะสายนำสัญญาณเข้ากับผิวหนังเพื่อลดการเคลื่อนที่ของสาย อย่างไรก็ตามไม่สามารถกำจัดสัญญาณรบกวนดังกล่าวได้หมด ดังนั้นจึงต้องมีการกรองสัญญาณโดยใช้ตัวกรองอีกครั้งหนึ่ง

สัญญาณรบกวนจากสัญญาณไฟฟ้ากล่อมเนื้อจากกล้ามเนื้อที่อยู่ติดกัน ปัญหาของสัญญาณรบกวนประเภทนี้แก้โดยการเลือกใช้ขนาดอิเล็กโทรดที่เหมาะสม ซึ่งไม่ควรใหญ่กว่าขนาดมัดกล้ามเนื้อที่ต้องการวัด และระยะทางระหว่างอิเล็กโทรดสำหรับการใช้อิเล็กโทรดแบบสอง

ชั่วคราวจะห่างกัน 20 มิลลิเมตร [66] ซึ่งจากงานวิจัยส่วนใหญ่ก็ได้พิจารณาการใช้งานอิเล็กทรอนิกส์โทรดตามหลักการดังกล่าว ทำให้สามารถลดสัญญาณรบกวนประเภทนี้ได้

นอกจากนี้ยังมีสัญญาณรบกวนอีกประเภทหนึ่งที่ปนมากับสัญญาณไฟฟ้ากล้ามเนื้อ คือ สัญญาณทางชีวภาพอื่นๆ ได้แก่ สัญญาณคลื่นไฟฟ้าหัวใจ (Electrocardiography, ECG) ซึ่งสำหรับในงานวิจัยนี้จะพบสัญญาณประเภทนี้ก็ต่อเมื่อวัดสัญญาณไฟฟ้ากล้ามเนื้อจากตำแหน่งกล้ามเนื้อคอบางตำแหน่ง ซึ่งอาจทำให้มีสัญญาณคลื่นไฟฟ้าหัวใจแทรกอยู่ ดังนั้นจำเป็นต้องทำการกรองสัญญาณรบกวนดังกล่าวนี้ออกไป

ในกระบวนการกรองสัญญาณรบกวน แต่ละงานวิจัยมีแนวทางการกำจัดสัญญาณรบกวนที่แตกต่างกันออกไป เช่น งานวิจัยกลุ่มของเยอรมัน ในช่วงแรกจะกำจัดสัญญาณรบกวนโดยใช้ตัวกรองความถี่สูงผ่าน (high pass filter) ซึ่งมีความถี่ตัดผ่าน 1 เฮิร์ตซ์ และตัวกรองความถี่ต่ำผ่าน (low pass filter) ซึ่งมีความถี่ตัดผ่าน 300 เฮิร์ตซ์ และไม่มีการใช้ตัวกรองนอตซ์ในการกำจัดสัญญาณรบกวนล้อมรอบซึ่งมีความถี่ 50 เฮิร์ตซ์ หรือ 60 เฮิร์ตซ์ [6]-[16] ต่อมามีการเปลี่ยนรูปแบบการกรองสัญญาณรบกวนโดยใช้ตัวกรองความถี่สูงผ่าน ซึ่งมีความถี่ตัดผ่าน 60 เฮิร์ตซ์ เพียงอย่างเดียว [17]-[22] และในระยะหลังงานวิจัยกลุ่มนี้ได้คิดค้นรูปแบบอิเล็กทรอนิกส์แบบใหม่ซึ่งเรียกว่า อิเล็กโทรดแบบจัดเรียง [22]-[26] ทำให้ต้องเปลี่ยนแปลงรูปแบบการกรองสัญญาณรบกวน โดยใช้ตัวกรองความถี่สูงผ่านซึ่งมีความถี่ตัดผ่าน 3 เฮิร์ตซ์ และ ตัวกรองความถี่ต่ำผ่านซึ่งมีความถี่ตัดผ่าน 900 เฮิร์ตซ์ สำหรับงานวิจัยของกลุ่มอื่นๆ ส่วนใหญ่จะใช้ตัวกรองนอตซ์ และใช้ตัวกรองแถบความถี่ผ่าน (band pass filter) ซึ่งกำหนดช่วงความถี่ตัดผ่าน 20-450 เฮิร์ตซ์ [36]-[37], [41], [47], [55] หรือ 30-500 เฮิร์ตซ์ [56] ซึ่งสัญญาณรบกวนที่เกิดขึ้นสำหรับแต่ละงานวิจัยก็แตกต่างกันออกไปทำให้การกรองสัญญาณรบกวนแตกต่างกัน แต่จะคล้ายคลึงกันในส่วนหลัก นอกจากนี้พบว่าในขณะการออกเสียงพูดบางครั้งมีการกลืนน้ำลาย หรือ การไอจามเกิดขึ้น ดังนั้นมีบางงานวิจัยที่ทำการกำจัดสัญญาณที่เกิดจากการกลืนหรือการไอจามโดยใช้โปรแกรม SCAN 4 Neuroscan [56]-[57] กระบวนการกรองสัญญาณรบกวนเป็นส่วนสำคัญซึ่งมีผลต่อความแม่นยำของระบบ ดังนั้นจึงต้องมีแนวทางในการกำจัดสัญญาณรบกวนออกไปให้ได้มากที่สุด

### (ข) กระบวนการแบ่งส่วนของสัญญาณไฟฟ้ากล้ามเนื้อ

กระบวนการแบ่งส่วนของสัญญาณไฟฟ้ากล้ามเนื้อ เป็นการแบ่งส่วนของสัญญาณไฟฟ้ากล้ามเนื้อและดึงเอาเฉพาะส่วนที่เป็นสัญญาณจากการพูดไปประมวลผล สำหรับ

ระบบรู้จำการพูดที่เป็นคำโดดจะง่ายต่อการแบ่งส่วนสัญญาณมากกว่าการรู้จำการพูดที่เป็นหน่วยเสียง เช่น หน่วยเสียงพยัญชนะ หรือ หน่วยเสียงสระ ซึ่งปรากฏอยู่ในคำโดด เนื่องจากกระบวนการพูดคำหนึ่งคำนั้น เสียงพยัญชนะ เสียงสระ และเสียงพยัญชนะของตัวสะกด จะออกเสียงควบกันในเวลาสั้นๆ ทำให้ยากต่อการแยกสัญญาณที่เป็นเสียงพยัญชนะออกจากเสียงสระหรือเสียงวรรณยุกต์ ในบางงานวิจัย [6]-[29] จึงมีการจับสัญญาณเสียงควบคู่ไปกับสัญญาณไฟฟ้ากล้ามเนื้อเพื่อนำสัญญาณเสียงดังกล่าวไปกำหนดขอบเขตสำหรับการตัดสัญญาณไฟฟ้ากล้ามเนื้อ เพื่อแยกหน่วยเสียงแต่ละเสียงออกจากกัน สัญญาณเสียงถูกใช้ในกระบวนการแบ่งส่วนสัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียว ไม่ได้นำสัญญาณเสียงไปใช้ในกระบวนการจำแนกคำพูด อย่างไรก็ตามวิธีการตัดสัญญาณไฟฟ้ากล้ามเนื้อดังกล่าวจะเกิดปัญหาในกรณีที่ทำกรจำแนกการพูดแบบไม่ออกเสียง ซึ่งทำให้ไม่มีสัญญาณเสียงป้อนเข้าไปในระบบ นอกจากนี้ยังมีปัญหาที่เกิดมาจากการห้วงเวลาระหว่างสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง โดยปกติแล้วสัญญาณไฟฟ้ากล้ามเนื้อจะเกิดขึ้นก่อนสัญญาณเสียง ดังนั้นจึงมีการเลื่อนจุดที่จะต้องคำนวณของสัญญาณไฟฟ้ากล้ามเนื้อออกไปด้วยเวลาการห้วงของสัญญาณเสียง ซึ่งไม่สามารถกำหนดเวลาดังกล่าวที่แน่นอนได้ ดังนั้นมีหลายงานวิจัยที่ทำการศึกษาหาเวลาห้วงระหว่างสองสัญญาณ ซึ่งพบว่าเวลาดังกล่าวมีผลต่อประสิทธิภาพของระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ และเวลาที่ให้ประสิทธิภาพดีที่สุดแตกต่างกันสำหรับแต่ละงานวิจัย งานวิจัย [10] พบว่าเวลาห้วง 0.02-0.10 วินาที ให้ประสิทธิภาพดีที่สุด งานวิจัย [30] พบว่าเวลาห้วง 0.05 วินาที ให้ประสิทธิภาพดีที่สุด และ งานวิจัย [33] พบว่า 0.10 วินาที ให้ประสิทธิภาพดีที่สุด ดังนั้นถ้าต้องการแบ่งส่วนสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้สัญญาณเสียงกำหนดขอบเขตของการแบ่งสัญญาณ จำเป็นต้องมีการเลื่อนจุดของสัญญาณไฟฟ้ากล้ามเนื้อด้วยเวลาดังกล่าวก่อนจะนำไปประมวลผลในขั้นตอนต่อไป

นอกจากนี้ยังมีวิธีอื่น เช่น การแบ่งหน่วยเสียงแต่ละเสียงโดยมองภาพแบบสเปคโตรกราฟฟิก (spectrographic) และข้อมูลกลุ่มความถี่สั้นพ้อง (formant information) จากโปรแกรม MATLAB แล้วแบ่งสัญญาณแบบแมนนวล (manual) จากการสังเกตภาพและข้อมูลดังกล่าว [32]-[33] ซึ่งเป็นเทคนิคในการตัดสัญญาณไฟฟ้ากล้ามเนื้อสำหรับระบบรู้จำการพูดที่ต้องการรู้จำหน่วยเสียง

สำหรับในการรู้จำคำโดดมีความง่ายในการตัดสัญญาณ โดยทั่วไปการตัดสัญญาณไฟฟ้ากล้ามเนื้อเริ่มจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูดในแต่ละคำพูด ซึ่งในงานวิจัย ทำการตรวจจับโดยใช้ค่ารากเฉลี่ยกำลัง

สองของสัญญาณเป็นค่าขีดแบ่ง (threshold) เพื่อใช้ในการกำหนดจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อ [58] บางงานวิจัยใช้เทคนิคสถิติอันดับสาม (third order statistics, TOS) [36] เพื่อกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูด

ในกระบวนการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อเพื่อดึงเอาเฉพาะส่วนของเสียงพูดที่ต้องการนั้น พบว่ายังไม่มีวิธีใดที่ดีที่สุด บางงานวิจัยก็ยังคงใช้วิธีการตรวจจับแบบแมนนวลแต่บางงานวิจัยก็ใช้วิธีแบบอัตโนมัติ

### 1.2.3 กระบวนการสกัดลักษณะเด่นของสัญญาณ

กระบวนการนี้เป็นการนำสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูดแต่ละเสียงมาคำนวณหาลักษณะเด่น (feature) ของสัญญาณไฟฟ้ากล้ามเนื้อเพื่อเป็นตัวแทนของเสียงพูดแต่ละเสียง กระบวนการนี้เป็นการลดข้อมูลของสัญญาณและลดการซ้ำซ้อนของข้อมูล ลักษณะเด่นดังกล่าวจะถูกป้อนให้กับกระบวนการฝึกสอนระบบ และกระบวนการจำแนกข้อมูลต่อไป

การหาลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ สามารถแบ่งออกได้เป็น 3 ประเภท คือ ลักษณะเด่นในโดเมนเวลา (time domain feature) ลักษณะเด่นในโดเมนความถี่ (frequency domain feature) และ ลักษณะเด่นในโดเมนเวลาและความถี่ (time-frequency domain) สำหรับในงานวิจัยทางด้านนี้จะมีการใช้ลักษณะเด่นทั้ง 3 รูปแบบ ซึ่งลักษณะเด่นที่ถูกนำมาใช้ทั้งหมดในงานวิจัยถูกแสดงดังตารางที่ 1.1 ถึง ตารางที่ 1.3

สำหรับตารางที่ 1.1 แสดงลักษณะเด่นในโดเมนเวลา ตารางที่ 1.2 แสดงลักษณะเด่นในโดเมนความถี่ และ ตารางที่ 1.3 แสดงลักษณะเด่นในโดเมนเวลา-ความถี่ ซึ่งตารางดังกล่าวจะแสดงให้เห็นว่างานวิจัยนี้มีการนำลักษณะเด่นหลากหลายชนิดมาใช้ในการทดสอบเพื่อหาลักษณะเด่นที่เหมาะสมที่สุดในแต่ละงานวิจัย รูปที่ 1.8 แสดงลักษณะเด่นที่ถูกนำมาทดสอบมากที่สุด 10 อันดับแรก ซึ่งพบว่าลักษณะเด่นในโดเมนเวลาถูกนำมาใช้มากที่สุด ได้แก่ TD15 MN และ RMS

ลักษณะเด่น TD15 เดิมเรียกว่า E4 เป็นลักษณะเด่นที่ถูกนำเสนอโดยกลุ่มวิจัยของเยอรมัน [9] งานวิจัยนี้ได้เสนอลักษณะเด่นที่มีชื่อว่า EMG features ซึ่งเป็นลักษณะเด่นที่อาศัยหลักการของตัวกรองแบบบริบท (contextual filter) ซึ่งประกอบด้วยตัวกรอง 3 ชนิด คือ ตัวกรองเดลต้า (delta filter) ตัวกรองเทรนด์ (trend filter) และ ตัวกรองซ้อน (stacking filter) โดยที่



EMG features สามารถแบ่งเป็นลักษณะเด่นย่อยตามประเภทของตัวกรอง ได้แก่ E0 E1 E2 E3 และ E4 โดยที่ E0-E2 เป็นลักษณะเด่นที่ใช้หลักการของตัวกรองเดลด้าและตัวกรองเทรนต์ สำหรับ E3-E4 เป็นลักษณะเด่นที่ใช้หลักการของตัวกรองซ็อน จากงานวิจัย [9] ได้ทำการทดสอบประสิทธิภาพของ EMG features ทั้ง 5 ชนิด (E0-E4) พบว่า E4 มีประสิทธิภาพดีที่สุด หลักการคำนวณของ E4 คือ การแบ่งสัญญาณไฟฟ้ากล้ามเนื้อออกเป็นส่วนหรือการแบ่งกรอบของสัญญาณ สัญญาณในแต่ละส่วนถูกเรียกว่าเฟรม (frame) ซึ่งจะถูกนำมาคำนวณลักษณะเด่นในโดเมนเวลา 5 ชนิด ค่าที่คำนวณได้ในแต่ละเฟรมจะถูกนำมาซ้อนกันเพื่อให้ได้ลักษณะเด่นชนิดใหม่ จำนวนเฟรมที่นำมาซ้อนกันเรียกว่าความกว้างการซ้อน (stacking width) โดยที่ E4 จะใช้จำนวนเฟรมเท่ากับ 5 ต่อมางานวิจัย [13] พบว่าความกว้างการซ้อนที่เหมาะสมที่สุด คือ 15 เฟรม และได้เปลี่ยนชื่อจาก E4 เป็น TD15 ซึ่งกลุ่มวิจัยของเยอรมันใช้ลักษณะเด่นชนิดนี้ในการพัฒนาระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อมาตลอดจนถึงปัจจุบัน ซึ่งเหมาะสมสำหรับการรู้จำคำพูดต่อเนื่อง

ถึงแม้ว่าลักษณะเด่น E4 หรือ TD15 จะถูกนำมาใช้สำหรับกลุ่มวิจัยเพียงกลุ่มเดียว แต่กลุ่มวิจัยกลุ่มนี้ก็แสดงให้เห็นว่าลักษณะเด่นชนิดนี้มีประสิทธิภาพดีกว่าลักษณะเด่นชนิดอื่น งานวิจัย [9] ได้ทำการทดลองเพื่อเปรียบเทียบลักษณะเด่น 3 ชนิด (ดังตารางที่ 1.4) คือ สัมประสิทธิ์อนุพันธ์อันดับหนึ่งของการแปลงแบบฟูริเยร์ช่วงสั้น (delta coefficient of short time Fourier transform, delta STFT) สัมประสิทธิ์อนุพันธ์อันดับหนึ่งของการแปลงแบบฟูริเยร์ช่วงสั้นรวมกับค่าเฉลี่ยในโดเมนเวลา (delta STFT+Mean) และ ลักษณะเด่นแบบตัวกรองตามบริบทที่คำนวณจากลักษณะเด่นในโดเมนเวลา (EMG features) ซึ่งมีอยู่ 5 ชนิด คือ E0 E1 E2 E3 และ E4 จากการทดลองพบว่า ลักษณะเด่นชนิดสุดท้ายมีประสิทธิภาพดีที่สุด โดยที่ E4 ให้ประสิทธิภาพดีกว่าอีก 4 ชนิดตามที่ได้กล่าวไว้แล้วข้างต้น ซึ่งจากงานวิจัยนี้จะเห็นได้ว่า ลักษณะเด่นในโดเมนเวลามีประสิทธิภาพดีกว่าลักษณะเด่นในโดเมนความถี่ นอกจากนี้งานวิจัย [14] นำลักษณะเด่นแบบ TD15 (E4 with 15 stacking width) ไปเปรียบเทียบกับ ลักษณะเด่นแบบการแปลงเวฟเลตไม่ต่อเนื่องแบบรีดันแดนด์ (redundant discrete wavelet transform, RDWT) ซึ่งเป็นลักษณะเด่นในเชิงเวลา-ความถี่ พบว่าลักษณะเด่นแบบ TD15 มีประสิทธิภาพดีกว่าลักษณะเด่น RDWT จากงานวิจัยของกลุ่มนี้จะเห็นได้ว่าการใช้ลักษณะเด่นในโดเมนเวลาให้ประสิทธิภาพดีที่สุด

อย่างไรก็ตามบางงานวิจัยก็ให้ผลแตกต่างกับงานวิจัยของเยอรมัน เช่น งานวิจัย [35] ทำการเปรียบเทียบลักษณะเด่นแบบกระบวนการอัตโนมัติถดถอย (autoregressive, AR) กับ ลักษณะเด่นแบบการแปลงเวฟเลต (wavelet transform, WT) พบว่าลักษณะเด่นแบบ WT ให้

ประสิทธิภาพดีกว่า นอกจากนี้งานวิจัย [30] เปรียบเทียบการใช้ลักษณะเด่นแบบค่าเฉลี่ยในโดเมนเวลา (mean, MN) การแปลงฟูริเยร์ช่วงสั้น (STFT) การแปลงเวฟเลต (WT) และการแปลงเวฟเลตพอกเก็ต (wavelet pocket transform, WPT) ซึ่งพบว่าลักษณะเด่นแบบ WT ให้ประสิทธิภาพดีที่สุด จากงานวิจัยดังกล่าวจะเห็นได้ว่าลักษณะเด่นในเชิงเวลา-ความถี่มีประสิทธิภาพดีกว่าลักษณะเด่นในโดเมนเวลา อย่างไรก็ตามไม่สามารถสรุปได้ว่าการใช้ลักษณะเด่นชนิดใดดีที่สุดสำหรับงานวิจัยทางด้านนี้ แต่การใช้ลักษณะเด่นแบบ TD15 ดูเหมือนจะมีความเป็นไปได้มากกว่าเนื่องจากถูกนำไปใช้ในการจำแนกแบบคำพูดต่อเนื่อง และสามารถใช้ได้อย่างมีประสิทธิภาพ ถึงแม้ว่าความแม่นยำจะไม่สูงมากในขณะนี้ อีกทั้งงานวิจัยอื่นๆ ยังอยู่ในช่วงเริ่มต้นของงานวิจัย และยังคงทำการจำแนกเพียงคำพูดโดด จึงยังไม่สามารถสรุปได้ว่าลักษณะเด่นดังกล่าวจะสามารถใช้งานได้ดีเมื่อนำไปใช้กับระบบการจำแนกคำพูดต่อเนื่อง

สำหรับในระบบรู้จำการพูดแบบเดิมที่ใช้สัญญาณเสียงพูดเป็นอินพุตให้กับระบบพบว่าค่าลักษณะเด่นแบบสัมประสิทธิ์เซปสตรัมความถี่เมล (Mel-frequency cepstral coefficient, MFCC) ซึ่งเป็นลักษณะเด่นในเชิงความถี่ที่ถูกนำมาใช้มากที่สุด และพบว่ามีประสิทธิภาพมากที่สุด เนื่องจากเป็นลักษณะเด่นที่ปรับปรุงมาจากสัมประสิทธิ์เซปสตรัมแบบเดิมโดยการปรับสเกลของสเปคตรัมให้อยู่บนสเกลเมล ซึ่งเป็นสเกลที่ถูกออกแบบให้เก็บรายละเอียดของสัญญาณช่วงความถี่ต่ำ ซึ่งเป็นความถี่ที่สำคัญสำหรับการรับฟังเสียงของมนุษย์มากกว่าความถี่สูง สำหรับระบบรู้จำการพูดโดยสัญญาณไฟฟ้ากล่อมเนื้อ ได้มีการนำลักษณะเด่น MFCC มาใช้หลายงานวิจัย โดยไม่ได้ใช้ลักษณะเด่นชนิดนี้แบบเดี่ยว แต่จะรวมกับลักษณะเด่นในโดเมนเวลา เช่น ค่ารากเฉลี่ยกำลังสอง (RMS) เป็นต้น ทั้งนี้เนื่องจากในงานวิจัย [9] พบว่าลักษณะเด่นในโดเมนเวลาให้ประสิทธิภาพดีกว่าลักษณะเด่นในโดเมนความถี่ ดังนั้นงานวิจัย [34] จึงได้รวมเอาลักษณะเด่นในโดเมนเวลามาใช้รวมกับลักษณะเด่น MFCC ซึ่งพบว่าการรวมกันของสองลักษณะเด่นให้ประสิทธิภาพดีกว่าลักษณะเด่นแบบโดเมนเวลาเพียงอย่างเดียว

**ตารางที่ 1.1** ลักษณะเด่นในโดเมนเวลาทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม

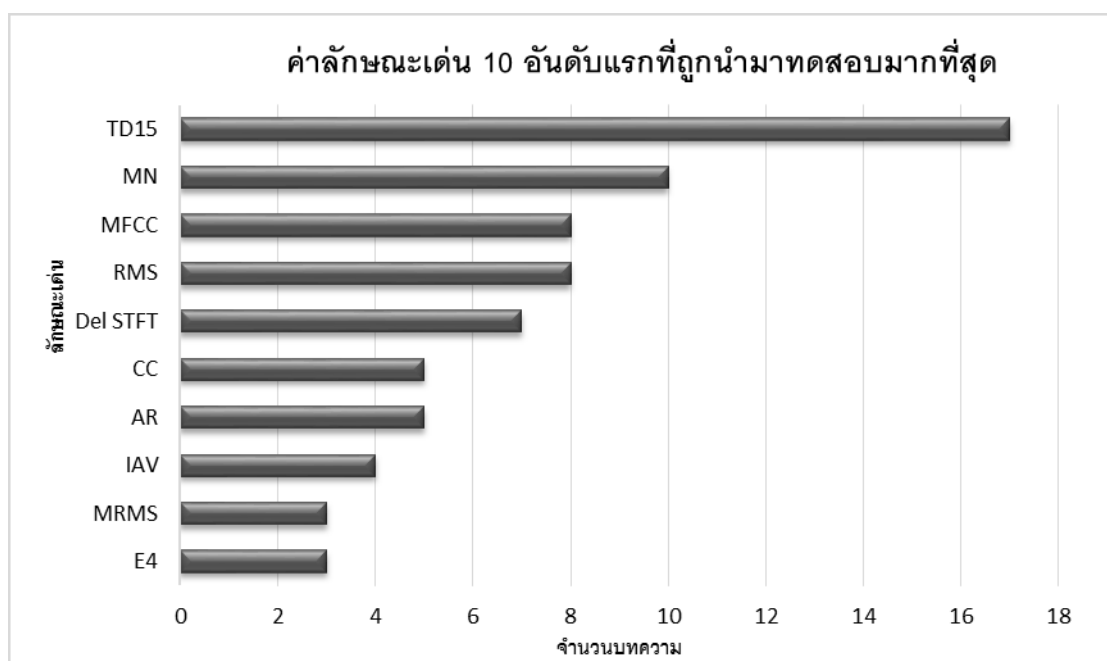
ลักษณะเด่นในโดเมนเวลา (Time domain feature)	
คำย่อ	ชื่อเต็ม
EMG features	EMG features (E0, E1, E2, E3, E4) using 3 types of contextual filters
E4	E4 with 5 frames stacking width
TD15	E4 with 15 frames stacking width (TD15)
MN	Mean
RMS	Root Mean Squared
MRMS	Normalized Integral Root Mean Squared
IV	Integrated Value
IAV	Integrated Absolute Value
MAV	Mean Absolute Value
ZC	Zero Crossing
ZCH	Zero Crossing of High pass signal
VAR	Variance
MX	Max Amplitude
TDC	Time Domain Context feature
MOA	Moving Average
Power	Signal Energy
SD	Standard Deviation
AR	Auto Regressive ( coefficient=2,10,20)
KT	Kurtosis
WD	Waveform Duration
MWL	Mean of Waveform Length
SSC	Slope Sign Change

**ตารางที่ 1.2** ลักษณะเด่นในโดเมนความถี่ทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม

ลักษณะเด่นในโดเมนความถี่ (Frequency domain feature)	
คำย่อ	ชื่อเต็ม
MFCC	Mel-Frequency Cepstral Coefficient
MFCC with M&MVN	Mel-Frequency Cepstral Coefficient with Mean and Variance Normalization
CC	Cepstral Coefficient
MSFB	Mel-Scale Filter Bank
CMN	Cepstral Mean Normalization
HT	Hartley Transform
LPC	Linear Predictive Coding
PSD	Power Spectral Density
FR	Frequency ratio (min/max magnitude FFT)
SM	Spectral Moment
SMA	Spectral Magnitude Average
MD	Median Frequency
MN	Mean Frequency
FFT	Fast Fourier Transform

**ตารางที่ 1.3** ลักษณะเด่นในโดเมนเวลา-ความถี่ทั้งหมดที่ถูกนำมาทดสอบในงานวิจัยจากการทบทวนวรรณกรรม

ลักษณะเด่นในโดเมนเวลา-ความถี่ (Time-Frequency domain feature)	
คำย่อ	ชื่อเต็ม
STFT	Short Time Fourier Transform
Del STFT	Delta coefficient of Short Time Fourier Transform
WT	Wavelet Transform
WPT	Wavelet Pocket Transform
FWT	Fast Wavelet Transform
DTWT	Dual Tree Wavelet Transform
DTCWT	Dual Tree Complex Wavelet Transform
2D DTCWT	2D Dual Tree Complex Wavelet Transform
RDWT	Redundant Discrete Wavelet Transform
SRDWT	Stacking filter of Redundant Discrete Wavelet Transform
CoWT	Coiflet Wavelet



**รูปที่ 1.8** ลักษณะเด่น 10 อันดับแรกที่ถูกนำมาทดสอบมากที่สุด จากการทบทวนวรรณกรรม

ตารางที่ 1.4 บทความที่มีการทดลองเพื่อเปรียบเทียบประสิทธิภาพของลักษณะเด่น

บทความ	ลักษณะเด่นที่ถูกนำมาทดสอบ	ลักษณะเด่นที่ดีที่สุด	ตัวจำแนก <sup>1</sup>	ประสิทธิภาพในการจำแนก (%)		ค่าที่ถูกจำแนก
				ACC <sup>2</sup>	WER <sup>3</sup>	
[8]	Del STFT+MN, TDC	TDC	HMM	82.9		คำโดด 32 คำ
[9]	Del STFT, Del STFT+MN, EMG features (E1, E2, E3, E4)	EMG feature (E4)	HMM		32	คำพูดต่อเนื่อง
[11]	Del STFT, DCTWT, FWT, RDWT, SRDWT	SRDWT	HMM		30.9	
[14]	TD15, RDWT	TD15	HMM		52.27	
[57]	DTWT, STFT, HT, MOA	DTWT	ANN	92		คำโดด 6 คำ
[56]	DTCWT, STFT	DTCWT	ANN	50		หน่วยเสียง 41 เสียง
[30]	MN, STFT, WT, WPT	WT	LDA	98		คำโดด 10 คำ
[47]	MSFB, LPC, MFCC	MFCC	HMM	80		คำ
[38]	MAV, RMS, ZCH, CC, delta of 4 features, delta delta of 4 features	CC	HMM	80-85		สระ 5 เสียง
[39]	2D DTCWT, CC	CC	HMM	-		
[3]	Power, MN, SD	MN	ML	-		10 คำโดด
[35]	AR, WT	WT		77		

<sup>1</sup> HMM: Hidden Markov Model, LDA: Linear Discriminant Analysis, ML: Maximum Likelihood, ANN: Artificial Neural Network

<sup>2</sup> ACC คือ อัตราความถูกต้องในการจำแนก (accuracy rate)

<sup>3</sup> WER คือ อัตราความผิดพลาดในการจำแนก (word error rate)

### 1.2.4 กระบวนการลดมิติของข้อมูล

สำหรับงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ มีบางงานวิจัยที่ใช้ลักษณะเด่นที่ให้ผลเป็นพารามิเตอร์จำนวนมากหรือมีการใช้ลักษณะเด่นหลายชนิด รวมถึงมีการใช้ช่องสัญญาณหลายช่องสัญญาณ ทำให้ข้อมูลที่ป้อนเข้าสู่ตัวจำแนกเพิ่มมากขึ้น ส่งผลให้ระบบทำงานช้าลงเนื่องจากต้องใช้เวลาในการคำนวณมากและซับซ้อนขึ้น ดังนั้นการลดมิติข้อมูลจึงจำเป็นในขั้นตอนนี้

เนื่องจากงานวิจัยทางด้านนี้ยังอยู่ในช่วงเริ่มแรก และบางงานวิจัยก็ใช้จำนวนช่องสัญญาณในการจับสัญญาณไฟฟ้ากล่อมเนื้อน้อย รวมถึงการใช้ลักษณะเด่นเพียงชนิดเดียว ทำให้ไม่จำเป็นต้องลดมิติข้อมูล แต่ก็มีอีกหลายงานวิจัยที่ใช้เทคนิคการลดมิติของข้อมูล จากการทบทวนวรรณกรรม พบว่าการใช้เทคนิคการวิเคราะห์การจำแนกประเภทแบบเชิงเส้น (linear discriminant analysis, LDA) ถูกนำมาใช้มากที่สุด LDA เป็นเทคนิคที่ผสมลักษณะเด่นในเวกเตอร์นำเข้าแบบเชิงเส้นเพื่อสร้างเวกเตอร์ใหม่ที่อยู่ในปริภูมิที่มีมิติต่ำกว่าเวกเตอร์เดิม ทำให้สามารถลดมิติของข้อมูลลงมาได้ โดยทำให้ข้อมูลจากกลุ่มเดียวกันเข้าใกล้กันมากขึ้น และข้อมูลที่อยู่ต่างกลุ่มห่างกันมากขึ้น นอกจากนี้ผลของการผสมลักษณะเด่นนี้สามารถใช้เป็นตัวจำแนกแบบเชิงเส้นได้อีกทางหนึ่ง ซึ่งในงานวิจัยด้านนี้ก็มีการใช้เทคนิคนี้เป็นตัวจำแนกข้อมูล สำหรับเทคนิคการลดมิติของข้อมูลวิธีอื่นที่นิยมใช้รองลงมาคือ การทำนายการเข้ารหัสแบบเชิงเส้น (linear predictive coding, LPC) เทคนิคการวิเคราะห์องค์ประกอบหลัก (principal component analysis, PCA)

### 1.2.5 กระบวนการรู้จำระบบ

ก่อนเข้าสู่กระบวนการจำแนกข้อมูล ระบบจะนำลักษณะเด่นที่ผ่านการลดมิติของอินพุตสู่กระบวนการฝึกสอน เพื่อให้ระบบสร้างแบบจำลองที่ใช้ในการเรียนรู้เสียงพูดแต่ละเสียง ก่อนที่จะนำแบบจำลองดังกล่าวเข้าสู่ขั้นตอนการทดสอบ เพื่อการจำแนกข้อมูล ซึ่งทั้งสองกระบวนการนี้อาศัยเครื่องมือที่เรียกว่าตัวจำแนก (classifier)

แบบจำลองที่ถูกสร้างจากกระบวนการเรียนรู้จากตัวจำแนกในระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล่อมเนื้ออยู่ 2 แบบจำลองหลัก คือ แบบจำลองเสียง และ แบบจำลองภาษา (language model) แบบจำลองเสียงเป็นแบบจำลองที่ใช้ในการจำแนกคำพูดแต่ละคำ ส่วน

แบบจำลองภาษาเป็นแบบจำลองในการจำแนกการเรียงกันของคำพูดแต่ละคำเพื่อสร้างเป็นประโยค ซึ่งในกระบวนการจำแนกข้อมูลดังกล่าวต้องอาศัยข้อมูลจากคลังคำศัพท์ร่วมด้วย ในกระบวนการจำแนกข้อมูลนี้บางครั้งจะถูกเรียกว่าการแปลรหัส ซึ่งจะใช้ลักษณะเด่นของสัญญาณกล้ำมเนื้อของเสียงพูดเป็นอินพุต และอาศัยข้อมูลจากทั้งแบบจำลองเสียง แบบจำลองภาษา และคลังคำศัพท์ เพื่อแปลรหัสออกมาเป็นคำพูดที่ต้องการ

ในงานวิจัยส่วนมากยังอยู่ในช่วงเริ่มต้นทำให้กระบวนการจำแนกข้อมูลยังจำกัดอยู่ที่การจำแนกคำโดด ดังนั้นจึงไม่มีการสร้างแบบจำลองภาษา แบบจำลองภาษาจะใช้ก็ต่อเมื่อต้องการจำแนกคำพูดที่เป็นประโยคหรือคำพูดต่อเนื่อง (continuous speech) ซึ่งมีอยู่เพียงสองกลุ่มงานวิจัยที่พัฒนาระบบนี้เป็นการรู้จำคำพูดต่อเนื่องตามที่ได้กล่าวไว้แล้วข้างต้น คือ กลุ่มเยอรมัน และ สหรัฐอเมริกา กลุ่ม 3

การพัฒนาระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อโดยรู้จำคำพูดต่อเนื่องของกลุ่มวิจัยเยอรมันนั้น ได้ใช้แบบจำลองภาษาในการจำแนกด้วย ซึ่งต่อมามีการพัฒนาเทคนิคในการจำแนกโดยนำแบบจำลองคุณลักษณะการออกเสียงพูด (articulatory features, AF or phonetic features, PF<sup>4</sup>) [7] มาใช้ร่วมกันกับแบบจำลองเสียง เพื่อทำให้ระบบมีความแม่นยำมากขึ้น ซึ่งคุณลักษณะการออกเสียงพูดเป็นคุณลักษณะต่างๆ (features) ที่แสดงถึงที่มาของการเกิดเสียงพูดแต่ละเสียง ซึ่งคุณลักษณะแต่ละเสียงจะอ้างอิงตารางสัทอักษรสากล (international phonetic alphabet, IPA) ซึ่งเป็นตารางที่แสดงถึงสัทอักษร<sup>5</sup> มาตรฐานสำหรับการแทนหน่วยเสียงของเสียงพูดทุกภาษา ซึ่งจะแบ่งตามคุณลักษณะของการออกเสียง คุณลักษณะดังกล่าว ได้แก่ เสียงก้อง (voiced) เสียงไม่ก้อง (unvoiced) เสียงที่เกิดที่ริมฝีปาก (bilabial) เสียงสระปิด (close) เสียงสระหน้า (front) เป็นต้น [67] โดยระบบจะทำการจำแนกว่าสัญญาณอินพุตที่เข้ามาจะเข้าคุณลักษณะดังกล่าวหรือไม่ ซึ่งต่อมากลุ่มงานวิจัยนี้ได้พัฒนาเทคนิค AF หรือ PF ไปเป็นการจัดกลุ่มของคุณลักษณะการออกเสียงพูด (bundled phonetic features, BDPFs) ซึ่งทำให้ปรับปรุงประสิทธิภาพของระบบมากขึ้นอีก [13]

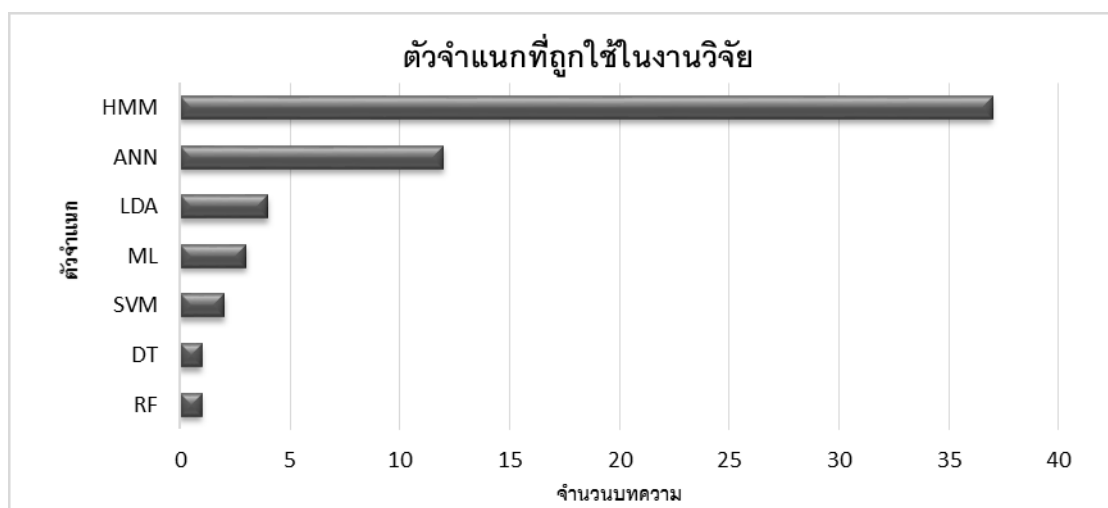
สำหรับตัวจำแนกที่ใช้ในการจำแนกเสียงพูดในระบบรู้จำเสียงพูดแบบเดิมที่ใช้สัญญาณเสียงเป็นอินพุตนั้น จะใช้ตัวจำแนกแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov

<sup>4</sup> ในงานวิจัยช่วงหลังจะเปลี่ยนจาก Articulatory features เป็น Phonetic features

<sup>5</sup> เครื่องหมายที่ชี้แทนระบบตัวเลขหรือเสียง



model, HMM) ซึ่งพบว่ามีประสิทธิภาพในการรู้จำเสียงพูดดีที่สุด เนื่องจากมีความยืดหยุ่นสูง ดังนั้นในงานวิจัยระบบรู้จำการพูดที่ใช้สัญญาณไฟฟ้ากล้ามเนื้อเป็นอินพุต จึงได้นำตัวจำแนกชนิดนี้ไปใช้ในหลายงานวิจัย ซึ่งพบว่าถูกนำมาใช้มากที่สุด ทั้งนี้เนื่องจาก HMM ไม่ส่งผลกระทบต่อการเปลี่ยนแปลงทางเวลา กล่าวคือ เมื่อคำพูดคำเดียวกันใช้เวลาในการพูดสั้นยาวต่างกัน ก็จะไม่ส่งผลต่อประสิทธิภาพในการจำแนก เนื่องจากตัวจำแนกดังกล่าวจะใช้ข้อมูลจากแต่ละลำดับเวลาของสัญญาณ ซึ่งจากการทบทวนวรรณกรรมสามารถสรุปตัวจำแนกที่ถูกใช้ในงานวิจัยดังรูปที่ 1.9 สำหรับตัวจำแนก ANN พบว่ามีการนำมาใช้รองลงมา ส่วนใหญ่เป็นงานวิจัยที่จำแนกคำโดด เนื่องจากตัวจำแนก ANN มีข้อดีในส่วนของคุณภาพไม่ซับซ้อนในการคำนวณ และ ANN เพียง 1 โครงข่ายจะใช้ในการจำแนกเสียงได้ทุกเสียง แต่สำหรับการใช้ตัวจำแนก HMM การจำแนกเสียง 1 เสียงจะต้องใช้ HMM 1 โมเดล ดังนั้นจึงต้องมีการสร้าง HMM หลายโมเดลเพื่อจำแนกเสียงให้ได้ครบทุกเสียง ซึ่งทำให้การทำงานของระบบช้าลง ดังนั้นตัวจำแนก ANN จึงเป็นตัวเลือกสำหรับการจำแนกคำพูดโดดเพียงไม่กี่คำ สำหรับตัวจำแนกอื่นๆ ที่พบในงานวิจัย ได้แก่ ตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้น (LDA) ตัวจำแนกค่าสูงสุดของไลค์ลิฮูด (maximum likelihood, ML) ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีน (SVM) ตัวจำแนกต้นไม้ตัดสินใจ (decision tree, DT) และ ตัวจำแนกแรนดอมฟอเรส (random forest, RF)



รูปที่ 1.9 ตัวจำแนกที่ถูกใช้ในงานวิจัยระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

### 1.2.6 การประยุกต์ใช้งาน

ขั้นตอนต่อไปหลังจากสร้างระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อแล้ว เป็นการนำระบบนี้ไปใช้งาน ซึ่งจะกล่าวถึงการนำระบบนี้ไปใช้งานโดยตรงโดยไม่ผ่านกระบวนการสังเคราะห์เสียงพูด และการนำระบบนี้ไปใช้งานโดยผ่านกระบวนการสังเคราะห์เสียงพูด

การนำระบบรู้จำดังกล่าวไปใช้งานโดยไม่ผ่านกระบวนการสังเคราะห์เสียงพูด เช่น การนำไปใช้สั่งงานการทำงานของคอมพิวเตอร์ด้วยสัญญาณไฟฟ้ากล้ามเนื้อของการพูด การส่งข้อความผ่านโทรศัพท์มือถือด้วยการใช้สัญญาณไฟฟ้ากล้ามเนื้อจากการพูด หรือ การสั่งงานให้หุ่นยนต์ทำงานโดยการใช้น้ำสัญญาณไฟฟ้ากล้ามเนื้อจากการพูด

สำหรับการนำระบบนี้ไปใช้ในการติดต่อสื่อสารระหว่างมนุษย์ จำเป็นต้องมีการสังเคราะห์เสียงซึ่งได้มาจากเอาต์พุตของระบบรู้จำก่อน เพื่อใช้เสียงนั้นในการสื่อสาร เทคนิคในการสังเคราะห์เสียงที่นิยมใช้กันมากที่สุดในปัจจุบันคือ การแปลงข้อความเป็นเสียงพูด

สำหรับการประยุกต์ใช้งานของแต่ละงานวิจัยก็มีหลากหลาย เช่น งานวิจัย [58] ได้ทำการจับสัญญาณไฟฟ้ากล้ามเนื้อบริเวณคอ เพื่อรู้จำเสียงพูดคำโดดที่ใช้ในการสั่งงาน โดยทดสอบกับผู้ใช้ที่สวมชุดดับเพลิงพร้อมหน้ากากออกซิเจน โดยออกเสียงพูดในห้องทดลองที่มีเสียงรบกวน 95 เดซิเบล ซึ่งเมื่อระบบทำการรู้จำสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูดดังกล่าวแล้ว คำพูดที่ผ่านการรู้จำจะไปปรากฏอยู่ที่หน้าจอของโทรศัพท์มือถือที่อยู่ภายนอกห้องทดลอง และข้อความนั้นจะถูกแปลงเป็นเสียงพูดเพื่อใช้สื่อสารระหว่างผู้พูดที่อยู่ในห้องทดลองที่มีเสียงรบกวนอย่างมากกับผู้ฟังซึ่งอยู่ภายนอกห้องทดลอง นอกจากนี้ยังนำคำพูดที่ผ่านการรู้จำดังกล่าวไปควบคุมการเคลื่อนที่ของหุ่นยนต์

นอกจากนี้ยังมีการประยุกต์ใช้ในการค้นหาเว็บ (Web browser interface) [56] โดยการรู้จำตัวเลขเพื่อใช้ในการสั่งงานการเชื่อมโยงหน้าเว็บหนึ่งไปยังอีกหน้าเว็บหนึ่ง (hyperlink) และทำการแปลรหัสตัวเลขให้เป็นตัวอักษรเพื่อใช้ตัวอักษรนั้นเป็นคำที่ใช้ในการค้นหาเว็บ สำหรับการประยุกต์ใช้งานนี้ไม่จำเป็นต้องมีสังเคราะห์เสียงพูด

งานวิจัย [37] ได้นำระบบนี้ไปใช้แก้ปัญหาคาการใช้โทรศัพท์มือถือในสถานที่ที่มีสัญญาณรบกวนอย่างมาก โดยใช้เสียงพูดแบบเงียบจากสัญญาณไฟฟ้ากล้ามเนื้อเป็นอินพุตแทนสัญญาณเสียงพูด ซึ่งงานวิจัยนี้ก็ได้ออกแบบอุปกรณ์ที่ใช้ในการติดต่อสื่อสารด้วยเทคนิคนี้ ซึ่งสามารถทำให้การใช้งานสะดวกขึ้น

### 1.2.7 สรุปการตรวจเอกสาร

ระบบรู้จำการพูดที่ใช้สัญญาณไฟฟ้ากล่อมเนื้อีมีการทำงานวิจัยมากกว่า 10 ปี โดยระยะเริ่มแรกจะเป็นการรู้จำคำโดดที่เป็นคำพูดสั้นๆ เพื่อใช้ในการสั่งงานเฉพาะ โดยจะใช้ข้อมูลจากผู้ทดลองเพียง 1-3 คน โดยจะรู้จำภาษาอังกฤษเป็นส่วนใหญ่ ลักษณะเด่นที่ถูกนำมาใช้ส่วนใหญ่จะเป็นลักษณะเด่นในโดเมนเวลา และตัวจำแนกที่ใช้ในระยะแรกๆ จะใช้ตัวจำแนก ANN เป็นส่วนใหญ่ ซึ่งในระยะหลังพบว่าตัวจำแนก HMM มีความเหมาะสมมากกว่า อย่างไรก็ตามก็ยังไม่มีการวิจัยที่ทำการทดลองเพื่อเปรียบเทียบการใช้ตัวจำแนกทั้งสองชนิดนี้ งานวิจัยในระยะแรกนี้ จะให้ความแม่นยำของระบบที่ต่ำ โดยเฉพาะอย่างยิ่งเมื่อใช้งานกับระบบที่เป็นระบบที่ไม่ขึ้นกับผู้พูด แต่จะให้ความแม่นยำสูงเมื่อระบบเป็นแบบขึ้นกับผู้พูด ในระยะต่อมากจึงมีการทำงานวิจัยเพื่อแก้ไขปัญหาดังกล่าว รวมถึงการเปรียบเทียบเพื่อคัดเลือกลักษณะเด่นที่เหมาะสม และปรับปรุงกระบวนการเตรียมสัญญาณเบื้องต้น เพื่อให้ได้ความแม่นยำของระบบเพิ่มมากขึ้น งานวิจัยส่วนใหญ่จะเน้นกระบวนการปรับปรุงการเตรียมสัญญาณเบื้องต้นและการเปรียบเทียบลักษณะเด่นที่เหมาะสม ซึ่งผลการปรับปรุงของระบบก็พบว่ามีความแม่นยำสูงในการจำแนกคำโดด แต่ความแม่นยำต่ำในการจำแนกคำพูดต่อเนื่อง อย่างไรก็ตามก็ถือว่าปรับปรุงประสิทธิภาพมากขึ้นกว่าเดิมมาก ทั้งนี้เนื่องจากในระยะหลังมีการเพิ่มจำนวนคำศัพท์ในการจำแนก และใช้คลังข้อมูลจากผู้ทดลองจำนวนมาก ทำให้ประสิทธิภาพลดลงจากงานวิจัยในช่วงเริ่มแรก สำหรับการนำระบบไปประยุกต์ใช้งาน ขณะนี้เป็นเพียงการสาธิตการทำงานระบบ ยังไม่มีรายงานว่ารระบบดังกล่าวถูกนำไปใช้งานจริง เนื่องจากระบบดังกล่าวยังอยู่ในช่วงเริ่มแรกของงานวิจัย ซึ่งจะต้องมีการปรับปรุงระบบอีกมากถึงจะนำมาใช้งานจริงได้

จากการทบทวนวรรณกรรมจะเห็นได้ว่าระบบรู้จำเสียงพูดที่ใช้สัญญาณไฟฟ้ากล่อมเนื้อี ควรแก่การพัฒนาาระบบให้สามารถใช้งานได้จริงในอนาคต ถึงแม้ว่างานวิจัยดังกล่าวจะประสบความสำเร็จ แต่ก็ยังจำกัดผู้ใช้ กล่าวคือ ผู้ใช้ต้องใช้งานระบบด้วยการพูดสำหรับภาษาที่ระบบกำหนดไว้ ผู้วิจัยจึงสังเกตเห็นว่าการพัฒนาระบบรู้จำดังกล่าวให้สามารถใช้ได้กับภาษาอื่นด้วยจะเป็นประโยชน์อย่างมาก ดังนั้นจึงได้พัฒนาระบบรู้จำการพูดภาษาไทยที่ใช้สัญญาณไฟฟ้ากล่อมเนื้อีขึ้นสำหรับการใช้งานกับกลุ่มผู้ใช้ที่พูดภาษาไทย เนื่องจากความแตกต่างระหว่างภาษาที่แตกต่างกัน ทำให้การพัฒนาาระบบให้เป็นภาษาไทยนั้นก็เป็นส่วนที่มีความท้าทายอยู่มาก โดยเฉพาะอย่างยิ่งเมื่อต้องการจำแนกในส่วนที่เป็นคำพูดต่อเนื่อง ซึ่งในการสร้างแบบจำลองภาษา จะแตกต่างกับภาษาอื่นๆ อย่างเห็นได้ชัด เนื่องจากโครงสร้างประโยค และการเรียงลำดับ

ของประโยคแตกต่างกัน พร้อมกันนั้นภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์ เสียงวรรณยุกต์ต่างกันความหมายของคำต่างกัน ดังนั้นจึงจำเป็นต้องมีการสร้างระบบรู้จำที่เป็นภาษาไทยขึ้นมา นอกจากส่วนที่เป็นการพัฒนาระบบดังกล่าวมาเป็นภาษาไทย การปรับปรุงโครงสร้างแต่ละส่วนของระบบก็เป็นส่วนสำคัญที่ผู้วิจัยจะทำการค้นคว้าวิจัย เพื่อจะทำให้ระบบมีประสิทธิภาพมากขึ้น

### 1.3 วัตถุประสงค์

- 1) เพื่อสร้างระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ
- 2) เพื่อสังเคราะห์เสียงพูดจากระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ
- 3) เพื่อออกแบบระบบรู้จำการพูดตัวเลขไทย 11 คำ และวรรณยุกต์ไทย 5 เสียง ให้มีความแม่นยำในการจำแนกสูงสุดเท่าที่จะเป็นไปได้
- 4) เพื่อสร้างแบบจำลองการทำงานของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้ต้นแบบของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งเป็นระบบรู้จำคำพูดโดด ซึ่งสามารถนำไปพัฒนาเป็นระบบรู้จำคำพูดแบบต่อเนื่องได้ในอนาคต
- 2) สามารถนำระบบนี้ไปใช้แทนระบบรู้จำการพูดแบบเดิมซึ่งใช้สัญญาณเสียงพูดเป็นอินพุต โดยระบบใหม่นี้สามารถใช้งานได้ดีถึงแม้อยู่ในสภาพแวดล้อมที่มีการรบกวนทางเสียงอย่างมาก
- 3) สามารถนำระบบนี้ไปใช้เป็นเครื่องมือสื่อสารในสถานการณ์ที่การสื่อสารด้วยเสียงพูดไม่มีความปลอดภัย เช่น การสื่อสารด้วยข้อมูลที่เป็นความลับ หรือ ข้อมูลที่มีความเป็นส่วนตัวสูง เป็นต้น
- 4) สามารถนำระบบนี้ไปใช้เป็นเครื่องมือสื่อสารในสถานการณ์ที่การสื่อสารด้วยเสียงพูดเป็นไปได้ยากหรือไม่สามารถกระทำได้ เช่น ขณะที่ผู้พูดสวมหน้ากากออกซิเจน สวมชุดป้องกันสารเคมี สวมชุดดับเพลิง เป็นต้น

5) สามารถนำระบบนี้ไปใช้เป็นเครื่องมือสื่อสารสำหรับผู้ป่วยหรือผู้สูงอายุที่ไม่สามารถพูดออกเสียงได้ หรือ ความสามารถในการพูดน้อยลง เนื่องจากปัญหาทางด้านสุขภาพ

### 1.5 ขอบเขตงานวิจัย

1) สร้างต้นแบบของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อเพื่อรู้จำคำไทยที่เป็นคำโดด และเป็นเสียงพูดของภาษาไทยภาคกลาง ซึ่งผ่านกระบวนการพูด 2 วิธี คือ การพูดแบบออกเสียง และการพูดแบบไม่ออกเสียง โดยระบบจะรู้จำคำพูดของตัวเลขไทย 11 คำ คือ "ศูนย์" "หนึ่ง" "สอง" "สาม" "สี่" "ห้า" "หก" "เจ็ด" "แปด" "เก้า" และ "สิบ" และ รู้จำเสียงวรรณยุกต์ของแต่ละคำพูดด้วย ซึ่งวรรณยุกต์ไทยประกอบด้วย 5 เสียง คือ "สามัญ" "เอก" "โท" "ตรี" และ "จัตวา" โดยการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจากผู้ทดลองหรืออาสาสมัครซึ่งใช้ภาษาไทยเป็นภาษาแม่จำนวนทั้งหมด 10 คน ซึ่งมีอายุระหว่าง 18-50 ปี สำหรับข้อมูลออกหรือเอาต์พุตของระบบส่วนนี้จะแสดงผลเป็นข้อความของคำพูดที่ผ่านการรู้จำ

2) สร้างแบบจำลองการทำงานของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อโดยการประยุกต์ใช้โปรแกรม MATLAB

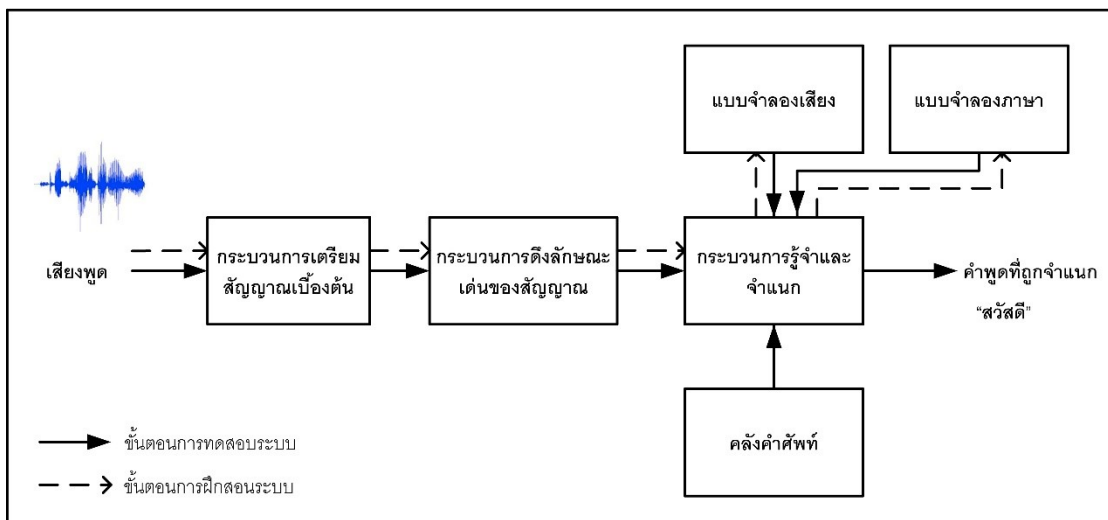
## บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

### 2.1 ระบบรู้จำเสียงพูดอัตโนมัติ

ระบบรู้จำเสียงพูดอัตโนมัติ (automatic speech recognition, ASR) คือ กระบวนการคัดแยกและตัดสินใจข้อมูลทางภาษาแบบอัตโนมัติ โดยข้อมูลทางภาษาที่กล่าวถึงนี้คือ สัญญาณเสียงพูด โดยกระบวนการของระบบนี้จะนำสัญญาณเสียงพูดไปผ่านกระบวนการรู้จำเพื่อให้คอมพิวเตอร์เข้าใจภาษาพูดของมนุษย์ ระบบนี้จะถูกนำไปประยุกต์ใช้ในงานด้านต่างๆ เช่น การควบคุมการทำงานของแป้นพิมพ์ด้วยเสียงเพื่อช่วยให้พิมพ์ง่ายและเร็วขึ้น การควบคุมการทำงานของอุปกรณ์ในรถยนต์โดยใช้คำสั่งเสียง การใช้คำสั่งเสียงเพื่อควบคุมการทำงานของระบบควบคุมการบินในห้องนักบิน การใช้คำสั่งเสียงเพื่อควบคุมการทำงานของโทรศัพท์มือถือ เป็นต้น

โครงสร้างพื้นฐานของระบบรู้จำเสียงพูดอัตโนมัติ ประกอบด้วยโครงสร้างหลักดังนี้ (ดังรูปที่ 2.1)

1) กระบวนการเตรียมสัญญาณเบื้องต้น กระบวนการนี้ประกอบด้วย การกำจัดสัญญาณรบกวน (noise reduction) การแบ่งช่วงสัญญาณ (frame blocking) การเน้นข้อมูลส่วนกลางเฟรมของข้อมูลที่จะนำไปวิเคราะห์ (windowing) และการหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณเสียงพูด



รูปที่ 2.1 โครงสร้างของระบบรู้จำเสียงพูดอัตโนมัติ

2) กระบวนการสกัดลักษณะเด่นของสัญญาณ เป็นการหาตัวแทนของเสียงพูด ซึ่งจะอยู่ในรูปของเวกเตอร์ กระบวนการดังกล่าวจะเป็นการลดจำนวนมิติของข้อมูลของเสียงพูดลง ด้วย โดยข้อมูลของเสียงพูดนั้นๆ จะแทนด้วยเวกเตอร์ค่าหนึ่ง โดยทั่วไปสำหรับระบบรู้จำเสียงพูดอัตโนมัติจะใช้วิธีสกัดลักษณะเด่นของข้อมูลแบบ MFCC

3) กระบวนการจำแนกรูปแบบ จะแบ่งเป็นสองขั้นตอน คือ กระบวนการฝึกสอนระบบ และ กระบวนการทดสอบระบบ เริ่มต้นระบบจะนำลักษณะเด่นของสัญญาณเสียงเข้าสู่กระบวนการฝึกสอนเพื่อสร้างแบบจำลองเสียงและแบบจำลองภาษา โดยที่แบบจำลองเสียงเป็นแบบจำลองที่ใช้ในการจำแนกเสียงแต่ละเสียง ส่วนแบบจำลองภาษาเป็นแบบจำลองที่ใช้ในการจำแนกการเรียงกันของแต่ละเสียงเพื่อสร้างเป็นประโยค สำหรับกระบวนการทดสอบระบบ เป็นการนำลักษณะเด่นของสัญญาณเสียงไปเปรียบเทียบกับแบบจำลองเสียงและแบบจำลองภาษาที่ถูกสร้างขึ้นจากกระบวนการเรียนรู้ เพื่อให้ระบบตัดสินใจว่าสัญญาณเสียงพูดที่เข้ามาเป็นเสียงพูดใด ซึ่งในกระบวนการจำแนกข้อมูลดังกล่าวต้องอาศัยข้อมูลจากคลังคำศัพท์ร่วมด้วย

## 2.2 ระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

สำหรับในงานวิจัยนี้เป็นการสร้างระบบรู้จำการพูดอัตโนมัติภาษาไทย โดยที่เปลี่ยนข้อมูลทางภาษาในการรู้จำจากสัญญาณเสียงไปเป็นสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งโครงสร้างการทำงานของระบบจะมีความคล้ายคลึงกับระบบรู้จำเสียงพูดแบบเดิมที่ใช้สัญญาณเสียงพูดในการรู้จำ แต่จะแตกต่างกันในส่วนของการรายละเอียด ซึ่งผู้วิจัยจะทำการศึกษาเพื่อปรับปรุงโครงสร้างในแต่ละส่วนของระบบ เพื่อให้ระบบมีประสิทธิภาพในการจำแนกที่ดีที่สุด และจะแปลงเอาต์พุตที่ได้จากระบบรู้จำไปเป็นเสียงพูดที่ใช้ในการสื่อสาร ซึ่งเรียกว่ากระบวนการสังเคราะห์เสียง

โครงสร้างของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ประกอบด้วย 2 ระบบหลัก (รูปที่ 2.2) คือ ระบบรู้จำการพูด และระบบสังเคราะห์เสียงพูด สำหรับระบบรู้จำการพูดนั้นประกอบด้วยโครงสร้างหลัก 3 ส่วน คือ กระบวนการเตรียมสัญญาณหรือประมวลผลสัญญาณเบื้องต้น กระบวนการสกัดลักษณะเด่นของสัญญาณและการคัดเลือกลักษณะเด่น และกระบวนการจำแนกรูปแบบ หลังจากนั้นก็จะเข้าสู่กระบวนการสังเคราะห์เสียงพูด ซึ่งจะแปลงข้อความจากคำพูดที่ถูกต้องจากระบบรู้จำการพูดดังกล่าวไปเป็นเสียงพูดเพื่อใช้ในการสื่อสาร

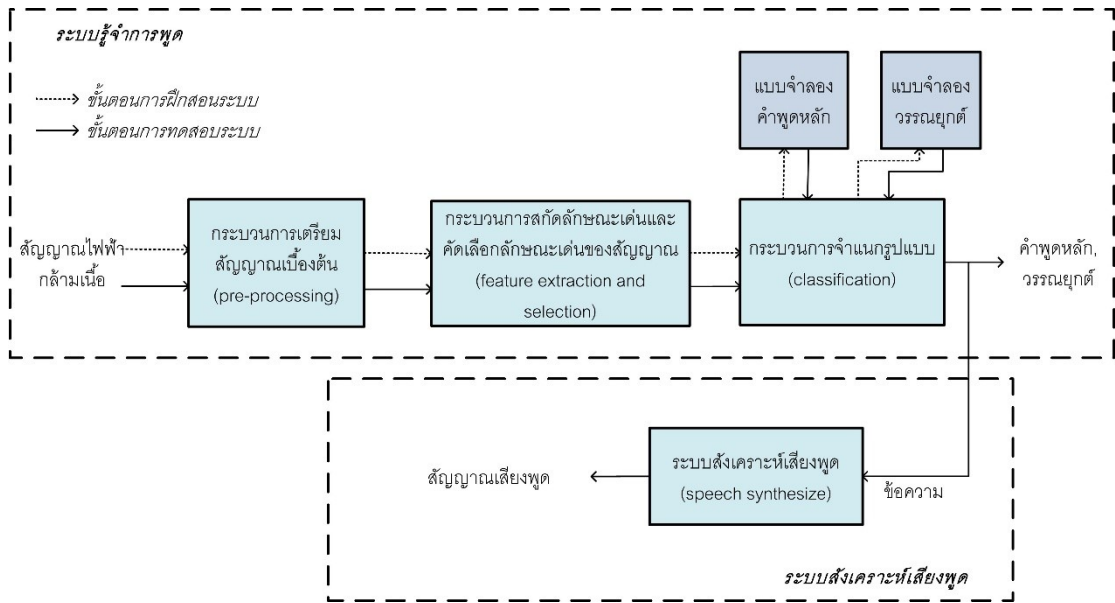
สำหรับกระบวนการเตรียมสัญญาณเบื้องต้นและกระบวนการสกัดลักษณะเด่นของสัญญาณจะคล้ายคลึงกับระบบรู้จำการพูดแบบเดิม สิ่งที่แตกต่างที่สุดคือ กระบวนการจำแนกรูปแบบ กระบวนการจำแนกข้อมูลสำหรับระบบรู้จำการพูดภาษาไทยที่ใช้สัญญาณไฟฟ้ากล้ามเนื้อ จะจำแนกโดยอาศัยแบบจำลองคำพูดหลัก (word model) และแบบจำลองเสียงวรรณยุกต์ (tone model) ที่ถูกสร้างขึ้นมาจากกระบวนการฝึกสอนระบบ ซึ่งในที่นี้เป็นารรู้จำคำโดดเพียง 11 คำ และวรรณยุกต์ 5 เสียง ดังนั้นจึงไม่จำเป็นต้องมีแบบจำลองภาษา ซึ่งเป็นแบบจำลองที่ใช้ในการจำแนกประโยค

สำหรับการทำงานของระบบ จะแบ่งออกเป็น 2 ขั้นตอน

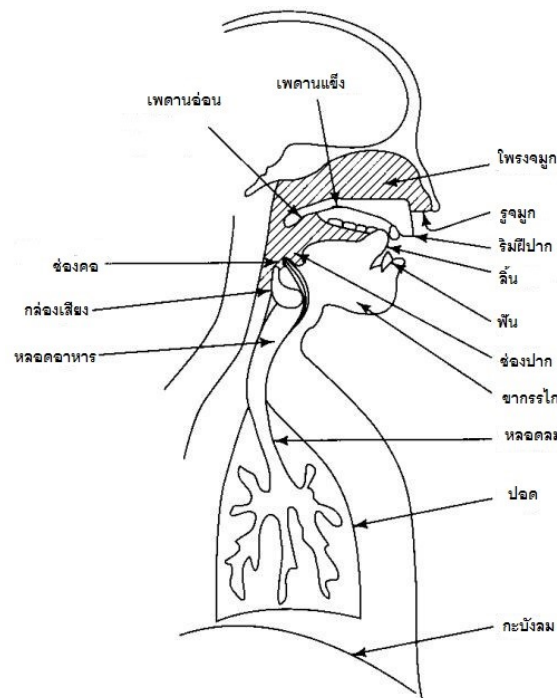
1) กระบวนการฝึกสอนระบบ เป็นขั้นตอนฝึกสอนให้ระบบเกิดการเรียนรู้และเข้าใจว่าสัญญาณไฟฟ้ากล้ามเนื้อที่ป้อนเข้าไปในระบบผ่านกระบวนการเตรียมสัญญาณเบื้องต้นและกระบวนการสกัดลักษณะเด่นจะถูกจำแนกไปเป็นคำใด โดยจะสร้างแบบจำลองที่ใช้ในการจดจำคำนั้นๆ (word model) และสร้างแบบจำลองเสียงวรรณยุกต์ (tone model) สำหรับใช้จำแนกเสียงวรรณยุกต์ไทย 5 เสียง ซึ่งเป็นวรรณยุกต์ของคำนั้นๆ

2) กระบวนการทดสอบระบบ เป็นขั้นตอนการใช้งานระบบเพื่อรู้จำการพูด หลังจากกระบวนการฝึกสอนระบบ ซึ่งจะนำสัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากการพูด ไปผ่านกระบวนการเตรียมสัญญาณเบื้องต้นและกระบวนการสกัดลักษณะเด่น และจะเข้าสู่ตัวจำแนกเพื่อจำแนกคำพูดที่ถูกต้องออกมา โดยอาศัยแบบจำลองคำพูดหลักและแบบจำลองเสียงวรรณยุกต์ โดยเอาต์พุตของระบบรู้จำจะแสดงผลในรูปแบบของข้อความ โดยเอาต์พุตดังกล่าว คือ คำพูดที่ถูกจำแนกออกมาพร้อมทั้งวรรณยุกต์ของคำนั้นๆ ต่อมาข้อความของคำพูดหลักจะถูกแปลงไปเป็นเสียงพูดโดยผ่านกระบวนการสังเคราะห์เสียง





รูปที่ 2.2 โครงสร้างของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากลั้วมเนื้อ



รูปที่ 2.3 อวัยวะในการสร้างเสียงพูด [68]

## 2.3 กลไกการสร้างเสียงพูด

กลไกการสร้างเสียงพูดเกิดจากกระบวนการหลัก 2 กระบวนการ [68] คือ กระบวนการคิดคำพูดซึ่งเกิดขึ้นที่สมอง และ กระบวนการบังคับอวัยวะที่ใช้ในการออกเสียงให้เปล่งเสียงออกมาตามที่สมองสั่ง อวัยวะเริ่มแรกในการออกเสียงพูด คือ ปอดและกะบังลม กระบังลมจะช่วยให้มีการหายใจเข้าออกเพื่อนำอากาศเข้าและออกจากปอด ปอดคือต้นกำเนิดของพลังงานที่ทำให้เกิดเสียง มนุษย์ใช้ลมจากปอดทำให้เกิดเสียงพูด โดยลมถูกบังคับให้ผ่าน อวัยวะต่างๆ ตั้งแต่หลอดลม กล่องเสียง เส้นเสียง คอหอย แล้วผ่านออกมาทางช่องปากหรือช่องจมูก ซึ่งอวัยวะทั้งหมดในช่องปากเป็นส่วนที่บังคับให้เสียงมีลักษณะแตกต่างกัน ซึ่งการที่อวัยวะเหล่านี้จะทำงานได้จะต้องอาศัยการสั่งงานของสมองเพื่อส่งกระแสประสาทไปควบคุมการทำงานของกล้ามเนื้อต่างๆ ที่จะไปควบคุมการทำงานของอวัยวะที่เกี่ยวข้องกับการพูดนั้น ซึ่งงานวิจัยนี้จะใช้สัญญาณไฟฟ้ากล้ามเนื้อจากกล้ามเนื้อดังกล่าวมาใช้ในการจดจำคำพูด

นอกจากนี้สามารถแบ่งอวัยวะที่ใช้ในการออกเสียง แบ่งเป็น 2 กลุ่ม [68] คือ อวัยวะที่เป็นส่วนกระทำอาการ หรือเรียกอีกอย่างหนึ่งว่า "กรณี" และ อวัยวะที่เป็นตำแหน่งที่เกิดเสียงต่างๆ หรือเรียกว่า "ฐาน" อวัยวะที่เป็นส่วนกระทำอาการจะทำหน้าที่เคลื่อนไหวเพื่อผลักดันไปยังที่ต่างๆ อวัยวะส่วนนี้ได้แก่ ลิ้น ส่วนอวัยวะที่เป็นตำแหน่งที่เกิดเสียงต่างๆ ได้แก่ ริมฝีปาก ฟัน ปุ่มเหงือก เพดานแข็ง เพดานอ่อน ลิ้นไก่ ช่องจมูก ส่วนต่างๆ ของลิ้นเช่น ปลายลิ้น หน้าลิ้น และหลังลิ้น ลิ้นปิดกล่องเสียง โฟรงคอ และเส้นเสียง เป็นต้น ซึ่งการทำงานของอวัยวะต่างๆ เหล่านี้จะทำให้เกิดเสียงที่แตกต่างกัน การแบ่งอวัยวะในการออกเสียงในรูปแบบของฐานและกรณี จะใช้ในการกำหนดสัทอักษรสากล ซึ่งจะกล่าวในหัวข้อต่อไป

## 2.4 กล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูด

กล้ามเนื้อต่างๆ ที่จะไปควบคุมการเคลื่อนที่ของอวัยวะที่เกี่ยวข้องกับการออกเสียงพูดสามารถแบ่งออกเป็น 7 กลุ่ม [64] ดังนี้

1) กล้ามเนื้อสำหรับการหายใจเพื่อควบคุมการไหลของอากาศสำหรับการพูด ซึ่งกล้ามเนื้อส่วนนี้ช่วยในการทำงานของกระบังลมและปอด

2) กล้ามเนื้อสำหรับการเคลื่อนที่ริมฝีปาก

- 3) กล้ามเนื้อสำหรับการเคลื่อนที่ของกระดูกขากรรไกรล่าง
- 4) กล้ามเนื้อสำหรับการเคลื่อนที่ลิ้น
- 5) กล้ามเนื้อในส่วนของเพดานอ่อน
- 6) กล้ามเนื้อในส่วนของคอหอย
- 7) กล้ามเนื้อสำหรับการทำงานของกล่องเสียง ซึ่งกล้ามเนื้อส่วนนี้ช่วยในการควบคุมการเปิดปิดของกล่องเสียง

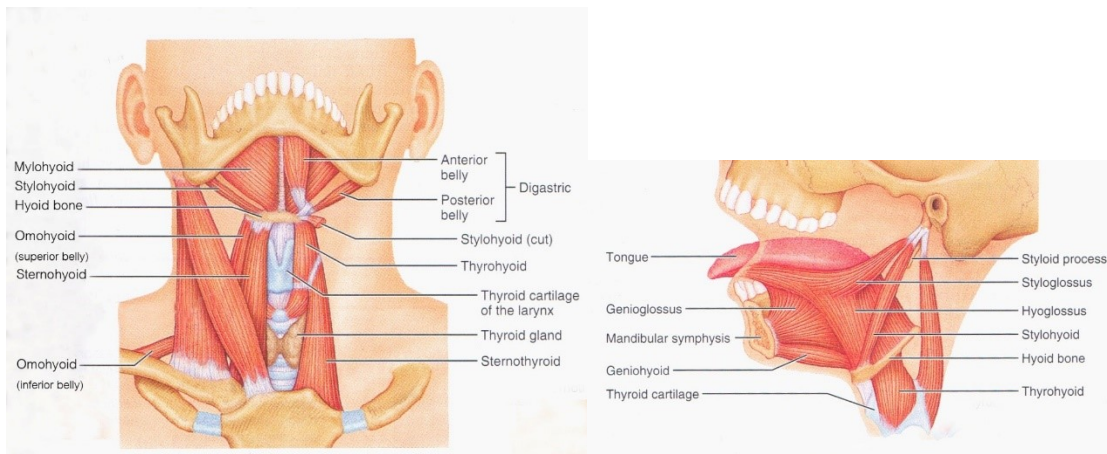
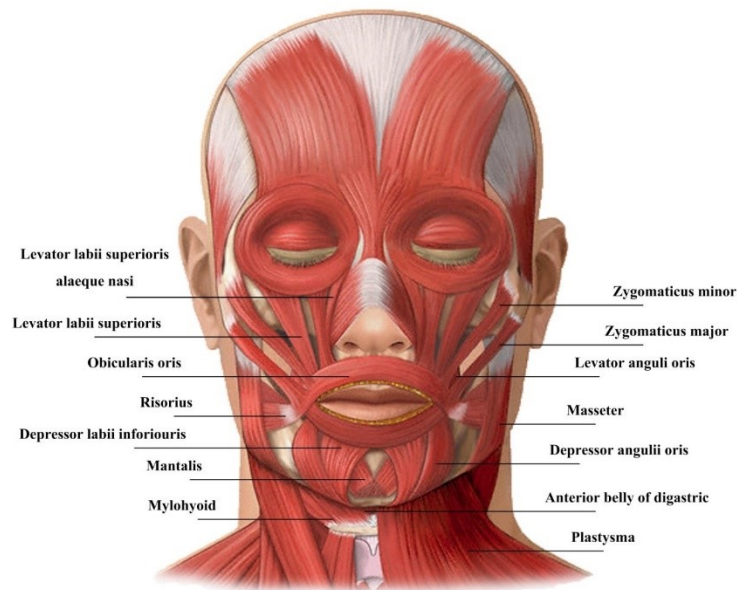
**ตารางที่ 2.1** รายชื่อและหน้าที่ของกล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูดทั้ง 7 กลุ่ม [64]

กลุ่ม	ชื่อกล้ามเนื้อ	หน้าที่
1	Diaphragm	ขยายทรวงอกเมื่อมีการหายใจเข้า
	External Intercostals	ทำงานร่วมกับกล้ามเนื้อกะบังลมเพื่อยกซี่โครงขึ้นสำหรับการสูดหายใจเข้า
	Pectoralis Major	ยกโพรงซี่โครงและกระดูกสันอก
	Pectoralis Minor	ยกโครงกระดูกขึ้นเป็นแรงในการสูดหายใจเข้า
	Internal Intercostals	ดึงโพรงซี่โครงลง
	Internal Obliques	กดช่องท้อง
	External Obliques	กดช่องท้องและดึงซี่โครงส่วนล่างให้ต่ำลง
	Rectus Abdominis	กล้ามเนื้อที่ช่วยเป็นแรงในการหายใจออก
Transversus Thoracic	กดซี่โครงสำหรับการหายใจออก	
2	Orbicularis Oris	การปิดริมฝีปาก และการห่อริมฝีปากเป็นวงกลม
	Levator Labii Superioris	ยกริมฝีปากบนขึ้น
	Levator Labii Superioris	ยกปีกจมูกหรือขยายรูจมูกพร้อมยกส่วนกลางของริมฝีปากบนขึ้น
	Alaeque Nasi	ปีกจมูก
	Zygomaticus Minor	ยกมุมปากด้านข้างขึ้น พร้อมกับยกริมฝีปากด้านบนขึ้น
	Depressor Labii Inferioris	ดึงริมฝีปากล่างลงและออกไปทางด้านข้างหรือใช้สำหรับการบ้วนปาก
	Mentalis	ยื่นริมฝีปากและทำปากเป็นวงกลม ทำงานร่วมกับ Orbicularis Oris

กลุ่ม	ชื่อกล้ามเนื้อ	หน้าที่
	Buccinator	ดึงผิวหนังบริเวณคางขึ้นไป ในขณะที่เดียวกันกับการยื่นริมฝีปากกลางและหันออกด้านใน ดึงมุมปากด้านข้างทำให้ร่องแก้มนูน
	Risorius	ดึงมุมปากด้านข้างช่วยในการขยายริมฝีปาก
	Zygomaticus Major	ดึงมุมปากขึ้นไปข้างบนและออกไปด้านข้าง ทำงานร่วมกับ Levator Anguli Oris
	Levator Anguli Oris	ดึงมุมปากขึ้นข้างบน
	Depressor Anguli Oris	ดึงมุมปากลงต่ำ ทำงานร่วมกับ Depressor Labii Inferioris
	Platysma	ดึงมุมปากลงไปที่ด้านล่าง และออกด้านข้าง กดกระดูกขากรรไกรล่าง ดึงริมฝีปากกลางกลับและลง ทำงานร่วมกับ Depressor Anguli Oris และ Depressor Labii Inferioris
3	Masseter	ยกและดึงมุมของขากรรไกรล่างไปด้านหน้า
	Medial Pterygoid	ยกและยื่นกระดูกขากรรไกรล่าง ทำงานร่วมกับ Masseter และ Temporalis
	Temporalis	ยกกระดูกขากรรไกรล่าง
	Anterior Belly of Digastric	ดึงกระดูกโคนลิ้นรูปตัวยูขึ้น (hybrid bone) และไปข้างหน้า การดึงกระดูกโคนลิ้นเป็นการดึงกล่องเสียงด้วย จึงทำให้เส้นเสียงมีการขยายและเพิ่มระดับเสียง
	Geniohyoid	ดึงกระดูกโคนลิ้นรูปตัวยูขึ้นและไปข้างหน้า
	Mylohyoid	ยกกระดูกโคนลิ้น
	Lateral Pterygoid	ยื่นกระดูกขากรรไกรล่างไปด้านข้างและลงล่าง
4	Superior Longitudinal	ดึงปลายลิ้นลงล่าง และ ยกปลายลิ้นไปที่เพดาน
	Inferior Longitudinal	ดึงปลายลิ้นลงล่าง และ หดกลับปลายลิ้น
	Transverse	ดึงมุมลิ้นไปข้างหน้า ทำให้ลิ้นแคบหรือยืดออก
	Vertical	ทำให้ลิ้นแบนราบและออกไปด้านข้าง
	Genioglossus	ยื่นลิ้นไปข้างหน้า และ ยกกระดูกโคนลิ้นรูปตัวยู

กลุ่ม	ชื่อกล้ามเนื้อ	หน้าที่
	Hyoglossus	กดลิ้น และดึงไปข้างๆ และ ลงล่าง
	Styloglossus	ยกและดึงหลังลิ้น
5	Levator Palatini	ยกเพดานอ่อนและดึงส่วนหลังสำหรับการออกเสียงไม่ขึ้นจมูก
	Musculus Uvulae	ยกเพดานอ่อนและลิ้นไก่ ช่วยปิดโพรงจมูก
	Tensor Palatini	ยืดขยายและตรึงแน่นเพดานอ่อน ช่วยปิดโพรงจมูก
	Palatoglossus	ยกหลังลิ้น ขณะลิ้นอยู่กับที่ กล้ามเนื้อนี้จะเป็นตัวกดเพดานอ่อน
	Palatopharyngeus	ทำให้เพดานอ่อนต่ำลง ขณะที่กระดูกอ่อนไฮรอยด์และผนังคอหอยอยู่กับที่ เมื่อเพดานอ่อนอยู่กับที่ กระดูกอ่อนไฮรอยด์จะยกขึ้นซึ่งเกิดขึ้นเวลากลิ้น
6	Superior Pharyngeal Constrictor	ทำให้ผนังคอหอยด้านบนแคบลง
	Medial Pharyngeal Constrictor	หดเกร็งคอหอยระหว่างการกลืน
	Inferior Pharyngeal Constrictor	หดเกร็งคอหอยส่วนล่างสำหรับการกลืน
7	Anterior Belly of Digastric	ยกกล่องเสียง
	Posterior Belly of Digastric	ดึงกระดูกโคนลิ้นรูปตัวยูและกล่องเสียง ช่วยในการนำลิ้นไปในตำแหน่งการออกเสียงที่เพดานอ่อน
	Genioglossus	ยกกล่องเสียง
	Hyoglossus	ยกกล่องเสียง
	Mylohyoid	ยกกล่องเสียง
	Medial Pharyngeal Constrictor	ยกกล่องเสียง
	Stylohyoid	ยกและดึงกระดูกโคนลิ้นและกล่องเสียง
	Omohyoid	ทำให้กระดูกโคนลิ้นและกล่องเสียงต่ำลง

กลุ่ม	ชื่อกล้ามเนื้อ	หน้าที่
	Sternohyoid	ดึงกระดูกโคนลิ้นและกล่องเสียงลงด้านล่างและไปข้างหน้า
	Thyrohyoid	การหดเกร็งกล้ามเนื้อส่วนนี้ จะไปลดระยะทางระหว่างกระดูกอ่อนไทรอยด์และกระดูกโคนลิ้น กล้ามเนื้อนี้จะไปกดกระดูกโคนลิ้นเมื่อกระดูกอ่อนไทรอยด์อยู่กับที่ แต่ถ้ากระดูกโคนลิ้นอยู่กับที่ กล้ามเนื้อจะไปยกกระดูกอ่อนไทรอยด์ขึ้น ทำให้เพิ่มระดับเสียง
	Aryepiglottic	ดึงกลับลิ้นเพื่อปิดปากทางเข้าของกล่องเสียง
	Thyroepiglottic	กดลิ้นปิดกล่องเสียง
	Posterior Cricoarytenoid	กางและหมุนกระดูกอ่อนไทรอยด์ออกด้านข้าง ดึงให้สายเสียงกางออกและไปด้านหน้า และ เปิดช่องว่างระหว่างสายเสียง
	Lateral Cricoarytenoid	หมุนกระดูกอ่อน Arytenoid เข้าข้างในและดึงให้สายเสียงหุบเข้าและไปด้านหลัง เมื่อสายเสียงหุบเข้าแล้ว การเพิ่มความตึงของกล้ามเนื้อนี้จะทำให้เพิ่มระดับเสียง และปิดช่องว่างระหว่างสายเสียง
	Interarytenoids	ดึงกระดูกอ่อน Arytenoid เข้าหากันและยกเล็กน้อย หุบสายเสียงโดยนำยอดปลายของกระดูกอ่อน Arytenoid เข้าหากัน
	Cricothyroid	ทำให้สายเสียงตึงขึ้นและยาวขึ้นเพื่อเพิ่มระดับเสียง
	Vocalis	ช่วยในการหุบสายเสียงระหว่างการพูด



รูปที่ 2.4 กล้ามเนื้อที่เกี่ยวข้องสำหรับกลไกการออกเสียงพูด [69], [70]

## 2.5 ระบบเสียงพูดในภาษาไทย

ตามหลักสัทศาสตร์สากล หน่วยเสียง เป็นลักษณะของเสียงที่เล็กที่สุดของภาษา ซึ่งสำหรับหน่วยเสียงของภาษาไทยแบ่งออกเป็น 3 ประเภท [68] คือ หน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ หน่วยเสียงพยัญชนะของไทยมีทั้งหมด 21 หน่วยเสียง ซึ่งแสดงดังตารางที่ 2.2 ซึ่งแบ่งออกเป็น พยัญชนะต้นเดี่ยว พยัญชนะต้นควบ และพยัญชนะท้าย ส่วนหน่วยเสียงสระ มีทั้งหมด 18 หน่วยเสียงดังตารางที่ 2.3 ซึ่งแบ่งออกเป็นสระเดี่ยว สระผสม

สระเสียงสั้น และสระเสียงยาว และหน่วยเสียงวรรณยุกต์ซึ่งมีทั้งหมด 5 หน่วยเสียง คือ เสียง สามัญ เอก โท ตรี จัตวา

จากตารางที่ 2.2 และ ตารางที่ 2.3 แสดงสัญลักษณ์หน่วยเสียงพยัญชนะและสระ ในภาษาไทย ซึ่งถูกเขียนด้วยสัญลักษณ์ตามสัทอักษรสากล สัทอักษรสากลเป็นสัญกรณ์มาตรฐาน สำหรับใช้แทนเสียงพูดในทุกภาษา สัทอักษรแต่ละตัวจะใช้แทนหน่วยเสียงแต่ละเสียง การแบ่ง หน่วยเสียงพยัญชนะจะแบ่งตามประเภทการออกเสียง และ ฐานกรณ์ (อวัยวะที่ใช้ในการออก เสียงพูด) สำหรับการแบ่งหน่วยเสียงสระจะแบ่งตามลักษณะของริมฝีปาก ส่วนของลิ้น และ ระดับ ลิ้นขณะออกเสียง สำหรับเสียงวรรณยุกต์ตามหลักสัทศาสตร์จะแทนเสียงวรรณยุกต์ สามัญ เอก โท ตรี และ จัตวา โดยใช้สัญลักษณ์ /ē/ /è/ /ê/ /é/ และ /ě/ ตามลำดับ เช่น คำว่า "นา" ซึ่งเป็นเสียง วรรณยุกต์สามัญ แทนด้วย สัญลักษณ์แสดงหน่วยเสียง คือ "/nāː/" (หน่วยเสียงพยัญชนะ คือ /n/, หน่วยเสียงสระ คือ /aː/ และ หน่วยเสียงวรรณยุกต์ คือ /ē/) เป็นต้น

**ตารางที่ 2.2** หน่วยเสียงพยัญชนะภาษาไทยซึ่งแทนด้วยสัญลักษณ์ของแต่ละหน่วยเสียงตาม IPA [68]

ประเภท การออกเสียง	ฐานกรณ์	ริมฝีปาก	ริม	พินและ	เพดาน	เพดาน	ช่อง
		กับพิน	ฝีปาก	ปุ่ม	แข็ง	อ่อน	ระหว่าง เส้นเสียง
ไม่ก้อง		p (ป)		t (ต)	c (จ)	k (ก)	ʔ (อ)
ระเบิด	ไม่ก้อง มีลม	ph (พ)		th (ท)	ch (ช)	kh (ค)	
	ก้อง	b (บ)		d (ด)			
นาสิก		m (ม)		n (น)		ŋ (ง)	
รัว				r (ร)			
เสียดแทรก			f (ฟ)	s (ซ)			h (ฮ)
เสียงเลื่อนหรือครึ่งสระ		w (ว)			j (ย)		
ข้างลิ้น				l (ล)			



ตารางที่ 2.3 หน่วยเสียงสระภาษาไทย ซึ่งแทนด้วยสัญลักษณ์ของแต่ละหน่วยเสียงตาม IPA [68]

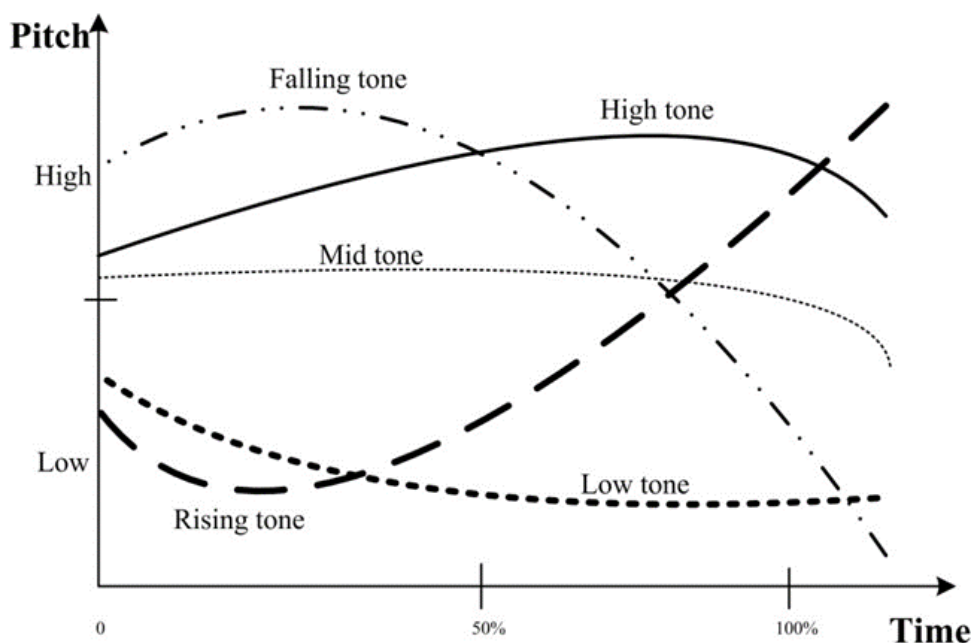
ริมฝีปาก	รี		ห่อ	
	ส่วนของลิ้น	หน้า	หลังค่อนไปทางกลาง	หลัง
ระดับลิ้น				
สูง	i (อิ), i: (อี)		ɯ (อึ), ɯ: (อือ)	u (อุ), u: (อู)
กลางสูง	e (เอะ), e: (เอ)		ɤ (เออะ), ɤ: (เออ)	ɔ (เอาะ), ɔ: (ออ)
กลางต่ำ	ɛ (แอะ), ɛ: (แอ)			o (โอะ), o: (โอ)
ต่ำ			a (อะ), a: (อา)	

สำหรับงานวิจัยนี้จะทำการจำแนกเสียงวรรณยุกต์ด้วย ซึ่งเป็นเอกลักษณ์เฉพาะของภาษาไทย การศึกษาที่มาของการเกิดเสียงวรรณยุกต์จึงจำเป็นเพื่อใช้ในการหาตำแหน่งที่เหมาะสมของกล้ามเนื้อสำหรับการจำแนกเสียงวรรณยุกต์

วรรณยุกต์ คือ การเปลี่ยนแปลงของระดับเสียง โดยที่ผลของการเปลี่ยนแปลงแต่ละแบบนั้นจะทำให้แต่ละพยางค์หรือคำมีความหมายแตกต่างกันอย่างสิ้นเชิง แตกต่างกับทำนอง ซึ่งเป็นการเปลี่ยนแปลงของระดับเสียงเช่นเดียวกัน แต่การเปลี่ยนแปลงนั้นไม่ส่งผลให้แต่ละพยางค์หรือคำมีความหมายแตกต่างกัน

รูปแบบการเปลี่ยนแปลงของระดับเสียงจะทำให้เกิดวรรณยุกต์ที่ต่างกัน 5 เสียงได้แก่

- 1) เสียงสามัญ เป็นการเปลี่ยนแปลงระดับเสียงจากเสียงกึ่งสูง ไปเป็นระดับเสียงกลาง
- 2) เสียงเอก เป็นการเปลี่ยนแปลงระดับเสียงจากเสียงกึ่งต่ำ ไปเป็นระดับเสียงต่ำ
- 3) เสียงโท เป็นการเปลี่ยนแปลงระดับเสียงจากเสียงสูง ไปเป็นระดับเสียงต่ำ
- 4) เสียงตรี เป็นการเปลี่ยนแปลงระดับเสียงจากเสียงกึ่งสูง ไปเป็นระดับเสียงสูง
- 5) เสียงจัตวา เป็นการเปลี่ยนแปลงระดับเสียงจากเสียงกึ่งต่ำ ไปเป็นระดับเสียงต่ำ แล้วไปเป็นระดับเสียงกึ่งสูง



รูปที่ 2.5 กราฟแสดงรูปแบบการเปลี่ยนแปลงระดับเสียงซึ่งทำให้เกิดเสียงวรรณยุกต์ 5 เสียง

สำหรับการเปลี่ยนแปลงของระดับเสียงนั้น เกิดจาก 2 กระบวนการหลัก กระบวนการแรกคือ การสั่นของเส้นเสียง ซึ่งเกิดจากการยืดตัวและเกร็งตัวของเส้นเสียง [71] ทำให้เส้นเสียงมีการเปิดปิด ซึ่งการเปิดปิดของเส้นเสียงจะเกิดขึ้นหลายๆ ครั้ง ซึ่งใช้เวลาน้อยมากในการเปิดปิดแต่ละครั้ง จึงเรียกได้ว่าเป็นการสั่นของเส้นเสียง ซึ่งการที่เส้นเสียงสั่นด้วยความถี่สูงจะทำให้เสียงนั้นมีระดับเสียงสูง และถ้าสั่นด้วยความถี่น้อยจะทำให้เสียงนั้นมีระดับเสียงต่ำ นอกจากนี้ การเปลี่ยนแปลงความตึงตัวของเส้นเสียง เป็นอีกกระบวนการที่ทำให้เกิดการเปลี่ยนแปลงระดับเสียง โดยถ้ามีการเปลี่ยนแปลงความตึงมากจะทำให้ระดับเสียงจะสูง กระบวนการเหล่านี้เกิดจากการทำงานร่วมกันของกล้ามเนื้อบริเวณคอ ได้แก่ Cricothyroid Sternothyroid Sternohyoid Omohyoid และ Thyrohyoid ซึ่งกล้ามเนื้อเหล่านี้จะทำให้เกิดการยืดตัวเกร็งตัวของเส้นเสียง

## 2.6 สัญญาณไฟฟ้ากล้ามเนื้อ

สัญญาณไฟฟ้ากล้ามเนื้อเกิดจากการที่เส้นใยมัดกล้ามเนื้อถูกกระตุ้นโดยระบบควบคุมเส้นประสาทส่วนกลาง ซึ่งจะเกิดกระแสไฟฟ้าขนาดเล็กโดยอยู่ในรูปของการไหลของไอออน กระแสไฟฟ้าเหล่านี้จะเคลื่อนที่ไปยังเนื้อเยื่อของร่างกายผ่านความต้านทานไฟฟ้าภายในเนื้อเยื่อ ทำให้เกิดสนามไฟฟ้าและความต่างศักย์ไฟฟ้าขึ้นบริเวณผิวของร่างกาย [65] สามารถจับ

สัญญาณเหล่านี้ได้ด้วยอิเล็กทรอนิกส์ สัญญาณไฟฟ้ากล้ามเนื้อที่เกิดขึ้นนี้จะมีขนาดเล็กมาก โดยทั่วไปจะมีขนาดแอมพลิจูด 0-10 มิลลิโวลต์ยอดถึงยอด (mVp-p) หรือ 0-1.5 มิลลิโวลต์อาร์เอ็มเอส (mVrms) [65] ซึ่งจำเป็นจะต้องมีการขยายสัญญาณในกระบวนการประมวลผลสัญญาณเบื้องต้น สัญญาณไฟฟ้ากล้ามเนื้อเป็นสัญญาณแบบสุ่มในธรรมชาติ มีความถี่อยู่ในช่วง 0-500 เฮิรตซ์ โดยความถี่หลักอยู่ในช่วง 50-150 เฮิรตซ์ [65]

สัญญาณไฟฟ้ากล้ามเนื้อที่ตรวจวัดได้นั้นจะมีสัญญาณรบกวนปนอยู่มาก จึงจำเป็นต้องมีการกรองสัญญาณรบกวนหรือสัญญาณที่ไม่ต้องการออกไปด้วยตัวกรองความถี่ชนิดต่างๆ ตามความเหมาะสม ซึ่งสำหรับสัญญาณไฟฟ้ากล้ามเนื้อโดยทั่วไป เมื่อมีการขยายสัญญาณไฟฟ้ากล้ามเนื้อให้มีแอมพลิจูดสูงขึ้นแล้ว จะนำสัญญาณมาผ่านตัวกรองความถี่ โดยผ่านตัวกรองความถี่ต่ำผ่านเพื่อเอาสัญญาณที่ความถี่มากกว่า 500 เฮิรตซ์ หรือ 600 เฮิรตซ์ออกไป จากนั้นจะผ่านตัวกรองความถี่สูงผ่านเพื่อกำจัดสัญญาณที่ความถี่ต่ำกว่า 10-20 เฮิรตซ์ ซึ่งเป็นความถี่ของสัญญาณรบกวนจากการเคลื่อนไหวที่

## 2.7 การคำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ

อินพุตของระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ คือ สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งการใช้สัญญาณไฟฟ้ากล้ามเนื้อเป็นอินพุตให้กับตัวจำแนกโดยตรงโดยไม่ผ่านกระบวนการลดมิติของข้อมูลไม่เหมาะสมอย่างยิ่ง เนื่องจากเป็นการบ่อนข้อมูลจำนวนมากให้กับระบบ ซึ่งเป็นข้อมูลที่มากเกินไปและบางข้อมูลอาจไม่มีความสัมพันธ์กัน ทำให้สิ้นเปลืองเวลาในการประมวลผลและลดประสิทธิภาพของการจำแนก [72] ดังนั้นสัญญาณไฟฟ้ากล้ามเนื้อจึงจำเป็นต้องผ่านกระบวนการสกัดลักษณะเด่นของสัญญาณเพื่อลดจำนวนมิติของข้อมูล และใช้บ่งชี้ลักษณะเฉพาะตัวของสัญญาณนั้นๆ

กระบวนการสกัดลักษณะเด่นของสัญญาณ คือ กระบวนการสกัดลักษณะสำคัญของข้อมูลจากสัญญาณดิบ ซึ่งทำให้ขนาดของข้อมูลลดลง แต่ยังคงเก็บลักษณะสำคัญของข้อมูลไว้ ซึ่งลักษณะเด่นดังกล่าวจะถูกใช้เป็นตัวแทนของสัญญาณไฟฟ้ากล้ามเนื้อในกระบวนการจำแนก โดยลักษณะเด่นต้องเป็นปริมาณที่สามารถวัดค่าได้และมีค่าเป็นตัวเลข นอกจากนี้กระบวนการสกัดลักษณะเด่นของสัญญาณอาจจะหมายถึง กระบวนการแปลงลักษณะเด่นเดิมไปยังเซตย่อยของลักษณะเด่นโดยผ่านฟังก์ชันต่าง ซึ่งผลของกระบวนการแปลงจะทำให้ลักษณะเด่น

มีมิติของข้อมูลน้อยลงหรือเท่ากับจำนวนลักษณะเด่นเดิม [73] อย่างไรก็ตามในงานวิจัยนี้จะอ้างอิงตามความหมายแรก

การคำนวณลักษณะเด่นของสัญญาณสามารถแบ่งออกได้เป็น 3 กลุ่ม คือ การวิเคราะห์ในโดเมนเวลา การวิเคราะห์ในโดเมนความถี่ และ การวิเคราะห์ทั้งในโดเมนเวลาและความถี่ ซึ่งลักษณะเด่นทั้งหมดที่ใช้ในการทดลองสำหรับงานวิจัยนี้ ได้แก่

### 2.7.1 ลักษณะเด่นที่วิเคราะห์ในโดเมนเวลา

ลักษณะเด่นที่วิเคราะห์ในโดเมนเวลาซึ่งถูกนำมาใช้ในงานวิจัยนี้จะพิจารณาจากลักษณะเด่นที่ถูกนำมาใช้ในหลากหลายงานวิจัยที่เกี่ยวข้องกับสัญญาณไฟฟ้ากล้ามเนื้อ [9], [30], [38], [74]-[86]

กำหนดให้  $x_n$  คือ ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่คอมพิวเตอร์ทำการชักตัวอย่าง (sampling) เข้ามาครั้งที่  $n$  โดยที่  $N$  คือ จำนวนข้อมูลทั้งหมด

1) ค่ารากเฉลี่ยกำลังสอง (root mean square, RMS) [38], [74-75]

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2} \quad (2.1)$$

2) ค่าความแปรปรวน (variance, VAR) [76-77]

$$VAR = \frac{1}{N-1} \sum_{n=1}^N x_n^2 \quad (2.2)$$

3) ค่าความยาวรูปคลื่น (waveform length, WL) [78]-[79]

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \quad (2.3)$$

4) ค่าแอมพลิจูดของวิลลิสัน (Willison amplitude, WAMP) [77], [80]

$$WAMP = \sum_{n=1}^{N-1} f(|x_n - x_{n+1}|), \quad (2.4)$$

$$f(x) = 1 \text{ if } x \geq TH \text{ (threshold), otherwise } f(x) = 0$$

5) ค่าจุดตัดศูนย์ (zero crossing, ZC) [78], [80]

$$ZC = \sum_{n=1}^{N-1} \text{sgn}(x_n \times x_{n+1}) \cap |x_n - x_{n+1}| \geq TH, \quad (2.5)$$

$$\text{sgn}(x) = 1 \text{ if } x \geq 0, \text{ otherwise } \text{sgn}(x) = 0$$

6) ค่าการเปลี่ยนแปลงความชัน (slope sign change, SSC) [78], [80]

$$SSC = \sum_{n=2}^{N-1} f[(x_n - x_{n-1}) \times (x_n - x_{n+1})], \quad (2.6)$$

$$f(x) = 1 \text{ if } x \geq TH, \text{ otherwise } f(x) = 0$$

7) ค่าเฉลี่ยของสัญญาณ (mean, MN) [9], [30]

$$MN = \frac{1}{N} \sum_{n=1}^N x_n \quad (2.7)$$

8) ค่าผลรวมของค่าสัมบูรณ์ (integrated absolute value, IAV) [81]-[82]

$$IAV = \sum_{n=1}^N |x_n| \quad (2.8)$$

9) ค่าเฉลี่ยของค่าสัมบูรณ์ (mean absolute value, MAV) [38], [77]-[78]

$$MAV = \frac{1}{N} \sum_{n=1}^N |x_n| \quad (2.9)$$

10) ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 1 (modified mean absolute value type 1, MMAV1) [79], [83]

$$MMAV1 = \frac{1}{N} \sum_{n=1}^N w_n |x_n|, \quad (2.10)$$

$$w_n = 1 \text{ if } 0.25N \leq n \leq 0.75N, \text{ otherwise } w_n = 0.5$$

11) ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 2 (modified mean absolute value type 2, MMAV2) [79], [83]

$$MMAV2 = \frac{1}{N} \sum_{n=1}^N w_n |x_n|, \quad (2.11)$$

$$w_n = 1 \text{ if } 0.25N \leq n \leq 0.75N, \quad w_n = 4n/N \text{ if } 0.25N > n,$$

$$w_n = 4(n - N) / N \text{ if } 0.75N < n$$

12) ค่าความชันของค่าเฉลี่ยค่าสัมบูรณ์ (mean absolute value slope, MAVSi) [84]-[85]

$$MAVS_i = MAV_{i+1} - MAV_i, \quad (2.12)$$

$i = 1, 2, 3, \dots, s-1$  เมื่อ  $s$  คือ จำนวนของ segment-1

13) ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน (frame-based time domain feature with stacking filter, TDn) จะแบ่งสัญญาณออกเป็นส่วนๆ ที่เท่ากัน แต่ละส่วนเรียกว่า เฟรม สัญญาณในแต่ละเฟรมจะถูกนำไปคำนวณลักษณะเด่นแล้วนำค่าลักษณะเด่นที่ได้ในแต่ละเฟรมมาซ้อนกันเพื่อให้ได้ลักษณะเด่นตัวใหม่ขึ้นมา ซึ่งเป็นการใช้เทคนิคของตัวกรองซ้อน ซึ่งถูกคิดค้นโดยกลุ่มวิจัยของชาวเยอรมัน [9]

ขั้นตอนการคำนวณลักษณะเด่น TDn

13.1) นำสัญญาณไฟฟ้ากล่อมเนื้อในแต่ละสัญญาณมาทำการนอร์มัลไลซ์ (normalize) ด้วยค่าเฉลี่ยของสัญญาณนั้นๆ

กำหนดให้  $x_m$  คือ สัญญาณไฟฟ้ากล่อมเนื้อที่ผ่านจากกระบวนการนอร์มัลไลซ์  $m=1$  ถึง  $N$  และ  $n=1$  ถึง  $N$

$$x_m = \frac{x_n - \frac{1}{N} \sum_{n=1}^N x_n}{\sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2 - \left(\frac{1}{N} \sum_{n=1}^N x_n\right)^2}} = \frac{x_n - \mu}{\sigma} \quad (2.13)$$

13.2) แยกองค์ประกอบความถี่ต่ำและความถี่สูงของสัญญาณออกจากกันโดยใช้ตัวกรองค่าเฉลี่ยเคลื่อนที่ (moving average filter) ในที่นี้จะใช้การเฉลี่ยแบบ 9 จุดสองรอบ

กำหนดให้  $w_m$  คือ องค์ประกอบความถี่ต่ำของสัญญาณ  $p_m$  คือ องค์ประกอบความถี่สูงของสัญญาณ และ  $r_m$  คือ องค์ประกอบความถี่สูงของสัญญาณที่ผ่านการกรอง

$$w_m = \frac{1}{9} \sum_{m=-4}^4 v_m, \text{ โดยที่ } v_m = \frac{1}{9} \sum_{m=-4}^4 x_m \quad (2.14)$$

$$p_m = x_m - w_m \quad (2.15)$$

$$r_m = |p_m| \quad (2.16)$$

13.3) นำสัญญาณไฟฟ้ากล่อมเนื้อในแต่ละสัญญาณแบ่งออกเป็นส่วนที่เท่ากัน ซึ่งแต่ละส่วนถูกเรียกว่า เฟรม โดยแต่ละเฟรมมีขนาด 27 มิลลิวินาที (frame size) และ แต่ละเฟรมซ้อนทับกัน 10 มิลลิวินาที (frame shift) กำหนดให้อัตราการซีกตัวอย่างของสัญญาณ (sampling rate) 1024 ตัวอย่าง/วินาที (sample/second) ดังนั้นในแต่ละเฟรมจะมีจำนวน 27 ตัวอย่าง แต่ละเฟรมซ้อนกัน 10 ตัวอย่าง

13.4) นำสัญญาณไฟฟ้ากล่อมเนื้อในแต่ละเฟรมมาคำนวณลักษณะเด่น 5 ชนิด ซึ่งลักษณะเด่นดังกล่าวเป็นลักษณะเด่นในแกนเวลา ซึ่งถูกเรียกว่าลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน (frame-based time domain features) ลักษณะเด่นทั้ง 5 ชนิด ได้แก่

(ก) ค่าเฉลี่ยขององค์ประกอบความถี่ต่ำของสัญญาณ (mean of low frequency part, MNL)

$$\bar{w} = \frac{1}{N} \sum_{m=1}^N w_m \quad (2.17)$$

(ข) ค่าเฉลี่ยขององค์ประกอบความถี่สูงของสัญญาณ (mean of high frequency part, MNH)

$$\bar{r} = \frac{1}{N} \sum_{m=1}^N r_m \quad (2.18)$$

(ค) ค่ากำลังขององค์ประกอบความถี่ต่ำของสัญญาณ (power of low frequency part, PWL)

$$p_w = \sum_{m=1}^N |w_m|^2 \quad (2.19)$$

(ง) ค่ากำลังขององค์ประกอบความถี่สูงของสัญญาณ (power of high frequency part, PWH)

$$p_r = \sum_{m=1}^N |r_m|^2 \quad (2.20)$$

(จ) ค่าจุดตัดผ่านศูนย์ขององค์ประกอบความถี่สูงของสัญญาณ (zero crossing of high frequency part, ZCH)

$$z_p = \sum_{m=1}^{N-1} \text{sgn}(p_m \times p_{m+1}), \tag{2.21}$$

$$\text{sgn}(x) = 1 \text{ if } x \geq 0, \text{ otherwise } \text{sgn}(x) = 0$$

13.5) สร้างลักษณะเด่นตัวใหม่ขึ้นมาโดยใช้เทคนิคของตัวกรองแบบบริบท ซึ่งประกอบด้วยตัวกรอง 3 ชนิด คือ ตัวกรองเดลต้า ตัวกรองเทรนต์ และ ตัวกรองซ้อน ในที่นี้จะใช้เทคนิคตัวกรองซ้อน ซึ่งสมการของตัวกรองซ้อน คือ

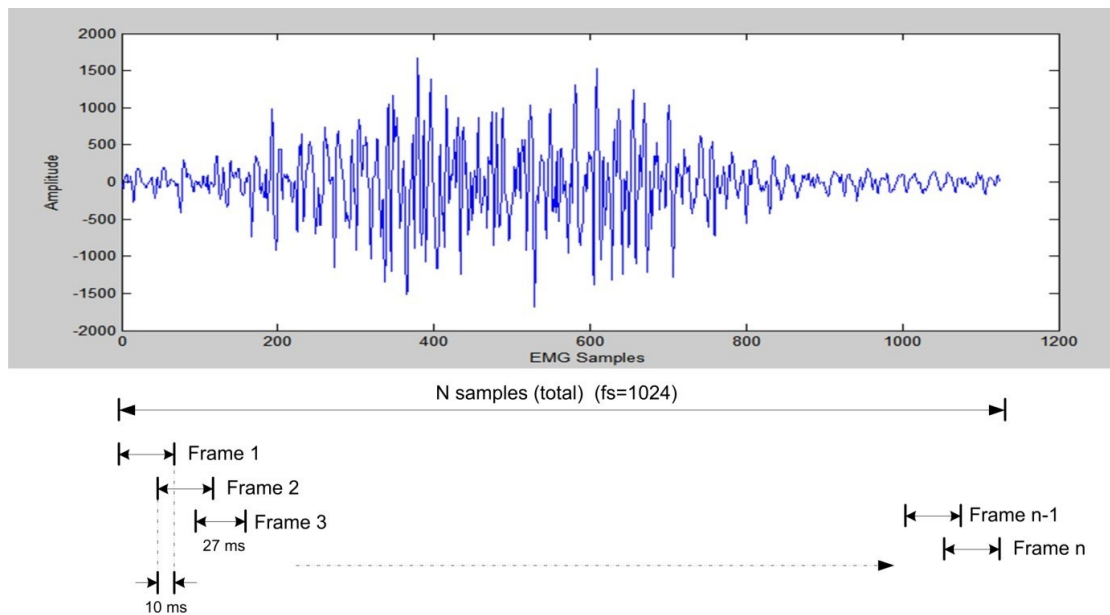
$$S(f_j, k) = [f_{j-k}, f_{j-k+1}, \dots, f_{j+k-1}, f_{j+k}] \tag{2.22}$$

โดยที่  $f = [\bar{w}, \bar{r}, p_w, p_r, z_p]$ ,  $j$  คือ ตัวบ่งชี้เฟรม (frame index),  $k$  คือ ความกว้างการซ้อน (stacking width) (จำนวนเฟรมที่อยู่ติดกับเฟรมปัจจุบันที่จะนำค่าลักษณะเด่นทั้ง 5 ชนิดมาซ้อนกัน ทั้งเฟรมก่อนหน้าและเฟรมถัดไป)

ค่าลักษณะเด่นสุดท้ายที่คำนวณได้จะอยู่ในรูปของ  $TDn$  [23]

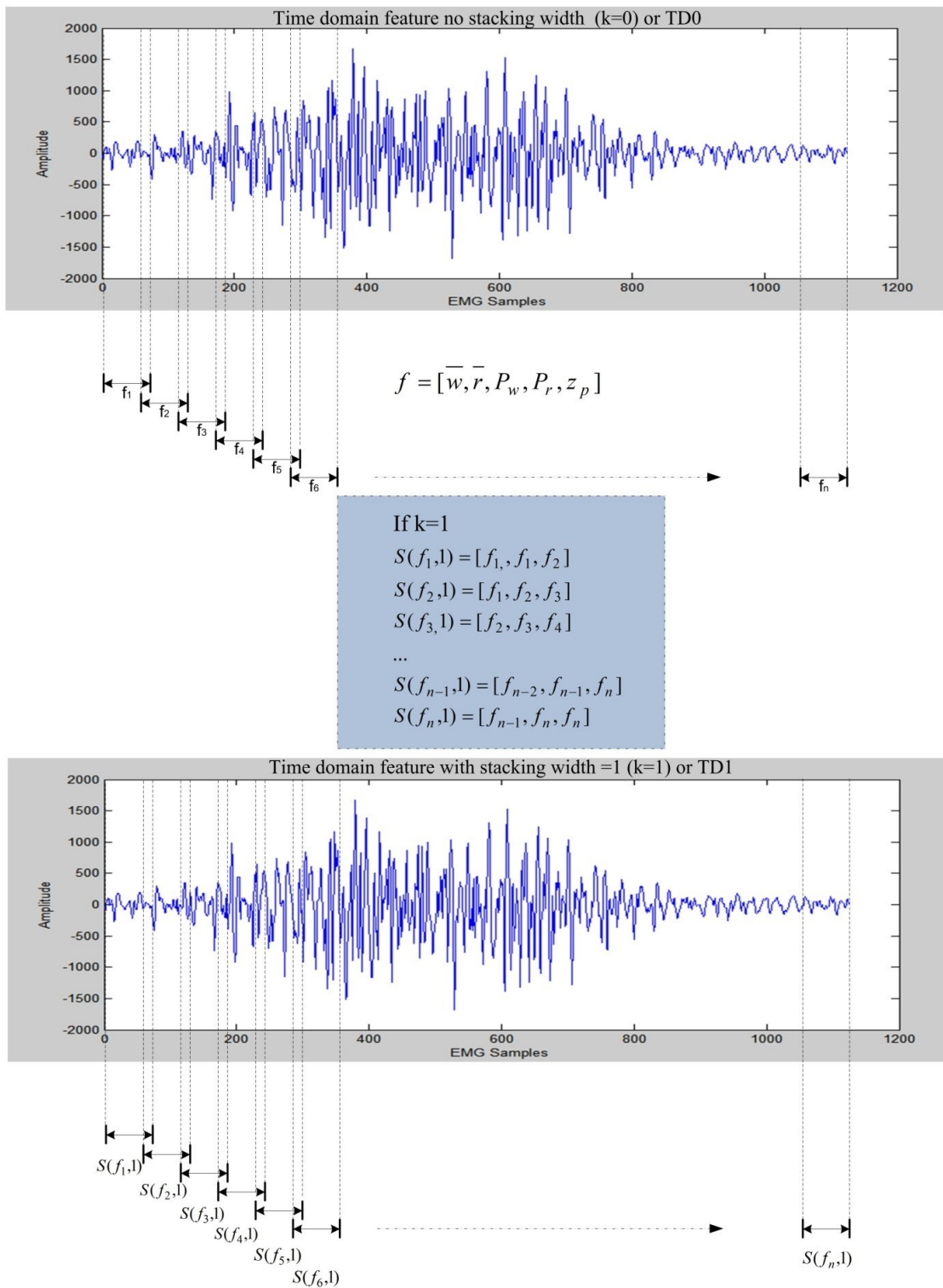
$$TDn = S(TD0, n) \text{ ซึ่ง } TD0 = [\bar{w}, \bar{r}, p_w, p_r, z_p] \tag{2.23}$$

โดยที่  $n$  คือ ความกว้างการซ้อน,  $TD0$  คือ การคำนวณค่าลักษณะเด่นที่  $n = 0$  กล่าวคือ ไม่มีการนำลักษณะเด่นที่คำนวณได้ของแต่ละเฟรมมาซ้อนกัน (TD ย่อมาจาก time domain feature) สำหรับในงานวิจัยนี้จะทำการทดลองโดยปรับจำนวน  $n$  ให้เหมาะสม



รูปที่ 2.6 แสดงการแบ่งเฟรมของสัญญาณไฟฟ้ากล้ามเนื้อเนื้อแต่ละสัญญาณ





รูปที่ 2.7 การคำนวณลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน ภาพบน: ไม่ได้ใช้ตัวกรองซ้อน (TD0) ภาพล่าง: แสดงลักษณะเด่นแบบเดียวกันแต่มีการใช้ตัวกรองซ้อนโดยปรับ  $k=1$  ลักษณะเด่นชนิดนี้ถูกเรียกว่า TD1

## 2.7.2 ลักษณะเด่นที่วิเคราะห์ในโดเมนความถี่

กำหนดให้  $P_i$  คือ กำลังไฟฟ้าสเปกตรัมที่ความถี่  $i$  และ  $f_i$  คือ ความถี่ที่ตำแหน่ง  $i$  และ  $M$  คือ  $(N/2)-1$

1) ค่าความถี่กลาง (median frequency, MDF) [79]

$$\sum_{i=1}^{MDF} P_i = \sum_{i=MDF}^M P_i = \frac{1}{2} \sum_{i=1}^M P_i \quad (2.24)$$

2) ค่าเฉลี่ยของความถี่ (mean frequency, MNF) [76], [86]

$$MNF = \frac{\sum_{i=1}^M f_i P_i}{\sum_{i=1}^M P_i} \quad (2.25)$$

3) ค่าสเปกตรัลโมเมนต์ (spectral moment, SMn) [86]

$$SMn = \sum_{i=1}^M f_i^n P_i \quad (2.26)$$

โดยที่  $n$  คือ จำนวนอันดับของค่าสเปกตรัลโมเมนต์

## 2.8 การคัดเลือกลักษณะเด่น

การคัดเลือกลักษณะเด่น (feature selection) คือ กระบวนการเลือกกลุ่มของลักษณะเด่นที่ดีที่สุดจากลักษณะเด่นทั้งหมดที่มีอยู่เพื่อให้ได้กลุ่มของลักษณะเด่นที่ดีที่สุดที่ใช้ในการจำแนกข้อมูล หรือกระบวนการเลือกเซตย่อยของลักษณะเด่นเดิมโดยไม่ผ่านกระบวนการแปลงผ่านฟังก์ชันต่างๆ [87] กระบวนการนี้จะลดจำนวนของลักษณะเด่นที่จะป้อนเป็นอินพุตให้กับตัวจำแนก โดยจะกำจัดข้อมูลที่ไม่สัมพันธ์กัน ข้อมูลที่ไม่จำเป็น ซึ่งทำให้เพิ่มความเร็วในการประมวลผลและอาจจะเพิ่มประสิทธิภาพในการจำแนกข้อมูล [87]-[88]

การแบ่งประเภทของเทคนิคการคัดเลือกลักษณะเด่นสามารถแบ่งได้เป็น 2 กลุ่ม [89]

กลุ่มแรก จะแบ่งประเภทตามเกณฑ์ประเมิน (evaluation criterion) ที่ใช้ในการประเมินลักษณะเด่นหรือเซตย่อยของลักษณะเด่น โดยจะแบ่งออกเป็น 3 ประเภท คือ

1) เทคนิคฟิลเตอร์ (filter method, FM) จะใช้คุณลักษณะภายในที่มีอยู่ของข้อมูลในการประเมินโดยไม่อาศัยผลการจำแนกของตัวจำแนก

2) เทคนิคแรปเปอร์ (wrapper method, WM) จะใช้ผลความแม่นยำในการจำแนกของตัวจำแนกในการประเมิน

3) เทคนิคฝังตัว (embedded method, EM) [88] หรือ เทคนิคไฮบริด (hybrid method, HM) [87] ประเมินลักษณะเด่นด้วยการประยุกต์ใช้ข้อดีของเทคนิคของฟิลเตอร์และแรปเปอร์รวมกัน โดยใช้การประเมินคุณลักษณะทั่วไปของข้อมูลสำหรับเลือกเซตย่อยของลักษณะเด่นและใช้ผลความแม่นยำในการจำแนกของตัวจำแนกสำหรับเลือกเซตย่อยสุดท้าย

กลุ่มสอง จะแบ่งตามแนวทางที่ลักษณะเด่นจะถูกประเมิน โดยจะแบ่งออกเป็น 2 ประเภท [89]-[90] คือ

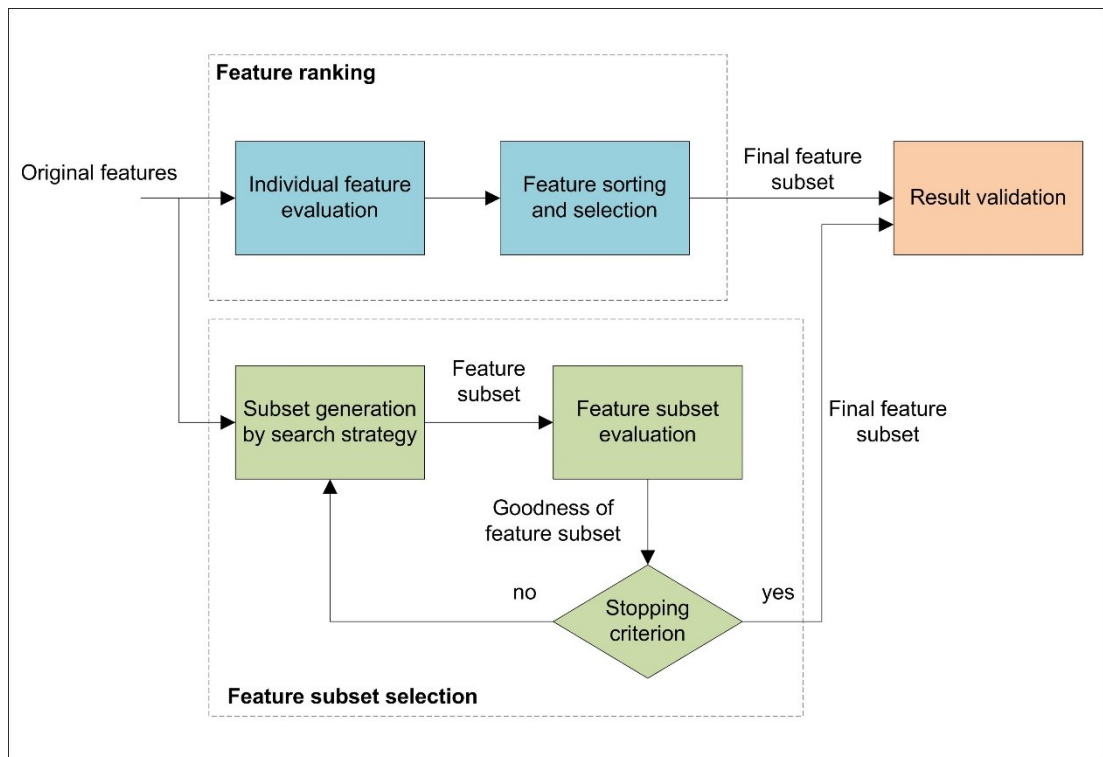
1) การเรียงลำดับลักษณะเด่น (feature ranking, FR) หรือ เป็นการประเมินลักษณะเด่นแต่ละตัวแยกกัน เรียงลำดับลักษณะเด่นตามผลการประเมิน และเลือกลักษณะเด่นที่ดีที่สุด  $k$  ตัว โดยเกณฑ์ที่ใช้ในการกำหนดจำนวนลักษณะเด่นมีหลายวิธี ได้แก่ การกำหนดจำนวนคงที่ การกำหนดค่าขีดแบ่ง เป็นต้น [91]

2) การเลือกเซตย่อยของลักษณะเด่น (feature subset selection, FSS) เป็นการเลือกเซตย่อยของลักษณะเด่นโดยผ่านกระบวนการค้นหา (search strategy) แล้วทำการประเมินเซตย่อยของลักษณะเด่นด้วยเกณฑ์การประเมิน

รูปที่ 2.8 แสดงขั้นตอนการคัดเลือกลักษณะเด่น ซึ่งจะแสดงขั้นตอนของการเลือกลักษณะเด่นสำหรับเทคนิค FR และ FSS สำหรับเทคนิคการเลือกลักษณะเด่นแบบ FR จะพิจารณาลักษณะเด่นแต่ละตัว แล้วประเมินลักษณะเด่นแต่ละตัวแยกกัน ซึ่งได้กล่าวไว้แล้วข้างต้น

สำหรับเทคนิค FSS จะประกอบด้วย 4 ขั้นตอนหลัก [87]-[88]

1) การสร้างเซตย่อยของลักษณะเด่น (subset generation) ซึ่งต้องผ่านกระบวนการค้นหาลักษณะเด่น โดยค้นหาจากเซตของลักษณะเด่นทั้งหมด ในกระบวนการค้นหาลักษณะเด่นต้องหาจุดเริ่มต้นของการค้นหา ซึ่งอาจจะเริ่มจากเซตว่าง แล้วค่อยเพิ่มจำนวนสมาชิกของเซตย่อย หรือเริ่มจากเซตเต็ม แล้วค่อยลดจำนวนสมาชิกของเซตย่อย เป็นต้น ซึ่งจุดเริ่มต้นดังกล่าวขึ้นอยู่กับวิธีการเลือกใช้เทคนิคการค้นหาลักษณะเด่น ซึ่งมีหลายวิธี ได้แก่



รูปที่ 2.8 กระบวนการคัดเลือกลักษณะเด่น (feature selection process)

### 1.1) การค้นหาแบบสมบูรณ์ (complete search)

เป็นการค้นหาแบบเอกซ์โพเนนเชียล (exponential search) การค้นหาเซตย่อยวิธีนี้ทำให้ได้เซตย่อยที่เหมาะสมที่สุดซึ่งสอดคล้องกับผลการประเมิน เช่น การค้นหาแบบบรานช์แอนด์บาวด์ (branch and bound search) การค้นหาแบบบีม (beam search) เป็นต้น

### 1.2) การค้นหาตามลำดับ (sequential search)

เป็นวิธีการค้นหาที่ง่ายและเร็ว เช่น การค้นหาแบบไปข้างหน้า (forward search) การค้นหาแบบไปข้างหลัง (backward search) การค้นหาแบบสองทิศทาง (bidirectional selection) การค้นหาแบบบวกแอลลบอาร์ (plus L take away R search) เป็นต้น วิธีนี้จะทำการเพิ่มหรือลดจำนวนสมาชิกของเซตย่อยครั้งละ 1 ตัว

2) การประเมินเซตย่อย (subset evaluation) เซตย่อยที่สร้างขึ้นมาน่าจะถูกนำมาประเมินด้วยเกณฑ์การประเมิน ซึ่งเกณฑ์การประเมินแบ่งออกเป็น 2 กลุ่ม คือ เกณฑ์การประเมินที่ไม่ขึ้นกับตัวจำแนก และ เกณฑ์การประเมินที่ขึ้นกับตัวจำแนก

2.1) เกณฑ์การประเมินที่ไม่ขึ้นกับตัวจำแนก จะใช้คุณลักษณะภายในที่มีอยู่ของ ข้อมูลในการประเมิน ได้แก่ การวัดความสามารถในการแบ่งกลุ่มของข้อมูล โดยวัดระยะทาง ระหว่างกลุ่มข้อมูลและภายในกลุ่มข้อมูล การวัดเกนข้อมูล (information gain) ซึ่งใช้หลักการของ เอนโทรปี (entropy) ซึ่งวัดระดับความเจือปนของกลุ่มข้อมูล การวัดความสัมพันธ์ของข้อมูล (correlation) เป็นต้น ตัวประเมินที่ไม่ขึ้นอยู่กับตัวจำแนกซึ่งถูกใช้ในงานวิจัยนี้มีดังนี้

2.1.1) ดัชนีอาร์อีเอส (ratio of Euclidean and standard deviation, RES) [92]

$$RES = \frac{\overline{ED}}{\sigma} \quad (2.27)$$

ซึ่ง  $\overline{ED}$  คือ ค่าเฉลี่ยของระยะทางแบบยูคลิด (Euclidean distance) เป็นการวัด ระยะทางระหว่างกลุ่มข้อมูลโดยวัดจากจุดศูนย์กลางหรือเซนทรอยด์ (centroid) ของข้อมูลแต่ละ กลุ่มโดยใช้ระยะทางแบบยูคลิด เพื่อวัดการกระจายของข้อมูลระหว่างกลุ่ม ( $M_{ij}$ ) และนำมาหา ค่าเฉลี่ย และ  $\sigma$  คือ ค่าเฉลี่ยของค่าเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งวัดการกระจาย ภายในกลุ่มข้อมูล ( $S_i^k$ ) แล้วนำมาหาค่าเฉลี่ยซึ่งแสดงดังสมการต่อไปนี้

$$\overline{ED} = \frac{2}{N(N+1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N M_{ij} \quad (2.28)$$

$$M_{ij} = \sqrt{\sum_{h=1}^Y |a_{hi} - a_{hj}|^2} \quad (2.29)$$

$$\sigma = \frac{1}{YN} \sum_{i=1}^Y \sum_{k=1}^N S_i^k \quad (2.30)$$

$$S_i^k = \frac{1}{Z_k} \sum_{j=1}^{Z_k} (x_{ij}^k - a_{ik})^2 \quad (2.31)$$

โดยที่  $N$  คือ จำนวนของกลุ่มข้อมูลทั้งหมด (จำนวนกลุ่มข้อมูลที่ต้องการจำแนก ในที่นี้คือ จำนวนคำพูดที่ต้องการจำแนก)  $Y$  คือ จำนวนช่องสัญญาณของสัญญาณไฟฟ้า กล้ามเนื้อ  $Z_k$  คือ จำนวนข้อมูลของแต่ละกลุ่มในแต่ละช่องสัญญาณ (ในที่นี้คือ จำนวนครั้งในการ พูดซ้ำของแต่ละคำ) ซึ่ง  $k = 1, 2, 3, \dots, N$

โดยกำหนดให้  $C_k$  คือ เมตริกซ์ของกลุ่มข้อมูลที่  $k$  จำนวนแถวของเมตริกซ์แสดง จำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ จำนวนคอลัมน์ของเมตริกซ์แสดงจำนวนข้อมูล ของแต่ละกลุ่มในแต่ละช่องสัญญาณ จะได้ว่า

$$\mathbf{C}_k = \begin{bmatrix} x_{11}^k & x_{12}^k & \cdots & x_{1Z}^k \\ x_{21}^k & x_{22}^k & \cdots & x_{2Z}^k \\ \vdots & \vdots & \ddots & \vdots \\ x_{Y1}^k & x_{Y2}^k & \cdots & x_{YZ}^k \end{bmatrix} \quad (2.32)$$

$x_{ij}^k$  คือ สัญญาณไฟฟ้าที่ผ่านกระบวนการคำนวณลักษณะเด่นของคำพูดที่  $k$  (cluster  $k^{th}$  จากทั้งหมด  $N$  กลุ่ม) ซึ่งพูดซ้ำครั้งที่  $j$  (ทั้งหมด  $Z$  ครั้ง) จากช่องสัญญาณที่  $i$  (จากทั้งหมด  $Y$  ช่อง)

กำหนดให้  $\mathbf{A}$  คือ เมตริกซ์ของเซนทรอยด์ของแต่ละกลุ่ม และ  $\mathbf{A}_k$  คือ เวกเตอร์ของเซนทรอยด์กลุ่มที่  $k$  ซึ่ง  $k = 1, 2, 3, \dots, N$

$$\mathbf{A}_k = \begin{bmatrix} \frac{x_{11}^k + x_{12}^k + \cdots + x_{1Z}^k}{Z} \\ \frac{x_{21}^k + x_{22}^k + \cdots + x_{2Z}^k}{Z} \\ \vdots \\ \frac{x_{Y1}^k + x_{Y2}^k + \cdots + x_{YZ}^k}{Z} \end{bmatrix} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{Yk} \end{bmatrix} \quad (2.33)$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{Y1} & a_{Y2} & \cdots & a_{YN} \end{bmatrix} \quad (2.34)$$

$a_{ik}$  คือ เซนทรอยด์ของกลุ่มที่  $k$  สำหรับช่องสัญญาณที่  $i$

2.1.2) เกณฑ์ระยะทางภายในกลุ่มและระหว่างกลุ่มข้อมูล (inter-intra distance criterion, in-in)

เป็นการประเมินลักษณะเด่นโดยเลือกลักษณะเด่นที่ให้ข้อมูลการกระจายตัวในภายในคลาสต่ำสุดและการกระจายตัวระหว่างคลาสสูงสุด โดยค่าการประเมินคำนวณโดยใช้สมการ (2.35)

$$J = \text{Tr} \{ \mathbf{S}_W^{-1} \mathbf{S}_B \} \quad (2.35)$$

โดยที่  $J$  คือ ค่าการประเมินซึ่งได้มาจากผลรวมของค่าที่อยู่ในแนวทแยงของเมตริกซ์  $\mathbf{S}_W^{-1} \mathbf{S}_B$  (trace of matrix) หรือ ผลรวมของค่าไอเกน (Eigen value) ของเมตริกซ์นั้น

$\mathbf{S}_W$  คือ เมตริกซ์การกระจายภายในคลาส (within-class scatter matrix) และ  $\mathbf{S}_B$  คือ เมตริกซ์โควาเรียนซ์ระหว่างคลาส (between-class covariance matrix) จะได้ว่า

$$\mathbf{S}_W = \sum_{i=1}^C \frac{n_i}{n} \hat{\Sigma}_i \quad (2.36)$$

$$\mathbf{S}_B = \sum_{i=1}^C \frac{n_i}{n} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (2.37)$$

$\hat{\Sigma}_i$  คือ ค่าประมาณไคลส์ติคสูงสุดของเมตริกซ์โควาเรียนซ์ของคลาส  $\omega_i$ ,  $C$  คือ จำนวนคลาสของข้อมูล

$$\hat{\Sigma}_i = \frac{1}{n_i} \sum_{j=1}^n z_{ij} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \quad (2.38)$$

$$z_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_j \in \omega_i \\ 0 & \text{otherwise} \end{cases}, n_i = \sum_{j=1}^n z_{ij}$$

$\mathbf{m}_i$  คือ ค่าเฉลี่ยของข้อมูลของคลาส  $\omega_i$ ,  $n$  คือ จำนวนข้อมูลทั้งหมด

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{j=1}^n z_{ij} \mathbf{x}_j \quad (2.39)$$

$$\mathbf{m} = \sum_{i=1}^C \frac{n_i}{n} \mathbf{m}_i \quad (2.40)$$

2.1.3) เกณฑ์ผลรวมของระยะทางมาฮาลานอบิสแบบประมาณ (sum of estimated Mahalanobis distance criterion, Maha-s)

ระยะทางแบบมาฮาลานอบิส หรือ ระยะทางแบบควอดราติก (quadratic distance) วัดการแยกกันของกลุ่มข้อมูล ซึ่งวัดระยะทางระหว่างค่าเฉลี่ยของกลุ่มข้อมูล สมมติให้มีจำนวนกลุ่มข้อมูล 2 กลุ่มหรือ 2 คลาส สามารถคำนวณระยะทางแบบมาฮาลานอบิสได้จากสมการที่ (2.41)

$$d = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2)} \quad (2.41)$$

ซึ่งการประเมินลักษณะเด่นโดยใช้เกณฑ์ผลรวมของระยะทางมาฮาลานอบิสแบบประมาณ ค่าการประเมินจะได้จากค่าผลรวมของระยะทางแบบมาฮาลานอบิสที่คำนวณได้จาก

สมการ (2.41) ซึ่งลักษณะเด่นที่มีค่าการประเมินสูง นั้นคือระยะทางระหว่างกลุ่มข้อมูลแยกกันมากที่สุด ซึ่งมีผลทำให้การจำแนกข้อมูลได้ดีที่สุด

2.1.4) เกณฑ์ค่าต่ำสุดของระยะทางมาฮาลานอบิสแบบประมาณ (minimum of estimated Mahalanobis distance criterion, Maha-m)

วิธีนี้จะเลือกค่าต่ำสุดของระยะทางมาฮาลานอบิสดังสมการที่ (2.41) เป็นตัวประเมินลักษณะเด่น ซึ่งลักษณะเด่นที่มีค่าการประเมินสูงสุด นั้นคือระยะทางระหว่างกลุ่มข้อมูลที่ต่ำที่สุดมีค่ามากที่สุด ซึ่งมีผลทำให้การจำแนกข้อมูลได้ดีที่สุด

2.1.5) เกณฑ์ผลรวมของระยะทางแบบยูคลิดกำลังสอง (sum of squared Euclidean distance criterion, Eucl-s)

จากสมการที่ (2.41) ถ้า  $S_w^{-1}$  เป็นเมตริกซ์เอกลักษณ์ (identity matrix) สมการดังกล่าวจะถูกเรียกว่าระยะทางแบบยูคลิด สมมุติให้มีจำนวนกลุ่มข้อมูล 2 กลุ่มหรือ 2 คลาส สามารถคำนวณระยะทางแบบยูคลิดกำลังสองจากสมการที่ (2.42)

$$d^2 = (\mathbf{m}_1 - \mathbf{m}_2)^2 \quad (2.42)$$

ซึ่งการประเมินลักษณะเด่นโดยใช้เกณฑ์ผลรวมของระยะทางแบบยูคลิดกำลังสอง ค่าการประเมินจะได้จากค่าผลรวมของระยะทางแบบยูคลิดกำลังสองที่คำนวณได้จากสมการที่ (2.42)

2.1.6) เกณฑ์ค่าต่ำสุดของระยะทางแบบยูคลิดกำลังสอง (minimum of squared Euclidean distance criterion, Eucl-m)

วิธีนี้จะเลือกค่าต่ำสุดของระยะทางแบบยูคลิดกำลังสองดังสมการที่ (2.42) เป็นตัวประเมินลักษณะเด่น

2.1.7) เกณฑ์ค่าสหสัมพันธ์แบบเชิงเส้นของเพียร์สัน (Pearson's linear correlation criterion, CorrP)

สหสัมพันธ์ (correlation) เป็นการหาความสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป ซึ่งค่าที่ใช้ในการวัดความสัมพันธ์ระหว่างตัวแปรดังกล่าวจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ซึ่งในการประเมินลักษณะเด่นวิธีนี้จะวัดค่าความสัมพันธ์ระหว่างเวกเตอร์ของลักษณะเด่นที่พิจารณากับเวกเตอร์คลาสของข้อมูล



สำหรับค่าการประเมินลักษณะเด่นโดยใช้ค่าสหสัมพันธ์แบบเชิงเส้นของเพียสัน จะให้ค่าอยู่ระหว่าง -1 ถึง +1 โดยถ้าให้ค่าการประเมินเท่ากับ 0 แสดงว่าไม่มีความสัมพันธ์ระหว่างสองตัวแปร สมการที่ใช้ในการคำนวณสัมประสิทธิ์ค่าสหสัมพันธ์แบบเชิงเส้นของเพียสันแสดงดังสมการที่ (2.43)

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{nS_x S_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} \sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (2.43)$$

$r_{xy}$  คือ สัมประสิทธิ์สหสัมพันธ์ของเวกเตอร์  $\mathbf{x}$  และ  $\mathbf{y}$  โดยเวกเตอร์แต่ละตัวมีจำนวนสมาชิกเท่ากับ  $n$ ,  $S_x$  และ  $S_y$  คือ ค่าเบี่ยงเบนมาตรฐานของเวกเตอร์  $\mathbf{x}$  และ  $\mathbf{y}$  ตามลำดับ

2.1.8) เกณฑ์ค่าสหสัมพันธ์แบบสเปียร์แมน (Spearman rank correlation criterion, CorrS)

การหาค่าสหสัมพันธ์วิธีนี้ถูกพัฒนามาจากการหาค่าสหสัมพันธ์เชิงเส้นแบบเพียสัน วิธีนี้จะใช้หลักการจัดอันดับของข้อมูล ซึ่งคำนวณจากสมการที่ (2.44)

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (2.44)$$

$$d_i = rg(x_i) - rg(y_i)$$

$r_s$  คือ สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนของเวกเตอร์  $\mathbf{x}$  และ  $\mathbf{y}$

$rg(x_i)$  และ  $rg(y_i)$  คือ ค่าของสมาชิกในเวกเตอร์  $\mathbf{x}$  และ  $\mathbf{y}$  ซึ่งถูกแปลงข้อมูลให้อยู่ในรูปของการจัดอันดับ

2.1.9) เกณฑ์ค่าสหสัมพันธ์แบบเคนดอล (Kendall rank correlation criterion, CorrK) ใช้หลักการหาความสัมพันธ์ของตัวแปรที่ถูกจัดอันดับเช่นเดียวกับการหาค่าสหสัมพันธ์แบบสเปียร์แมน

กำหนดให้  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  เป็นสมาชิกของของเวกเตอร์  $\mathbf{x}$  และ  $\mathbf{y}$  สำหรับคู่ของข้อมูล  $(x_i, y_i)$  กับ  $(x_j, y_j)$  ซึ่ง  $i \neq j$  จะสอดคล้องกันถ้าข้อมูลการจัดอันดับของทั้งคู่ตรงกัน ถ้าคู่ของ  $x_i > x_j$  กับ  $y_i > y_j$  หรือ  $x_i < x_j$  กับ  $y_i < y_j$  ข้อมูลของคู่นั้นจะไม่สอดคล้อง

กัน อย่างไรก็ตามข้อมูลของคู่นั้นจะไม่ถูกจัดอยู่ในทั้งสองกรณี ถ้าคู่ของ  $x_i > x_j$  กับ  $y_i < y_j$  หรือ  $x_i < x_j$  กับ  $y_i > y_j$

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)} \quad (2.45)$$

$n_c$  คือ จำนวนคู่ของข้อมูลที่สอดคล้องกัน และ  $n_d$  คือ จำนวนคู่ของข้อมูลที่ไม่สอดคล้องกัน

### 2.1.10) เกณฑ์ค่าเกินความรู้ (information gain criterion, IG)

เกินความรู้ (information gain) ใช้หลักการของเอนโทรปี (entropy) ในการประเมินลักษณะเด่นที่เหมาะสมที่สุดในการจำแนก เอนโทรปี คือ การวัดระดับความเจือปนของกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างมีค่าเอนโทรปีต่ำหรือระดับความเจือปนต่ำ แสดงว่ากลุ่มตัวอย่างนั้นสามารถแยกออกจากกลุ่มตัวอย่างกลุ่มอื่นได้ดี

$$Entropy = \sum_i -p_i \log_2 p_i \quad (2.46)$$

$p_i$  คือ ความน่าจะเป็นของข้อมูล  $i$  ซึ่งคำนวณจากสัดส่วนของจำนวนข้อมูลของ  $i$  ต่อจำนวนข้อมูลทั้งหมดในกลุ่มตัวอย่างที่พิจารณา

$$Information\ Gain = Entropy(target) - Entropy(target|feature) \quad (2.47)$$

$$IG(T, f) = H(T) - H(T|f)$$

กำหนดให้  $T$  คือ เซตของข้อมูลที่ใช้ในการฝึกสอนซึ่งอยู่ในรูปของ  $(\mathbf{x}, y) = (x_1, x_2, x_3, \dots, x_k, y)$  โดยที่  $x_f \in vals(f)$  ซึ่งคือ ค่าของลักษณะเด่นตัวที่  $f$  และ  $y$  คือ ค่าของคลาสที่สอดคล้องกัน ค่าเกินความรู้ของลักษณะเด่น  $f$  ถูกแสดงดังสมการ (2.48)

$$IG(T, f) = H(T) - \sum_{v \in vals(f)} \frac{|\{\mathbf{x} \in T | x_f = v\}|}{|T|} \cdot H(\{\mathbf{x} \in T | x_f = v\}) \quad (2.48)$$

2.2) เกณฑ์การประเมินที่ขึ้นอยู่กับตัวจำแนก จะใช้ประสิทธิภาพในการจำแนกของตัวจำแนกเป็นเกณฑ์การประเมิน ซึ่งเซตย่อยของลักษณะเด่นจะถูกนำเข้าสู่กระบวนการเรียนรู้และการทดสอบระบบของตัวจำแนก เซตย่อยที่ให้ผลการจำแนกที่ดีที่สุดคือ เซตย่อยที่เหมาะสมที่สุด

3) การหยุดกระบวนการเลือกลักษณะเด่น (stopping criteria) กระบวนการเลือกลักษณะเด่นจะทำซ้ำการค้นหาและประเมินเซตย่อยของลักษณะเด่นไปเรื่อยๆ จนกว่าจะถึงเกณฑ์ที่กำหนดไว้ ซึ่งเกณฑ์ดังกล่าวได้แก่

3.1) กระบวนการค้นหาเสร็จสมบูรณ์

3.2) ถึงขอบเขตที่กำหนดให้ ขอบเขตดังกล่าว ได้แก่ จำนวนสมาชิกของเซตย่อยที่ต้องการ จำนวนสูงสุดของการทำซ้ำ เป็นต้น

3.3) เมื่อการเพิ่มหรือลดจำนวนสมาชิกของเซตย่อย ไม่ทำให้ประสิทธิภาพดีขึ้นอีก

3.4) เมื่อได้ประสิทธิภาพที่เพียงพอ

4) การตรวจสอบผล (result validation) เป็นการตรวจสอบประสิทธิภาพของเทคนิคการเลือกของลักษณะเด่น โดยใช้เซตย่อยสุดท้าย เพื่อเป็นอินพุตให้กับตัวจำแนกในการตรวจสอบ กระบวนการนี้ไม่ได้เป็นส่วนหนึ่งของกระบวนการเลือกลักษณะเด่น [88]

## 2.9 ตัวจำแนกรูปแบบที่ใช้ในการจำแนกสัญญาณไฟฟ้ากล้ามเนื้อ

จากการทบทวนวรรณกรรมที่ได้กล่าวไว้ในบทที่ 1 จะเห็นได้ว่า มีการใช้ตัวจำแนกหลายชนิดด้วยกัน ซึ่งที่นิยมมากที่สุด คือ HMM และ ANN ซึ่งในงานวิจัยนี้ จะนำตัวจำแนกสองชนิดมาใช้ในงานวิจัย และจะทำการเปรียบเทียบประสิทธิภาพตัวจำแนกชนิดอื่นอีก 7 ชนิด ซึ่งในหัวข้อนี้จะกล่าวถึงหลักการทั่วไปของตัวจำแนกแต่ละชนิด

### 2.9.1 ตัวจำแนกแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model classifier, HMM)

HMM เป็นเทคนิคหนึ่งที่ใช้สำหรับการจำแนกข้อมูลสำหรับระบบรู้จำต่างๆ โดยการหาค่าลักษณะเฉพาะของสัญญาณจากคุณลักษณะทางสถิติ ซึ่ง HMM จัดเป็นคุณลักษณะหนึ่งของสัญญาณทางสถิติ HMM เหมาะสำหรับจัดการกับปัญหาของสัญญาณที่ขึ้นอยู่กับเวลา ซึ่งถูกพัฒนามาจากแบบจำลองมาร์คอฟ (Markov) ดังนั้น ก่อนที่จะเรียนรู้หลักการของ HMM จะอธิบายหลักการของแบบจำลองมาร์คอฟก่อน

สำหรับสัญญาณหนึ่งๆ เมื่อเราพิจารณาลำดับของสถานะ (sequence of states) ที่เวลาต่อเนื่องกัน สถานะที่เวลา  $t$  ถูกนิยามเป็น  $\omega(t)$  โดยที่ลำดับของสถานะทั้งหมดเท่ากับ  $T$  ซึ่งถูกนิยามเป็น  $\omega^T = \{\omega(1), \omega(1), \dots, \omega(T)\}$  กำหนดให้ความน่าจะเป็นในการเปลี่ยนจากสถานะหนึ่งไปยังอีกสถานะหนึ่ง (transition probability) คือ  $a_{ij}$  จะได้ว่า

$$a_{ij} = P(\omega_j(t+1)|\omega_i(t)) \quad (2.49)$$

$a_{ij}$  คือ ความน่าจะเป็นในการเปลี่ยนจากสถานะ  $\omega_i$  ที่เวลา  $t+1$

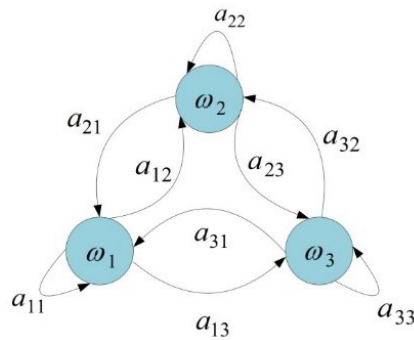
โดยที่  $a_{ij} \geq 0$  และ  $\sum_{j=1}^n a_{ij} = 1$  [93]

รูปที่ 2.9 แสดงแผนภาพสถานะแบบไม่ต่อเนื่องของแบบจำลองมาร์คอฟพื้นฐาน ซึ่งในที่นี้มี 3 สถานะ (states) คือ  $\omega_1, \omega_2$  และ  $\omega_3$  สมมุติกำหนดให้ลำดับข้อมูลในการเปลี่ยนสถานะของข้อมูล  $\omega^T = \{\omega(1), \omega(2), \dots, \omega(T)\} = \{\omega_1, \omega_3, \omega_3, \omega_1, \omega_2\}$  และ กำหนดให้  $\theta$  คือ แบบจำลองที่ต้องการหาความน่าจะเป็นของการลำดับข้อมูล (sequence) จะได้ว่า ความน่าจะเป็นที่แบบจำลองจะมีลำดับของข้อมูลเป็น  $\omega^T$  ข้างต้น คือ

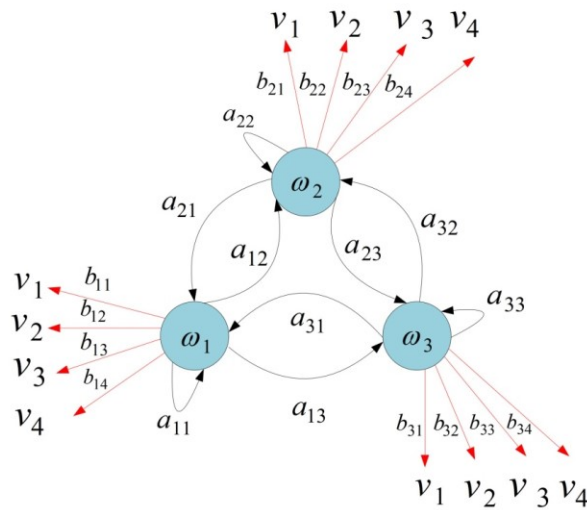
$$\begin{aligned} P(\omega^T|\theta) &= P(\omega_1)P(\omega_3|\omega_1)P(\omega_3|\omega_3)P(\omega_1|\omega_3)P(\omega_2|\omega_1) \\ &= \pi_1 a_{13} a_{33} a_{31} a_{12} \end{aligned} \quad (2.50)$$

โดยที่  $\pi_i$  คือ ความน่าจะเป็นที่สถานะเริ่มต้น (prior probability)

ยกตัวอย่างการใช้แบบจำลองมาร์คอฟกับการจำแนกคำพูด สมมุติมีสัญญาณของคำพูดคำว่า "hidden" กำหนดให้แต่ละสถานะแทนหน่วยเสียงของคำ ได้แก่ /h/ /i/ /d/ /e/ และ /n/ ซึ่งก็จะหาความน่าจะเป็นในการเปลี่ยนสถานะจาก /h/ ไปสถานะ /i/ จากสถานะ /i/ ไปเป็นสถานะ /d/ ตามลำดับ สำหรับแบบจำลองมาร์คอฟสามารถสังเกตเหตุการณ์ที่เกิดขึ้นในการเปลี่ยนสถานะได้ เหตุผลนี้ทำให้มีข้อจำกัดในการนำไปใช้มากขึ้นไป [94] ทำให้ต่อมาได้มีการพัฒนาไปเป็นแบบจำลองฮิดเดนมาร์คอฟ (hidden Markov หรือ HMM)



รูปที่ 2.9 แสดงแผนภาพสถานะแบบไม่ต่อเนื่องของแบบจำลองมาร์คอฟพื้นฐาน



รูปที่ 2.10 แสดงแผนภาพ HMM 3 สถานะ

สำหรับ HMM จะไม่สามารถสังเกตเหตุการณ์ที่เกิดขึ้นในการเปลี่ยนสถานะได้ มันจึงถูกเรียกว่า "Hidden" จากรูปที่ 2.10 แสดงแผนภาพ HMM ซึ่งในที่นี้มี 3 สถานะ คือ  $\omega_1, \omega_2$  และ  $\omega_3$  และ  $a_{ij}$  คือ ความน่าจะเป็นในการเปลี่ยนจากสถานะหนึ่งเป็นสถานะหนึ่ง ในแต่ละสถานะจะมีการปล่อย symbol 4 ตัว คือ  $v_1, v_2, v_3$  และ  $v_4$  ซึ่งความน่าจะเป็นที่แต่ละสถานะจะปล่อย symbol คือ  $b_{jk}$  จะสังเกตได้ว่าในแต่ละสถานะ  $\omega(t)$  จะมีการปล่อย  $v(t)$  ออกมา ซึ่งคือ symbol โดยที่สามารถสังเกตการเปลี่ยนสถานะของ symbol นี้ได้ จึงเรียกว่า สถานะที่มองเห็นได้ (visible state)

กำหนดให้ลำดับของสถานะที่มองเห็นได้ คือ  $V^T = \{v(1), v(2), \dots, v(T)\}$  สำหรับที่สถานะ  $\omega(t)$  จะมีความน่าจะเป็นของการปล่อย  $v_k(t)$  ซึ่งความน่าจะเป็นนี้ถูกเรียกว่า  $b_{jk}$  (emission probabilities)

$$b_{jk} = P(v_k(t) | \omega_j(t)) \quad (2.51)$$

$$\text{โดยที่ } \sum_{k=1}^n b_{jk} = 1$$

จะเห็นได้ว่าเราสามารถสังเกตเห็นการเปลี่ยนแปลงของลำดับของสถานะที่มองเห็นได้  $v_k(t)$  แต่ไม่สามารถสังเกตการณ์เปลี่ยนแปลงลำดับของสถานะซ่อน (hidden state) หรือ  $\omega_i$  ได้ จึงถูกเรียกว่า "Hidden Markov model"

สามารถสรุปค่าพารามิเตอร์ทั้งหมดที่จะใช้ในปัญหาในการคำนวณด้วย HMM ดังนี้

- 1) จำนวนสถานะในแบบจำลอง
- 2) จำนวนชนิดของข้อมูล (สถานะที่มองเห็นได้)
- 3) ค่าการกระจายความน่าจะเป็นของการเปลี่ยนสถานะ (transition probability)
- 4) ค่าการกระจายความน่าจะเป็นของการปล่อย symbol แต่ละสถานะ (emission probability)

สามารถกำหนดตัวแปรของแบบจำลองได้ดังนี้  $\lambda = (A, B, \pi)$  สำหรับตัวแปรที่แสดงจำนวนสถานะและชนิดของข้อมูล สามารถรู้ได้จากตัวแปร  $A$  และ  $B$

ประเด็นหลักในการคำนวณ HMM มี 3 ข้อคือ

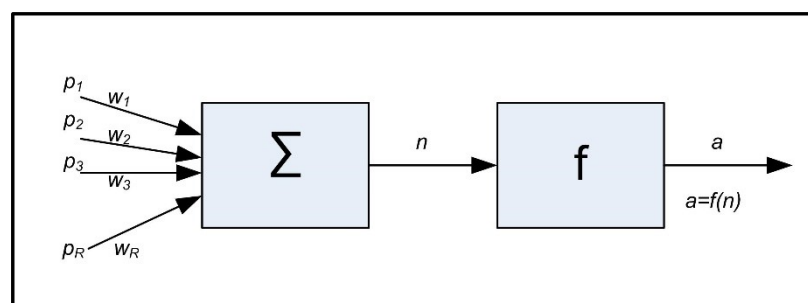
- 1) ปัญหาการประเมิน (evaluation problem) เป็นการหาความน่าจะเป็นของลำดับสถานะที่มองเห็นได้  $V^T$  ที่ถูกกำหนดโดยแบบจำลอง  $\lambda = (A, B, \pi)$
- 2) ปัญหาการถอดรหัส (decoding problem) เป็นการหาลำดับของสถานะซ่อน  $\omega^T$  หรือเส้นทางของลำดับสถานะที่ให้ค่าความเป็นไปมากที่สุด
- 3) ปัญหาการเรียนรู้ (learning problem) เป็นการปรับค่าพารามิเตอร์ต่างๆ ในแบบจำลอง เพื่อให้ได้ค่าความน่าจะเป็นสูงสุดโดยใช้ชุดข้อมูลในการฝึกฝนหลายๆ รอบ

## 2.9.2 ตัวจำแนกโครงข่ายประสาทเทียม (artificial neural network classifier, ANN)

โครงข่ายประสาทเทียม คือ แบบจำลองทางคณิตศาสตร์ที่ใช้ในการจำแนกกลุ่มของข้อมูลโดยเลียนแบบหลักการทำงานของเครือข่ายประสาทในสมองมนุษย์ ซึ่งในสมองประกอบด้วยเซลล์ประสาท และจุดประสานประสาท (synapse) ในเซลล์ประสาทแต่ละตัวประกอบด้วย 1) เดนไดรท์ (dendrite) เป็นอินพุต 2) ตัวเซลล์ (cell body) เป็นตัวประมวลผล และ 3) แอกซอน (axon) เป็นเอาต์พุตในการส่งสัญญาณประสาทไปยังเซลล์ประสาทตัวอื่น

รูปที่ 2.11 แสดงแบบจำลองเซลล์ประสาท ซึ่งประกอบด้วยอินพุต ( $p$ ) ซึ่งก็คือลักษณะเด่นที่จะป้อนให้กับระบบในการจำแนก ค่าน้ำหนัก ( $w$ ) ซึ่งเปรียบเสมือนกับจุดประสานประสาทในเซลล์ประสาทที่ปรับเปลี่ยนความเข้มข้นของสารละลายไปตามการเรียนรู้ของแต่ละบุคคล เช่นเดียวกับค่าน้ำหนักในแบบจำลองเซลล์ประสาทนี้จะเปลี่ยนค่าน้ำหนักไปตามอินพุตซึ่งเป็นส่วนสำคัญในกระบวนการเรียนรู้ ฟังก์ชันการรวม (summation function) เปรียบเสมือนกับตัวเซลล์ ซึ่งจะรวมผลคูณของค่าน้ำหนักกับอินพุตเข้าด้วยกัน ฟังก์ชันการแปลง (transfer function) จะนำเอาต์พุตที่ได้จากฟังก์ชันการรวม ( $n$ ) มาแปลงโดยใช้ฟังก์ชันการแปลงซึ่งให้เอาต์พุตออกมา ( $a$ ) ซึ่งเอาต์พุตนี้เปรียบเสมือนกับแอกซอนที่จะส่งสัญญาณไปยังเซลล์ประสาทอื่นๆ

สถาปัตยกรรมของตัวจำแนก ANN แบ่งออกเป็น โครงข่ายที่มีเซลล์ประสาทชั้นเดียว (single of neuron) และ โครงข่ายที่มีเซลล์ประสาทหลายชั้น (multiple layers of neurons) โครงข่ายที่มีเซลล์ประสาทชั้นเดียวจะมีจำนวนชั้นอินพุต 1 ชั้น ชั้นซ่อน 1 ชั้น (hidden layer) และ ชั้นเอาต์พุต 1 ชั้น โดยจำนวนเซลล์ในชั้นซ่อนอาจจะมีหลายเซลล์ (hidden cell) สำหรับโครงข่ายที่มีเซลล์ประสาทหลายชั้นจะมีชั้นซ่อนมากกว่า 1 ชั้น โครงข่ายประเภทนี้เหมาะสำหรับแก้ไขปัญหงานที่มีความซับซ้อน



รูปที่ 2.11 แบบจำลองเซลล์ประสาทชั้นเดียว

### การเรียนรู้ของตัวจำแนก ANN

1) การเรียนรู้แบบมีผู้สอน กระบวนการฝึกสอนระบบจำเป็นต้องใช้ชุดข้อมูลอินพุตและเอาต์พุตเป้าหมาย เพื่อใช้ในการปรับค่าน้ำหนักของโครงข่าย ซึ่งการปรับค่าน้ำหนักจะพิจารณาจากค่าความผิดพลาดที่น้อยที่สุดระหว่างเอาต์พุตของโครงข่ายที่คำนวณได้กับเอาต์พุตเป้าหมาย

2) การเรียนรู้แบบไม่มีผู้สอน กระบวนการฝึกสอนระบบจะใช้ชุดข้อมูลอินพุตเท่านั้น โดยจะใช้ค่าทางสถิติของชุดฝึกสอนในการเรียนรู้ และโครงข่ายประสาทเทียมจะหาเอาต์พุตของระบบเองโดยพิจารณาจากความสัมพันธ์ระหว่างข้อมูลอินพุตและเอาต์พุต

โครงข่ายประสาทเทียมมีหลายชนิด แต่สำหรับในงานวิจัยนี้จะใช้โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าโดยมีการเรียนรู้แบบย้อนกลับ (feed forward backpropagation) ซึ่งเป็นเทคนิคที่มีประสิทธิภาพและนิยมใช้กันในหลากหลายงานวิจัย กระบวนการนี้เป็นการฝึกสอนแบบมีผู้สอน และมีการเรียนรู้แบบย้อนกลับ ซึ่งกระบวนการเรียนรู้แบบย้อนกลับประกอบด้วย 2 ขั้นตอน ขั้นตอนแรก คือ การส่งผ่านไปข้างหน้า อินพุตจะถูกส่งผ่านเข้ามาในชั้นอินพุตแล้วส่งต่อไปยังชั้นซ่อน และส่งออกไปยังชั้นเอาต์พุต ขั้นตอนที่สอง คือ การส่งผ่านย้อนกลับ จะเป็นการส่งกลับค่าน้ำหนักที่คำนวณจากค่าความผิดพลาด ซึ่งคือผลต่างระหว่างเอาต์พุตเป้าหมายกับเอาต์พุตที่แท้จริง ซึ่งจะถูกส่งย้อนกลับในทิศทางตรงข้ามกับขั้นตอนแรก ค่าน้ำหนักจะถูกปรับจนค่าเอาต์พุตที่แท้จริงใกล้เคียงกับค่าเอาต์พุตเป้าหมายมากที่สุด

### 2.9.3 ตัวจำแนกค่าเฉลี่ยที่ใกล้ที่สุด (nearest mean classifier, NM)

ตัวจำแนกชนิดนี้ง่ายและเร็วในการประมวลผล ซึ่งจำแนกประเภทโดยพิจารณาค่าเฉลี่ยข้อมูลของแต่ละคลาส โดยระบุคลาสของชุดทดสอบที่ใกล้กับค่าเฉลี่ยของคลาสมากที่สุด กำหนดให้  $\mu_k$  คือ ค่าเฉลี่ยของข้อมูลคลาส  $\omega_k$

$$\mu_k = \frac{1}{n_k} \sum_{i \in I_k} \mathbf{x}_i \quad (2.52)$$

กำหนดให้  $\mathbf{z}$  คือ ข้อมูลของชุดทดสอบ  $\mathbf{z}$  จะถูกระบุว่าเป็นคลาส  $\omega_k$  เมื่อ

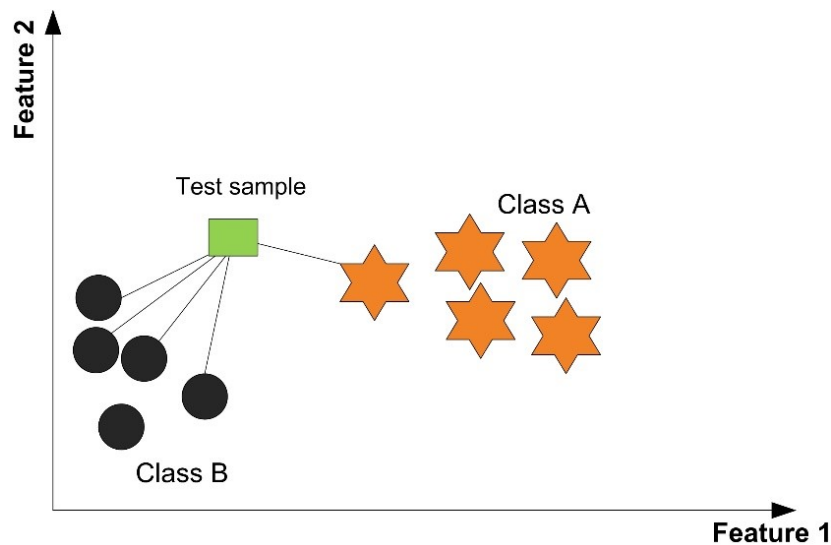
$$k = \arg \min_k d(\mu_k, \mathbf{z}) \quad (2.53)$$



ซึ่ง  $d(x, y)$  คือ ฟังก์ชันระยะทาง

#### 2.9.4 ตัวจำแนกเพื่อนบ้านใกล้ที่สุด K ตัว (K-nearest neighbor classifier, KNN)

จำแนกประเภทโดยพิจารณาข้อมูลที่ใกล้กับข้อมูลที่นำมาทดสอบมากที่สุด K ตัว ถ้าข้อมูล K ตัวนั้นอยู่ในกลุ่มใดมากที่สุด ข้อมูลที่นำมาทดสอบจะถูกจัดให้อยู่ในกลุ่มเดียวกัน รูปที่ 2.12 แสดงตัวอย่างการจำแนกโดยใช้ KNN โดยกำหนดให้  $K=5$  จากตัวอย่างมีข้อมูลทั้งหมด 2 กลุ่ม คือ กลุ่ม A และ กลุ่ม B ต้องการจัดกลุ่มข้อมูลทดสอบ พิจารณาข้อมูลที่ใกล้กับข้อมูลทดสอบที่สุด 5 ตัว (โดยคำนวณระยะทางระหว่างข้อมูลทดสอบกับข้อมูลในกลุ่ม A และ กลุ่ม B ทุกตัว) ซึ่งพบว่าข้อมูล 4 ตัวอยู่ในกลุ่ม B และข้อมูล 1 ตัวอยู่ในกลุ่ม A ดังนั้นข้อมูลทดสอบจะถูกจัดในกลุ่ม B สำหรับการเลือกค่า K ที่เหมาะสมไม่ควรเลือก K น้อยเกินไปเพราะจะทำให้เบี่ยงเบนสูง และไม่ควรเลือก K มากเกินไปเพราะจะทำให้ข้อมูลเกิดความลำเอียง วิธีหนึ่งในการเลือกค่า K คือ การตรวจสอบแบบไขว้ (cross validation) เป็นการแบ่งข้อมูลออกเป็นชุดฝึกสอนและชุดทดสอบ เลือกค่า K โดยพิจารณาค่าที่ทำให้จำนวนตัวอย่างในชุดทดสอบถูกจำแนกผิดพลาดน้อยที่สุด



รูปที่ 2.12 การจำแนกโดยตัวจำแนก KNN

### 2.9.5 ตัวจำแนกเชิงเส้นแบบเบย์นอร์มัล (linear Bayes normal classifier, LBN)

จำแนกประเภทโดยใช้หลักการความน่าจะเป็นตามกฎของเบย์ (Bayes' theorem) โดยที่ความหนาแน่นอย่างมีเงื่อนไขของคลาส (class-conditional density) มีการกระจายแบบปกติ (normal distribution) และเมตริกซ์โควาเรียนซ์ (covariance matrix,  $\Sigma$ ) เท่ากันทุกคลาส กฎของเบย์แสดงดังสมการที่ (2.54)

$$P(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j)P(\omega_j)}{p(\mathbf{x})} \quad (2.54)$$

$P(\omega_j)$  คือ ความน่าจะเป็นก่อนหน้า (prior probability) ของคลาส  $\omega_j$

$p(\mathbf{x} | \omega_j)$  คือ ความน่าจะเป็นที่จะเป็นคลาส  $\omega_j$  เมื่อพิจารณาลักษณะเด่น  $\mathbf{x}$  โดยทั่วไปเรียกว่า ไลค์ลิฮูด (likelihood) หรือ ฟังก์ชันความหนาแน่นของความน่าจะเป็นอย่างมีเงื่อนไข (class-conditional probability density function) หรือ ความหนาแน่นอย่างมีเงื่อนไขของคลาส

$P(\omega_j | \mathbf{x})$  คือ ความน่าจะเป็นหลัง (posterior probability)

$p(\mathbf{x})$  คือ ผลรวมของ  $p(\mathbf{x} | \omega_j)P(\omega_j)$  โดยทั่วไปเรียกว่า อีวิดนซ์เฟกเตอร์ (evidence factor) ซึ่งเป็นค่าคงที่ไม่มีผลต่อกระบวนการตัดสินใจจำแนกประเภท

สำหรับกฎการตัดสินใจ (decision rule) ในการจำแนกประเภท สมมติต้องการจำแนกข้อมูลที่มี 2 คลาส คือ  $\omega_1$  และ  $\omega_2$  จะตัดสินใจว่าข้อมูลที่ต้องการจำแนกเป็นคลาส  $\omega_1$  เมื่อ  $P(\omega_1 | \mathbf{x}) > P(\omega_2 | \mathbf{x})$  ในทางกลับกัน จะตัดสินใจว่าข้อมูลที่ต้องการจำแนกเป็นคลาส  $\omega_2$  เมื่อ  $P(\omega_2 | \mathbf{x}) > P(\omega_1 | \mathbf{x})$

สำหรับตัวจำแนก LBN ซึ่งจำแนกประเภทตามกฎของเบย์โดยสมมติให้มีการกระจายแบบปกติ นั้น จะใช้กฎการตัดสินใจแบบค่าสูงสุดของความน่าจะเป็นหลัง (Maximum A Posterior, MAP) ซึ่งเป็นการเลือกคลาสที่ให้ค่าความน่าจะเป็นหลังสูงสุด โดยที่เป็นการทำให้ค่าความผิดพลาดของความน่าจะเป็นต่ำสุด (minimum probability error) กฎการตัดสินใจดังกล่าว มีฟังก์ชันการจำแนก (discriminant function) ดังนี้ [95]

$$g_i(\mathbf{x}) = P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{\sum_{j=1}^c p(\mathbf{x} | \omega_j)P(\omega_j)} \quad (2.55)$$

$$g_i(\mathbf{x}) = p(\mathbf{x} | \omega_i)P(\omega_i) \quad (2.56)$$

$$g_i(\mathbf{x}) = \ln p(\mathbf{x} | \omega_i) + \ln P(\omega_i) \quad (2.57)$$

ผลการตัดสินใจจะแบ่งเนื้อที่ลักษณะเด่น (feature space) ไปเป็นพื้นที่ตัดสินใจ  $c$  กลุ่ม (decision regions) คือ  $R_1, \dots, R_c$  ถ้า  $g_i(\mathbf{x}) > g_j(\mathbf{x})$  สำหรับ  $j \neq i$  จะได้ว่า เวกเตอร์ลักษณะเด่น  $\mathbf{x}$  อยู่ในพื้นที่  $R_i$  ซึ่งผลการตัดสินใจ คือ เวกเตอร์ลักษณะเด่น  $\mathbf{x}$  ถูกจัดให้อยู่ในคลาส  $\omega_i$

โครงสร้างของตัวจำแนกโดยใช้หลักการของเบย์ (Bayes classifier) จะตัดสินใจโดยใช้  $p(\mathbf{x} | \omega_i)$  ถ้าสมมุติให้  $p(\mathbf{x} | \omega_i)$  มีการกระจายแบบปกติแบบหลายตัวแปร (multivariate normal)  $p(\mathbf{x} | \omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$  จะได้ว่า

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \quad (2.58)$$

สำหรับตัวจำแนก LBN สมมุติให้  $\boldsymbol{\Sigma}$  เท่ากันทุกคลาส ( $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ ) จะได้ฟังก์ชันการจำแนก ดังนี้

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \ln P(\omega_i) \quad (2.59)$$

โดยที่รูปแบบฟังก์ชันการจำแนกจะเป็นแบบเชิงเส้น

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0} \quad (2.60)$$

ซึ่ง  $\mathbf{w}_i = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i$  เป็นเวกเตอร์น้ำหนัก และ  $w_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + \ln P(\omega_i)$  เป็นค่าขีดแบ่ง

### 2.9.6 ตัวจำแนกควอดราติกเบย์นอร์มัล (quadratic Bayes normal classifier, QBN)

จำแนกประเภทโดยใช้หลักการความน่าจะเป็นตามกฎของเบย์เช่นเดียวกับตัวจำแนก LBN แตกต่างกันในส่วนของฟังก์ชันการจำแนกโดยตัวจำแนก QBN จะจัดอยู่ในกรณีทั่วไป ซึ่ง  $\Sigma_i \neq \Sigma_j$  [95]

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \quad (2.61)$$

ซึ่งรูปแบบฟังก์ชันการจำแนกเป็นฟังก์ชันแบบควอดราติก (quadratic function)

$$g_i(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0} \quad (2.62)$$

$$\text{ซึ่ง } \mathbf{W}_i = -\frac{1}{2} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i, \quad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i$$

$$\text{และ } w_{i0} = -\frac{1}{2} \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

### 2.9.7 ตัวจำแนกเชิงเส้นแบบลอจิสติก (logistic linear classifier, LOGL)

จำแนกประเภทโดยการทำให้ค่าไลค์ลิฮูดมีค่าสูงสุด (maximum likelihood) โดยการปรับค่าน้ำหนักของไลค์ลิฮูดจะใช้ฟังก์ชันการจำแนกลอจิสติก (logistic discriminant function) สำหรับปัญหาการจำแนกแบบหลายคลาส สมมุติฐานพื้นฐานพื้นฐานสำหรับการจำแนกข้อมูลซึ่งมีจำนวนคลาสเท่ากับ  $C$  แสดงดังสมการที่ (2.63) ซึ่งอัตราของล็อกไลค์ลิฮูด (log-likelihood ratio) ทุกคู่เป็นเชิงเส้น [96]

$$\log \left( \frac{p(\mathbf{x} | \omega_s)}{p(\mathbf{x} | \omega_C)} \right) = \beta_{s0} + \boldsymbol{\beta}_s^T \mathbf{x}, \quad s = 1, \dots, C-1 \quad (2.63)$$

สำหรับความน่าจะเป็นหลัง  $P(\omega_s | \mathbf{x})$  และ  $P(\omega_C | \mathbf{x})$  สามารถแสดงให้อยู่ในรูปแบบดังสมการที่ (2.64) และ (2.65)

$$P(\omega_s | \mathbf{x}) = \frac{\exp(\beta'_{s0} + \boldsymbol{\beta}_s^T \mathbf{x})}{1 + \sum_{s=1}^{C-1} \exp(\beta'_{s0} + \boldsymbol{\beta}_s^T \mathbf{x})}, \quad s = 1, \dots, C-1 \quad (2.64)$$

$$P(\omega_C | \mathbf{x}) = \frac{1}{1 + \sum_{s=1}^{C-1} \exp(\beta'_{s0} + \beta'_s{}^T \mathbf{x})} \quad (2.65)$$

$$\text{โดยที่ } \beta'_{s0} = \beta_{s0} + \log\left(\frac{P(\omega_s)}{P(\omega_C)}\right)$$

กฎการตัดสินใจในการจำแนกจะขึ้นอยู่กับฟังก์ชันเชิงเส้น  $\beta'_{s0} + \beta'_s{}^T \mathbf{x}$  เท่านั้น โดยจะตัดสินใจให้  $\mathbf{x}$  ถูกจัดอยู่ในคลาส  $\omega_j$  ถ้ามีเงื่อนไขดังต่อไปนี้ คือ

$$\max\{\beta_{s0} + \beta'_s{}^T \mathbf{x}\} = \beta_{j0} + \beta'_j{}^T \mathbf{x} > 0, \quad s=1, \dots, C-1 \quad (2.66)$$

ถ้าไม่ตรงตามเงื่อนไขนี้  $\mathbf{x}$  จะถูกจัดอยู่ในคลาส  $\omega_C$

กำหนดให้จำนวนข้อมูลสุ่มตัวอย่างเท่ากับ  $n_i$  และ จำนวนคลาสเท่ากับ  $C$  สามารถคำนวณไลค์ลิฮูด ( $L$ ) ได้ดังนี้

$$L = \prod_{i=1}^C \prod_{r=1}^{n_i} p(\mathbf{x}_{ir} | \omega_i) \quad (2.67)$$

จากสมการที่ (2.67) เมื่อพิจารณาการจำแนกแบบสองคลาส ( $\omega_1, \omega_2$ ) จะได้ว่า

$$L = \prod_{r=1}^{n_1} p(\mathbf{x}_{1r} | \omega_1) \prod_{r=1}^{n_2} p(\mathbf{x}_{2r} | \omega_2) \quad r=1, \dots, n_s; s=1, 2 \quad (2.68)$$

ซึ่งสามารถเขียนให้อยู่ในรูปของ

$$\begin{aligned} L &= \prod_{r=1}^{n_1} P(\omega_1 | \mathbf{x}_{1r}) \frac{p(\mathbf{x}_{1r})}{P(\omega_1)} \prod_{r=1}^{n_2} P(\omega_2 | \mathbf{x}_{2r}) \frac{p(\mathbf{x}_{2r})}{P(\omega_2)} \\ &= \frac{1}{P(\omega_1)^{n_1} P(\omega_2)^{n_2}} \prod_{\text{all } \mathbf{x}} p(\mathbf{x}) \prod_{r=1}^{n_1} P(\omega_1 | \mathbf{x}_{1r}) \prod_{r=1}^{n_2} P(\omega_2 | \mathbf{x}_{2r}) \end{aligned} \quad (2.69)$$

พารามิเตอร์  $\frac{1}{P(\omega_1)^{n_1} P(\omega_2)^{n_2}} \prod_{\text{all } \mathbf{x}} p(\mathbf{x})$  เป็นอิสระต่อโมเดลนี้ ดังนั้น การทำให้ไลค์ลิฮูด  $L$  มีค่าสูงสุดแสดงได้ดังสมการดังต่อไปนี้

$$L' = \prod_{r=1}^{n_1} P(\omega_1 | \mathbf{x}_{1r}) \prod_{r=1}^{n_2} P(\omega_2 | \mathbf{x}_{2r}) \quad \text{หรือ} \quad (2.70)$$

$$\log(L') = \sum_{r=1}^{n_1} \log(P(\omega_1 | \mathbf{x}_{1r})) + \sum_{r=1}^{n_2} \log(P(\omega_2 | \mathbf{x}_{2r})) \quad (2.71)$$

จากสมการที่ (2.71) เมื่อพิจารณาการจำแนกแบบหลายคลาสจะได้ว่า

$$\log(L') = \sum_{s=1}^C \sum_{r=1}^{n_s} \log(P(\omega_s | \mathbf{x}_{sr})) \quad (2.72)$$

หาค่าอนุพันธ์ของ  $\log(L')$  เพื่อหาค่าสูงสุดของล็อกไลค์ลิฮูดจะได้ว่า

$$\frac{\partial \log(L')}{\partial \beta'_{j0}} = n_j - \sum_{all \mathbf{x}} P(\omega_j | \mathbf{x}) \quad (2.73)$$

$$\frac{\partial \log(L')}{\partial (\beta'_j)_l} = \sum_{r=1}^{n_j} (x_{jr})_l - \sum_{all \mathbf{x}} P(\omega_j | \mathbf{x}) x_l \quad (2.74)$$

## 2.9.8 ตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้นของฟิชเชอร์ (Fisher's linear discriminant analysis classifier, FLDA)

เกณฑ์ของฟิชเชอร์ (Fisher's criterion) จะหาผลรวมเชิงเส้นของตัวแปร เพื่อแยกข้อมูลสองคลาสออกจากกันมากที่สุด โดยเกณฑ์ดังกล่าวจะพิจารณาอัตราระหว่างความแปรปรวนระหว่างคลาส (between-class) และภายในคลาส (within-class) ซึ่งเป้าหมายคือ ต้องหาทิศทางของ  $\mathbf{w}$  ที่ทำให้  $J_F$  มีค่าสูงสุด [96]

$$J_F = \frac{|\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)|^2}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \quad (2.75)$$

ซึ่ง  $\mathbf{m}_1, \mathbf{m}_2$  คือ ค่าเฉลี่ยของแต่ละกลุ่ม  $\mathbf{S}_W$  คือ เมตริกซ์โควาเรียนซ์การรวมกลุ่มภายในคลาสของข้อมูลตัวอย่าง (within-class sample covariance matrix)

การหาทิศทางของ  $\mathbf{w}$  ที่ทำให้  $J_F$  มีค่าสูงสุด ใช้วิธีการหาค่าอนุพันธ์ของ  $J_F$  เมื่อเทียบกับ  $\mathbf{w}$  แล้วกำหนดให้เท่ากับศูนย์ จะได้ว่า

$$\frac{\partial J_F}{\partial \mathbf{w}} = \frac{\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \left\{ 2(\mathbf{m}_1 - \mathbf{m}_2) + \frac{\mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \mathbf{S}_W \mathbf{w} \right\} = 0 \quad (2.76)$$

เนื่องจากสนใจทิศทางของ  $\mathbf{w}$  และ  $\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)/\mathbf{w}^T\mathbf{S}_w\mathbf{w}$  มีค่าเป็นสเกลาร์ ดังนั้น  $\mathbf{w} \propto \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$

สำหรับการจำแนกประเภทแบบหลายคลาสจะพิจารณาการทำให้การกระจายระหว่างกลุ่มมีค่าสูงสุดและการกระจายภายในกลุ่มที่ต่ำที่สุด โดยที่

$$J_F(\mathbf{a}) = \frac{\mathbf{a}^T \mathbf{S}_B \mathbf{a}}{\mathbf{a}^T \mathbf{S}_W \mathbf{a}} \quad (2.77)$$

โดยที่  $\mathbf{S}_B$  และ  $\mathbf{S}_W$  คือ

$$\mathbf{S}_B = \sum_{i=1}^C \frac{n_i}{n} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \quad (2.78)$$

$$\mathbf{S}_W = \sum_{i=1}^C \frac{n_i}{n} \hat{\Sigma}_i \quad (2.79)$$

$\mathbf{a}$  คือ เวกเตอร์ลักษณะเด่น  $\mathbf{m}_i$  และ  $\hat{\Sigma}_i$  ซึ่ง  $i=1, \dots, C$  คือ ค่าเฉลี่ยของตัวอย่าง (sample means) และเมตริกซ์โควาเรียนซ์ของแต่ละคลาส

### 2.9.9 ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีน (support vector machine classifier, SVM)

จำแนกประเภทโดยอาศัยหลักการที่หาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งกลุ่มข้อมูล โดยหาเส้นแบ่งกลุ่มที่สามารถแยกกลุ่มข้อมูลได้ดีที่สุด เส้นที่สามารถแบ่งกลุ่มได้ดีที่สุดคือ เส้นที่มีค่ามารจิน (margin) สูงสุด โดยที่มารจิน คือ ผลรวมของระยะห่างระหว่างเส้นแบ่งกลุ่มที่พิจารณาและเส้นที่ผ่านข้อมูลที่ใกล้ที่สุด ซึ่งข้อมูลที่อยู่บนเส้นที่ใกล้ที่สุดที่ถูกพิจารณาเรียกว่า ซัพพอร์ตเวกเตอร์ (support vector)

ปกติแล้ว SVM ถูกนำมาใช้กับข้อมูลที่เป็นเชิงเส้น อย่างไรก็ตามข้อมูลที่ถูกนำมาใช้ในการจำแนกส่วนใหญ่เป็นแบบไม่เชิงเส้น ดังนั้นจึงนำฟังก์ชันเคอร์เนล (kernel function) มาใช้ในการแก้ปัญหา

ในการจำแนกแบบไม่เป็นเชิงเส้นจะต้องแปลงกลุ่มข้อมูลตัวอย่างไปสู่ปริภูมิมิติที่สูงขึ้นโดยผ่านฟังก์ชันที่ไม่เป็นเชิงเส้นซึ่งเรียกว่า ฟังก์ชันเคอร์เนล สำหรับตัวอย่างของฟังก์ชันเคอร์เนล ได้แก่

ฟังก์ชันเคอร์เนลแบบโพลีโนมเมียล (polynomial)

$$K(x, y) = (x^T y + 1)^d \quad (2.80)$$

ฟังก์ชันเคอร์เนลแบบเอกโพเนนเชียล (exponential)

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \text{ ซึ่ง } 2\sigma^2 \text{ เป็นพารามิเตอร์อิสระ} \quad (2.81)$$

ฟังก์ชันเคอร์เนลแบบระยะทาง (distance)

$$K(x, y) = \|x - y\|^d \quad (2.82)$$

ซึ่งในงานวิจัยนี้ใช้ฟังก์ชันเคอร์เนลแบบระยะทางโดยที่  $d = 4$

## 2.10 การสังเคราะห์เสียงพูด

กระบวนการสังเคราะห์เสียงพูด คือ กระบวนการสร้างเสียงเทียมของเสียงพูดของมนุษย์ [97] โดยระบบคอมพิวเตอร์ที่ถูกใช้สำหรับวัตถุประสงค์นี้ เรียกว่าตัวสังเคราะห์เสียง (speech synthesizer) ปัจจุบันทฤษฎีสำหรับการสังเคราะห์เสียงพูดสามารถแบ่งออกเป็น 3 ชนิด [97] ได้แก่

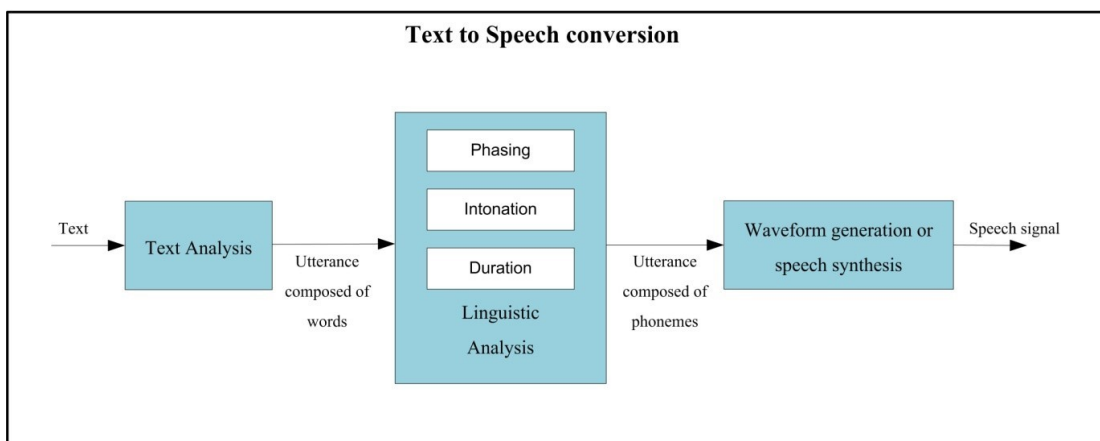
1) การสังเคราะห์เสียงโดยอาศัยการแปลงรหัสรูปคลื่นสัญญาณ (waveform coding) ซึ่งจะใช้สัญญาณเสียงพูดของมนุษย์ในช่วงสั้นๆ เช่น เสียงคำ หรือ วลี ที่ถูกบันทึกไว้ และสังเคราะห์เสียงที่เป็นประโยชน์ออกมา โดยอาศัยการต่อกันของเสียงพูดดังกล่าว

2) การสังเคราะห์เสียงโดยอาศัยทฤษฎีการวิเคราะห์และสังเคราะห์ (analysis-synthesis method) วิธีนี้จะแก้ปัญหาเรื่องขนาดของความจำของข้อมูลสำหรับวิธีแรกซึ่งต้องใช้เนื้อที่ในการเก็บข้อมูลมาก ซึ่งวิธีนี้จะแปลงคำหรือวลีของเสียงพูดของมนุษย์ไปเป็นค่าพารามิเตอร์ซึ่งถูกวิเคราะห์โดยใช้โมเดลการสร้างเสียงพูด (speech production model) โดยประโยคที่ถูกสังเคราะห์เสียงพูดออกมาจะมาจากการเชื่อมต่อกันของพารามิเตอร์ดังกล่าว แต่วิธีนี้จะลดความเป็นธรรมชาติของเสียงมากกว่าวิธีก่อนหน้า



3) การสังเคราะห์เสียงโดยอาศัยกฎ (synthesis by rule) วิธีนี้จะสร้างเสียงพูด โดยการใช้ลำดับของสัญลักษณ์หรือตัวอักษรของหน่วยเสียงหรือพยางค์เป็นอินพุต ซึ่งเสียงพูดที่สังเคราะห์ออกมาจะมาจากพารามิเตอร์ของหน่วยเสียงที่นำมาต่อกันด้วยกฎทางเสียงและภาษา

กระบวนการแปลงข้อความเป็นเสียงพูด (text to speech conversion) เป็นเป้าหมายหลักสำหรับงานด้านการสังเคราะห์เสียงในปัจจุบัน กระบวนการนี้เป็นกระบวนการสังเคราะห์เสียงโดยอาศัยกฎ การแปลงข้อความเป็นเสียงพูดนี้ถูกใช้ประยุกต์ใช้งานในหลากหลายงาน เช่น การเข้าถึงข้อมูลของจดหมายอิเล็กทรอนิกส์โดยรับข้อมูลผ่านเสียง การแปลงข้อความในเอกสารต่างๆ เป็นภาษาพูดซึ่งเหมาะสำหรับผู้ใช้ที่ตาบอด เป็นต้น กระบวนการนี้จะประกอบด้วย 2 ส่วนหลัก [98] ในส่วนแรกจะเป็นการนอร์มัลไลซ์ข้อความ ซึ่งจะประกอบด้วย การวิเคราะห์ข้อความ (text analysis) และส่วนประมวลผลจังหวะและทำนอง (prosody analysis) สำหรับในส่วนที่สองจะเป็นส่วนสังเคราะห์เสียงพูด (speech synthesis)



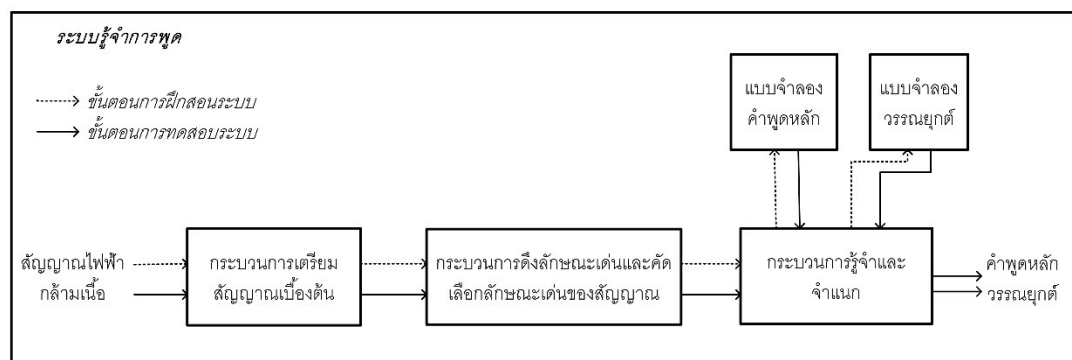
รูปที่ 2.13 ระบบสร้างเสียงพูดโดยใช้การแปลงข้อความเป็นเสียงพูด

### บทที่ 3 วิธีการวิจัย

ในบทนี้จะกล่าวถึง วิธีการและขั้นตอนการวิจัย เพื่อสร้างระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งในบทนี้จะเป็นการออกแบบและทดลองเพื่อสร้างโครงสร้างหลักของระบบในส่วนของระบบรู้จำ ได้แก่ การออกแบบรูปแบบการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อซึ่งรวมถึงวัสดุและอุปกรณ์ที่ใช้ในการทดลอง รูปแบบการเตรียมสัญญาณเบื้องต้น รูปแบบการสกัดลักษณะเด่นและการประเมินลักษณะเด่น และรูปแบบในการจำแนกของตัวจำแนก หลังจากการออกแบบโครงสร้างหลักของระบบ จะทำการปรับปรุงประสิทธิภาพของระบบเพื่อให้ได้ความแม่นยำในการจำแนกที่ดีที่สุดเท่าที่จะเป็นไปได้ ซึ่งในส่วนของการปรับปรุงระบบจะกล่าวในบทที่ 4

#### 3.1 การออกแบบรูปแบบการเก็บข้อมูลของสัญญาณ

ในหัวข้อนี้จะกล่าวถึงการออกแบบและทดลองเพื่อหาตำแหน่งอิเล็กโทรดที่เหมาะสมสำหรับการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ การออกแบบอุปกรณ์เพื่อเก็บข้อมูลของสัญญาณทั้งในส่วนที่เป็นสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อ และออกแบบขั้นตอนการเก็บข้อมูลทั้งหมด



รูปที่ 3.1 โครงสร้างหลักของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ในส่วน  
ของระบบรู้จำ

### 3.1.1 การออกแบบตำแหน่งอิเล็กโทรด

อินพุตของระบบรู้จำการพูดในงานวิจัยนี้ คือ สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งสัญญาณไฟฟ้ากล้ามเนื้อที่ถูกใช้ในงานวิจัยนี้ต้องเป็นสัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากกล้ามเนื้อที่เกี่ยวข้องกับการพูดภาษาไทย ดังนั้นสิ่งสำคัญอันดับแรกสำหรับงานวิจัยนี้ คือ การออกแบบตำแหน่งอิเล็กโทรดสำหรับตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ

อุปกรณ์ที่ใช้ในการทดลองเบื้องต้นเพื่อออกแบบตำแหน่งของอิเล็กโทรด ได้แก่ เครื่องขยายและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ Mobi6 ซึ่งผลิตโดยบริษัท TMS international ประเทศเนเธอร์แลนด์ โดยสามารถบันทึกสัญญาณไฟฟ้ากล้ามเนื้อได้ 6 ช่องสัญญาณ อิเล็กโทรดแบบผิวหน้าชนิด Ag/AgCl ยี่ห้อ ARBO ซึ่งเป็นอิเล็กโทรดแบบมีเจลอยู่ภายในและใช้ครั้งเดียวทิ้ง มีขนาดเส้นผ่านศูนย์กลางภายนอก 2.4 เซนติเมตร และขนาดเส้นผ่านศูนย์กลางภายใน (ส่วนของเจล) 1 เซนติเมตร ซึ่งผลิตโดยบริษัท Tyco healthcare Group LP ประเทศเยอรมัน



รูปที่ 3.2 เครื่องขยายและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อยี่ห้อ Mobi6



รูปที่ 3.3 อิเล็กโทรดแบบผิวหน้า Ag/AgCl ชนิดใช้แล้วทิ้งยี่ห้อ ARBO

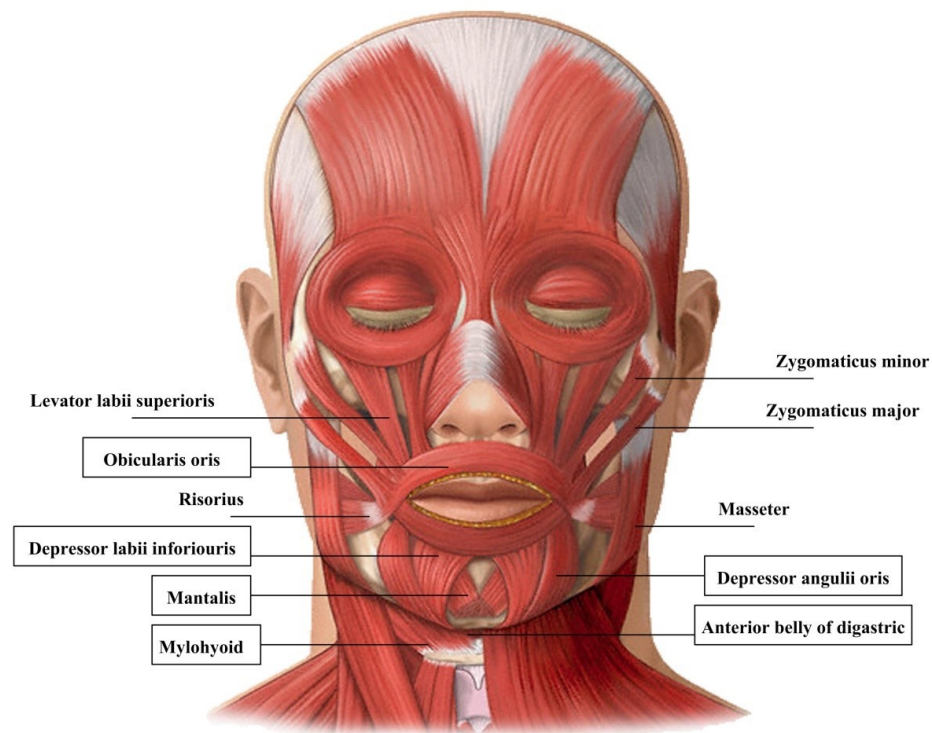
### ขั้นตอนการออกแบบตำแหน่งอิเล็กโทรด

1) ศึกษากลไกการพูดและกล้ามเนื้อที่มีผลสำหรับการออกเสียงพูดภาษาไทย ซึ่งพบว่ากล้ามเนื้อเหล่านี้ส่วนใหญ่อยู่บริเวณใบหน้าและลำคอ ซึ่งมีอยู่จำนวนมากและเป็นกล้ามเนื้อขนาดเล็ก ชับซ้อน และกล้ามเนื้อบางมัดอยู่ลึกเข้าไปซึ่งยากต่อการตรวจจับโดยใช้อิเล็กโทรดแบบผิวหน้า

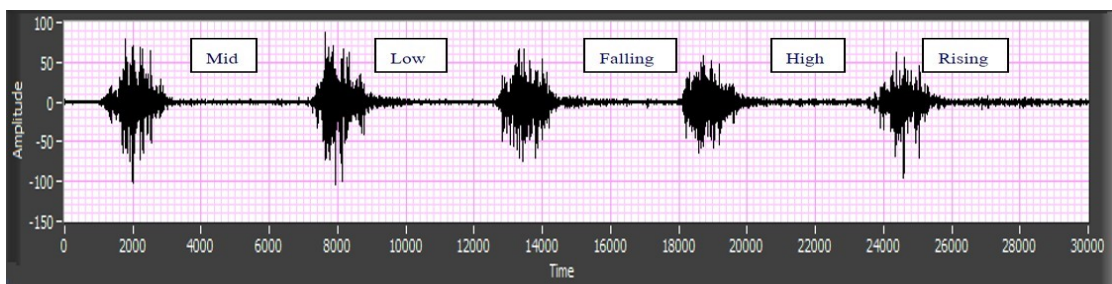
2) เลือกตำแหน่งของกล้ามเนื้อที่มีผลต่อการออกเสียงพูดภาษาไทยซึ่งเป็นตำแหน่งของกล้ามเนื้อที่มีขนาดไม่เล็กจนเกินไป ซึ่งสามารถจับสัญญาณได้ง่าย จากการศึกษาพบว่ากล้ามเนื้อที่มีขนาดใหญ่และมีการซ้อนทับกับกล้ามเนื้อชนิดอื่นน้อย และสามารถจับสัญญาณได้ง่ายด้วยอิเล็กโทรดแบบผิวหน้ามีทั้งหมด 11 มัดกล้ามเนื้อดังรูปที่ 3.4

3) ทำการทดลองจับสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 11 ตำแหน่ง จากอาสาสมัคร 2 คน ซึ่งเป็นผู้หญิง 1 คน และ ผู้ชาย 1 คน อายุระหว่าง 20-30 ปี โดยใช้อิเล็กโทรดแบบผิวหน้า ขณะบันทึกสัญญาณไฟฟ้ากล้ามเนื้ออาสาสมัครจะออกเสียงพูดหน่วยเสียงภาษาไทยทั้งหมด 35 หน่วยเสียง ซึ่งครอบคลุมเสียงพยัญชนะ 21 หน่วยเสียง เสียงสระ 9 หน่วยเสียง และเสียงวรรณยุกต์ 5 เสียง พิจารณาลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากการตรวจจับสัญญาณทั้ง 11 ตำแหน่ง พบว่าบางตำแหน่งมีขนาดของสัญญาณสูง บางตำแหน่งมีขนาดของสัญญาณต่ำมาก ซึ่งตำแหน่งที่ให้ขนาดของสัญญาณต่ำมากไม่เหมาะสมในการนำมาใช้ในงานวิจัยนี้ เนื่องจากขนาดของสัญญาณที่ต่ำนั้น ทำให้ไม่สามารถแยกแยะความแตกต่างของคำพูดแต่ละคำได้อย่างชัดเจน จากการทดลองนี้พบว่าตำแหน่งของอิเล็กโทรดที่ใช้ในการจับสัญญาณไฟฟ้ากล้ามเนื้อที่เหมาะสมมีอยู่ 6 ตำแหน่ง ซึ่งเป็นตำแหน่งที่ให้ขนาดของสัญญาณสูง ได้แก่ ตำแหน่งของกล้ามเนื้อ

1) Orbicularis Oris 2) Depressor Labii Inferioris 3) Mentalis 4) Depressor Anguli Oris 5) Anterior Belly of Digastric และ 6) Mylohyoid ซึ่งแสดงดังรูปที่ 3.4 (มัดกล้ามเนื้อที่อยู่ในกรอบสี่เหลี่ยม)

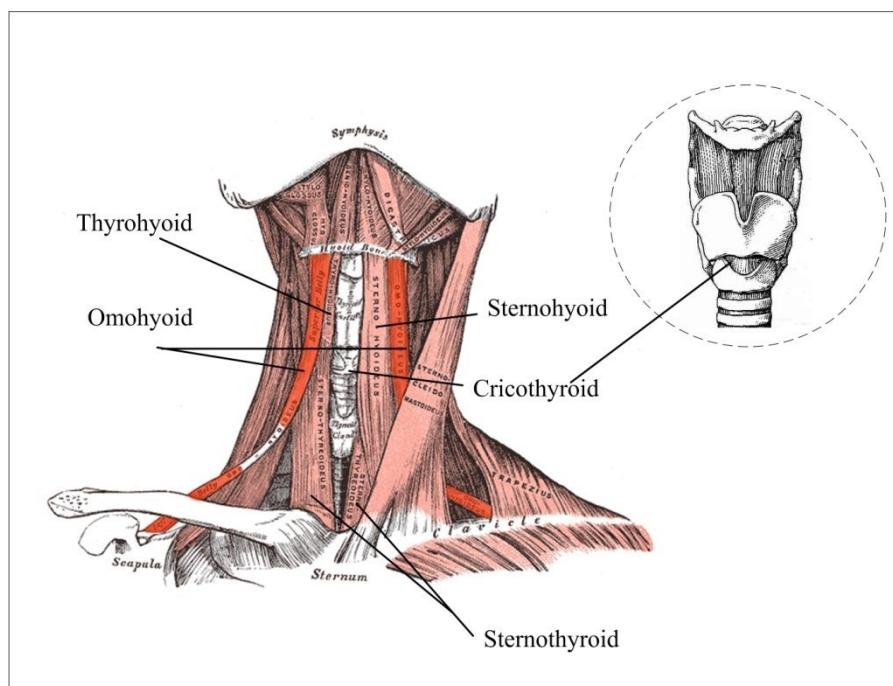


รูปที่ 3.4 ตำแหน่งของกล้ามเนื้อที่เกี่ยวข้องกับการสร้างเสียงพูด กล้ามเนื้อที่อยู่ในกรอบสี่เหลี่ยมเป็นกล้ามเนื้อที่ถูกคัดเลือก [69]



รูปที่ 3.5 สัญญาณไฟฟ้ากล้ามเนื้อของการพูดเสียงวรรณยุกต์ 5 เสียง "ดู, ดู, ดู, ดู, ดู" ซึ่งได้จากกล้ามเนื้อบริเวณหน้า (Depressor Anguli Oris)

อย่างไรก็ตาม จากการทดลองพบว่าตำแหน่งของกล้ามเนื้อทั้ง 6 ตำแหน่งนั้นไม่สามารถแยกเสียงวรรณยุกต์ได้ กล่าวคือ ลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อของเสียงวรรณยุกต์แต่ละเสียงมีความคล้ายคลึงกันมาก (ดังรูปที่ 3.5) ดังนั้นจึงต้องทำการทดลองเพื่อหาตำแหน่งของสัญญาณไฟฟ้ากล้ามเนื้อสำหรับการแยกเสียงวรรณยุกต์โดยเฉพาะ



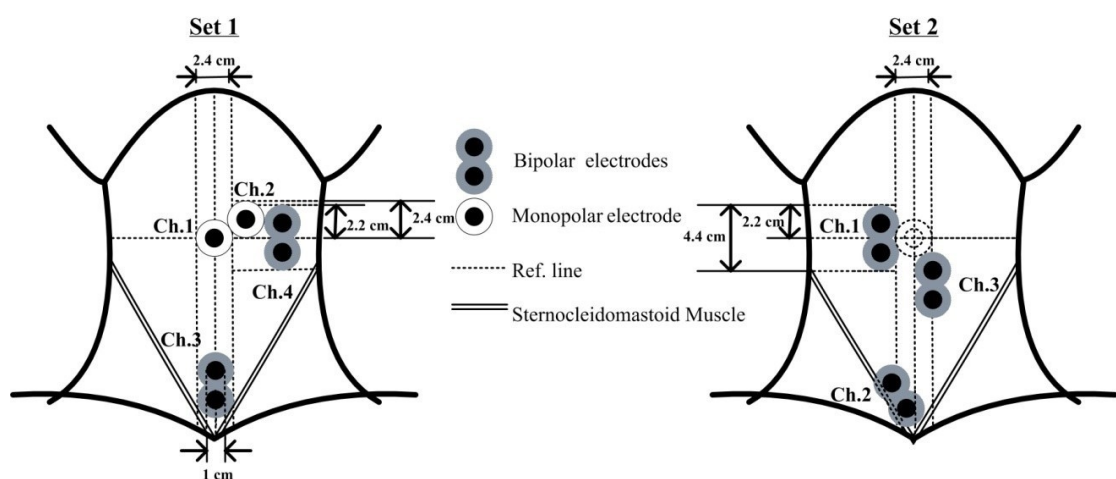
**รูปที่ 3.6** กล้ามเนื้อบริเวณคอ (strap muscles) ซึ่งมีผลในการออกเสียงวรรณยุกต์ [99]

4) ทำการทดลองเพื่อหาตำแหน่งของอิเล็กโทรดสำหรับการจับสัญญาณไฟฟ้ากล้ามเนื้อเพื่อแยกเสียงวรรณยุกต์ จากการศึกษาพบว่ากล้ามเนื้อที่มีผลในการออกเสียงวรรณยุกต์คือ กล้ามเนื้อบริเวณคอ (strap muscles) ดังรูปที่ 3.6 ซึ่งสอดคล้องกับงานวิจัยของ Erickson [100] โดยเขาได้จับสัญญาณไฟฟ้ากล้ามเนื้อบริเวณคอโดยใช้อิเล็กโทรดแบบเข็ม (hooked-wire electrode) งานวิจัยดังกล่าวพบว่ารูปแบบของสัญญาณไฟฟ้ากล้ามเนื้อบริเวณคอสัมพันธ์กับลักษณะของรูปแบบความถี่มูลฐานของวรรณยุกต์ไทย (fundamental frequency) ดังนั้นงานวิจัยนี้จึงอาศัยแนวคิดนี้ เพื่อหาตำแหน่งอิเล็กโทรดที่เหมาะสมสำหรับการแยกเสียงวรรณยุกต์ โดยใช้อิเล็กโทรดแบบผิวหนัง

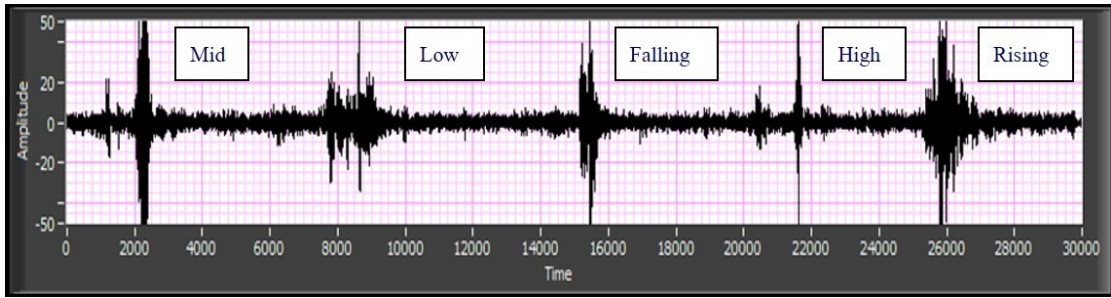
เนื่องจากกล้ามเนื้อบริเวณคอเป็นกล้ามเนื้อที่มีลักษณะการซ้อนทับกันอยู่ กล่าวคือ ณ ตำแหน่งใดๆ ของผิวหนังจะมีกล้ามเนื้อหลายชนิดซ้อนกันอยู่ ดังนั้นจึงไม่สามารถระบุตำแหน่งของอิเล็กโทรดเป็นชื่อมัดกล้ามเนื้อได้ การทดลองนี้จึงได้ระบุตำแหน่งของอิเล็กโทรดดังรูปที่ 3.7 ซึ่งมีทั้งหมด 7 ตำแหน่ง อย่างไรก็ตามอุปกรณ์ที่ใช้ในการขยายสัญญาณและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ (Mobi 6) สามารถตรวจวัดสัญญาณได้เพียง 6 ช่องสัญญาณ ดังนั้นจึงแบ่งการทดลองออกเป็น 2 ชุด โดยชุดที่ 1 ตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ 4 ตำแหน่ง และ ชุดที่ 2 ตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อที่เหลืออีก 3 ตำแหน่ง โดยใช้อิเล็กโทรดแบบผิวหนังทั้งแบบขั้วเดียว (monopolar electrode) และแบบสองขั้ว (bipolar electrode) ขึ้นอยู่กับขนาดของกล้ามเนื้อ

ที่ต้องการวัด ถ้าเป็นกล้ามเนื้อขนาดเล็กจะใช้อิเล็กโทรดแบบขั้วเดียว ถ้าเป็นกล้ามเนื้อขนาดใหญ่ จะใช้อิเล็กโทรดแบบสองขั้ว

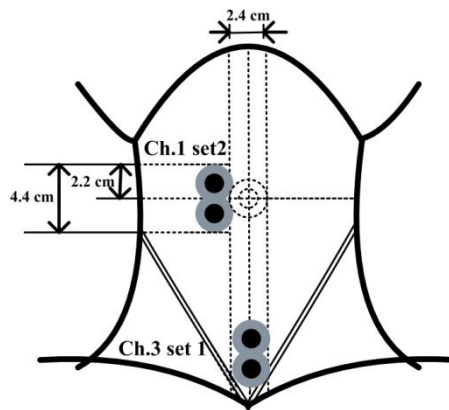
ทำการทดลองวัดสัญญาณไฟฟ้ากล้ามเนื้อจากตำแหน่งกล้ามเนื้อคอทั้งหมด 7 ตำแหน่ง จากอาสาสมัคร 1 คน ซึ่งเป็นเพศหญิงอายุ 25 ปี ซึ่งพูดภาษาไทยเป็นภาษาแม่ อาสาสมัครออกเสียงพูดคำโดดทั้งหมด 9 คำ ("กา", "ใจ", "ดู", "เบ", "อ", "ดี", "เป", "กาน", "กาง") ซึ่งแต่ละคำออกเสียงวรรณยุกต์ทั้ง 5 เสียง พิจารณาลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อจาก ตำแหน่งกล้ามเนื้อคอ พบว่าสัญญาณไฟฟ้ากล้ามเนื้อของเสียงวรรณยุกต์แต่ละเสียงมีลักษณะที่ แตกต่างกันได้ชัด ดังรูปที่ 3.8 นั้นแสดงว่าสัญญาณไฟฟ้ากล้ามเนื้อคอสามารถแยกเสียง วรรณยุกต์ได้ ซึ่งจะทดสอบความสามารถในการแยกเสียงวรรณยุกต์ของสัญญาณไฟฟ้ากล้ามเนื้อ แต่ละตำแหน่งโดยการพิจารณาจากกราฟการกระจาย กล่าวคือ นำสัญญาณไฟฟ้ากล้ามเนื้อแต่ ละตำแหน่ง มาคำนวณลักษณะเด่นของสัญญาณ ซึ่งการทดลองนี้เลือกใช้ลักษณะเด่น 8 ชนิด คือ RMS VAR WL WAMP MDF SM1 SM2 และ SM3 หลังจากนั้น พล็อตกราฟการกระจายของ ลักษณะเด่นทั้งหมดเพื่อจัดกลุ่มวรรณยุกต์ พิจารณาความสามารถในการแยกแยะข้อมูลหรือ ความสามารถในการจัดกลุ่มของวรรณยุกต์แต่ละเสียงจากกราฟการกระจาย จากการทดลอง พบว่าสัญญาณไฟฟ้ากล้ามเนื้อจากตำแหน่งกล้ามเนื้อคอสามารถแยกเสียงวรรณยุกต์ได้โดย เลือกใช้ลักษณะเด่นที่เหมาะสม ซึ่งตำแหน่งที่สามารถแยกเสียงวรรณยุกต์ได้ดีที่สุด คือ ตำแหน่ง บริเวณข้างลูกกระเดือก (สัญญาณช่องที่ 1 ของกล้ามเนื้อชุดที่ 2) และตำแหน่งส่วนล่างของลำคอ (สัญญาณช่องที่ 3 ของกล้ามเนื้อชุดที่ 1) ดังรูปที่ 3.9



รูปที่ 3.7 ตำแหน่งของอิเล็กโทรดที่ใช้สำหรับการทดลองเพื่อเลือกตำแหน่งที่เหมาะสมสำหรับการแยกเสียงวรรณยุกต์



รูปที่ 3.8 สัญญาณไฟฟ้ากล้ามเนื้อจากกล้ามเนื้อบริเวณส่วนล่างของลำคอสำหรับการออกเสียงวรรณยุกต์แต่ละเสียง



รูปที่ 3.9 ตำแหน่งอิเล็กโทรดที่ดีที่สุดจากการทดลอง เพื่อใช้ในการแยกเสียงวรรณยุกต์

ดังนั้นตำแหน่งของอิเล็กโทรดที่ถูกเลือกมีทั้งหมด 8 ตำแหน่ง คือ ตำแหน่งกล้ามเนื้อหน้าและใต้คาง 6 ตำแหน่ง และกล้ามเนื้อคอ 2 ตำแหน่ง ซึ่งในส่วนของกล้ามเนื้อคอใช้สำหรับการแยกวรรณยุกต์โดยเฉพาะ อย่างไรก็ตามจำนวนช่องสัญญาณของ Mobi6 ที่ใช้ในการตรวจจับและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อมีเพียง 6 ช่องสัญญาณ และ 1 ช่องสัญญาณจะถูกใช้สำหรับการรับสัญญาณการกดสวิตช์เพื่อกำหนดขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อ ดังนั้นในงานวิจัยนี้จำเป็นต้องลดจำนวนอิเล็กโทรดให้เหลือเพียง 5 ตำแหน่ง

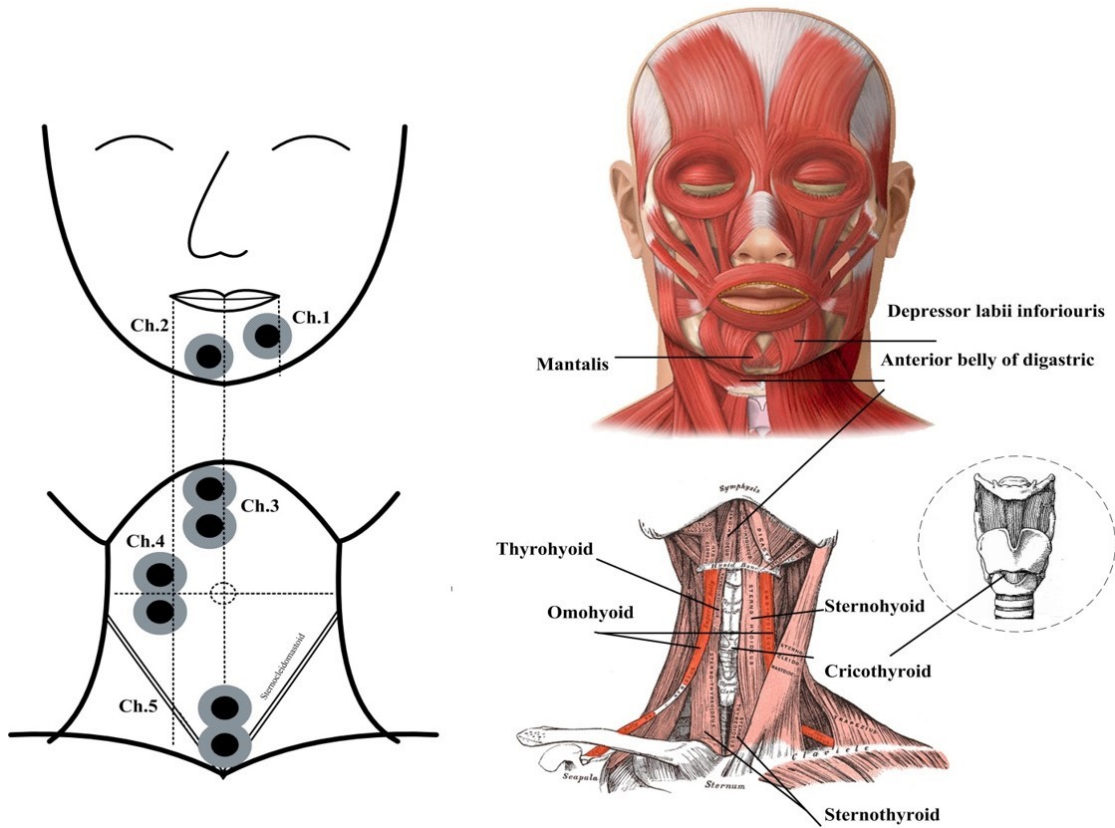
5) เลือกตำแหน่งของอิเล็กโทรดให้เหลือเพียง 5 ตำแหน่ง ซึ่งครอบคลุมตำแหน่งการแยกเสียงวรรณยุกต์ 2 ตำแหน่ง ดังนั้นตำแหน่งอิเล็กโทรดบริเวณกล้ามเนื้อหน้าและใต้คางที่ถูกเลือกไว้ 6 ตำแหน่งจะถูกลดเหลือเพียง 3 ตำแหน่ง โดยการเลือกจะพิจารณาจากหน้าที่ของกล้ามเนื้อในแต่ละมัดกล้ามเนื้อ และจัดกลุ่มกล้ามเนื้อที่มีหน้าที่การทำงานคล้ายคลึงกัน ซึ่งสามารถจัดกลุ่มกล้ามเนื้อที่คล้ายคลึงกันดังตารางที่ 3.1 ในกลุ่มที่ 1 เลือก Depressor Labii Inferioris เนื่องจากสัญญาณมีแอมพลิจูดสูงกว่าและมีความชัดเจนของสัญญาณมากกว่า ในกลุ่ม



ที่ 2 เลือก Mentalis เพราะ Orbicularis Oris มีสัญญาณรบกวนเยอะและไม่ค่อยเสถียรขณะตรวจจับสัญญาณ สำหรับกลุ่มที่ 3 เลือก Anterior Belly of Digastric เพราะกล้ามเนื้อนี้ถูกนำมาใช้ในหลายงานวิจัย และเป็นกล้ามเนื้อหลักที่ทุกงานวิจัยต้องมี ดังนั้นตำแหน่งอิเล็กโทรด 5 ตำแหน่งที่จะถูกใช้ในงานวิจัยนี้แสดงดังรูปที่ 3.10 ซึ่งสัญญาณช่องที่ 4 และช่องที่ 5 จะถูกใช้สำหรับการจำแนกวรรณยุกต์ สำหรับสัญญาณช่องที่ 1-5 ถูกใช้สำหรับการจำแนกคำพูดหลัก (ไม่สนใจวรรณยุกต์) สัญญาณช่องที่ 1 และ ช่องที่ 2 ใช้อิเล็กโทรดแบบขั้วเดียว ซึ่งจะวัดสัญญาณไฟฟ้ากล้ามเนื้อเทียบกับตำแหน่งอ้างอิง โดยตำแหน่งอ้างอิงจะวัดที่ตำแหน่งผิวหนังบริเวณกระดูกหลังหู สำหรับสัญญาณช่องที่ 3-5 ใช้อิเล็กโทรดแบบสองขั้ว สำหรับตำแหน่งของกราวด์ของระบบจะวัดที่ตำแหน่งของผิวหนังบริเวณข้อมือ

### ตารางที่ 3.1 การจัดกลุ่มของมัดกล้ามเนื้อที่มีหน้าที่คล้ายกัน

กลุ่ม	กล้ามเนื้อ	หน้าที่	เลือก
1	Depressor Anguli Oris (DAO)	ดึงมุมปากต่ำลง ทำงานร่วมกับ DLI	
	Depressor Labii Inferioris (DLI)	ดึงริมฝีปากล่างลงและออกไปด้านข้าง	✓
2	Mentalis (MTL)	ย่นริมฝีปากและทำปากเป็นวงกลม	✓
	Orbicularis Oris (OCO)	การปิดริมฝีปาก หรือริมฝีปากเป็นวงกลม	
3	Anterior Belly of Digastric (ABD)	ดึงกระดูก Hyoid ขึ้น และไปข้างหน้า	✓
	Mylohyoid (MLH)	ยกกระดูก Hyoid	



รูปที่ 3.10 สรุปตำแหน่งอิเล็กโทรด 5 ตำแหน่งที่ถูกคัดเลือก [69], [99]

### 3.1.2 การออกแบบอุปกรณ์และวิธีการเก็บข้อมูลของสัญญาณ

หลังจากการออกแบบตำแหน่งอิเล็กโทรดสำหรับตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ ขั้นตอนต่อไปเป็นการออกแบบอุปกรณ์เพื่อใช้ในการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งอุปกรณ์มีดังนี้

1) เครื่องขยายและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อยี่ห้อ Mobie6 มีทั้งหมด 6 ช่องสัญญาณ 5 ช่องสัญญาณแรกใช้สำหรับการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อจาก 5 ตำแหน่งตามที่ได้ออกแบบไว้ สำหรับสัญญาณช่องสุดท้ายถูกใช้สำหรับการรับสัญญาณการกดสวิทช์

2) อิเล็กโทรดแบบผิวหน้าชนิด Ag/AgCl ยี่ห้อ ARBO ซึ่งเป็นอิเล็กโทรดแบบมีเจลอยู่ภายในและใช้ครั้งเดียวทิ้ง มีขนาดเส้นผ่านศูนย์กลางภายนอก 2.4 เซนติเมตร และขนาดเส้นผ่านศูนย์กลางภายใน (ส่วนของเจล) 1 เซนติเมตร

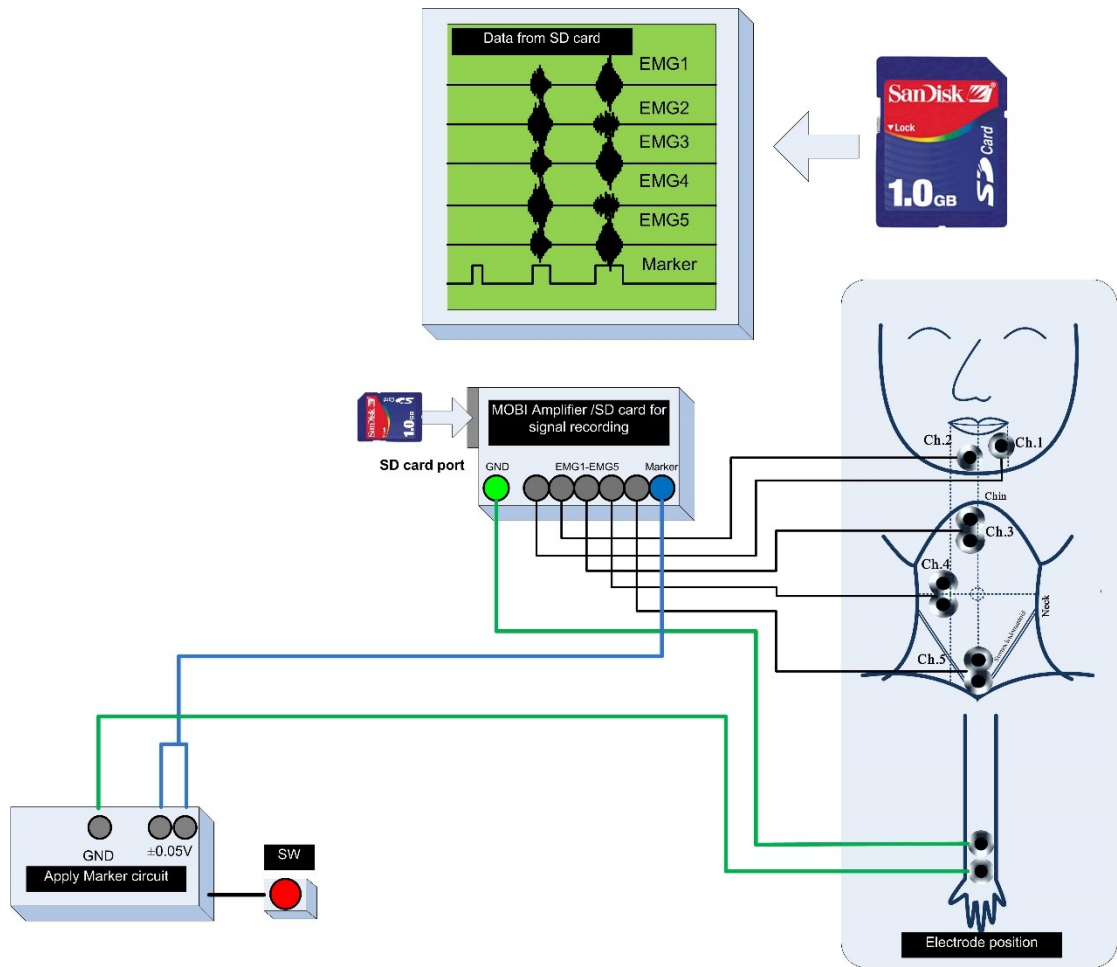
3) การ์ดบันทึกข้อมูล (SD card) ซึ่งใช้กับเครื่องวัดสัญญาณ Mobi6 สำหรับบันทึกข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ

4) วงจรสร้างสัญญาณการกดสวิตช์ และสวิตช์แบบกดติดปล่อยดับ เพื่อสร้างสัญญาณสำหรับกำหนดจุดเริ่มต้นของการบันทึกสัญญาณทั้งไฟฟ้ากล้ามเนื้อและใช้กำหนดขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด

5) โปรแกรม Portilab ซึ่งมาพร้อมกับเครื่องวัดสัญญาณ Mobi6 สำหรับใช้ดึงข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจากเครื่องวัดสัญญาณ Mobi6

6) โปรแกรม Photoshop สำหรับสร้างวิดีโอ เพื่อใช้ในการเก็บข้อมูลของสัญญาณให้มีจังหวะการพูดที่สม่ำเสมอ

รูปแบบการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อแสดงดังรูปที่ 3.11 ซึ่งจะตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ 5 ช่องสัญญาณและสัญญาณการกดสวิตช์ 1 ช่องสัญญาณโดยใช้ Mobi6 สำหรับสัญญาณการกดสวิตช์จะถูกใช้สำหรับการกำหนดจุดเริ่มต้นของการบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ และการกำหนดขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อของคำพูดแต่ละเสียง โดยจะกดสวิตช์เมื่อเริ่มพูดและปล่อยสวิตช์เมื่อพูดจบ ทำให้สามารถระบุตำแหน่งสัญญาณที่อยู่ในช่วงการกดสวิตช์เป็นสัญญาณที่เกิดจากการพูดจริง ไม่ใช่สัญญาณอื่นๆ เช่น สัญญาณที่เกิดจากการไอ จาม กลืนน้ำลาย หรือการขยับกล้ามเนื้อบริเวณที่จับสัญญาณโดยไม่ได้ตั้งใจ เป็นต้น ซึ่งสัญญาณการกดสวิตช์นี้จะถูกส่งไปยังช่องสัญญาณที่ 6 ของเครื่องวัดสัญญาณ Mobi6 สัญญาณนี้จะถูกเรียกว่า "marker" โดยสัญญาณนี้จะถูกบันทึกไปพร้อมกับสัญญาณไฟฟ้ากล้ามเนื้ออีก 5 ช่องสัญญาณ ซึ่งสัญญาณไฟฟ้ากล้ามเนื้อกับสัญญาณการกดสวิตช์จะเป็นสัญญาณที่ขนานกัน โดยความถี่ในการซักรับตัวอย่างของสัญญาณทั้ง 6 ช่องสัญญาณกำหนดไว้ 1024 เฮิรตซ์ โดยข้อมูลของสัญญาณทั้ง 6 ช่องสัญญาณจะถูกบันทึกลง SD card ภายใน Mobi6 หลังจากนั้นจะทำการดึงข้อมูลของสัญญาณโดยใช้โปรแกรม Portilab



รูปที่ 3.11 รูปแบบการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ

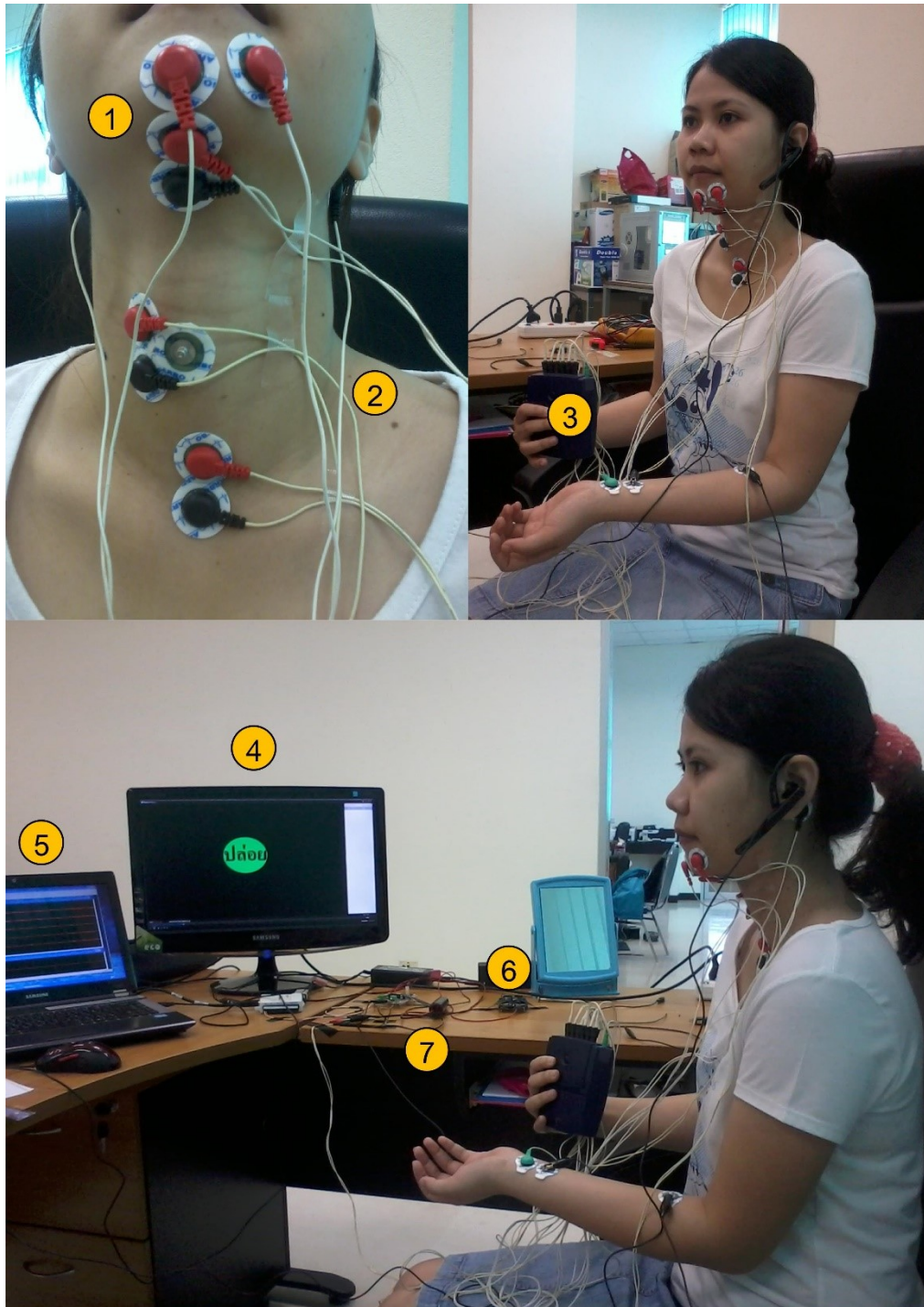


รูปที่ 3.12 ซ้าย: ตัวอย่างวิดีโอที่ใช้ขณะเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อขณะออกเสียงพูด  
ขวา: ลักษณะการกดสวิทช์ขณะการออกเสียงพูดเพื่อเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อ

ในกระบวนการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ ได้ออกแบบและสร้างวิดีโอ เพื่อใช้ในการเก็บข้อมูลของสัญญาณให้มีจังหวะการพูดที่สม่ำเสมอ กล่าวคือ อาสาสมัครจะออกเสียงพูดตามคำพูดที่ปรากฏบนวิดีโอขณะที่ทำการวัดสัญญาณ ซึ่งทำให้การบันทึกสัญญาณแต่ละคำ มีจังหวะสม่ำเสมอ ซึ่งวิดีโอนี้สร้างโดยใช้โปรแกรม Photoshop ตัวอย่างวิดีโอแสดงดังรูปที่ 3.12 (รูปซ้าย) ซึ่งในวิดีโอนี้จะบอกลำดับการกดสวิตช์และการออกเสียงพูดของแต่ละเสียง วงกลมใหญ่สีแดงแสดงถึงการกดสวิตช์ เมื่อกดสวิตช์แล้วจึงเริ่มออกเสียงพูด วงกลมใหญ่สีเขียวแสดงถึงการปล่อยสวิตช์ ซึ่งหมายถึงสิ้นสุดการออกเสียงคำนั้นๆ โดยก่อนการบันทึกแต่ละสัญญาณจะเริ่มจากการกดสวิตช์และปล่อย 1 ครั้งก่อนที่จะเริ่มออกเสียงพูดคำแรก ซึ่งสัญญาณการกดสวิตช์ครั้งแรกเป็นตัวกำหนดจุดเริ่มต้นของการบันทึกสัญญาณ ลักษณะการกดสวิตช์แสดงดังรูปที่ 3.12 (ขวา)

นอกจากนี้มีการเก็บข้อมูลของสัญญาณเสียงของคำพูด "ศูนย์" ถึง "สิบ" เพื่อใช้ในส่วนขอระบบสังเคราะห์เสียง ซึ่งอุปกรณ์ที่ใช้ในการเก็บข้อมูลของสัญญาณเสียง มีดังนี้

- 1) ไมโครโฟน สำหรับรับข้อมูลสัญญาณเสียง
- 2) เครื่องคอมพิวเตอร์ที่มีการ์ดเสียงอยู่ภายใน สำหรับเก็บข้อมูลและบันทึกสัญญาณเสียง



รูปที่ 3.13 ภาพขณะทำการทดลองเพื่อเก็บข้อมูลของสัญญาณจากอาสาสมัคร 1) อิเล็กโทรดแบบผิวหน้า 2) สายนำสัญญาณ 3) Mobi6 4) หน้าจอสำหรับแสดงวิดีโอ 5) คอมพิวเตอร์ 6) วงจรสร้างสัญญาณการกดสวิตช์ 7) สวิตช์แบบกดติดปล่อยดับ

### 3.1.3 การออกแบบคำพูด จำนวนคำพูด วิธีการพูด จำนวนอาสาสมัคร และ ประเภทอาสาสมัครที่ใช้สำหรับการเก็บข้อมูลสัญญาณ

งานวิจัยนี้ต้องการสร้างระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งต้องการจำแนกคำพูดของตัวเลขไทย 11 คำ ได้แก่ "ศูนย์" "หนึ่ง" "สอง" "สาม" "สี่" "ห้า" "หก" "เจ็ด" "แปด" "เก้า" และ "สิบ" และ รู้จำวรรณยุกต์ของแต่ละคำพูดด้วย ซึ่งวรรณยุกต์ไทยประกอบด้วย 5 เสียง คือ "สามัญ" "เอก" "โท" "ตรี" และ "จัตวา" ดังนั้นคำพูดที่ใช้ในการเก็บข้อมูลของสัญญาณจะประกอบไปด้วยคำต่างๆ เหล่านี้ สำหรับเสียงวรรณยุกต์จะเก็บข้อมูลการออกเสียงคำพูดของการผันวรรณยุกต์ของคำพูด 32 คำ ซึ่งครอบคลุมคำพูดที่ประกอบด้วยพยัญชนะอักษรต่ำ อักษรกลาง อักษรสูง รวมถึงการผันวรรณยุกต์ของคำพูด "ศูนย์" ถึง "สิบ" คำพูดทั้งหมดที่ถูกใช้ในกระบวนการเก็บข้อมูลของสัญญาณแสดงดังตารางที่ 3.3 โดยแบ่งเป็นสองส่วน คือ คำพูดสำหรับการจำแนกวรรณยุกต์ และคำพูดสำหรับการจำแนกคำพูดหลัก 11 คำ (ไม่คำนึงถึงวรรณยุกต์) คำพูดสำหรับการจำแนกวรรณยุกต์ทั้งหมด 148 คำ ซึ่งเกิดจากการผันวรรณยุกต์ของคำพูดหลัก 32 คำ

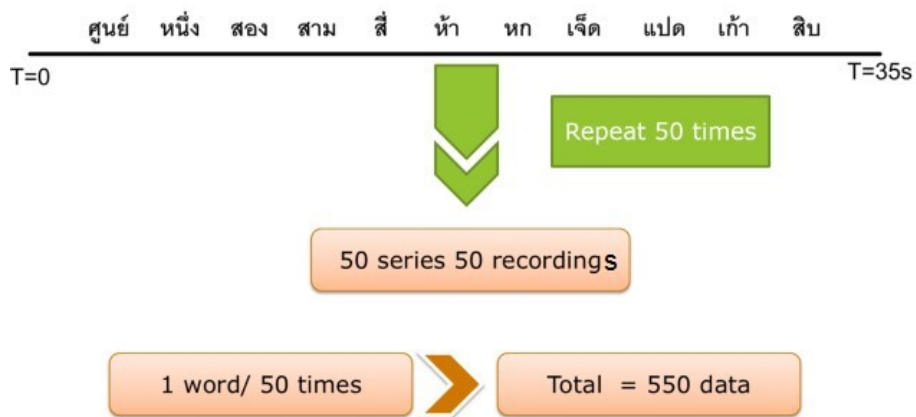
สำหรับจำนวนคำพูดที่ใช้ในการเก็บข้อมูลของสัญญาณ อาสาสมัครแต่ละคนจะออกเสียงพูดคำพูดทั้งหมด 2 ชุด

ชุดที่ 1) ออกเสียงคำพูด "ศูนย์" ถึง "สิบ" ต่อเนื่องกัน โดยเว้นช่วงพักระหว่างคำพูดแต่ละคำประมาณ 2 วินาที เมื่อพูดครบ 1 รอบ จะทำการบันทึกสัญญาณ 1 ครั้ง โดยที่สัญญาณมีความยาว 35 วินาที โดยที่อาสาสมัครแต่ละคนจะออกเสียงพูดทั้งหมด 50 รอบ จะได้สัญญาณที่ถูกบันทึกทั้งหมด 50 สัญญาณ ดังนั้นใน 1 คำ จะได้ข้อมูลทั้งหมด 50 ข้อมูล เพราะฉะนั้นในคำพูดชุดนี้จะมีทั้งหมด 550 ข้อมูล ( $11 \text{ คำ} \times 50 \text{ รอบ} = 550 \text{ ข้อมูล}$ ) ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อชุดนี้จะถูกนำไปใช้ในส่วนของการจำแนกคำพูดหลัก 11 คำพูด

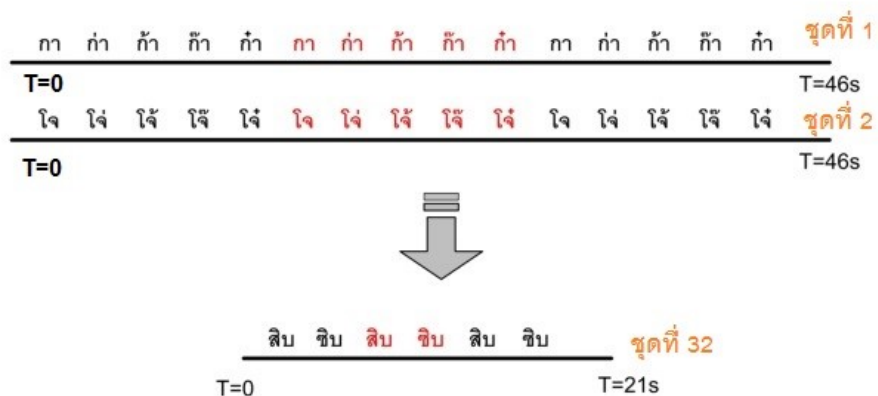
ชุดที่ 2) ออกเสียงคำพูดที่ใช้ในการจำแนกวรรณยุกต์ ซึ่งประกอบด้วย 32 ชุดคำพูด ดังตารางที่ 3.3 โดยในแต่ละชุดคำพูด จะประกอบด้วยเสียงวรรณยุกต์ทั้ง 5 เสียง ยกเว้นบางคำที่ไม่สามารถผันวรรณยุกต์ได้ครบ 5 เสียง ในการบันทึกสัญญาณ 1 ครั้งจะออกเสียงคำพูดในแต่ละชุด 3 ครั้งต่อเนื่องกัน เช่น ออกเสียง "กา ก่า ก้า ก๊า ก๋า ก่า ก้า ก๊า ก๋า ก่า ก้า ก๊า ก๋า" ใน 1 สัญญาณ (1 ครั้งของการบันทึก) จะมีความยาว 46 วินาทีสำหรับคำที่ผันได้ครบ 5 เสียง และความยาว 21 วินาทีสำหรับคำที่ผันได้ไม่ครบ 5 เสียง โดยจะบันทึกสัญญาณเรียงจากคำพูดชุดที่ 1 ถึงชุดที่ 32 หลังจากนั้นจะวนกลับมาบันทึกสัญญาณในชุดที่ 1 อีกครั้ง และทำแบบนี้จนครบ 3

รอบ จะได้สัญญาณที่ถูกบันทึกทั้งหมด 96 สัญญาณ (32x3) ดังนั้น สำหรับ 1 คำพูด จะมีข้อมูลทั้งหมด 9 ข้อมูล (1คำx3ครั้งx3รอบ=9 ข้อมูล) เพราะฉะนั้นในคำพูดชุดนี้จะมีข้อมูลทั้งหมด 1,332 ข้อมูล (9 x148คำ=1,332 ข้อมูล) ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อชุดนี้จะถูกนำไปใช้ในส่วนของการจำแนกวรรณยุกต์ 5 เสียง

ในส่วนของวิธีการพูด อาสาสมัครจะออกเสียงพูด 2 แบบ คือ การพูดแบบออกเสียง (พูดตามปกติ) และการพูดแบบไม่ออกเสียง โดยจะพูดตามจำนวนคำพูดที่กล่าวไว้ข้างต้น ซึ่งจะพูดแบบออกเสียงก่อน เมื่อพูดจนครบจำนวนคำพูดที่กำหนดแล้ว ก็ให้พูดแบบไม่ออกเสียง โดยไม่ต้องเปลี่ยนตำแหน่งหรือถอดอิล็กโทรดออก



รูปที่ 3.14 ลักษณะการบันทึกสัญญาณการออกเสียงพูด "ศูนย์" ถึง "สิบ" ที่ใช้ในการจำแนกคำพูดหลัก



รูปที่ 3.15 ลักษณะการบันทึกสัญญาณการออกเสียงพูดคำพูดที่ใช้ในการจำแนกวรรณยุกต์



ในส่วนของอาสาสมัคร ในงานวิจัยนี้จะเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัคร 10 คน โดยที่อาสาสมัคร 8 คน อายุตั้งแต่ 18-30 ปี เป็นผู้ชาย 4 คน และผู้หญิง 4 คน ส่วนอาสาสมัครที่เหลืออีก 2 คน อายุ 50-60 ปี โดยเป็นผู้หญิง 1 คน และผู้ชาย 1 คน เพื่อให้สามารถครอบคลุมสัญญาณไฟฟ้ากล้ามเนื้อจากกลุ่มวัยรุ่นและผู้สูงอายุ โดยอาสาสมัครทั้งหมดต้องเป็นคนที่พูดภาษาไทยเป็นภาษาแม่ สัญลักษณ์ที่ใช้แทนข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคนใช้ชื่อว่า "S01" ถึง "S10" ดังตารางที่ 3.2 ซึ่ง "S01" หมายถึง ข้อมูลของอาสาสมัคร คนที่ 1 เป็นต้น

ตารางที่ 3.2 ข้อมูลอาสาสมัคร

ข้อมูลอาสาสมัคร	อายุ 18-30 ปี	หญิง	"S01"
			"S02"
			"S03"
			"S04"
	อายุ 50-60 ปี	ชาย	"S05"
			"S06"
		หญิง	"S07"
			"S08"
		ชาย	"S09"
			"S10"

ตารางที่ 3.3 คำพูดทั้งหมดที่ถูกใช้สำหรับการเก็บข้อมูลของสัญญาณ

คำพูด (Words)	วรรณยุกต์ (Tones)	สามัญ	เอก	โท	ตรี	จัตวา	หมายเหตุ	
		(Mid)	(Low)	(Falling)	(High)	(Rising)		
คำพูดที่ใช้สำหรับ กระบวนการจำแนกวรรณยุกต์ (tone classification)	อักษรกลาง	1	กา	ก่า	ก้า	ก๊า	ก๋า	
		2	ใจ	ใจ๋	ใจั	ใจ็	ใจุ	
		3	คู	คู๋	คูั	คู็	คูุ	
		4	เบ	เบ๋	เบั	เบ็	เบุ	
		5	ออ	ออ๋	ออั	ออ็	ออุ	
		6	ตี	ตี๋	ตีั	ตี็	ตีุ	
	อักษรต่ำ	7	เปอ	เปอ๋	เปอั	เปอ็	เปอุ	
		8	เพอ	เพอ๋	เพอั	เพอ็	เพอุ	
		9	ฟา	ฟา๋	ฟาั	ฟา็	ฟาุ	
		10	ทอ	ทอ๋	ทอั	ทอ็	ทอุ	
		11	คู	คู๋	คูั	คู็	คูุ	
		12	เชือ	เชือ๋	เชือั	เชือ็	เชือุ	
	อักษรสูง	13	โฮ	โฮ๋	โฮั	โฮ็	โฮุ	
		14	แฮร์	แฮร์	แฮร์	แฮร์	แฮร์	
		15	งู	งู๋	งูั	งู็	งูุ	
		16	ยา	ยา๋	ยาั	ยา็	ยาุ	
		17	นี่	นี่๋	นี่ั	นี่็	นีุ่	
		18	เวือ	เวือ๋	เวือั	เวือ็	เวือุ	
		19	ว่ว	ว่ว๋	ว่วั	ว่ว็	ว่วุ	
		20	มื่อ	มื่อ๋	มื่อั	มื่อ็	มื่อุ	
		21	ลอ	ลอ๋	ลอั	ลอ็	ลอุ	
คำพูดที่ใช้สำหรับ กระบวนการจำแนกคำพูดหลัก (word classification)	22	ซูน	ซูน	ซูน	ซูน	ซูน	1 ซูน	
	23	นึ่ง	นึ่ง	นึ่ง	นึ่ง	นึ่ง	2 นึ่ง	
	24	ซอ	ซอ	ซอ	ซอ	ซอ	3 ซอ	
	25	ทรา	ทรา	ทรา	ทรา	ทรา	4 ทรา	
	26	ชี	ชี	ชี	ชี	ชี	5 ชี	
	27	ฮา	ฮา	ฮา	ฮา	ฮา	6 ฮา	
	28	-	หก	-	หก	-	7 หก	
	29	-	เจ็ด	-	เจ็ด	-	8 เจ็ด	
	30	-	แปด	-	แปด	-	9 แปด	
	31	กาว	กาว	กาว	กาว	กาว	10 กาว	
	32	-	สิบ	-	สิบ	-	11 สิบ	

**ตารางที่ 3.4** สรุปข้อมูลจำนวนอาสาสมัคร วิธีการพูด จำนวนการพูด และข้อมูลทั้งหมดสำหรับการเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง

จำนวนอาสาสมัคร	ประเภทของคำพูด	วิธีการพูด	จำนวนคำ	จำนวนครั้งในแต่ละรอบ	จำนวนรอบบันทึก	สรุปจำนวนข้อมูลแต่ละประเภท	จำนวนข้อมูลทั้งหมดสำหรับอาสาสมัคร 1 คน	ข้อมูลทั้งหมด
10	ศูนย์-สิบ	ออกเสียง	11	1	50	11x1x50 =550	550(2)+1332(1) =2432	2432x10 =24320
		ไม่ออกเสียง	11	1	50	11x1x50 =550		
	การผันวรรณยุกต์	ออกเสียง	148	3	3	148x3x3 =1332		
		ไม่ออกเสียง						

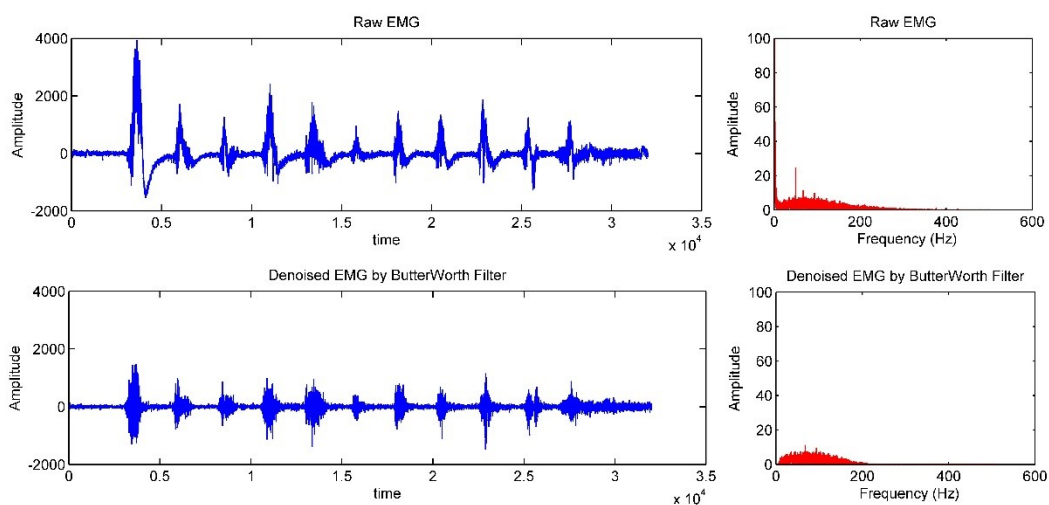
### 3.2 กระบวนการเตรียมสัญญาณเบื้องต้น

หลังจากเก็บข้อมูลของสัญญาณทั้งหมด กระบวนการต่อไปคือ กระบวนการเตรียมสัญญาณเบื้องต้นสำหรับสัญญาณไฟฟ้ากล้ามเนื้อ ได้แก่ การกรองสัญญาณรบกวน และกระบวนการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งจะใช้โปรแกรม LabVIEW และ MATLAB ในการประมวลผล

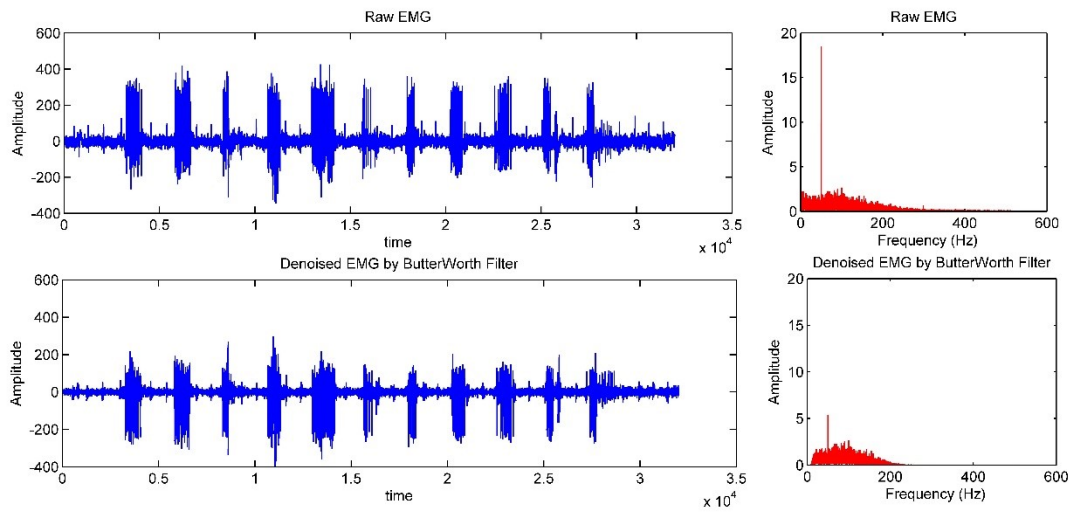
#### 3.2.1 การกรองสัญญาณรบกวน

จากการพิจารณาสัญญาณไฟฟ้ากล้ามเนื้อในโดเมนเวลาพบว่า สัญญาณช่องที่ 4 และช่องที่ 5 ซึ่งเป็นสัญญาณไฟฟ้ากล้ามเนื้อบริเวณกล้ามเนื้อคอ มีสัญญาณรบกวนประเภทสัญญาณคลื่นไฟฟ้าหัวใจ (ECG) อย่างชัดเจน สำหรับสัญญาณช่องที่ 1 ถึง ช่องที่ 3 ซึ่งเป็นสัญญาณกล้ามเนื้อบริเวณใบหน้าและใต้คาง จะเห็นได้ว่ามีสัญญาณรบกวนจากการเคลื่อนไหวที่

อย่างชัดเจน ซึ่งเกิดจากการขยับสายอิเล็กโทรด นอกจากนั้นเมื่อพิจารณารูปแบบของสัญญาณในโดเมนความถี่จะเห็นได้ชัดเจนว่า สัญญาณไฟฟ้ากล้ามเนื้อช่องที่ 1 ถึง 3 มีสัญญาณรบกวนในช่วงที่ต่ำกว่า 10 เฮิร์ตซ์ สูงมาก และทุกช่องสัญญาณมีสัญญาณรบกวนความถี่ 50 เฮิร์ตซ์ ปนอยู่ ดังนั้นจึงแบ่งการกำจัดสัญญาณรบกวนออกเป็น 2 ส่วน สำหรับสัญญาณช่องที่ 1 ถึง 3 จะใช้ตัวกรองความถี่ผ่านซึ่งมีช่วงความถี่ตัดผ่านอยู่ในช่วง 20-350 เฮิร์ตซ์ ซึ่งสามารถกำจัดสัญญาณรบกวนจากการเคลื่อนไหวและสัญญาณรบกวนที่ความถี่สูงได้ สำหรับสัญญาณช่องที่ 4 และช่องที่ 5 จะใช้ตัวกรองความถี่ผ่านซึ่งมีช่วงความถี่ตัดผ่านอยู่ในช่วง 30-350 เฮิร์ตซ์ ซึ่งสามารถลดทอนสัญญาณรบกวนคลื่นไฟฟ้าหัวใจได้ โดยตัวกรองความถี่ผ่านที่ใช้ คือ ตัวกรองชนิดบัตเตอร์เวิร์ท (Butterworth) อันดับที่ 5 ซึ่งได้ทำการทดสอบปรับอันดับของตัวกรอง โดยอันดับที่ 5 สามารถกำจัดสัญญาณรบกวนคลื่นไฟฟ้าหัวใจได้มากที่สุด นอกจากนั้นตัวกรองนอตช์จะถูกใช้สำหรับกำจัดสัญญาณรบกวน 50 เฮิร์ตซ์ สำหรับ 5 ช่องสัญญาณ ซึ่งตัวกรองความถี่ที่ใช้ทั้งหมดเป็นตัวกรองความถี่แบบดิจิทัล



**รูปที่ 3.16** สัญญาณไฟฟ้ากล้ามเนื้อช่องที่ 1 สำหรับการพูด "ศูนย์" ถึง "สิบ" ภาพบน: สัญญาณดิบ ภาพล่าง: สัญญาณที่ผ่านการกรองโดยใช้ Butterworth filter และ Notch filter



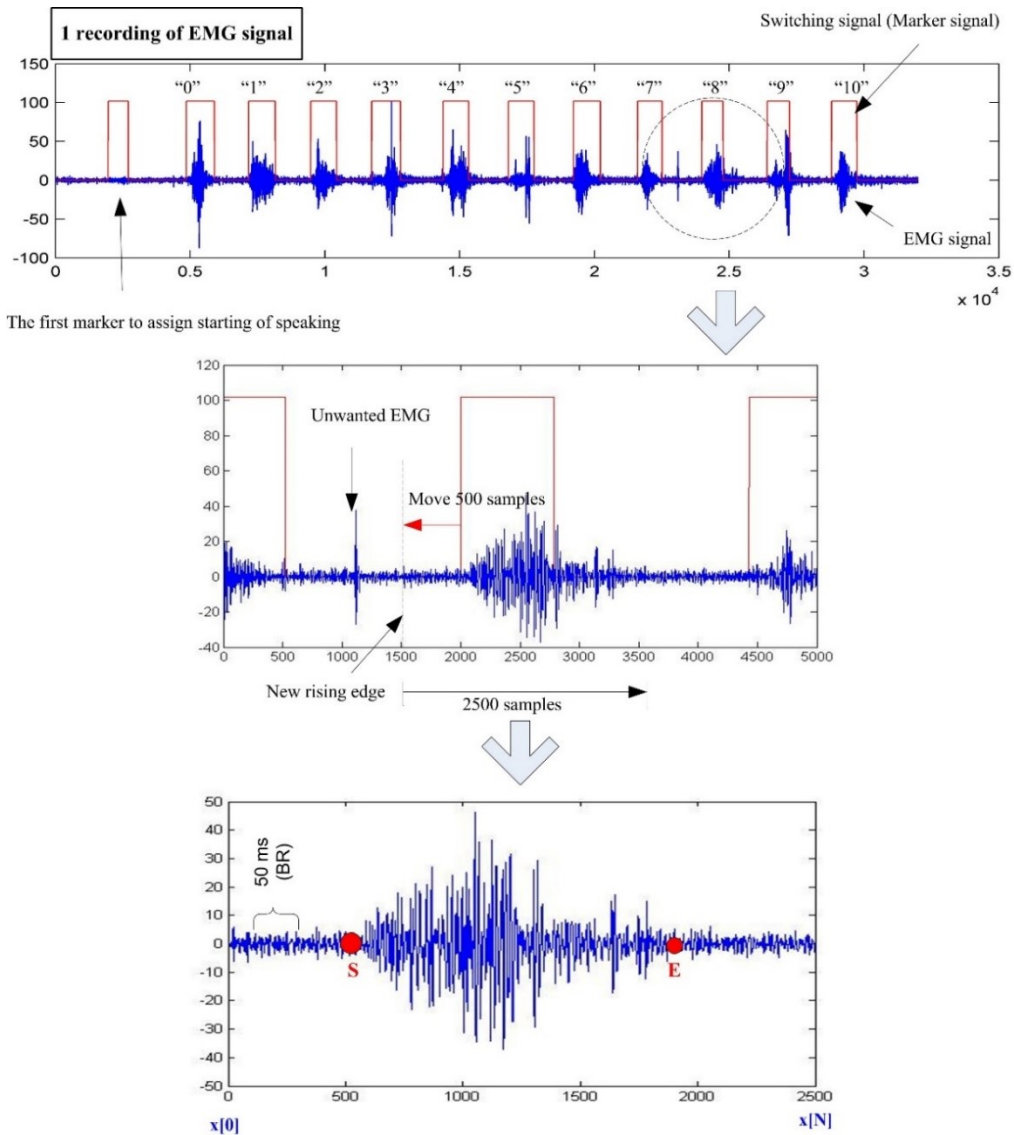
**รูปที่ 3.17** สัญญาณไฟฟ้ากล้ามเนื้อเนื้อของที่ 4 สำหรับการพูด "ศูนย์" ถึง "สิบ" ภาพบน: สัญญาณดิบ ภาพล่าง: สัญญาณที่ผ่านการกรองโดยใช้ Butterworth filter และ Notch filter

### 3.2.2 การตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ

หลังจากผ่านการกรองสัญญาณรบกวนแล้ว สัญญาณไฟฟ้ากล้ามเนื้อจะถูกนำมาทำการวิเคราะห์หาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด โดยเริ่มจากการตัดสัญญาณไฟฟ้ากล้ามเนื้อของแต่ละคำพูดออกมาก่อน ซึ่งไม่ได้ตัดมาจากจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อที่แท้จริง แต่จะตัดมาจากการใช้สัญญาณจากการกดสวิตช์เป็นตัวกำหนดขอบเขต โดยจะดึงสัญญาณมา 2500 ตัวอย่าง (samples) จากขอบเขตดังกล่าว เนื่องจากพบว่าขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อแต่ละคำพูดไม่เกิน 2500 ตัวอย่าง ลักษณะการตัดสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้สัญญาณการกดสวิตช์เป็นตัวกำหนดขอบเขตสามารถอธิบายได้ดังรูปที่ 3.18 ซึ่งการตัดสัญญาณเบื้องต้นโดยใช้สัญญาณการกดสวิตช์ตัวกำหนดขอบเขต เป็นการลดปัญหาการหลีกเลี่ยงสัญญาณที่ไม่ต้องการระหว่างการบันทึกสัญญาณ เช่น สัญญาณไฟฟ้ากล้ามเนื้อที่ไม่ได้เกิดจากการออกเสียงพูดจริงๆ ได้แก่ การไอ การจาม การหาว การขยับปาก เป็นต้น

เขียนโค้ดเพื่อตัดสัญญาณอัตโนมัติโดยใช้โปรแกรม LabVIEW โดยการออกแบบโปรแกรมให้เริ่มตัดสัญญาณจากการกดสวิตช์ครั้งที่ 2 ของสัญญาณ (สัญญาณการกดสวิตช์ครั้งที่

1 เป็นการกดสวิตช์เพื่อเตรียมความพร้อมสำหรับการบันทึกสัญญาณ สัญญาณที่เริ่มพุดอยู่ในช่วงของการกดสวิตช์ครั้งที่ 2) โดยโปรแกรมจะตรวจสอบว่าเจอช่วงขาขึ้นของสัญญาณสวิตช์หรือไม่ ถ้าใช่และเป็นช่วงขาขึ้นของสัญญาณสวิตช์ครั้งที่ 2 เป็นต้นไปจะทำการบันทึกสัญญาณ โดยจะบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ 2500 ตัวอย่าง โดยเริ่มจากตัวอย่างก่อนหน้าสัญญาณขาขึ้นของสัญญาณสวิตช์ 500 ตัวอย่าง เนื่องจากพบว่าสัญญาณการกดสวิตช์เกิดขึ้นหลังสัญญาณไฟฟ้ากล้ามเนื้อสำหรับบางสัญญาณ ดังนั้นจึงดึงสัญญาณไฟฟ้ากล้ามเนื้อโดยครอบคลุมสัญญาณที่เกิดก่อนการกดสวิตช์ 500 ตัวอย่าง หลังจากนั้นสัญญาณที่ได้จะถูกนำไปหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อเพื่อนำไปใช้ในการประมวลผลต่อไป



รูปที่ 3.18 การตัดสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้สัญญาณการกดสวิตช์เป็นตัวกำหนดขอบเขต

หลังจากตัดสัญญาณไฟฟ้ากล้ามเนื้อออกมาเป็นสัญญาณสำหรับแต่ละคำพูดแล้ว จะทำการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูดของแต่ละคำ โดยใช้วิธีที่เรียกว่า ค่าขีดแบ่งของสามเท่าของค่าเบี่ยงเบนมาตรฐาน หรือ Three standard deviation threshold ซึ่งคำนวณจากค่าผลรวมของค่าเฉลี่ย (mean) และสามเท่าของค่าเบี่ยงเบนมาตรฐานของสัญญาณอ้างอิง ( $3 \times$  standard deviation) ซึ่งใช้ค่านี้ในการตรวจสอบ (threshold, TH) เพื่อหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะออกเสียงพูด เนื่องจากเป็นวิธีที่ง่ายและนิยมใช้กันมากในหลายงานวิจัย โดยงานวิจัยนี้ประยุกต์เทคนิคของ Di Fabio [101] เพื่อให้เหมาะสมกับลักษณะของสัญญาณของงานวิจัยนี้ โดยเพิ่มเงื่อนไขในการตรวจสอบอีก 1 เงื่อนไข

กำหนดให้  $x_{BR}$  คือ สัญญาณอ้างอิง (base line reference, BR) ซึ่งเลือกจากสัญญาณไฟฟ้ากล้ามเนื้อในขณะที่ไม่มีการออกเสียงพูดหรือขณะเงียบ ซึ่งจะดึงสัญญาณมา 50 มิลลิวินาที

ค่าที่ใช้ในการตรวจสอบเพื่อหาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะออกเสียงพูด คำนวณจาก

$$TH = \bar{x}_{BR} + 3SD_{BR} \quad (3.1)$$

$$\bar{x}_{BR} = \frac{1}{M} \sum_{i=1}^M x_{BR}[n] \quad (3.2)$$

$$SD_{BR} = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (x_{BR}[i] - \bar{x}_{BR})^2} \quad (3.3)$$

ซึ่ง คือ  $M$  จำนวนตัวอย่างของสัญญาณอ้างอิง

กำหนดให้  $\mathbf{x}$  คือ สัญญาณไฟฟ้ากล้ามเนื้อของหนึ่งคำพูด ซึ่งถูกตัดมาจากกระบวนการแบ่งส่วนสัญญาณเบื้องต้น โดยมีจำนวน 2500 ตัวอย่าง ( $N = 2500$ )

การหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด เริ่มพิจารณาจาก  $x[0]$  ซึ่ง  $x[n]$  จะเป็นจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด เมื่อเข้าเงื่อนไขทั้งสามเงื่อนไขดังต่อไปนี้

- 1)  $x[n] > TH$
- 2) ทุกตัวอย่างตั้งแต่  $x[n], x[n+1], \dots, [x+9] > TH$

3) ต้องมีอย่างน้อย 20 ตัวอย่าง ตั้งแต่  $x[n], x[n+1], \dots, [x+49] > TH$  (เงื่อนไขเพิ่มเติม)

ในทางกลับกันการพิจารณาหาจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อขณะพูด เริ่มพิจารณาจาก  $x[N]$  ซึ่ง  $x[n]$  จะเป็นจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อเข้าเงื่อนไขทั้งสามเงื่อนไขดังต่อไปนี้

$$1) x[n] > TH$$

$$2) \text{ ทุกตัวอย่างตั้งแต่ } x[n], x[n-1], \dots, [x-9] > TH$$

3) ต้องมีอย่างน้อย 20 ตัวอย่าง ตั้งแต่  $x[n], x[n-1], \dots, [x-49] > TH$  (เงื่อนไขเพิ่มเติม)

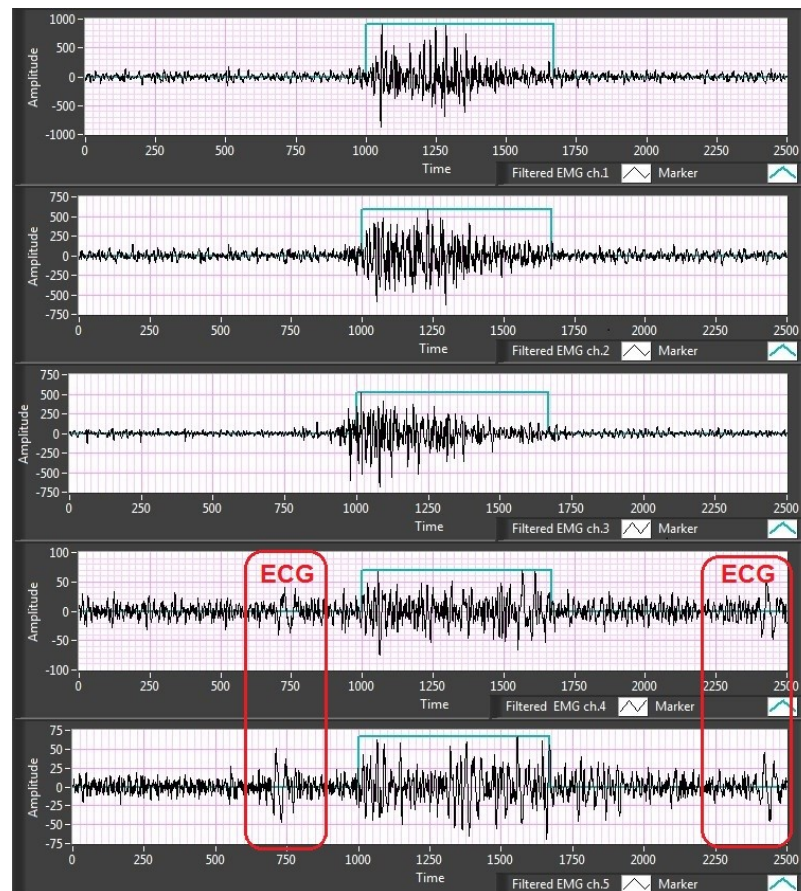
สาเหตุที่งานวิจัยนี้ต้องเพิ่มเงื่อนไขที่สามขึ้นมา เนื่องจากพบความผิดปกติของลักษณะของสัญญาณไฟฟ้ากล่อมเนื้อที่เกิดขึ้นในบางสัญญาณ ซึ่งทำให้อาจจะเกิดความผิดพลาดในการตรวจสอบโดยใช้เงื่อนไขสองเงื่อนไขดั้งเดิม

1) ปัญหาที่หนึ่ง การที่สัญญาณรบกวนคลื่นไฟฟ้าหัวใจยังคงปรากฏอยู่ในสัญญาณไฟฟ้ากล่อมเนื้อ ทำให้การตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณอาจเกิดความผิดพลาด เนื่องจากโปรแกรมจะไปตรวจจับสัญญาณรบกวนแทน ซึ่งสัญญาณรบกวนดังกล่าวจะพบในสัญญาณช่องที่ 4 และช่องที่ 5 แสดงดังรูปที่ 3.19

2) ปัญหาที่สอง การเกิดขึ้นของสัญญาณไฟฟ้ากล่อมเนื้อที่ไม่ได้เกิดจากการพูด ซึ่งอยู่นอกช่วงสัญญาณการกวดสวิตซ์ทำให้การตรวจจับจุดเริ่มต้นและจุดสิ้นสุดผิดพลาด ซึ่งจะไปตรวจจับสัญญาณในช่วงนั้นแทน แสดงดังรูปที่ 3.20

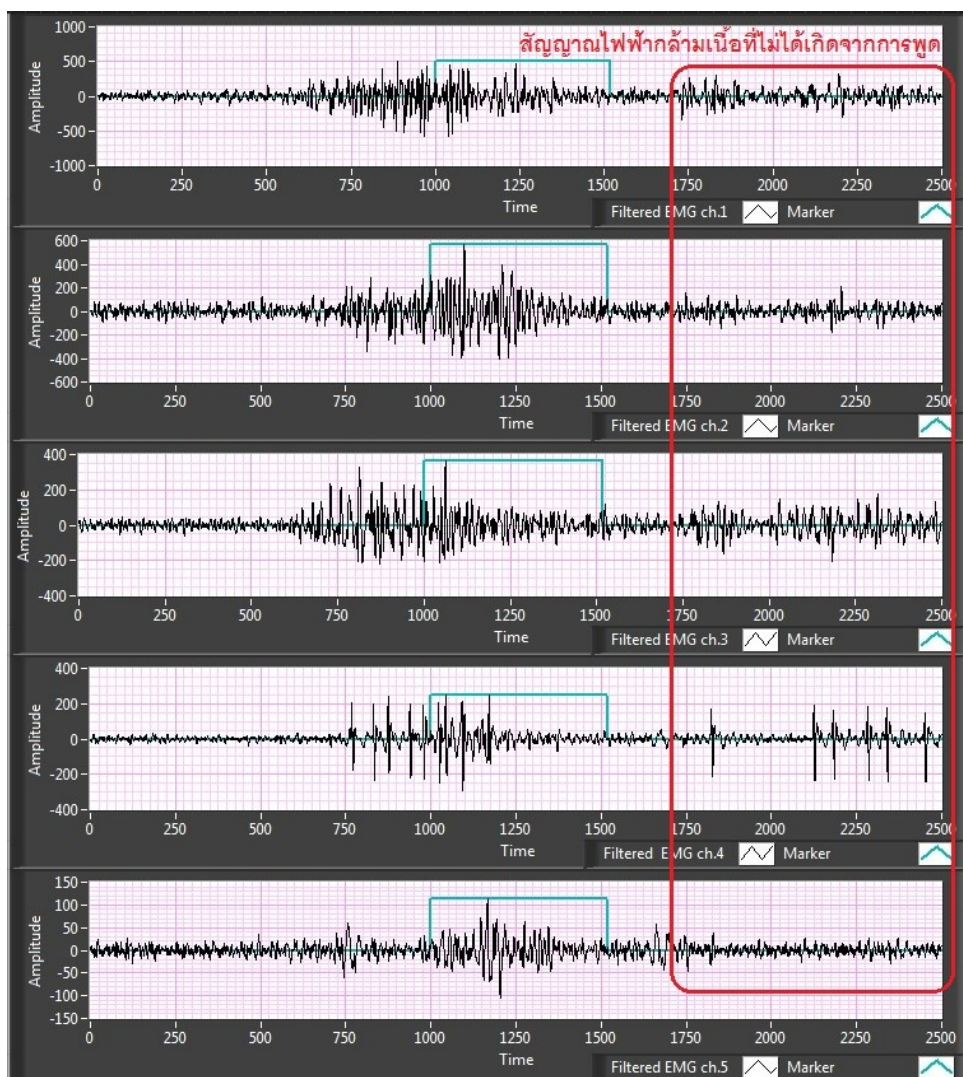
จากปัญหาดังกล่าว ในการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อขณะพูด จะต้องมีการประยุกต์ใช้วิธีข้างต้นให้เหมาะสม เนื่องจากพบว่าส่วนใหญ่สัญญาณที่ไม่ต้องการที่ปรากฏขึ้นนั้น เกิดขึ้นแค่ช่วงเวลาสั้นๆ ไม่เกิน 20 มิลลิวินาที (หรือประมาณ 20 samples) ซึ่งส่วนใหญ่สัญญาณในช่วงนี้จะมีแอมพลิจูดเกินค่า TH และตำแหน่งที่เกิดของสัญญาณอยู่ห่างจากสัญญาณไฟฟ้ากล่อมเนื้อจริงที่ต้องการจะตรวจจับ ดังนั้นในโปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล่อมเนื้อขณะพูดจึงเพิ่มเงื่อนไขที่สามเพื่อลดความผิดพลาดในการตรวจจับ ซึ่งได้กล่าวไว้แล้วข้างต้น





**รูปที่ 3.19** การปรากฏของสัญญาณคลื่นไฟฟ้าหัวใจที่ไม่สามารถกำจัดได้หมด ซึ่งจะปรากฏเฉพาะสัญญาณช่องที่ 4 และ ช่องที่ 5

อย่างไรก็ตามในโปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูดจะใช้วิธีตรวจจับกึ่งอัตโนมัติ กล่าวคือ ทำการตรวจจับโดยใช้โปรแกรม (automatic) และตรวจจับแบบแมนนวล (manual) ไปพร้อมๆ กัน เพื่อลดความผิดพลาดในการตรวจจับอีกทางหนึ่ง โดยการเพิ่มเงื่อนไขอีกหนึ่งเงื่อนไข คือ ถ้าโปรแกรมแสดงเอาต์พุตว่า "การตรวจจับอาจจะผิดพลาด" โปรแกรมการตรวจจับจะหยุดลง และระบุสาเหตุของปัญหา หลังจากนั้นต้องทำการตัดสัญญาณนั้นเองด้วยมือ แต่ถ้าโปรแกรมไม่แสดงเอาต์พุตความผิดพลาดในการตรวจจับ ระบบจะตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณขณะพูดและตัดสัญญาณเฉพาะส่วนที่ต้องการอัตโนมัติ



รูปที่ 3.20 การปรากฏของสัญญาณไฟฟ้ากล้ามเนื้อที่ไม่ได้เกิดจากการหดซึ่งอยู่นอกช่วงการกด สวิตช์

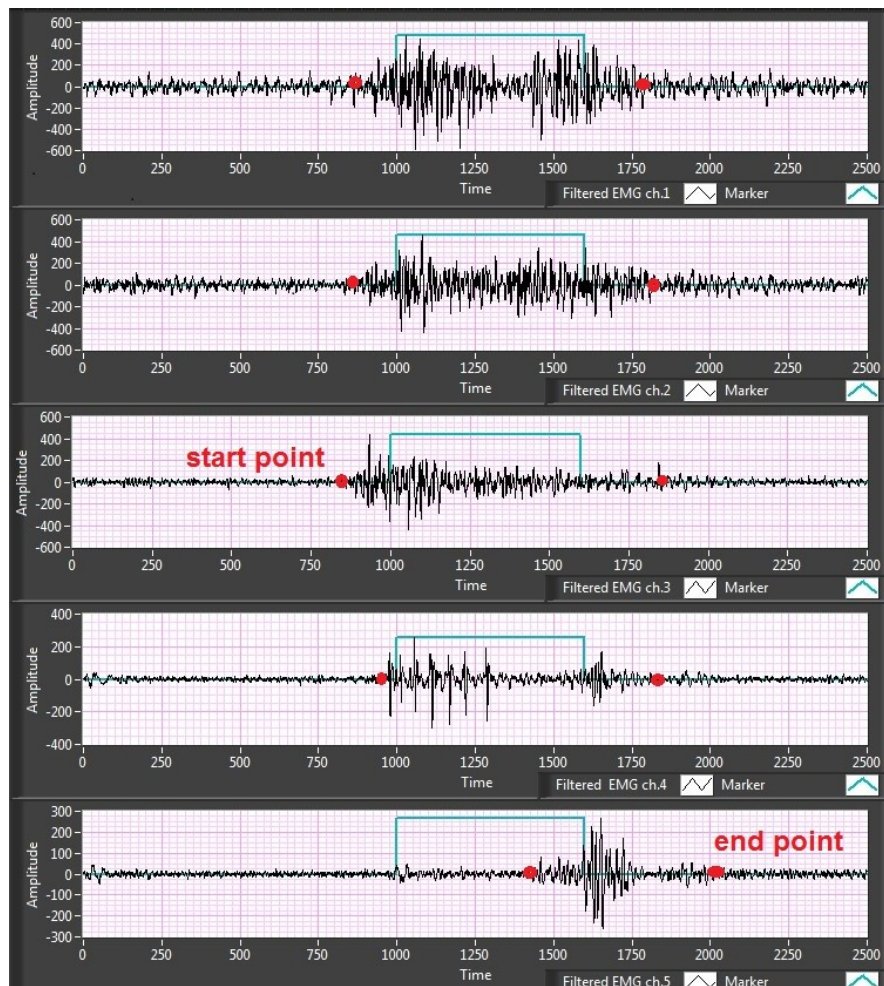
สำหรับสาเหตุที่จะระบุให้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ระบุว่า "การตรวจจับอาจจะผิดพลาด" มีสาเหตุดังนี้

1) ความยาวของสัญญาณที่ถูกตัดออกมาหลังจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ มีค่ามากกว่าสองเท่าของความยาวของสัญญาณการกด สวิตช์ ซึ่งหมายความว่าจุดที่ถูกตรวจจับไม่ใช่จุดที่ต้องการ

2) ความยาวของสัญญาณที่ถูกตัดออกมาหลังจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ มีค่าน้อยกว่าครึ่งหนึ่งของความยาวของสัญญาณการกด สวิตช์ ซึ่งหมายความว่าจุดที่ตรวจจับไม่ใช่จุดที่ต้องการ

3) สำหรับการตรวจจับจุดเริ่มต้น ถ้าจุดเริ่มต้นที่ได้  $x[n]$  อยู่ในตำแหน่งของ  $n \leq 500$  จุดนั้นอาจไม่ใช่สัญญาณไฟฟ้ากล้ามเนื้อขณะพูด แต่อาจจะเป็นสัญญาณรบกวน

นอกจากนี้ เนื่องจากสัญญาณไฟฟ้ากล้ามเนื้อที่เกิดขึ้นในแต่ละช่องสัญญาณมีการเริ่มต้นและสิ้นสุดของสัญญาณที่ต่างกันไป ดังนั้นในการกำหนดจุดเริ่มต้นจะเลือกจุดเริ่มต้นจากช่องสัญญาณที่มีจุดเริ่มต้นเกิดขึ้นก่อน สำหรับจุดสิ้นสุดจะเลือกจุดสิ้นสุดจากช่องสัญญาณที่มีการสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อครั้งสุดท้าย จากรูปที่ 3.21 แสดงตำแหน่งของจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูดสำหรับแต่ละช่องสัญญาณ ซึ่งมีเวลาเริ่มต้นและสิ้นสุดไม่พร้อมกัน สำหรับตัวอย่างสัญญาณนี้จะเลือกตัดสัญญาณโดยใช้จุดเริ่มต้นจากช่องสัญญาณที่ 3 และจุดสิ้นสุดของสัญญาณจากช่องสัญญาณที่ 5



รูปที่ 3.21 แสดงตำแหน่งของจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด สำหรับแต่ละช่องสัญญาณ ซึ่งมีเวลาเริ่มต้นและสิ้นสุดไม่พร้อมกัน

สรุปขั้นตอนการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ

- 1) ตรวจหาสัญญาณอ้างอิง ( $x_{BR}$ ) ทั้ง 5 ช่องสัญญาณ
- 2) คำนวณค่าที่ใช้ในการตรวจสอบ ( $TH$ ) จากสัญญาณอ้างอิง ทั้ง 5 ช่องสัญญาณ โดยใช้สมการที่ (3.1) ( $TH_1, TH_2, TH_3, TH_4, TH_5$ )
- 3) หาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อแต่ละช่องสัญญาณ ซึ่งก็คือ  $S_1, S_2, S_3, S_4, S_5$  โดยพิจารณาจากค่า  $TH_1, TH_2, TH_3, TH_4, TH_5$  และเงื่อนไข 3 เงื่อนไข ตามที่ได้กล่าวไว้แล้วข้างต้น

4) หาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อ ( $S$ ) เป็นการหาจุดที่มีเกิดสัญญาณขึ้นก่อนจากทั้ง 5 ช่องสัญญาณ

$$S = \min(S_1, S_2, S_3, S_4, S_5) \quad (3.4)$$

5) หาจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อแต่ละช่องสัญญาณ ซึ่งก็คือ  $E_1, E_2, E_3, E_4, E_5$  โดยพิจารณาจากค่า  $TH_1, TH_2, TH_3, TH_4, TH_5$  และเงื่อนไข 3 เงื่อนไข ตามที่ได้กล่าวไว้แล้วข้างต้น

6) หาจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ( $E$ ) เป็นการหาจุดที่มีการสิ้นสุดของสัญญาณหลังสุด จากทั้ง 5 ช่องสัญญาณ

$$E = \max(E_1, E_2, E_3, E_4, E_5) \quad (3.5)$$

7) ตรวจสอบว่า จุดที่ตรวจจับนั้นมีโอกาสผิดพลาดหรือไม่ โดยพิจารณาตามเงื่อนไขดังนี้คือ

- ความยาวของสัญญาณที่ถูกตัดออกมาหลังจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ( $E - S$ ) มีค่ามากกว่าสองเท่าของความยาวของสัญญาณการกวดสวิตช์

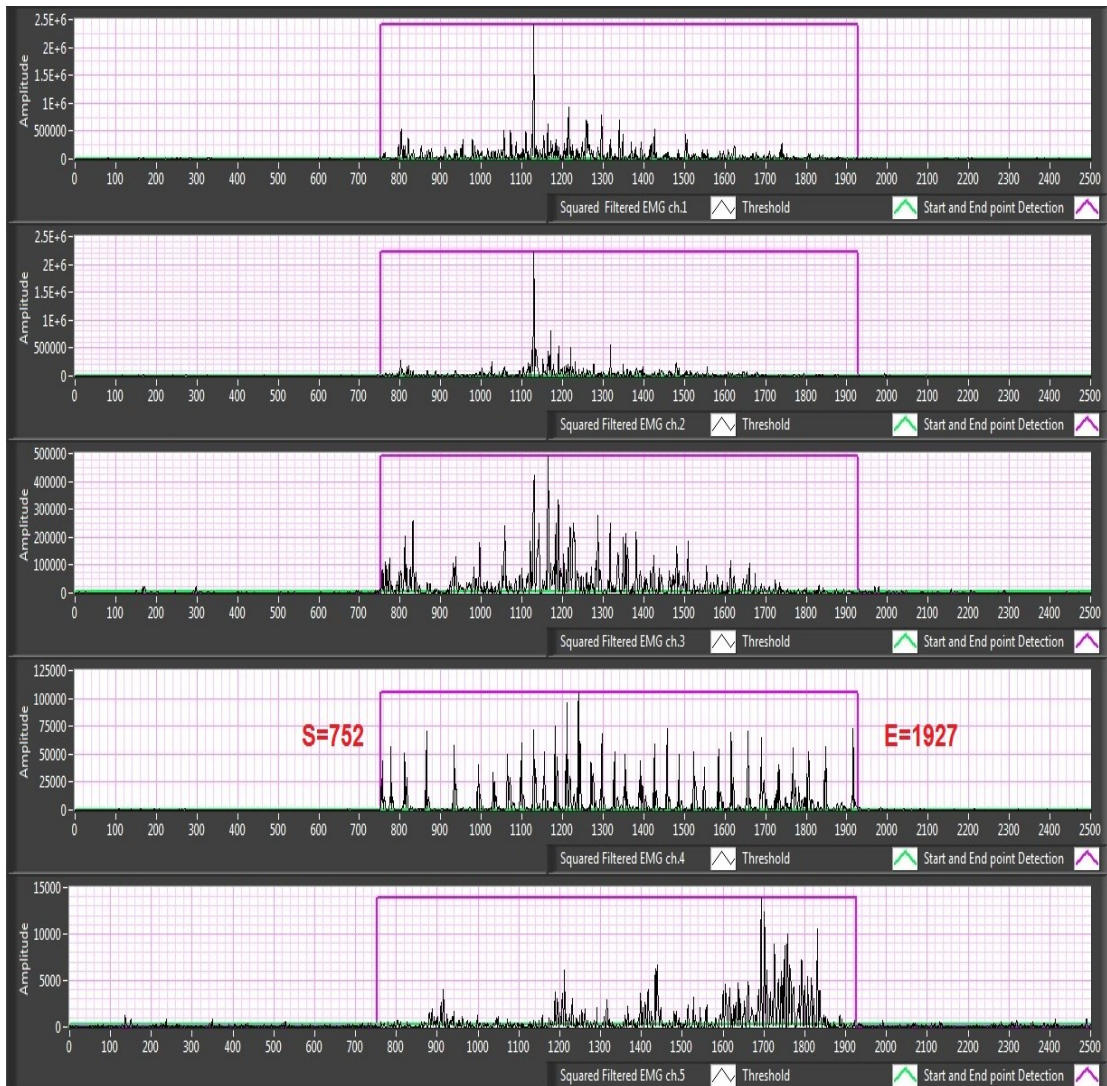
- ความยาวของสัญญาณที่ถูกตัดออกมาหลังจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ( $E - S$ ) มีค่าน้อยกว่าครึ่งหนึ่งของความยาวของสัญญาณการกวดสวิตช์

- ถ้าจุดเริ่มต้นที่ได้  $x[n]$  อยู่ในตำแหน่งของ  $n \leq 500$  ( $S \leq 500$ )

ถ้าตรงตามเงื่อนไขใดเงื่อนไขหนึ่งหรือทั้งสามเงื่อนไข โปรแกรมจะระบุออกมาว่าการตรวจจับผิดพลาดเนื่องจากสาเหตุใด ซึ่งโปรแกรมจะไม่ทำการตัดและบันทึกสัญญาณ โดยเรา

จะทำการพิจารณาหาจุดเริ่มต้นและจุดสิ้นสุดเอง โดยการสังเกต แต่ถ้าระบบการตรวจจับถูกต้อง โปรแกรมจะทำการตัดสัญญาณและบันทึกสัญญาณนั้นอัตโนมัติ

รูปที่ 3.22 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ จากสัญญาณของคำว่า "สี" สำหรับการพูดแบบออกเสียง ซึ่งผลของการตรวจจับไม่เข้าเงื่อนไขของการตรวจจับผิดพลาด ดังนั้นโปรแกรมจะตัดสัญญาณโดยกำหนดขอบเขตจากจุดดังกล่าว สำหรับรูปที่ 3.23 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อจากสัญญาณของคำว่า "สิบ" สำหรับการพูดแบบออกเสียง ซึ่งผลของการตรวจจับเข้าเงื่อนไขของการตรวจจับผิดพลาด ซึ่งโปรแกรมระบุว่า ความยาวของสัญญาณที่ถูกตัดออกมาหลังจากการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ( $E - S$ ) มีค่ามากกว่าสองเท่าของความยาวของสัญญาณการกวดสิทธิ์ ซึ่งแสดงว่าจุดที่ตรวจจับนั้นยังไม่ถูกต้อง จะต้องหยุดโปรแกรมเพื่อตัดสัญญาณแบบอัตโนมัติ แล้วทำการตัดสัญญาณแบบแมนนวล



รูปที่ 3.22 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ จากสัญญาณของคำว่า "สี่" ซึ่งผลการตรวจจับถูกต้อง



รูปที่ 3.23 แสดงผลของการใช้โปรแกรมตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งผลของการตรวจจับเข้าเงื่อนไขของการตรวจจับผิดพลาด

### 3.3 การประมวลผลสัญญาณไฟฟ้ากล้ำมเนื้อเพื่อออกแบบระบบรับรู้การพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อ

หลังจากผ่านกระบวนการเก็บข้อมูลของสัญญาณและการเตรียมสัญญาณเบื้องต้นแล้ว สัญญาณที่ได้จะถูกนำมาประมวลผลเพื่อสร้างระบบรับรู้การพูดโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อ ซึ่งขั้นตอนแรกในการประมวลผล คือ การสกัดลักษณะเด่นของสัญญาณไฟฟ้ากล้ำมเนื้อ ซึ่งนำไปสู่การประเมินและเลือกลักษณะเด่นของสัญญาณในขั้นตอนต่อไป หลังจากนั้นลักษณะเด่นที่ถูกเลือกจะถูกนำไปใช้ในกระบวนการฝึกสอนระบบเพื่อสร้างแบบจำลองในการจำแนก และเข้าสู่กระบวนการจำแนกในขั้นตอนสุดท้าย

ซึ่งในหัวข้อนี้จะกล่าวถึง การทดลองเพื่อออกแบบระบบรับรู้การพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ำมเนื้อให้ได้ความถูกต้องในการจำแนกสูงสุดเท่าที่จะเป็นไปได้ ซึ่งจะแบ่งออกเป็นสองการทดลองหลัก คือ การออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด) และการออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง)

#### 3.3.1 การออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด)

3.3.1.1 การทดลองเบื้องต้นเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกคำพูด 11 คำพูด โดยใช้ RES เป็นตัวประเมินลักษณะเด่น และ เปรียบเทียบผลการประเมินที่ได้กับผลการจำแนกของตัวจำแนก ANN

##### (ก) วัตถุประสงค์

- เพื่อประเมินและคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล้ำมเนื้อสำหรับการจำแนกคำพูด 11 คำพูด
- เพื่อเปรียบเทียบความสามารถในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้ากล้ำมเนื้อแต่ละช่องสัญญาณ
- เพื่อเปรียบเทียบความสามารถในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้ากล้ำมเนื้อ เมื่อใช้จำนวนช่องสัญญาณที่แตกต่างกัน
- เพื่อเปรียบเทียบผลการประเมินลักษณะเด่นโดยใช้เกณฑ์ประเมิน RES และประสิทธิภาพในการจำแนกของตัวจำแนก ANN



### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครเพียง 1 คน (S01) โดยใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงและแบบไม่ออกเสียง โดยใช้สัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ สำหรับการทดลองนี้จะใช้โปรแกรม MATLAB สำหรับประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อ

#### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อมาคำนวณลักษณะเด่น 24 ชนิด ซึ่งแสดงดังตารางที่ 3.5 ซึ่งครอบคลุมลักษณะเด่นที่ถูกวิเคราะห์ในแกนเวลา (F1-F19) และลักษณะเด่นที่ถูกวิเคราะห์ในแกนความถี่ (F20-F24) โดยที่ F16-F19 เป็นลักษณะเด่นย่อยที่ถูกใช้ในงานวิจัยของเยอรมัน แสดงดังสมการ (2.17)-(2.20)

2) ทำการประเมินลักษณะเด่น 24 ชนิด โดยใช้ตัวประเมิน ซึ่งคำนวณโดยใช้สมการที่ (2.27)-(2.31) ตัวประเมินนี้เป็นการวัดความสามารถในการแบ่งกลุ่มของข้อมูล โดยวัดระยะทางระหว่างกลุ่มข้อมูลและภายในกลุ่มข้อมูล โดยใช้การวัดระยะทางแบบยุคลิด ซึ่งพบว่า การประเมินวิธีนี้สามารถประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อที่ใช้ในการจำแนกท่าทางการเคลื่อนไหวของมือได้ดี ซึ่งผลการประเมินเป็นไปทิศทางเดียวกันกับผลการจำแนกของตัวจำแนก [92] ดังนั้นจึงนำตัวประเมินนี้มาทดสอบกับงานวิจัยนี้

#### ตารางที่ 3.5 ลักษณะเด่นที่ใช้ในการทดลอง

ลักษณะเด่น		ลักษณะเด่น		ลักษณะเด่น		ลักษณะเด่น	
F1	RMS	F7	IAV	F13	MAVS_2	F19	PWH
F2	VAR	F8	MN	F14	MAVS_3	F20	MDF
F3	WL	F9	MAV	F15	MAVS_4	F21	SM1
F4	WAMP	F10	MMAV1	F16	MNH	F22	SM2
F5	ZC	F11	MMAV2	F17	MNL	F23	SM3
F6	SSC	F12	MAVS_1	F18	PWL	F24	MNF

การประเมินลักษณะเด่นจะแบ่งเป็น 4 รูปแบบ คือ

- การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้่า  
กล้่ามเนื้อทั้ง 5 ช่องสัญญาณ ( $Y = 5$ )

- การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้่า  
กล้่ามเนื้อ 3 ช่องสัญญาณ (ch1-ch3) จากตำแหน่งกล้่ามเนื้อหน้า ( $Y = 3$ )

- การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้่า  
กล้่ามเนื้อ 2 ช่องสัญญาณ (ch4-ch5) จากตำแหน่งกล้่ามเนื้อคอ ( $Y = 2$ )

- การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้่า  
กล้่ามเนื้อ 1 ช่องสัญญาณ (ประเมินแต่ละช่องสัญญาณแยกกัน) ( $Y = 1$ )

โดยปกติแล้วสัญญาณไฟฟ้่ากล้่ามเนื้อจาก 2 ช่องสัญญาณสุดท้าย ถูกออกแบบ  
มาเพื่อใช้สำหรับการจำแนกเสียงวรรณยุกต์ และสัญญาณไฟฟ้่ากล้่ามเนื้อจาก 3 ช่องสัญญาณ  
แรก ถูกออกแบบสำหรับการจำแนกคำพูดหลัก 11 คำพูด อย่างไรก็ตามการทดลองนี้จะทำการ  
ทดลองเพื่อวิเคราะห์ความสามารถในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้่า  
กล้่ามเนื้อจากช่องสัญญาณ 2 ช่องสัญญาณสุดท้ายด้วย

3) เลือกลักษณะเด่นที่ดีที่สุด 3 อันดับแรกจากผลการประเมินด้วย RES ทั้ง 4  
รูปแบบ เพื่อเข้าสู่กระบวนการฝึกสอนระบบและจำแนกระบบ (training and testing) โดยใช้ตัว  
จำแนก ANN ทำการเปรียบเทียบผลการประเมินลักษณะเด่นโดยใช้ RES กับผลความแม่นยำใน  
การจำแนกที่ได้จากตัวจำแนก ANN โดยเปรียบเทียบทั้ง 4 รูปแบบ

- รูปแบบที่หนึ่งเป็นการใช้ข้อมูลของสัญญาณไฟฟ้่ากล้่ามเนื้อทั้ง 5 ช่องสัญญาณ  
เป็นอินพุตให้กับระบบ โดยใช้ลักษณะเด่นเพียง 1 ตัว ดังนั้นจำนวนอินพุตของตัวจำแนก ANN จะ  
ประกอบไปด้วย 5 อินพุต

- รูปแบบที่สองของการฝึกสอนและทดสอบระบบ เป็นการนำข้อมูลของ  
สัญญาณไฟฟ้่ากล้่ามเนื้อจาก 3 ช่องสัญญาณแรกเป็นอินพุตให้กับระบบ โดยใช้ลักษณะเด่นเพียง  
1 ตัว ดังนั้นจำนวนอินพุตของตัวจำแนก ANN จะประกอบไปด้วย 3 อินพุต

- รูปแบบที่สามของการฝึกสอนและทดสอบระบบ เป็นการนำข้อมูลของ  
สัญญาณไฟฟ้่ากล้่ามเนื้อจาก 2 ช่องสัญญาณสุดท้ายเป็นอินพุตให้กับระบบ โดยใช้ลักษณะเด่น  
เพียง 1 ตัว ดังนั้น จำนวนอินพุตของ ANN จะประกอบไปด้วย 2 อินพุต

- รูปแบบที่สี่ของการฝึกสอนและทดสอบระบบ เป็นการให้ข้อมูลของสัญญาณไฟฟ้ากล่ามเนื้อจาก 1 ช่องสัญญาณเป็นอินพุตให้กับระบบ โดยใช้ลักษณะเด่นเพียง 1 ตัว ซึ่งพิจารณาแต่ละช่องสัญญาณ ดังนั้นจำนวนอินพุตของตัวจำแนก ANN จะประกอบไปด้วย 1 อินพุต

ในขั้นตอนนี้เลือกใช้ตัวจำแนก ANN เนื่องจากเป็นตัวจำแนกที่ถูกนำมาใช้ในงานวิจัยทางด้านนี้มากเป็นอันดับสองรองจากตัวจำแนก HMM โดยในช่วงเริ่มแรกของงานวิจัยส่วนใหญ่จะนิยมใช้ตัวจำแนก ANN มากกว่า เนื่องจากเป็นตัวจำแนกที่สามารถนำไปใช้ได้ง่าย ไม่ซับซ้อนและมีประสิทธิภาพสูง ซึ่งเหมาะกับงานวิจัยนี้ เนื่องจากต้องการจำแนกคำโดดเพียงไม่กี่คำ ไม่ได้เป็นคำพูดต่อเนื่อง และไม่ได้คำนึงถึงความสั้นความยาวของแต่ละคำพูด ซึ่งตัวจำแนก ANN จะสามารถนำมาใช้งานได้ง่ายกว่า

ตัวจำแนก ANN ที่ถูกใช้ในการทดลองนี้เป็นโครงข่ายแบบไปข้างหน้า (feed forward network) โดยมีจำนวนของชั้นซ่อน 1 ชั้น และจำนวนชั้นของเอาต์พุต 1 ชั้น โดยทั้งสองชั้นใช้ฟังก์ชันการแปลงชนิดแทนซิกมอยด์ (Tan-sigmoid function) และปรับจำนวนเซลล์ในชั้นซ่อนตามความเหมาะสม ซึ่งจากการทดลองพบว่าการใช้จำนวนเซลล์ในชั้นซ่อน 20 เซลล์เหมาะสมที่สุด

ทำการฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก ANN ทั้ง 4 รูปแบบ เช่นเดียวกับรูปแบบการประเมินความสามารถในการแยกแยะกลุ่มข้อมูลด้วย RES จากการทดลองก่อนหน้า จำนวนข้อมูลที่ใช้ในการฝึกสอนและทดสอบระบบสำหรับลักษณะเด่นแต่ละตัวมีทั้งหมด 550 ข้อมูล โดยใช้เทคนิคการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (percentage split หรือ hand out method) ซึ่งจะแบ่งข้อมูลทั้งหมดออกเป็น 2 กลุ่มตามอัตราส่วนที่กำหนดให้ ข้อมูลกลุ่มแรกจะถูกใช้สำหรับการฝึกสอนระบบและข้อมูลที่เหลือจะถูกใช้สำหรับการทดสอบระบบ ซึ่งจะแบ่งข้อมูลสำหรับฝึกสอนระบบ 80% และทดสอบระบบ 20%

4) ทำการทดลองเพิ่มความแม่นยำของระบบโดยการเพิ่มจำนวนอินพุตให้กับตัวจำแนก กล่าวคือ เพิ่มจำนวนลักษณะเด่นจากเดิม 1 ตัว เป็น 2 ตัว และ 3 ตัว ตามลำดับ

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

#### 1) ผลการประเมินลักษณะเด่นโดยใช้ RES

รูปที่ 3.24 และ รูปที่ 3.25 แสดงผลการประเมินลักษณะเด่นสำหรับ 3 รูปแบบแรก ซึ่งเป็นการเปรียบเทียบผลการประเมินโดยใช้ RES สำหรับการทดลองสามรูปแบบ คือ การใช้ข้อมูลทั้ง 5 ช่องสัญญาณ การใช้ข้อมูล 3 ช่องสัญญาณแรก และการใช้ข้อมูล 2 ช่องสัญญาณสุดท้าย สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ

สำหรับการประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อทั้ง 5 ช่องสัญญาณ จะเห็นได้ว่าลักษณะเด่นที่ให้ผลการประเมินที่ดีที่สุด คือ PWH ซึ่งให้ผลเหมือนกันทั้งการพูดแบบออกเสียงและไม่ออกเสียง

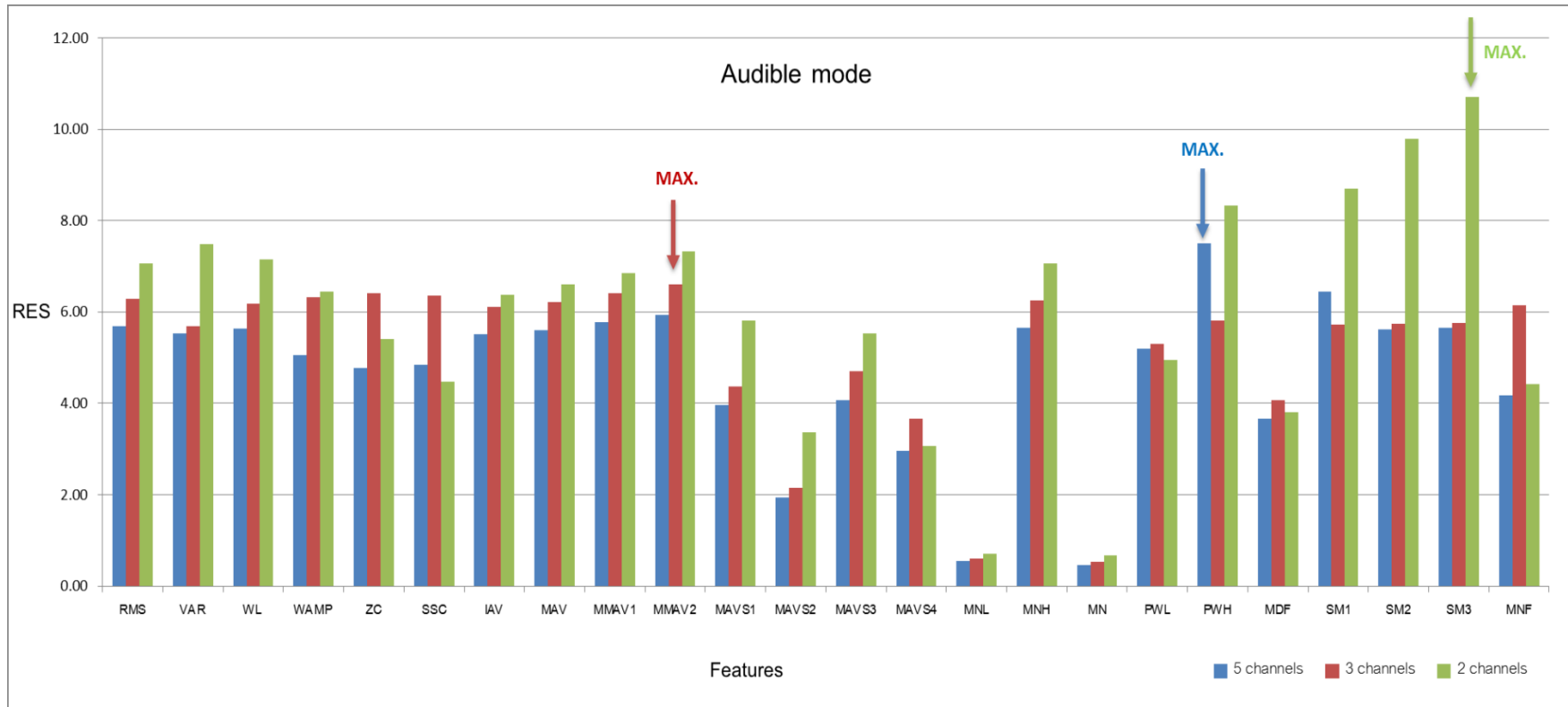
สำหรับการประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจาก 3 ช่องสัญญาณแรก ลักษณะเด่นที่ให้ผลการประเมินที่ดีที่สุดสำหรับการพูดแบบออกเสียงและไม่ออกเสียงให้ผลแตกต่างกัน สำหรับการพูดแบบออกเสียง ลักษณะเด่นที่ให้ผลการประเมินดีที่สุด คือ MMAV2 สำหรับการพูดแบบไม่ออกเสียง ลักษณะเด่นที่ให้ผลการประเมินดีที่สุดคือ MNF

สำหรับการประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 2 ช่องสัญญาณสุดท้าย ซึ่งได้ออกแบบไว้สำหรับการแยกเสียงวรรณยุกต์นั้น พบว่าเมื่อประเมินความสามารถในการจำแนกคำพูดหลัก 11 คำ ด้วย RES พบว่าให้ผลการประเมินที่ดีเช่นเดียวกับการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจาก 3 ช่องสัญญาณแรก ซึ่งแสดงว่าสัญญาณไฟฟ้ากล่อมเนื้อ 2 ช่องสัญญาณนี้สามารถใช้ในการจำแนกคำพูดหลักได้ด้วย โดยลักษณะเด่นที่ให้ผลการประเมินสูงสุด คือ SM3 ซึ่งให้ผลเหมือนกันทั้งการพูดแบบออกเสียงและไม่ออกเสียง

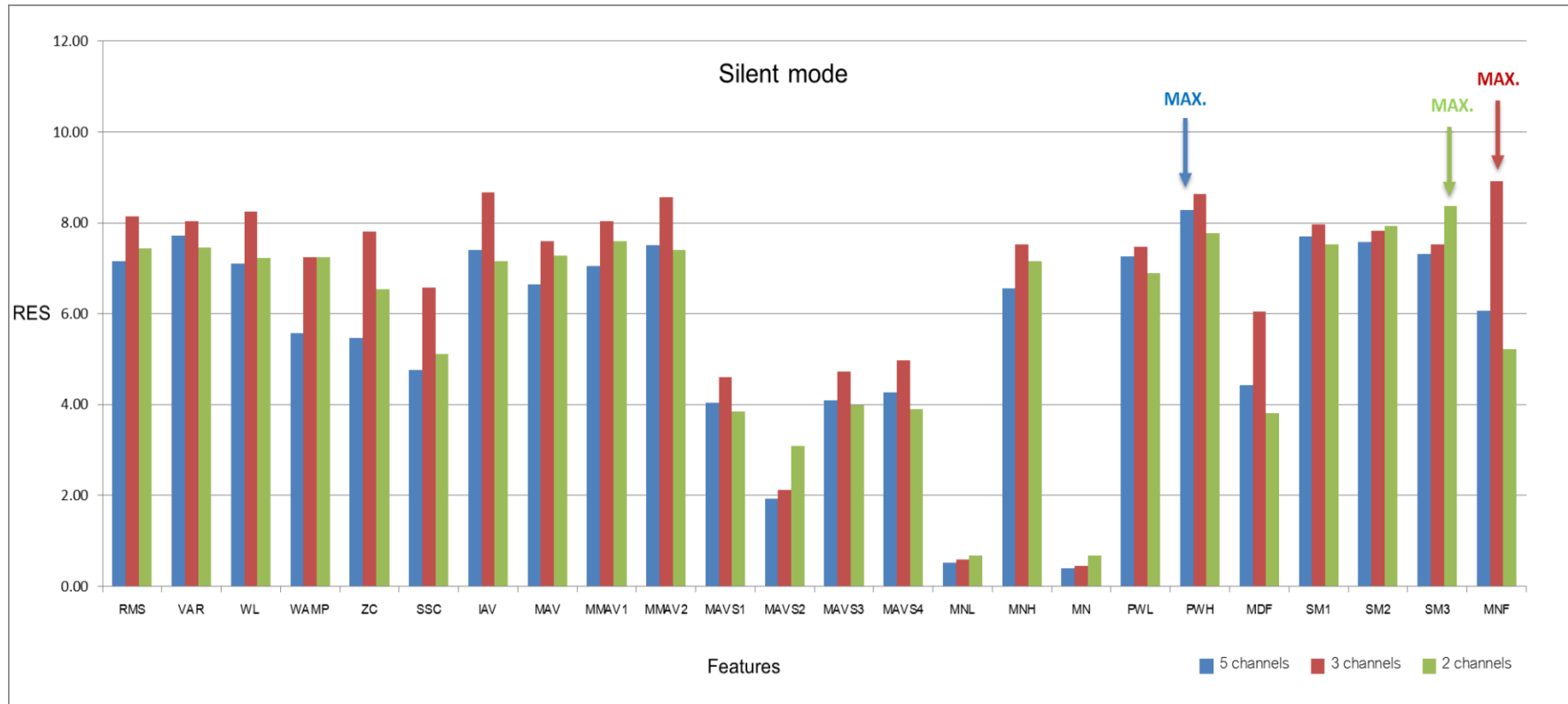
เมื่อเปรียบเทียบผลการประเมินทั้ง 3 รูปแบบ จะเห็นได้ว่าสำหรับการพูดแบบออกเสียง การประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 2 ช่องสัญญาณสุดท้าย ให้ผลการประเมินที่สูงที่สุด (RES สูงสุด) สำหรับการพูดแบบไม่ออกเสียงให้ผลที่แตกต่างกับการพูดแบบออกเสียง กล่าวคือ การประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 3 ช่องสัญญาณแรกให้ผลการประเมินที่ดีที่สุด

จากผลการทดลองนี้จะเห็นได้ว่า สัญญาณไฟฟ้ากล่อมเนื้อจาก 2 ช่องสัญญาณสุดท้ายซึ่งเป็นตำแหน่งของกล่อมเนื้อคอ ซึ่งถูกออกแบบไว้สำหรับการจำแนกรรณยุกต์ สามารถนำมาใช้ในการจำแนกคำพูดหลัก (11 คำพูด) ได้ด้วย นอกจากนี้ยังพบว่าให้ผลการประเมินที่ดีกว่าการใช้สัญญาณไฟฟ้ากล่อมเนื้อ 3 ช่องสัญญาณแรก อย่างไรก็ตามสำหรับการพูดแบบไม่ออกเสียง สัญญาณไฟฟ้ากล่อมเนื้อจาก 3 ช่องสัญญาณแรกให้ผลการประเมินที่ดีกว่า ทั้งนี้เนื่องจากการพูดแบบไม่ออกเสียงนั้น ไม่ค่อยมีทำงานของกล่อมเนื้อบริเวณคอ ทำให้ขนาดของสัญญาณน้อย

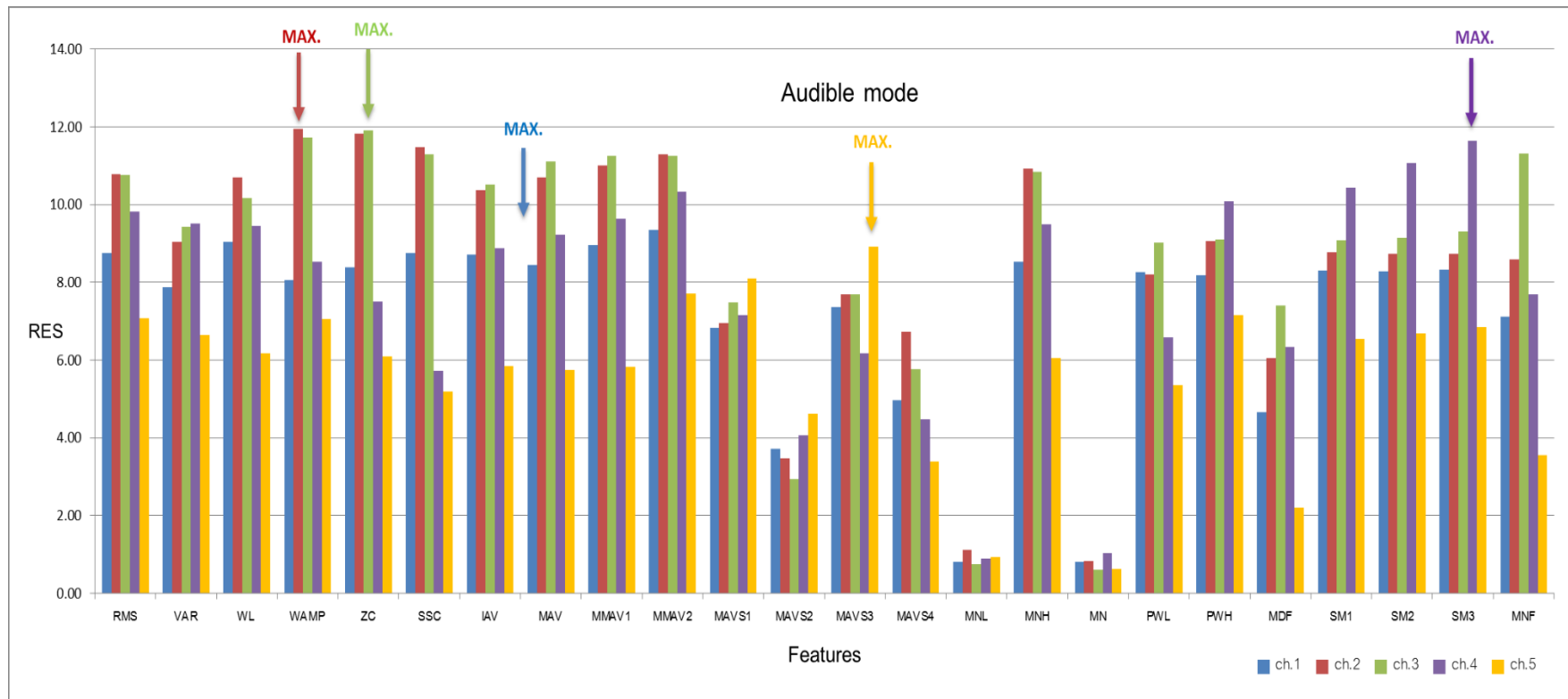
จากนั้นพิจารณาผลการประเมินลักษณะเด่นสำหรับแต่ละช่องสัญญาณแยกกัน ซึ่งผลการประเมินแสดงดังรูปที่ 3.26 และ รูปที่ 3.27 สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ เมื่อเปรียบเทียบผลการประเมินสำหรับแต่ละช่องสัญญาณ พบว่าลักษณะเด่นที่มีผลการประเมินที่ดีที่สุดในแต่ละช่องสัญญาณแตกต่างกัน นอกจากนี้เมื่อพิจารณาจากลักษณะเด่นส่วนใหญ่ พบว่าสัญญาณช่องที่ 2 และ 3 มีผลการประเมินที่ดีกว่าสัญญาณช่องอื่น โดยที่สัญญาณช่องที่ 4 และ 5 มีผลการประเมินที่ต่ำที่สุด ซึ่งให้ผลการประเมินตรงกันข้ามกับการทดลองก่อนหน้า กล่าวคือ ผลการประเมินสำหรับการทดลองสามรูปแบบแรก พบว่าการใช้สัญญาณไฟฟ้ากล่อมเนื้อจากช่องสัญญาณที่ 4 และ ช่องสัญญาณที่ 5 ให้ผลการประเมินที่ดีกว่าการใช้สัญญาณไฟฟ้ากล่อมเนื้อจาก 3 ช่องสัญญาณแรก (สำหรับการพูดแบบออกเสียง) ดังนั้นในเบื้องต้น สามารถตั้งสมมุติฐานได้ว่าไม่สามารถนำผลการประเมินที่คำนวณจากข้อมูลที่มีจำนวนมิติต่างกันมาเปรียบเทียบกันได้ กล่าวคือ การทดลองสามรูปแบบแรก มีจำนวนมิติของข้อมูล 5 มิติ 3 มิติ 2 มิติ ตามลำดับ การทดลองรูปแบบสุดท้ายมีจำนวนมิติของข้อมูล 1 มิติ ซึ่งไม่สามารถนำค่า RES ของทั้ง 4 รูปแบบมาเปรียบเทียบกันได้ อย่างไรก็ตามค่า RES ที่คำนวณจากข้อมูลที่มีมิติเดียวกันสามารถนำมาเปรียบเทียบกันได้ ซึ่งจะทำการทดสอบสมมุติฐานนี้ จากผลการจำแนกโดยตัวจำแนก ANN



รูปที่ 3.24 กราฟแสดงการเปรียบเทียบผลการประเมินโดยใช้ RES สำหรับการทดลองสามรูปแบบแรก คือ การใช้ข้อมูล 5 ช่องสัญญาณ 3 ช่องสัญญาณ และ 2 ช่องสัญญาณสุดท้าย สำหรับการพูดแบบออกเสียง

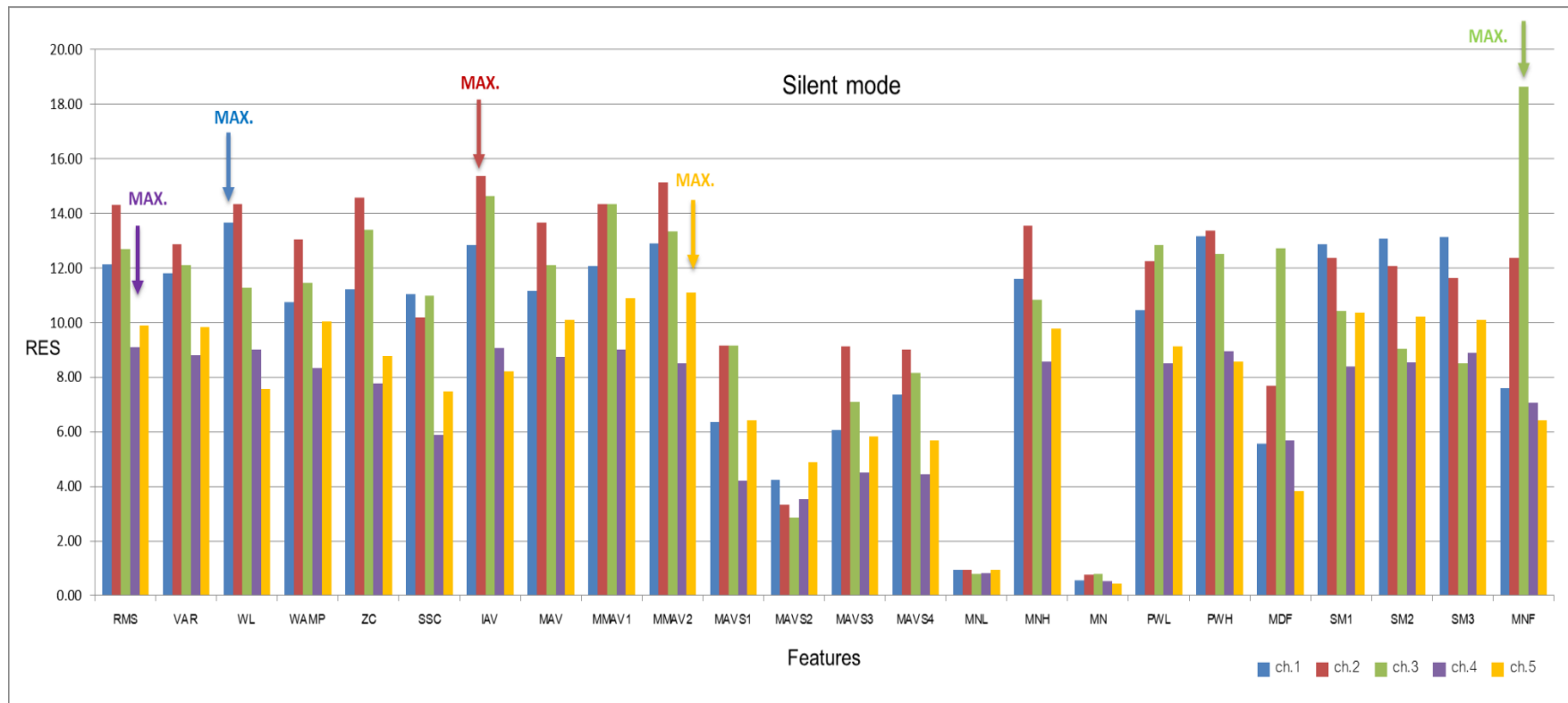


รูปที่ 3.25 กราฟแสดงการเปรียบเทียบผลการประเมินโดยใช้ RES สำหรับการทดลองสามรูปแบบแรก คือ การใช้ข้อมูล 5 ช่องสัญญาณ 3 ช่องสัญญาณ และ 2 ช่องสัญญาณสุดท้าย สำหรับการพูดแบบไม่ออกเสียง



รูปที่ 3.26 กราฟแสดงผลการประเมินโดยใช้ RES สำหรับการทดลองรูปแบบสุดท้าย การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า กล้ามเนื้อ 1 ช่องสัญญาณ (การพิจารณาแต่ละช่องสัญญาณ) สำหรับการพูดแบบออกเสียง





รูปที่ 3.27 กราฟแสดงผลการประเมินโดยใช้ RES สำหรับการทดลองรูปแบบสุดท้าย การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า กล้ามเนื้อ 1 ช่องสัญญาณ (การพิจารณาแต่ละช่องสัญญาณ) สำหรับการพูดแบบไม่ออกเสียง

## 2) ผลการจำแนกโดยใช้ตัวจำแนก ANN

ตารางที่ 3.6 แสดงการเปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับรูปแบบที่หนึ่ง ซึ่งพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า 5 ช่องสัญญาณ โดยเปรียบเทียบประสิทธิภาพของลักษณะเด่น 3 ชนิดที่มีผลการประเมินด้วย RES ดีที่สุด จะเห็นได้ว่าผลความแม่นยำในการจำแนกโดยใช้ตัวจำแนก ANN มีทิศทางเดียวกันกับผลการประเมินด้วย RES โดยที่ลักษณะเด่นแบบ PWH ให้ความแม่นยำสูงสุด ทั้งการพูดออกเสียงและไม่ออกเสียง สำหรับการพูดแบบออกเสียงให้ความแม่นยำ 80.73% ส่วนการพูดแบบไม่ออกเสียงให้ความแม่นยำ 81.93 % ซึ่งสูงกว่าแบบออกเสียงเล็กน้อย

**ตารางที่ 3.6** เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า 5 ช่องสัญญาณ (รูปแบบที่หนึ่ง)

รูปแบบที่หนึ่ง : การใช้ข้อมูลทั้ง 5 ช่องสัญญาณ					
ออกเสียง			ไม่ออกเสียง		
ลักษณะเด่น	RES	Accuracy rate (%)	ลักษณะเด่น	RES	Accuracy Rate (%)
PWH	7.51	80.73	PWH	8.29	81.93
SM1	6.45	78.31	VAR	7.72	79.52
MMAV2	5.93	77.11	SM1	7.70	78.31

**ตารางที่ 3.7** เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้า 3 ช่องสัญญาณแรก (รูปแบบที่สอง)

รูปแบบที่สอง : การใช้ข้อมูล 3 ช่องสัญญาณแรก					
ออกเสียง			ไม่ออกเสียง		
ลักษณะเด่น	RES	Accuracy rate (%)	ลักษณะเด่น	RES	Accuracy Rate (%)
MMAV2	6.59	65.06	MNF	8.91	68.67
ZC	6.41	65.06	IAV	8.67	69.87
MMAV1	6.41	62.65	PWH	8.63	69.87

**ตารางที่ 3.8** เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 2 ช่องสัญญาณสุดท้าย (รูปแบบที่สาม)

รูปแบบที่สาม : การใช้ข้อมูล 2 ช่องสัญญาณสุดท้าย					
ออกเสียง			ไม่ออกเสียง		
ลักษณะเด่น	RES	Accuracy rate (%)	ลักษณะเด่น	RES	Accuracy Rate (%)
SM3	10.70	61.44	SM3	8.37	56.62
SM2	9.80	57.83	SM2	7.92	56.62
SM1	8.70	57.83	PWH	7.77	55.83

ตารางที่ 3.7 แสดงการเปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับรูปแบบที่สอง ซึ่งพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 3 ช่องสัญญาณแรก โดยเปรียบเทียบประสิทธิภาพของลักษณะเด่น 3 ชนิดที่มีผลการประเมินด้วย RES ดีที่สุด จะเห็นได้ว่าผลความแม่นยำในการจำแนกโดยตัวจำแนก ANN ของลักษณะเด่นทั้ง 3 ชนิดมีผลความแม่นยำที่ใกล้เคียงกันมาก ซึ่งสอดคล้องกับผลการประเมินด้วย RES นอกจากนี้จะเห็นได้ว่า เมื่อเปรียบเทียบความแม่นยำในการจำแนกกับรูปแบบแรกซึ่งใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อทั้ง 5 ช่องสัญญาณ รูปแบบที่สองให้ผลความแม่นยำต่ำกว่า

ตารางที่ 3.8 แสดงการเปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับรูปแบบที่สาม ซึ่งพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 2 ช่องสัญญาณสุดท้าย โดยเปรียบเทียบประสิทธิภาพของลักษณะเด่น 3 ชนิดที่มีผลการประเมินด้วย RES ดีที่สุด จะเห็นได้ว่าผลความแม่นยำในการจำแนกโดยตัวจำแนก ANN มีทิศทางเดียวกันกับผลการประเมินด้วย RES เช่นเดียวกัน

ตารางที่ 3.9 แสดงการเปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับรูปแบบที่สี่ ซึ่งพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อ 1 ช่องสัญญาณ โดยเปรียบเทียบประสิทธิภาพของลักษณะเด่น 3 ชนิดที่มีผลการประเมินด้วย RES ดีที่สุด จะเห็นได้ว่าผลความแม่นยำในการจำแนกโดยตัวจำแนก ANN ค่อนข้างต่ำมาก กล่าวคือ ส่วนใหญ่ให้ความแม่นยำต่ำกว่า 50% ซึ่งตรงกันข้ามกับผลการประเมินโดยใช้ RES ซึ่งพบว่าผลการประเมินรูปแบบนี้ให้ค่า RES ค่อนข้างสูงกว่าสามรูปแบบแรก เมื่อ

พิจารณาช่องสัญญาณที่ให้ผลความแม่นยำในการจำแนกสูงสุด พบว่าสัญญาณจากช่องที่ 3 และ 4 ให้ความแม่นยำสูงสุดสำหรับการพูดแบบออกเสียง สำหรับการพูดแบบไม่ออกเสียง ช่องสัญญาณที่ 2 และ 3 ให้ความแม่นยำในการจำแนกสูงสุด

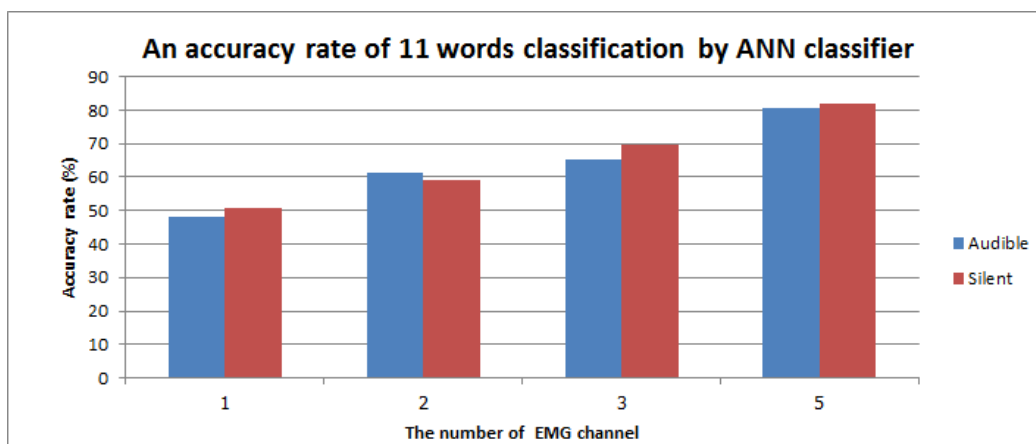
**ตารางที่ 3.9** เปรียบเทียบผลการประเมินโดยใช้ RES และผลการจำแนกโดยใช้ตัวจำแนก ANN สำหรับลักษณะเด่น 3 ชนิด เมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อสำหรับแต่ละช่องสัญญาณ (รูปแบบที่สี่)

รูปแบบที่สี่ : การใช้ข้อมูล 1 ช่องสัญญาณ						
ช่องสัญญาณ	ออกเสียง			ไม่ออกเสียง		
	ลักษณะเด่น	RES	Accuracy rate (%)	ลักษณะเด่น	RES	Accuracy rate (%)
ch.1	MMAV2	9.35	38.55	WL	13.67	49.40
	WL	9.05	37.35	PWH	13.18	49.40
	MMAV1	8.96	37.35	SM3	13.12	44.58
ch.2	WAMP	11.94	46.99	IAV	15.37	50.60
	ZC	11.83	43.37	MMAV2	14.14	50.60
	SSC	11.48	43.37	ZC	14.58	50.60
ch.3	ZC	11.90	48.19	MNF	18.64	37.34
	WAMP	11.72	46.99	IAV	14.65	49.39
	MNF	11.32	39.75	MMAV1	14.34	50.60
ch.4	SM3	11.65	48.19	RMS	9.10	39.76
	MMAV2	10.33	48.19	WL	9.01	38.55
	PWH	10.08	46.98	MMAV1	9.01	39.76
ch.5	MAVS_3	8.91	38.55	MMAV2	11.09	45.78
	MAVS_1	8.10	38.55	MMAV1	10.89	45.78
	MMAV2	7.71	34.93	SM1	10.37	43.37

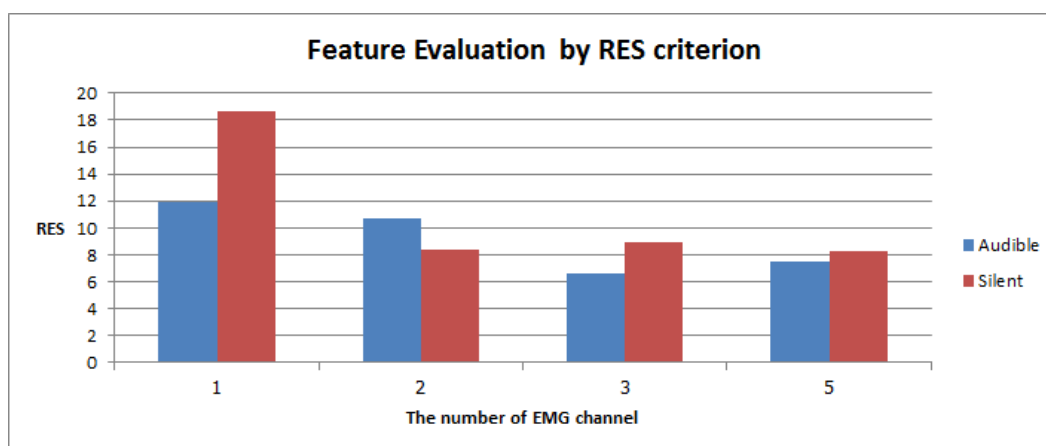
จากทั้ง 4 รูปแบบของการฝึกสอนและทดสอบระบบโดยใช้จำนวนช่องสัญญาณที่ต่างกันเป็นอินพุตให้กับตัวจำแนก ANN พบว่าการใช้จำนวนช่องสัญญาณมากขึ้นทำให้ระบบมีความแม่นยำมากขึ้นซึ่งสอดคล้องกับผลการวิจัยของงานวิจัยอื่นๆ [18], [41] โดยการทดลองนี้พบว่าการใช้จำนวนช่องสัญญาณ 5 ช่องสัญญาณเป็นอินพุตให้กับระบบได้รับความแม่นยำมากที่สุด ดังรูปที่ 3.28 ในทางตรงกันข้ามเมื่อพิจารณาผลการประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้ RES ดังรูปที่ 3.29 พบว่าการใช้จำนวนช่องสัญญาณมากขึ้นทำให้ค่า RES มีค่าลดลง นั่นคือความสามารถในการจำแนกลดลง ซึ่งเป็นผลการประเมินที่ไม่ถูกต้องเมื่อเทียบกับผลการจำแนกโดยใช้ตัวจำแนก ANN ดังนั้นจากสมมุติฐานก่อนหน้านี้นี้ สามารถสรุปได้ว่าไม่สามารถใช้ RES สำหรับการประเมินความสามารถในการคัดแยกกลุ่มของข้อมูลที่มีมิติต่างกันได้ กล่าวคือ ไม่สามารถนำผลการประเมิน RES ของข้อมูลที่มาจากการคำนวณ 5 ช่องสัญญาณ 3 ช่องสัญญาณ 2 ช่องสัญญาณ และ 1 ช่องสัญญาณมาเปรียบเทียบกันได้ อย่างไรก็ตามสามารถนำผลการประเมิน RES เปรียบเทียบกันได้ภายใต้ข้อมูลที่มีมิติเดียวกัน เช่น การเปรียบเทียบค่า RES ของลักษณะเด่นชนิดต่างๆ ที่มาจากการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ เป็นต้น เนื่องจากผลการทดลองส่วนใหญ่พบว่าข้อมูลที่มีมิติเดียวกัน ให้ผลการประเมินโดยใช้ค่า RES ไปในทิศทางเดียวกันกับผลการจำแนกโดยใช้ตัวจำแนก ANN

ในส่วนของการทดลองนี้ ความแม่นยำสูงสุดในการจำแนกคำพูดหลัก 11 คำ โดยใช้ตัวจำแนก ANN คือ 81.93% ซึ่งได้จากการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ โดยใช้ลักษณะเด่นแบบ PWH

ตารางที่ 3.10 แสดงผลการความแม่นยำในการจำแนก เมื่อทำการทดลองเพิ่มความแม่นยำของระบบโดยการเพิ่มจำนวนอินพุตให้กับตัวจำแนก ANN กล่าวคือ เพิ่มจำนวนลักษณะเด่นจากเดิม 1 ตัว เป็น 2 ตัว และ 3 ตัว ตามลำดับ โดยที่ยังคงใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณเหมือนเดิม ดังนั้นจำนวนอินพุตของโครงข่ายจะถูกเพิ่มเป็น 10 และ 15 ตามลำดับ ซึ่งจะเลือกใช้ลักษณะเด่น 3 ชนิดที่มีผลการประเมินที่ดีที่สุดจากรูปแบบการประเมินที่พิจารณาจากการใช้จำนวนข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ (รูปแบบที่หนึ่ง) โดยแต่ละช่องสัญญาณจะใช้ลักษณะเด่นเดียวกัน ซึ่งได้ผลการทดลองแสดงให้เห็นว่าการเพิ่มจำนวนลักษณะเด่นสามารถปรับปรุงความแม่นยำของระบบได้ สำหรับการพูดแบบออกเสียงให้ผลความแม่นยำเพิ่มขึ้นเป็น 92.77% และ 85.54% สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ นอกจากนี้จะสังเกตได้ว่า การพูดแบบออกเสียงจะให้ความแม่นยำที่สูงกว่าการพูดแบบไม่ออกเสียงประมาณ 7%



**รูปที่ 3.28** กราฟเปรียบเทียบผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูดโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวนช่องสัญญาณที่แตกต่างกัน



**รูปที่ 3.29** กราฟเปรียบเทียบผลการประเมินความสามารถในการจำแนกกลุ่มของลักษณะเด่นสัญญาณไฟฟ้ากล้ามเนื้อ โดยใช้ RES เมื่อใช้จำนวนช่องสัญญาณที่แตกต่างกัน

**ตารางที่ 3.10** ผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวนลักษณะเด่นที่แตกต่างกัน และใช้จำนวนข้อมูลทั้ง 5 ช่องสัญญาณ

จำนวนของ ลักษณะเด่น	ออกเสียง		ไม่ออกเสียง	
	ลักษณะเด่น	Accuracy rate (%)	ลักษณะเด่น	Accuracy rate (%)
1	PWH	80.73	PWH	81.93
	PWH+SM1	91.57	PWH+VAR	84.34
2	PWH+MMAV2	89.16	PWH+SM1	84.34
	SM1+MMAV2	91.56	VAR+SM1	84.34
3	PWH+SM1+MMAV2	92.77	PWH+VAR+SM1	85.54

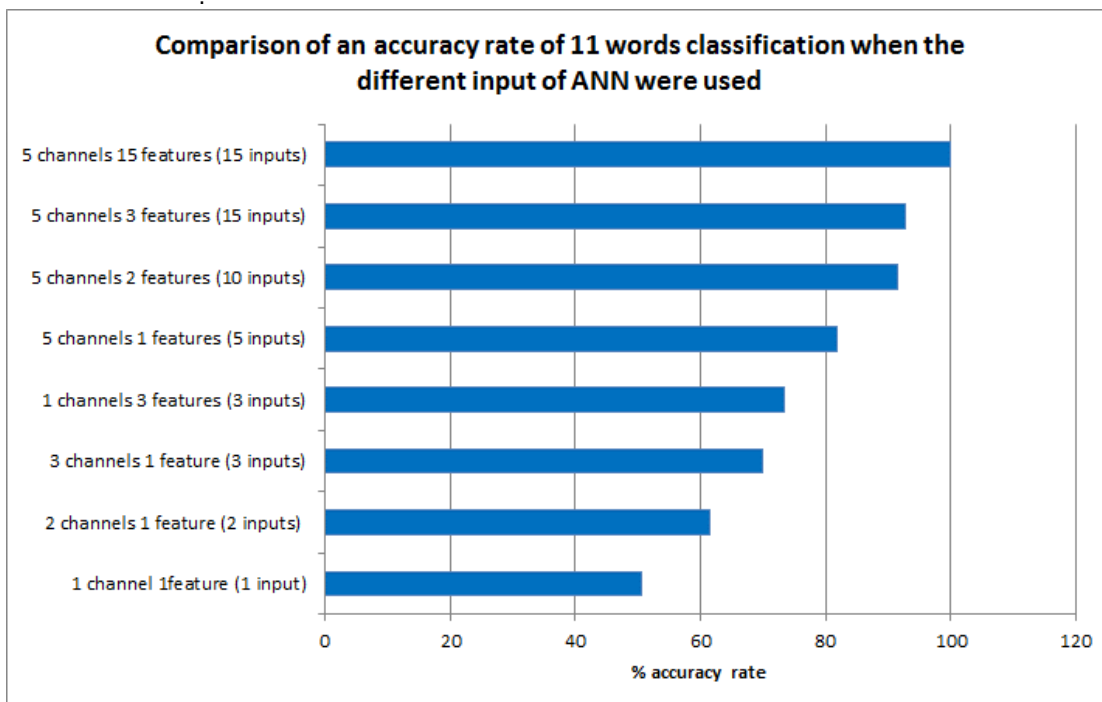
เนื่องจากผลการทดลองก่อนหน้านี้ จะเห็นได้ว่าลักษณะเด่นที่ดีที่สุดในแต่ละช่องสัญญาณแตกต่างกัน เป็นไปได้ว่าการรวมกันของลักษณะที่ดีที่สุดจากแต่ละช่องสัญญาณน่าจะเพิ่มความแม่นยำให้กับระบบมากกว่าการใช้ลักษณะเด่นชนิดเดียวกันทั้ง 5 ช่องสัญญาณ ถึงแม้ว่าจำนวนอินพุตของระบบจะเท่ากัน ดังนั้นทำการทดลองเพิ่มเติมโดยเลือกลักษณะเด่นที่ดีที่สุด 3 ชนิดของแต่ละช่องสัญญาณจากผลการประเมินในรูปแบบที่สี่ เปรียบเทียบผลความแม่นยำในการจำแนกโดยใช้ตัวจำแนก ANN เมื่อเพิ่มจำนวนลักษณะเด่นในแต่ละช่องสัญญาณจาก 1 ตัว เป็น 2 ตัว และ 3 ตัวตามลำดับ

ผลการทดลองแสดงดังตารางที่ 3.11 จะเห็นได้ว่าการใช้จำนวนลักษณะเด่นเพียง 1 ตัวสำหรับแต่ละช่องสัญญาณ ผลความแม่นยำในการจำแนกต่ำมาก เมื่อเพิ่มจำนวนลักษณะเด่นเป็น 3 ตัว จะทำให้แต่ละช่องสัญญาณมีความแม่นยำเพิ่มขึ้น แต่ก็ถือว่ายังให้ความแม่นยำที่ต่ำอยู่ โดยจะเห็นได้ว่าช่องสัญญาณที่ 4 ให้ความแม่นยำดีที่สุดสำหรับการพูดแบบออกเสียง และช่องสัญญาณที่ 2 ให้ความแม่นยำสูงสุดสำหรับการพูดแบบไม่ออกเสียง ต่อมาเมื่อนำลักษณะเด่นที่ดีที่สุดทั้ง 3 ตัวจากทั้ง 5 ช่องสัญญาณมารวมกัน (ซึ่งจะเห็นได้ว่าเป็นลักษณะเด่นที่แตกต่างกันในแต่ละช่องสัญญาณ) เพื่อใช้เป็นอินพุตให้กับตัวจำแนก ANN จะเห็นว่าสามารถเพิ่มความแม่นยำสูงสุด 100 % สำหรับการพูดแบบออกเสียง และ 95.18% สำหรับการพูดแบบไม่ออกเสียง

ตารางที่ 3.11 ผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อใช้จำนวนลักษณะเด่นที่แตกต่างกัน โดยพิจารณาลักษณะเด่นที่ดีที่สุดในแต่ละช่องสัญญาณ

ช่อง สัญญาณ	ลักษณะ เด่น	ออกเสียง			ไม่ออกเสียง			
		Accuracy rate (%)			ลักษณะ เด่น	Accuracy rate (%)		
		จำนวนลักษณะเด่น				จำนวนลักษณะเด่น		
		1	3	15	1	3	15	
ch.1	MMAV2	38.55			WL	43.37		
	WL	42.17	57.83		PWH	49.40	69.88	
	MMAV1	37.35			SM3	44.58		
ch.2	WAMP	46.99			IAV	48.19		
	ZC	43.37	65.06		MMAV2	50.60	73.49	
	SSC	43.37			ZC	50.60		
ch.3	ZC	46.99			MNF	37.34		
	WAMP	48.19	69.88	100	IAV	49.39	68.67	95.18
	MNF	39.75			MMAV1	50.60		
ch.4	SM3	40.96			RMS	39.76		
	MMAV2	48.19	72.29		WL	38.55	50.60	
	PWH	46.98			MMAV1	39.76		
ch.5	MAVS_3	37.35			MMAV2	40.96		
	MAVS_1	38.55	54.22		MMAV1	45.78	54.22	
	MMAV2	34.93			SM1	43.37		
*หมายเหตุ จำนวนลักษณะเด่น 15 ตัว = 3 ลักษณะเด่น x 5 ช่องสัญญาณ								





**รูปที่ 3.30** กราฟเปรียบเทียบผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูด สำหรับการใช้จำนวนอินพุตของตัวจำแนก ANN ที่แตกต่างกัน

จากรูปที่ 3.30 เป็นกราฟเปรียบเทียบผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูด สำหรับการใช้จำนวนอินพุตของตัวจำแนก ANN ที่แตกต่างกัน ซึ่งจะเห็นได้ว่าการป้อนจำนวนอินพุต 15 ตัว ให้กับโครงข่ายประสาทให้ความแม่นยำสูงสุด โดยที่อินพุตดังกล่าวได้มาจากลักษณะเด่นที่ดีที่สุด 3 ตัวในแต่ละช่องสัญญาณ

### (ง) สรุปผลการทดลอง

การทดลองนี้เป็นการทดลองเบื้องต้นสำหรับการออกแบบระบบจำแนกคำพูดหลัก 11 คำ ("ศูนย์" ถึง "สิบ") โดยใช้สัญญาณไฟฟ้ากล้ามเนื้อเป็นอินพุตให้กับระบบ เพื่อคัดเลือกลักษณะเด่นที่เหมาะสม โดยใช้ RES เป็นตัวประเมินลักษณะเด่น และเปรียบเทียบผลการประเมินที่ได้กับความถูกต้องในการจำแนกของตัวจำแนก ANN

คำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ 24 ชนิด และประเมินความสามารถในการแยกแยะกลุ่มออกเป็น 11 กลุ่ม โดยใช้ตัวประเมิน RES ซึ่งประเมิน 4 รูปแบบ คือ การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5

ช่องสัญญาณ 3 ช่องสัญญาณจากตำแหน่งกล่อมเนื้อหน้า 2 ช่องสัญญาณจากตำแหน่งกล่อมเนื้อคอ และ การประเมินแต่ละช่องสัญญาณแยกกัน หลังจากนั้นนำลักษณะเด่นที่มีผลการประเมินที่ดี 3 อันแรกของแต่ละรูปแบบการประเมินมาทำการฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก ANN

จากการทดลองพบว่า ผลการประเมินลักษณะเด่นด้วย RES และผลความแม่นยำในการจำแนกโดยใช้ตัวจำแนก ANN มีทิศทางเดียวกัน ดังนั้น RES มีประสิทธิภาพในการประเมินลักษณะเด่น นอกจากนี้เมื่อเปรียบเทียบความสามารถในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้ากล่อมเนื้อแต่ละช่องสัญญาณ พบว่าสัญญาณช่องที่ 2 และ 3 ให้ผลความแม่นยำที่ดีที่สุด สำหรับการพูดแบบออกเสียง และ สำหรับการพูดแบบไม่ออกเสียง สัญญาณช่องที่ 1 และ 2 ให้ผลความแม่นยำที่ดีที่สุด จากผลการประเมินจะเห็นได้ว่าสัญญาณจากช่องสัญญาณที่ 4 และ 5 ซึ่งเป็นสัญญาณที่ออกแบบไว้สำหรับการจำแนกวรรณยุกต์สามารถนำมาจำแนกคำพูดหลัก 11 คำพูดได้เช่นเดียวกัน เนื่องจากให้ผลการประเมินใกล้เคียงกับสัญญาณจากสัญญาณ 3 ช่องแรก

จากผลความแม่นยำในการจำแนกโดยใช้ตัวจำแนก ANN พบว่าการเพิ่มจำนวนของช่องสัญญาณมากขึ้น สามารถเพิ่มความแม่นยำให้กับระบบ ซึ่งในที่นี้การใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อทั้ง 5 ช่องสัญญาณโดยใช้ลักษณะเด่น 1 ตัว เป็นอินพุตให้กับตัวจำแนก ANN ให้ค่าความแม่นยำสูงสุด 81.93% หลังจากนั้นปรับปรุงระบบโดยการเพิ่มจำนวนลักษณะเด่นเป็น 3 ตัว โดยแต่ละช่องสัญญาณใช้ลักษณะเด่นชนิดเดียวกัน พบว่าสามารถเพิ่มความแม่นยำได้ถึง 92.77%

นอกจากนี้ การทดลองนี้ได้ทำการปรับปรุงระบบเพิ่มเติม ซึ่งทำการทดลองโดยใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อทั้ง 5 ช่องสัญญาณเช่นเดียวกัน แต่เลือกลักษณะเด่นที่ดีที่สุด 3 ตัว จากผลการประเมินด้วย RES ของแต่ละช่องสัญญาณ ดังนั้นลักษณะเด่นทั้งหมดที่ใช้เป็นอินพุตให้กับตัวจำแนก ANN มีจำนวน 15 ตัว จากผลการจำแนกโดยใช้ตัวจำแนก ANN พบว่าสามารถให้ความแม่นยำสูงสุดถึง 100 %

อย่างไรก็ตามในการทดลองนี้เป็นการทดลองเบื้องต้นสำหรับการออกแบบระบบการจำแนกสัญญาณไฟฟ้ากล่อมเนื้อของการพูดตัวเลขาภาษาไทย 11 คำ โดยขึ้นอยู่กับผู้พูด ซึ่งใช้ข้อมูลจากอาสาสมัครเพียง 1 คน ดังนั้นการทดลองต่อไปจะกล่าวถึงการออกแบบโดยอ้างอิงข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน

### 3.3.1.2 การทดลองเพื่อคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกคำพูด 11 คำพูด สำหรับสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน โดยใช้ RES เป็นตัวประเมินลักษณะเด่น และวัดประสิทธิภาพในการจำแนกโดยใช้ตัวจำแนก ANN

การทดลองนี้เป็นการทดลองเพื่อออกแบบระบบการจำแนกสัญญาณไฟฟ้ากล่อมเนื้อของการพูดตัวเลขภาษาไทย 11 คำ โดยอ้างอิงข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน เพื่อให้เทคนิคที่ออกแบบสามารถนำไปใช้ได้กับข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมด

#### (ก) วัตถุประสงค์

- เพื่อประเมินและคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อสำหรับการจำแนกคำพูด 11 คำ โดยอ้างอิงข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน
- เพื่อหาจำนวนลักษณะเด่นที่เหมาะสม สำหรับการจำแนกคำพูด 11 คำ เพื่อเพิ่มความแม่นยำในการจำแนก
- เพื่อเปรียบเทียบวิธีการประเมินระหว่างการประเมินลักษณะเด่นในแต่ละช่องสัญญาณแยกกัน และ การประเมินลักษณะเด่นโดยพิจารณาข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อรวมกันทั้ง 5 ช่องสัญญาณ

#### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) โดยใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงและแบบไม่ออกเสียง สำหรับอาสาสมัครแต่ละคนจะข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจะมีทั้งหมด 550 ข้อมูล (สำหรับอาสาสมัครแต่ละคนและสำหรับแต่ละรูปแบบการออกเสียง) โดยที่จำนวน 80% ของข้อมูลจะถูกใช้ในกระบวนการฝึกสอนระบบ และ 20% ของข้อมูลจะถูกใช้ในกระบวนการทดสอบระบบ ซึ่งเป็นการฝึกสอนระบบและทดสอบระบบแบบขึ้นกับผู้พูด กล่าวคือ ข้อมูลในการฝึกสอนระบบและทดสอบระบบต้องเป็นข้อมูลที่มาจากอาสาสมัครคนเดียวกัน

### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้าทุกแฉกมาคำนวณลักษณะเด่น 26 ชนิด ซึ่งแสดงดังตารางที่ 3.12 ซึ่งเพิ่มเติมจากการทดลองแรก (หัวข้อ 3.3.1.1) 2 ชนิด คือ ZCH และ SM5

2) ทำการประเมินลักษณะเด่นทั้งหมด โดยใช้ตัวประเมิน RES ซึ่งการทดลองนี้จะประเมิน 2 รูปแบบ คือ

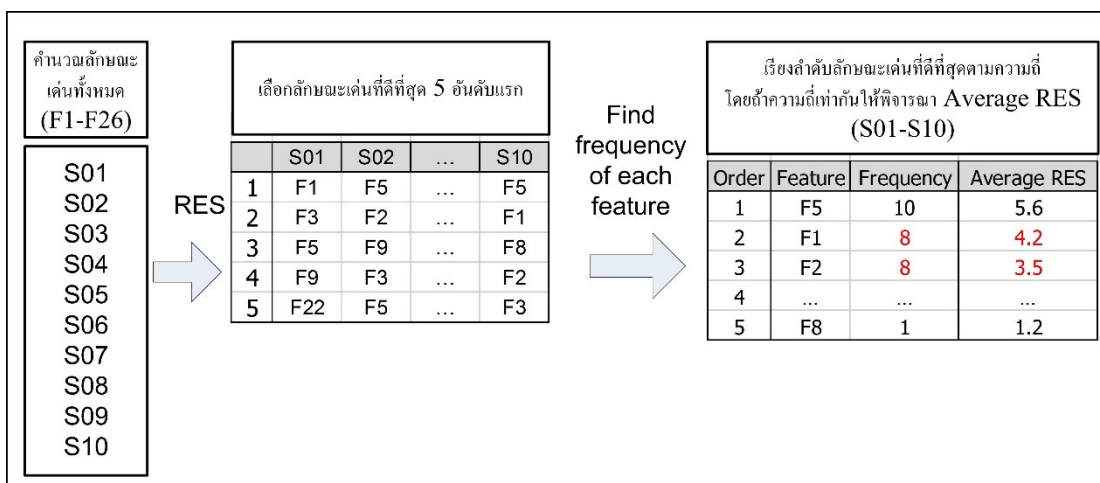
- รูปแบบที่หนึ่ง ประเมินลักษณะเด่นในแต่ละช่องสัญญาณแยกกัน ซึ่งรูปแบบนี้ในแต่ละช่องสัญญาณจะได้ลักษณะเด่นที่ดีที่สุดแตกต่างกัน

- รูปแบบที่สอง ประเมินลักษณะเด่นโดยพิจารณาข้อมูลของสัญญาณไฟฟ้าทุกแฉกมาคำนวณร่วมกันทั้ง 5 ช่องสัญญาณ ในรูปแบบนี้ทุกช่องสัญญาณจะได้ลักษณะเด่นชนิดเดียวกัน

3) เลือกลักษณะเด่นที่ดีที่สุด 5 อันดับแรกของสัญญาณไฟฟ้าทุกแฉกของอาสาสมัครแต่ละคนจากวิธีการประเมินข้างต้น นำลักษณะเด่นที่ถูกเลือกทั้งหมดในกลุ่มนี้มาพิจารณาหาความถี่ นั่นคือ จำนวนครั้งที่ลักษณะเด่นนั้นๆ จะปรากฏอยู่ในกลุ่มของลักษณะเด่นที่ดีที่สุด 5 อันดับแรกของสัญญาณไฟฟ้าทุกแฉกของอาสาสมัครแต่ละคน (ทั้ง 10 คน) ซึ่งความถี่สูงสุดเท่ากับ 10 หลังจากนั้นเรียงลำดับลักษณะเด่นดังกล่าวโดยพิจารณาจากความถี่ ซึ่งถ้าลักษณะเด่นมีความถี่เท่ากันให้พิจารณาจากค่าเฉลี่ยของ RES (คำนวณจากค่าเฉลี่ยของ RES จาก S01 ถึง S10) สำหรับขั้นตอนการประเมินและเลือกลักษณะเด่นดังกล่าวแสดงดังรูปที่ 3.31

ตารางที่ 3.12 ลักษณะเด่นที่ใช้ในการทดลอง

ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น
F1 RMS	F8 MN	F15 MAVS_4	F21 MDF
F2 VAR	F9 MAV	F16 MNH	F22 SM1
F3 WL	F10 MMAV1	F17 MNL	F23 SM2
F4 WAMP	F11 MMAV2	F18 PWL	F24 SM3
F5 ZC	F12 MAVS_1	F19 PWH	F25 SM5
F6 SSC	F13 MAVS_2	F20 ZCH	F26 MNF
F7 IAV	F14 MAVS_3		



**รูปที่ 3.31** แผนภาพแสดงขั้นตอนการประเมินและเลือกลักษณะเด่นโดยอ้างอิงข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อเนื้อของอาสาสมัครทั้ง 10 คน

4) ฝึกสอนระบบและทดสอบระบบโดยใช้ตัวจำแนก ANN เช่นเดียวกับการทดลองในหัวข้อ 3.3.1.1 โดยเลือกลักษณะเด่นที่ดีที่สุด 3 อันดับแรกเป็นอินพุตให้กับตัวจำแนก ANN โดยจะทดสอบ 2 รูปแบบเช่นเดียวกับรูปแบบวิธีการประเมิน กล่าวคือ รูปแบบแรก แต่ละช่องสัญญาณจะใช้ลักษณะเด่นที่ดีที่สุดสำหรับแต่ละช่องสัญญาณ รูปแบบที่สอง ทุกช่องสัญญาณจะใช้ลักษณะเด่นชุดเดียวกัน หลังจากนั้นทดลองปรับจำนวนลักษณะเด่นที่ป้อนให้กับตัวจำแนก ANN โดยเพิ่มจำนวนลักษณะเด่นเป็น 5 ตัว 7 ตัว และ 9 ตัว ตามลำดับ

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

#### 1) ผลการประเมินลักษณะเด่นโดยใช้ RES

ตารางที่ 3.13 และ ตารางที่ 3.14 แสดงผลการประเมินและเลือกลักษณะเด่นที่ได้จากการประเมินรูปแบบที่หนึ่ง ซึ่งจะพิจารณาเลือกลักษณะเด่นแยกกันในแต่ละช่องสัญญาณ จากตารางจะเป็นผลการประเมินของช่องสัญญาณที่ 1 สำหรับการพูดแบบออกเสียง จะเห็นได้ว่ามีจำนวนลักษณะเด่น 11 ชนิดที่ปรากฏอยู่ในกลุ่มของลักษณะเด่นที่มีผลการประเมินด้วย RES ที่ดีที่สุด 5 อันดับแรกของสัญญาณไฟฟ้ากล้ามเนื้อเนื้อของอาสาสมัครแต่ละคน หากความถี่ของลักษณะเด่นดังกล่าวและเรียงลำดับด้วยความถี่ของการปรากฏของลักษณะเด่นนั้นๆ ได้ผลดัง ตารางที่ 3.14 เมื่อประเมินและเลือกลักษณะเด่นครบทุกช่องสัญญาณ ทั้งการพูดแบบออกเสียงและไม่ออกเสียง สามารถสรุปผลการประเมินดังตารางที่ 3.16

สำหรับการเลือกลักษณะเด่นที่ได้จากการประเมินรูปแบบที่สอง ซึ่งประเมินลักษณะเด่นโดยพิจารณาข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อรวมกันทั้ง 5 ช่องสัญญาณ ในรูปแบบนี้ทุกช่องสัญญาณจะใช้ลักษณะเด่นชนิดเดียวกัน ผลการประเมินและเลือกลักษณะเด่นแสดงดังตารางที่ 3.15

**ตารางที่ 3.13** ลักษณะเด่นที่มีผลการประเมินดีที่สุด 5 อันดับแรก(การประเมินรูปแบบที่หนึ่ง) สำหรับสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน จากสัญญาณช่องสัญญาณที่ 1 สำหรับการพูดแบบออกเสียง

S01		S02		S03		S04		S05	
ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES
MMAV2	1.87	MNF	2.06	RMS	2.40	MMAV2	1.81	WAMP	2.30
WL	1.81	WL	1.59	VAR	2.29	MMAV1	1.67	MMAV1	2.30
MMAV1	1.78	WAMP	1.54	MMAV1	2.28	MAV	1.48	MAV	2.29
SSC	1.76	ZC	1.53	MMAV2	2.23	SSC	1.46	RMS	2.25
RMS	1.75	SSC	1.49	MAV	2.21	RMS	1.38	MMAV2	2.16
S06		S07		S08		S09		S10	
ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES
WAMP	2.49	WL	2.79	WAMP	2.98	ZC	1.74	ZC	1.52
ZC	2.46	WAMP	2.65	RMS	2.95	WAMP	1.72	RMS	1.49
WL	2.15	ZC	2.60	ZC	2.94	SSC	1.65	MAV	1.46
IAV	2.15	RMS	2.58	WL	2.88	IAV	1.48	MMAV1	1.40
SSC	2.04	MMAV2	2.57	MMAV1	2.87	WL	1.45	VAR	1.35

ตารางที่ 3.14 การเรียงลำดับของลักษณะเด่นจากตารางที่ 3.13 โดยพิจารณาจากความถี่และค่าเฉลี่ย RES

อันดับ	ลักษณะเด่น	ความถี่	ค่าเฉลี่ย RES
1	RMS	7	1.91
2	WAMP	6	1.97
3	ZC	6	1.95
4	MMAV1	6	1.92
5	WL	6	1.89
6	MMAV2	5	1.83
7	SSC	5	1.75
8	MAV	4	1.87
9	IAV	2	1.8
10	VAR	2	1.69
11	MNF	1	1.07

ตารางที่ 3.15 ผลการประเมินและเลือกลักษณะเด่น จากการประเมินรูปแบบที่สอง

ช่องสัญญาณ	อันดับ	ลักษณะเด่นที่ถูกเลือก 9 อันดับแรกสำหรับทุกช่องสัญญาณ (จากการประเมินรูปแบบที่สอง)	
		ออกเสียง	ไม่ออกเสียง
1-5	1	VAR	IAV
	2	SM2	MMAV1
	3	MMAV1	RMS
	4	WL	WL
	5	MAV	MMAV2
	6	MMAV2	SM1
	7	WAMP	ZC
	8	ZC	WAMP
	9	SSC	SSC

ตารางที่ 3.16 ผลการประเมินและเลือกลักษณะเด่น จากการประเมินรูปแบบที่หนึ่ง

วิธีการพูด	อันดับ	ลักษณะเด่นที่ถูกเลือก 9 อันดับแรกสำหรับแต่ละช่องสัญญาณ (จากการประเมินรูปแบบที่หนึ่ง)				
		Ch.1	Ch.2	Ch.3	Ch.4	Ch.5
ออกเสียง	1	RMS	MMAV1	RMS	WL	MMAV2
	2	WAMP	RMS	MAV	MMAV2	IAV
	3	ZC	MAV	MMAV2	SSC	WAMP
	4	MMAV1	WAMP	MNF	ZC	RMS
	5	WL	ZC	ZC	MMAV1	WL
	6	MMAV2	MMAV2	MMAV1	IAV	ZC
	7	SSC	SSC	WL	WAMP	MMAV1
	8	MAV	SM1	WAMP	RMS	MAVS_4
	9	MNF	WL	SSC	SM5	SSC
ไม่ออกเสียง	1	IAV	WAMP	MNF	MMAV2	RMS
	2	MMAV1	MMAV1	IAV	RMS	MMAV2
	3	RMS	RMS	MMAV1	WL	MAV
	4	WL	ZC	ZC	MMAV1	MMAV1
	5	ZC	MAV	WAMP	IAV	ZC
	6	MMAV2	WL	MDF	MNF	WL
	7	WAMP	MMAV2	RMS	WAMP	WAMP
	8	SSC	SSC	WL	SSC	SM1
	9	MNF	MNF	SSC	MAVS_3	MNF



## 2) ผลการจำแนกโดยใช้ตัวจำแนก ANN

รูปที่ 3.32 แสดงผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อปรับจำนวนลักษณะเด่นซึ่งเป็นอินพุตของตัวจำแนก ANN เป็นจำนวน 3 5 7 และ 9 ตามลำดับ โดยแสดงผลความแม่นยำในการจำแนกสำหรับข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 10 คน (S01-S10) กราฟด้านซ้ายมือ (ก) และ (ค) เป็นผลความแม่นยำในการจำแนกเมื่อใช้ลักษณะเด่นที่ถูกประเมินจากการประเมินรูปแบบที่หนึ่ง (ลักษณะเด่นในแต่ละช่องสัญญาณแตกต่างกัน) โดยที่ (ก) เป็นการพูดแบบออกเสียง และ (ค) เป็นการพูดแบบไม่ออกเสียง สำหรับกราฟด้านขวามือ (ข) และ (ง) เป็นผลความแม่นยำในการจำแนกเมื่อใช้ลักษณะเด่นที่ถูกประเมินจากการประเมินรูปแบบที่สอง (ลักษณะเด่นเหมือนกันทุกช่องสัญญาณ) โดยที่ (ข) เป็นการพูดแบบออกเสียง และ (ง) เป็นการพูดแบบไม่ออกเสียง จากกราฟจะเห็นได้ว่าผลการจำแนกของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้ง 10 คน แตกต่างกันอย่างเห็นได้ชัด โดยที่ S08 ให้ผลความแม่นยำดีที่สุดสำหรับการพูดแบบออกเสียง (ทั้งสองรูปแบบการประเมิน) ซึ่งให้ผลความแม่นยำสูงถึง 100% และ S01 ให้ผลความแม่นยำดีที่สุดสำหรับการพูดแบบไม่ออกเสียง (ทั้งสองรูปแบบการประเมิน) ซึ่งให้ผลความแม่นยำสูงสุด 98.18% (รูปแบบที่สอง, ลักษณะเด่น 9 ตัว) สำหรับสัญญาณที่ให้ความแม่นยำต่ำสุด คือ S06 กับ S10 ซึ่งให้ผลเหมือนกันทั้งการพูดแบบออกเสียงและไม่ออกเสียง (ทั้งสองรูปแบบการประเมิน) โดยความแม่นยำต่ำสุด คือ 59.09% (รูปแบบที่หนึ่ง, ลักษณะเด่น 3 ตัว, ไม่ออกเสียง)

การพิจารณาความแม่นยำในการจำแนกคำพูด 11 คำ โดยพิจารณาจากความแม่นยำของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคนดังรูปที่ 3.32 ยังไม่สามารถระบุได้ชัดเจนว่าการเพิ่มจำนวนลักษณะเด่นมีผลต่อความแม่นยำในการจำแนกอย่างไร ดังนั้นเพื่อให้สามารถวิเคราะห์จำนวนลักษณะเด่นที่เหมาะสมและระบุอัตราการเปลี่ยนแปลงของความแม่นยำขณะที่มีการปรับจำนวนลักษณะเด่นได้อย่างชัดเจน จึงนำผลความแม่นยำในการจำแนกคำพูดทั้ง 11 คำ ของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด (S01-S10) มาคำนวณหาค่าเฉลี่ย

รูปที่ 3.33 แสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ ซึ่งคำนวณจากข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ซึ่งแสดงในรูปของแผนภูมิแท่ง (แกนปฐมภูมิ) โดยที่กราฟเส้นตรงแสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (แกนทุติยภูมิ) กราฟด้านซ้ายมือ (ก) และ (ค) เป็นผลการทดลองเมื่อใช้ลักษณะเด่นที่ถูกประเมินจากการประเมินรูปแบบที่หนึ่ง (ลักษณะเด่นในแต่ละช่องสัญญาณแตกต่างกัน) โดยที่ (ก) เป็นการพูดแบบออก

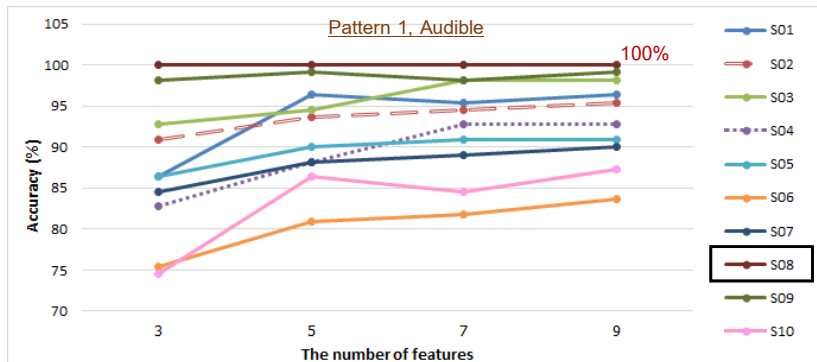
เสียง และ (ค) เป็นการพูดแบบไม่ออกเสียง สำหรับกราฟด้านขวามือ (ข) และ (ง) เป็นผลการทดลองเมื่อใช้ลักษณะเด่นที่ถูกระบุจากการประเมินรูปแบบที่สอง (ลักษณะเด่นเหมือนกันทุกช่องสัญญาณ) โดยที่ (ข) เป็นการพูดแบบออกเสียง และ (ง) เป็นการพูดแบบไม่ออกเสียง จากกราฟจะเห็นได้ว่า เมื่อเพิ่มจำนวนลักษณะเด่นเพื่อเป็นอินพุตให้กับตัวจำแนก ANN ความแม่นยำเฉลี่ยในการจำแนกจะเพิ่มขึ้นด้วย ซึ่งให้ผลไปในทิศทางเดียวกันสำหรับทั้งสองรูปแบบการประเมินและสองรูปแบบวิธีการพูด เมื่อพิจารณากราฟโดยส่วนใหญ่แล้วพบว่า ความแม่นยำเฉลี่ยจะเพิ่มขึ้นมากที่สุดเมื่อจำนวนลักษณะเด่นถูกเพิ่มจาก 3 ไป 5 หลังจากนั้นความแม่นยำเฉลี่ยจะเพิ่มขึ้นเล็กน้อย โดยที่การใช้ลักษณะเด่นที่ได้จากการประเมินรูปแบบที่สองสามารถเพิ่มความแม่นยำได้มากกว่ารูปแบบที่หนึ่ง จากผลการทดลองนี้สามารถสรุปได้ว่า การใช้จำนวนลักษณะเด่น 9 ตัวสามารถเพิ่มความแม่นยำเฉลี่ยได้มากที่สุด

รูปที่ 3.34 แสดงการเปรียบเทียบผลความแม่นยำเฉลี่ยระหว่างการพูดแบบออกเสียงและการพูดแบบไม่ออกเสียง สำหรับการใช้ลักษณะเด่นที่ได้จากการประเมินทั้งสองรูปแบบ โดยที่ (ก) คือ รูปแบบที่หนึ่ง และ (ข) คือ รูปแบบที่สอง จากกราฟจะเห็นได้ว่าผลความแม่นยำเฉลี่ยของการพูดแบบออกเสียงให้ผลดีกว่าการพูดแบบไม่ออกเสียงอย่างเห็นได้ชัด

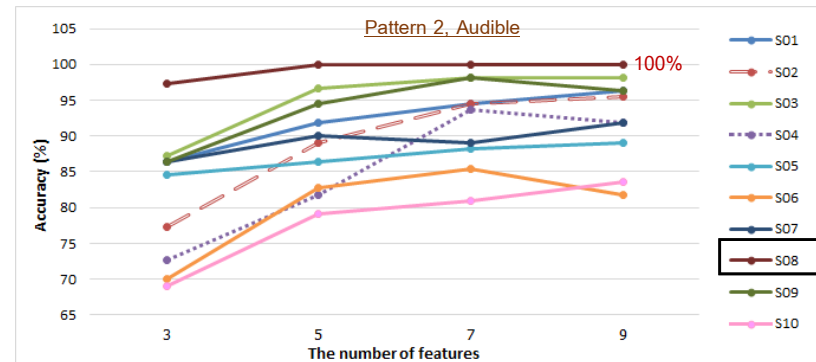
รูปที่ 3.35 แสดงการเปรียบเทียบผลความแม่นยำเฉลี่ยระหว่างการใช้ลักษณะเด่นที่ได้จากการประเมินรูปแบบที่หนึ่งและรูปแบบที่สอง โดยที่ (ก) คือ การพูดแบบออกเสียง และ (ข) คือ การพูดแบบไม่ออกเสียง เมื่อพิจารณา ณ ตำแหน่งการใช้จำนวนลักษณะเด่น 3 ตัว พบว่าผลความแม่นยำเฉลี่ยทั้งสองรูปแบบแตกต่างกันอย่างเห็นได้ชัด โดยที่รูปแบบที่หนึ่งซึ่งใช้ลักษณะเด่นแตกต่างกันในแต่ละช่องสัญญาณให้ผลความแม่นยำเฉลี่ยสูงกว่า (5.45% สำหรับการพูดแบบออกเสียง และ 3.09% สำหรับการพูดแบบไม่ออกเสียง) อย่างไรก็ตามเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น พบว่าความแม่นยำเฉลี่ยทั้งสองรูปแบบแตกต่างกันน้อยมาก ดังนั้นผลการทดลองในส่วนนี้สรุปได้ว่าเมื่อจำนวนลักษณะเด่นที่ป้อนให้กับตัวจำแนกเพิ่มมากขึ้น การเลือกใช้รูปแบบการประเมินทั้งสองรูปแบบให้ผลไม่แตกต่างกัน ดังนั้นควรเลือกใช้รูปแบบที่สองในการประเมินเนื่องจากใช้เวลาในการประมวลผลน้อยกว่าและความซับซ้อนน้อยกว่า (ประเมินลักษณะเด่นโดยพิจารณาข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อรวมกันทั้ง 5 ช่องสัญญาณ ในรูปแบบนี้ทุกช่องสัญญาณจะได้ลักษณะเด่นชนิดเดียวกัน)

สำหรับความแม่นยำเฉลี่ยสูงสุดที่ได้จากการทดลองนี้ ได้จากการใช้จำนวนลักษณะเด่น 9 ตัว จากการใช้ลักษณะเด่นที่ได้จากการประเมินรูปแบบที่หนึ่ง คือ 93.36% และ 81.64% (สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ) อย่างไรก็ตามการใช้ลักษณะ

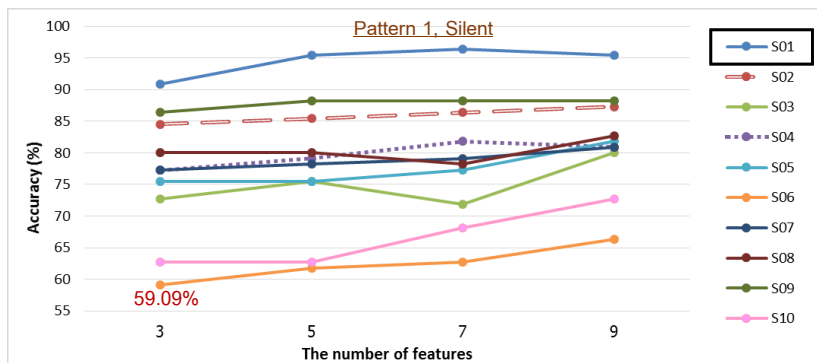
เด่นที่ได้จากการประเมินรูปแบบที่สองได้รับความแม่นยำแตกต่างกันเพียงเล็กน้อย คือ 92.45% และ 81.45% (สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ)



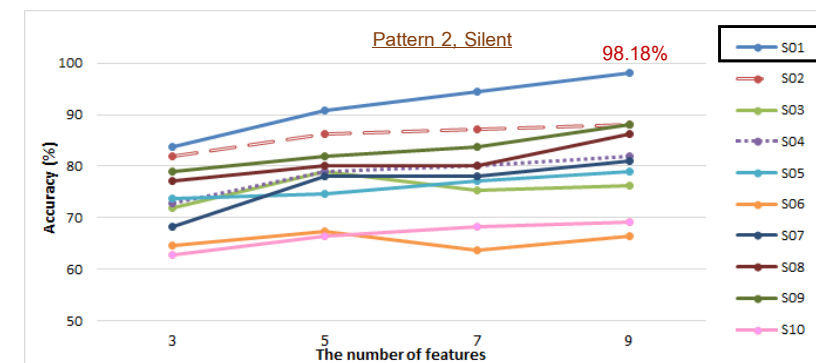
(ก)



(ข)

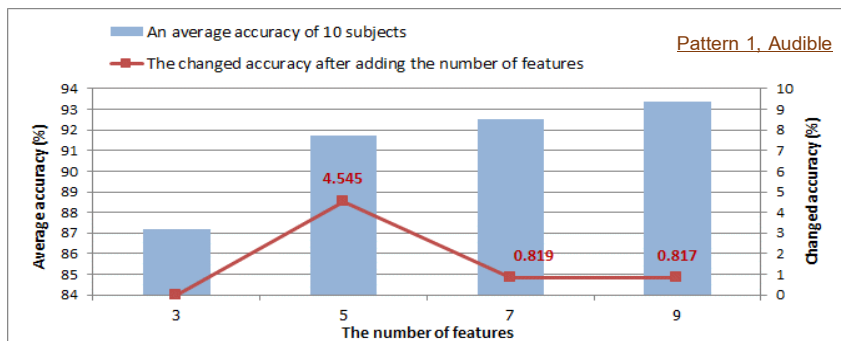


(ค)

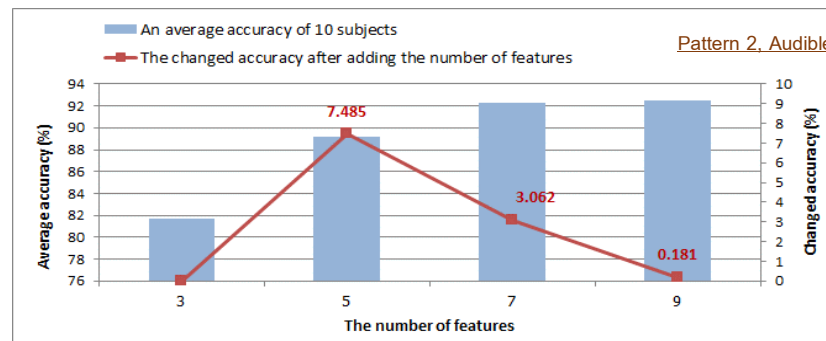


(ง)

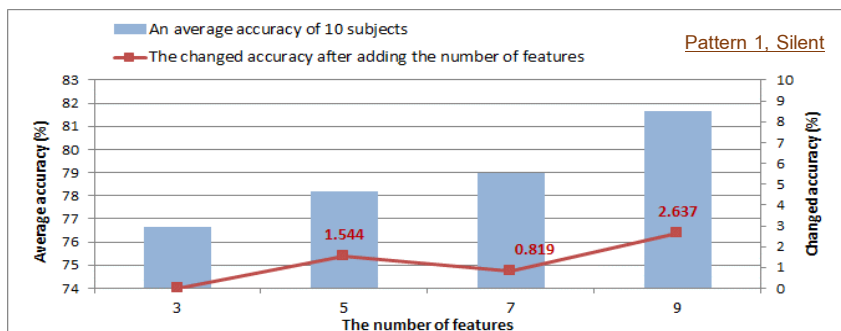
รูปที่ 3.32 กราฟแสดงผลความแม่นยำในการจำแนกคำพูด 11 คำโดยใช้ตัวจำแนก ANN เมื่อปรับจำนวนลักษณะเด่น 3 5 7 9 ตามลำดับ ซ้าย: รูปแบบที่หนึ่ง ขวา: รูปแบบที่สอง บน: ออกเสียง ล่าง: ไม่ออกเสียง



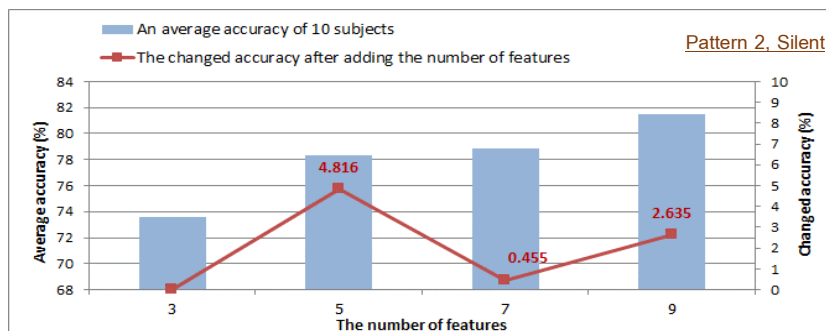
(ก)



(ข)

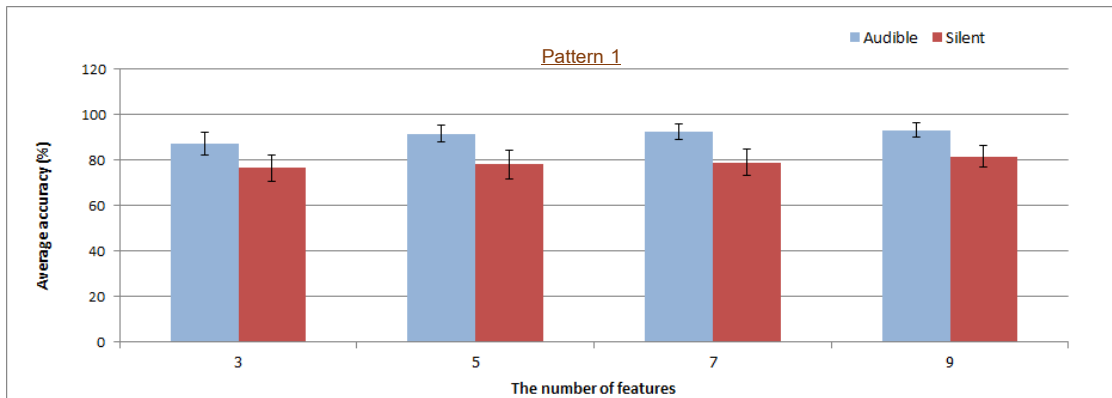


(ค)

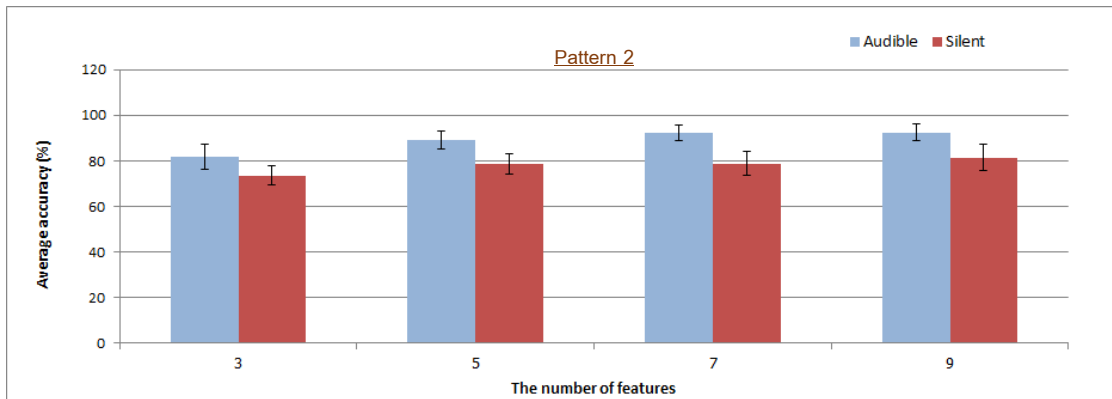


(ง)

รูปที่ 3.33 แผนภูมิแท่งแสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ (แกนปฐมภูมิ) และ กราฟเส้นตรงแสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (แกนทุติยภูมิ) ซ้าย: รูปแบบที่หนึ่ง ขวา: รูปแบบที่สอง บน: ออกเสียง ล่าง: ไม่ออกเสียง

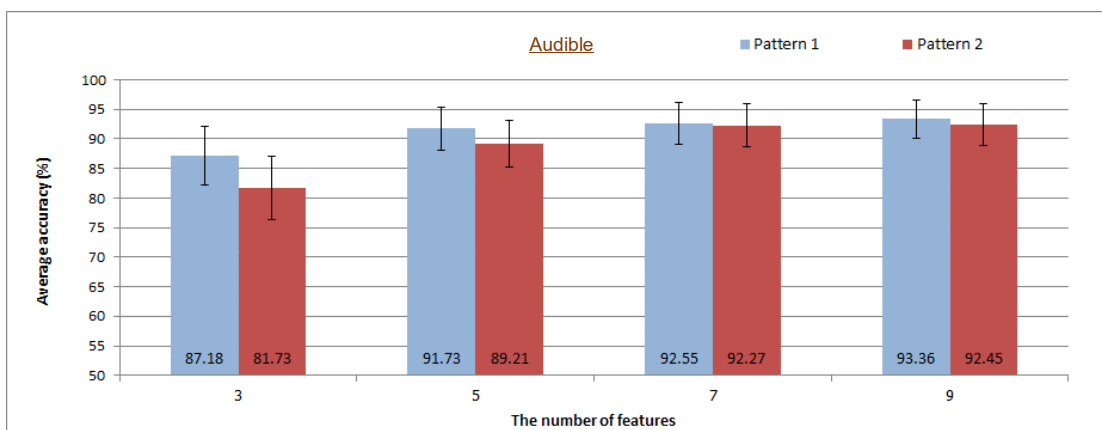


(ก)

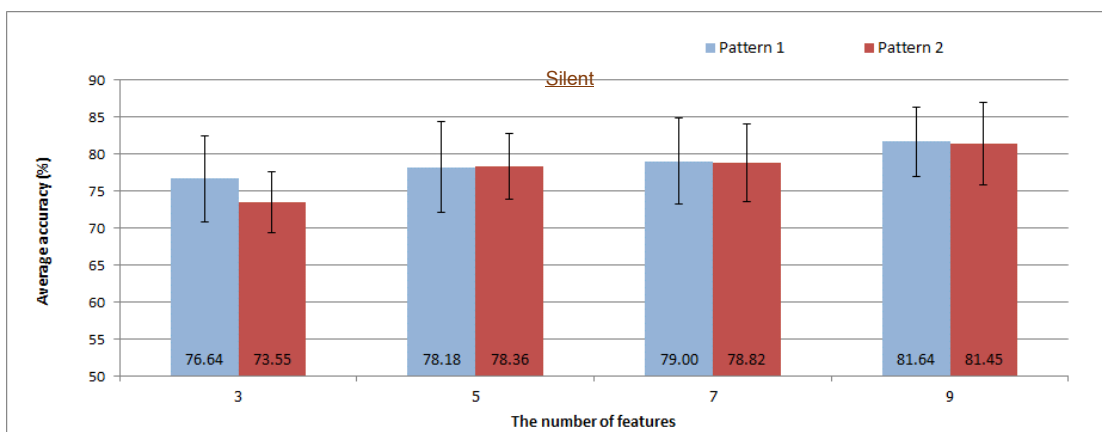


(ข)

**รูปที่ 3.34** แผนภูมิแท่งแสดงผลการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำ เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ระหว่างวิธีการพูดแบบออกเสียง และไม่ออกเสียง ก) รูปแบบที่หนึ่ง ข) รูปแบบที่สอง หมายถึง: แถบค่าคลาดเคลื่อน (error bar) แสดงช่วงความเชื่อมั่น (confidence interval) 95%



(ก)



(ข)

รูปที่ 3.35 แผนภูมิแท่งแสดงผลการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูด 11 คำคน เมื่อมีการเพิ่มจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ระหว่างการใช้ลักษณะเด่นจากการประเมินรูปแบบที่หนึ่ง (pattern 1) และรูปแบบที่สอง (pattern 2) ก) ออกเสียง ข) ไม่ออกเสียง  
หมายเหตุ: แถบค่าคลาดเคลื่อน (error bar) แสดงช่วงความเชื่อมั่น (confidence interval) 95%

### (ง) สรุปผลการทดลอง

การทดลองนี้เป็นการทดลองเพื่อออกแบบระบบการจำแนกสัญญาณไฟฟ้า กล้ามเนื้อของการพูดตัวเลขภาษาไทย 11 คำ โดยอ้างอิงข้อมูลสัญญาณไฟฟ้างแล้ามเนื้อจากอาสาสมัครทั้ง 10 คน เพื่อให้เทคนิคที่ออกแบบสามารถนำไปใช้ได้กับข้อมูลของสัญญาณไฟฟ้างแล้ามเนื้อทั้งหมด

นำสัญญาณไฟฟ้างแล้ามเนื้อจากอาสาสมัครทั้ง 10 คน เพื่อมาคำนวณลักษณะเด่นของสัญญาณไฟฟ้างแล้ามเนื้อ 26 ชนิด และประเมินความสามารถในการแยกแยะกลุ่มออกเป็น 11 กลุ่ม โดยใช้ตัวประเมิน RES ซึ่งประเมิน 2 รูปแบบ คือ การประเมินลักษณะเด่นเมื่อพิจารณาการใช้ข้อมูลของสัญญาณไฟฟ้างแล้ามเนื้อทั้ง 5 ช่องสัญญาณ (รูปแบบที่หนึ่ง) และ การประเมินแต่ละช่องสัญญาณแยกกัน (รูปแบบที่สอง) เลือกลักษณะเด่นที่ดีที่สุด 5 อันดับแรกของสัญญาณไฟฟ้างแล้ามเนื้อของอาสาสมัครแต่ละคนจากวิธีการประเมินข้างต้น นำลักษณะเด่นที่ถูกเลือกทั้งหมดในกลุ่มนี้มาพิจารณาหาความถี่ หลังจากนั้นเรียงลำดับลักษณะเด่นดังกล่าวโดยพิจารณาจากค่าความถี่ ซึ่งถ้าลักษณะเด่นมีความถี่เท่ากันให้พิจารณาจากค่าเฉลี่ยของ RES นำลักษณะเด่นดังกล่าวเพื่อใช้เป็นอินพุตให้กับตัวจำแนก โดยปรับจำนวนลักษณะเด่น 3 5 7 และ 9 ตามลำดับ ฝึกสอนระบบและทดสอบระบบโดยใช้ตัวจำแนก ANN โดยจะทดสอบ 2 รูปแบบ เช่นเดียวกับรูปแบบวิธีการประเมิน กล่าวคือ รูปแบบแรก แต่ละช่องสัญญาณจะใช้ลักษณะเด่นที่ดีที่สุดสำหรับแต่ละช่องสัญญาณ รูปแบบที่สอง ทุกช่องสัญญาณจะใช้ลักษณะเด่นชุดเดียวกัน

จากการทดลองพบว่า การเพิ่มจำนวนลักษณะเด่นสามารถเพิ่มความแม่นยำในการจำแนก โดยที่การเพิ่มลักษณะเด่น 9 ตัวให้ความแม่นยำสูงสุด โดยที่ความแม่นยำเฉลี่ยสูงสุดได้จากการใช้ลักษณะเด่นที่ได้จากการประเมินรูปแบบที่หนึ่ง (93.36% และ 81.64% สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ) เมื่อเปรียบเทียบความแม่นยำในการจำแนกสำหรับการใช้ลักษณะเด่นจากวิธีการประเมินทั้งสองรูปแบบ พบว่ารูปแบบที่หนึ่งซึ่งใช้ลักษณะเด่นแตกต่างกันสำหรับแต่ละช่องสัญญาณ ให้ความแม่นยำดีกว่ารูปแบบที่สอง เมื่อใช้จำนวนลักษณะเด่น 3 ตัว อย่างไรก็ตามเมื่อใช้จำนวนลักษณะเด่นเพิ่มมากขึ้นความแม่นยำในการจำแนกระหว่างสองรูปแบบแตกต่างกันน้อยมาก ดังนั้นการเลือกใช้รูปแบบที่สองในการประเมินเหมาะสมมากกว่ารูปแบบแรกเมื่อจำนวนลักษณะเด่นที่ใช้มีจำนวนมาก เนื่องจากใช้เวลาในการประมวลผลน้อยกว่าเมื่อเปรียบเทียบความแม่นยำในการจำแนกสำหรับข้อมูลสัญญาณไฟฟ้างแล้ามเนื้อของอาสาสมัครแต่



ละคนพบว่า S08 ให้ผลความแม่นยำดีที่สุดสำหรับการพูดแบบออกเสียง และ S01 ให้ผลความแม่นยำดีที่สุดสำหรับการพูดแบบไม่ออกเสียง

ในการทดลองนี้เป็นการใช้ตัวประเมินลักษณะเด่นและตัวจำแนกเพียงชนิดเดียว ถึงแม้ว่าจะให้ผลความแม่นยำในการจำแนกที่ดี อย่างไรก็ตามเพื่อให้ได้เทคนิคที่ดีที่สุดและเหมาะสมที่สุดสำหรับงานวิจัยนี้ จึงต้องมีการทดลองเพื่อเปรียบเทียบประสิทธิภาพของตัวจำแนกและตัวประเมินลักษณะเด่นหลากหลายชนิด นอกจากนี้ยังต้องมีการปรับปรุงในส่วนของกระบวนการเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสม ซึ่งในการทดลองนี้เป็นการเลือกลักษณะเด่นโดยการจัดอันดับลักษณะเด่นตามผลการประเมินด้วย RES ซึ่งการใช้วิธีนี้ทำให้ลักษณะเด่นที่ถูกเลือกอาจมีความสัมพันธ์กันทำให้ข้อมูลซ้ำซ้อน ซึ่งไม่ได้ทำให้ความแม่นยำเพิ่มขึ้นแต่ยังทำให้เวลาในการประมวลผลช้าลง ดังนั้นในบทต่อไปจะได้กล่าวถึงส่วนของการทดลองเพื่อปรับปรุงระบบ

### 3.3.2 การออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง)

3.3.2.1 การทดลองเบื้องต้นเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้ RES เป็นตัวประเมินลักษณะเด่น และเปรียบเทียบผลการประเมินที่ได้กับผลการจำแนกของตัวจำแนก ANN

#### (ก) วัตถุประสงค์

- เพื่อคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อที่เหมาะสมสำหรับการจำแนกวรรณยุกต์ไทยทั้ง 5 เสียง
- เพื่อเปรียบเทียบผลการประเมินลักษณะเด่นโดยใช้เกณฑ์ประเมิน RES และประสิทธิภาพในการจำแนกของตัวจำแนก ANN
- เพื่อออกแบบเทคนิคการจำแนกวรรณยุกต์

#### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครเพียง 1 คน (S01) โดยใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ซึ่งเป็นคำพูดของการออกเสียงวรรณยุกต์ ซึ่งการทดลองนี้จะใช้คำพูดชุดที่ 1 ถึง 21 ดังตารางที่ 3.3 เป็นคำพูดโดด 105 คำซึ่งครอบคลุมการออก

เสียงวรรณยุกต์ทั้ง 5 เสียง และครอบคลุมหน่วยเสียงพยัญชนะทั้งหมด ในแต่ละคำจะถูกพูดซ้ำ 3 ครั้ง (945 ข้อมูล) คำพูดชุดที่ 1 2 8 9 และ 15 ซึ่งมีการผันวรรณยุกต์ทั้ง 5 เสียงถูกใช้สำหรับการทดสอบระบบ (225 ข้อมูล) สำหรับข้อมูลชุดที่เหลือจะถูกใช้สำหรับการฝึกสอนระบบ (720 ข้อมูล)

### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อมาคำนวณลักษณะเด่น ซึ่งในการทดลองนี้จะแบ่งวิธีการคำนวณลักษณะเด่นออกเป็นสองรูปแบบ ดังนี้

รูปแบบที่หนึ่ง คำนวณโดยใช้ข้อมูลของสัญญาณทั้งหมด (ตั้งแต่ช่วงเริ่มต้นและสิ้นสุดของคำพูดแต่ละเสียง) แต่ละสัญญาณจะมีความยาวเท่ากับ  $N$  ซึ่งเป็นวิธีเดิมที่ใช้ในการคำนวณลักษณะเด่นจากการทดลองที่ผ่านมา ลักษณะเด่นทั้งหมดในรูปแบบที่หนึ่งถูกแสดงดังตารางที่ 3.17

รูปแบบที่สอง คำนวณโดยแบ่งสัญญาณไฟฟ้ากล่อมเนื้อออกเป็น 6 ส่วนเท่าๆ กัน ซึ่งแต่ละส่วนจะมีความยาวของสัญญาณเท่ากับ  $N/6$  แล้วนำสัญญาณในแต่ละส่วนมาคำนวณลักษณะเด่น ซึ่งจะใช้ลักษณะเด่น 8 ชนิด คือ RMS WL WAMP SSC ZC MAV SM1 และ MAVS ในการคำนวณ ลักษณะเด่นในแต่ละส่วนจะถูกเรียกชื่อใหม่ โดยใช้ตัวเลขซึ่งอยู่ด้านหลังเครื่องหมายอัฒภาค (underscore) เป็นตัวบ่งชี้ส่วนของสัญญาณที่นำมาคำนวณ เช่น RMS\_1 คือ ลักษณะเด่นชนิด RMS ของสัญญาณส่วนที่หนึ่ง เป็นต้น จากการคำนวณลักษณะเด่นชนิด MAVS สำหรับทั้ง 6 ส่วนของสัญญาณ จะได้ลักษณะเด่นตัวใหม่ 5 ตัว สำหรับลักษณะเด่นชนิด RMS WL และ MAV จะได้ลักษณะเด่นชนิดละ 6 ตัว ลักษณะเด่นทั้งหมดที่ในรูปแบบที่สองถูกแสดงดังตารางที่ 3.18

ตารางที่ 3.17 ลักษณะเด่นรูปแบบที่หนึ่ง

ลักษณะเด่น		ลักษณะเด่น		ลักษณะเด่น	
F1	RMS	F7	IAV	F13	MNF
F2	VAR	F8	MN	F14	SM1
F3	WL	F9	MAV	F15	SM2
F4	WAMP	F10	MMAV1	F16	SM3
F5	ZC	F11	MMAV2	F17	SM5
F6	SSC	F12	MDF		

ตารางที่ 3.18 ลักษณะเด่นรูปแบบที่สอง

ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น
F18 RMS_1	F24 WL_1	F30 WAMP_1	F36 ZC_1
F19 RMS_2	F25 WL_2	F31 WAMP_1	F37 ZC_1
F20 RMS_3	F26 WL_3	F32 WAMP_1	F38 ZC_1
F21 RMS_4	F27 WL_4	F33 WAMP_1	F39 ZC_1
F22 RMS_5	F28 WL_5	F34 WAMP_1	F40 ZC_1
F23 RMS_6	F29 WL_6	F35 WAMP_1	F41 ZC_1
ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น	ลักษณะเด่น
F42 SSC_1	F48 MAV_1	F54 SM1_1	F60 MAVS_1
F43 SSC_1	F49 MAV_2	F55 SM1_1	F61 MAVS_2
F44 SSC_1	F50 MAV_3	F56 SM1_1	F62 MAVS_3
F45 SSC_1	F51 MAV_4	F57 SM1_1	F63 MAVS_4
F46 SSC_1	F52 MAV_5	F58 SM1_1	F64 MAVS_5
F47 SSC_1	F53 MAV_6	F59 SM1_1	

สาเหตุที่ต้องมีการคำนวณลักษณะเด่นในรูปแบบที่สอง เนื่องจากตำแหน่งของเสียงวรรณยุกต์ที่เกิดขึ้นในสัญญาณไฟฟ้ากล่อมเนื้อของคำพูดแต่ละคำไม่ได้เกิดตลอดช่วงสัญญาณ แต่จะเกิดในบางตำแหน่งของสัญญาณเท่านั้น เพราะในหนึ่งสัญญาณของคำพูดหนึ่งคำจะประกอบด้วยเสียงพยัญชนะ สระ และวรรณยุกต์ ดังนั้นจึงไม่สามารถระบุได้ว่าเสียงวรรณยุกต์เกิดขึ้น ณ ตำแหน่งใดของสัญญาณ ดังนั้นการคำนวณลักษณะเด่นวิธีนี้ จะใช้ในการตรวจสอบว่าเสียงวรรณยุกต์เกิดตรงตำแหน่งใดของสัญญาณ

จากการคำนวณลักษณะเด่นรูปแบบแรกจะได้ลักษณะเด่นทั้งหมด 17 ตัว (F1-F17) และรูปแบบที่สองจะได้ลักษณะเด่นทั้งหมด 47 ตัว (F18-F64) ดังนั้นในการทดลองนี้ทำการทดสอบลักษณะเด่นทั้งหมด 64 ตัว

2) ทำการประเมินลักษณะเด่นทั้งหมด โดยใช้ตัวประเมิน RES ซึ่งประเมินความสามารถในการจำแนกวรรณยุกต์ 5 เสียง และเลือกลักษณะเด่นที่มีผลการประเมินดีที่สุด 5 อันดับแรกเพื่อเป็นอินพุตให้กับตัวจำแนก ANN

3) ฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก ANN เพื่อวัดประสิทธิภาพในการจำแนกวรรณยุกต์ 5 เสียง

4) พล็อตแผนภาพการกระจายของลักษณะเด่นทั้ง 64 ตัว วิเคราะห์ความสามารถในการแยกแยะข้อมูลของวรรณยุกต์ทั้ง 5 เสียง และออกแบบเทคนิคการจำแนกวรรณยุกต์จากการวิเคราะห์แผนภาพการกระจาย

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

1) ผลการประเมินลักษณะเด่นด้วย RES และผลความแม่นยำในการจำแนกโดยใช้ตัวจำแนก ANN

รูปที่ 3.36 แสดงผลการประเมินความสามารถในการแยกแยะกลุ่มข้อมูลวรรณยุกต์ทั้ง 5 เสียง โดยใช้ตัวประเมิน RES โดยเปรียบเทียบผลการประเมินของลักษณะเด่นทั้ง 64 ตัว จากผลการประเมินพบว่าลักษณะเด่น MMAV2 มีค่า RES สูงสุด ซึ่งแสดงให้เห็นว่าการประเมินที่ดีที่สุด และลักษณะเด่น SSC\_2 มีค่า RES ต่ำสุด เมื่อพิจารณาลักษณะเด่นที่มีผลการประเมินที่ดีที่สุด 10 อันดับแรก พบว่าส่วนใหญ่เป็นลักษณะเด่นที่ได้จากการคำนวณรูปแบบที่สอง ซึ่งมีการแบ่งสัญญาณออกเป็นส่วนๆ แล้วคำนวณลักษณะเด่นในแต่ละส่วนของสัญญาณ ซึ่งแสดงให้เห็นว่าการคำนวณลักษณะเด่นในรูปแบบที่สองน่าจะเหมาะสมในการจำแนกวรรณยุกต์มากกว่าลักษณะเด่นที่คำนวณในรูปแบบที่หนึ่งซึ่งใช้ทุกส่วนของสัญญาณในการคำนวณ โดยที่พบว่าลักษณะเด่นที่คำนวณจากส่วนที่ 3 และส่วนที่ 4 ของสัญญาณให้ผลการประเมินดีที่สุด นั่นก็แสดงว่าตำแหน่งการเกิดเสียงวรรณยุกต์น่าจะอยู่ตรงส่วนดังกล่าว ในทางกลับกัน พิจารณาลักษณะเด่น 10 อันดับสุดท้ายที่ให้ผลการประเมินต่ำสุด จะเห็นส่วนว่าส่วนใหญ่เป็นลักษณะเด่นที่ได้มาจากการคำนวณจากส่วนที่ 1 และส่วนที่ 2 ของสัญญาณ นั่นก็แสดงว่าตำแหน่งดังกล่าวไม่มีผลต่อการจำแนกวรรณยุกต์

รูปที่ 3.37 แสดงแผนภาพการกระจายของลักษณะเด่นที่มีผลการประเมินดีที่สุด MMAV2 (ซ้าย) และผลการประเมินต่ำที่สุด SSC\_2 (ขวา) จากรูปจะเห็นได้ชัดเจนว่า MMAV2 สามารถแยกแยะกลุ่มข้อมูลของวรรณยุกต์ทั้ง 5 เสียงได้ ถึงแม้ว่าจะไม่สามารถแยกได้ชัดเจน ส่วน SSC\_2 ไม่สามารถแยกแยะกลุ่มข้อมูลของวรรณยุกต์ทั้ง 5 เสียงได้เลย

ตารางที่ 3.19 แสดงผลการประเมินและผลการจำแนกของลักษณะเด่นที่ดีที่สุด 5 อันดับแรก จะเห็นได้ว่าผลการประเมินด้วย RES เป็นไปในทิศทางเดียวกันกับความแม่นยำในการ

จำแนกของตัวจำแนก ANN ซึ่งลักษณะเด่นที่มีความแม่นยำสูงสุด คือ MMAV2 โดยให้ความแม่นยำ 56.2% ซึ่งความแม่นยำที่ได้ยังต่ำอยู่มาก ซึ่งจำเป็นจะต้องหาเทคนิคเพื่อปรับปรุงความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง

## 2) ผลการวิเคราะห์แผนภาพการกระจายของลักษณะเด่นทั้ง 64 ตัว

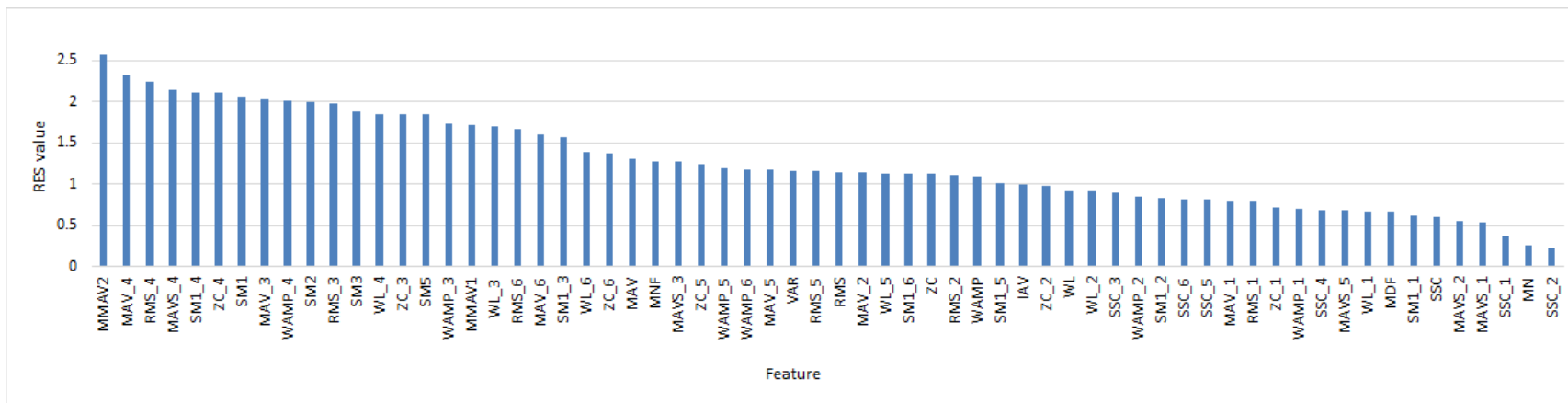
จากการพิจารณาความสามารถในการแยกแยะเสียงวรรณยุกต์โดยวิธีการสังเกตแผนภาพการกระจายของลักษณะเด่นทั้ง 64 ตัว พบว่าลักษณะเด่นที่สามารถแยกแยะกลุ่มของวรรณยุกต์ได้ ส่วนใหญ่จะสามารถแยกแยะวรรณยุกต์ได้ค่อนข้างชัดเจนเพียง 3 กลุ่ม คือ วรรณยุกต์เอก (low tone) วรรณยุกต์จัตวา (rising tone) และ กลุ่มของวรรณยุกต์อีกสามเสียงที่เหลือซึ่งไม่สามารถแยกได้อย่างชัดเจน แสดงดังรูปที่ 3.38 (ก) เมื่อพิจารณาแผนภาพการกระจายเฉพาะวรรณยุกต์สามเสียงที่เหลือโดยไม่สนใจวรรณยุกต์เอก และจัตวา (เอาข้อมูลวรรณยุกต์ดังกล่าวออกจากแผนภาพการกระจาย) จะเห็นได้ว่าสามารถแยกเสียงวรรณยุกต์ออกเป็น 2 กลุ่ม คือ วรรณยุกต์ตรี (high tone) และกลุ่มของวรรณยุกต์สองเสียงที่เหลือซึ่งไม่สามารถแยกได้อย่างชัดเจน แสดงดังรูปที่ 3.38 (ข) โดยที่ลักษณะเด่นบางชนิดเท่านั้นที่สามารถแยกเสียงวรรณยุกต์ในกลุ่มนี้ได้ เช่น ลักษณะเด่น MAV\_5 เป็นต้น และเมื่อพิจารณาแผนภาพการกระจายเฉพาะวรรณยุกต์สองเสียงที่เหลือโดยไม่สนใจวรรณยุกต์อื่นๆ จะเห็นได้ว่าสามารถแยกเสียงวรรณยุกต์สามัญ (mid tone) และโท (falling tone) ออกจากกันได้ แสดงดังรูปที่ 3.38 (ค) โดยที่ลักษณะเด่นบางชนิดเท่านั้นที่สามารถแยกเสียงวรรณยุกต์ในกลุ่มนี้ได้ เช่น ลักษณะเด่น SSC\_4 เป็นต้น

**ตารางที่ 3.19** ผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง ของลักษณะเด่นที่ดีที่สุด 5 อันดับแรกจากผลการประเมินด้วย RES

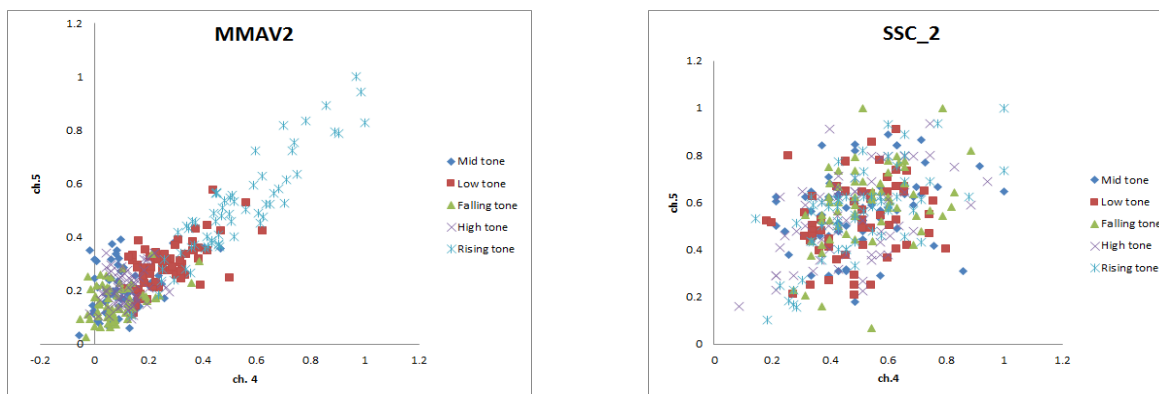
อันดับ	ลักษณะเด่น	RES	Accuracy rate (%)
1	MMAV2	2.57	56.2
2	MAV_4	2.33	55.2
3	RMS_4	2.25	55.2
4	MAVS_5	2.14	54.3
5	SM1_4	2.12	51.4

จากผลการทดลองข้างต้น สามารถแบ่งลักษณะเด่นได้เป็น 3 กลุ่ม คือ กลุ่มแรก เป็นกลุ่มของลักษณะเด่นที่มีความสามารถในการแยกแยะเสียงเอกและเสียงจัตวาออกจากเสียงวรรณยุกต์อื่นได้ดี โดยที่วรรณยุกต์ที่เหลือถูกจัดเป็นกลุ่มเดียวกัน (สามัญ โท และตรี) กลุ่มที่สอง เป็นกลุ่มของลักษณะเด่นที่มีความสามารถในการแยกเสียงตรีออกจากอีกสองเสียงที่เหลือ (สามัญ และโท) และกลุ่มที่สามเป็นลักษณะเด่นที่มีความสามารถในการแยกแยะเสียงสามัญและโทออกจากกัน จากผลการทดลองนี้จึงได้ออกแบบเทคนิคการจำแนกวรรณยุกต์โดยการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย

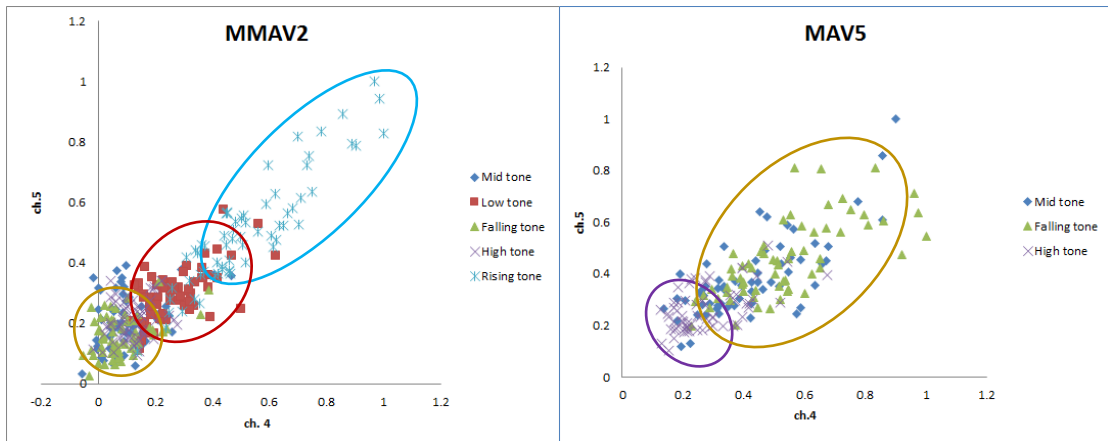
รูปที่ 3.39 แสดงรูปแบบการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย ในแต่ละโครงข่ายจะใช้ลักษณะเด่นที่แตกต่างกัน โดยโครงข่ายที่หนึ่งจะใช้ในการจำแนกวรรณยุกต์เอกและจัตวา ในกรณีที่ผลการจำแนกออกมาเป็นวรรณยุกต์อื่นๆ ระบบจะส่งต่อไปที่โครงข่ายที่สองซึ่งใช้ในการจำแนกวรรณยุกต์ตรี ในกรณีที่ผลการจำแนกออกมาเป็นวรรณยุกต์อื่นๆ ระบบก็จะส่งต่อไปที่โครงข่ายที่สามซึ่งใช้ในการจำแนกวรรณยุกต์สามัญและโท



รูปที่ 3.36 ผลการประเมินความสามารถในการคัดแยกกลุ่มวรรณยุกต์ 5 เสียงของลักษณะเด่น 64 ตัว โดยการใช้ตัวประเมิน RES

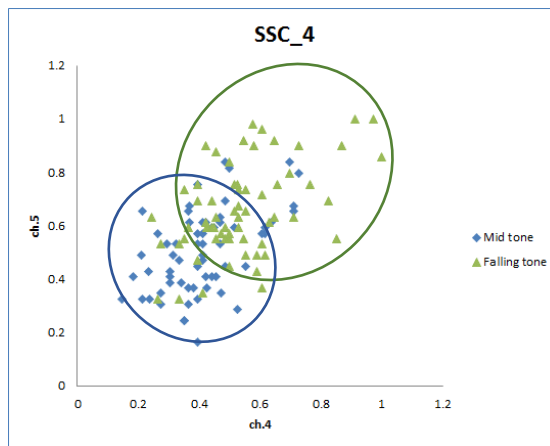


รูปที่ 3.37 แผนภาพการกระจายของลักษณะเด่น (ซ้าย) RES สูงสุด (ขวา) RES ต่ำสุด



(ก)

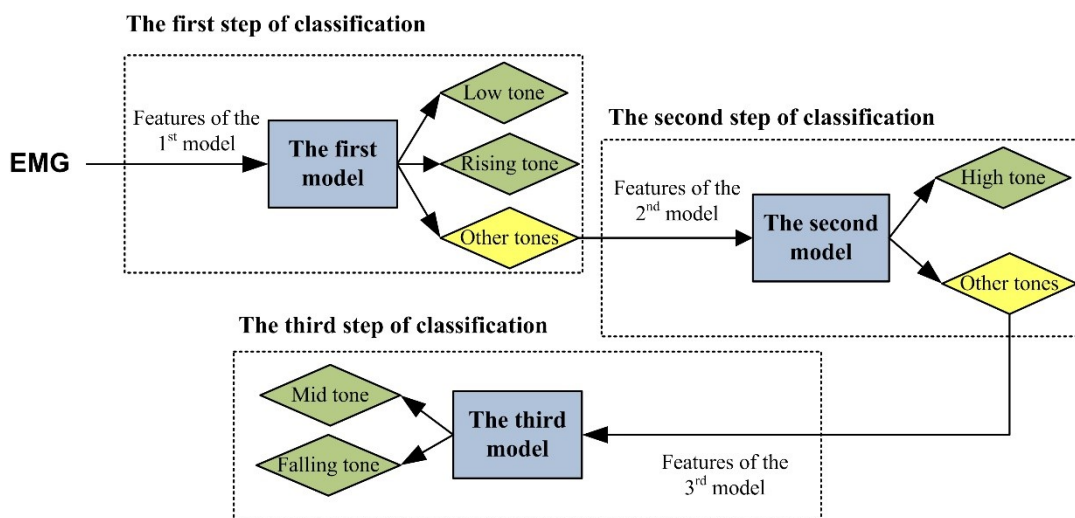
(ข)



(ค)

**รูปที่ 3.38** แผนภาพการกระจายแสดงการแยกแยะกลุ่มของวรรณยุกต์ 5 เสียง (ก) พิจารณา  
วรรณยุกต์ทั้ง 5 เสียง (ข) พิจารณาเฉพาะวรรณยุกต์สามัญ โท และ ตี (ค) พิจารณาเฉพาะ  
วรรณยุกต์สามัญและโท





รูปที่ 3.39 แผนภาพเทคนิคการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย

### (ง) สรุปผลการทดลอง

การทดลองนี้เป็นการทดลองเบื้องต้นเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้ RES เป็นตัวประเมินลักษณะเด่น 64 ตัว และเปรียบเทียบผลการประเมินที่ได้กับผลการจำแนกของตัวจำแนก ANN ผลการทดลองพบว่าลักษณะเด่นที่ให้ผลการประเมินดีที่สุด คือ MMAV2 โดยที่ได้รับความแม่นยำ 56.2% จากการจำแนกโดยใช้ตัวจำแนก ANN

อย่างไรก็ตามประสิทธิภาพในการจำแนกวรรณยุกต์ไทย 5 เสียง ยังต่ำอยู่มาก ดังนั้นจึงได้นำเสนอเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย ซึ่งเป็นผลมาจากการพิจารณาแผนภาพการกระจายแสดงการแยกแยะกลุ่มวรรณยุกต์ทั้ง 5 เสียง สำหรับลักษณะเด่นทั้งหมด โดยเทคนิคนี้จะจำแนกวรรณยุกต์สามขั้นตอน ขั้นตอนแรกจะจำแนกวรรณยุกต์ออกมาสามกลุ่ม (โดยพิจารณาวรรณยุกต์ทั้ง 5 เสียง) คือ กลุ่มวรรณยุกต์เอก กลุ่มวรรณยุกต์จัตวา และ กลุ่มวรรณยุกต์อื่นๆ (สามัญ โท ตรี) ซึ่งถ้าผลการจำแนกในขั้นตอนนี้ออกมาเป็นวรรณยุกต์อื่นๆ ระบบจะส่งต่อไปที่ขั้นตอนที่สอง ขั้นตอนนี้เป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่ม (โดยพิจารณาเฉพาะวรรณยุกต์สามัญ โท และ ตรี) คือ กลุ่มวรรณยุกต์ตรี และกลุ่มวรรณยุกต์อื่นๆ (สามัญ โท) ซึ่งถ้าผลการจำแนกในขั้นตอนนี้ออกมาเป็นวรรณยุกต์อื่นๆ ระบบจะส่งต่อไปที่ขั้นตอนที่สาม ขั้นตอนนี้เป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่มเช่นเดียวกัน (โดย

พิจารณาเฉพาะวรรณยุกต์สามัญและโท) คือ กลุ่มวรรณยุกต์สามัญ และ กลุ่มวรรณยุกต์โท ซึ่งในแต่ละขั้นตอนของการจำแนกจะใช้ลักษณะเด่นที่แตกต่าง

การทดลองต่อไปจะเป็นการทดสอบประสิทธิภาพของเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่ายที่ได้นำเสนอ โดยการประเมินและคัดเลือกลักษณะเด่นในแต่ละโครงข่ายของการจำแนก และทดสอบประสิทธิภาพในการจำแนกโดยใช้ตัวจำแนก ANN

### 3.3.2.2 การทดลองเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์สำหรับเทคนิคการจำแนกแบบสามขั้นตอนซึ่งใช้ตัวจำแนก ANN สามโครงข่าย

#### (ก) วัตถุประสงค์

- เพื่อประเมินและคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อ สำหรับเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย

- เพื่อทดสอบประสิทธิภาพในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้เทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย

#### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อเช่นเดียวกับการทดลองในหัวข้อ 3.3.2.1 ซึ่งใช้ข้อมูลของอาสาสมัครเพียง 1 คน (S01) โดยใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ซึ่งใช้คำพูดชุดที่ 1 ถึง 21 ดังตารางที่ 3.3 เป็นคำพูดโดด 105 คำซึ่งครอบคลุมการออกเสียงวรรณยุกต์ทั้ง 5 เสียง และครอบคลุมหน่วยเสียงพยัญชนะทั้งหมด ในแต่ละคำจะถูกพูดซ้ำ 3 ครั้ง (945 ข้อมูล) คำพูดชุดที่ 1 2 8 9 และ 15 ซึ่งมีการผันวรรณยุกต์ทั้ง 5 เสียงถูกใช้สำหรับการทดสอบระบบ (225 ข้อมูล) สำหรับข้อมูลชุดที่เหลือจะถูกใช้สำหรับการฝึกสอนระบบ (720 ข้อมูล)

#### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อมาคำนวณลักษณะเด่น 64 ตัว เช่นเดียวกับการทดลองในหัวข้อ 3.3.2.1 ซึ่งแสดงดังตารางที่ 3.17 และตารางที่ 3.18

2) ประเมินลักษณะเด่นทั้งหมดโดยใช้ตัวประเมิน RES โดยจะแบ่งการประเมินออกเป็น 3 กลุ่มซึ่งแสดงดังรูปที่ 3.40

กลุ่มแรก เป็นการประเมินลักษณะเด่นเพื่อใช้ในขั้นตอนที่หนึ่ง (โครงข่ายที่หนึ่ง) สำหรับเทคนิคการจำแนกแบบสามขั้นตอน ซึ่งจะพิจารณาวรรณยุกต์ทั้ง 5 เสียง แล้วประเมินความสามารถในการแยกแยะกลุ่มข้อมูลวรรณยุกต์ออกเป็น 3 กลุ่ม คือ วรรณยุกต์เอก วรรณยุกต์จัตวา และวรรณยุกต์อื่นๆ (สามัญ โท และ ตรี)

กลุ่มสอง เป็นการประเมินลักษณะเด่นเพื่อใช้ในขั้นตอนที่สอง (โครงข่ายที่สอง) สำหรับเทคนิคการจำแนกแบบสามขั้นตอน ซึ่งจะพิจารณาวรรณยุกต์เฉพาะวรรณยุกต์สามัญ โท และ ตรี แล้วประเมินความสามารถในการแยกแยะกลุ่มข้อมูลวรรณยุกต์ออกเป็น 2 กลุ่ม คือ วรรณยุกต์ตรี และ วรรณยุกต์อื่นๆ (สามัญ และ โท)

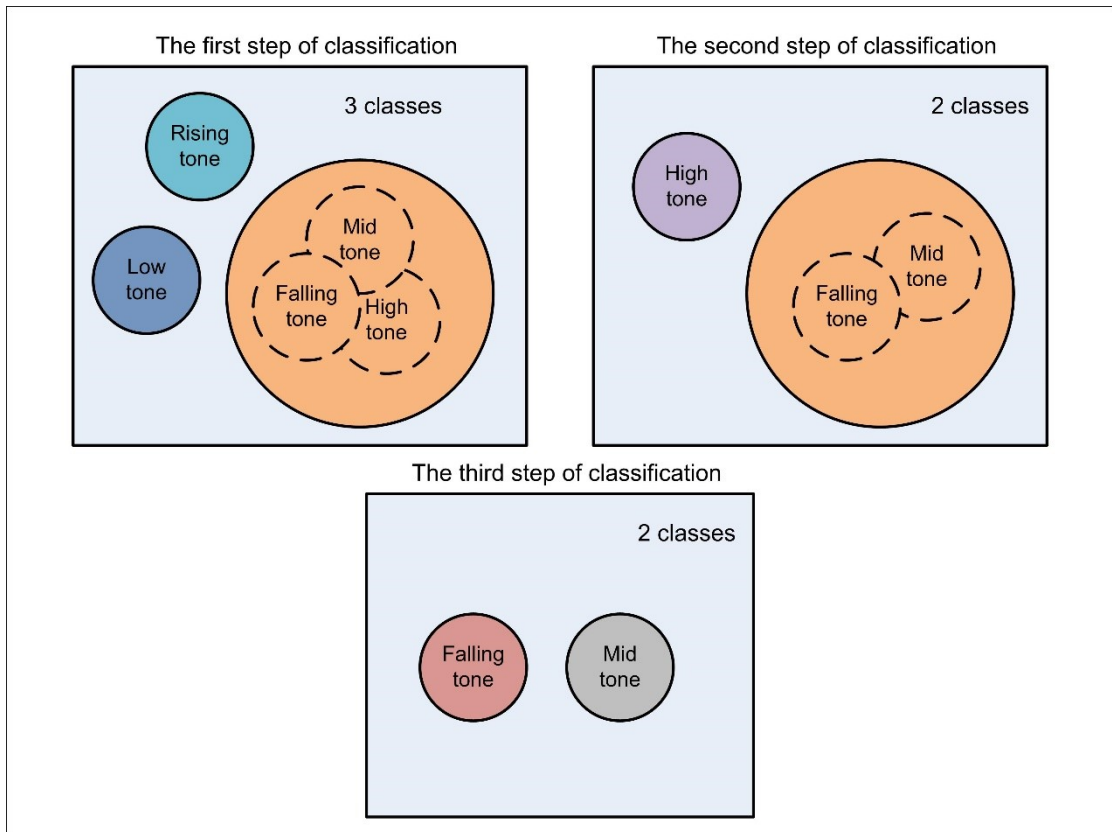
กลุ่มสาม เป็นการประเมินลักษณะเด่นเพื่อใช้ในขั้นตอนที่สาม (โครงข่ายที่สาม) สำหรับเทคนิคการจำแนกแบบสามขั้นตอน ซึ่งจะพิจารณาวรรณยุกต์เฉพาะวรรณยุกต์สามัญและ โท แล้วประเมินความสามารถในการแยกแยะกลุ่มข้อมูลวรรณยุกต์ออกเป็น 2 กลุ่ม คือ วรรณยุกต์สามัญ และ โท

3) เลือกลักษณะเด่นที่มีผลการประเมินดีที่สุดในแต่ละกลุ่มเพื่อเป็นอินพุตให้กับตัวจำแนก ANN ในแต่ละโครงข่าย หลังจากนั้นฝึกสอนและทดสอบระบบเพื่อวัดประสิทธิภาพในการจำแนกวรรณยุกต์ในแต่ละโครงข่าย

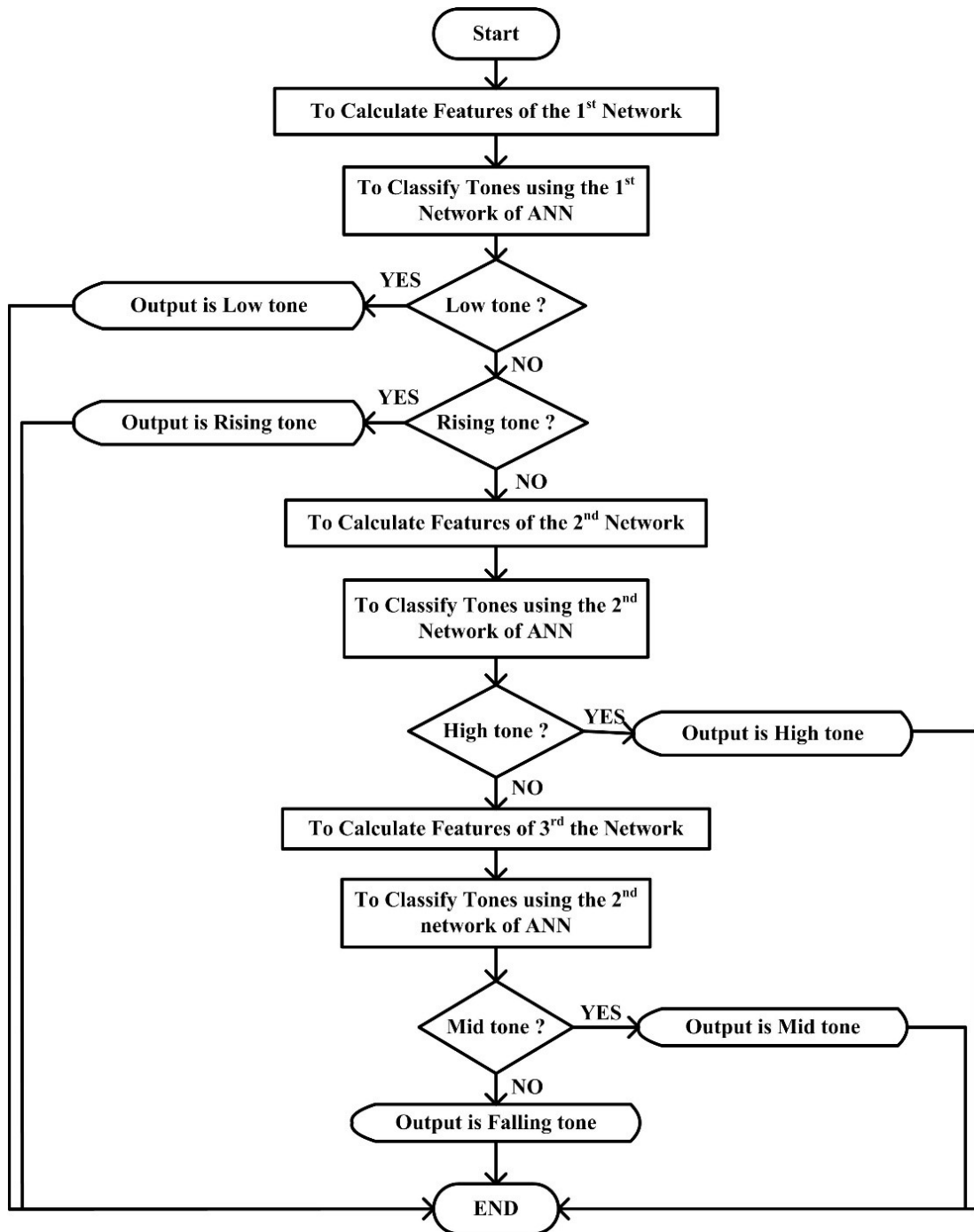
4) วัดประสิทธิภาพของการจำแนกวรรณยุกต์ทั้ง 5 เสียง (ความแม่นยำรวมของระบบ) โดยใช้โปรแกรมการจำแนกวรรณยุกต์แบบสามขั้นตอนซึ่งใช้ตัวจำแนก ANN อนุกรมกันสามโครงข่าย ซึ่งถูกเขียนขึ้นโดยใช้โปรแกรม MATLAB โดยขั้นตอนการทำงานของโปรแกรมแสดงดังรูปที่ 3.41

การทำงานของโปรแกรมเริ่มจากลักษณะเด่นกลุ่มที่หนึ่ง จะถูกป้อนให้กับตัวจำแนก ANN โครงข่ายที่หนึ่ง ถ้าผลการจำแนก คือ วรรณยุกต์เอก หรือ วรรณยุกต์จัตวา โปรแกรมจะแสดงผลการจำแนกออกมาและสิ้นสุดการทำงาน แต่ถ้าผลการจำแนกคือ วรรณยุกต์อื่น ระบบจะทำงานต่อไปโดยจะนำลักษณะเด่นในกลุ่มที่สองป้อนให้กับตัวจำแนก ANN โครงข่ายที่สอง ถ้าผลการจำแนก คือ วรรณยุกต์ตรี โปรแกรมจะแสดงผลการจำแนกออกมาและสิ้นสุดการทำงาน แต่ถ้าผลการจำแนกคือ วรรณยุกต์อื่น ระบบจะทำงานต่อไปโดยจะนำลักษณะเด่นในกลุ่มที่สามป้อนให้กับตัวจำแนก ANN โครงข่ายที่สาม ซึ่งจะจำแนกวรรณยุกต์สามัญ หรือ วรรณยุกต์ตรี หลังจากนั้นโปรแกรมจะแสดงผลการจำแนกออกมาและสิ้นสุดการทำงาน

5) เพิ่มจำนวนอินพุตให้กับตัวจำแนก ANN แต่ละโครงข่าย โดยเพิ่มจำนวนลักษณะเด่น 2 ตัว และ 3 ตัว หลังจากนั้นทดสอบประสิทธิภาพในการจำแนก



รูปที่ 3.40 แผนภาพแสดงการแบ่งกลุ่มวรรณยุกต์ในแต่ละขั้นตอนของเทคนิคการจำแนกแบบสามขั้นตอน



รูปที่ 3.41 ผังงานแสดงการทำงานของโปรแกรมการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

#### 1) ผลการประเมินลักษณะเด่นทั้งสามกลุ่ม

ตารางที่ 3.20 แสดงผลการประเมินลักษณะเด่นในแต่ละกลุ่มของเทคนิคการจำแนกแบบสามขั้นตอน โดยการใช้ตัวประเมิน RES พิจารณาลักษณะเด่นในกลุ่มที่ 1 พบว่าลักษณะเด่น MMAV2 ให้ผลการประเมินดีที่สุด (RES=2.90) สำหรับลักษณะเด่นในกลุ่มที่ 2 MAV\_5 ให้ผลการประเมินดีที่สุด (RES=1.90) สำหรับลักษณะเด่น SSC\_4 คือ ลักษณะเด่นที่มีผลการประเมินที่ดีที่สุดในกลุ่มที่ 3 (RES=1.33) พิจารณาลักษณะเด่นที่ดีที่สุดในแต่ละกลุ่ม พบว่ากลุ่มที่ 1 ให้ค่า RES สูงที่สุด ซึ่งผลการจำแนกควรจะให้ความแม่นยำสูงสุดเช่นเดียวกัน

#### 2) ผลความแม่นยำในการจำแนกโดยใช้เทคนิคการจำแนกแบบสามขั้นตอน

ตารางที่ 3.21 แสดงผลความแม่นยำในการจำแนกวรรณยุกต์โดยใช้ตัวจำแนก ANN 3 โครงข่าย โดยบ่อนลักษณะเด่นที่มีผลการประเมินที่ดีที่สุดในแต่ละกลุ่มให้กับตัวจำแนก ANN สำหรับโครงข่ายที่ 1 ซึ่งใช้ในการจำแนกวรรณยุกต์เอกและจัตวา ให้ผลความแม่นยำในการจำแนก 84.8% สำหรับโครงข่ายที่ 2 ซึ่งใช้ในการจำแนกวรรณยุกต์ตรี ให้ผลความแม่นยำในการจำแนก 82.5% และโครงข่ายที่ 3 ซึ่งใช้ในการจำแนกวรรณยุกต์สามัญและโทให้ผลความแม่นยำในการจำแนก 78.6 % หลังจากนั้น เมื่อนำโครงข่ายที่ได้จากการฝึกสอนและทดสอบระบบเข้าสู่โปรแกรมการจำแนกวรรณยุกต์แบบสามขั้นตอนที่ได้ออกแบบไว้จะได้ความแม่นยำรวมของระบบในการจำแนกวรรณยุกต์ทั้ง 5 เสียง คือ 68.61% จากการทดลองนี้จะเห็นได้ว่า เทคนิคการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่ายสามารถปรับปรุงความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียงได้ โดยความแม่นยำเพิ่มขึ้น 12.41% เมื่อเปรียบเทียบกับผลการจำแนกที่ได้จากการทดลองในหัวข้อ 3.3.2.1 ซึ่งจำแนกโดยใช้ตัวจำแนก ANN เพียง 1 โครงข่าย ซึ่งได้รับความแม่นยำในการจำแนกเพียง 56.2%

รูปที่ 3.42 แสดงผลความแม่นยำในการจำแนกวรรณยุกต์เมื่อเพิ่มจำนวนอินพุตให้กับตัวจำแนก ANN แต่ละโครงข่าย (ก) โครงข่ายที่ 1 (ข) โครงข่ายที่ 2 (ค) โครงข่ายที่ 3 โดยเพิ่มจำนวนลักษณะเด่น 2 ตัว และ 3 ตัว ซึ่งเลือกจากลักษณะเด่นที่มีผลการประเมินดีที่สุด 3 อันดับแรก สำหรับโครงข่ายที่ 1 จะเห็นได้ว่าเมื่อเพิ่มจำนวนลักษณะเด่น 3 ตัว ให้ผลความแม่นยำในการจำแนกเพิ่มขึ้นเป็น 91.4% สำหรับโครงข่ายที่ 2 ลักษณะเด่นที่เพิ่มขึ้นไม่สามารถเพิ่มความแม่นยำในการจำแนก โดยการใช้ลักษณะเด่น MAV\_5 เพียง 1 ตัวให้ความแม่นยำสูงสุด 82.5% สำหรับโครงข่ายที่ 3 การเพิ่มลักษณะเด่น 3 ตัวให้ผลความแม่นยำสูงสุดเช่นเดียวกับโครงข่ายที่ 1 โดยให้

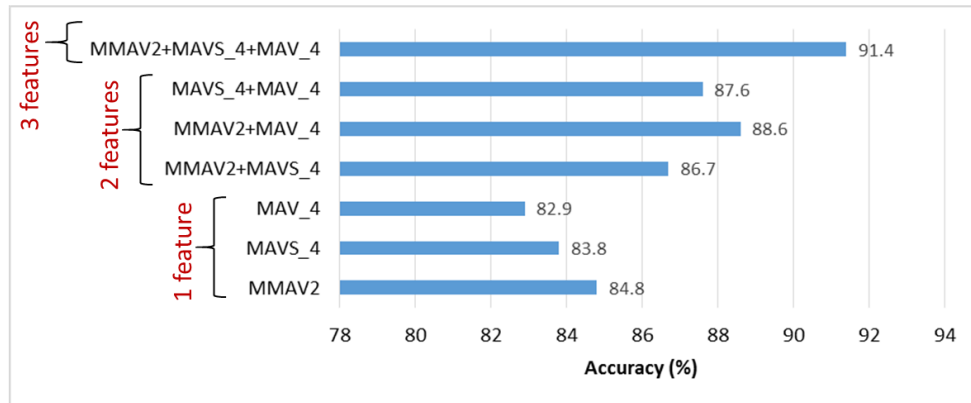
ความแม่นยำ 88.1% ดังนั้น สรุปได้ว่าโครงข่ายที่ 1 เลือกใช้ลักษณะเด่น 3 ตัว คือ MMAV2 MAV\_4 และ MAVS\_4 โครงข่ายที่ 2 เลือกใช้ลักษณะเด่นเพียง 1 ตัว คือ MAV\_5 สำหรับโครงข่ายที่ 3 เลือกใช้ลักษณะเด่น 3 ตัว คือ SSC\_3 SSC\_4 และ MAVS\_2 ทดสอบประสิทธิภาพเทคนิคการจำแนกแบบสามชั้นตอนโดยใช้โปรแกรมที่ได้ออกแบบไว้ ได้ความแม่นยำรวมของระบบ คือ 73.11% ซึ่งเพิ่มขึ้นจากการใช้ลักษณะเด่นเพียง 1 ตัวสำหรับแต่ละโครงข่าย (68.61%)

**ตารางที่ 3.20** ลักษณะเด่นที่มีผลการประเมินด้วย RES ดีที่สุด 8 อันดับแรกในแต่ละกลุ่ม

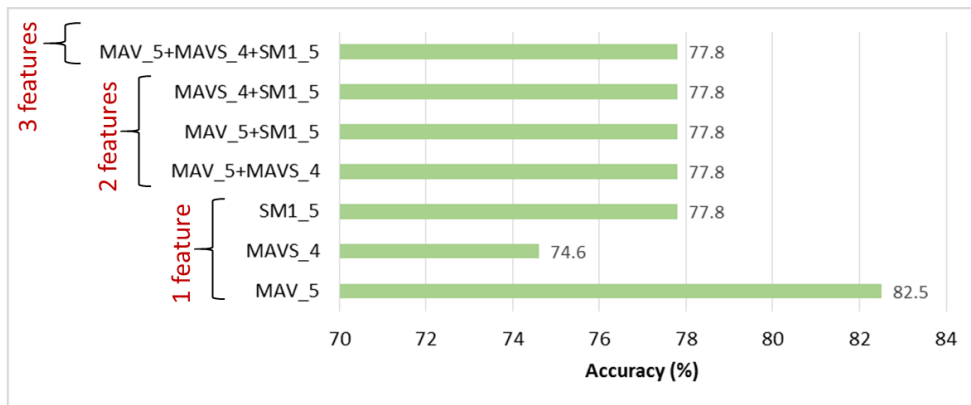
อันดับ	กลุ่ม 1		กลุ่ม 2		กลุ่ม 3	
	ลักษณะเด่น	RES	ลักษณะเด่น	RES	ลักษณะเด่น	RES
1	MMAV2	2.90	MAV_5	1.90	SSC_4	1.33
2	MAVS_4	2.50	MAVS_4	1.87	SSC_3	1.29
3	MAV_4	2.48	SM1_5	1.86	MAVS_2	1.12
4	RMS_4	2.42	RMS_5	1.82	SSC	1.08
5	MAV_3	2.30	RMS	1.62	WAMP_5	1.00
6	RMS_3	2.30	VAR	1.55	RMS_1	0.90
7	ZC_3	2.25	MAV	1.45	MAV_1	0.85
8	MAVS_3	2.22	MAVS_2	1.28	RMS_3	0.83

**ตารางที่ 3.21** ผลความแม่นยำในการจำแนกวรรณยุกต์โดยใช้ตัวจำแนก ANN

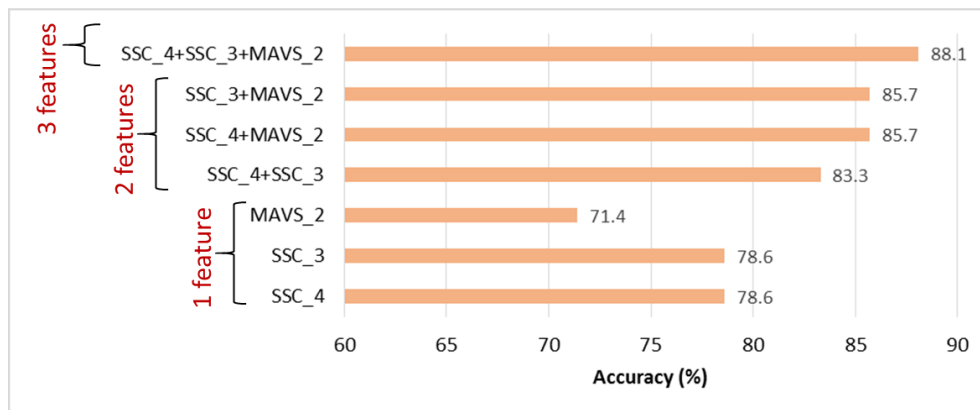
	ลักษณะเด่น	Accuracy (%)	Total Accuracy (%)
โครงข่ายที่ 1	MMAV2	84.8	
โครงข่ายที่ 2	MAV_5	82.5	68.61%
โครงข่ายที่ 3	SSC_4	78.6	



(ก)



(ข)



(ค)

รูปที่ 3.42 ผลความแม่นยำในการจำแนกในแต่ละโครงข่ายสำหรับเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย (ก) โครงข่ายที่ 1 (ข) โครงข่ายที่ 2 (ค) โครงข่ายที่ 3



### (ง) สรุปผลการทดลอง

การทดลองนี้เป็นการทดลองเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์สำหรับเทคนิคการจำแนกแบบสามขั้นตอนซึ่งใช้ตัวจำแนก ANN สามโครงข่าย ซึ่งในแต่ละโครงข่ายจะใช้ลักษณะเด่นที่แตกต่างกันสำหรับจำแนกวรรณยุกต์ที่แตกต่างกัน ประเมินลักษณะเด่น 64 ตัวโดยใช้ตัวประเมิน RES โดยเลือกลักษณะเด่นที่มีผลการประเมินที่ดีที่สุดในแต่ละโครงข่ายเพื่อเป็นอินพุตให้กับตัวจำแนก ANN สำหรับโครงข่ายที่หนึ่งที่ใช้ในการจำแนกวรรณยุกต์เอกและจัตวา ลักษณะเด่นที่ดีที่สุดคือ MMAV2 สำหรับโครงข่ายที่สองที่ใช้ในการจำแนกวรรณยุกต์ตรี ลักษณะเด่นที่ดีที่สุดคือ MAV\_5 และโครงข่ายที่สามที่ใช้ในการจำแนกวรรณยุกต์สามัญและโท ลักษณะเด่นที่ดีที่สุดคือ SSC\_4 โดยที่การจำแนกวรรณยุกต์ 5 เสียงสำหรับเทคนิคดังกล่าวให้ความแม่นยำ 68.61% ซึ่งสามารถปรับปรุงความแม่นยำ 12.41% นอกจากนั้นทำการปรับจำนวนลักษณะเด่นในแต่ละโครงข่าย พบว่าการเพิ่มจำนวนลักษณะเด่น 3 ตัวสามารถเพิ่มความแม่นยำให้กับระบบ ยกเว้นโครงข่ายที่สองซึ่งการใช้ MAV\_5 เพียงตัวเดียวให้ความแม่นยำสูงสุด โดยความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียงโดยใช้เทคนิคการจำแนกแบบสามขั้นตอนซึ่งมีการปรับจำนวนลักษณะเด่นเป็น 3 ตัว (สำหรับโครงข่ายที่ 1 และ โครงข่ายที่ 3) มีผลการจำแนกเท่ากับ 73.11%

การทดลองนี้แสดงให้เห็นว่าเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่ายที่ได้นำเสนอสามารถปรับปรุงความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง อย่างไรก็ตามเทคนิคดังกล่าวถูกออกแบบโดยใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครเพียง 1 คน (S01) ดังนั้นเพื่อให้สามารถนำเทคนิคที่ออกแบบไปใช้ได้กับข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครที่ครอบคลุมมากขึ้น จึงจำเป็นต้องมีการปรับปรุงเทคนิคการจำแนกวรรณยุกต์ด้วยเทคนิคดังกล่าวโดยการออกแบบจะอ้างอิงข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน (S01-S10) ซึ่งจะนำเสนอในการทดลองต่อไป

3.3.2.3 การทดลองเพื่อคัดเลือกลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์สำหรับรูปแบบการจำแนกแบบสามขั้นตอนซึ่งใช้ตัวจำแนก ANN สามโครงข่าย ซึ่งออกแบบจากสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน โดยใช้ RES เป็นตัวประเมินลักษณะเด่น

#### (ก) วัตถุประสงค์

- เพื่อปรับปรุงเทคนิคการจำแนกวรรณยุกต์แบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย โดยออกแบบจากสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน เพื่อให้สามารถนำเทคนิคดังกล่าวไปใช้ได้กับข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครที่ครอบคลุมมากขึ้น

- เพื่อประเมินและคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อ สำหรับเทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย โดยอ้างอิงข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน

- เพื่อทดสอบประสิทธิภาพในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้เทคนิคการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย และเปรียบเทียบกับเทคนิคการจำแนกแบบเดิมที่ใช้การจำแนกแบบขั้นตอนเดียวโดยใช้ตัวจำแนก ANN 1 โครงข่าย

#### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อ โดยใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ทั้งหมด ดังตารางที่ 3.3 ในแต่ละคำจะถูกพูดซ้ำ 9 ครั้ง (1,332 ข้อมูล สำหรับสัญญาณของอาสาสมัครแต่ละคน) สำหรับคำพูดชุดที่ 1 8 15 22 และ 23 จะถูกใช้สำหรับการทดสอบระบบ (225 ข้อมูล) สำหรับคำพูดชุดที่เหลือจะถูกใช้สำหรับกระบวนการฝึกสอนระบบ (1,107 ข้อมูล) การแบ่งชุดคำพูดสำหรับการฝึกสอนและทดสอบระบบดังกล่าว เพื่อเป็นการทดสอบว่าระบบที่ออกแบบสามารถจำแนกวรรณยุกต์ได้ กล่าวคือ ถึงแม้ว่าจะออกเสียงพูดคำพูดที่มีองค์ประกอบของพยัญชนะและสระต่างกัน แต่วรรณยุกต์เดียวกัน ระบบจะสามารถจัดกลุ่มคำนั้นให้อยู่ในกลุ่มของวรรณยุกต์เดียวกันได้

#### ขั้นตอนการทดลอง

- 1) นำข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมดมาคำนวณลักษณะเด่น ซึ่งในการทดลองนี้จะแบ่งวิธีการคำนวณลักษณะเด่นออกเป็นสองรูปแบบเช่นเดียวกับการทดลองใน

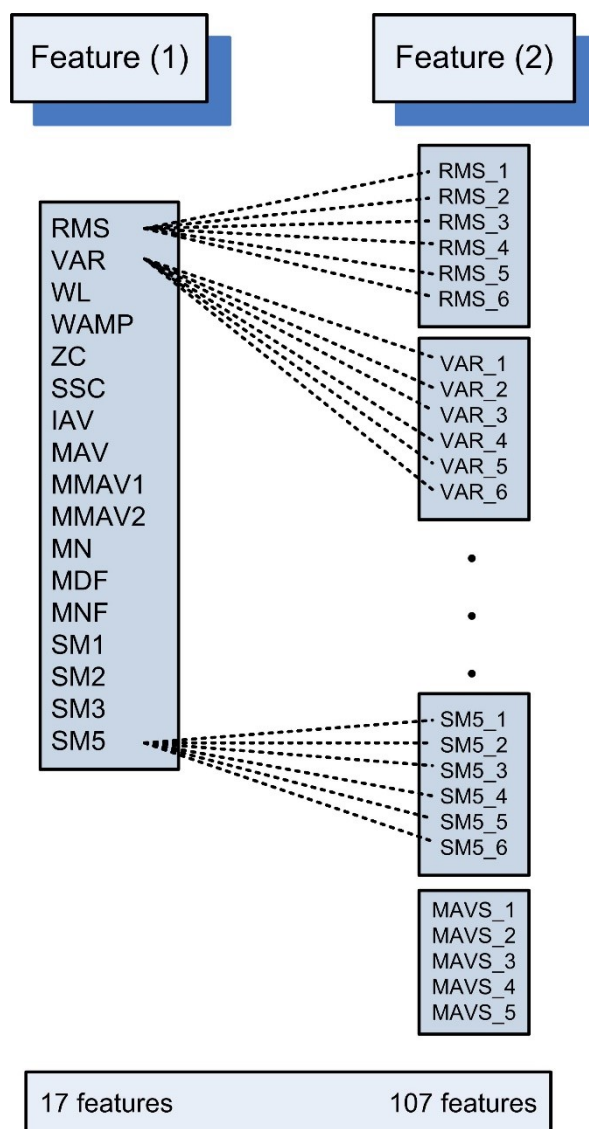
หัวข้อที่ 3.3.2.1 และ 3.3.2.2 โดยรูปแบบที่หนึ่งเป็นการคำนวณโดยใช้ข้อมูลของสัญญาณทั้งหมด (ตั้งแต่ช่วงเริ่มต้นและสิ้นสุดของคำพูดแต่ละเสียง) ซึ่งลักษณะเด่นทั้งหมดในรูปแบบนี้ถูกแสดงดังตารางที่ 3.17 ซึ่งมีทั้งหมด 17 ชนิด(ตัว) และ รูปแบบที่สองเป็นการคำนวณโดยแบ่งสัญญาณไฟฟ้ากล่อมเนื้อออกเป็น 6 ส่วนเท่าๆ กัน แล้วนำสัญญาณในแต่ละส่วนมาคำนวณลักษณะเด่น ซึ่งทำให้ได้ลักษณะเด่นตัวใหม่เพิ่มขึ้นมา ซึ่งในการทดลองในหัวข้อ 3.3.2.1 และ 3.3.2.2 จะนำสัญญาณในแต่ละส่วนมาคำนวณลักษณะเด่น 8 ชนิดจากตารางที่ 3.17 สำหรับในการทดลองนี้จะนำสัญญาณในแต่ละส่วนมาคำนวณลักษณะเด่นทั้งหมด (17 ชนิด) จากตารางที่ 3.17 ซึ่งจะทำให้ได้ลักษณะเด่นในรูปแบบที่สอง 107 ตัว

ลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมดในการทดลองนี้ถูกแสดงดังรูปที่ 3.43 ซึ่งมีทั้งหมด 124 ตัว

2) ทำการประเมินลักษณะเด่นทั้งหมดสำหรับสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน โดยใช้ตัวประเมิน RES ซึ่งการประเมินลักษณะเด่นถูกแบ่งออกเป็น 2 รูปแบบ ดังนี้

รูปแบบที่หนึ่ง การประเมินลักษณะเด่นสำหรับใช้ในระบบจำแนกแบบขั้นตอนเดียว (single step of classification) โดยประเมินความสามารถในการแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่ม ซึ่งเป็นกลุ่มของเสียงวรรณยุกต์ทั้ง 5 เสียง รูปแบบนี้จะจำแนกวรรณยุกต์ออกมา 5 เสียงภายในขั้นตอนเดียว

รูปแบบที่สอง การประเมินลักษณะเด่นสำหรับใช้ในระบบจำแนกแบบสามขั้นตอน (three steps of classification) รูปแบบนี้จะจำแนกวรรณยุกต์ออกมาทีละขั้นตอน โดยขั้นตอนแรกจะจำแนกวรรณยุกต์ออกมาสามกลุ่ม (โดยพิจารณาวรรณยุกต์ทั้ง 5 เสียง) คือ กลุ่มวรรณยุกต์เอก กลุ่มวรรณยุกต์จัตวา และ กลุ่มวรรณยุกต์อื่นๆ (สามัญ โท ตรี) ขั้นตอนที่สองเป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่ม (โดยพิจารณาเฉพาะวรรณยุกต์สามัญ โท และ ตรี) คือ กลุ่มวรรณยุกต์ตรี และ กลุ่มวรรณยุกต์อื่นๆ (สามัญ โท) ขั้นตอนที่สามเป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่มเช่นเดียวกัน (โดยพิจารณาเฉพาะวรรณยุกต์สามัญ และ โท) คือ กลุ่มวรรณยุกต์สามัญ และ กลุ่มวรรณยุกต์โท ดังนั้นในรูปแบบนี้จะแบ่งการประเมินออกเป็น 3 กลุ่มย่อยตามขั้นตอนของการจำแนก ซึ่งจะประเมินความสามารถในการแบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่ม 2 กลุ่ม และ 2 กลุ่ม ตามลำดับ



**รูปที่ 3.43** ลักษณะเด่นของสัญญาณไฟฟ้าก่ล้ามนเนื่องจากการคำนวณทั้งสองรูปแบบ ทั้งหมด 124 ตัว ซึ่งถูกใช้ในการทดลองในหัวข้อ 3.3.2.3

3) เลือกลักษณะเด่นเพื่อใช้เป็นอินพุตให้กับตัวจำแนก โดยทำการเลือกลักษณะเด่นในแต่ละรูปแบบการประเมิน โดยมีขั้นตอนในการเลือกลักษณะเด่น ดังต่อไปนี้

3.1) เลือกลักษณะเด่นที่มีค่า RES สูงที่สุด 10 อันดับแรกสำหรับสัญญาณไฟฟ้าก่ล้ามนี้อของอาสาสมัครแต่ละคน (10 คน) ดังนั้นในขั้นตอนนี้จะได้ลักษณะเด่นทั้งหมด 100 ตัว

3.2) พิจารณาลักษณะเด่นทั้งหมด (100 ตัว) ที่ถูกเลือกจากหัวข้อที่ 3.1) จากนั้นหาความถี่ของลักษณะเด่นแต่ละตัว ซึ่งความถี่สูงสุดเท่ากับ 10 นั่นคือ ข้อมูลสัญญาณไฟฟ้า

กล่ามเนื้อของอาสาสมัครทั้ง 10 คน มีลักษณะเด่นตัวนั้นปรากฏอยู่ในกลุ่มของลักษณะเด่นที่มี RES สูงสุด 10 อันดับแรกจากหัวข้อที่ 3.1)

3.3) เรียงลำดับลักษณะเด่นจากความถี่สูงสุดไปหาความถี่ต่ำสุด อย่างไรก็ตาม ในกรณีที่ความถี่ของลักษณะเด่นเท่ากัน จะเรียงลำดับโดยพิจารณาค่าเฉลี่ยของ RES ซึ่งคำนวณจากค่าเฉลี่ยของ RES ของสัญญาณไฟฟ้ากล่ามเนื้อของอาสาสมัครทั้งหมด

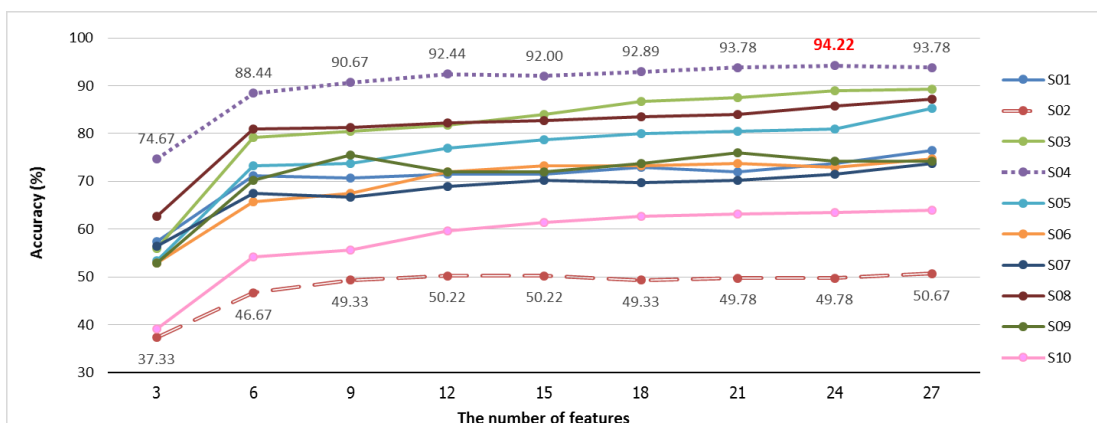
การเลือกลักษณะเด่นจากวิธีดังกล่าว จะทำให้ได้ลักษณะเด่นเพียงชุดเดียวที่สามารถนำไปใช้ได้กับสัญญาณไฟฟ้ากล่ามเนื้อของอาสาสมัครทั้งหมด เนื่องจากกระบวนการเลือกลักษณะเด่นข้างต้น เลือกลงจากการพิจารณาข้อมูลของสัญญาณไฟฟ้ากล่ามเนื้อจากอาสาสมัครทั้งหมด

4) ทำการฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก ANN เพื่อจำแนกวรรณยุกต์ 5 เสียง โดยนำลักษณะเด่นที่ได้จากผลการประเมินและเลือกลักษณะเด่นมาเป็นอินพุตให้กับตัวจำแนก โดยทำการทดลองปรับจำนวนอินพุตของตัวจำแนก ANN โดยใช้จำนวนลักษณะเด่น 3 5 7 9 12 15 18 21 24 และ 27 ตามลำดับ โดยเรียงลำดับลักษณะเด่นที่ดีที่สุด ในการทดลองนี้จะทดสอบประสิทธิภาพของรูปแบบการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย โดยจะเปรียบเทียบกับการจำแนกแบบขั้นตอนเดียวซึ่งใช้ตัวจำแนก ANN หนึ่งโครงข่าย

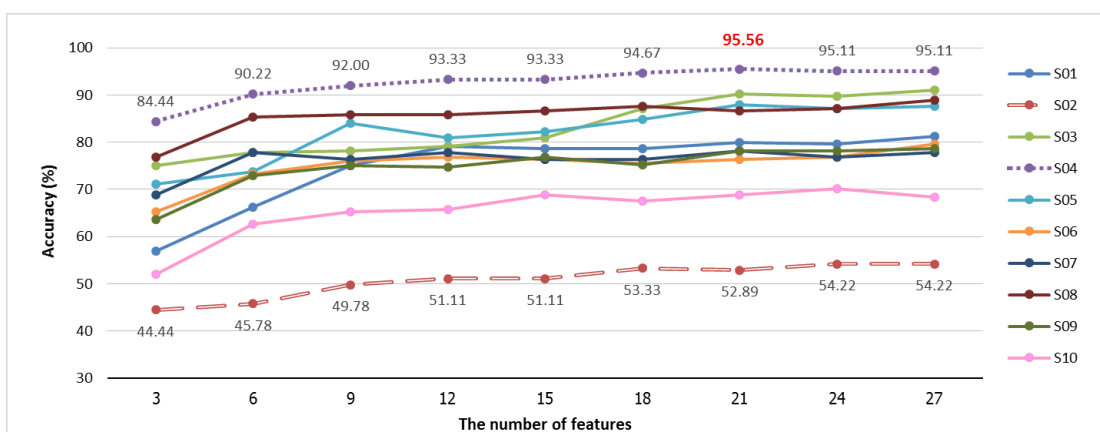
### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

#### 1) เปรียบเทียบผลการจำแนกโดยใช้รูปแบบการจำแนกทั้งสองรูปแบบ

รูปที่ 3.44 แสดงผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง สำหรับรูปแบบการจำแนกทั้งสองรูปแบบ คือ (ก) การจำแนกแบบขั้นตอนเดียว และ (ข) การจำแนกแบบสามขั้นตอน โดยเปรียบเทียบความแม่นยำในการจำแนกของสัญญาณไฟฟ้ากล่ามเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) ขณะที่มีการเพิ่มจำนวนลักษณะเด่นจาก 3 ตัว จนถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว จะเห็นได้ว่าสัญญาณไฟฟ้ากล่ามเนื้อของอาสาสมัครเกือบทั้งหมดจากทั้งสองรูปแบบการจำแนก ให้ผลไปในทิศทางเดียวกัน กล่าวคือ เมื่อจำนวนลักษณะเด่นเพิ่มขึ้นความแม่นยำในการจำแนกจะเพิ่มขึ้นด้วย โดยความแม่นยำจะเพิ่มขึ้นอย่างมากในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 และความแม่นยำจะเพิ่มขึ้นเล็กน้อยและค่อยๆ คงที่ในช่วงการปรับจำนวนลักษณะเด่นตั้งแต่ 12 ตัวเป็นต้นไป



(ก)



(ข)

รูปที่ 3.44 กราฟแสดงผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง ซึ่งเปรียบเทียบความแม่นยำของข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน

เมื่อพิจารณาในแต่ละช่วงของการปรับจำนวนลักษณะเด่น จะเห็นได้ว่าสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคนให้ผลความแม่นยำที่แตกต่างกัน ซึ่งมีความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุดอย่างเห็นได้ชัด พิจารณาผลการจำแนกแบบขั้นตอนเดียวซึ่งแสดงในรูปที่ 3.44(ก) ณ ตำแหน่งการใช้จำนวนลักษณะเด่น 3 ตัว จะเห็นได้ว่าความแม่นยำสูงสุด คือ 76.67% เป็นผลความแม่นยำของ S04 ในขณะที่ความแม่นยำต่ำสุด คือ 37.33% ได้จากสัญญาณของ S02 จะเห็นได้ว่าเป็นความแตกต่างกัน 39.34% และเมื่อพิจารณา ณ ตำแหน่งการใช้จำนวนลักษณะเด่นที่ให้ผลความแม่นยำสูงสุดจากรูปแบบการจำแนกแบบขั้นตอนเดียว พบว่าการใช้จำนวนลักษณะเด่น 24 ตัวให้ผลความแม่นยำสูงสุดซึ่งได้รับจากสัญญาณ S04

เช่นเดียวกัน โดยความแม่นยำที่ได้ คือ 94.22% ในขณะที่ความแม่นยำต่ำสุด คือ 49.78% ซึ่งเป็นสัญญาณของ S02 เช่นเดียวกัน โดยที่ความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุด ณ ตำแหน่งการใช้ลักษณะเด่น 24 ตัว คือ 44.44% ดังนั้นแสดงให้เห็นว่าสัญญาณของ S04 ให้ผลความแม่นยำสูงสุด และ S02 ให้ผลความแม่นยำต่ำสุด โดยที่มีความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุดมาก ซึ่งให้ผลเช่นเดียวกันกับรูปแบบการจำแนกแบบสามขั้นตอนในรูปแบบที่ 3.44(ข) นอกจากนี้เมื่อพิจารณาการจำแนกแบบสามขั้นตอน จะเห็นได้ว่า S02 ให้ผลความแม่นยำที่แตกต่างกับสัญญาณอื่นๆ อย่างเห็นได้ชัด โดยให้ความแม่นยำที่ต่ำมาก อย่างไรก็ตามการเพิ่มจำนวนลักษณะเด่นสามารถเพิ่มความแม่นยำในการจำแนก โดยให้ผลการเปลี่ยนแปลงไปในทิศทางเดียวกันสำหรับข้อมูลสัญญาณทั้งหมด (S01-S10)

พิจารณาเฉพาะสัญญาณ S04 ซึ่งให้ผลการจำแนกดีที่สุดจากทั้งสองรูปแบบการจำแนก พบว่าสำหรับการจำแนกแบบขั้นตอนเดียว การเพิ่มจำนวนลักษณะเด่น 24 ตัว ให้ความแม่นยำสูงสุด (94.22%) ซึ่งสามารถเพิ่มความแม่นยำ 19.55% เมื่อเปรียบเทียบกับการใช้จำนวนลักษณะเด่นเพียง 3 ตัว ในขณะที่การจำแนกแบบสามขั้นตอน การเพิ่มจำนวนลักษณะเด่น 21 ตัว ให้ผลความแม่นยำสูงสุด (95.56%) ซึ่งความแม่นยำเพิ่มขึ้น 10.67% เมื่อเปรียบเทียบกับการใช้จำนวนลักษณะเด่นเพียง 3 ตัว ดังนั้นจะเห็นได้ว่ารูปแบบการจำแนกแบบสามขั้นตอนให้ความแม่นยำที่ดีกว่าการจำแนกแบบขั้นตอนเดียว และใช้จำนวนลักษณะเด็มน้อยกว่า

อย่างไรก็ตามยังไม่สามารถสรุปจำนวนลักษณะเด่นที่เหมาะสมในงานวิจัยนี้ได้ อย่างชัดเจน เนื่องจากความแม่นยำสูงสุดที่เกิดขึ้นในแต่ละสัญญาณเกิดจากการใช้จำนวนลักษณะเด่นที่ต่างกันไป ดังนั้นเพื่อให้สามารถวิเคราะห์จำนวนลักษณะเด่นที่เหมาะสม และระบุอัตราการเปลี่ยนแปลงของความแม่นยำในขณะที่มีการปรับจำนวนลักษณะเด่นได้อย่างชัดเจน จึงนำผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมด (S01-S10) มาคำนวณหาค่าเฉลี่ย

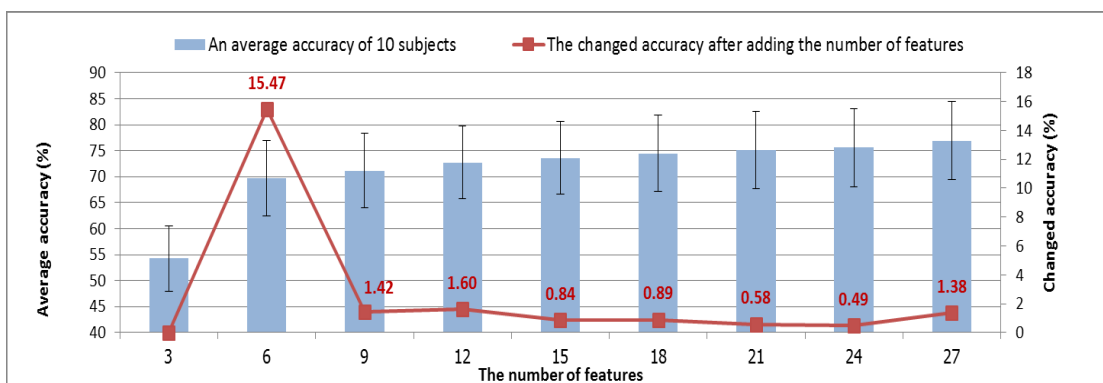
รูปที่ 3.45 เป็นผลความแม่นยำเฉลี่ยซึ่งถูกแสดงในรูปแบบของแผนภูมิแท่ง โดยที่กราฟเส้นตรงแสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น พิจารณาการจำแนกแบบขั้นตอนเดียว รูปที่ 3.45(ก) จะเห็นว่าในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 ความแม่นยำเพิ่มขึ้นมากที่สุด โดยเพิ่มขึ้น 15.47% หลังจากนั้นความแม่นยำจะเพิ่มขึ้นเพียงเล็กน้อย โดยที่ความแม่นยำสูงสุดได้จากการใช้จำนวนลักษณะเด่น 27 ตัว สำหรับการจำแนกแบบสามขั้นตอน รูปที่ 3.45(ข) จะเห็นว่าในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 ความแม่นยำเพิ่มขึ้นมากที่สุดเช่นเดียวกัน โดยเพิ่มขึ้น 6.71% หลังจากนั้นความแม่นยำจะ

เพิ่มขึ้นเพียงเล็กน้อย โดยที่ความแม่นยำสูงสุดได้จากการใช้จำนวนลักษณะเด่น 27 ตัว อย่างไรก็ตาม ถึงแม้ว่าอัตราการเพิ่มขึ้นของความแม่นยำเฉลี่ยของการจำแนกแบบสามขั้นตอนในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 จะต่ำกว่าการจำแนกแบบขั้นตอนเดียว แต่ความแม่นยำเฉลี่ยของการจำแนกแบบสามขั้นตอนก็ยังคงสูงกว่าการจำแนกแบบขั้นตอนเดียว

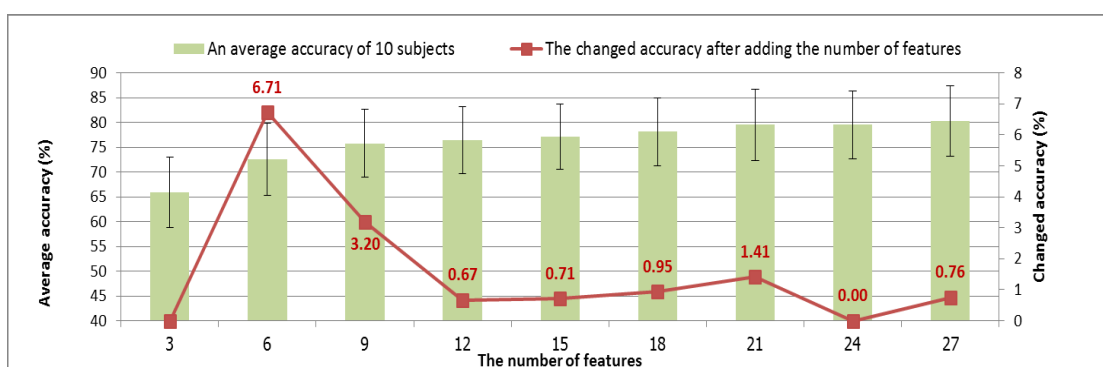
จากการพิจารณาความแม่นยำเฉลี่ยของการจำแนกวรรณยุกต์ 5 เสียงทั้งสองรูปแบบการจำแนก พบว่าการเพิ่มจำนวนลักษณะเด่นสามารถเพิ่มความแม่นยำให้กับระบบ โดยที่การทดลองนี้การเพิ่มลักษณะเด่น 27 ตัวให้ความแม่นยำเฉลี่ยสูงสุดสำหรับทั้งสองรูปแบบการจำแนก (76.93%, 80.27%) สำหรับการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอนตามลำดับ) อย่างไรก็ตามการเพิ่มจำนวนอินพุตให้กับตัวจำแนก ยิ่งจำนวนอินพุตมากขึ้นจะทำให้เวลาในการประมวลผลสัญญาณช้าลง ดังนั้นจำนวนลักษณะเด่นที่เหมาะสมในงานวิจัยนี้ จึงต้องพิจารณาทั้งในแง่ของความแม่นยำและเวลาในการประมวลผล ซึ่งจากรูปที่ 3.45(ก) และ รูปที่ 3.45(ข) แสดงให้เห็นว่าจำนวนลักษณะเด่นที่เหมาะสมสำหรับการจำแนกแบบขั้นตอนเดียวและสามขั้นตอน คือ 12 ตัว และ 9 ตัว ตามลำดับ ทั้งนี้เนื่องจากการเพิ่มจำนวนลักษณะเด่นหลังจากจำนวนดังกล่าวทำให้ความแม่นยำเพิ่มขึ้นเพียงเล็กน้อยเท่านั้น กล่าวคือ สำหรับการจำแนกแบบขั้นตอนเดียว ความแม่นยำที่เพิ่มขึ้นหลังจากเพิ่มจำนวนลักษณะเด่น 12 ตัวเป็นต้นไป มีอัตราการเพิ่มขึ้นอยู่ในช่วงระหว่าง 0.49-1.38 % สำหรับการจำแนกแบบสามขั้นตอน ความแม่นยำที่เพิ่มขึ้นหลังจากเพิ่มจำนวนลักษณะเด่น 9 ตัวเป็นต้นไป มีอัตราการเพิ่มขึ้นอยู่ในช่วงระหว่าง 0-1.41%

รูปที่ 3.46 แสดงการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง ระหว่างรูปแบบการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอน จะเห็นได้ว่าการจำแนกแบบสามขั้นตอนมีความแม่นยำสูงกว่าการจำแนกแบบขั้นตอนเดียว ซึ่งจะแตกต่างกันอย่างเห็นได้ชัดเมื่อพิจารณา ณ ตำแหน่งที่มีการใช้จำนวนลักษณะเด่นเพียง 3 ตัว โดยที่การจำแนกแบบสามขั้นตอนมีความแม่นยำเฉลี่ยสูงกว่า 11.6% เมื่อเพิ่มจำนวนลักษณะเด่นมากขึ้นจะเห็นได้ว่า มีความแตกต่างเพียงเล็กน้อย ระหว่างความแม่นยำเฉลี่ยของทั้งสองรูปแบบ



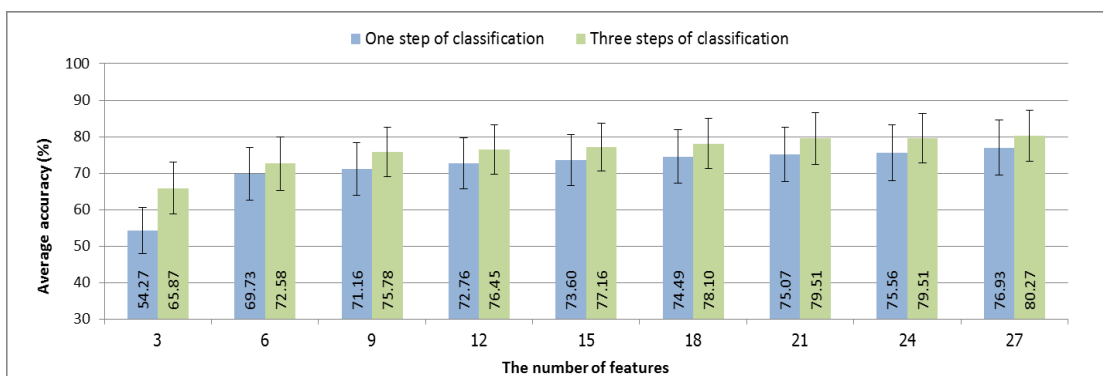


(ก)



(ข)

รูปที่ 3.45 กราฟแกนปฐมภูมิ (ซ้าย) แสดงผลความแม่นยำเฉลี่ยในการจำแนกรรณยุกต์ 5 เสียง เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว กราฟแกนทุติยภูมิ (ขวา) แสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน



รูปที่ 3.46 แผนภูมิแท่งแสดงการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกระหว่างรูปแบบการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอน

ตารางที่ 3.22 ลักษณะเด่นที่ถูกเลือก สำหรับใช้ในรูปแบบการจำแนกแบบสามขั้นตอน

ลำดับ	ลักษณะเด่นสำหรับการจำแนกแบบสามขั้นตอน		
	ขั้นตอนที่ 1	ขั้นตอนที่ 2	ขั้นตอนที่ 3
1	MMAV2	IAV_6	SSC_3
2	MAV_3	WL_6	SSC_4
3	MMAV1_3	SSC_6	SSC_2
4	ZC_3	ZC_6	WAMP_3
5	RMS_3	RMS_6	ZC_2
6	ZC	WAMP_6	RMS_2
7	MMAV1	MMAV1_6	MNF_2
8	ZC_4	SSC	MAV_2
9	WAMP_4	SSC_5	WAMP_2

เปรียบเทียบกับผลการทดลองในหัวข้อ 3.3.2.2 ซึ่งได้นำเสนอเทคนิคการจำแนกแบบสามขั้นตอน โดยออกแบบจากสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครเพียง 1 คน ซึ่งพบว่าเทคนิคนี้สามารถจำแนกวรรณยุกต์ 5 เสียงได้ โดยมีความแม่นยำในการจำแนก 73.11% อย่างไรก็ตาม การทดลองดังกล่าวยังไม่สามารถยืนยันได้ว่า เทคนิคนี้จะมีประสิทธิภาพเพียงพอเมื่อนำไปใช้กับสัญญาณไฟฟ้ากล่อมเนื้ออื่นๆ จากผลการทดลองนี้จะเห็นได้ว่าเทคนิคการจำแนกแบบสามขั้นตอนที่ได้นำเสนอมีประสิทธิภาพเพียงพอในการจำแนกวรรณยุกต์ 5 เสียง ซึ่งได้ปรับปรุงเทคนิคดังกล่าวในส่วนของการคัดเลือกลักษณะเด่น และการเพิ่มจำนวนลักษณะเด่นให้กับระบบ

## 2) คุณลักษณะของลักษณะเด่นที่ถูกเลือก

จากผลการทดลองแสดงให้เห็นว่า การจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN ตามโครงข่ายมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว โดยให้ความแม่นยำสูงสุดเมื่อเพิ่มจำนวนลักษณะเด่น 27 ตัว อย่างไรก็ตามจะพิจารณาลักษณะเด่นเพียง 9 ตัว เนื่องจากการเพิ่มจำนวนลักษณะเด่นมากกว่า 9 ตัวมีผลในการเพิ่มความแม่นยำเพียงเล็กน้อย

ตารางที่ 3.22 แสดงลักษณะเด่นที่ถูกใช้สำหรับการจำแนกแบบสามขั้นตอน ซึ่งจะแบ่งเป็นลักษณะเด่นย่อย 3 ชุด ในแต่ละขั้นตอนจะใช้ลักษณะเด่นแตกต่างกัน จะเห็นว่า

ลักษณะเด่นที่ถูกเลือกส่วนใหญ่เป็นลักษณะเด่นในรูปแบบที่สอง ซึ่งมาจากการคำนวณโดยแบ่งสัญญาณไฟฟ้ากัลลัมเนื้อออกเป็น 6 ส่วนเท่าๆ กัน ดังนั้นจะเห็นได้ว่าลักษณะเด่นในรูปแบบนี้มีประสิทธิภาพในการจำแนกวรรณยุกต์ได้ดี เนื่องจากเป็นลักษณะเด่นที่คำนวณมาจากบางส่วนของสัญญาณ ซึ่งในแต่ละส่วนของสัญญาณมีผลในการจำแนกวรรณยุกต์แตกต่างกัน

พิจารณาลักษณะเด่นที่ใช้ในแต่ละขั้นตอนของการจำแนก ในการจำแนกขั้นตอนที่หนึ่งซึ่งใช้ในการจำแนกวรรณยุกต์เอกและจัตวา ลักษณะเด่นส่วนใหญ่เป็นลักษณะเด่นที่คำนวณจากส่วนที่ 3 และส่วนที่ 4 ของสัญญาณ ในการจำแนกขั้นตอนที่สองซึ่งใช้ในการจำแนกวรรณยุกต์ตรี ลักษณะเด่นส่วนใหญ่เป็นลักษณะเด่นที่คำนวณจากส่วนที่ 6 ของสัญญาณ สำหรับการจำแนกขั้นตอนที่สามซึ่งใช้ในการจำแนกวรรณยุกต์สามัญและโท ลักษณะเด่นส่วนใหญ่เป็นลักษณะเด่นที่คำนวณจากส่วนที่ 2 และส่วนที่ 3 ของสัญญาณ ดังนั้นจะเห็นได้ว่าในแต่ละส่วนของสัญญาณมีผลในการจำแนกวรรณยุกต์แตกต่างกัน กล่าวคือ ส่วนที่ 3 และส่วนที่ 4 ของสัญญาณมีผลในการจำแนกวรรณยุกต์เอกและจัตวาออกจากกลุ่มวรรณยุกต์อื่นๆ ส่วนที่ 6 ของสัญญาณมีผลในการจำแนกวรรณยุกต์ตรีออกจากกลุ่มของวรรณยุกต์โทและสามัญ และ ส่วนที่ 2 และส่วนที่ 3 ของสัญญาณ มีผลในการจำแนกวรรณยุกต์สามัญและโทออกจากกัน

### (ง) สรุปผลการทดลอง

การทดลองนี้เป็นการปรับปรุงเทคนิคการจำแนกแบบสามขั้นตอนซึ่งใช้ตัวจำแนก ANN สามโครงข่ายในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยการใช้ข้อมูลของสัญญาณไฟฟ้ากัลลัมเนื้อของอาสาสมัคร 10 คนในการออกแบบ โดยได้ปรับปรุงในส่วนของการคำนวณลักษณะเด่น คัดเลือกลักษณะเด่น และเพิ่มจำนวนลักษณะเด่นเพื่อเป็นอินพุตให้กับตัวจำแนก ANN นอกจากนี้ทำการเปรียบเทียบรูปแบบการจำแนกสองรูปแบบ คือ การจำแนกแบบสามขั้นตอน และการจำแนกแบบขั้นตอนเดียว จากผลการทดลองพบว่าการจำแนกแบบสามขั้นตอนมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว โดยให้ความแม่นยำสูงสุด 95.56% จากข้อมูลสัญญาณไฟฟ้ากัลลัมเนื้อ S04 และความแม่นยำเฉลี่ย 80.27% จากข้อมูลสัญญาณไฟฟ้ากัลลัมเนื้อทั้งหมด นอกจากนี้เมื่อพิจารณาจากความแม่นยำเฉลี่ย การใช้จำนวนลักษณะเด่น 27 ตัวให้ความแม่นยำสูงสุด อย่างไรก็ตามจำนวนลักษณะเด่นที่เหมาะสม คือ 9 ตัว เนื่องจากการเพิ่มจำนวนลักษณะเด่นมากกว่า 9 ตัว จะเพิ่มประสิทธิภาพในการจำแนกเพียงเล็กน้อย โดยที่การใช้จำนวนลักษณะเด่น 9 ตัว ให้ความแม่นยำเฉลี่ย 75.78%

อย่างไรก็ตาม การทดลองนี้เป็นการคัดเลือกลักษณะเด่นโดยการจัดอันดับลักษณะเด่นตามผลการประเมินด้วยตัวประเมิน RES ซึ่งวิธีนี้ทำให้ลักษณะเด่นที่ใช้อาจมีความสัมพันธ์กัน นั่นก็แสดงว่าข้อมูลมีความซ้ำซ้อน ซึ่งมีผลทำให้ความแม่นยำในการจำแนกไม่เพิ่มขึ้น แต่จะไปเพิ่มพารามิเตอร์เข้าไปในระบบ ทำให้การประมวลผลช้าลง นอกจากนี้ ในการเพิ่มจำนวนลักษณะเด่นครั้งละ 3 ตัว ทำให้ไม่สามารถระบุได้ชัดว่าลักษณะเด่นตัวใดมีผลในการเพิ่มหรือลดประสิทธิภาพของระบบ ดังนั้นงานต่อไปจะปรับปรุงในส่วนของเทคนิคการเลือกลักษณะเด่นโดยใช้เทคนิคการค้นหา เพื่อให้ได้ลักษณะเด่นที่เหมาะสม ซึ่งจะกล่าวถึงในบทที่ 4

### 3.3.3 สรุปผลการออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด) และ การออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง)

**ตารางที่ 3.23** สรุปผลการออกแบบระบบสำหรับการจำแนกคำพูดหลัก (11 คำพูด) และ การออกแบบระบบสำหรับการจำแนกวรรณยุกต์ (5 เสียง)

ระบบการจำแนก	ความแม่นยำเฉลี่ย (%)	เทคนิคที่ใช้	
คำพูดหลัก 11 คำ	ออกเสียง	92.45±3.57	ใช้จำนวนลักษณะเด่น 9 ตัวในแต่ละช่องสัญญาณ โดยใช้ลักษณะเด่นชุดเดียวกันในแต่ละช่องสัญญาณ ซึ่งได้จากการประเมินด้วยตัวประเมิน RES และใช้ตัวจำแนก ANN เป็นตัวจำแนกระบบ
	ไม่ออกเสียง	81.45±5.57	
วรรณยุกต์ 5 เสียง	ออกเสียง	75.78±6.85	ใช้จำนวนลักษณะเด่น 9 ตัวในแต่ละช่องสัญญาณ (ในแต่ละโครงข่าย) ซึ่งได้จากการประเมินด้วยตัวประเมิน RES และใช้รูปแบบการจำแนกแบบสามขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่ายเป็นตัวจำแนกระบบ

## บทที่ 4

### การปรับปรุงประสิทธิภาพของระบบรู้จำ

ในบทนี้จะกล่าวถึง การทดลองและผลทดลองเพื่อปรับปรุงประสิทธิภาพของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ ซึ่งได้ทำการออกแบบมาแล้วในบทที่ 3 โดยจะทำการปรับปรุงใน 2 ส่วน คือ การคำนวณลักษณะเด่นและการคัดเลือกลักษณะเด่นที่เหมาะสม และการเปรียบเทียบตัวจำแนกที่เหมาะสม

#### 4.1 การจำแนกคำพูดหลัก 11 คำ และวรรณยุกต์ 5 เสียง โดยใช้ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน (TDn) ซึ่งใช้ตัวจำแนก HMM ในการจำแนกระบบ

เนื่องจากงานวิจัยนี้ได้ทำงานวิจัยร่วมกับห้องทดลอง Cognitive Systems Lab (CSL) จากประเทศเยอรมัน ซึ่งมีการพัฒนาระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ ตั้งแต่ปี ค.ศ. 2005 จนถึงปัจจุบันซึ่งมีการพัฒนามาอย่างต่อเนื่อง โดยในปี ค.ศ. 2006 ได้พัฒนาระบบรู้จำดังกล่าวให้สามารถรู้จำคำพูดต่อเนื่องขึ้นเป็นครั้งแรก ดังนั้นการทดลองนี้จะนำเทคนิคต่างๆ ที่ถูกใช้ในห้องทดลองนี้ เพื่อทดสอบกับระบบรู้จำการพูดภาษาไทยที่ได้ออกแบบไว้จากบทที่ 3 โดยจะเลือกใช้ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน (Frame-based time domain feature with stacking filter, TDn) ซึ่งจะใช้ตัวจำแนก HMM ในการฝึกสอนและทดสอบระบบ

#### (ก) วัตถุประสงค์

- เพื่อทดสอบประสิทธิภาพของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ เมื่อใช้ลักษณะเด่น TDn และใช้ตัวจำแนก HMM ในการฝึกสอนและทดสอบระบบ

#### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) สำหรับการจำแนกคำพูดหลัก 11 คำ จะใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงและแบบไม่ออกเสียง สำหรับการจำแนกวรรณยุกต์ 5 เสียง จะใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ทั้งหมด ดังตารางที่ 3.3 ซึ่งเป็นข้อมูลของการออกเสียงวรรณยุกต์ โดยการแบ่งข้อมูลชุดฝึกสอนระบบและทดสอบระบบ จะแบ่งข้อมูลเช่นเดียวกับหัวข้อ 3.3.1.2 และ 3.3.2.3

### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อเนื้อทั้งหมดมาคำนวณลักษณะเด่นในแกนเวลาแบบตัวกรองซ็อน หรือ TDn ซึ่งคำนวณโดยใช้สมการที่ (2.13)-(2.23) โดยการคำนวณลักษณะเด่นชนิดนี้จะแบ่งสัญญาณออกเป็นส่วนๆ ที่เท่ากัน แต่ละส่วนเรียกว่า "เฟรม" แล้วคำนวณลักษณะเด่นในแต่ละเฟรม โดยจะคำนวณลักษณะเด่น 5 ชนิด คือ MNL MNH PWL PWH และ ZCH ซึ่งแสดงดังสมการ (2.17)-(2.21) แล้วนำค่าที่ได้ในแต่ละเฟรมมาซ็อนกันเพื่อให้ได้ลักษณะเด่นตัวใหม่ขึ้นมา ซึ่งเป็นการใช้เทคนิคของตัวกรองซ็อนซึ่งแสดงดังสมการ (2.22) ลักษณะเด่นสุดท้ายที่คำนวณได้จะอยู่ในรูปของ TDn ดังสมการ(2.23) โดยที่ n คือ จำนวนการซ็อนกันของลักษณะเด่นที่อยู่ติดกันที่คำนวณได้จากแต่ละเฟรม (stacking width) โดยการทดลองนี้จะปรับจำนวนการซ็อนกัน 5 ระดับ โดยที่  $n=0, 1, 3, 5, 7$  โดยลักษณะเด่นที่ได้จะถูกเรียกว่า TD0 TD1 TD3 TD5 และ TD7 ตามลำดับ รายละเอียดการคำนวณลักษณะเด่น TDn ถูกแสดงในบทที่ 2 หัวข้อ 2.7.1

ในการคำนวณลักษณะเด่นดังกล่าวทำให้จำนวนมิติของเวกเตอร์ลักษณะเด่นเพิ่มขึ้นโดยที่จำนวนมิติของเวกเตอร์ลักษณะเด่นขึ้นอยู่กับจำนวนช่องสัญญาณ และจำนวนการซ็อนกันของลักษณะเด่นที่อยู่ติดกันที่คำนวณได้จากแต่ละเฟรม

2) ลดจำนวนมิติข้อมูลของเวกเตอร์ลักษณะเด่นในแต่ละเฟรมโดยใช้การวิเคราะห์การจำแนกประเภทแบบเชิงเส้น (linear discriminant analysis, LDA) ซึ่งเป็นเทคนิคหนึ่งของการเรียนรู้แบบมีผู้สอน เทคนิคนี้จะหาค่าสูงสุดของอัตราค่าความแปรปรวนระหว่างกลุ่มข้อมูลกับความแปรปรวนภายในกลุ่มข้อมูล นั่นคือทำให้ระยะทางระหว่างกลุ่มมากที่สุด และระยะทางภายในกลุ่มน้อยที่สุด ค่านี้จะยืนยันความสามารถในการแยกกลุ่มสูงสุด [102] โดยจำนวนมิติข้อมูลที่ลดต้องไม่ต่ำกว่า  $C-1$  ซึ่ง  $C$  คือ จำนวนของกลุ่มข้อมูลหรือคลาส สำหรับระบบการจำแนกคำพูดหลัก 11 คำ จะลดจำนวนมิติข้อมูลจากเดิมลงเหลือ 12 22 และ 33 ตามลำดับ สำหรับระบบการจำแนกวรรณยุกต์ LDA จะลดจำนวนมิติข้อมูลจากเดิมลงเหลือ 10 15 20 และ 24 ตามลำดับ สำหรับระบบการจำแนกวรรณยุกต์ LDA จะลดจำนวนมิติข้อมูลจากเดิมลงเหลือ 10 15 20 และ 24 ตามลำดับ ซึ่งตารางที่ 4.1 แสดงจำนวนมิติของเวกเตอร์ลักษณะเด่นทั้งหมดที่ใช้ในการทดลองนี้

3) ฝึกสอนระบบและทดสอบระบบโดยใช้ HMM จะทำการฝึกสอนและทดสอบแยกกันสำหรับข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน

**ตารางที่ 4.1** จำนวนมิติข้อมูลของเวกเตอร์ลักษณะเด่นในแต่ละเฟรม

ลักษณะเด่น	จำนวนมิติของข้อมูล	
	การจำแนก 11 คำพูดหลัก (5 ช่องสัญญาณ)	การจำแนกวรรณยุกต์ 5 เสียง (2 ช่องสัญญาณ)
TD0	$(5 \times 5 \times 1) = 25$	$(2 \times 5 \times 1) = 10$
TD1	$(5 \times 5 \times 3) = 75$	$(2 \times 5 \times 3) = 30$
TD3	$(5 \times 5 \times 7) = 175$	$(2 \times 5 \times 7) = 70$
TD5	$(5 \times 5 \times 11) = 275$	$(2 \times 5 \times 11) = 110$
TD7	$(5 \times 5 \times 15) = 375$	$(2 \times 5 \times 15) = 150$

สำหรับการจำแนกคำพูดหลัก 11 คำ ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคนมีทั้งหมด 550 ข้อมูล (ในแต่ละรูปแบบการออกเสียง) โดยที่ 440 ข้อมูลจะถูกใช้สำหรับการฝึกสอนระบบ และ 110 ข้อมูลจะถูกใช้สำหรับการทดสอบระบบ โดยแต่ละกลุ่มข้อมูลจะประกอบด้วยคำพูด "ศูนย์" ถึง "สิบ" จำนวนเท่าๆ กัน

สำหรับการจำแนกวรรณยุกต์ 5 เสียง ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคนมีทั้งหมด 1,332 ข้อมูล โดยที่ข้อมูล 225 ข้อมูลจะถูกใช้สำหรับทดสอบระบบ ซึ่งเป็นคำพูดชุดที่ 1 8 15 22 และ 23 จากตารางที่ 3.3 ข้อมูลที่เหลือจะถูกใช้สำหรับการฝึกสอนระบบ

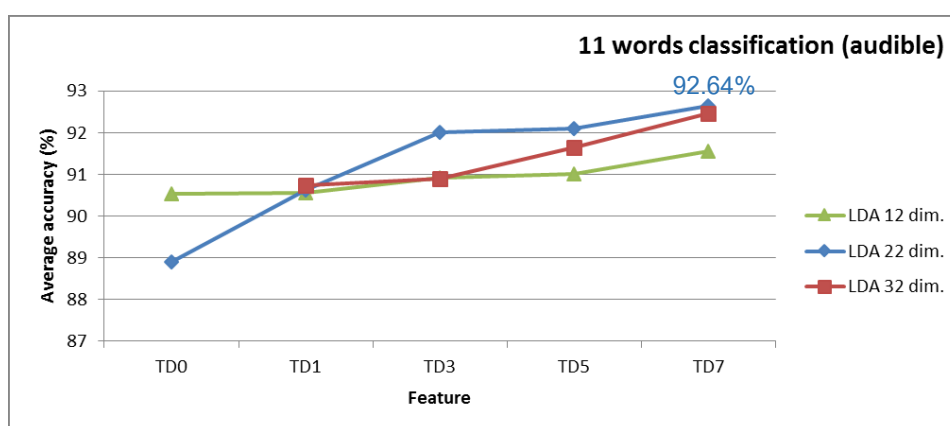
#### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

##### 1) การจำแนกคำพูดหลัก 11 คำพูด

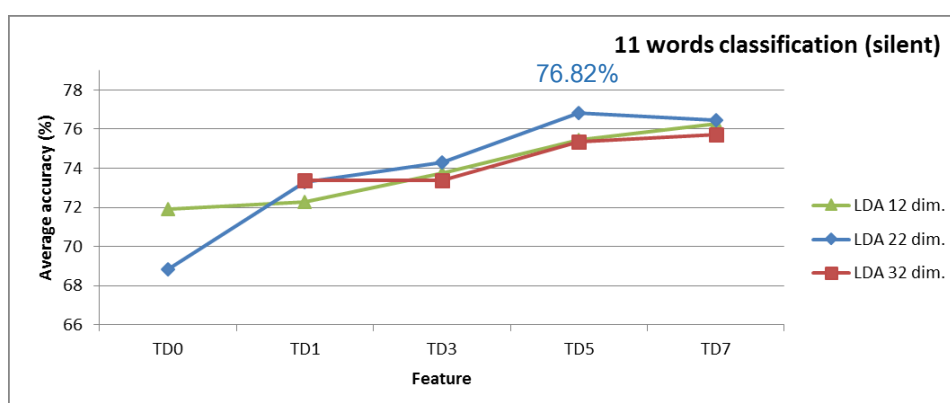
รูปที่ 4.1 แสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ (คำนวณจากสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน) โดยใช้ลักษณะเด่น TDn ซึ่งมีการปรับค่า n 5 ระดับ ( $n=0$ ,  $n=1$ ,  $n=3$ ,  $n=5$  และ  $n=7$ ) ซึ่งถูกเรียกว่า TD0 TD1 TD3 TD5 และ TD7 จากการทดลองพบว่าความแม่นยำในการจำแนกเพิ่มขึ้นเมื่อค่า n เพิ่มขึ้น ซึ่งเป็นไปในทิศทางเดียวกันทั้งรูปแบบการพูดแบบออกเสียง (รูปที่ 4.1 (ก)) และไม่ออกเสียง (รูปที่ 4.1 (ข)) นอกจากนี้เปรียบเทียบความแม่นยำเมื่อมีการลดมิติของข้อมูลโดยใช้ LDA ในจำนวนที่แตกต่างกัน (กล่าวคือ การลดมิติข้อมูลลงมา 3 ระดับ คือ 12 22 และ 32) พบว่าการลดมิติข้อมูลลงมาเท่ากับ 22 ให้ผล

ความแม่นยำที่ดีที่สุดสำหรับ TD3 TD5 และ TD7 เมื่อพิจารณา TD0 การลดมิติข้อมูลลงมาเท่ากับ 12 ให้ผลความแม่นยำที่ดีที่สุด (เมื่อพิจารณา TD0 จะเห็นได้ว่า TD0 ไม่มีข้อมูลของการลดมิติลงมาเท่ากับ 32 เนื่องจากข้อมูลเดิม TD0 มีจำนวนมิติของข้อมูลเท่ากับ 25)

สำหรับผลความแม่นยำเฉลี่ยสูงสุดในการจำแนกคำพูดหลัก 11 คำที่ได้จากการทดลองนี้ คือ 92.64% เมื่อใช้ลักษณะเด่น TD7 และลดมิติข้อมูลลงมาเท่ากับ 22 โดยใช้ LDA ซึ่งเป็นข้อมูลของการพูดแบบออกเสียง สำหรับการพูดแบบไม่ออกเสียง ความแม่นยำเฉลี่ยสูงสุดเท่ากับ 76.82% เมื่อใช้ TD5 และลดมิติข้อมูลลงมาเท่ากับ 22 โดยใช้ LDA



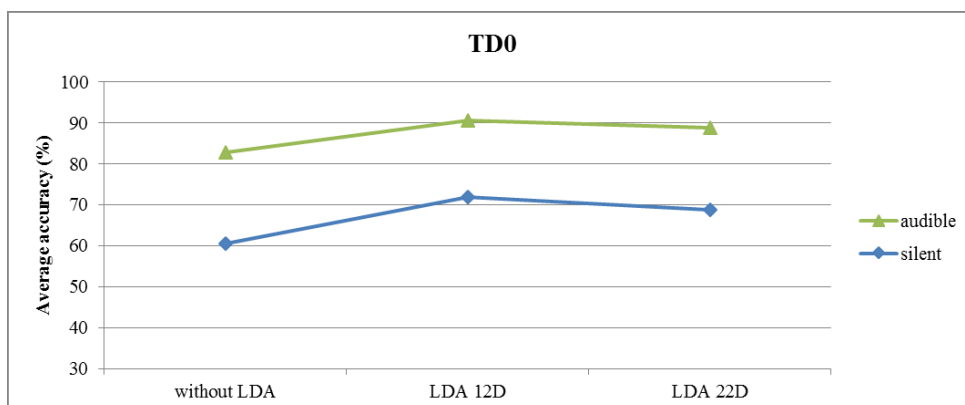
(ก)



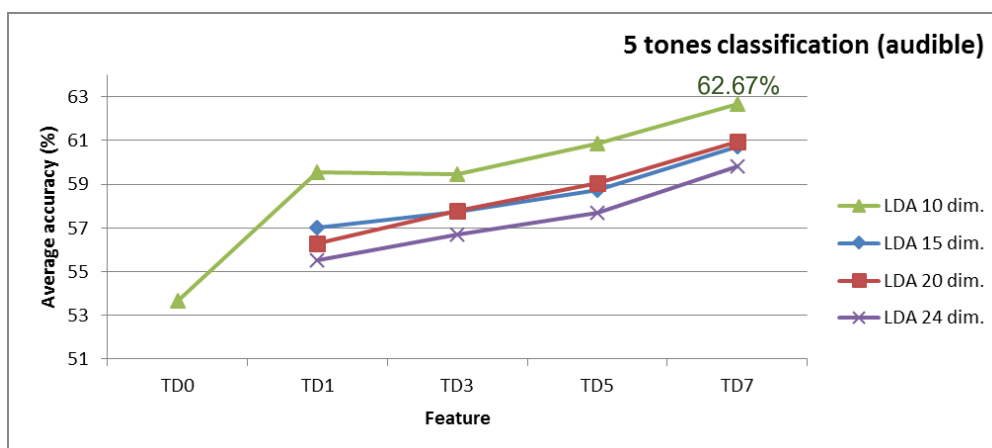
(ข)

รูปที่ 4.1 ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ (ก) การพูดออกเสียง (ข) การพูดไม่ออกเสียง





รูปที่ 4.2 ผลความแม่นยำเฉลี่ยของลักษณะเด่น TD0 โดยเปรียบเทียบความแม่นยำที่ได้ระหว่างการไม่ใช้ LDA ในการลดมิติของข้อมูล และการใช้ LDA ในการลดมิติของข้อมูล



รูปที่ 4.3 ผลความแม่นยำเฉลี่ยในการจำแนกรรณยุกต์ 5 เสียง

รูปที่ 4.2 แสดงผลความแม่นยำเฉลี่ยของลักษณะเด่น TD0 โดยเปรียบเทียบความแม่นยำที่ได้ระหว่างการไม่ใช้ LDA ในการลดมิติของข้อมูล และการใช้ LDA ในการลดมิติของข้อมูล ซึ่งจะเห็นได้ว่าการใช้ LDA ในการลดมิติของข้อมูล นอกจากจะช่วยลดจำนวนข้อมูลที่จะป้อนให้กับตัวจำแนก ยังช่วยเพิ่มความแม่นยำในการจำแนกอีกด้วย โดยเพิ่มความแม่นยำ 7.22% และ 11.27% สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ (เปรียบเทียบระหว่างการไม่ใช้ LDA และการลดมิติของข้อมูลเท่ากับ 12 โดยใช้ LDA)

สาเหตุที่นำลักษณะเด่น TD0 มาพิจารณาเพื่อเปรียบเทียบความแม่นยำระหว่างการไม่ใช้ LDA ในการลดมิติของข้อมูลและ การใช้ LDA ในการลดมิติของข้อมูลเพียงชนิดเดียว เนื่องจากข้อจำกัดของโปรแกรมที่ใช้ในการประมวลผลสัญญาณไม่สามารถป้อนลักษณะเด่นที่มี

จำนวนมิติของข้อมูลเกิน 32 ได้ ซึ่ง TD0 เป็นลักษณะเด่นตัวเดียวที่มีจำนวนมิติของข้อมูลเท่ากับ 25

### 2) การจำแนกวรรณยุกต์ 5 เสียง

รูปที่ 4.3 แสดงผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง (คำนวณจากสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน) โดยใช้ลักษณะเด่น TDn ซึ่งมีการปรับค่า n 5 ระดับ (n=0, n=1, n=3, n=5 และ n=7) ซึ่งถูกเรียกว่า TD0 TD1 TD3 TD5 และ TD7 จากการทดลองพบว่า ความแม่นยำในการจำแนกเพิ่มขึ้นเมื่อค่า n เพิ่มขึ้นเช่นเดียวกับผลการทดลองการจำแนกคำพูดหลัก 11 คำ นอกจากนี้เปรียบเทียบความแม่นยำเมื่อมีการลดมิติของข้อมูลโดยใช้ LDA ในจำนวนที่แตกต่างกัน (กล่าวคือ การลดมิติข้อมูลลงมา 4 ระดับ คือ 10 15 20 และ 24) พบว่า การลดมิติข้อมูลลงมาเท่ากับ 10 ให้ผลความแม่นยำดีที่สุด (เมื่อพิจารณา TD0 จะเห็นได้ว่า TD0 ไม่มีข้อมูลของการลดมิติลงมาเท่ากับ 15 20 และ 24 เนื่องจากข้อมูลเดิม TD0 มีจำนวนมิติของข้อมูลเท่ากับ 10) เมื่อพิจารณาความแม่นยำ

สำหรับผลความแม่นยำเฉลี่ยสูงสุดในการจำแนกวรรณยุกต์ 5 เสียงที่ได้จากการทดลองนี้ คือ 62.67% เมื่อใช้ลักษณะเด่น TD7 และลดมิติข้อมูลลงมาเท่ากับ 10 โดยใช้ LDA

### 3) การเปรียบเทียบประสิทธิภาพกับเทคนิคเดิม

ตารางที่ 4.2 แสดงการเปรียบเทียบผลความแม่นยำเฉลี่ยที่ได้จากการใช้ลักษณะเด่น TDn และ ตัวจำแนก HMM (เทคนิคใหม่) กับผลความแม่นยำเฉลี่ยที่ได้จากการใช้ลักษณะเด่นในแกนเวลา 9 ตัวจากผลการประเมินด้วย RES และใช้ตัวจำแนก ANN (เทคนิคเดิม จากตารางที่ 3.23) พิจารณาระบบการจำแนกคำพูดหลัก 11 คำ สำหรับการพูดแบบออกเสียง ความแม่นยำเฉลี่ยระหว่างสองเทคนิคไม่แตกต่างกัน สำหรับการพูดแบบออกเสียงเทคนิคเดิมให้ความแม่นยำเฉลี่ยสูงกว่า 4.63% สำหรับการจำแนกวรรณยุกต์ 5 เสียง เทคนิคเดิมให้ความแม่นยำสูงกว่า 13.11% จะเห็นได้ว่าเทคนิคใหม่นี้ยังไม่สามารถปรับปรุงความแม่นยำเฉลี่ยของระบบได้ และยังทำให้ความแม่นยำเฉลี่ยในการจำแนกลดลงในส่วนของการพูดแบบไม่ออกเสียงสำหรับการจำแนกคำพูดหลัก 11 คำ และในส่วนของการจำแนกวรรณยุกต์

ตารางที่ 4.2 เปรียบเทียบความแม่นยำเฉลี่ยระหว่างการใช้เทคนิคเดิมและเทคนิคใหม่

ระบบการจำแนก		ความแม่นยำเฉลี่ย (%)	
		เทคนิคเดิม	เทคนิคใหม่
คำพูดหลัก 11 คำ	ออกเสียง	92.45±3.57	92.64±4.24
	ไม่ออกเสียง	81.45±5.57	76.82±4.67
วรรณยุกต์ 5 เสียง	ออกเสียง	75.78±6.85	62.67±8.12

### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการทดสอบประสิทธิภาพของระบบการจำแนกคำพูดหลัก 11 คำ และระบบการจำแนกวรรณยุกต์ 5 เสียง โดยการใช้ลักษณะเด่น TDn ลดมิติของข้อมูลโดยใช้ LDA และใช้ตัวจำแนก HMM ในการฝึกสอนและทดสอบระบบ จากผลการทดลองพบว่า การจำแนกคำพูดหลัก 11 คำ ให้ความแม่นยำเฉลี่ยสูงสุด 92.64% สำหรับการพูดแบบออกเสียง และ 76.82% สำหรับการพูดแบบไม่ออกเสียง การจำแนกวรรณยุกต์ 5 เสียง ให้ความแม่นยำเฉลี่ยสูงสุด 62.67% ซึ่งเมื่อเปรียบเทียบกับเทคนิคที่ได้ออกแบบไว้ในบทที่ 3 (ใช้ลักษณะเด่นในแกนเวลา 9 ตัว จากผลการประเมินด้วย RES และใช้ตัวจำแนก ANN) จะเห็นได้ว่าเทคนิคนี้ยังไม่สามารถปรับปรุงความแม่นยำของระบบได้ ทั้งนี้อาจจะเป็นเพราะว่าเทคนิคดังกล่าวถูกออกแบบมาสำหรับการจำแนกคำพูดแบบต่อเนื่อง นอกจากนี้เทคนิคดังกล่าวอาจจะไม่เหมาะสมกับงานวิจัยนี้ เนื่องจากงานวิจัยนี้ จำนวนคำพูดที่ต้องการจำแนกน้อย และเป็นการจำแนกคำพูดโดดเท่านั้น ซึ่งการใช้เทคนิคนี้ทำให้เพิ่มความซับซ้อนให้กับระบบเกินความจำเป็น

## 4.2 การเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

จากการทดลองที่ผ่านมา งานวิจัยนี้ได้ออกแบบระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งทำการฝึกสอนระบบและทดสอบระบบโดยใช้ตัวจำแนก 2 ชนิด คือ ANN และ HMM ซึ่งจากการทบทวนวรรณกรรมที่เกี่ยวข้องพบว่าเป็นตัวจำแนกที่นิยมใช้สำหรับงานวิจัยทางด้านนี้ และจากผลการทดลองที่ผ่านมาจะเห็นได้ว่าตัวจำแนก ANN ให้ความแม่นยำในการจำแนกดีกว่าตัวจำแนก HMM อย่างไรก็ตาม ข้อเสียของตัวจำแนก ANN คือ ใช้เวลาในการฝึกสอนระบบนาน ดังนั้นการทดลองนี้จะทำการเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด

เพื่อหาตัวจำแนกที่เหมาะสมที่สุดในงานวิจัยนี้ กล่าวคือ ตัวจำแนกที่มีประสิทธิภาพในการจำแนกที่ดี และใช้เวลาในการประมวลผลน้อย

### (ก) วัตถุประสงค์

- เพื่อเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อเนื้อ (การจำแนกคำพูดหลัก 11 คำ และการจำแนกวรรณยุกต์ 5 เสียง)

### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) สำหรับการจำแนกคำพูดหลัก 11 คำ จะใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงและแบบไม่ออกเสียง สำหรับการจำแนกวรรณยุกต์ 5 เสียง จะใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ทั้งหมด ดังตารางที่ 3.3 ซึ่งเป็นข้อมูลของการออกเสียงวรรณยุกต์

การทดลองนี้จะใช้เทคนิคที่ได้ออกแบบไว้ในบทที่ 3 ซึ่งแสดงในตารางที่ 3.23 สำหรับตัวจำแนกจะทำการฝึกสอนระบบและทดสอบระบบโดยใช้ตัวจำแนก 8 ชนิด ซึ่งรายละเอียดการทดลองมีดังนี้

#### ขั้นตอนการทดลอง

1) นำสัญญาณข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อเนื้อทั้งหมดมาคำนวณคุณลักษณะเด่นเพื่อใช้เป็นอินพุตให้กับตัวจำแนก สำหรับระบบการจำแนกคำพูดหลัก 11 คำ จะคำนวณคุณลักษณะเด่นในตารางที่ 3.15 สำหรับการจำแนกวรรณยุกต์ 5 เสียงจะคำนวณคุณลักษณะเด่นในตารางที่ 3.22

2) ฝึกสอนระบบและทดสอบระบบโดยใช้ตัวจำแนก 8 ชนิด ซึ่งแสดงดังตารางที่ 4.3 สำหรับการจำแนกวรรณยุกต์ 5 เสียง จะใช้เทคนิคการจำแนกแบบสามขั้นตอนตามที่ได้ออกแบบไว้ในบทที่ 3

3) วัดประสิทธิภาพในการจำแนกของระบบโดยใช้เทคนิคการตรวจสอบไขว้ (K-fold cross validation) โดยการแบ่งชุดข้อมูลออกเป็น K กลุ่ม สำหรับการทดลองนี้ K=10 (ซึ่งเป็นค่าที่นิยมใช้ในหลายงานวิจัย) ซึ่งเป็นการเลือกข้อมูลสำหรับการฝึกสอนและทดสอบระบบโดยแบ่ง

ชุดข้อมูลออกเป็น 10 กลุ่ม และทำการทดลอง 10 ครั้ง โดยการทดลองครั้งที่ 1 จะใช้ข้อมูลชุดที่ 1 เป็นข้อมูลสำหรับทดสอบระบบ และข้อมูลชุดที่ 2-9 เป็นข้อมูลสำหรับฝึกสอนระบบ ในการทดลองครั้งที่ 2 จะใช้ชุดข้อมูลชุดที่ 2 เป็นข้อมูลสำหรับทดสอบระบบ และข้อมูลที่เหลือถูกใช้สำหรับฝึกสอนระบบ ทำการทดลองจนครบ 10 ครั้ง จากนั้นนำผลความแม่นยำในการจำแนกที่ได้ทั้ง 10 การทดลองมาหาค่าเฉลี่ย ซึ่งการใช้วิธีนี้ทำให้ข้อมูลทุกตัวถูกใช้สำหรับฝึกสอนและทดสอบระบบ จากการทดลองที่ผ่านมาใช้เทคนิคการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (percentage split หรือ hand out method) ซึ่งจะแบ่งข้อมูลทั้งหมดออกเป็น 2 กลุ่มตามอัตราส่วนที่กำหนดให้ ข้อมูลกลุ่มแรกจะถูกใช้สำหรับการฝึกสอนระบบและข้อมูลที่เหลือจะถูกใช้สำหรับการทดสอบระบบ ซึ่งเป็นวิธีที่ง่ายและใช้เวลาในการคำนวณน้อย แต่ข้อเสียคือ ค่าความแม่นยำที่ได้มีความแปรปรวนสูง เนื่องจากข้อมูลทุกตัวไม่ได้ถูกใช้ในกระบวนการฝึกสอนและทดสอบระบบ ความแม่นยำที่วัดได้ ได้จากการทดสอบประสิทธิภาพของข้อมูลเพียงชุดเดียว ดังนั้นการทดลองนี้จึงเลือกใช้วิธีการตรวจสอบไขว้สำหรับการแบ่งข้อมูลเพื่อวัดความแม่นยำในการจำแนกของระบบ เพื่อลดความแปรปรวนของค่าความแม่นยำในการจำแนก

**ตารางที่ 4.3** ตัวจำแนกทั้งหมดที่ใช้ในการทดลองนี้

ตัวจำแนก (classifier)		
	ชื่อเต็ม	ชื่อย่อ
C1	Nearest Mean	NM
C2	K-Nearest Neighbor	KNN
C3	Linear Bayes Normal	LBN
C4	Logistic Linear	LOGL
C5	Quadratic Bayes Normal	QBN
C6	Fisher's Least Square Linear Discriminant	FLDA
C7	Support Vector Machine	SVM
C8	Artificial Neural Network	ANN

**ตารางที่ 4.4** เปรียบเทียบผลความแม่นยำเฉลี่ยระหว่างเทคนิคการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละและเทคนิคการตรวจสอบไขว้

ระบบการจำแนก	ความแม่นยำเฉลี่ย(10 subjects)(%)	
	Percentage spit	10-fold cross validation
คำพูดหลัก 11 คำ	ออกเสียง	92.45±3.57
	ไม่ออกเสียง	81.45±5.57
วรรณยุกต์ 5 เสียง	ออกเสียง (one step)	71.16±7.66
	ออกเสียง (three steps)	75.78±6.85

#### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

1) เปรียบเทียบความแม่นยำในการจำแนกระหว่างการใช้เทคนิคการตรวจสอบไขว้และการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ

ตารางที่ 4.4 แสดงผลความแม่นยำเฉลี่ยของระบบการจำแนกคำพูดหลัก 11 คำ และวรรณยุกต์ 5 เสียง ซึ่งใช้ตัวจำแนก ANN ในการฝึกสอนและทดสอบระบบ โดยเปรียบเทียบระหว่างการใช้เทคนิคแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (ผลการทดลองที่ผ่านมาในตารางที่ 3.23) และการตรวจสอบไขว้ (ผลการทดลองนี้) เพื่อวัดประสิทธิภาพในการจำแนกของระบบ จะเห็นได้ว่าเทคนิคการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละให้ความแม่นยำสูงกว่าการแบ่งข้อมูลแบบตรวจสอบไขว้ เนื่องจากการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละจะใช้ข้อมูลเพียง 1 ชุดในการทดสอบระบบ (เป็นข้อมูลที่ไม่ได้ถูกใช้ในกระบวนการฝึกสอนระบบ) ซึ่งข้อมูลที่ถูกละเลือกสำหรับใช้เป็นข้อมูลชุดทดสอบอาจจะเป็นข้อมูลชุดที่ดีที่สุดซึ่งทำให้ผลการจำแนกดีที่สุด อย่างไรก็ตามถ้า นำชุดทดสอบชุดอื่นมาทดสอบผลการจำแนกอาจจะเปลี่ยนแปลงได้ ซึ่งอาจจะให้ผลการจำแนกสูงหรือต่ำกว่านี้มาก ดังนั้นข้อเสียของวิธีนี้คือ มีความแปรปรวนสูง พิจารณาผลการจำแนกเมื่อใช้เทคนิคการแบ่งข้อมูลแบบตรวจสอบไขว้ ถึงแม้ว่าผลการจำแนกจะต่ำกว่าเล็กน้อย แต่ความน่าเชื่อถือสูงกว่า มีความแปรปรวนน้อยกว่า เนื่องจากข้อมูลทุกตัวถูกใช้สำหรับฝึกสอนและทดสอบระบบ

พิจารณาการจำแนกวรรณยุกต์ 5 เสียง เดิมเมื่อวัดประสิทธิภาพโดยใช้การแบ่งข้อมูลแบบร้อยละ จะเห็นว่ากรจำแนกแบบสามขั้นตอนมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว 4.62% แต่เมื่อวัดประสิทธิภาพโดยใช้การแบ่งข้อมูลแบบตรวจสอบไขว้ พบว่าการจำแนกแบบขั้นตอนเดียวมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว 2.22% ทั้งนี้การวัดประสิทธิภาพโดยการแบ่งข้อมูลแบบตรวจสอบไขว้มีความน่าเชื่อถือมากกว่า เนื่องจากข้อมูลทุกตัวถูกใช้สำหรับการฝึกสอนและทดสอบระบบ ดังนั้นในส่วนนี้จึงสรุปได้ว่าการจำแนกแบบขั้นตอนเดียวดีกว่าสามขั้นตอน

## 2) เปรียบเทียบประสิทธิภาพในการจำแนกของตัวจำแนก 8 ชนิด

รูปที่ 4.4 แสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ ระหว่างการใช้ตัวจำแนก 8 ชนิด โดยที่ (ก) เป็นการพูดแบบออกเสียง และ (ข) เป็นการพูดแบบไม่ออกเสียง จะเห็นได้ว่าตัวจำแนกที่ให้ผลการจำแนกสูงสุดทั้งสองรูปแบบการออกเสียงแตกต่างกัน สำหรับการพูดแบบออกเสียง ตัวจำแนก FLDA ให้ผลความแม่นยำเฉลี่ยสูงสุด (90.01%) ซึ่งให้ผลการจำแนกสูงกว่าตัวจำแนก ANN 2.19% สำหรับการพูดแบบไม่ออกเสียง ตัวจำแนก LBN ให้ผลความแม่นยำเฉลี่ยสูงสุด (79.18%) ซึ่งให้ผลการจำแนกสูงกว่าตัวจำแนก ANN 1.95% จะเห็นได้ว่าสำหรับการจำแนกคำพูดหลัก 11 คำ ตัวจำแนกที่มีประสิทธิภาพสูงกว่า ANN คือ FLDA และ LBN นอกจากนี้ตัวจำแนกทั้งสองตัวนี้ใช้เวลาในการฝึกสอนและทดสอบระบบน้อยกว่า ANN

รูปที่ 4.5 แสดงผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง โดยใช้ (ก) เทคนิคการจำแนกแบบขั้นตอนเดียว และ (ข) เทคนิคการจำแนกแบบสามขั้นตอน โดยเปรียบเทียบประสิทธิภาพของการใช้ตัวจำแนก 8 ชนิด จะเห็นได้ว่าตัวจำแนก LOGL มีประสิทธิภาพในการจำแนกดีที่สุดสำหรับการจำแนกวรรณยุกต์ทั้งสองรูปแบบ โดยให้ความแม่นยำเฉลี่ย 68.36% สำหรับรูปแบบการจำแนกแบบขั้นตอนเดียว และ 67.69% สำหรับการจำแนกแบบสามขั้นตอน ซึ่งประสิทธิภาพระหว่างสองรูปแบบการจำแนกไม่แตกต่างกันมากนักเมื่อสำหรับตัวจำแนก LOGL

อย่างไรก็ตามเมื่อพิจารณาตัวจำแนกที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรก สำหรับการจำแนกคำพูดหลัก 11 คำ ทั้งรูปแบบการพูดแบบออกเสียงและไม่ออกเสียง และการจำแนกวรรณยุกต์ 5 เสียง ซึ่งแสดงดังตารางที่ 4.5 จะเห็นได้ว่าความแม่นยำแตกต่างกันเล็กน้อย ดังนั้นในการทดลองนี้ยังไม่สามารถสรุปตัวจำแนกที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ได้ ซึ่งจะต้องทำการทดสอบประสิทธิภาพของตัวจำแนกเหล่านี้อีกครั้ง โดยพิจารณาเฉพาะตัวจำแนก

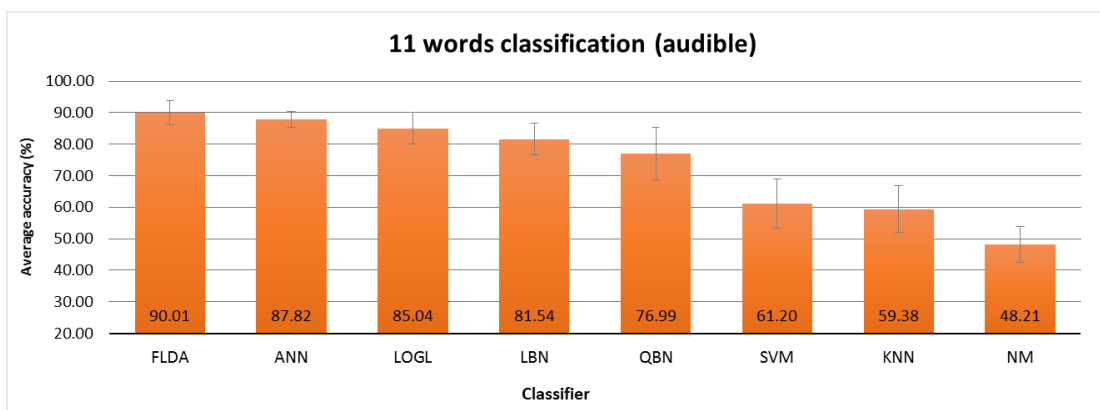
ที่แสดงตารางที่ 4.5 ซึ่งพบว่ามีทั้งหมด 4 ชนิดคือ FLDA ANN LBN และ LOGL ซึ่งจะทำการทดสอบในการทดลองต่อไป

พิจารณาเวลาในการประมวลผลสำหรับการใช้ตัวจำแนกแต่ละชนิด แสดงดังรูปที่ 4.6 ซึ่งจะพิจารณาเฉพาะเวลาในการฝึกสอนระบบสำหรับการจำแนกคำพูดหลัก 11 คำ แสดงให้เห็นว่าตัวจำแนกที่ใช้เวลาในการประมวลผลมากที่สุด คือ SVM รองลงมา คือ ANN ตัวจำแนกชนิดอื่นนอกจากนั้นใช้เวลาในการประมวลผลใกล้เคียงกัน โดยจะเห็นได้ว่าตัวจำแนก SVM และ ANN ใช้เวลาในการประมวลผลมากกว่าตัวจำแนกชนิดอื่นอย่างมาก จากการทดลองนี้จะเห็นได้ว่า LBN FLDA และ LOGL มีประสิทธิภาพในการจำแนกใกล้เคียงกับ ANN แต่เวลาที่ใช้ในการประมวลผลน้อยกว่ามาก ดังนั้นตัวจำแนก 3 ชนิดนี้น่าจะมีความเหมาะสมมากกว่า ซึ่งจะทำการทดสอบประสิทธิภาพของตัวจำแนก 4 ชนิดนี้อีกครั้งในการทดลองต่อไป

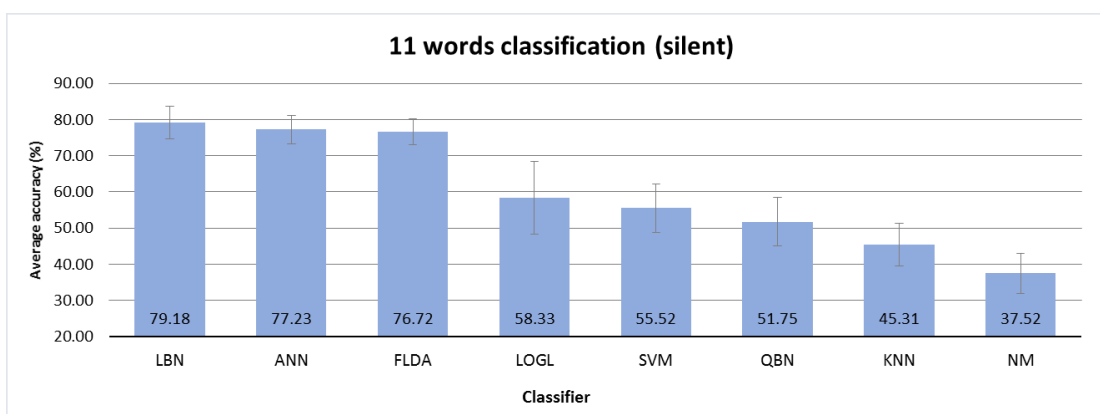
**ตารางที่ 4.5** สรุปตัวจำแนกที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรก

	การจำแนกคำพูดหลัก 11 คำ				การจำแนกรรณยุกต์ 5 เสียง	
	ออกเสียง		ไม่ออกเสียง			
ตัวจำแนก	ความแม่นยำเฉลี่ย (%)	ตัวจำแนก	ความแม่นยำเฉลี่ย (%)	ตัวจำแนก	ความแม่นยำเฉลี่ย (%)	
					one step	three steps
FLDA	90.01±3.84	LBN	79.18±4.49	LOGL	68.36±7.38	67.69±7.43
ANN	87.82±2.49	ANN	77.23±3.84	ANN	68.16±7.66	65.94±7.80
LOGL	85.04±5.04	FLDA	76.72±3.60	LBN	67.17±7.16	65.69±7.23



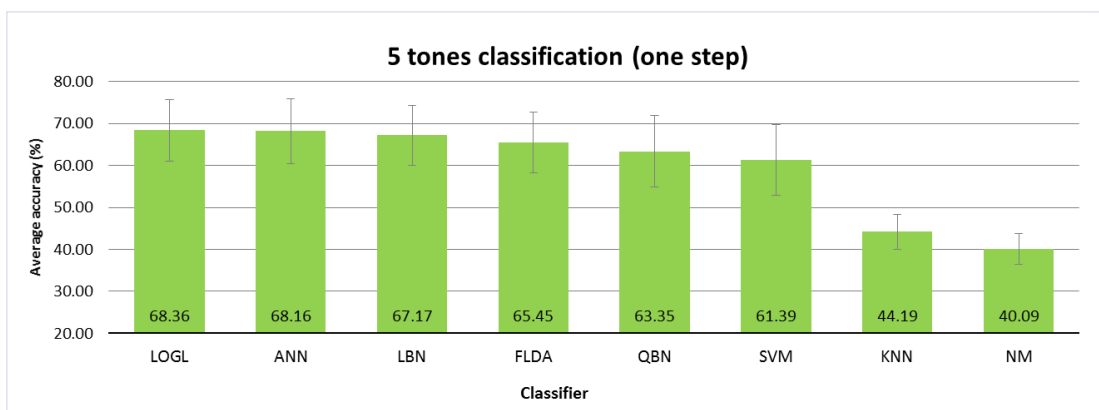


(ก)

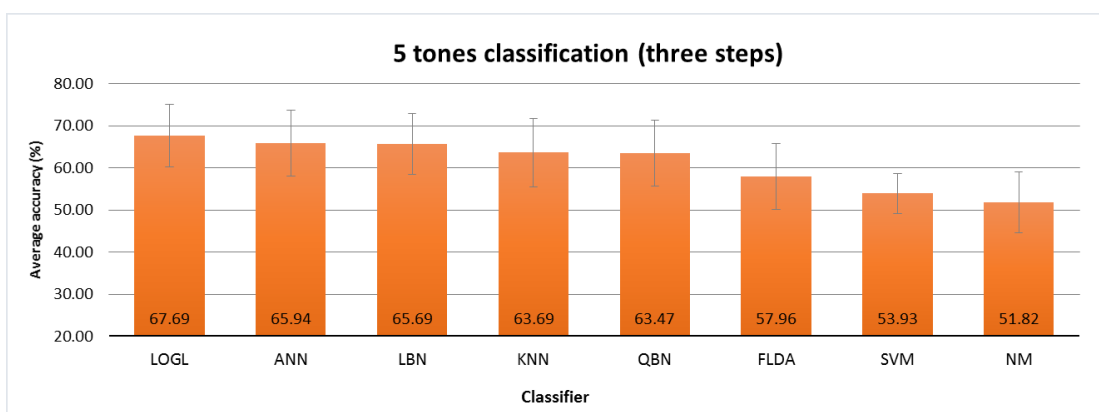


(ข)

รูปที่ 4.4 เปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ ระหว่างการใช้ตัว  
จำแนก 8 ชนิด (ก) การพูดแบบออกเสียง (ข) การพูดแบบไม่ออกเสียง

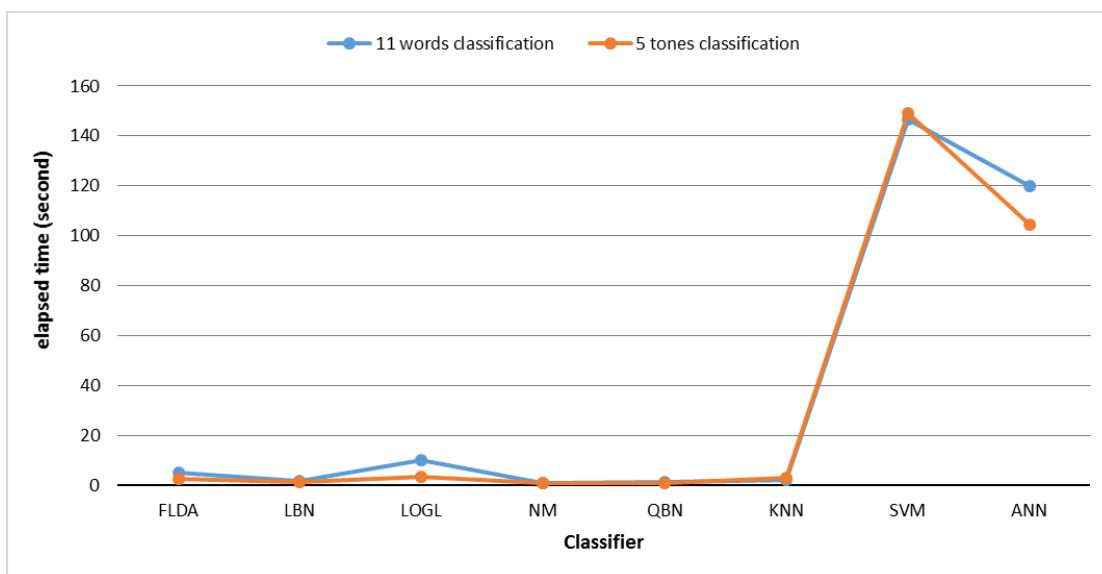


(ก)



(ข)

รูปที่ 4.5 เปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกรรณยุกต์ 5 เสียง ระหว่างการใช้ตัวจำแนก 8 ชนิด (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน



รูปที่ 4.6 เวลาในการประมวลผล (elapsed time) สำหรับการจำแนกคำพูดหลัก 11 คำ ของตัวจำแนกแต่ละตัว

#### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ ซึ่งแบ่งออกเป็น 2 ส่วน คือ การจำแนกคำพูดหลัก 11 คำ และการจำแนกวรรณยุกต์ 5 เสียง โดยใช้เทคนิคตามที่ได้ออกแบบไว้ในบทที่ 3 ซึ่งใช้ลักษณะเด่น 9 ตัว สำหรับแต่ละช่องสัญญาณเป็นอินพุตให้กับตัวจำแนก และเปรียบเทียบระหว่างการใช้เทคนิคการจำแนกแบบขั้นตอนเดียวและสามขั้นตอนในการจำแนกวรรณยุกต์ 5 เสียง ซึ่งการทดลองนี้จะวัดประสิทธิภาพในการจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 สำหรับการแบ่งชุดข้อมูลสำหรับฝึกสอนและทดสอบระบบ โดยจะทดสอบประสิทธิภาพในการจำแนกของตัวจำแนก 12 ชนิด จากผลการทดลองพบว่า ตัวจำแนกที่ให้ผลการจำแนกที่ดีที่สุดสำหรับการจำแนกคำพูดหลัก 11 คำ คือ FLDA (ออกเสียง) และ LBN (ไม่ออกเสียง) สำหรับการจำแนกวรรณยุกต์ 5 เสียงพบว่าตัวจำแนก LOGL ให้ผลการจำแนกที่ดีที่สุด จะเห็นได้ว่าตัวจำแนกดังกล่าวให้ผลการจำแนกดีกว่าตัวจำแนก ANN และใช้เวลาในการประมวลผลน้อยกว่า นอกจากนี้ยังพบว่าเมื่อวัดประสิทธิภาพในการจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง พบว่าประสิทธิภาพในการจำแนกแบบขั้นตอนเดียวใกล้เคียงกับการจำแนกแบบสามขั้นตอน ดังนั้นการใช้เทคนิคการจำแนกแบบขั้นตอนเดียวมีความเหมาะสมมากกว่าเนื่องจากเป็นเทคนิคที่ง่าย ไม่ซับซ้อนและใช้เวลาในการประมวลผลน้อยกว่า

อย่างไรก็ตามเมื่อพิจารณาตัวจำแนกที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรก สำหรับการจำแนกคำพูดหลัก 11 คำ ทั้งรูปแบบการพูดแบบออกเสียงและไม่ออกเสียง และการจำแนกวรรณยุกต์ 5 เสียง ความแม่นยำเฉลี่ยแตกต่างกันเล็กน้อย ดังนั้นในการทดลองนี้ยังไม่สามารถสรุปตัวจำแนกที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ได้ ซึ่งจะต้องทำการทดสอบประสิทธิภาพของตัวจำแนกเหล่านี้อีกครั้ง โดยพิจารณาเฉพาะตัวจำแนกที่แสดงตารางที่ 4.5 ซึ่งพบว่าทั้งหมด 4 ชนิดคือ FLDA ANN LBN และ LOGL ซึ่งจะทำการทดสอบในการทดลองต่อไป

#### 4.3 การเปรียบเทียบตัวประเมินลักษณะเด่น 14 ชนิด สำหรับระบบรู้จำการพูดภาษาไทย โดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

จากการทดลองที่ผ่านมา งานวิจัยนี้ได้ออกแบบระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งทำการคำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและประเมินลักษณะเด่นโดยใช้ตัวประเมิน RES และเลือกลักษณะเด่นที่ดีที่สุด 9 อันดับแรกโดยพิจารณาจากผลการประเมินด้วย RES เพื่อเป็นอินพุตให้กับตัวจำแนก ซึ่งเป็นเทคนิคการเลือกลักษณะเด่นแบบ FR ซึ่งในการทดลองที่ผ่านมาประเมินลักษณะเด่นโดยใช้ตัวประเมินเพียงชนิดเดียว ดังนั้นในการทดลองนี้จะทำการเปรียบเทียบตัวประเมินลักษณะเด่น 14 ชนิด เพื่อหาตัวประเมินลักษณะเด่นที่เหมาะสมที่สุดสำหรับงานวิจัยนี้

##### (ก) วัตถุประสงค์

- เพื่อเปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ (การจำแนกคำพูดหลัก 11 คำ)
- เพื่อเปรียบเทียบประสิทธิภาพของตัวจำแนก 4 ชนิด

##### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) โดยจะใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พุด "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงและไม่ออกเสียง ซึ่งเป็นข้อมูลสำหรับการจำแนกคำพูดหลัก 11 คำ

### ขั้นตอนการทดลอง

1) นำข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมดมาคำนวณลักษณะเด่น 22 ชนิด ซึ่งแสดงดังตารางที่ 4.6

2) ทำการประเมินลักษณะเด่นแต่ละตัวโดยใช้ตัวประเมิน 14 ชนิด ซึ่งแบ่งเป็น 2 ประเภท

ตัวประเมินกลุ่มแรก เป็นตัวประเมินที่ไม่ขึ้นกับตัวจำแนกซึ่งถูกใช้ในเทคนิคการเลือกลักษณะเด่นแบบฟิลเตอร์ ตัวประเมินกลุ่มนี้จะใช้คุณลักษณะภายในที่มีอยู่ของข้อมูลในการประเมิน ได้แก่ การวัดความสามารถในการแบ่งกลุ่มของข้อมูล โดยวัดระยะทางระหว่างกลุ่มข้อมูล และภายในกลุ่มข้อมูล การวัดเกณฑ์ความรู้ ซึ่งใช้หลักการของเอนโทรปีที่วัดระดับความเจือปนของกลุ่มข้อมูล การวัดความสัมพันธ์ของข้อมูล เป็นต้น

ตัวประเมินกลุ่มที่สอง เป็นตัวประเมินที่ขึ้นอยู่กับตัวจำแนกซึ่งถูกใช้ในเทคนิคการเลือกลักษณะเด่นแบบแรปเปอร์ ตัวประเมินกลุ่มนี้จะใช้ผลความแม่นยำในการจำแนกของตัวจำแนกในการประเมิน ซึ่งในการทดลองนี้จะใช้ผลความแม่นยำของตัวจำแนก 4 ชนิด คือ FLDA ANN LBN และ LOGL ซึ่งเป็นตัวจำแนกที่มีผลการจำแนกที่ดีที่สุดจากการทดลองก่อนหน้านี้

สำหรับการทดลองนี้จะกำหนดชื่อเรียกของตัวประเมินแต่ละกลุ่ม โดยกำหนดให้ตัวประเมินกลุ่มแรกขึ้นต้นด้วยตัวอักษร "F" เช่น F\_RES เป็นต้น ตัวประเมินกลุ่มที่สองขึ้นต้นด้วยตัวอักษร "W" เช่น W\_ANN เป็นต้น ตัวประเมินทั้ง 14 ชนิดถูกแสดงดังตารางที่ 4.7 ตัวประเมิน C1-C10 เป็นตัวประเมินกลุ่มแรก C11-C14 เป็นตัวประเมินกลุ่มที่สอง

3) เลือกลักษณะเด่นเพื่อใช้เป็นอินพุตให้กับตัวจำแนก โดยทำการเลือกลักษณะเด่นสำหรับแต่ละตัวประเมิน โดยมีขั้นตอนในการเลือกลักษณะเด่น ดังต่อไปนี้ (รูปแบบเดียวกับการเลือกลักษณะเด่นที่ออกแบบไว้ในบทที่ 3)

3.1) เลือกลักษณะเด่นที่มีผลการประเมินด้วยตัวประเมินนั้นๆ สูงที่สุด 5 อันดับแรกสำหรับสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน (10 คน)

3.2) พิจารณาลักษณะเด่นทั้งหมดที่ถูกเลือกจากหัวข้อที่ 3.1) จากนั้น หาความถี่ของลักษณะเด่นแต่ละตัว ซึ่งความถี่สูงสุดเท่ากับ 10 นั่นคือ ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน มีลักษณะเด่นตัวนั้นปรากฏอยู่ในกลุ่มของลักษณะเด่นที่มี ผลการประเมินสูงสุด 10 อันดับแรกจากหัวข้อที่ 3.1)

3.3) เรียงลำดับลักษณะเด่นจากความถี่สูงสุดไปหาความถี่ต่ำสุด อย่างไรก็ตาม ในกรณีที่ความถี่ของลักษณะเด่นเท่ากัน จะเรียงลำดับโดยพิจารณาค่าเฉลี่ยของตัวประเมน

3.4) เลือกลักษณะเด่นที่ดีที่สุด 9 อันดับแรกจากหัวข้อที่ 3.3) เพื่อเป็นอินพุตให้กับตัวจำแนก

4) ฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก 4 ชนิด FLDA ANN LBN และ LOGL วัดประสิทธิภาพในการจำแนกของตัวจำแนกใช้เทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 เปรียบเทียบประสิทธิภาพตัวประเมนทั้ง 14 ชนิด และ ตัวจำแนก 4 ชนิด

ในกระบวนการฝึกสอนและทดสอบระบบ ชุดข้อมูลสำหรับฝึกสอนและทดสอบระบบต้องเป็นชุดข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครคนเดียวกัน

#### ตารางที่ 4.6 ลักษณะเด่น 22 ชนิด

ลักษณะเด่น		ลักษณะเด่น		ลักษณะเด่น		ลักษณะเด่น	
F1	RMS	F7	MN	F13	MNF	F19	MNH
F2	VAR	F8	IAV	F14	SM1	F20	PWL
F3	WL	F9	MAV	F15	SM2	F21	PWH
F4	WAMP	F10	MMAV1	F16	SM3	F22	ZCH
F5	ZC	F11	MMAV2	F17	SM5		
F6	SSC	F12	MDF	F18	MNL		

ตารางที่ 4.7 ตัวประเมินลักษณะเด่น 14 ชนิด

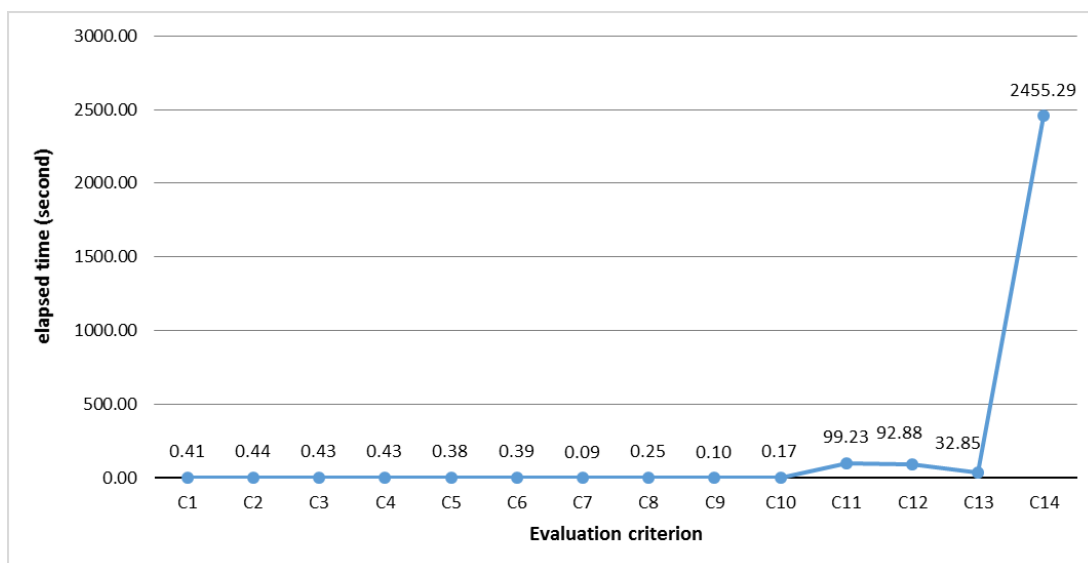
ตัวประเมินลักษณะเด่น 14 ชนิด			
ประเภท	สัญลักษณ์	ตัวย่อ	ชื่อเต็ม
ไม่ขึ้นอยู่กับตัวจำแนก (filter)	C1	F_RES	เกณฑ์อัตราส่วนระหว่างระยะทางแบบยูคลิดและค่าเบี่ยงเบนมาตรฐาน
	C2	F_in-in	เกณฑ์ระยะทางภายในกลุ่มและระหว่างกลุ่มข้อมูล
	C3	F_Maha-s	เกณฑ์ผลรวมของระยะทางมาฮาลานอบิสแบบประมาณ
	C4	F_Maha-m	เกณฑ์ค่าต่ำสุดของระยะทางมาฮาลานอบิสแบบประมาณ
	C5	F_Eucl-s	เกณฑ์ผลรวมของระยะทางแบบยูคลิดกำลังสอง
	C6	F_Eucl-m	เกณฑ์ค่าต่ำสุดของระยะทางแบบยูคลิดกำลังสอง
	C7	F_CorrP	เกณฑ์ค่าสหสัมพันธ์แบบเชิงเส้นของเพียร์สัน
	C8	F_CorrK	เกณฑ์ค่าสหสัมพันธ์แบบเคนดอลล์
	C9	F_CorrS	เกณฑ์ค่าสหสัมพันธ์แบบสเปียร์แมน
	C10	F_IG	เกณฑ์ค่าเอนโทรปี
ขึ้นอยู่กับตัวจำแนก (wrapper)	C11	W_FLDA	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกการวิเคราะห์การจำแนกประเภทแบบเชิงเส้นของฟิชเชอร์
	C12	W_LOGL	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกเชิงเส้นแบบลอจิสติก
	C13	W_LBN	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกเชิงเส้นแบบเบย์นอร์มัล
	C14	W_ANN	เกณฑ์การประเมินซึ่งขึ้นอยู่กับผลการจำแนกของตัวจำแนกโครงข่ายประสาทเทียม

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

รูปที่ 4.8 แสดงผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูด สำหรับการพูดแบบออกเสียง ซึ่งเปรียบเทียบประสิทธิภาพในการจำแนกของตัวประเมินลักษณะเด่น 14 ชนิด (C1-C14) เมื่อทดสอบด้วยตัวจำแนก 4 ชนิด (ก) FLDA (ข) LOGL (ค) LBN (ง) ANN จะเห็นได้ว่า ตัวประเมินที่ให้ผลการจำแนกที่ดีที่สุดสำหรับตัวจำแนกแต่ละตัวแตกต่างกัน กล่าวคือ ตัวประเมิน C11 (W\_FLDA) ให้ผลการจำแนกที่ดีที่สุดเมื่อทดสอบกับตัวจำแนก FLDA และ LBN (90.42% และ 93.25%) ตัวประเมิน C1 (F\_RES) ให้ผลการจำแนกที่ดีที่สุดเมื่อทดสอบกับตัวจำแนก ANN และ LOGL (87.82% และ 85.04%) สำหรับ รูปที่ 4.9 เป็นผลการทดลองสำหรับการพูดแบบไม่ออกเสียงซึ่งให้ผลเช่นเดียวกับการพูดแบบออกเสียง กล่าวคือ ตัวประเมินที่ให้ผลการจำแนกที่ดีที่สุดสำหรับตัวจำแนกแต่ละตัวแตกต่างกัน โดยที่ตัวประเมิน C14 (W\_ANN) ให้ผลการจำแนกที่ดีที่สุดเมื่อทดสอบกับตัวจำแนก ANN และ FLDA (77.35% และ 78.10%) ตัวประเมิน C11 (W\_FLDA) ให้ผลการจำแนกที่ดีที่สุดเมื่อทดสอบกับตัวจำแนก LBN (80.12%) และตัวประเมิน C2 (F\_in-in) ให้ผลการจำแนกที่ดีที่สุดเมื่อทดสอบกับตัวจำแนก LOGL (64.02%)

พิจารณาผลการจำแนกสูงสุดสำหรับการทดลองนี้ พบว่าตัวจำแนก LBN ให้ผลการจำแนกสูงสุดเมื่อใช้ตัวประเมิน C11 (W\_FLDA) ซึ่งให้ผลเช่นเดียวกันทั้งรูปแบบการพูดแบบออกเสียงและไม่ออกเสียง (93.25% สำหรับการพูดแบบออกเสียง และ 80.12% สำหรับการพูดแบบไม่ออกเสียง) พิจารณาประสิทธิภาพของตัวจำแนก LBN (รูปที่ 4.8(ค) และ รูปที่ 4.9(ค)) จะเห็นได้ว่า ผลการจำแนกของตัวประเมินแต่ละชนิดค่อนข้างแตกต่างกันอย่างเห็นได้ชัด โดยที่ค่าพิสัย (range) ของข้อมูลอยู่ในช่วงที่กว้างมาก (59.28% สำหรับการพูดแบบออกเสียง และ 46.92% สำหรับการพูดแบบไม่ออกเสียง) ค่าเบี่ยงเบนมาตรฐานของผลการจำแนกของตัวประเมินทั้ง 14 ชนิด คือ 23.21% (ออกเสียง) และ 19.64% (ไม่ออกเสียง) ดังนั้นแสดงให้เห็นว่า การเลือกตัวประเมินที่เหมาะสมสำหรับการเลือกลักษณะเด่นเพื่อให้เป็นอินพุตให้กับตัวจำแนกมีผลอย่างมากสำหรับประสิทธิภาพในการจำแนกของระบบ โดยเฉพาะอย่างยิ่งตัวจำแนก LBN ในทางตรงกันข้าม เมื่อพิจารณาประสิทธิภาพในการจำแนกของตัวจำแนก ANN จะเห็นได้ว่าผลการจำแนกของตัวประเมินแต่ละชนิดแตกต่างกันเพียงเล็กน้อย โดยที่ค่าเบี่ยงเบนมาตรฐานของผลการจำแนกของตัวประเมินทั้ง 14 ชนิดมีค่าต่ำมาก (1.41% สำหรับการพูดแบบออกเสียง และ 2.3% สำหรับการพูดแบบไม่ออกเสียง)



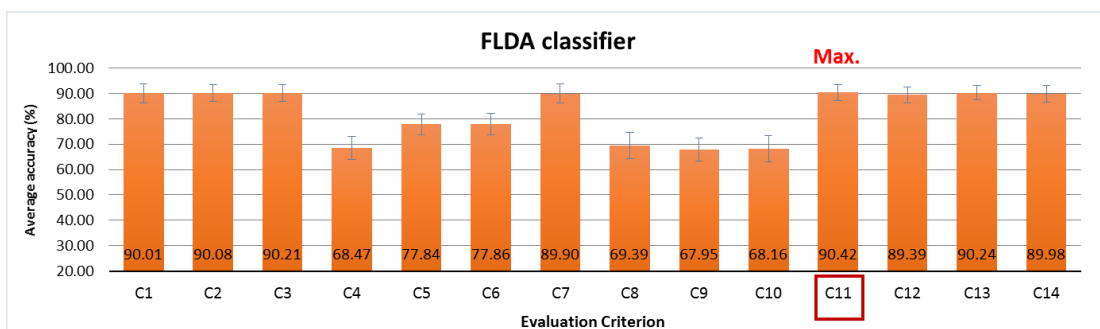


รูปที่ 4.7 เวลาในการประมวลผล (elapsed time) สำหรับการใช้ตัวประเมินแต่ละชนิด

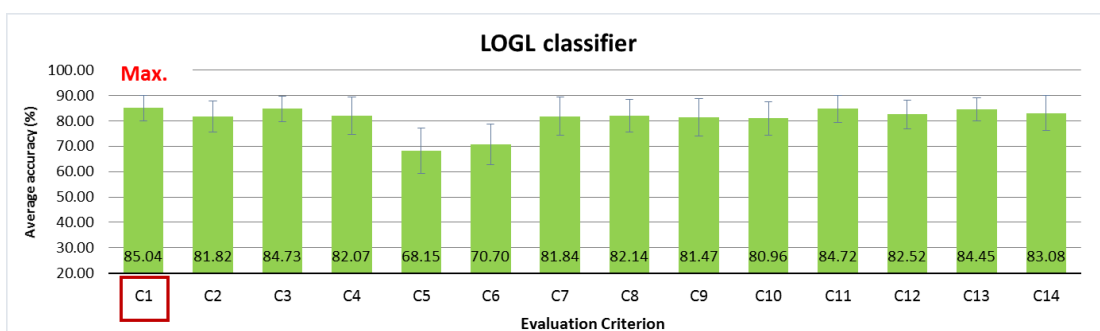
ถึงแม้ว่าตัวประเมิน C11 (ซึ่งเป็นตัวประเมินในกลุ่มที่ 2 ที่ใช้ผลความแม่นยำในการจำแนกของตัวจำแนกในการประเมิน) จะให้ผลการความแม่นยำเฉลี่ยในการจำแนกสูงสุด อย่างไรก็ตามตัวประเมินชนิดนี้ใช้เวลาในการประมวลผลค่อนข้างมาก (แสดงดังรูปที่ 4.7) ดังนั้นการใช้ตัวประเมินในกลุ่มแรกซึ่งใช้คุณลักษณะภายในที่มีอยู่ของข้อมูลในการประเมิน ซึ่งมีผลการจำแนกใกล้เคียงกับตัวประเมิน C11 น่าจะเป็นทางเลือกที่ดีกว่า จากตารางที่ 4.8 แสดงตัวประเมินที่ดีที่สุด 3 อันดับแรกสำหรับตัวจำแนกแต่ละตัว พบว่ามีตัวประเมินในกลุ่มแรก 2 ชนิด ซึ่งให้ผลการจำแนกใกล้เคียงกับตัวประเมิน C11 คือ C2 (F<sub>in-in</sub>) และ C3 (F<sub>Maha-s</sub>) ซึ่งตัวประเมินทั้งสามชนิดนี้ให้ความแม่นยำที่สูงและใช้เวลาในการประมวลผลน้อยเมื่อเปรียบกับตัวประเมินในกลุ่มที่สอง

ตารางที่ 4.8 ตัวประเมินที่มีผลการจำแนกที่ดีที่สุด 3 อันดับแรกสำหรับตัวจำแนกแต่ละชนิด

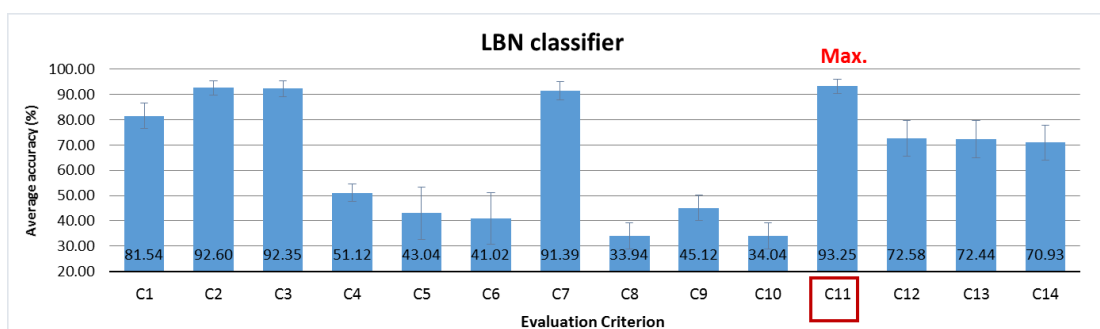
ตัวจำแนก	การพูดแบบออกเสียง		การพูดแบบไม่ออกเสียง	
	ตัวประเมิน	ความแม่นยำเฉลี่ย (%)	ตัวประเมิน	ความแม่นยำเฉลี่ย (%)
FLDA	W_FLDA	90.42±3.11	W_ANN	78.10±3.79
	W_LBN	90.24±2.88	W_FLDA	76.74±4.19
	F_Maha-s	90.21±3.26	F_RES	76.72±3.60
LOGL	F_RES	85.04±5.04	F_in-in	64.02±6.70
	F_Maha-s	84.73±5.02	F_Maha-s	63.02±7.00
	W_FLDA	84.72±5.42	W_LBN	61.50±6.34
<i>LBN</i>	<i>W_FLDA</i>	<i>93.25±2.77</i>	<i>W_FLDA</i>	<i>80.12±3.98</i>
	F_in-in	92.60±2.81	F_in-in	79.40±4.06
	F_Maha-s	92.35±3.16	F_Maha-s	79.35±4.14
ANN	R_RES	87.82±2.49	W_ANN	77.35±3.32
	W_FLDA	87.81±2.05	F_RES	77.23±3.84
	F_Maha-s	87.78±1.99	F_Maha-s	76.42±3.74



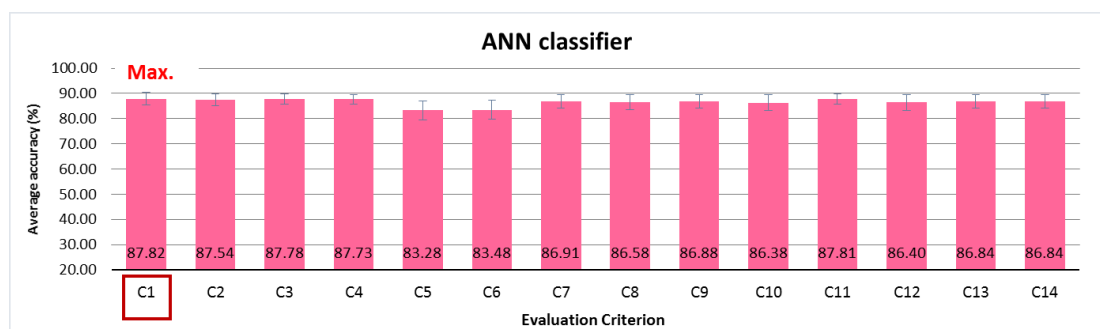
(ก)



(ข)

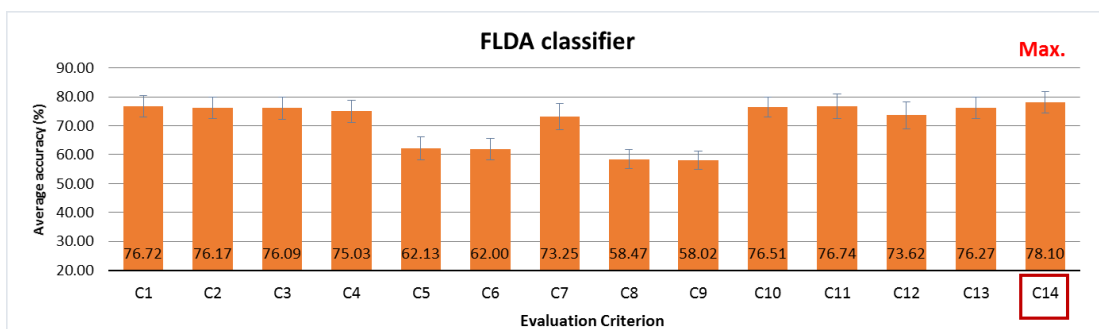


(ค)

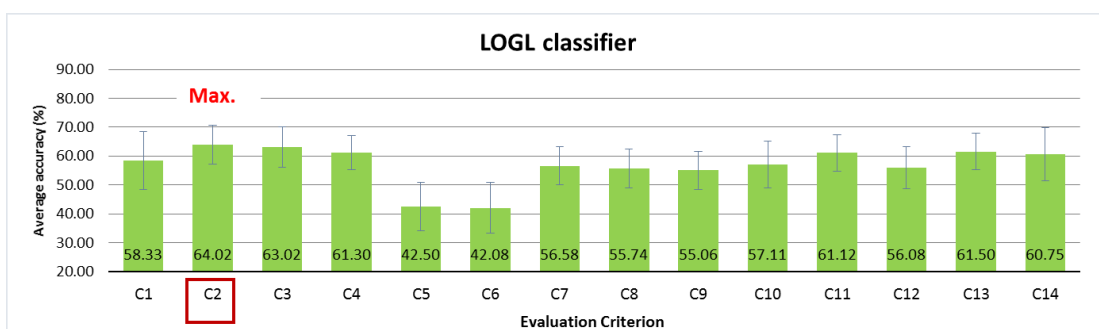


(ง)

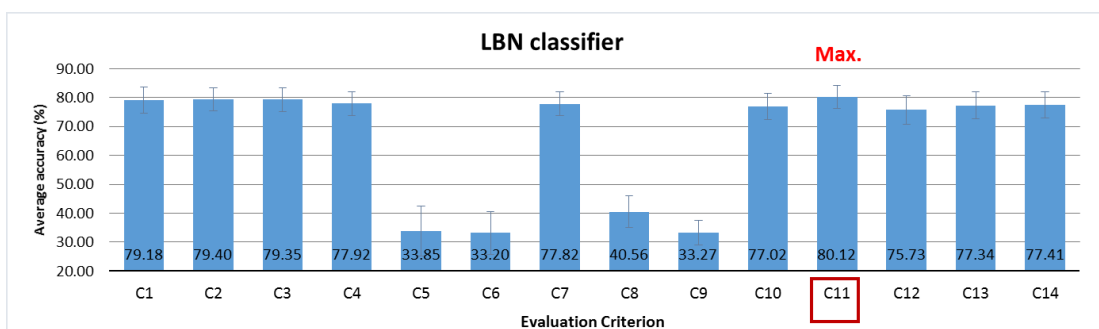
รูปที่ 4.8 เปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด (C1-C14) เมื่อทดสอบด้วยตัวจำแนก 4 ชนิด (ก) FLDA (ข) LOGL (ค) LBN (ง) ANN (สำหรับการพูดแบบออกเสียง)



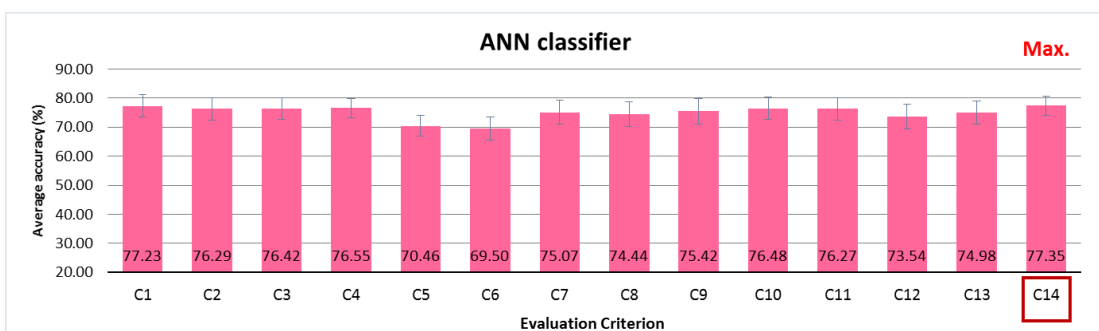
(ก)



(ข)



(ค)



(ง)

รูปที่ 4.9 เปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด (C1-C14) เมื่อทดสอบด้วยตัวจำแนก 4 ชนิด (ก) FLDA (ข) LOGL (ค) LBN (ง) ANN (สำหรับการพูดแบบไม่ออกเสียง)

### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการเปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด และตัวจำแนก 4 ชนิด สำหรับการจำแนกคำพูดหลัก 11 คำพูด จากการทดลองพบว่าการใช้ W\_FLDA เป็นตัวประเมินเพื่อเลือกลักษณะเด่น และ ตัวจำแนก LBN เพื่อจำแนกระบบให้ผลการจำแนกสูงสุด โดยที่ผลความแม่นยำเฉลี่ย คือ 93.25% สำหรับการพูดแบบออกเสียง และ 80.12% สำหรับการพูดแบบไม่ออกเสียง

อย่างไรก็ตามจากผลการทดลองนี้จะเห็นได้ว่าตัวประเมิน W\_FLDA กับ ตัวจำแนก LBN อาจจะไม่ใช่ตัวประเมินและตัวจำแนกที่ดีที่สุดสำหรับงานวิจัยนี้ถึงแม้ว่าจะให้ผลการจำแนกดีที่สุด เหตุผลแรก คือ ตัวประเมิน W\_FLDA ใช้เวลาในการประมวลผลค่อนข้างมาก ซึ่งมีตัวประเมินอีก 2 ชนิด (ตัวประเมินในกลุ่มแรก) ที่มีประสิทธิภาพในการจำแนกใกล้เคียงกับตัวประเมิน W\_FLDA แต่ใช้เวลาในการประมวลผลต่ำ ซึ่งเป็นตัวเลือกสำหรับการใช้งานในระบบที่มีข้อจำกัดด้านเวลา เหตุผลที่สอง ข้อเสียของตัวจำแนก LBN คือ ประสิทธิภาพในการจำแนกลดลงอย่างมากเมื่อลักษณะเด่นที่ถูกป้อนให้กับระบบไม่มีความเหมาะสม ถึงแม้ว่าจะให้ประสิทธิภาพสูงที่สุดเมื่อเปรียบเทียบกับตัวจำแนกชนิดอื่นเมื่อลักษณะเด่นที่ถูกป้อนให้กับระบบมีความเหมาะสม ดังนั้นการทดลองต่อไปจะเป็นการปรับปรุงระบบในส่วนของกระบวนการเลือกลักษณะเด่น โดยจะใช้เทคนิคการเลือกเซตย่อยของลักษณะเด่น (FSS) โดยผ่านกระบวนการค้นหาเพื่อให้ได้ลักษณะเด่นที่เหมาะสมที่สุด

#### 4.4 การเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกเซตย่อยของลักษณะเด่นโดยผ่านกระบวนการค้นหา

การทดลองที่ผ่านมาใช้เทคนิคการจัดอันดับลักษณะเด่น (feature ranking) หรือ FR ในกระบวนการเลือกเซตย่อยของลักษณะเด่นเพื่อใช้เป็นอินพุตให้กับตัวจำแนก ซึ่งวิธีดังกล่าวเป็นวิธีที่ง่ายและใช้เวลาในการประมวลผลน้อย อย่างไรก็ตามข้อเสียของวิธีดังกล่าว คือ ไม่สามารถกำจัดลักษณะเด่นที่มีความซ้ำซ้อนออกไปได้ เนื่องจากผลการประเมินของลักษณะเดนนั้นๆ มีค่าเท่ากัน ดังนั้นการทดลองนี้การเลือกลักษณะเด่นจะใช้เทคนิคการเลือกเซตย่อยของลักษณะเด่น (feature subset selection) หรือ FSS ซึ่งผ่านกระบวนการค้นหาลักษณะเด่น เพื่อเลือกเซตย่อยหรือลักษณะเด่นที่มีความสัมพันธ์กันและกำจัดลักษณะเด่นที่มีความซ้ำซ้อนออกไปเพื่อเพิ่มประสิทธิภาพในการจำแนกของระบบและลดเวลาในการประมวลผล

### (ก) วัตถุประสงค์

- เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาเซตย่อยของลักษณะเด่น 3 ชนิด
- เพื่อเปรียบเทียบตัวประเมินลักษณะเด่น 4 ชนิด
- เพื่อเปรียบเทียบประสิทธิภาพของตัวจำแนก 2 ชนิด

### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S08 ในการออกแบบ เนื่องจากเป็นสัญญาณที่ให้ผลการจำแนกคำพูดหลัก 11 คำที่ดีที่สุดจากการทดลองที่ผ่านมา โดยจะใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงเท่านั้น

สาเหตุที่ใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครเพียง 1 คน เนื่องจากเพื่อให้สามารถเปรียบเทียบผลของการใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ได้อย่างถูกต้อง

สำหรับการทดลองที่ผ่านมาได้ทำการเลือกลักษณะเด่น โดยการพิจารณาลักษณะเด่นที่มีผลการประเมินที่ดีที่สุดของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน และนำลักษณะเด่นดังกล่าวมาหาความถี่ เพื่อเลือกเซตย่อยของลักษณะเด่นที่ดีที่สุดที่สามารถนำมาใช้ได้กับสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด (S01-S10) ซึ่งถ้าใช้วิธีนี้กับเทคนิคการเลือกลักษณะเด่นแบบ FSS โดยผ่านกระบวนการค้นหา เซตย่อยของลักษณะเด่นสุดท้ายจะไม่ได้มาจากกระบวนการค้นหาที่แท้จริง ดังนั้นการทดลองนี้จึงพิจารณาข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่ดีที่สุดเพียง 1 คน คือ S08 แล้วนำเซตย่อยของลักษณะเด่นที่ได้ไปใช้กับข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้ออื่นๆ

#### ขั้นตอนการทดลอง

- 1) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ S08 มาคำนวณลักษณะเด่น 22 ชนิด เช่นเดียวกับการทดลองในหัวข้อ 4.3 ซึ่งแสดงดังตารางที่ 4.6
- 2) ทำการเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS โดยใช้ตัวประเมิน 4 ชนิด และ กระบวนการค้นหาลักษณะเด่น 3 วิธี โดยกำหนดจำนวนลักษณะเด่นที่จะ

ถูกเลือก 9 ตัวสำหรับแต่ละช่องสัญญาณ เพื่อให้สามารถเปรียบเทียบกับผลการทดลองที่ผ่านมาได้

ตัวประเมินลักษณะเด่น 4 ชนิด ที่ใช้ในการทดลองนี้เป็นตัวประเมินที่มีผลการจำแนกที่ดีที่สุด 3 ชนิดจากการทดลองที่ผ่านมา (หัวข้อ 4.3) ได้แก่ F\_in-in F\_Maha-s และ W\_FLDA นอกจากนี้ เนื่องจากการทดลองนี้จะทำการเปรียบเทียบประสิทธิภาพของตัวจำแนก 2 ชนิด คือ LBN และ FLDA ดังนั้นจะนำตัวประเมิน W\_LBN มาใช้ในการทดลองนี้ด้วย เนื่องจากตัวประเมินแบบ W\_LBN เป็นตัวประเมินที่ใช้ผลการจำแนกของตัวจำแนกสำหรับการประเมินลักษณะเด่น ดังนั้นลักษณะเด่นที่ถูกระบุจากตัวประเมินดังกล่าวน่าจะมีประสิทธิภาพในการจำแนกที่ดีที่สุดเมื่อใช้ตัวจำแนกชนิดเดียวกัน

กระบวนการค้นหาลักษณะเด่น จะใช้วิธีการค้นหา 3 วิธีดังตารางที่ 4.9 ซึ่งเป็นวิธีการค้นหาแบบลำดับ

3) นำเซตย่อยของลักษณะเด่นที่ได้จากการออกแบบโดยใช้ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัคร S08 ไปใช้กับข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้งหมดสำหรับทั้งแบบการพูดออกเสียงและไม่ออกเสียง

4) ฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก 2 ชนิด FLDA และ LBN เนื่องจากเป็นตัวจำแนกที่มีประสิทธิภาพที่ดีใกล้เคียงกันจากผลการทดลองที่ผ่านมา (หัวข้อที่ 4.3) วัดประสิทธิภาพในการจำแนกของตัวจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 เปรียบเทียบผลการจำแนกของกระบวนการเลือกลักษณะเด่นแบบ FSS โดยเปรียบเทียบเทคนิคการค้นหา 3 วิธี ตัวประเมิน 4 ชนิด เปรียบเทียบประสิทธิภาพของตัวจำแนก 2 ชนิด นอกจากนั้นเปรียบเทียบประสิทธิภาพเทคนิคการเลือกลักษณะเด่นระหว่างเทคนิค FR (เทคนิคเดิม) และ FSS ในกระบวนการฝึกสอนและทดสอบระบบ ชุดข้อมูลสำหรับฝึกสอนและทดสอบระบบต้องเป็นชุดข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครคนเดียวกัน

#### ตารางที่ 4.9 เทคนิคการค้นหาลักษณะเด่น 3 วิธี

ประเภท	ตัวย่อ	ชื่อเต็ม
Sequential search	FW	forward search
	BWs	backward search
	LR	plus L take away R search (L=2, R=1)

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

1) การเปรียบเทียบประสิทธิภาพเทคนิคการเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหา 3 วิธี และ ตัวประเมิน 4 วิธี ซึ่งใช้ข้อมูลของสัญญาณ S08 (ออกเสียง) ในการออกแบบ

รูปที่ 4.10 แสดงผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณ S08 สำหรับการพูดแบบออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหาหลักเกณฑ์เด่น 3 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนกสองชนิด คือ (ก) FLDA (ข) LBN เมื่อพิจารณาผลความแม่นยำของตัวจำแนก FLDA วิธีการค้นหาหลักเกณฑ์เด่นที่ให้ความแม่นยำดีที่สุด คือ FW โดยใช้ตัวประเมิน W\_FLDA ซึ่งให้ผลความแม่นยำ 97.82% เมื่อเปรียบเทียบกับผลการทดลองที่ผ่านมาซึ่งใช้เทคนิค FR โดยไม่มีการค้นหาหลักเกณฑ์เด่น ให้ผลความแม่นยำ 96.55% (พิจารณาตัวประเมิน W\_FLDA เหมือนกัน) ดังนั้นความแม่นยำเพิ่มขึ้น 1.27% ซึ่งความแม่นยำเพิ่มขึ้นเพียงเล็กน้อยเมื่อพิจารณาผลความแม่นยำของตัวจำแนก LBN เทคนิคการค้นหาที่ให้ความแม่นยำดีที่สุด คือ FW เช่นเดียวกัน โดยใช้ตัวประเมินแบบ W\_LBN โดยให้ผลความแม่นยำ 99.45% เมื่อเปรียบเทียบกับผลการทดลองที่ผ่านมาซึ่งใช้เทคนิค FR โดยไม่มีการค้นหาหลักเกณฑ์เด่น ให้ผลความแม่นยำ 92.73% (พิจารณาตัวประเมิน W\_LBN เหมือนกัน) ดังนั้นความแม่นยำเพิ่มขึ้น 6.73% ดังนั้นจะเห็นว่าตัวจำแนก LBN ให้ผลการจำแนกดีกว่าตัวจำแนก FLDA

จะเห็นว่าจากการทดลองนี้การประเมินลักษณะเด่นโดยใช้ W\_LBN (ประเมินโดยใช้ผลการจำแนกของตัวจำแนก LBN) ให้ผลดีที่สุดเมื่อทดสอบกับตัวจำแนก LBN และ ตัวประเมิน W\_FLDA (ประเมินโดยใช้ผลการจำแนกของตัวจำแนก FLDA) ให้ผลดีที่สุดเมื่อทดสอบกับตัวจำแนก FLDA ซึ่งโดยทั่วไปตัวจำแนกที่ใช้ในขั้นตอนการประเมินลักษณะเด่นกับตัวจำแนกในกระบวนการจำแนกข้อมูล ถ้าเป็นตัวจำแนกชนิดเดียวกันจะให้ความแม่นยำสูงสุด [87] แต่จากการทดลองก่อนหน้านี้ (หัวข้อที่ 4.3) ผลการทดลองไม่เป็นไปตามนี้ ทั้งนี้อาจเป็นเพราะว่าการทดลองดังกล่าวมีการประยุกต์วิธีการเลือกลักษณะเด่น ซึ่งไม่ได้เลือกลักษณะเด่นโดยใช้ตัวประเมินนั้นๆ โดยตรง แต่จะเลือกลักษณะเด่นที่ผ่านการประเมินด้วยตัวประเมินที่มีค่าการประเมินดีที่สุดสำหรับสัญญาณของอาสาสมัครแต่ละคน แล้วหาค่าความถี่ของลักษณะเด่นในกลุ่มนั้น แล้วเลือกลักษณะเด่นอีกครั้งตามค่าความถี่ของลักษณะเด่น เพื่อให้ได้จำนวนลักษณะเด่นเพียง 1 ชุด สำหรับใช้กับข้อมูลของอาสาสมัครทั้ง 10 คน แต่สำหรับการทดลองนี้ใช้ข้อมูลของอาสาสมัครเพียง 1 คน แล้วเลือกลักษณะเด่นโดยตรงจากเทคนิคการเลือกลักษณะเด่นดังกล่าว



อย่างไรก็ตามจากผลการทดลองที่ผ่านมา (หัวข้อ 4.3) พบว่าตัวจำแนก LBN มีความอ่อนไหวต่อการเลือกลักษณะเด่นมาก กล่าวคือ ถ้าเลือกลักษณะเด่นพลาดไปเพียงเล็กน้อย ความแม่นยำจะลดลงอย่างมาก เช่นเดียวกับการทดลองนี้ ผลความแม่นยำในการเลือกลักษณะเด่นแต่ละวิธีให้ความแม่นยำที่แตกต่างกันอย่างมาก ถ้าเลือกลักษณะเด่นที่เหมาะสมจะให้ความแม่นยำที่สูงที่สุด ซึ่งสูงกว่าตัวจำแนกตัวอื่นๆ ในการทดลองนี้พบว่า ตัวจำแนก LBN ให้ความแม่นยำสูงกว่าตัวจำแนก FLDA โดยที่ความแม่นยำสูงสุดจากตัวจำแนก FLDA คือ 97.82% ส่วนความแม่นยำสูงสุดจากตัวจำแนก LBN คือ 99.45%

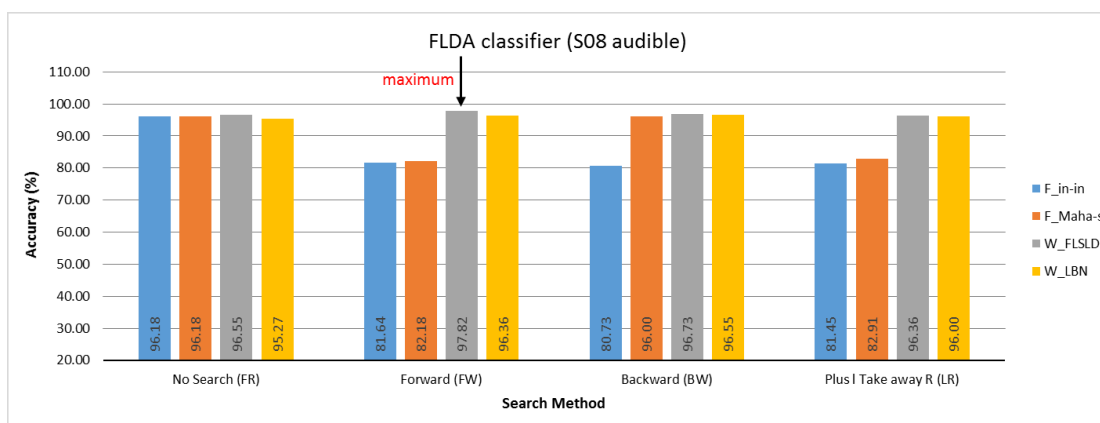
จากการทดลองที่ผ่านมา (หัวข้อที่ 4.3) พบว่าวิธีการประเมินแบบ F\_in-in และ F\_Maha-s ให้ผลการประเมินเหมือนกัน และความแม่นยำเท่ากัน แต่จากการทดลองนี้พบว่าให้ผลความแม่นยำแตกต่างกันมากสำหรับทั้ง 3 รูปแบบของวิธีการค้นหา โดยที่ F\_in-in จะให้ความแม่นยำต่ำกว่า F\_Maha-s มาก โดยเฉพาะอย่างยิ่งเมื่อจำแนกโดยใช้ตัวจำแนก LBN วิธีการค้นหาแบบ BW จะให้ความแม่นยำระหว่างตัวประเมินทั้งสองรูปแบบแตกต่างกันอย่างมาก กล่าวคือ ตัวประเมิน F\_in-in ให้ความแม่นยำ 38.91% และ ตัวประเมิน F\_Maha-s ให้ความแม่นยำ 84.73% ซึ่งแตกต่างกัน 45.82%

2) ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูด สำหรับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน ทั้งแบบออกเสียงและไม่ออกเสียง

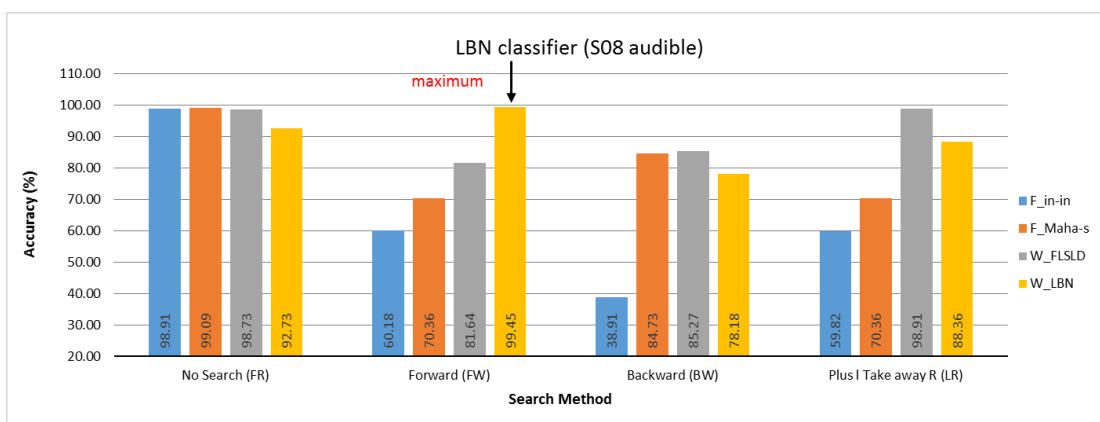
เนื่องจากผลการทดลองข้างต้น ทำการคัดเลือกลักษณะเด่นโดยการให้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครที่ดีที่สุดเพียง 1 คน (S08) เท่านั้น ดังนั้นเพื่อทดสอบว่าเซตย่อยของลักษณะเด่นที่เลือกมานั้นสามารถนำมาใช้ได้กับข้อมูลของอาสาสมัครคนอื่นได้หรือไม่ การทดลองนี้จะนำเซตย่อยของลักษณะเด่นดังกล่าวซึ่งถูกเลือกด้วยวิธีการต่างๆ ซึ่งถูกออกแบบโดยใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ S08 มาทดสอบกับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน ทั้งแบบการพูดแบบออกเสียงและไม่ออกเสียง ซึ่งผลการทดลองแสดงดังรูปที่ 4.11 และ รูปที่ 4.12

รูปที่ 4.11 แสดงผลความแม่นยำเฉลี่ยสำหรับการพูดแบบออกเสียง ซึ่งคำนวณจากความแม่นยำของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน จากผลการทดลองจะเห็นได้ว่าให้ผลไปในทิศทางเดียวกันกับการทดลองก่อนหน้าซึ่งใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อเพียง 1 คน กล่าวคือ วิธีการเลือกลักษณะเด่นที่ให้ความแม่นยำเฉลี่ยสูงสุด คือ การใช้เทคนิคค้นหาแบบ FW โดยใช้ W\_LBN เป็นตัวประเมิน โดยที่ให้ผลความแม่นยำเฉลี่ยสูงสุด 94.00% เมื่อ

จำแนกโดยใช้ตัวจำแนก LBN สำหรับการพูดแบบออกไม่ออกเสียงซึ่งแสดงในรูปที่ 4.12 ให้ผลเช่นเดียวกัน กล่าวคือ วิธีการเลือกลักษณะเด่นที่ให้ความแม่นยำเฉลี่ยสูงสุด คือ การใช้เทคนิคค้นหาแบบ FW โดยใช้ W\_LBN เป็นตัวประเมิน โดยที่ให้ความแม่นยำเฉลี่ยสูงสุด 81.06% เมื่อจำแนกโดยใช้ตัวจำแนก LBN



(ก)



(ข)

รูปที่ 4.10 ผลความแม่นยำในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณ S08 สำหรับการพูดแบบออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหาลักษณะเด่น 3 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN

เมื่อเปรียบเทียบความแม่นยำเฉลี่ยสูงสุดระหว่างเทคนิคการเลือกลักษณะเด่นแบบ FR และ FSS ซึ่งใช้เทคนิคการค้นหาในการคัดเลือกเซตย่อยของลักษณะเด่น ซึ่งแสดงดังตารางที่ 4.10 จะเห็นว่ามีความแตกต่างกันเพียงเล็กน้อยเท่านั้นทั้งแบบการพุดแบบออกเสียงและไม่ออกเสียง ดังนั้นสำหรับงานที่ต้องการความเร็วในการประมวลผล วิธีการเลือกลักษณะเด่นแบบ FR เป็นทางเลือกหนึ่งที่มีความเหมาะสม อย่างไรก็ตามสำหรับงานที่ไม่คำนึงถึงความเร็วในการประมวลผลมากนัก การเลือกใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS โดยใช้วิธีการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ W\_LBN เป็นตัวประเมิน มีความเหมาะสมมากกว่า เนื่องจากวิธีการเลือกเซตย่อยของลักษณะเด่นโดยใช้กระบวนการค้นหา สามารถกำจัดลักษณะเด่นที่มีความซ้ำซ้อนออกไปได้ ซึ่งช่วยลดจำนวนลักษณะเด่นที่จะป้อนให้กับตัวจำแนก แต่ยังคงประสิทธิภาพในการจำแนกไว้เท่าเดิม

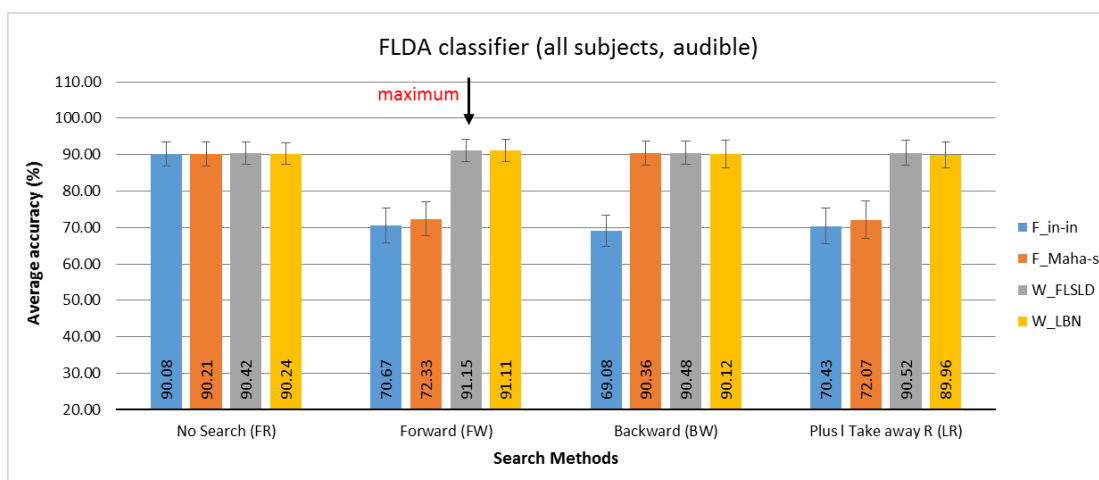
ตารางที่ 4.11 แสดงการเปรียบเทียบประสิทธิภาพของตัวจำแนก FLDA และ LBN โดยพิจารณาจากความแม่นยำสูงสุดที่ได้จากตัวจำแนกแต่ละชนิด จะเห็นได้ว่าตัวจำแนก LBN มีความแม่นยำเฉลี่ยสูงกว่า FLDA 2.85% สำหรับการพุดแบบออกเสียง และ 2.84% สำหรับการพุดแบบไม่ออกเสียง อย่างไรก็ตามจากผลการทดลองจะเห็นได้ว่าตัวจำแนก LBN มีความอ่อนไหวต่อการเลือกลักษณะเด่นมาก ดังนั้นการเลือกใช้ตัวจำแนก LBN ให้มีประสิทธิภาพสูงสุดต้องเลือกเทคนิคเลือกลักษณะเด่นให้มีความเหมาะสม ซึ่งจากการทดลองนี้พบว่า เทคนิคการเลือกลักษณะเด่นที่เหมาะสมคือ การเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาแบบ FW ซึ่งใช้ตัวประเมินชนิด W\_LBN

**ตารางที่ 4.10** เปรียบเทียบเทคนิคการเลือกลักษณะเด่นแบบ FR และ FSS

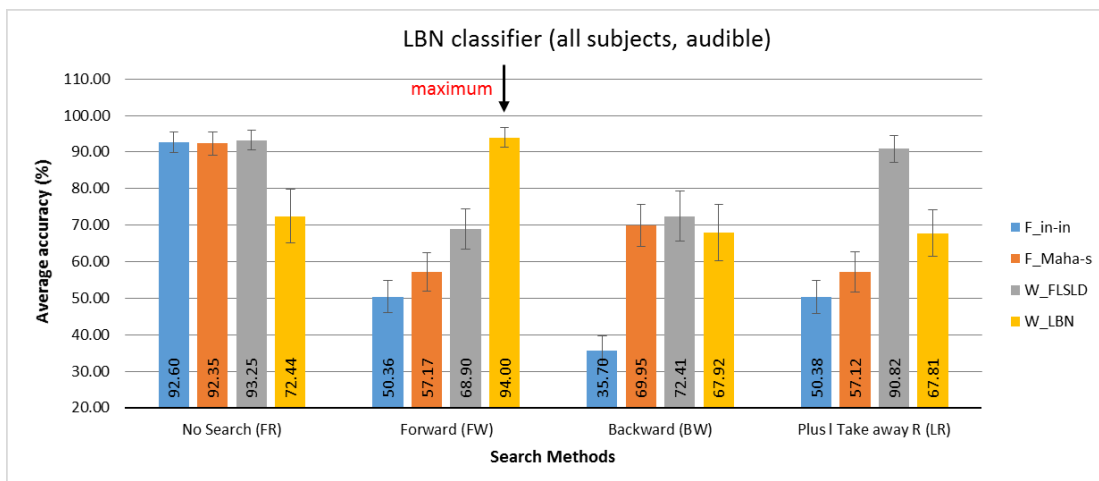
เทคนิคการเลือก ลักษณะเด่น	ความแม่นยำเฉลี่ยสูงสุด (%)		เทคนิคที่ให้ความแม่นยำเฉลี่ย สูงสุด
	ออกเสียง	ไม่ออกเสียง	
FR	93.25±2.77	80.12±3.98	W_FLDA, LBN
FSS	94.00±2.70	81.06±4.05	FW, W_LBN, LBN

ตารางที่ 4.11 เปรียบเทียบประสิทธิภาพของตัวจำแนก FLDA และ LBN

ตัวจำแนก	ความแม่นยำเฉลี่ยสูงสุด (%)		การเลือกลักษณะเด่น
	ออกเสียง	ไม่ออกเสียง	
FLDA	91.15±3.04	78.22±3.75	FW, W_FLDA
LBN	94.00±2.70	81.06±4.05	FW, W_LBN

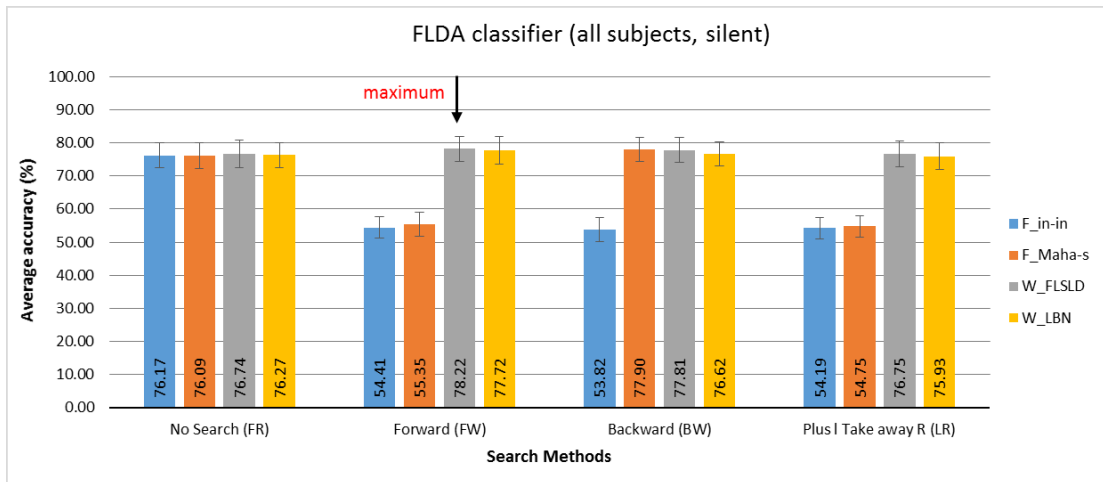


(ก)

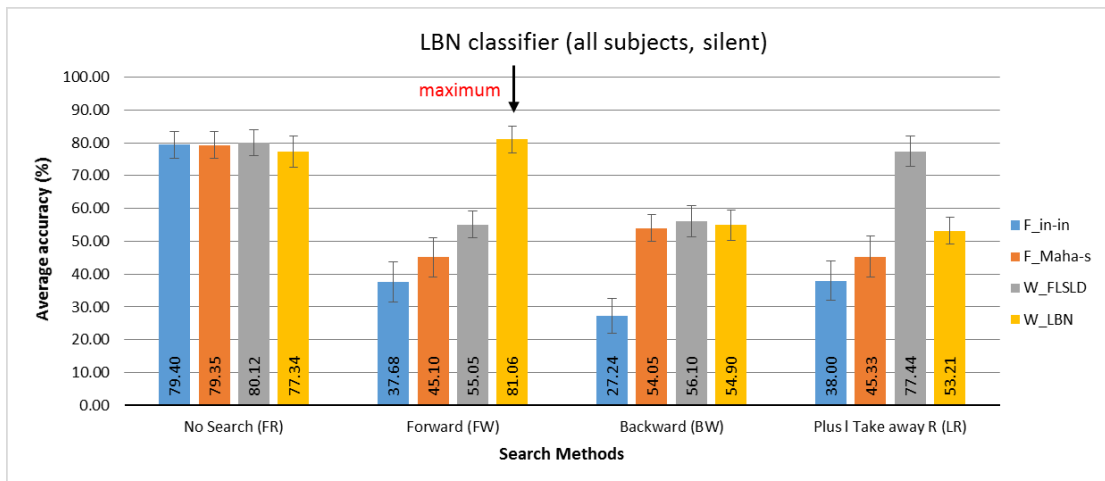


(ข)

รูปที่ 4.11 ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด สำหรับการพูดแบบออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหาลักษณะเด่น 3 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN



(ก)



(ข)

รูปที่ 4.12 ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำพูดของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด สำหรับการพูดแบบไม่ออกเสียง ซึ่งเปรียบเทียบเทคนิคการค้นหาลักษณะเด่น 4 วิธี โดยการใช้ตัวประเมิน 4 วิธี เมื่อใช้ตัวจำแนก (ก) FLDA (ข) LBN

### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการเปรียบเทียบประสิทธิภาพของเทคนิคเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาเซตย่อยของลักษณะเด่น 3 ชนิด คือ FW BW และ LR และเปรียบเทียบการใช้ตัวประเมิน 4 ชนิด คือ F\_in-in F-Maha-s W\_FLDA และ W\_LBN นอกจากนี้ทดสอบประสิทธิภาพของตัวจำแนก 2 ชนิด คือ LBN และ FLDA จากผลการทดลองพบว่า เทคนิคที่ให้ประสิทธิภาพสูงสุด คือ การเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ตัวประเมิน W\_LBN โดยให้ความแม่นยำเฉลี่ยสูงสุดในการจำแนกคำพูดหลัก 11 คำเมื่อจำแนกโดยใช้ตัวจำแนก LBN เท่ากับ 94% สำหรับการพูดแบบออกเสียง) และ 81.06% สำหรับการพูดแบบไม่ออกเสียง

อย่างไรก็ตามจากการทดลองดังกล่าว เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาเซตย่อยของลักษณะเด่น มีการกำหนดจำนวนของลักษณะเด่นที่ใช้เป็นอินพุตให้กับตัวจำแนก 9 ตัวสำหรับแต่ละช่องสัญญาณ (จำนวนที่เหมาะสมที่ได้ออกแบบไว้ในบทที่ 3) ซึ่งจำนวนลักษณะเด่นดังกล่าวอาจจะไม่เหมาะสมสำหรับเทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งจะทำการทดลองเพื่อหาจำนวนลักษณะเด่นที่เหมาะสมในการทดลองต่อไป

#### 4.5 การคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกตัวเลขไทย 11 คำ โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN

จากการทดลองที่ผ่านมา (หัวข้อที่ 4.1-4.3) ได้กำหนดจำนวนลักษณะเด่นสำหรับป้อนเป็นอินพุตให้กับตัวจำแนก จำนวน 9 ตัวสำหรับแต่ละช่องสัญญาณ (5 ช่องสัญญาณ) โดยอ้างอิงมาจากผลการออกแบบในบทที่ 3 ซึ่งใช้เทคนิคการเลือกลักษณะเด่นแบบ FR ทั้งนี้เพื่อให้ง่ายต่อการเปรียบเทียบประสิทธิภาพของเทคนิคการเลือกลักษณะเด่นแต่ละวิธี อย่างไรก็ตามจำนวนดังกล่าวอาจจะไม่ใช่จำนวนที่เหมาะสมสำหรับวิธีการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN (เทคนิคที่ให้ความแม่นยำสูงสุดจากการทดลองที่ผ่านมาในหัวข้อที่ 4.4) ดังนั้นการทดลองนี้จะทำการคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมโดยใช้เทคนิคดังกล่าว

### (ก) วัตถุประสงค์

- เพื่อคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกคำพูดหลัก 11 คำพูด โดยใช้เทคนิคการคัดเลือกลักษณะเด่นแบบ FSS ซึ่งค้นหาเซตย่อยของลักษณะเด่นโดยวิธี FW และใช้ตัวประเมิน W\_LBN

### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S08 ในการออกแบบ เนื่องจากเป็นสัญญาณที่ให้ผลการจำแนกคำพูดหลัก 11 คำที่ดีที่สุดจากการทดลองที่ผ่านมา โดยจะใช้ข้อมูลการออกเสียงพูดชุดที่ 1 ซึ่งเป็นข้อมูลของคำพูด 11 พยางค์ "ศูนย์" ถึง "สิบ" โดยเป็นข้อมูลของการพูดแบบออกเสียงเท่านั้น หลังจากนั้นจะนำผลที่ได้จากการออกแบบโดยใช้ข้อมูลของ S08 ไปใช้กับข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด (S01-S10)

#### ขั้นตอนการทดลอง

- 1) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ S08 มาคำนวณลักษณะเด่น 22 ชนิด เช่นเดียวกับการทดลองในหัวข้อ 4.3 และ 4.4 ซึ่งแสดงดังตารางที่ 4.6
- 2) ทำการประเมินและคัดเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งผ่านกระบวนการค้นหาวิธี FW โดยใช้ W\_LBN เป็นตัวประเมิน
- 3) พล็อตกราฟแสดงความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและค่าการประเมิน (feature curve) พิจารณาลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมที่จะป้อนเป็นอินพุตให้กับตัวจำแนก
- 4) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่เหลือ (S01-S10) มาคำนวณลักษณะเด่น ซึ่งถูกเลือกมาจากหัวข้อที่ 3) แล้วป้อนเป็นอินพุตให้กับตัวจำแนก LBN
- 5) ฝึกสอนและทดสอบระบบโดยใช้ตัวจำแนก LBN โดยเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว เรียงลำดับลักษณะเด่นตามผลที่ได้จากหัวข้อที่ 3) วัดประสิทธิภาพในการจำแนกของตัวจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 คำนวณความแม่นยำเฉลี่ยในการจำแนกสำหรับข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด ในกระบวนการฝึกสอนและทดสอบระบบ ชุดข้อมูลสำหรับฝึกสอนและทดสอบระบบต้องเป็นชุดข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครคนเดียวกัน

6) พล็อตกราฟแสดงความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและผลความแม่นยำเฉลี่ยในการจำแนก พิจารณาจำนวนลักษณะเด่นที่เหมาะสมที่จะป้อนเป็นอินพุตให้กับตัวจำแนกอีกครั้ง โดยพิจารณาจากความแม่นยำเฉลี่ยในการจำแนกซึ่งคำนวณจากข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้งหมด

### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

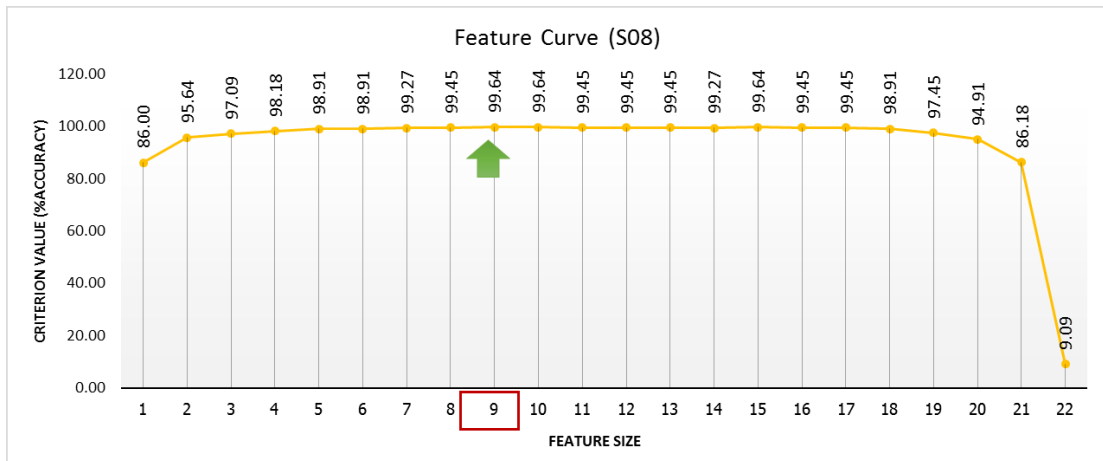
รูปที่ 4.13 แสดงกราฟความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและค่าการประเมิน เมื่อใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW และใช้ W\_LBN เป็นตัวประเมิน สำหรับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S08 ในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่น จำนวนลักษณะเด่นจะถูกเพิ่มครั้งละ 1 ตัว ก่อนที่เซตย่อยของลักษณะเด่นจะถูกประเมินด้วยตัวประเมิน จากกราฟแสดงให้เห็นว่าเมื่อจำนวนลักษณะเด่นเพิ่มขึ้นค่าการประเมินจะเพิ่มขึ้นเรื่อยๆ โดยที่ค่าการประเมินสูงสุดได้รับเมื่อจำนวนลักษณะเด่นเท่ากับ 9 ตัว ซึ่งให้ค่าการประเมิน 99.64 % หลังจากนั้นค่าการประเมินจะค่อยๆ คงที่และจะค่อยๆ ลดลง โดยจะลดลงอย่างมากเมื่อเพิ่มลักษณะเด่นตัวที่ 22 เข้าไป ซึ่งก็คือ ลักษณะเด่น SM5 โดยให้ผลการประเมินเพียง 9.09% ซึ่งผลการประเมินลดลง 77.09% นอกจากนี้ลักษณะเด่นที่มีผลในการลดความแม่นยำในการจำแนก คือ SM2 และ SM3 ซึ่งผลการประเมินลดลง 2.55% เมื่อเพิ่มลักษณะเด่น SM2 เข้าไป และผลการประเมินลดลง 8.73% เมื่อเพิ่มลักษณะเด่น SM3 เข้าไป ซึ่งตารางที่ 4.12 แสดงลักษณะเด่นที่ถูกเลือกในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่น

จากการทดลองนี้จะเห็นได้ว่า จำนวนลักษณะเด่นที่ให้ผลการประเมินสูงสุดเท่ากับ 9 ตัว ซึ่งเซตย่อยของลักษณะเด่นดังกล่าวคือ WL WAMP MMAV2 MNH MNF MMAV1 VAR RMS และ MDF นำลักษณะเด่นชุดนี้ไปทดสอบกับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครอีก 9 คน ทั้งการพูดแบบออกเสียงและออกเสียง โดยปรับเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว หาความแม่นยำเฉลี่ยในการจำแนกโดยใช้ตัวจำแนก LBN ซึ่งผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ แสดงดังรูปที่ 4.14

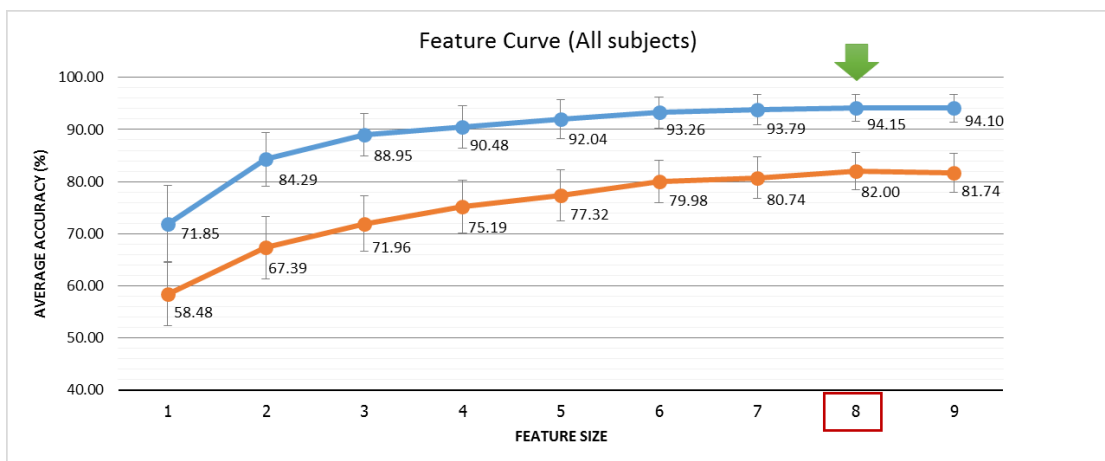


**ตารางที่ 4.12** ลักษณะเด่นที่ถูกเลือกในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่นด้วยวิธี FW และผลการประเมินเซตย่อยของลักษณะเด่นในแต่ละรอบด้วยวิธีการใช้ตัวประเมิน W\_LBN

รอบในการค้นหาเซตย่อยของลักษณะเด่น (จำนวนลักษณะเด่น)	ลักษณะเด่นที่ถูกเพิ่มในแต่ละรอบ	ค่าการประเมิน (%)
1	WL	86.00
2	WAMP	95.64
3	MMAV2	97.09
4	MNH	98.18
5	MNF	98.91
6	MMAV1	98.91
7	VAR	99.27
8	RMS	99.45
9	MDF	<b>99.64</b>
10	SM1	99.64
11	PWL	99.45
12	ZCH	99.45
13	MAV	99.45
14	MNL	99.27
15	ZC	99.64
16	IAV	99.45
17	SSC	99.45
18	MN	98.91
19	PWH	97.45
20	SM2	94.91
21	SM3	86.18
<b>22</b>	<b>SM5</b>	<b>9.09</b>



รูปที่ 4.13 ผลการประเมินและเลือกลักษณะเด่นโดยใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน  $W\_LBN$  สำหรับสัญญาณ S08



รูปที่ 4.14 ผลความแม่นยำเฉลี่ยในการจำแนกของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้ง 10 คน เมื่อเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว

จากรูปที่ 4.14 แสดงให้เห็นว่า ค่าความแม่นยำเฉลี่ยจะเพิ่มขึ้นสูงสุดเมื่อเพิ่มจำนวนลักษณะเด่น 8 ตัว ทั้งการพูดแบบออกเสียง (94.15%) และไม่ออกเสียง (82.00%) ซึ่งเมื่อเปรียบเทียบกับ การทดลองที่ผ่านมา (หัวข้อที่ 4.4) ซึ่งใช้เทคนิคการเลือกลักษณะเด่นวิธีเดียวกัน แต่กำหนดจำนวนลักษณะเด่น 9 ตัว การทดลองให้ผลความแม่นยำเพิ่มขึ้นเล็กน้อย อย่างไรก็ตาม จำนวนลักษณะเด่นที่ใช้น้อยกว่า

### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกตัวเลขไทย 11 คำ โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN จากการทดลองพบว่าการใช้จำนวนลักษณะเด่น 8 ตัวให้ความแม่นยำเฉลี่ยสูงสุด โดยให้ความแม่นยำ 94.15% สำหรับการพูดแบบออกเสียง และ 82.00% สำหรับการพูดแบบไม่ออกเสียง โดยเซตของลักษณะเด่นที่ถูกใช้ คือ WL WAMP MMAV2 MNH MNF MMAV1 VAR และ RMS

เปรียบเทียบกับผลการทดลองที่ผ่านมาในหัวข้อที่ 4.4 จะให้ความแม่นยำที่ใกล้เคียงกันมาก อย่างไรก็ตามในการทดลองดังกล่าวทำการกำหนดจำนวนลักษณะเด่นโดยอ้างอิงจากผลการทดลองในบทที่ 3 แต่สำหรับการทดลองนี้จำนวนลักษณะเด่นที่ถูกเลือกได้มาจากการพิจารณารูปความสัมพันธ์ระหว่างจำนวนลักษณะเด่นและผลการประเมินจากการคัดเลือกลักษณะเด่นโดยวิธี FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN ซึ่งให้ผลถูกต้องมากกว่า

#### 4.6 การคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW

จากการทดลองในหัวข้อที่ 4.3-4.5 ได้ออกแบบเทคนิคการเลือกลักษณะเด่นสำหรับการจำแนกคำพูดหลัก 11 คำพูด ซึ่งพบว่าวิธีที่มีประสิทธิภาพดีที่สุด คือ เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้การค้นหาเซตย่อยของลักษณะเด่นแบบ FW โดยใช้ตัวประเมิน W\_LBN นอกจากนี้ได้เปรียบเทียบประสิทธิภาพของตัวจำแนก ซึ่งพบว่าตัวจำแนก LBN ให้ผลการจำแนกที่ดีที่สุด และใช้เวลาในการประมวลผลน้อยที่สุด ดังนั้นในการทดลองนี้จะนำเทคนิคดังกล่าวมาใช้ในการจำแนกวรรณยุกต์ 5 เสียง เนื่องจากการทดลองในหัวข้อที่ 4.2 พบว่าตัวจำแนก LOGL มีประสิทธิภาพดีที่สุดสำหรับการจำแนกวรรณยุกต์ ดังนั้นการทดลองนี้จะทดสอบประสิทธิภาพของตัวจำแนกชนิดนี้ด้วย โดยจะใช้ตัวประเมิน W\_LOGL สำหรับเลือกลักษณะเด่นเพื่อใช้ในการจำแนกด้วยตัวจำแนก LOGL

### (ก) วัตถุประสงค์

- เพื่อคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ 5 เสียง โดยใช้เทคนิคการคัดเลือกลักษณะเด่นแบบ FSS ซึ่งค้นหาเซตย่อยของลักษณะเด่นโดยวิธี FW โดยเปรียบเทียบการใช้ตัวประเมิน 2 ชนิด คือ W\_LBN และ W\_LOGL และเปรียบเทียบตัวจำแนก 2 ชนิด คือ LBN และ LOGL

### (ข) การทดลอง

การทดลองนี้จะใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S04 ในการออกแบบเพื่อเลือกลักษณะเด่นที่เหมาะสม เนื่องจากเป็นสัญญาณที่ให้ผลการจำแนกวรรณยุกต์ 5 เสียงที่ดีที่สุดจากการทดลองที่ผ่านมา โดยจะใช้ข้อมูลการออกเสียงพูดชุดที่ 2 ซึ่งเป็นข้อมูลของการพูดเสียงวรรณยุกต์ ดังตารางที่ 3.3 หลังจากนั้นจะนำผลที่ได้จากการออกแบบโดยใช้ข้อมูลของ S04 ไปใช้กับข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด (S01-S10)

#### ขั้นตอนการทดลอง

- 1) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ S04 มาคำนวณลักษณะเด่น 124 ชนิด เช่นเดียวกับการทดลองในหัวข้อ 3.3.2.3 ซึ่งแสดงดังรูปที่ 3.43
- 2) ทำการประเมินและคัดเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งผ่านกระบวนการค้นหาวิธี FW โดยใช้ตัวประเมิน 2 ชนิด คือ W\_LBN และ W\_LOGL
- 3) พล็อตกราฟแสดงความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและค่าการประเมิน พิจารณาลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมที่จะป้อนเป็นอินพุตให้กับตัวจำแนก
- 4) นำข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่เหลือ (S01-S10) มาคำนวณลักษณะเด่น ซึ่งถูกเลือกมาจากหัวข้อที่ 3) แล้วป้อนเป็นอินพุตให้กับตัวจำแนก LBN และ LOGL โดยที่ลักษณะเด่นที่ถูกเลือกมาจากตัวประเมิน W\_LBN จะถูกป้อนให้กับตัวจำแนก LBN และลักษณะเด่นที่ถูกเลือกมาจาก W\_LOGL จะถูกป้อนให้กับตัวจำแนก LOGL
- 5) ฝึกสอนและทดสอบระบบ (ใช้เทคนิคการจำแนกแบบขั้นตอนเดียว เนื่องจาก การทดลองในหัวข้อ 4.2 พบว่าการจำแนกแบบขั้นตอนเดียวมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียวเมื่อทดสอบประสิทธิภาพโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้งโดยที่  $K=10$ ) โดย

เพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว เรียงลำดับลักษณะเด่นตามผลที่ได้จากหัวข้อที่ 3) วัดประสิทธิภาพในการจำแนกของตัวจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 คำนวณความแม่นยำเฉลี่ยในการจำแนกสำหรับข้อมูลของสัญญาณไฟฟ้ากล้ำมเนื้อทั้งหมด ในกระบวนการฝึกสอนและทดสอบระบบ ชุดข้อมูลสำหรับฝึกสอนและทดสอบระบบต้องเป็นชุดข้อมูลของสัญญาณไฟฟ้ากล้ำมเนื้อของอาสาสมัครคนเดียวกัน

6) พล็อตกราฟแสดงความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและผลความแม่นยำเฉลี่ยในการจำแนก พิจารณาจำนวนลักษณะเด่นที่เหมาะสมที่จะป้อนเป็นอินพุตให้กับตัวจำแนกอีกครั้ง โดยพิจารณาจากความแม่นยำเฉลี่ยในการจำแนกซึ่งคำนวณจากข้อมูลของสัญญาณไฟฟ้ากล้ำมเนื้อจากอาสาสมัครทั้งหมด

7) เปรียบเทียบประสิทธิภาพของตัวจำแนก 2 ชนิด และตัวประเมิน 2 ชนิดสำหรับการจำแนกวรรณยุกต์ 5 เสียง

#### (ค) ผลการทดลองและวิเคราะห์ผลการทดลอง

รูปที่ 4.15 แสดงกราฟความสัมพันธ์ระหว่างจำนวนของลักษณะเด่นและค่าการประเมิน เมื่อใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยเปรียบเทียบระหว่างการใช้ตัวประเมิน W\_LBN และ W\_LOGL สำหรับสัญญาณไฟฟ้ากล้ำมเนื้อของอาสาสมัคร S04 ในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่น จำนวนลักษณะเด่นจะถูกเพิ่มครั้งละ 1 ตัว ก่อนที่เซตย่อยของลักษณะเด่นจะถูกประเมินด้วยตัวประเมิน พิจารณาผลการเลือกลักษณะเด่นของตัวประเมิน W\_LBN จะเห็นว่าผลการประเมินจะเพิ่มขึ้นอย่างรวดเร็วในช่วงแรกของการเพิ่มจำนวนลักษณะเด่นและจะให้ผลการประเมินสูงสุดเมื่อใช้จำนวนลักษณะเด่น 18 ตัว (93.98%) หลังจากนั้นผลการประเมินจะค่อยๆ คงที่ไปเรื่อยๆ และจะลดต่ำลงอย่างรวดเร็วเมื่อเพิ่มลักษณะเด่นตัวที่ 107 เป็นต้นไป ซึ่งลักษณะเด่นในกลุ่มนี้เป็นลักษณะเด่นชนิด SM2 SM3 และ SM5 ซึ่งเป็นลักษณะเด่นชุดเดียวกับการจำแนกคำพูดหลัก 11 คำซึ่งให้ผลการจำแนกต่ำมาก สำหรับผลการประเมินของตัวประเมิน W\_LOGL จะเห็นว่าให้ค่าการประเมินใกล้เคียงกับตัวประเมิน W\_LBN แต่ผลการประเมินของ W\_LOGL จะให้ผลการประเมินสูงสุดเมื่อใช้จำนวนลักษณะเด่นน้อยกว่า กล่าวคือ ให้ผลการประเมินสูงสุด 93.88% เมื่อใช้จำนวนลักษณะเด่นเพียง 14 ตัว และค่าการประเมินจะเริ่มลดลงเมื่อเพิ่มลักษณะเด่นตัวที่ 20 เป็นต้นไป และจะลดลงเรื่อยๆ อย่างช้าๆ และค่อยๆ คงที่

อย่างไรก็ตามจะพิจารณาจำนวนลักษณะเด่นที่เหมาะสมอีกครั้งโดยการนำลักษณะเด่นที่ได้จากการเลือกซึ่งพิจารณาเฉพาะสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S04 อย่างเดียว ไปทดสอบกับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครอีก 9 คน โดยจะพิจารณาเฉพาะลักษณะเด่น 20 ตัวแรกเท่านั้น โดยปรับเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว หากความแม่นยำเฉลี่ยในการจำแนกโดยใช้ตัวจำแนก LBN และ LOGL โดยลักษณะเด่นที่ถูกเลือกโดยใช้ตัวประเมิน W\_LBN จะถูกจำแนกโดยใช้ตัวจำแนก LBN และ ลักษณะเด่นที่ถูกเลือกโดยใช้ตัวประเมิน W\_LOGL จะถูกจำแนกโดยใช้ตัวจำแนก LOGL ซึ่งผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียง แสดงดังรูปที่ 4.16

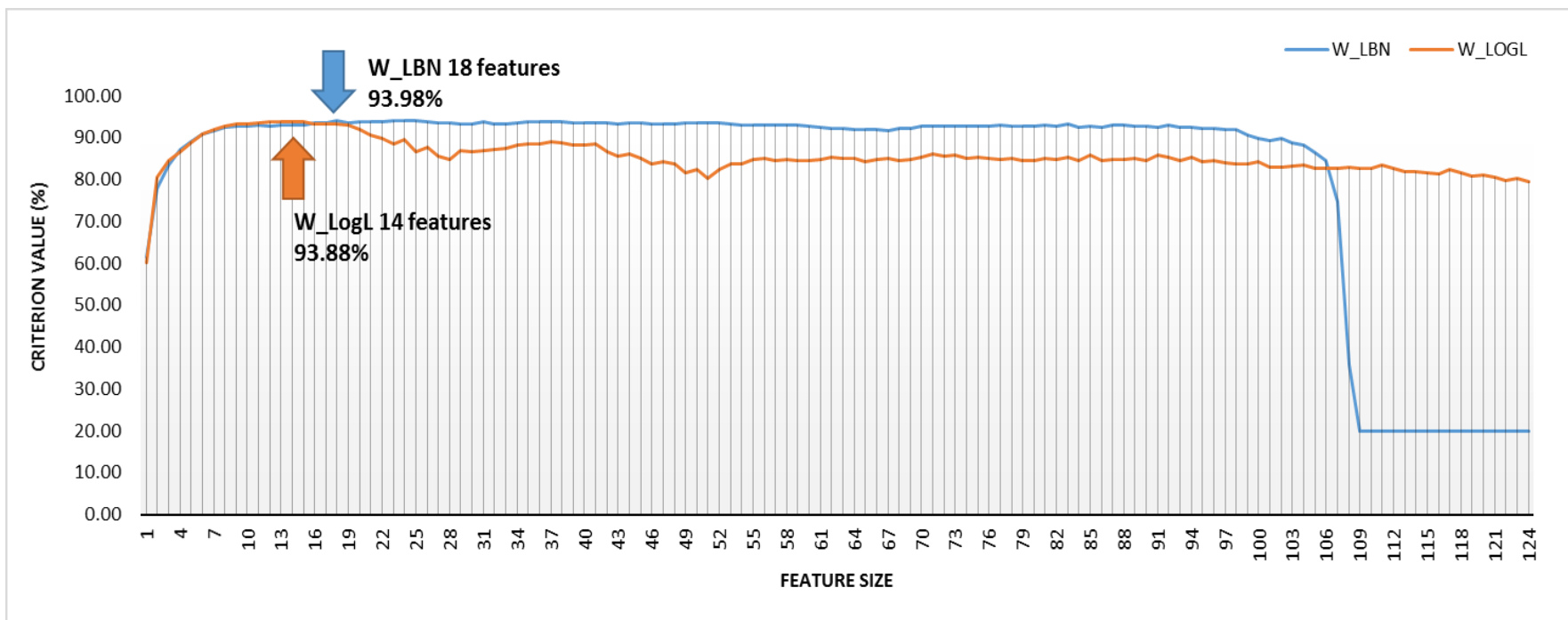
รูปที่ 4.16 แสดงผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้ง 10 คน เมื่อเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว สำหรับการใช้ตัวประเมินและตัวจำแนก 2 ชนิด พิจารณาผลความแม่นยำเฉลี่ยในการจำแนกของตัวจำแนก LBN ให้ประสิทธิภาพสูงสุด 72.73% เมื่อใช้จำนวนลักษณะเด่น 20 ตัว สำหรับตัวจำแนก LOGL ให้ความแม่นยำเฉลี่ยสูงสุด 72.14% เมื่อใช้จำนวนลักษณะเด่น 11 ตัว จะเห็นได้ว่าความแม่นยำเฉลี่ยในการจำแนกของตัวจำแนก 2 ชนิดแตกต่างกันเพียงเล็กน้อยเท่านั้น อย่างไรก็ตามตัวจำแนก LBN ใช้เวลาในการประมวลผลน้อยกว่า ถึงแม้ว่าจะใช้จำนวนลักษณะเด่นมากกว่า LOGL ก็ตาม (ตารางที่ 4.13) ดังนั้นการเลือกใช้เทคนิคการเลือกลักษณะเด่นโดยใช้เทคนิคการค้นหาแบบ FW ซึ่งใช้ W\_LBN เป็นตัวประเมินและจำแนกโดยตัวจำแนก LBN มีความเหมาะสมมากกว่า

**ตารางที่ 4.13** เวลาในการฝึกสอนและทดสอบระบบสำหรับการใช้ตัวจำแนกสองชนิด

เทคนิคที่ใช้	เวลาในการประมวลผล (Second)	ความแม่นยำเฉลี่ย (%)
20 features, LBN classifier	0.875	72.73±7.29
11 feature, LOGL classifier	3.890	72.14±7.28

**ตารางที่ 4.14** ลักษณะเด่นที่ถูกเลือกในแต่ละรอบของการค้นหาเซตย่อยของลักษณะเด่นด้วยวิธี FW และผลการประเมินเซตย่อยของลักษณะเด่นในแต่ละรอบด้วยวิธีการใช้ตัวประเมิน W\_LBN

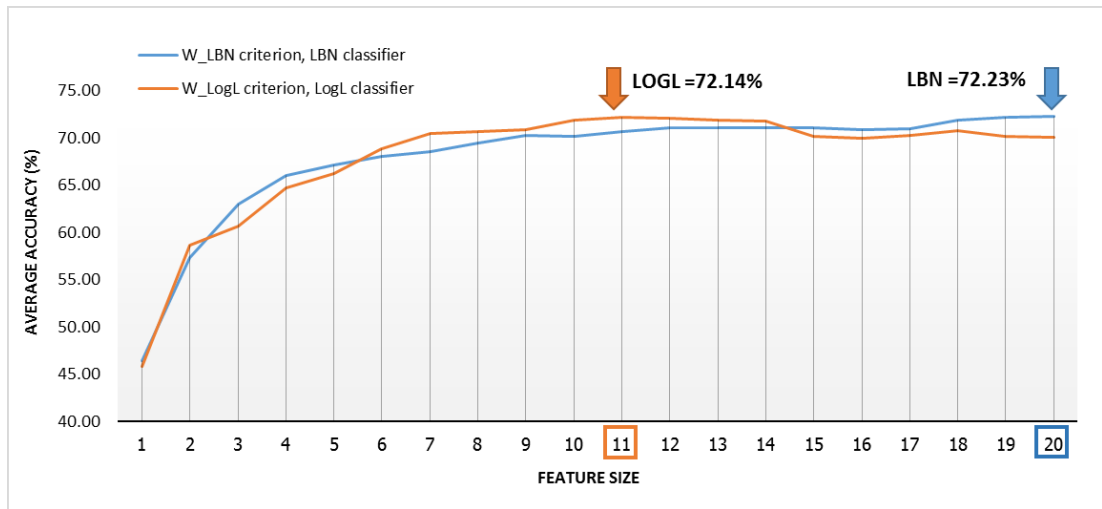
รอบในการค้นหาเซต ย่อยของลักษณะเด่น (จำนวนลักษณะเด่น)	W_LBN		W_LOGL	
	ลักษณะเด่นที่ถูก เพิ่มในแต่ละรอบ	ค่าการ ประเมิน (%)	ลักษณะเด่นที่ถูก เพิ่มในแต่ละรอบ	ค่าการ ประเมิน (%)
1	MMAV2	61.53	MMAV2	60.26
2	MMAV1	77.83	MMAV1	80.55
3	SSC	83.48	MAVS_3	84.49
4	IAV	87.13	SSC_3	86.69
5	WAMP_3	88.99	WL_1	88.82
6	RMS_4	90.74	SSC	90.75
7	MNF_4	91.78	WL	91.82
8	ZC	92.36	VAR_3	92.62
9	RMS	92.64	IAV_2	93.29
10	MDF_4	92.62	SSC_6	93.29
11	WL_4	92.86	VAR	93.56
12	MMAV1_1	92.76	VAR_1	93.66
13	MAV_4	93.11	SM2_5	93.86
14	MMAV2_1	92.94	<b>MAV_6</b>	<b>93.88</b>
15	MAV	92.88	WAMP_3	93.78
16	MN_2	93.58	ZC_1	93.31
17	VAR_1	93.46	MDF_6	93.23
18	<b>SSC_6</b>	<b>93.98</b>	WL_3	93.13
19	MMAV1_6	93.64	VAR_6	92.93
20	MMAV1_4	93.66	SM1_4	91.86



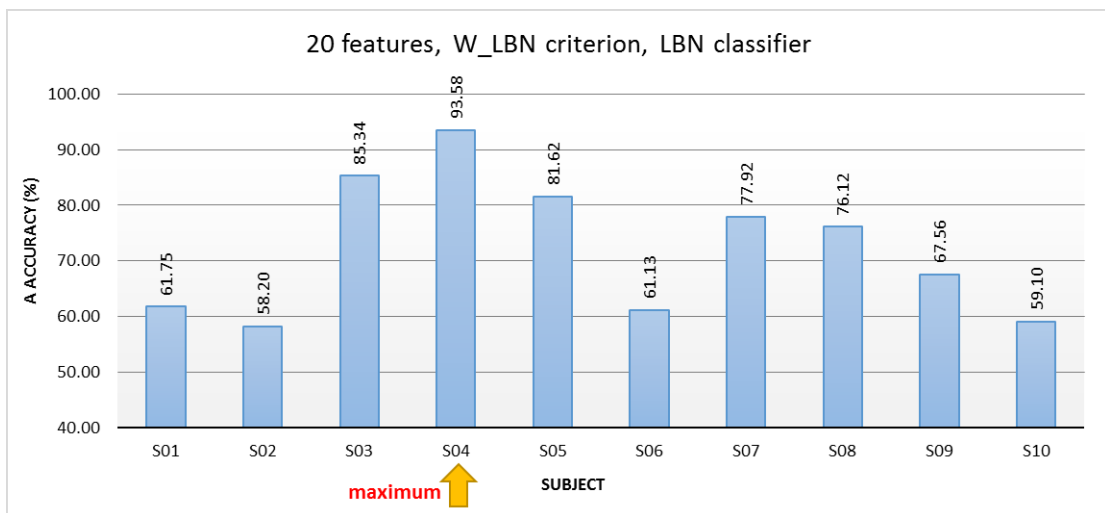
รูปที่ 4.15 ผลการประเมินและเลือกลักษณะเด่น (124 ตัว) โดยใช้เทคนิคการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN และ W\_LOGL สำหรับสัญญาณ

S04





รูปที่ 4.16 ผลความแม่นยำเฉลี่ยในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้ง 10 คน เมื่อเพิ่มจำนวนลักษณะเด่นครั้งละ 1 ตัว สำหรับการใช้ตัวประเมินและตัวจำแนก 2 ชนิด



รูปที่ 4.17 ผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน โดยพิจารณาจากการใช้จำนวนลักษณะเด่น 20 ตัว ซึ่งคัดเลือกจากวิธีการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ W\_LBN เป็นตัวประเมินและจำแนกโดยใช้ LBN

### (ง) สรุปผลการทดลอง

การทดลองนี้นำเสนอการคัดเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมในการจำแนกวรรณยุกต์ 5 เสียง โดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW โดยเปรียบเทียบการใช้ตัวประเมิน 2 ชนิด คือ W\_LBN และ W\_LOGL ซึ่งจำแนกโดยใช้ตัวจำแนก LBN และ LOGL จากการทดลองพบว่า การใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยตัวจำแนก LBN มีความเหมาะสมมากกว่า เนื่องจากประสิทธิภาพในการจำแนกใกล้เคียงกัน แต่การใช้ LBN ใช้เวลาในการประมวลผลน้อยกว่ามาก ซึ่งในการทดลองนี้ให้ความแม่นยำเฉลี่ยสูงสุด คือ 72.23%

### 4.7 สรุปผลการปรับปรุงระบบ

ตารางที่ 4.15 และ ตารางที่ 4.16 แสดงผลสรุปการปรับปรุงระบบโดยเปรียบเทียบกับระบบที่ได้ออกแบบไว้ในบทที่ 3 ซึ่งใช้เทคนิคการประเมินลักษณะเด่นแบบ FR โดยใช้ตัวประเมิน RES และใช้ตัวจำแนก ANN โดยจะพิจารณาความแม่นยำเฉลี่ยในการจำแนกซึ่งคำนวณจากสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน โดยการวัดประสิทธิภาพในการจำแนกใช้วิธีเทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 เทคนิคที่ใช้ในการปรับปรุงระบบ คือ การเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW และใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN ซึ่งสามารถปรับปรุงความแม่นยำเฉลี่ย 6.33% และ 4.77% สำหรับการจำแนกคำพูดหลัก 11 คำ แบบออกเสียงและไม่ออกเสียงตามลำดับ สำหรับการจำแนกวรรณยุกต์สามารถปรับปรุงความแม่นยำเฉลี่ย 4.57% เมื่อพิจารณาความแม่นยำเฉลี่ยซึ่งคำนวณจากความแม่นยำของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน จะเห็นได้ว่าการจำแนกวรรณยุกต์ประสิทธิภาพยังไม่ดีนัก แต่เมื่อพิจารณาความแม่นยำจากสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครที่ให้ผลการจำแนกที่ดีที่สุด (S04) จะเห็นได้ว่าให้ผลความแม่นยำที่ดีใกล้เคียงกับการจำแนกคำพูดหลัก 11 คำ นอกจากการปรับปรุงประสิทธิภาพในการจำแนกของระบบสำหรับการใช้เทคนิคคือ การเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW และใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN การใช้เทคนิคนี้ยังใช้เวลาในการประมวลผลน้อยกว่าเทคนิคเดิมอย่างมาก ซึ่งใช้ ANN ในการจำแนก ดังนั้นงานวิจัยนี้จึงเลือกใช้วิธีนี้ในการประมวลผลสัญญาณไฟฟ้ากล่อมเนื้อเพื่อจำแนกคำพูดหลัก 11 คำ และจำแนกวรรณยุกต์ 5 เสียง ซึ่งจะใช้เทคนิคนี้ในการรวมระบบและสร้างแบบจำลองการทำงานของระบบที่จะกล่าวในบทต่อไป

ในส่วนของการสร้างแบบจำลองการทำงานของระบบ จะนำโมเดลการจำแนกคำพูดหลัก 11 คำ และโมเดลการจำแนกวรรณยุกต์ 5 เสียง ซึ่งฝึกสอนระบบโดยการให้ตัวจำแนก LBN มาใช้ในการสร้างแบบจำลองการทำงานของระบบ โดยลักษณะเด่นที่ใช้เป็นอินพุตให้กับตัวจำแนกสำหรับโมเดลจำแนกคำพูดหลักจะใช้ลักษณะเด่น 8 ตัวตามที่ได้ออกแบบไว้ในหัวข้อ 4.5 สำหรับโมเดลการจำแนกวรรณยุกต์จะใช้ลักษณะเด่น 20 ตัวตามที่ได้ออกแบบไว้ในหัวข้อ 4.6 ซึ่งเป็นลักษณะเด่นที่ถูกคัดเลือกโดยใช้เทคนิคการเลือกลักษณะเด่นแบบ FSS ซึ่งใช้วิธีการค้นหาแบบ FW โดยใช้ตัวประเมิน W\_LBN

**ตารางที่ 4.15** สรุปผลการปรับปรุงระบบ (ความแม่นยำเฉลี่ยของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 10 คน)

เทคนิค	ความแม่นยำเฉลี่ยในการจำแนก (10 folds cross validation, average 10 subjects)		
	คำพูดหลัก 11 คำ		วรรณยุกต์ 5 เสียง
	ออกเสียง	ไม่ออกเสียง	
FR, W_RES, ANN	87.82±2.49	77.23±3.84	68.16±7.66
FSS, FW, W_LBN, LBN	94.15±2.62	82.00±3.60	72.73±7.29

**ตารางที่ 4.16** สรุปผลการปรับปรุงระบบ (ความแม่นยำสูงสุด)

เทคนิค	ความแม่นยำสูงสุดในการจำแนก (10 folds cross validation, the best subject)		
	คำพูดหลัก 11 คำ		วรรณยุกต์ 5 เสียง
	ออกเสียง	ไม่ออกเสียง	
FR, W_RES, ANN	93.33 (S09)	88.36 (S09)	88.71 (S04)
FSS, FW, W_LBN, LBN	99.64 (S08)	95.09 (S01)	93.58 (S04)

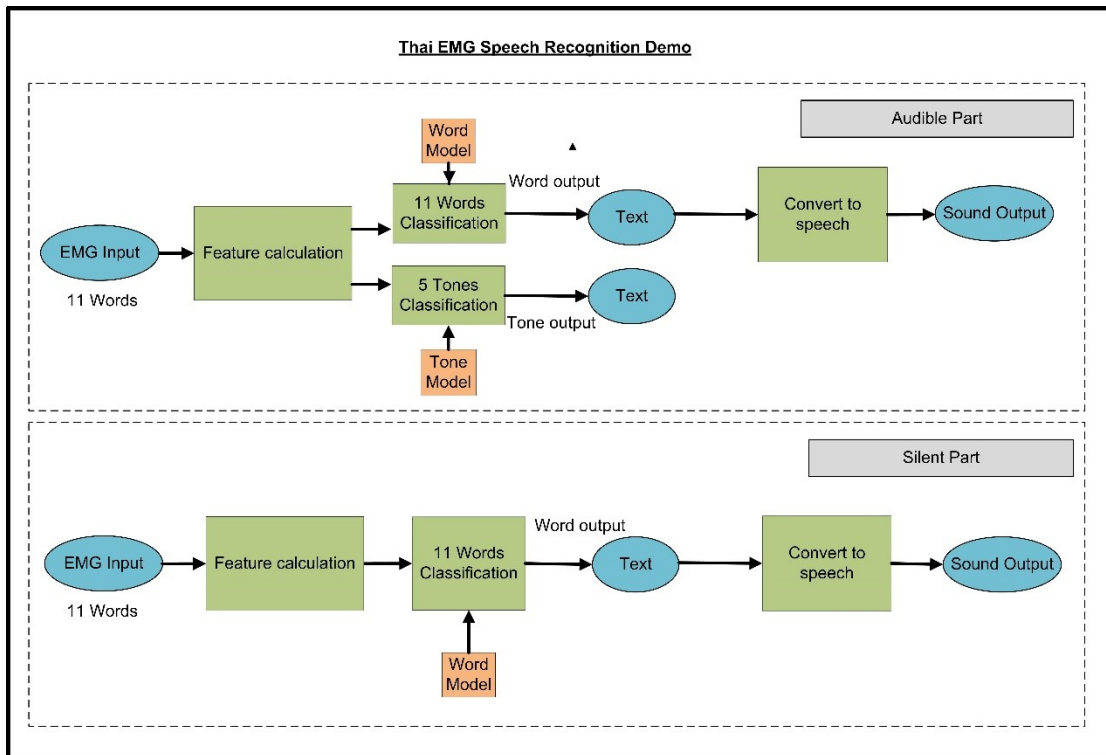
## บทที่ 5

### การสร้างแบบจำลองการทำงานของระบบ

ในบทนี้จะกล่าวถึงการสร้างแบบจำลองการทำงานของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งในบทที่ผ่านมาได้กล่าวถึงการออกแบบในส่วนของระบบรู้จำ ซึ่งได้ออกแบบระบบการจำแนกคำพูดหลัก 11 คำ และระบบการจำแนกวรรณยุกต์ 5 เสียง ซึ่งในการประยุกต์ใช้งานจริง อาจจะต้องแปลงเอาต์พุตที่ได้จากระบบรู้จำไปเป็นเสียงพูดที่ใช้ในการสื่อสาร ซึ่งกระบวนการนี้เรียกว่ากระบวนการสังเคราะห์เสียง

การสร้างแบบจำลองการทำงานของระบบ จะจำลองการทำงานของระบบโดยใช้โปรแกรม MATLAB ซึ่งจะสร้างฟังก์ชัน "ThaiEMGSpeechDemo" เพื่อใช้ในการทดสอบการทำงานของระบบ โดยจะป้อนอินพุตของสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูดที่ใช้สำหรับการทดสอบ โปรแกรมจะทำการประมวลผลและแสดงผลการจำแนกเป็นตัวอักษร หลังจากนั้นจะทำการแปลงผลลัพธ์ที่ได้จากระบบจำแนกไปเป็นเสียงพูดที่ใช้ในการสื่อสาร ซึ่งการแปลงจะใช้วิธีการแปลงข้อความเป็นเสียงพูด การแปลงวิธีนี้เป็นเทคนิคหนึ่งของการสังเคราะห์เสียงโดยอาศัยกฎ ซึ่งโดยปกติแล้ววิธีการแปลงข้อความเป็นเสียงพูดจะประกอบด้วย 2 ส่วนหลัก [98] ในส่วนแรกจะเป็นการนอร์มัลไลซ์ข้อความ ซึ่งประกอบด้วย การวิเคราะห์ข้อความ และส่วนประมวลผลจังหวะและทำนอง สำหรับในส่วนที่สองจะเป็นส่วนสังเคราะห์เสียงพูด ซึ่งแสดงดังรูปที่ 2.13 อย่างไรก็ตาม เนื่องจากงานวิจัยนี้ต้องการจำแนกคำพูดโดดเพียง 11 คำ ซึ่งเอาต์พุตของคำพูดมีเพียง 11 คำ และไม่ได้เป็นคำพูดต่อเนื่อง ดังนั้นไม่จำเป็นต้องมีส่วนประมวลผลจังหวะและทำนอง สำหรับกระบวนการแปลงข้อความเป็นเสียงพูดในงานวิจัยนี้จะเก็บข้อมูลสัญญาณเสียงพูดของเสียงพูดทั้ง 11 คำเพื่อเป็นคลังข้อมูลสำหรับเอาต์พุตของเสียงพูด ทำการวิเคราะห์เอาต์พุตที่ได้จากระบบรู้จำ และแสดงผลสัญญาณเสียงพูดให้ตรงกับเอาต์พุตที่ได้จากระบบรู้จำ

ในการสร้างแบบจำลองจะแบ่งระบบออกเป็น 2 ส่วน คือ แบบจำลองสำหรับการพูดแบบออกเสียง และ แบบจำลองสำหรับการพูดแบบไม่ออกเสียง ดังรูปที่ 5.1 สำหรับการพูดแบบออกเสียง จะทำการจำแนกเสียงวรรณยุกต์ของคำนั้นๆ ด้วย แต่สำหรับการพูดแบบไม่ออกเสียง จะจำแนกเฉพาะคำพูดหลัก 11 คำเท่านั้น เนื่องจาก งานวิจัยนี้ยังไม่สามารถออกแบบระบบการจำแนกวรรณยุกต์สำหรับการพูดแบบไม่ออกเสียงได้



รูปที่ 5.1 รูปแบบการทำงานของแบบจำลองระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

ในส่วน of แบบจำลองสำหรับการพูดแบบออกเสียง จะจำลองการทำงานโดยทดสอบป้อนสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูด 11 คำพูด ได้แก่ คำพูด "ศูนย์" ถึง "สิบ" เมื่อป้อนอินพุตเข้ามา ระบบจะคำนวณลักษณะเด่นของสัญญาณตามที่ได้ออกแบบไว้ในบทที่ 4 และจะทำการจำแนกสัญญาณโดยใช้โมเดลที่ได้ฝึกสอนระบบไว้แล้วซึ่งประกอบด้วย 2 โมเดล คือ โมเดลการจำแนกคำพูดหลัก 11 คำ (word model) และโมเดลการจำแนกวรรณยุกต์ (tone model) โปรแกรมจะแสดงเอาต์พุตออกมาเป็น 2 ส่วน คือ ข้อความของคำพูดที่ได้ และ ข้อความวรรณยุกต์ของคำนั้น โดยเอาต์พุตในส่วนแรกจะถูกแปลงไปเป็นเสียงพูดเพื่อใช้ในการสื่อสารสำหรับในส่วน of แบบจำลองการพูดแบบไม่ออกเสียง จะคล้ายกับแบบจำลองการพูดแบบออกเสียง แต่จะไม่มี การจำแนกวรรณยุกต์ของคำนั้นๆ

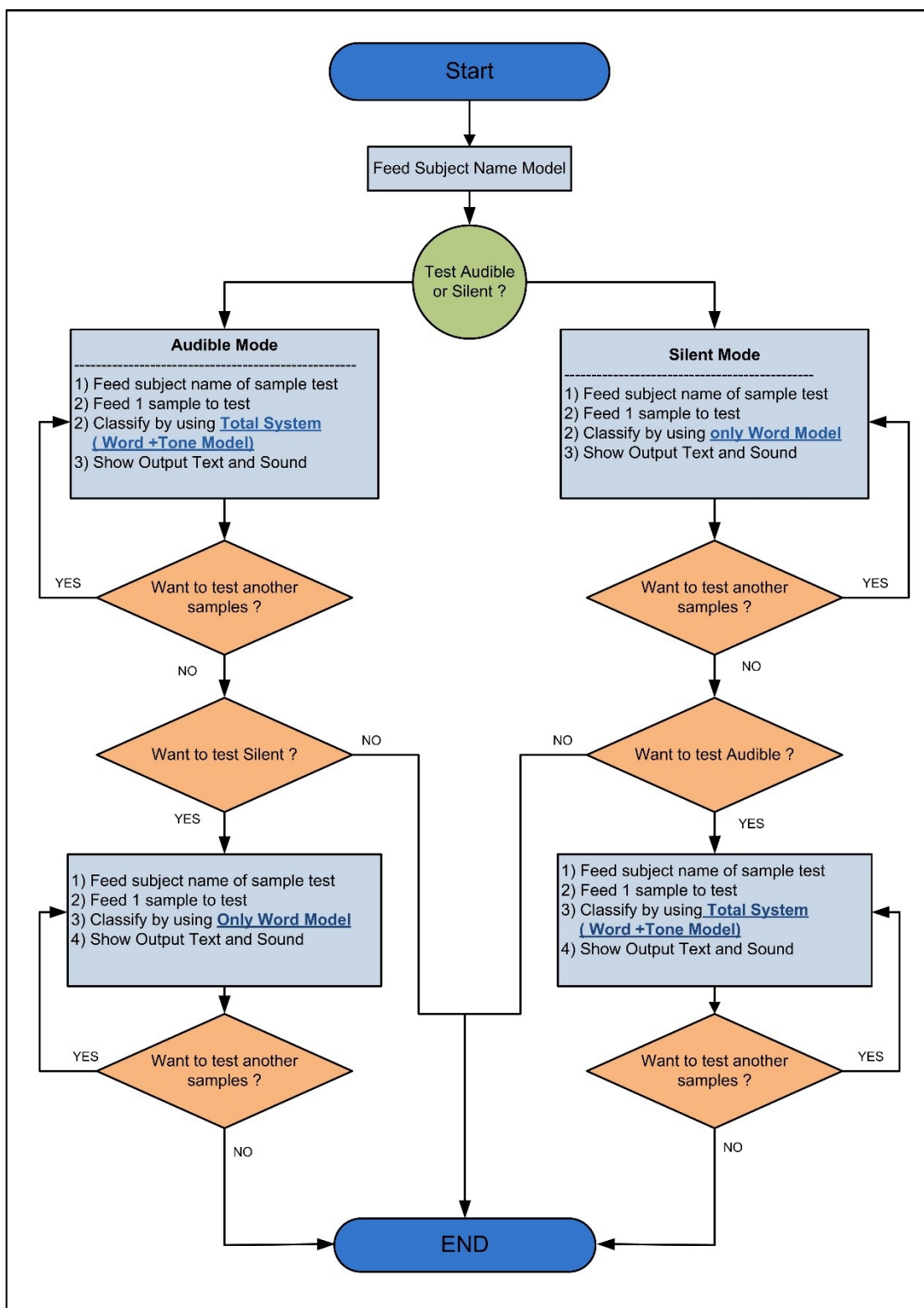
โมเดลในการทดสอบที่ผ่านการฝึกสอนระบบแล้วโดยใช้ตัวจำแนก LBN ประกอบด้วย

1) โมเดลการจำแนกคำพูดหลักแบบออกเสียง 10 โมเดล ซึ่งเป็นโมเดลของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน (S01-S10)

2) โมเดลการจำแนกคำพูดหลักแบบไม่ออกเสียง 10 โมเดล ซึ่งเป็นโมเดลของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน (S01-S10)

3) โมเดลการจำแนกวรรณยุกต์ 5 เสียงแบบออกเสียง 10 โมเดล ซึ่งเป็นโมเดลของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคน (S01-S10)

แผนผังแสดงการทำงานของโปรแกรมแบบจำลองระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อแสดงดังรูปที่ 5.2 โดยเริ่มจากการป้อนชื่อโมเดลที่จะใช้ทดสอบ เช่น "S01" คือ โมเดลของอาสาสมัครคนที่ 1 ซึ่งผ่านการฝึกสอนระบบแล้ว หลังจากนั้นเลือกว่าต้องการทดสอบแบบจำลองการพูดแบบออกเสียงหรือไม่ออกเสียง ป้อน "A" คือ แบบจำลองแบบออกเสียง ป้อน "S" คือ แบบจำลองแบบไม่ออกเสียง ถ้าต้องการทดสอบแบบจำลองแบบออกเสียง ระบบจะเข้าสู่ "audio mode" และจะเริ่มทำการจำแนกโดยป้อนข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อที่จะทดสอบ ซึ่งสามารถทดสอบโดยใช้ข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน ขั้นตอนแรก ป้อนชื่อของข้อมูลอาสาสมัครที่ต้องการจะทดสอบ เช่น "S10" คือ ต้องการทดสอบสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครคนที่ 10 ต่อมาเลือกสัญญาณที่จะทดสอบ ในที่นี้ข้อมูลของอาสาสมัครแต่ละคนที่ใช้ในการทดสอบจะมีทั้งหมด 110 ข้อมูล ถ้าพิมพ์ "1" คือเลือกสัญญาณที่ 1 มาทดสอบ หลังจากนั้นโปรแกรมจะทำการประมวลผล โดยถ้าอยู่ในโหมดการพูดแบบออกเสียง เอาต์พุตที่ได้จะแสดงข้อความของคำพูดและวรรณยุกต์ของคำนั้น ถ้าอยู่ในโหมดการพูดแบบไม่ออกเสียง เอาต์พุตที่ได้จะแสดงข้อความของคำพูดเท่านั้น หลังจากนั้นโปรแกรมจะแปลงเอาต์พุตที่ได้จากระบบรู้จำไปเป็นเสียงพูด รูปที่ 5.3 แสดงตัวอย่างผลการทำงานของโปรแกรมจำลองการทำงานของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ



รูปที่ 5.2 แผนผังแสดงการทำงานของแบบจำลองระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล่อมเนื้อ

```

>> ThaiEMGspeechDemo('H')
#####
Now we are testing ThaiEMGspeechDemo to classify Thai words from EMG signal for AUDIBLE and SILENT speech
#####
Please Feed Speaker Name of Model using for this classification ('S01'or 'S02' or ... 'S10')
'S01'
Do you want to classify Audible speech or Silent speech? ('A' or 'S')
'A'
*****
*****START AUDIBLE EMG Speech Classification*****
*****
Now we are testing AUDIBLE ThaiEMGspeechDemo to classify 11 Thai words including 5 tones
We have 110 samples for each speaker(10 speakers)
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S01'
Please Feed Sample test (1-110)
1
-----START CLASSIFY-(Speaker S01 Model)-----

*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S01 for Sample <<< กุ้ย >>>

The result of Word classification is " กุ้ย " ---> Target is " กุ้ย " ***CORRECTLY CLASSIFY WORD***
The result of Tone classification is " Rising Tone " ---> Target is " Rising Tone " ***CORRECTLY CLASSIFY TONE***

-----END CLASSIFY-----

Do you want to test other samples? ('y' or 'n')
'y'
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S08'
Please Feed Sample test (1-110)
9
-----START CLASSIFY-(Speaker S01 Model)-----

*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S08 for Sample <<< กุ้ย >>>

The result of Word classification is " เก้า " ---> Target is " กุ้ย " ***!INCORRECTLY CLASSIFY WORD***
The result of Tone classification is " Rising Tone " ---> Target is " Rising Tone " ***CORRECTLY CLASSIFY TONE***

-----END CLASSIFY-----

Do you want to test other samples? ('y' or 'n')
'n'
*****
*****FINISH AUDIBLE EMG Speech Classification*****
*****
Do you want to test SILENT Thai EMG speech Classification? ('y' or 'n')
'y'
*****
*****START SILENT EMG Speech Classification*****
*****
Now we are testing SILENT ThaiEMGspeechDemo to classify 11 Thai words
We have 110 samples for each speaker(10 speakers)
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S01'
Please Feed Sample test (1-110)
1
-----START CLASSIFY-(Speaker S01 Model)-----

*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S01 for Sample <<< กุ้ย >>>

The result of Word classification is " กุ้ย " ---> Target is " กุ้ย " ***CORRECTLY CLASSIFY WORD***
***The Final---> CORRECTLY CLASSIFY***

-----END CLASSIFY-----

```





```

Do you want to test other samples? ('y' or 'n')
'y'
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S01'
Please Feed Sample test (1-110)
80
-----START CLASSIFY-(Speaker S01 Model)-----
*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S01 for Sample <<< เจ็ด >>>
The result of Word classification is " เจ็ด " ---> Target is " เจ็ด " ***CORRECTLY CLASSIFY WORD***
***The Final---> CORRECTLY CLASSIFY***
-----END CLASSIFY-----
Do you want to test other samples? ('y' or 'n')
'y'
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S01'
Please Feed Sample test (1-110)
110
-----START CLASSIFY-(Speaker S01 Model)-----
*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S01 for Sample <<< สิบ >>>
The result of Word classification is " สิบ " ---> Target is " สิบ " ***CORRECTLY CLASSIFY WORD***
***The Final---> CORRECTLY CLASSIFY***
-----END CLASSIFY-----
Do you want to test other samples? ('y' or 'n')
'y'
Please Feed Speaker Name of sample to test ('S01'or 'S02' or ... 'S10')
'S07'
Please Feed Sample test (1-110)
45
-----START CLASSIFY-(Speaker S01 Model)-----
*****
* Classification based on Speaker S01 Model *
*****
Now feeding EMG input sample of speaker S07 for Sample <<< สี่ >>>
The result of Word classification is " สิบ " ---> Target is " สี่ " ***!INCORRECTLY CLASSIFY WORD***
***The Final---> !INCORRECTLY CLASSIFY***
-----END CLASSIFY-----
Do you want to test other samples? ('y' or 'n')
'n'
*****
*****FINISH SILENT EMG Speech Classification*****
*****
#####
##### FINISH Testing ThaiEMGspeechDemo #####
#####
>>

```

รูปที่ 5.3 แสดงตัวอย่างผลการการทำงานของโปรแกรมจำลองการทำงานของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

## บทที่ 6 บทสรุปและข้อเสนอแนะ

### 6.1 สรุปผลการวิจัย

งานวิจัยนี้สร้างต้นแบบของระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ เพื่อรู้จำคำไทยที่เป็นคำโดด และเป็นเสียงพูดของภาษาไทยภาคกลาง ซึ่งผ่านกระบวนการพูด 2 วิธี คือ การพูดแบบออกเสียง และการพูดแบบไม่ออกเสียง โดยระบบจะรู้จำคำพูดของตัวเลขไทย 11 คำ ได้แก่ "ศูนย์" "หนึ่ง" "สอง" "สาม" "สี่" "ห้า" "หก" "เจ็ด" "แปด" "เก้า" และ "สิบ" และ รู้จำวรรณยุกต์ของแต่ละคำพูดด้วย ซึ่งวรรณยุกต์ไทยประกอบด้วย 5 เสียง ได้แก่ "สามัญ" "เอก" "โท" "ตรี" และ "จัตวา" โดยการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจากผู้ทดลองหรืออาสาสมัครซึ่งใช้ภาษาไทยเป็นภาษาแม่ทั้งหมด 10 คน ซึ่งมีอายุระหว่าง 18-30 ปี

การออกแบบระบบรู้จำจะแบ่งออกเป็น 2 ส่วน คือ ส่วนของระบบการรู้จำคำพูดหลัก 11 คำพูด และ ส่วนของระบบการรู้จำวรรณยุกต์ 5 เสียง การออกแบบทั้งสองส่วนจะเริ่มจากการออกแบบตำแหน่งอิเล็กโทรดที่ใช้ในการเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด การออกแบบอุปกรณ์และวิธีการเก็บข้อมูลของสัญญาณ ซึ่งตำแหน่งที่ออกแบบมีทั้งหมด 5 ตำแหน่ง (5 ช่องสัญญาณ) ประกอบด้วยกล้ามเนื้อหน้าและใต้คาง 3 ตำแหน่ง และกล้ามเนื้อคอ 2 ตำแหน่ง ซึ่งกล้ามเนื้อคอจะใช้ในการรู้จำวรรณยุกต์เพียงอย่างเดียว สำหรับกระบวนการเตรียมสัญญาณเบื้องต้น ในส่วนของการกรองสัญญาณรบกวนจะใช้ตัวกรองชนิดบัตเตอร์เวิร์คอันดับที่ 5 ซึ่งจะใช้ทั้งตัวกรองความถี่ผ่านและตัวกรองนอตช์สำหรับการกำจัดและลดทอนสัญญาณรบกวน ในส่วนของการตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อมีการประยุกต์ใช้เทคนิคค่าขีดแบ่งของสามเท่าของค่าเบี่ยงเบนมาตรฐาน หลังจากการออกแบบในส่วนแรกทีกล่าวมาแล้วนั้น จะทำการประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อเพื่อออกแบบระบบรู้จำ ซึ่งประกอบด้วยการสกัดลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ การคัดเลือกลักษณะเด่น การฝึกสอนและทดสอบระบบรู้จำ และการวัดประสิทธิภาพของระบบรู้จำ โดยในการออกแบบเบื้องต้นจะใช้เทคนิคการเลือกลักษณะเด่นแบบ FR โดยใช้ตัวประเมินชนิด RES ซึ่งเป็นเทคนิคที่ง่ายและมีประสิทธิภาพในการเลือกลักษณะเด่น สำหรับตัวจำแนกจะใช้ตัวจำแนกชนิด ANN ซึ่งเป็นเทคนิคที่มีประสิทธิภาพและนิยมใช้ในงานวิจัยทางด้านนี้ สำหรับการวัดประสิทธิภาพในการรู้จำหรือจำแนกระบบจะใช้เทคนิคการแบ่งข้อมูลแบบร้อยละ สำหรับการจำแนกวรรณยุกต์ได้ออกแบบระบบการจำแนกแบบสาม

ขั้นตอนโดยใช้ตัวจำแนก ANN สามโครงข่าย เพื่อจำแนกวรรณยุกต์ออกมาที่ละขั้นตอน ซึ่งพบว่ามีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียวซึ่งใช้ ANN เพียงโครงข่ายเดียว สำหรับผลการออกแบบเบื้องต้น การออกแบบระบบรู้จำคำพูดหลัก 11 คำ โดยใช้เทคนิคดังกล่าวให้ความแม่นยำเฉลี่ยในการจำแนก 92.45% และ 81.45% สำหรับการพูดแบบออกเสียงและไม่ออกเสียงตามลำดับ (คำนวณจากความแม่นยำของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน) ซึ่งใช้จำนวนลักษณะเด่น 9 ตัวเป็นอินพุตให้กับตัวจำแนก สำหรับการออกแบบระบบรู้จำวรรณยุกต์ได้ออกแบบรูปแบบการจำแนกแบบสามขั้นตอน โดยให้ความแม่นยำเฉลี่ย 75.78%

หลังจากการออกแบบระบบรู้จำในเบื้องต้นแล้วนั้น ได้มีการปรับปรุงระบบเพื่อเพิ่มความแม่นยำในการจำแนกและหาเทคนิคที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ โดยเริ่มจากการเลือกใช้ลักษณะเด่นในแกนเวลาแบบตัวกรองซ้อน โดยใช้ตัวจำแนก HMM ในการฝึกสอนและทดสอบระบบ โดยให้ผลความแม่นยำเฉลี่ยในการจำแนกคำพูดหลัก 11 คำ 92.64% สำหรับการพูดแบบออกเสียง และ 76.82% สำหรับการพูดแบบไม่ออกเสียง การจำแนกวรรณยุกต์ 5 เสียงให้ความแม่นยำเฉลี่ยสูงสุด 62.67% อย่างไรก็ตามจากผลการทดสอบระบบเทคนิคนี้ยังไม่สามารถปรับปรุงความแม่นยำของระบบได้ ทั้งนี้อาจจะเป็นเพราะว่าเทคนิคดังกล่าวถูกออกแบบมาสำหรับการจำแนกคำพูดแบบต่อเนื่อง นอกจากนี้เทคนิคดังกล่าวอาจจะไม่เหมาะสมกับงานวิจัยนี้ เนื่องจากงานวิจัยนี้จำนวนคำพูดที่ต้องการจำแนกน้อย และเป็นการจำแนกคำพูดโดดเท่านั้น ซึ่งการใช้เทคนิคนี้ทำให้เพิ่มความซับซ้อนให้กับระบบเกินความจำเป็น

ต่อมาได้ทำการเปรียบเทียบประสิทธิภาพของตัวจำแนก 8 ชนิด โดยทดสอบประสิทธิภาพในการจำแนกคำพูดหลัก 11 คำ และวรรณยุกต์ 5 เสียง เพื่อหาตัวจำแนกที่เหมาะสมที่สุดในงานวิจัยนี้ โดยการวัดประสิทธิภาพในการจำแนกของตัวจำแนกจะใช้เทคนิคการแบ่งข้อมูลแบบตรวจสอบไขว้ K ครั้ง ซึ่งค่า K เท่ากับ 10 จากผลการทดลองพบว่าตัวจำแนกที่ให้ผลการจำแนกที่ดีที่สุดสำหรับการจำแนกคำพูดหลัก 11 คำ คือ FLDA (การพูดแบบออกเสียง) และ LBN (การพูดแบบไม่ออกเสียง) สำหรับการจำแนกวรรณยุกต์ 5 เสียงพบว่าตัวจำแนก LOGL ให้ผลการจำแนกที่ดีที่สุด ซึ่งจะเห็นได้ว่าตัวจำแนกดังกล่าวให้ผลการจำแนกดีกว่าตัวจำแนก ANN และใช้เวลาในการประมวลผลน้อยกว่า นอกจากนี้ยังพบว่าเมื่อวัดประสิทธิภาพในการจำแนกโดยใช้เทคนิคการตรวจสอบไขว้ K ครั้ง พบว่าประสิทธิภาพในการจำแนกแบบขั้นตอนเดียวใกล้เคียงกับการจำแนกแบบสามขั้นตอน ดังนั้นการใช้เทคนิคการจำแนกแบบขั้นตอนเดียวมีความเหมาะสมมากกว่าเนื่องจากเป็นเทคนิคที่ง่าย ไม่ซับซ้อนและใช้เวลาในการประมวลผลน้อยกว่า

หลังจากนั้นทำการปรับปรุงระบบในส่วนของการคัดเลือกลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งในการออกแบบเบื้องต้นได้ใช้เทคนิคการเลือกลักษณะเด่นแบบ FR โดยใช้ตัวประเมิน RES เพียงชนิดเดียว เพื่อทดสอบว่าตัวประเมินดังกล่าวมีประสิทธิภาพเพียงพอและเหมาะสมในงานวิจัยนี้ จึงได้ทำการทดลองเพื่อเปรียบเทียบประสิทธิภาพของตัวประเมินลักษณะเด่น 14 ชนิด ซึ่งประกอบด้วยตัวประเมินที่ขึ้นอยู่กับตัวจำแนกและตัวประเมินที่ไม่ขึ้นอยู่กับตัวจำแนก นอกจากนั้นจะเปรียบเทียบประสิทธิภาพของตัวจำแนก 4 ชนิดที่ดีที่สุดจากผลการเปรียบเทียบตัวจำแนกจากการทดลองก่อนหน้านี้ สำหรับการจำแนกคำพูดหลัก 11 คำพูด จากการทดลองพบว่า การใช้ W\_FLDA เป็นตัวประเมินเพื่อเลือกลักษณะเด่น และ ตัวจำแนก LBN เพื่อจำแนกระบบให้ผลการจำแนกสูงสุด โดยที่ผลความแม่นยำเฉลี่ย คือ 93.25% สำหรับการพูดแบบออกเสียง และ 80.12% สำหรับการพูดแบบไม่ออกเสียง

หลังจากการเปรียบเทียบประสิทธิภาพของตัวประเมินทั้ง 14 ชนิด ได้มีการปรับปรุงระบบในส่วนของการคัดเลือกลักษณะเด่นโดยใช้เทคนิคการเลือกเซตย่อยของลักษณะเด่น (FSS) โดยผ่านกระบวนการค้นหา ซึ่งการเลือกลักษณะเด่นวิธีนี้จะทำให้สามารถกำจัดลักษณะเด่นที่มีความซ้ำซ้อนออกไปได้ โดยเปรียบเทียบเทคนิคการค้นหาเซตย่อยของลักษณะเด่น 3 ชนิด คือ FW BW และ LR และเปรียบเทียบการใช้ตัวประเมิน 4 ชนิด คือ F\_in-in F\_Maha-s W\_FLDA และ W\_LBN นอกจากนี้ทดสอบประสิทธิภาพของตัวจำแนก 2 ชนิด คือ LBN และ FLDA จากผลการทดลองพบว่า เทคนิคที่ให้ประสิทธิภาพสูงสุด คือ การเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ตัวประเมิน W\_LBN โดยให้ความแม่นยำเฉลี่ยสูงสุดในการจำแนกคำพูดหลัก 11 คำเมื่อจำแนกโดยใช้ตัวจำแนก LBN โดยมีความแม่นยำเท่ากับ 94% สำหรับการพูดออกเสียง และ 81.06% สำหรับการพูดแบบไม่ออกเสียง

จากการปรับปรุงระบบทั้งในส่วนของการคัดเลือกลักษณะเด่นและการจำแนกของตัวจำแนกทำให้ได้เทคนิคที่สามารถปรับปรุงระบบได้ดีที่สุดและใช้เวลาในการประมวลผลน้อย นั่นคือ การเลือกลักษณะเด่นแบบ FSS โดยใช้เทคนิคการค้นหาลักษณะเด่นแบบ FW ซึ่งใช้ตัวประเมิน W\_LBN และมีประสิทธิภาพดีที่สุดเมื่อจำแนกโดยตัวจำแนก LBN ดังนั้นงานวิจัยนี้ได้ใช้เทคนิคดังกล่าวสำหรับการเลือกลักษณะเด่นและจำนวนลักษณะเด่นที่เหมาะสมอีกครั้งสำหรับระบบการจำแนกทั้งสองส่วน สำหรับในส่วนของระบบการจำแนกคำพูดหลัก 11 คำ ทำการคัดเลือกลักษณะเด่นโดยวิธีดังกล่าวพบว่า การใช้จำนวนลักษณะเด่น 8 ตัวให้ความแม่นยำเฉลี่ยสูงสุด โดยให้ความแม่นยำเฉลี่ย 94.15% สำหรับการพูดแบบออกเสียง และ 82.00% สำหรับการ

พูดแบบไม่ออกเสียง สำหรับในส่วนของการจำแนกวรรณยุกต์ 5 เสียง การใช้ลักษณะเด่น 20 ตัวให้ความแม่นยำเฉลี่ยสูงสุด 72.23%

เมื่อเปรียบเทียบประสิทธิภาพในการจำแนกของระบบระหว่างก่อนการปรับปรุงระบบ (ใช้เทคนิคการประเมินลักษณะเด่นแบบ FR โดยใช้ตัวประเมิน RES และใช้ตัวจำแนก ANN) และหลังจากการปรับปรุง (โดยใช้การเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW และใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN) โดยจะพิจารณาความแม่นยำเฉลี่ยในการจำแนกซึ่งคำนวณจากสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน ซึ่งมีการวัดประสิทธิภาพในการจำแนกโดยใช้วิธีเทคนิคการตรวจสอบไขว้ K ครั้ง โดยที่ K เท่ากับ 10 พบว่าสามารถปรับปรุงความแม่นยำเฉลี่ย 6.33% และ 4.77% สำหรับการจำแนกคำพูดหลัก 11 คำแบบออกเสียงและไม่ออกเสียงตามลำดับ สำหรับการจำแนกวรรณยุกต์สามารถปรับปรุงความแม่นยำเฉลี่ย 4.57% เมื่อพิจารณาความแม่นยำเฉลี่ยซึ่งคำนวณจากความแม่นยำของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน จะเห็นได้ว่าการจำแนกวรรณยุกต์ประสิทธิภาพยังไม่ดีนัก แต่เมื่อพิจารณาความแม่นยำจากสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครที่ให้ผลการจำแนกที่ดีที่สุด (S04) จะเห็นได้ว่าให้ผลความแม่นยำที่ดีใกล้เคียงกับการจำแนกคำพูดหลัก 11 คำ นอกจากการปรับปรุงความแม่นยำในการจำแนกของระบบ การใช้เทคนิคดังกล่าว (การเลือกลักษณะเด่นแบบ FSS ซึ่งใช้เทคนิคการค้นหาแบบ FW และใช้ตัวประเมิน W\_LBN ซึ่งจำแนกโดยใช้ตัวจำแนก LBN) ยังใช้เวลาในการประมวลผลน้อยกว่าเทคนิคเดิมอย่างมาก (ซึ่งใช้ ANN ในการจำแนก) ดังนั้นงานวิจัยนี้จึงเลือกใช้วิธีนี้ในการประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อเพื่อจำแนกคำพูดหลัก 11 คำ และจำแนกวรรณยุกต์ 5 เสียง และใช้เทคนิคนี้ในการรวมระบบและสร้างแบบจำลองการทำงานจากระบบ สำหรับประสิทธิภาพในการจำแนกของระบบที่ได้ออกแบบไม่สามารถนำไปเปรียบเทียบกับงานวิจัยอื่นๆ ได้ เนื่องจากลักษณะของคำพูด จำนวนคำพูดที่ใช้ในการจำแนก จำนวนอาสาสมัครที่ใช้ในการทดสอบแตกต่างกัน ทำให้ไม่สามารถนำผลความแม่นยำในการจำแนกไปเปรียบเทียบได้โดยตรง

หลังจากได้ออกแบบและปรับปรุงในส่วนขอระบบรู้จำแล้วนั้น ได้ทำการรวมระบบและสร้างแบบจำลองการทำงานจากระบบขึ้นโดยใช้โปรแกรม MATLAB ซึ่งสร้างฟังก์ชัน "ThaiEMGSpeechDemo" เพื่อใช้ในการทดสอบการทำงานระบบ โดยจะป้อนอินพุตของสัญญาณไฟฟ้ากล้ามเนื้อของเสียงพูดที่ใช้สำหรับการทดสอบ โปรแกรมจะทำการประมวลผลและจะแสดงผลการจำแนกเป็นตัวอักษร (Text) หลังจากนั้นจะทำการแปลงผลลัพธ์ที่ได้จากระบบจำแนกไปเป็นเสียงพูดที่ใช้ในการสื่อสาร

## 6.2 ข้อจำกัดและแนวทางวิจัยในอนาคต

เนื่องจากในการวิจัยนี้เป็นเพียงต้นแบบระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อซึ่งสามารถจำแนกคำพูดตัวเลขไทย "ศูนย์" ถึง "สิบ" และสามารถจำแนกวรรณยุกต์ของคำนั้นๆ ซึ่งเป็นช่วงเริ่มต้นของงานวิจัยทางด้านนี้สำหรับระบบที่เป็นภาษาไทย ดังนั้นงานวิจัยนี้จึงมีข้อจำกัดอยู่หลายประการ ซึ่งจะได้กล่าวในหัวข้อนี้ โดยจะแนะนำแนวทางงานวิจัยในอนาคตเพื่อที่จะแก้ไขข้อจำกัดดังกล่าว โดยข้อจำกัดดังกล่าวจะแบ่งเป็นหัวข้อย่อยต่างๆ ดังนี้

### 1) ขอบเขตของคำพูด

งานวิจัยนี้รู้จำเฉพาะคำพูดภาษาไทย 11 คำ และรู้จำวรรณยุกต์ 5 เสียง ซึ่งเป็นจำนวนคำที่น้อยและเป็นการรู้จำคำโดด เหมาะสำหรับการประยุกต์ใช้งานที่ต้องการจำแนกคำพูดเพียงเล็กน้อย เช่น การบอกรหัสเอทีเอ็ม การบอกรหัสเพื่อควบคุมการทำงานของคอมพิวเตอร์ โดยใช้รหัส เป็นต้น

แนวทางการพัฒนาระบบในอนาคต ควรพัฒนาระบบให้สามารถรู้จำคำพูดได้จำนวนมากขึ้น เพื่อให้สามารถนำไปประยุกต์ใช้งานได้หลากหลาย โดยเฉพาะอย่างยิ่งเพื่อให้สามารถนำไปประยุกต์ใช้งานกับผู้ป่วยหรือผู้สูงอายุที่มีปัญหาการออกเสียงพูด ซึ่งจำเป็นต้องใช้คำพูดจำนวนมากในการสื่อสารและเป็นคำพูดต่อเนื่อง เพื่อให้ระบบสามารถรู้จำคำพูดได้จำนวนมากโดยที่ไม่จำเป็นต้องสร้างโมเดลการรู้จำตามจำนวนคำที่จำแนก งานวิจัยในอนาคตควรจะออกแบบระบบให้สามารถรู้จำหน่วยเสียง ซึ่งเป็นส่วนที่เล็กที่สุดของภาษา หน่วยเสียงแต่ละหน่วยเสียงจะประกอบกันเป็นคำพูด โดยการรู้จำหน่วยเสียงจะทำให้พัฒนาระบบไปเป็นระบบที่สามารถจำแนกคำพูดแบบต่อเนื่องได้

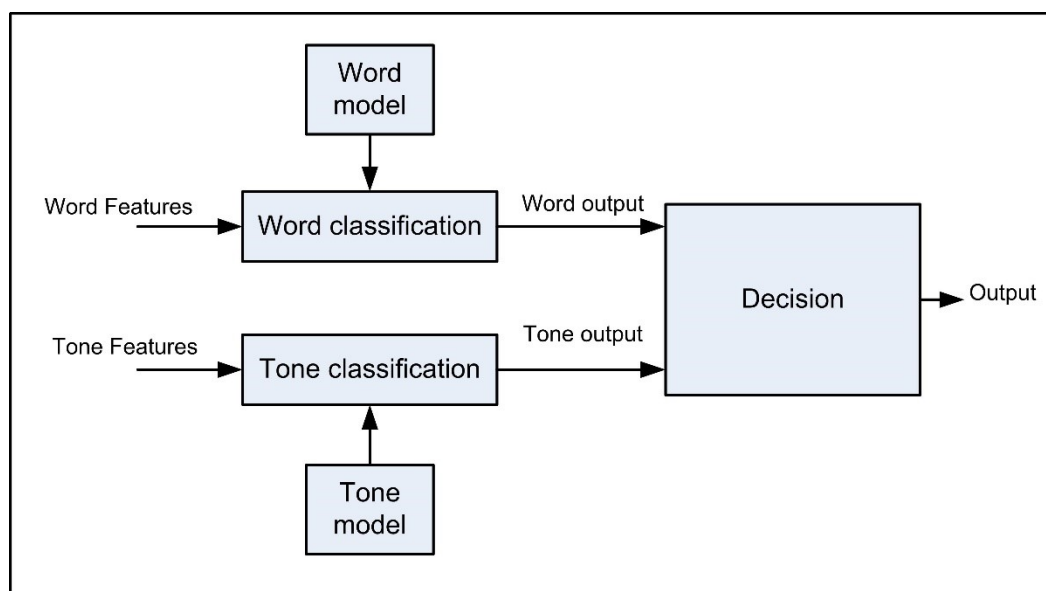
### 2) การฝึกสอนระบบเพื่อสร้างโมเดลการจำแนกและทดสอบระบบ

ในการออกแบบระบบเพื่อวัดประสิทธิภาพในการจำแนก งานวิจัยนี้ฝึกสอนระบบและทดสอบระบบโดยใช้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อคนเดียวกัน กล่าวคือเป็นระบบที่ขึ้นกับผู้พูด ถ้านำข้อมูลชุดทดสอบของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครอีกคนไปทดสอบกับโมเดลการจำแนกของอาสาสมัครอีกคน ผลการจำแนกอาจจะต่ำลง

งานวิจัยในอนาคตควรทำการออกแบบระบบโดยทำการฝึกสอนและทดสอบระบบแบบไม่ขึ้นกับผู้พูด เพื่อให้ระบบที่ออกแบบมีประสิทธิภาพ มีความน่าเชื่อถือและสามารถนำไปใช้ได้กับข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อเนื้ออื่นๆ

### 3) ระบบการจำแนกวรรณยุกต์

สำหรับในส่วนขอระบบการจำแนกวรรณยุกต์ที่ได้ออกแบบมานั้น ได้ออกแบบมาเพื่อเป็นต้นแบบสำหรับการพัฒนาระบบรู้จำในอนาคตที่ต้องการจำแนกคำพูดที่มีจำนวนมากขึ้น และเป็นคำพูดต่อเนื่อง ซึ่งในภาษาไทย คำพูดที่มีโครงสร้างของพยัญชนะ สระ และตัวสะกดเหมือนกัน แต่วรรณยุกต์ต่างกันจะให้ความหมายของคำที่ต่างกันโดยสิ้นเชิง ถ้ามีคำเหล่านี้อยู่ในคลังข้อมูลที่ต้องการจำแนกจำเป็นต้องจำแนกวรรณยุกต์ด้วย เนื่องจากคำพูดที่มีโครงสร้างเหมือนกันแต่ต่างที่วรรณยุกต์ จะมีลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อเนื้อที่คล้ายกันมาก ทำให้ระบบจำแนกคำพูดอาจจะจำแนกผิดพลาดได้ โดยอาจจะจำแนกเป็นคำที่คล้ายกันในกลุ่มนั้นๆ ซึ่งการจำแนกวรรณยุกต์จะช่วยในการตัดสินใจเพื่อหาคำพูดที่ถูกต้องมากขึ้น สำหรับลักษณะโครงสร้างของระบบดังกล่าว จะเพิ่มตัวตัดสินใจเข้ามาในส่วนของการจำแนก ในส่วนนี้จะต้องออกแบบให้มีความแม่นยำสูงสุดเท่าที่จะเป็นไปได้ ระบบการจำแนกวรรณยุกต์จึงจะสามารถปรับปรุงประสิทธิภาพรวมของระบบได้ รูปแบบของการนำระบบจำแนกวรรณยุกต์เข้ามาช่วยในการตัดสินใจแสดงดังรูปที่ 6.1



**รูปที่ 6.1** ตัวอย่างโครงสร้างรวมของระบบรู้จำเมื่อใช้ระบบจำแนกวรรณยุกต์เข้ามามีส่วนในการจำแนกร่วมกับการจำแนกคำพูดหลัก

#### 4) แบบจำลองการทำงานของระบบ

งานวิจัยนี้สร้างแบบจำลองการทำงานของระบบแบบออฟไลน์ (offline) ซึ่งเป็นการทดสอบการทำงานของระบบโดยป้อนอินพุตที่มีการเก็บข้อมูลของสัญญาณไว้แล้ว เนื่องจากข้อจำกัดของอุปกรณ์ในการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ (Mobi6) ไม่สามารถแก้ไขโปรแกรมที่ติดมากับอุปกรณ์ได้ ดังนั้นจึงไม่สามารถสร้างแบบจำลองการทำงานแบบออนไลน์ (online) ซึ่งทำการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อและป้อนอินพุตของสัญญาณในขณะนั้นเพื่อมาทดสอบกับระบบ

แนวทางวิจัยในอนาคต ควรจะมีการพัฒนาอุปกรณ์ตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อขึ้นมาเองเพื่อให้สามารถเชื่อมต่ออุปกรณ์กับโปรแกรมการทำงานของระบบได้

### 6.3 ข้อเสนอแนะ

1) การกำจัดสัญญาณรบกวนจากสัญญาณไฟฟ้ากล้ามเนื้อจากกล้ามเนื้อที่อยู่ติดกัน (cross talk noise)

สัญญาณรบกวนประเภทนี้เป็นอีกหนึ่งปัญหาของการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อโดยใช้อิเล็กโทรดแบบผิวหน้า ซึ่งเป็นการติดอิเล็กโทรดลงบนผิวหนังตรงตำแหน่งที่ต้องการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ การตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อโดยการใช้อิเล็กโทรดดังกล่าวมีผลทำให้สัญญาณที่ได้มีการปนกันระหว่างกล้ามเนื้อที่ต้องการตรวจจับกับกล้ามเนื้อที่อยู่ติดกัน ซึ่งจำเป็นจะต้องกำจัดสัญญาณรบกวนดังกล่าวเพื่อให้ได้สัญญาณจากตำแหน่งของกล้ามเนื้อที่ต้องการเพียงสัญญาณเดียว วิธีหนึ่งที่ยอมรับใช้ในการกำจัดสัญญาณรบกวนดังกล่าว คือ การวิเคราะห์องค์ประกอบอิสระ (independent component analysis, ICA) ซึ่งเป็นเทคนิคที่ใช้ในการแยกองค์ประกอบของสัญญาณ และใช้ในการลดจำนวนมิติของข้อมูล ซึ่งในงานวิจัยที่ใช้เทคนิคดังกล่าวในการกำจัดสัญญาณรบกวนประเภทนี้ และพบว่ามีประสิทธิภาพเพียงพอในการกำจัดสัญญาณรบกวนดังกล่าว [103] นอกจากนี้ยังพบว่าเทคนิคนี้สามารถกำจัดสัญญาณรบกวนจากสัญญาณคลื่นไฟฟ้าหัวใจได้อีกด้วย [104]

2) ลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ

จากงานวิจัยนี้จะเห็นได้ว่า เมื่อพิจารณาผลความแม่นยำในการจำแนกของระบบ ลักษณะเด่นที่วิเคราะห์ในแกนเวลามีประสิทธิภาพในการจำแนกดีกว่าลักษณะเด่นที่วิเคราะห์ใน



แกนความถี่ การคำนวณลักษณะเด่นในแกนเวลาจะพิจารณาคุณลักษณะของสัญญาณไฟฟ้า กล้ามเนื้อในแกนเวลา ซึ่งเป็นการเปลี่ยนแปลงขนาดของสัญญาณตามเวลา ในการออกเสียง คำพูดแต่ละเสียง กล้ามเนื้อที่ใช้ในกระบวนการออกเสียงพูดแต่ละเสียงแตกต่างกัน และการ ทำงานของกล้ามเนื้อแต่ละชนิดจะเกิดการหดตัวและคลายตัวในจังหวะที่ต่างกันอย่างชัดเจน ดังนั้นการพิจารณาจากสัญญาณไฟฟ้ากล้ามเนื้อในแกนเวลาสำหรับการจำแนกคำพูดแต่ละเสียงเหมาะสมมากกว่าการวิเคราะห์ในแกนความถี่ ซึ่งในการวิเคราะห์สัญญาณไฟฟ้า กล้ามเนื้อในแกนความถี่ พบว่าลักษณะของสัญญาณเมื่อถูกแปลงให้อยู่ในรูปของสัญญาณใน แกนความถี่ เช่น การแปลงแบบฟูริเยร์ช่วงสั้น เป็นต้น ลักษณะของสัญญาณที่ได้มีความแตกต่าง กันน้อยสำหรับคำพูดแต่ละเสียง ทำให้ลักษณะเด่นประเภทนี้มีประสิทธิภาพในการจำแนกต่ำกว่า ลักษณะเด่นในแกนเวลา

อย่างไรก็ตาม เนื่องจากคุณสมบัติของสัญญาณไฟฟ้ากล้ามเนื้อเป็นสัญญาณไม่ นิ่ง (non-stationary) แต่การคำนวณลักษณะเด่นในแกนเวลาเป็นการสมมุติให้สัญญาณดังกล่าว เป็นสัญญาณนิ่ง (stationary) ทำให้ลักษณะเด่นในกลุ่มนี้มีความแปรปรวนสูงเมื่อสัญญาณไฟฟ้า กล้ามเนื้อถูกบันทึกโดยผ่านอุปกรณ์ที่มีการเคลื่อนที่ [105] นอกจากนี้การวิเคราะห์สัญญาณใน แกนเวลา ปัญหาของสัญญาณรบกวนที่ปรากฏอยู่ในสัญญาณไฟฟ้ากล้ามเนื้อเป็นปัญหาหลักที่ ทำให้ประสิทธิภาพในการจำแนกถูกลดทอนด้วยสัญญาณรบกวนดังกล่าว ดังนั้นการเลือกใช้ ลักษณะเด่นในแกนเวลาให้มีประสิทธิภาพสูงสุดจำเป็นต้องพิจารณาในส่วนของการกำจัด สัญญาณรบกวนดังกล่าวด้วย

### บรรณานุกรมหรือเอกสารอ้างอิง

- [1] N. Sugie and K. Tsunoda, "A speech Prosthesis Employing a Speech Synthesizer-vowel Discrimination from Perioral Muscle Activities and Vowel Production," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp.485-490, 1985.
- [2] M.S. Morse and E.M. O'Brien, "Research Summary of a Scheme to Ascertain the Availability of Speech Information in the Myoelectric Signals of Neck and Head Muscles Using Surface Electrodes," *Comput. Biol. Med.*, vol.6, no. 6, pp. 399-410, 1986.
- [3] M.S. Morse, S.H. Day, B. Trull, and H. Morse, "Use of Myoelectric Signals to Recognize Speech," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Images 21<sup>st</sup> Century*, vol. 6, 1989, pp.1793-1794.
- [4] M.S. Morse, S.H. Day, and J. May, "Time Domain Analysis of the Myoelectric Signal Secondary to Speech," in *Proc. Annu. Inter. Conf. IEEE. Eng. Med. Biol. Soc.*, vol 12, 1990, pp. 1318-1319.
- [5] M.S. Morse, Y.N. Gopalan, and M. Wright, "Speech Recognition using Myoelectric Signals with Neural Networks," in *Proc. 13<sup>th</sup> Annu. Inter. Conf. IEEE Eng. Med. Biol. Soc.*, vol 13, Piscataway, 1991, pp. 1877-1878.
- [6] L. Maier-Hein, F. Metze, T. Schultz, and A. Waibel, "Session Independent Non-Audible Speech Recognition using Surface Electromyography," in *IEEE workshop on Automatic speech recognition and understanding*, SanJuan, Puerto Rico, 2005, pp. 331-336.
- [7] S.C. Jou, L. Marier-Hein, T. Schultz, and A. Waibel, "Articulatory Feature Classification using Surface Electromyography," in *Proc. IEEE Inter. Conf. on Acoustic specch and signal processing (ICASSP)*, Toulouse, 2006, pp. 605-608.
- [8] M. Walliczek, F. Kraft, S.C. Jou, T. Schultz, and A. Waibel, "Sub-Word Unit based Non-Audible Speech Recognition using Surface Electromyography," in *Proc. Interspeech(ICSLP)*, Pittsburgh, Pennsylvania, 2006, pp. 1487-1490

- [9] S.C. Jou, T. Schultz, M. Walliczek, F. Kraft, and A. Waibel, "Towards Continuous Speech Recognition Using Surface Electromyography," in *Proc. Interspeech*, Pittsburgh, PA, 2003, pp. 573-576.
- [10] S.C. Jou, T. Schultz, and A. Waibel, "Continuous Electromyographic Speech Recognition with a Multi-Stream Decoding Architecture," in *Proc. IEEE Inter. Conf. on Acoustic speech and signal processing (ICASSP)*, Honolulu, Hawaii, 2007, pp.401-404.
- [11] M. Wand, S.C. Jou, and T. Schultz, "Wavelet-based Front-End for Electromyographic Speech Recognition," in *Proc. Interspeech*, Antwerp, Belgium, 2007, pp. 686-689.
- [12] S.C. Jou and T. Schultz, "EARS: Electromyographical Automatic Recognition of Speech," in *Proc. Biosignals*, 2008, pp. 3-12.
- [13] T. Schultz and M. Wand, "Modeling Coarticulation in EMG-based Continuous Speech Recognition," *Speech communication*, vol. 52, pp.341-353, 2010.
- [14] M. Wand and T. Schultz, "Towards Speaker-Adaptive Speech Recognition based on Surface Electromyography," in *Proc. Biosignals*, Porto, Portugal, 2009, pp. 155-162.
- [15] M. Wand, S.C. Jou, A.R. Toth, and T. Schultz, "Impact of Different Specking Modes on EMG-Based Speech Recognition," in *Proc. Interspeech*, Brighton, United Kingdom, 2009.
- [16] M. Janke, M. Wand, and T. Schultz, "A Spectral Mapping Method for EMG-based Recognition of Silent Speech," in *Proc. B-INTERFACE*, 2010.
- [17] M. Janke, M. Wand and T. Schultz, "Impact of Lack of Acoustic Feedback in EMG-Based Silent Speech Recognition," in *Proc. Interspeech*, 2010
- [18] M. Wand and T. Schultz, "Session-Independent EMG-based Speech Recognition," in *Proc. Biosignals*, 2011, pp. 295-300.
- [19] M. Wand and T. Schultz, "Analysis of Confusion in EMG-based Speech Recognition," in *Proc. IEEE Inter. Conf., Acoustics, Speech and Signal Processing (ICASSP)*, Prague, May 2011, pp. 1520-6149.

- [20] C. Herff, M. Janke, M. Wand, and T. Schultz, "Impact of Different Feedback Mechanisms in EMG-based Speech Recognition," in *Proc. Interspeech*, Florence, Italy, 2011, pp.2213-2216.
- [21] M. Wand, M. Janke, and T. Schultz, "Investigation on Speaking Mode Discrepancies in EMG-based Speech Recognition," in *Proc. Interspeech*, Florence, Italy, 2011, pp. 601-604.
- [22] M. Wand, M. Janke and T. Schultz, "Decision-tree based Analysis of Speaking Mode Discrepancies in EMG-based Speech Recognition," in *BIOSIGNALS*, Algarve, Portugal, 2012, pp. 101-109.
- [23] M. Wand, C. Schulte, M. Janke and T. Schultz, "Array-based Electromyographic Silent Speech Interface," in *Biosignals*, Barcelona, Spain, 2013, pp. 89-96
- [24] M. Wand, A. Himmelsbach, T. Heistermann, M. Janke, and T. Schultz, "Artifact removal algorithm for an EMG-based Silent Speech Interface," in *Proc. Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, 2013, pp. 5750–5753
- [25] M. Wand and T. Schultz, "Pattern Learning with Deep Neural Networks in EMG-based Speech Recognition," in *Proc. Annu. Int. Conf. IEEE Eng. in Medicine and Biology Soc.*, 2014, pp. 4200–4203.
- [26] T. Heistermann, M. Janke, M. Wand, and T. Schultz, "Spatial Artifact Detection for Multi-Channel EMG-Based Speech Recognition," in *Proc. BIOSIGNALS*, 2014, pp. 189–196.
- [27] M. Wand and T. Schultz, "Towards real-life application of EMG-based speech recognition by using unsupervised adaptation," in *Proc. INTERSPEECH*, 2014, pp. 1189–1193.
- [28] M. Wand, M. Janke, and T. Schultz, "The EMG-UKA corpus for electromyographic speech processing," in *Proc. INTERSPEECH*, 2014, pp. 1593–1597.
- [29] M. Wand, M. Janke, and A. T. Schultz, "Tackling speaking mode varieties in EMG-based speech recognition," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2515–2526, 2014.

- [30] A.D.C. Chan, K. Englehart, B. Hugins, and D.F. Lovely, "Myo-electric Signal to Augment Speech Recognition," *Med. Biol. Eng. Comput.*, vol. 39, no. 4, pp. 500-504, 2001.
- [31] A.D.C. Chan, K. Englehart, B. Hugins and D.F. Lovely, "Hidden Markov Model Classification of Myoelectric Signals in Speech," *IEEE EMBS Magazine*, vol. 21, no. 4, pp. 143-146, 2002.
- [32] A.D.C. Chan, K. Englehart, B. Hudgins, and D.F. Lovely, "Multiexpert Automatic Speech Recognition using Acoustic and Myoelectric Signals," *Ph.D. dissertation, Dept. Electr. Comput. Eng.*, Univ. New Brunswick, Fredericton, NB, Canada, 2003.
- [33] E.J. Scheme, B. Hudgins, and P.A. Parker, "Myoelectric Signal Classification for Phoneme-based Speech Recognition," *IEEE Trans. Biomed. Eng.*, Vol. 54, no. 4, pp. 694-699, Apr. 2007.
- [34] Q. Zhou, N. Jiang, K. Englehart, and B. Hudgins, "Improved Phoneme-based Myoelectric Speech Recognition," *IEEE Trans. Bio. Med. Eng.*, vol. 56, no. 8, pp. 2016-2023, 2009.
- [35] R.S. Kumaran, K. Narayanan, and J.N. Gowdy, "Myoelectric Signals for Multimodal Speech Recognition," in *Proc. Interspeech(ISCA)*, 2005, pp. 1189-1192
- [36] Y. Deng, R. Patel, J.T. Heaton, G. Colby, L.D. Gilmore, J. Cabrera. S.H. Roy, C.J. De Luca, and G.S. Meltzner, "Disordered Speech Recognition Using Acoustic and sEMG Signals," in *Proc. Interspeech*, 2009, pp. 644-647.
- [37] H. Manabe, A. Hiraiwa, and T. Sugimura, " Unvoiced Speech Recognition using EMG--Mime Speech Recognition," in *Proc. Conf. computer Human Interaction*, Ft. Lauderdale, FL, 2003, pp. 794-795.
- [38] T. Kubo, T. Toda, M. Yoshida, T. Hattori and K. Ikeda, "Vowel recognition based on surface electromyography with electrode grid on submental region," *Trans. Jpn. Soc. Med. Biol. Eng.*, vol. 50, no. 1, pp. 38-46, 2012.

- [39] T. Kubo, M. Yoshida, T. Hattori, and K. Ikeda, "Shift invariant feature extraction for sEMG-based speech recognition with electrode grid," in *Proc. Annu. Inter. Conf. IEEE Eng. Medicine and Biology Soc.*, 2013, pp. 5797–5800.
- [40] T. Kubo, M. Yoshida, T. Hattori and K. Ikeda, "Towards excluding redundancy in electrode grid for automatic speech recognition based on surface EMG," *Neurocomputing*, vol. 134, pp. 15-19, 2014.
- [41] G. Colby, J.T. Heaton, L.D. Gilmore, J. Sroka, Y. Deng, J. Cabrera, S. Roy, C.J. De Luca, and G.S. Meltzer, "Sensor Subset Selection for Surface Electromyography Based Speech Recognition," in *Proc. IEEE Inter. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Taipei, 2009, pp.473-476.
- [42] X. Jia, X. Wang, J. Li, and Y. Du, "Unvoiced Speech Recognition Based on One-Channel Facial Myoelectric Signal," in *Proc. Word Congress on Intelligent control and Automation*, Dalian ,China, 2006, pp.9362-9366.
- [43] N. Bu, T. Tsuji, J. Arita, and M. Ohgh, "Phoneme Classification for Speech Synthesizer using Differential EMG Signals between Muscles," in *Proc. IEEE Eng. Med. Biol*, Shanghai, China, 2005.
- [44] Y. Deng, G. Colby, J. T. Heaton, and G. S. Meltzner, "Signal processing advances for the MUTE sEMG-based silent speech recognition system," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, 2012, pp. 1-6.
- [45] X. Jia, X. Wang, J. Li, D. Yang, and Y. Song, "Unvoiced Chinese Digital Recognition Based on Facial Myoelectric Signal," in *Proc. Inter. Conf. Comm., Circuits and Systems*, Guilin, 2006, pp. 598-601.
- [46] M. Lyu, C. Xiong and Q. Zhang, "Electromyography (EMG)-based Chinese voice command recognition," in *2014 IEEE Inter. Conf. Inform. Automation (ICIA)*, 2014.
- [47] H. Manabe and Z. Zhang, "Multi-Stream HMM for EMG-based Speech Recognition," in *Proc. 26<sup>th</sup> Annu. Inter. Conf. IEEE EMBS*, San Francisco, CA, USA, 2004, pp. 4389-4392.

- [48] M. Okamoto and Y. Matsubara, "Neuro-based Pattern Classification Method for EMG Signals not Belonging to Predefined Classes," in *Proc. Inter. Conf. Human and Computers*, 2008, pp. 163-168.
- [49] K.S. Lee, "EMG-based Speech Recognition Using Hidden Markov Models with Global Control Variables," *IEEE Trans. Bio. Med. Eng.*, vol. 55, no. 3, pp.930-940, 2008.
- [50] J.A.G. Mendes, R.R. Robson, S. Labidi, and A.K. Barros, "Subvocal Speech Recognition Based on EMG Signal Using Independent Component Analysis and Neural Network MLP," in *Proc. Image and Signal Processing (CISP)*, Sanya, Hainan, 2008, pp. 221-224.
- [51] J. Freitas, A. Teixeira and M. S. Dias, "Towards a silent speech interface for Portuguese," in *Proc. Biosignals*, 2012, pp. 91-100.
- [52] E. Lopez-Larraz, O.M. Mozos, J.M. Antelis, and J. Minguez, "Syllable-Based Speech Recognition Using EMG," in *Proc. Annu. Inter. Conf. IEEE Eng. Med. Biol. Soc.*, Buenos Aires, 2010, pp. 4699-4702.
- [53] L. Fraiwan, K. Lweesy, A. Al-Nemrawi, S. Addabass, and R. Saifan, "Voiceless Arabic Vowels Recognition Using Facial EMG," *Med. Biol. Eng. Comput.*, vol. 49, pp. 811-818, 2011.
- [54] M. Topalović, Đ. Damnjanović, A. Peulić, M. Blagojević and N. Filipović, "Syllable-based speech recognition using Electromyography and decision set classifier," *Biomed. Eng. Appl. basis Commun.*, vol. 27, no. 02, p. 1550020, 2015.
- [55] G.S. Meltzner, J. Sroka, J.T. Heaton, L.D. Gilmore, G. Colby, S. Roy, N. Chen, and C.J. De Luca, "Speech Recognition for Vocalized and Subvocal Modes of Production using Surface EMG Signals from the Neck and Face," in *Proc. Interspeech*, Brisbane, Australia, 2008.
- [56] C. Jorgensen and K. Binsted, "Web Browser Control using EMG based Sub Vocal Speech Recognition," in *Proc. Int. Conf. System Sci.*, Hawaii, 2005.

- [57] C. Jorgensen, D.D. Lee, and S. Agabon, "Sub Auditory Speech Recognition based on EMG/EPG Signals," in *Proc. Int. Conf. Neural Network*, Portland, OR, 2003, pp. 3128-3133.
- [58] B.J. Betts and C. Jorgensen, "Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment," *Moffett Field CA NASA Ames Research Center*, vol. 18, no.6, pp. 1242-1259, 2005.
- [59] S. Kumar, D.K. Kumar, M. Alemu, and M. Burry, "EMG Based Voiced Recognition," in *Proc. Intelligent Sensors, Sensor Networks and Information*, Melbourne, 2004, pp. 593-597.
- [60] W.C. Yau, S.P. Arjunan, and D.K. Kumar, "Classification of Voiceless Speech Using Facial Muscle Activity and Vision based Techniques," in *Proc. TENCON*, Hyderabad, 2008, pp. 1-6.
- [61] S.P. Arjunan, H. Weghorn, D.K. Kumar, and W.C. Yau, "Unspoken Vowel Recognition Using Facial Electromyogram," in *Proc. IEEE EMBS Annual Inter. Conf.*, New York City, USA, 2006, pp. 2191-2194
- [62] S.P. Arjunan, H. Weghorn, D.K. Kumar, G. Naik, and W.C. Yau, "Recognition of Human Voice Utterances from Facial Surface EMG without Using Audio Signals," in *Proc. Inter. Conf. Enterprise Infor. Sys.*, 2008, pp. 366-378.
- [63] G. S. Meltzner, G. Colby, Y. Deng, and J. T. Heaton, "Signal acquisition and processing techniques for sEMG based silent speech recognition," in *Proc. Annu. Inter. Conf. IEEE Eng. in Medicine and Biology Soc.*, 2011, pp. 4848–4851.
- [64] M. Epstein, N. Hacopian, and P. Ladefoged, "Dissection of the speech production mechanism," Department of Linguistics, UCLA, 2002.
- [65] C.J. De Luca, "Surface Electromyography: Detection and Recording," DelSys Incorporated, 2002.
- [66] S. Day, "Important factors in surface EMG measurement," Bortec Biomed. Lt., Calgary, Tech. Rep., 2002.



- [67] S. Stuker, T. Schultz, F. Metze, and A. Waibel, "Multilingual Articulatory Features," in *Proc. Inter. Conf. Acoust. Speech Signal Process.*, 2003, pp. 144-147.
- [68] ก. นาคสกุล, ระบบเสียงภาษาไทย. กรุงเทพฯ: สำนักพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย, 2551.
- [69] (2012, Jan. 6). *Human Facial Muscles* [online].  
Available:<http://www.corbisimages.com/stock-photo/rights-managed/42-177799652/human-facial-muscles>
- [70] E.N. Marieb, *Human Anatomy & Physiology*. USA: Benjamin Cummings, 2001.
- [71] J.J. Ohala, "The physiology of tone," L.M. Hyman (ed.), *Consonant type and Tone*, Southern California Occasional Papers in Linguistics, University of Southern California, LA, 1972.
- [72] H. Liu and H. Motoda, "Less is More," in *Feature Extraction, Construction and Selection*, New York: Kluwer Academic, 1993, pp. 3-12.
- [73] A.K. Jain, P.W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern anal. Mach. Intell.*, vol. 22, no. 1, pp.4-37, 2000.
- [74] R. Boostani and M. H. Moradi, "Evaluation of the forearm EMG signal features for the control of a prosthetic hand," *Physiol. Meas.*, vol. 24, no. 2, pp. 309–319, 2003.
- [75] K.S. Kim, H.H. Choi, C.S. Moon, and C.W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Curr. Appl. Phys.*, vol. 11, no. 3, pp. 740-745, 2011.
- [76] S.H. Park and S.P. Lee, "EMG pattern recognition based on artificial intelligence techniques," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 4, pp. 400–405, 1998.
- [77] M. Zardoshti-Kermani, B.C. Wheeler, K. Badie, and R.M. Hashemi, "EMG feature evaluation for movement control of upper extremity prostheses," *IEEE Trans. Rehabil. Eng.*, vol. 3, no. 4, pp. 324–333, 1995.

- [78] B. Hudgins, P. Parker, and R. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.
- [79] M.A. Oskoei and H. Hu, "Support vector machine based classification scheme for myoelectric control applied to upper limb," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1956–1965, 2008.
- [80] L. Philipson, "The electromyographic signal used for control of upper extremity prostheses and for quantification of motor blockade during epidural anaesthesia," Ph.D. thesis, Linköping University, Linköping, Sweden, 1987.
- [81] H.P. Huang and C.Y. Chen, "Development of a myoelectric discrimination system for a multi-degree prosthetic hand," in *Proc. IEEE Inter. Conf. Robot. Autom.*, vol. 3, 1999, pp. 2392–2397.
- [82] R. Merletti, "Standards for reporting EMG data," *J. Electromyogr. Kinesiol.*, vol. 6, no. 1, pp. III–IV, 1996.
- [83] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *J. Comput.*, vol. 1, no. 1, pp. 71-80, 2009.
- [84] C.J. Miller, "Real-time feature extraction and classification of prehensile EMG signals," M.S. thesis, San Diego State University, San Diego, CA, US, 2008.
- [85] M. Zecca, S. Micera, M.C. Carrozza, and P. Dario, "A novel feature extraction for robust EMG pattern recognition," *Cri. Rev. Biomed. Eng.*, vol. 30, no. 4-6, pp. 459–485, 2002.
- [86] S. Du and M. Vuskovic, "Temporal vs. spectral approach to feature extraction from prehensile EMG signals," in *Proc. IEEE Inter. Conf. Inf. Reuse Integr.*, 2004, pp. 344–350.
- [87] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491-502, 2005.
- [88] V. Kumar and S. Minz, "Feature Selection: A literature Review," *Smart Comput. Review*, vol. 4, no. 3, pp. 211-229, 2014.

- [89] R. Ruiz, J.C. Riquelme and J.S. Aguilar-Ruiz, "Best Agglomerative Ranked Subset for Feature Selection," In *FSDM, 2008*, pp. 148-162.
- [90] I. Yuyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157-1182, 2003.
- [91] A. Arauzo-Azofra, J.L. Aznarte M and J.M. benitez, "Empirical study of feature selection methods based on individual feature evaluation for classification problems," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8170-8177, 2011.
- [92] A. Phinyomark, S. Hiramviriyaya, C. Limsakul, and P. Phukpattaranont, "Evaluation of EMG feature extraction for hand movement recognition based on euclidean distance and standard deviation," In *Proc. Inter conf. Electr. Eng. Comput. Telecommun. Inf. Technol.*, 2010, pp.856-860.
- [93] R.O. Duda, P.E. Hart, and D.G. Stork, "Maximum Likelihood and Bayesian Estimation," in *Pattern Classification*, 2nd Ed. New York: Wiley, 2000, pp. 42-53.
- [94] C. Whaichana and S. Jitapnakul, "Basic theory for Hidden Markov Models," Elect. Eng. Dep., Chulalongkorn Univ., Bangkok, Seminar Report, 2002.
- [95] R.O. Duda, P.E. Hart, and D.G. Stork, "Bayesian decision theory," in *Pattern Classification*, 2nd ed. John Wiley & Sons, 2000.
- [96] AR. Webb, "Linear discriminant analysis," in *Statistical Pattern Recognition*, 2nd ed., UK: John Wiley & Sons, 2002, pp. 123-168.
- [97] S. Furui, "Speech Synthesis," in *Digital Speech Processing, Synthesis and Recognition*, K.J. Ray Liu, 2nd Ed. New York: Marcel Dekker, 2001, pp.213-237.
- [98] (2012, Jan. 10). Speech synthesis [online].  
Available: [https://en.wikipedia.org/wiki/Speech\\_synthesis](https://en.wikipedia.org/wiki/Speech_synthesis)
- [99] (2010, Oct. 20). Omohyoid Muscle [online].  
Available: [https://en.wikipedia.org/wiki/Omohyoid\\_muscle](https://en.wikipedia.org/wiki/Omohyoid_muscle)
- [100] D. Erickson, " Laryngeal muscles activity in connection with Thai tones," Annual Bulletin Research Institute Logopedics and Phoniatics, University of Tokyo, Japan, 1993, pp. 135-149.

- [101] R.P. Di Fabio, "Reliability of computerized surface electromyography for determining the onset of muscle activity," *Phys. Ther.*, vol. 16, no. 2, pp.175-187, 2006.
- [102] S. Balakrishnama and A. Ganapathi, "Linear Discriminant Analysis-A Brief Tutorial," *Elect. Eng. Com. Eng. Dep.*, Missisipi State Univ., Missisipi..
- [103] R.M. Howard, A. Teixeira, and A.J. Harrison, "An exploration of eliminating cross-talk in surface electromyography using independent component analysis," in *Proc. Signals and Systems (ISSC)*, 2015, pp. 1-6.
- [104] N.W. Willigenburg, A. Daffertshofer, I. Kingma, and J.H. van Dieen, "Removing ECG contamination from EMG recording: a comparison of ICA-based and other filtering procedures," *J. Electromyogr. Kinesiol.*, vol. 22, pp. 485-493, 2012.
- [105] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420-7431, 2012.

## ภาคผนวก

ภาคผนวก ก

สัทอักษรสากล (International Phonetic Alphabet, IPA)

the international phonetic alphabet (2005)

consonants (pulmonic)	LABIAL		CORONAL				DORSAL			LARYNGEAL		
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Alveolo-palatal	Palatal	Velar	Uvular	Pharyngeal	Glottal
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ			
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ	ʔ	ʕ	
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ	
Approximant		ʋ		ɹ		ɻ	j	ɰ				
Tap, flap		ⱱ		ɾ		ɽ						
Trill	B			r						R	H	ʕ
Lateral fricative				ɬ ɮ		ɮ̝ ɮ̞	ɬ̝ ɬ̞	ɮ̥ ɮ̦				
Lateral approximant				l		ɭ	ʎ	ʟ				
Lateral flap				ɺ		ɻ̥						

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured *ɦ*. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

consonants (non-pulmonic)

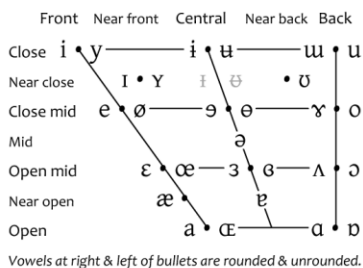
clicks	implosives	ejectives
⦿ Bilabial fricated	ɓ Bilabial	ʼ examples:
Laminar alveolar fricated ("dental")	ɗ Dental or alveolar	p' Bilabial
! Apical (post)alveolar abrupt ("retroflex")	ɖ Retroflex	t' Dental or alveolar
‡ Laminar postalveolar abrupt ("palatal")	ɟ Palatal	k' Velar
‡ Lateral alveolar fricated ("lateral")	ɠ Velar	ʈ' Lateral affricate
ɰ Velar (back released)	ɠ Uvular	s' Alveolar fricative

consonants (co-articulated)

M	Voiceless labialized velar approximant	//morphophonemic//
W	Voiced labialized velar approximant	/phonemic/
ɥ	Voiced labialized palatal approximant	[phonetic]
ɥ̟	Simultaneous x and ʃ (existence disputed)	<orthographic>
t͡	Affricates and double articulations	
ɡ͡b	may be joined by a tie bar	

brackets

vowels



## ประวัติผู้เขียน

ชื่อ สกุล	ว่าที่ร้อยตรีหญิง นียวดี ศรีสุวรรณ		
รหัสประจำตัวนักศึกษา	5310130010		
วุฒิการศึกษา			
	วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
	วิศวกรรมศาสตรบัณฑิต (วิศวกรรมไฟฟ้า)	มหาวิทยาลัยสงขลานครินทร์	2551

### ทุนการศึกษา (ที่ได้รับในระหว่างการการศึกษา)

ทุนผู้ช่วยวิจัยโครงการปริญญาเอกกาญจนาภิเษก ซึ่งเป็นโครงการร่วมให้ทุนระหว่างสำนักงานกองทุนสนับสนุนงานวิจัยและมหาวิทยาลัยสงขลานครินทร์

### การตีพิมพ์เผยแพร่ผลงาน

- Srisuwan, N., P. Phukpattaranont, and C Limsakul. 2011. "Feature selection for Thai tone classification based on surface EMG". *Procedia Engineering* 32: 253-259.
- Srisuwan, N., P. Phukpattaranont, and C. Limsakul. 2013. "Three steps of Neural Network classification for EMG-based Thai tones speech recognition." In *International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON)*. 1-6.
- Srisuwan, N., M. Wand, M. Janke, P. Phukpattaranont, T. Schultz, and C. Limsakul. 2014. "Enhancement of EMG-based Thai number words classification using frame-based time domain features with stacking filter." In *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. 1-6.