การจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความ

Classification of Thai syllables used for dysarthria rehabilitation

นิดา แซ่จอง

Nida Sae Jong

วิทยานิพนธ์นี้สำหรับการศึกษาตามหลักสูตรปริญญา
ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า
มหาวิทยาลัยสงขลานครินทร์

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

2562

การจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความ

Classification of Thai syllables used for dysarthria rehabilitation

นิดา แซ่จอง

Nida Sae Jong

วิทยานิพนธ์นี้สำหรับการศึกษาตามหลักสูตรปริญญา
ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Electrical Engineering
Prince of Songkla University
2562

ชื่อวิทยานิพนธ์ การจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความ
ผู้เขียน นางสาวนิดา แซ่จอง
สาขาวิชา วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(รองศาสตราจารย์ ดร.พรชัย พฤกษ์ภัทรานนต์)

คณะกรรมการสอบ

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ศุภฤกษ์ จันทร์จรัสจิตต์)

.....กรรมการ
(รองศาสตราจารย์ ดร.พรชัย พฤกษ์ภัทรานนต์)

.....กรรมการ
(ดร.รักกฤตว์ ดวงสร้อยทอง)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ดุจดาว บูรณะพาณิชย์กิจ)

.....กรรมการ
(ดร.อภิเดช บูรณวงศ์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้สำหรับการศึกษา ตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

.....
(ศาสตราจารย์ ดร.ดำรงศักดิ์ ฟ้ารุ่งสาง)
คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ.....
(รองศาสตราจารย์ ดร.พรชัย พฤกษ์ภัทรานนท์)
อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....
(นางสาวนิดา แซ่จอง)
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....
(นางสาวนิดา แซ่จอง)
นักศึกษา

ชื่อวิทยานิพนธ์ การจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความ
ผู้เขียน นางสาวนิดา แซ่จอง
สาขาวิชา วิศวกรรมไฟฟ้า
ปีการศึกษา 2561

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอระบบจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความด้วยสัญญาณไฟฟ้ากล้ามเนื้อ 5 ช่องสัญญาณและสัญญาณเสียง สำหรับจำแนก 12 พยางค์ไทย ระบบจำแนกพยางค์ไทยที่นำเสนอถูกแบ่งเป็น 4 ส่วนสำคัญคือ การปรับปรุงสัญญาณก่อนการประมวลผล การแทนที่ลักษณะเด่น การลดมิติของข้อมูลและการจำแนกพยางค์ อันดับแรกศึกษาคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อระหว่างอาสาสมัครคนปกติและผู้ที่มีอาการพูดไม่เป็นความ โดยการคำนวณลักษณะเด่น 3 กลุ่มคือ ขนาด ความถี่ และค่าทางสถิติ ลักษณะเด่น 2 ค่าจากแต่ละกลุ่มถูกกำหนดและวิเคราะห์ จากนั้น spectral regression extreme learning machine ซึ่งเป็นเทคนิคฉายลักษณะเด่นถูกใช้เพื่อลดขนาดของเวกเตอร์ลักษณะเด่น สุดท้ายลักษณะเด่นที่ผ่านการฉายลักษณะเด่นแล้วถูกจำแนกโดยใช้โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าร่วมกับการแบ่งข้อมูลออกเป็น 5 ส่วนเท่ากันเพื่อสร้างและทดสอบโมเดล (5-fold cross validation) ผลการทดลองแสดงให้เห็นว่าลักษณะเด่นกลุ่มขนาดและความถี่ส่งผลต่อประสิทธิภาพของการจำแนกพยางค์ อันดับสองประเมินประสิทธิภาพของระบบโดยการแยกพิจารณาช่องสัญญาณไฟฟ้ากล้ามเนื้อเดียวและการรวมข้อมูลจาก 2 3 4 และ 5 ช่องสัญญาณไฟฟ้ากล้ามเนื้อรวมกันโดยใช้ระบบที่นำเสนอ พบว่าเมื่อลดจำนวนช่องสัญญาณ ประสิทธิภาพในการจำแนกพยางค์จะลดลง อันดับสามในกรณีของสัญญาณเสียงจำนวนสัมประสิทธิ์ของลักษณะเด่น Mel-frequency cepstral coefficient (MFCC) คือ 8 13 และ 18 ถูกตรวจสอบ นอกจากนี้ลักษณะเด่น 2 กลุ่มระหว่างลักษณะเด่นในโดเมนเวลา 5 ค่าถูกเปรียบเทียบกับ MFCC ผลการทดลองชี้ให้เห็นว่า MFCC ดีกว่าลักษณะเด่นอีกกลุ่มหนึ่งและจำนวนสัมประสิทธิ์เท่ากับ 18 ให้ประสิทธิภาพสูงที่สุด สุดท้ายการจับคู่ที่ดีที่สุดของลักษณะเด่นและช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อถูกเลือกเพื่อรวมกับ MFCC ที่สกัดจากสัญญาณเสียง ผลการทดลองแสดงให้เห็นว่าการรวมข้อมูลจากหลายแหล่งที่มาให้ผลดีกว่าการใช้ข้อมูลจากแหล่งที่มาเดียวกัน โดยมีค่าความถูกต้องสูงสุดประมาณ 97 เปอร์เซ็นต์ กล่าวอีกนัยหนึ่งค่าความถูกต้องเพิ่มขึ้นสูงสุดถึง 51 เปอร์เซ็นต์ เมื่อเทียบกับการรวมข้อมูลจากแหล่งที่มาเดียวกัน นอกจากนี้ค่าเบี่ยงเบนมาตรฐานของค่าความถูกต้องในการจำแนกมีค่าต่ำเมื่อเปรียบเทียบกับการรวมข้อมูลจากแหล่งที่มาเดียวกัน บ่งชี้ว่าระบบมีความทนทานต่อสัญญาณรบกวนของการรู้จำพยางค์เพิ่มขึ้น

คำสำคัญ ระบบรู้จำพยางค์ สัญญาณไฟฟ้ากล้ามเนื้อ สัญญาณเสียง การรวมข้อมูลจากแหล่งที่มาเดียวกัน การรวมข้อมูลจากหลายแหล่งที่มา

Thesis Title Classification of Thai syllables used for dysarthria rehabilitation
Author Miss Nida Sae Jong
Major Program Electrical Engineering
Academic Year 2018

ABSTRACT

This thesis presented a Thai syllables classification system used for dysarthria rehabilitation based on five channels of surface electromyography (sEMG) and a channel of acoustic signal for classifying twelve Thai syllables. The proposed syllables classification system was divided into four important parts including signal pre-processing, feature representation, dimensionally reduction and classification. Firstly, we studied the characteristic of sEMG signal between healthy and dysarthric volunteers by calculating three feature groups as amplitude, frequency and probabilistic value. Two features from each feature group were determined and analyzed. Subsequently, a spectral regression extreme learning machine (SRELM) was used as the feature projection technique to reduce the dimension of the feature vector. Finally, the projected features were classified using a feed forward neural network (NN) classifier with 5-fold cross-validation. The results showed that amplitude and frequency feature affected to the syllable recognition performance. Secondly, the individual sEMG channel and the 2, 3, 4 and 5 combination sEMG channels were evaluated using the proposed system. The results found showed that when the channel of the electrode was reduced, the syllables classification performance was decreased. Thirdly, in case of the acoustic signal, the number of Mel frequency cepstral coefficients (MFCC) as 8, 13 and 18 were investigated. Moreover, two feature groups between five time domains and MFCC were compared. The results indicated that MFCC was better than another feature group and 18 coefficients gave the best performance. Finally, the best combination of features and channels of sEMG signal was chosen to be fused with the mel-frequency cepstral coefficients extracted from the acoustic signal. Results showed that the multimodal fusion outperformed the use of a single signal source achieving up to ~97% of accuracy. In other words, an accuracy

improvement up to 51% could be achieved when using the proposed multimodal fusion. Moreover, its low standard deviations in classification accuracy compared to those from the unimodal fusion indicated the improvement in the robustness of the syllable recognition.

Keyword speech recognition, surface electromyography (sEMG), acoustic, unimodal fusion, multimodal fusion

กิตติกรรมประกาศ

ขอขอบพระคุณ รองศาสตราจารย์ ดร.พรชัย พฤกษ์ภัทรานนต์ ประธานกรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาอุทิศเวลาให้คำปรึกษา แนะนำเอกสาร และข้อมูลต่างๆในการทำวิจัย รวมถึงการช่วยเหลือแก้ไขปัญหาและอุปสรรคต่างๆ ในการทำวิจัย ตลอดจนตรวจสอบและแก้ไขวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ศุภฤกษ์ จันทร์จรัสจิตต์ ประธานกรรมการสอบวิทยานิพนธ์และผู้ทรงคุณวุฒิจากภายนอก ที่ได้กรุณาสละเวลาเป็นกรรมการสอบวิทยานิพนธ์ ตลอดจนให้คำแนะนำที่เป็นประโยชน์และวิจารณ์ผลงานวิจัย รวมถึงการตรวจสอบวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ ดร.รักกฤตว์ ดวงสร้อยทอง ดร.อภิเดช บูรณวงศ์ ผู้ศาสตราจารย์ ดร.ดุจดาว บูรณะพาณิชย์กิจ และ กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาให้คำปรึกษาและคำแนะนำที่เป็นประโยชน์ในการทำวิจัย รวมถึงการตรวจสอบวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ อ.ดวงมน วงศ์จันทร์ดี สังกัดคลินิกโสตสัมผัสและแก้ไขการพูด โรงพยาบาลหาดใหญ่ ที่กรุณาให้คำปรึกษาเกี่ยวกับการฝึกบำบัดการพูด

ขอขอบพระคุณ ผู้ศาสตราจารย์ ดร.สุรพงษ์ ชาติพันธุ์ อาจารย์ภาควิชาวิศวกรรมชีว-การแพทย์ คณะแพทย์ศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ได้ให้ความอนุเคราะห์ด้านข้อมูลของผู้ที่มีอาการพูดไม่เป็นความ

ขอขอบพระคุณ โครงการพัฒนาอาจารย์และบุคลากรสำหรับสถาบันอุดมศึกษาในเขตพัฒนาเฉพาะกิจจังหวัดชายแดนภาคใต้ สำนักงานคณะกรรมการอุดมศึกษา ที่กรุณาสนับสนุนทุนการศึกษาและการทำวิจัย

ขอขอบพระคุณ บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ที่ได้ให้การสนับสนุนทุนในการทำวิจัยและให้ความช่วยเหลือด้านการประสานงานต่างๆ

ขอขอบพระคุณ คณาจารย์ บุคลากร นักศึกษาปริญญาโทและปริญญาเอกภาควิชาวิศวกรรมไฟฟ้าทุกคน รวมถึงเพื่อน พี่ น้อง ในภาควิชา ที่ให้ความช่วยเหลือและกำลังใจในการทำวิจัยมาโดยตลอด

สุดท้ายนี้ ข้าพเจ้าขอน้อมรำลึกถึงพระคุณของบิดามารดาและครอบครัว ที่ส่งเสริมสนับสนุน ให้กำลังใจ และทุนทรัพย์แก่ข้าพเจ้าตลอดมาจนสำเร็จการศึกษา

นิดา แซ่จอง

สารบัญ

สารบัญ (ต่อ)

สารบัญ (ต่อ)

รายการตาราง

รายการตาราง (ต่อ)

รายการภาพประกอบ

รายการภาพประกอบ (ต่อ)

รายการภาพประกอบ (ต่อ)

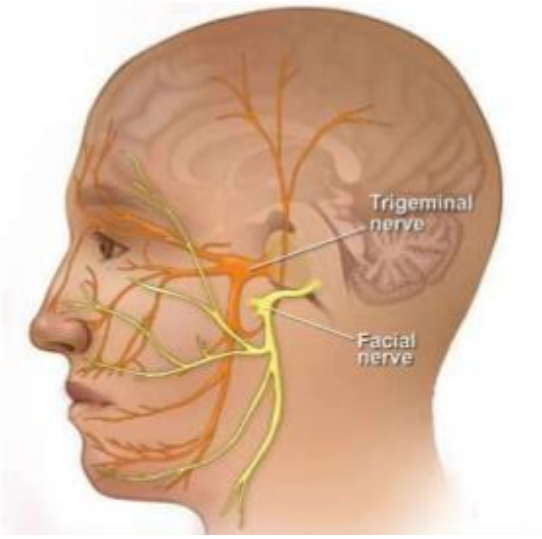
บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของงานวิจัย

โรคหลอดเลือดสมอง (stroke) เป็นสาเหตุของการสูญเสียชีวิตและความพิการที่พบบ่อยทั่วโลกโดยในแต่ละปีพบว่าประมาณ 0.4 เปอร์เซ็นต์ของประชากรที่มีอายุมากกว่า 45 ปีในสหรัฐอเมริกา ยุโรป และออสเตรเลียมีอัตราการเกิดโรคหลอดเลือดสมองขั้นต้น (first stroke) [1] เช่นเดียวกับสถานการณ์โรคหลอดเลือดสมองในเอเชียพบว่าอัตราการเกิดโรคและอัตราการตายจากโรคหลอดเลือดสมองมีค่าสูง โดยเฉพาะอย่างยิ่งประเทศในกลุ่มกำลังพัฒนา [2] ในส่วนของประเทศไทยพบว่าอัตราการเกิดความพิการจากโรคหลอดเลือดสมองมีค่าสูงถึง 1,108 คนต่อประชากร 100,000 คน สำหรับการสูญเสียปีสุขภาวะ (DALYs) ในปี พ.ศ. 2553 จำแนกตามกลุ่มอายุและเพศ [2] โดยทั่วไปโรคหลอดเลือดสมองทำให้เกิดความเสียหายต่อร่างกายหลายประการ หนึ่งในนั้นคือความผิดปกติเกี่ยวกับการพูดหรือโรคอาการพูดไม่เป็นความ (dysarthria) ซึ่งเกิดจากความผิดปกติของระบบประสาทควบคุมการพูด (motor speech disorder) ส่งผลให้เกิดความบกพร่องด้านการเคลื่อนไหวของกล้ามเนื้อที่ใช้สำหรับการกำเนิดคำพูด รวมถึงปาก ลิ้น และเส้นเสียง [3]

อาการพูดไม่เป็นความสามารถแบ่งออกเป็น 6 ชนิด ตามตำแหน่งพยาธิสภาพและลักษณะคำพูด อย่างไรก็ตามในการวิจัยนี้มุ่งเน้นเฉพาะภาวะพูดไม่เป็นความแบบอ่อนแรง (flaccid dysarthria) ซึ่งเกิดจากความอ่อนแอของระบบกล้ามเนื้อที่ใช้สำหรับการกำเนิดการพูด โดยเส้นประสาทสมองที่เกี่ยวข้องกับภาวะพูดไม่เป็นความแบบอ่อนแรงคือ เส้นประสาทไทรเจมินัล (trigeminal) เส้นประสาทเฟเชียล (facial) เส้นประสาทเวกัส (vagus) และเส้นประสาทกล้ามเนื้อลิ้น [4] ดังแสดงในภาพประกอบที่ 1-1 [5] สำหรับแนวทางการรักษาทางคลินิกโดยทั่วไปนักแก้ไขการพูด (speech-language) จะทำการรักษาโดยการพัฒนาวิธีออกเสียงและมุ่งเน้นเรื่องความชัดเจนของการออกเสียงของหน่วยเสียง [6] การฟื้นฟูสมรรถภาพของการพูดและภาษาจะได้ผลดีหากได้รับการฝึกอย่างเต็มที่และต่อเนื่องจากนักแก้ไขการพูดในช่วง 1 ปีหลังจากมีความผิดปกติทางการพูดและทางภาษา [7] อย่างไรก็ตามพบว่าจำนวนนักแก้ไขการพูดไม่เพียงพอกับความต้องการของผู้ป่วย [8] ส่งผลให้ผู้ป่วยไม่ได้รับการรักษาอย่างต่อเนื่อง เพื่อที่จะบรรเทาปัญหานี้การพัฒนาระบบอัจฉริยะแบบอัตโนมัติสำหรับการฟื้นฟูการพูดมีความจำเป็นอย่างยิ่งเพื่อสนับสนุนนักแก้ไขการพูดทำให้ผู้ป่วยฟื้นฟูเร็วขึ้น



ภาพประกอบที่ 1-1 เส้นประสาทที่เกี่ยวข้องกับภาวะพูดไม่เป็นความแบบอ่อนแรง เส้นประสาทคู่ที่ 5 และ 7 สั่งการกล้ามเนื้อในการแสดงสีหน้า [5]

จากการทบทวนวรรณกรรมซึ่งจะกล่าวถึงในหัวข้อ 1.5 พบว่างานวิจัยด้านระบบอัจฉริยะแบบอัตโนมัติสำหรับผู้ป่วยอาการพูดไม่เป็นความส่วนใหญ่เป็นงานด้านการประเมินระดับความรุนแรงของโรคโดยวิเคราะห์บนพื้นฐานของสัญญาณเสียงเป็นหลัก [8][9][10] อย่างไรก็ตามสัญญาณเสียงไม่เหมาะสำหรับการนำไปใช้ในบางสถานการณ์เช่น ในสภาวะแวดล้อมที่มีเสียงรบกวน [11] เนื่องจากสัญญาณเสียงถูกแทรกสอดจากสัญญาณรบกวนรอบข้างได้ง่าย ทำให้สัญญาณเสียงไม่เหมาะจะใช้งานที่บ้าน เพราะต้องทดสอบในระบบปิดซึ่งต้องลงทุนค่อนข้างสูง ด้วยเหตุนี้การรู้จำคำพูด (speech recognition) บนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อถูกนำเสนออย่างแพร่หลาย สัญญาณไฟฟ้ากล้ามเนื้อเกิดจากการหดตัวของมัดกล้ามเนื้อขณะที่ออกเสียง วิธีวัดสัญญาณไฟฟ้ากล้ามเนื้อเริ่มจากติดเซนเซอร์หรือที่เรียกว่าอิเล็กโทรดบริเวณผิวหนัง จากนั้นอิเล็กโทรดจะแปลงกระแสไฟฟ้าขนาดเล็กเป็นแรงดันไฟฟ้าเมื่อกระแสเคลื่อนผ่านมัดกล้ามเนื้อในสภาวะมัดกล้ามเนื้อหดตัว [12] ข้อดีของการใช้สัญญาณไฟฟ้ากล้ามเนื้อคือสามารถบ่งชี้ผลตอบสนองทางไฟฟ้าด้านสรีรวิทยาในหลายๆ กิจกรรม [13] และสามารถนำไปใช้เพื่อสอนมัดกล้ามเนื้อที่อ่อนแรงให้ทำงาน [14] [15] อย่างไรก็ตามการนำสัญญาณไฟฟ้ากล้ามเนื้อมาประยุกต์ใช้งานยังมีข้อจำกัดบางอย่างที่ทำให้เข้าถึงได้ยาก ยกตัวอย่างเช่น (1) ความยากในการวัดสัญญาณไฟฟ้ากล้ามเนื้อในผู้ป่วยพูดไม่เป็นความชนิดอ่อนแรงเนื่องจากขนาดของสัญญาณค่อนข้างต่ำ (2) การแทรกสอดของสัญญาณรบกวนจากหลายแหล่งเช่น การเคลื่อนที่ของสายนำสัญญาณ (motion artefact) แหล่งจ่ายไฟ (power line) หรือคลื่นไฟฟ้าหัวใจ (electro- cardiographic artefacts) ซึ่งทำให้คุณภาพของสัญญาณลดลงและส่งผลต่อประสิทธิภาพในการรู้จำคำพูด [16] (3) การใช้อิเล็กโทรดจำนวนมากทำให้ค่าใช้จ่ายเพิ่มขึ้นเนื่องจากเป็นแบบใช้แล้วทิ้งและไม่สะดวกในการประยุกต์ใช้งานด้านระบบฟื้นฟูการพูด นอกจากนี้ยัง

พบว่าระบบรู้จำคำพูดสำหรับผู้ป่วยพูดไม่เป็นความได้ถูกพัฒนาในภาษาต่างๆ เช่น ภาษาอังกฤษและภาษาสเปน [17][18][19] แต่สำหรับภาษาไทยงานวิจัยด้านนี้ยังไม่ค่อยแพร่หลาย [19] ประเด็นที่น่าสนใจคือความแตกต่างของการออกเสียงภาษาอังกฤษและภาษาไทย ในการออกเสียงพยัญชนะภาษาไทยกระแสลมที่ใช้ได้มาจากอวัยวะหลายแหล่งเช่น จากปอด กล่องเสียงและเพดานอ่อน แต่การออกเสียงพยัญชนะในภาษาอังกฤษล้วนแล้วแต่ใช้กระแสลมจากปอดทั้งสิ้น [21] นอกจากนี้ในระบบเสียงภาษาไทยมีหน่วยเสียงพยัญชนะเสียงกักจากฐานเพดานอ่อนคือ/k, kh/ ค ควาย ซึ่งแตกต่างจากหน่วยเสียงพยัญชนะเสียงกักของภาษาอังกฤษคือ /k - kh/ โดยออกเสียงเป็นเสียง ก ไก่ แทน [21]

จากที่มาและความสำคัญของงานวิจัย ปัญหาของการวิจัย และช่องว่างของงานวิจัยดังอธิบายในข้างต้น วิทยานิพนธ์นี้จึงนำเสนอระบบจำแนกพยางค์ภาษาไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความเพื่อการบำบัดการพูดเบื้องต้นจำนวน 12 พยางค์โดยใช้สัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้า สัญญาณเสียง และการรวมสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงเพื่อเปรียบเทียบประสิทธิภาพการจำแนกพยางค์ โดยมีเป้าหมายเพื่อลดจำนวนอิเล็กโทรดและเพิ่มความทนทานต่อสัญญาณรบกวนจากภายนอก

1.2 โจทย์การวิจัย

1.2.1 มีวิธีการเลือกลักษณะเด่นอย่างมีประสิทธิภาพอย่างไร ที่สามารถทำให้ประสิทธิภาพในการจำแนกพยางค์สูงแต่ใช้ลักษณะเด่นจำนวนน้อยสำหรับสัญญาณไฟฟ้ากล้ามเนื้อ

1.2.2 มีวิธีการลดจำนวนอิเล็กโทรดอย่างไร โดยที่ประสิทธิภาพในการจำแนกพยางค์ยังคงเท่าเดิมหรือดีกว่าเดิม

1.3 วัตถุประสงค์ของการวิจัย

ศึกษาระบบจำแนกพยางค์ที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความจำนวน 12 พยางค์ ได้แก่ อา อี อู คา คี คู นา นี นู มา มี และ มู จากสัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้า และ/หรือสัญญาณเสียงของคนปกติเพื่อใช้เป็นต้นแบบในการจำแนกพยางค์

1.4 ประโยชน์ที่คาดว่าจะได้รับ

ได้ระบบจำแนกพยางค์ไทยจำนวน 12 พยางค์เพื่อใช้เป็นพื้นฐานสำหรับการประยุกต์ใช้ในการฟื้นฟูอาการพูดไม่เป็นความ

1.5 การทบทวนวรรณกรรม บทความ และงานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยที่เกี่ยวข้องทางด้านการรู้จำคำพูดสามารถแบ่งออกเป็นสองส่วนหลักคือการรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง ปัญหาของงานวิจัยในปัจจุบันจากการสำรวจงานวิจัยที่เกี่ยวข้องสามารถเกริ่นสรุปช่องว่างของความรู้แยกตามที่มาของสัญญาณมีรายละเอียดดังนี้

1.5.1 การรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ ประเด็นวิจัยที่น่าสนใจสามารถแบ่งออกเป็น 3 ข้อหลักคือตำแหน่งการวางอิเล็กโทรด ลักษณะเด่นที่ใช้ในการรู้จำคำพูดและการรวมข้อมูลมีรายละเอียดดังนี้

ก) ตำแหน่งการวางอิเล็กโทรดเป็นองค์ประกอบหนึ่งที่มีความสำคัญซึ่งส่งผลต่อประสิทธิภาพของระบบรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ เนื่องจากงานวิจัยส่วนใหญ่มีการใช้อิเล็กโทรดจำนวนมาก [11][12][17][18][19] ทำให้การบันทึกสัญญาณไฟฟ้ากล้ามเนื้อบนใบหน้าปฏิบัติได้ยากส่งผลต่อการนำไปประยุกต์ใช้งานจริง[22] วรรณกรรม [23] รายงานความสัมพันธ์ระหว่างตำแหน่งของมัดกล้ามเนื้อบนใบหน้าและท่าทางการแสดงออกบนใบหน้าซึ่งมีความสำคัญต่อการฟื้นฟูการหดตัวของมัดกล้ามเนื้อเฉพาะจุด ต่อมาวรรณกรรม [24] เสนอความสัมพันธ์ของมัดกล้ามเนื้อบนใบหน้ากับกริยาท่าทางบนใบหน้า 29 รูปแบบซึ่งมีความสัมพันธ์กับการรักษาทางคลินิกขั้นสูงโดยสัมพันธ์กับการบำบัดอวัยวะที่เกี่ยวกับการพูดโดยการไม่ออกเสียง (non-speech oral motor treatment) ตำแหน่งการวางอิเล็กโทรดบนใบหน้าสำหรับการวัดสัญญาณไฟฟ้ากล้ามเนื้อที่ได้รับความนิยมมากที่สุดประกอบด้วยมัดกล้ามเนื้อดังต่อไปนี้ กล้ามเนื้อ Zygomaticus major กล้ามเนื้อ Levator anguli oris และกล้ามเนื้อ Depressor anguli oris เป็นต้น [17][18][19] โดยทั่วไปการวัดสัญญาณไฟฟ้ากล้ามเนื้อมี 2 รูปแบบคือแบบขั้วเดียวและแบบสองขั้ว [25] ดังนั้นระบบการได้ข้อมูลจะต้องต่อกับกราวด์และอิเล็กโทรดอ้างอิงเพื่อป้องกันสัญญาณรบกวนจากภายนอก นอกจากนี้อิเล็กโทรดที่ทำหน้าที่เป็นกราวด์จะถูกวางในตำแหน่งที่ไม่มีการทำงานมัดกล้ามเนื้อระหว่างการวัดเช่น บริเวณหน้าผาก ด้านหลังลำคอและข้อมือ ในส่วนของการวัดสัญญาณไฟฟ้ากล้ามเนื้อแบบขั้วเดียว อิเล็กโทรดอ้างอิงควรจะวางให้ห่างจากตำแหน่งของหัวใจเพื่อลดผลกระทบของคลื่นไฟฟ้าหัวใจระหว่างการวัดสัญญาณ [2] ดังนั้นอิเล็กโทรดอ้างอิงมักจะติดบริเวณติ่งหู

ข) ลักษณะเด่นที่ใช้ในการรู้จำคำพูด การคำนวณลักษณะเด่นเป็นกระบวนการที่จำเป็นเพื่อลดความซ้ำซ้อนของข้อมูลและเพิ่มความเกี่ยวเนื่องของข้อมูล ลักษณะเด่นหลากหลายถูกนำมาใช้ในการรู้จำคำพูดทั้งลักษณะเด่นในโดเมนเวลาและโดเมนความถี่ สำหรับงานก่อนหน้านี้ผู้วิจัย [26] คำนวณลักษณะเด่น 6 แบบประกอบด้วยค่าสัมบูรณ์ของค่าเฉลี่ย (MAV) ความยาวคลื่น (WL)

การตัดผ่านค่าศูนย์ (ZC) ค่าความถี่กลาง (MNF) ค่าความโด่ง ($L-KURT$) และค่าความเบ้ ($L-SKW$) ของการกระจายของข้อมูล โดยแบ่งลักษณะเด่นทั้งหมดเป็น 3 กลุ่มคือขนาด (amplitude based feature: ABF) ความถี่ (frequency based feature: FBF) และค่าทางสถิติของการกระจายของข้อมูล (statistic based feature: SBF) เพื่อศึกษาคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อ พบว่าลักษณะเด่นกลุ่ม ABF และ FBF ให้ค่าความถูกต้องในการจำแนกพยางค์สูงเมื่อเทียบกับลักษณะเด่นกลุ่ม SBF อย่างไรก็ตามพบว่าลักษณะเด่นที่นิยมนำมาใช้ในการรู้จำคำพูดคือสัมประสิทธิ์เซปสตรัลที่คำนวณบนแกนความถี่แบบเมล (mel-frequency cepstral coefficient: MFCC) [17][18][19][27] เนื่องจากชุดตัวกรองความถี่แบบเมลถูกออกแบบให้เลียนแบบการรับรู้ในการได้ยินของคน ส่งผลให้การแปรผันในช่วงความถี่ต่ำมีความสำคัญกว่าในช่วงความถี่สูง ดังนั้นความกว้างของแถบความถี่จะเพิ่มขึ้นเมื่อความถี่กลางของวงจรกรองเพิ่มขึ้น

ค) การรวมข้อมูลสามารถแบ่งเป็น 2 รูปแบบขึ้นกับธรรมชาติของข้อมูลคือ การรวมข้อมูลจากแหล่งที่มาเดียวกัน (unimodal fusion) และการรวมข้อมูลจากหลายแหล่งที่มา (multimodal fusion) [28] วิธีการรวมข้อมูลสามารถแบ่งได้เป็นหลายระดับประกอบด้วยระดับเซนเซอร์ ระดับลักษณะเด่น ระดับคะแนน และระดับการตัดสินใจ อย่างไรก็ตามการรวมข้อมูลระดับลักษณะเด่นถูกนำมาใช้อย่างแพร่หลายในเชิงการรู้จำรูปแบบ [13] การรวมข้อมูลจากแหล่งที่มาเดียวกันหมายถึงการรวมข้อมูลที่มาจากต้นทางเดียวกัน ในงานวิจัยนี้หมายถึงการรวมลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อแบบหลายช่องสัญญาณ ขณะที่การรวมข้อมูลจากหลายแหล่งที่มาเป็นการรวมข้อมูลที่มาจากต้นทางต่างกัน ในงานวิจัยนี้หมายถึงการรวมลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง ในส่วนของการรวมข้อมูลจากแหล่งที่มาเดียวกันถูกเผยแพร่อย่างกว้างขวางในการรู้จำคำพูดบนพื้นฐานสัญญาณไฟฟ้ากล้ามเนื้อยกตัวอย่างเช่น วรรณกรรม [27] นำเสนอการทดสอบความสัมพันธ์ของสัญญาณไฟฟ้ากล้ามเนื้อในแต่ละมัดกล้ามเนื้อด้วยโมเดลแบบผสมของการแจกแจงปกติ (gaussian mixture model: GMM) โดยความสัมพันธ์ระหว่างช่องสัญญาณถูกแทนที่ด้วยค่าสหสัมพันธ์ (cross correlation) และฟังก์ชันความหนาแน่นของค่าสหสัมพันธ์ระหว่างสัญญาณไฟฟ้ากล้ามเนื้อในแต่ละช่องสัญญาณถูกควบคุมด้วย global control variable (GCV) ดังนั้นวิธีการรวมเวกเตอร์ลักษณะเด่นจากมัดกล้ามเนื้อที่ต่างกันจึงมีความสำคัญและสามารถแบ่งการรวมเวกเตอร์ออกเป็นการรวมแบบก่อน (early integration: EI) และการรวมแบบหลัง (late Integration: LI) อย่างไรก็ตามวรรณกรรมนี้มุ่งเน้นไปที่โมเดล LI เนื่องจากฟังก์ชันความหนาแน่นถูกกำหนดสำหรับแต่ละลักษณะเด่น ขณะที่โมเดล EI ก่อตัวขึ้นจากเวกเตอร์ลักษณะเด่นหลายส่วน ในส่วนของการทดลองได้ทำการเปรียบเทียบประสิทธิภาพของโมเดล EI โมเดล LI แบบอิสระ และโมเดล LI แบบพึ่งพากัน โดยการทดสอบความเหมือน (likelihood) และการรู้จำ (recognition) ผลการทดลองพบว่าสัญญาณไฟฟ้ากล้ามเนื้อของกล้ามเนื้อแต่ละมัดมีลักษณะพึ่งพา

กัน ต่อมาวรรณกรรม [17] เชื่อมต่อเวกเตอร์ลักษณะเด่นจากสัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้าและลำคอทั้งหมด 8 ช่องสัญญาณ ดังนั้นผลของการรวมข้อมูลจากสัญญาณไฟฟ้ากล้ามเนื้อคือความยาวของเวกเตอร์ลักษณะเด่นเท่ากับ 328 (41 ลักษณะเด่น/ช่องสัญญาณ × 8 ช่องสัญญาณ) จากนั้นตัวจำแนกประเภทชนิดการตัดสินใจแบบต้นไม้ร่วมกับอัลกอริทึมเอดาบูสท์ถูกนำมาใช้เพื่อจำแนกคำพูดจำนวน 30 พยางค์ในภาษาสเปน ค่าความถูกต้องเฉลี่ยของอาสาสมัครทั้ง 3 คนเท่ากับ 70 เปอร์เซ็นต์

ในกรณีของการรวมข้อมูลจากหลายแหล่งที่มานิยมนำไปใช้ในการระบุตัวบุคคล [28] นอกจากนี้ในงานวิจัยด้านการรู้จำคำพูดพบว่า วรรณกรรม [18] และ [19] เป็นการรวมข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงด้วยวิธีการรวมข้อมูลระดับการตัดสินใจ ข้อดีคือสามารถรวมข้อมูลแม้ว่าชนิดของข้อมูลจากแต่ละแหล่งที่มามีความแตกต่างกัน ซึ่งส่งผลให้ลักษณะเด่นที่ได้มีความแตกต่างกันอย่างมากทำให้ยากต่อการนำมารวมกัน สุดท้ายกระบวนการเสียงส่วนใหญ่ (voting) ถูกนำมาใช้ในการจำแนกประเภท วิธีการดังกล่าวอยู่บนพื้นฐานของกรอบทางคณิตศาสตร์ของทฤษฎีพิสูจน์หลักฐาน (evidence theory) เรียกว่าวิธี plausibility [18] จุดประสงค์ของการรวมข้อมูลคือการพัฒนาผลของการจำแนกคำพูดที่ระดับของสัญญาณรบกวนต่างกันโดยที่รู้หรือไม่รู้ค่าสัญญาณต่อสัญญาณรบกวน ผลการทดลองแสดงให้เห็นว่าค่าความถูกต้องของการจำแนกคำหรือหน่วยเสียงจากการรวมข้อมูลโดยใช้วิธี plausibility ดีกว่าการใช้สัญญาณเสียงเพียงอย่างเดียว อย่างไรก็ตามการรวมข้อมูลระดับการตัดสินใจมีข้อจำกัดด้านจำนวนของข้อมูลที่มี [28]

ต่อมาวรรณกรรม [29] ทดสอบความถูกต้องของการจำแนกคำของคนที่มีอาการพูดไม่เป็นความด้วยวิธีเปรียบเทียบการรู้จำคำพูดของคนที่มีความผิดปกติด้านการพูดโดยใช้สัญญาณเสียง สัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้าและคอและการรวมข้อมูลจากสัญญาณทั้งสองรวมกัน วิธีการรวมข้อมูลเป็นแบบระดับคะแนน ผลการทดลองบ่งชี้ว่าโมเดลที่ใช้ทั้งสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อจำแนกดีกว่าการใช้สัญญาณเสียง หรือ สัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียวสำหรับการพูดออกเสียง

1.5.2 การรู้จำคำพูดบนพื้นฐานของสัญญาณเสียง เน้นกล่าวถึงวรรณกรรมที่เกี่ยวข้องกับผู้ป่วยที่มีอาการพูดไม่เป็นความเป็นหลัก การวิจัยส่วนใหญ่เป็นการสร้างระบบเพื่อวินิจฉัยความรุนแรงของโรคอาการพูดไม่เป็นความและระบบรู้จำคำพูดเพื่อให้ผู้ป่วยสามารถสื่อสารกับคนรอบข้างได้ วรรณกรรม [9] นำเสนอวิธีสร้างตัวบ่งชี้ระดับความรุนแรงของโรคแบบอัตโนมัติ ด้วยการใช้ตัวชี้วัดการกระจายของการพูด (Speech confusion index: Ø) โดยพิจารณาจากความง่ายของสัญญาณเสียงที่อาจจะผิดพลาดจากคำที่กำหนดไว้บนพื้นฐานของการประมวลผลสัญญาณและเปรียบเทียบความถูกต้องของการจำแนกระดับความรุนแรงของโรคกับการวินิจฉัยโดยผู้เชี่ยวชาญ (A-

score) และผู้ที่ไม่มีปัญหาด้านการได้ยิน (I-score) ลักษณะเด่นที่ใช้คือ MFCC โดยมีกรอบสัญญาณแฮมมิงเท่ากับ 25 มิลลิวินาทีต่อเฟรมและช่วงห่างระหว่างเฟรมเท่ากับ 10 มิลลิวินาที นอกจากนี้โมเดลมาร์คอฟซ่อนเร้น (hidden markov model: HMM) และโครงข่ายประสาทเทียม (artificial neuron network : ANN) ถูกใช้เป็นตัวจำแนกประเภท จากการทดลองผู้วิจัยใช้ตัวชี้วัดจำนวนสามตัวเพื่อเปรียบเทียบประสิทธิภาพระหว่าง Ø, A-score และ I-score ดังนี้ rank-order inconsistency (ROI) ซึ่งแสดงประสิทธิภาพของลำดับ (performance order) ค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson's correlation coefficient : R2) และค่ารากที่สองของค่าเฉลี่ยกำลังสองของผลต่าง (root-mean-square of difference : Δ rms) ผลปรากฏว่าตัวชี้วัดการกระจายของการพูดให้ค่าความถูกต้องของการจำแนกระดับของโรคได้ดีกว่าแบบ A-score และแบบ I-score ในเทอมของ ROI, R2 และ Δ rms ทั้งตัวจำแนกประเภทโมเดลมาร์คอฟซ่อนเร้น และโครงข่ายประสาทเทียม

วรรณกรรม [8] นำเสนอวิธีการจำแนกระดับความรุนแรงของโรค โดยมุ่งความคิดไปที่การจับคู่ การแทนที่ และการหักล้างของลักษณะเด่น เนื่องจากปัญหาของข้อมูลที่มีมากเกินไปและข้อจำกัดด้านประสิทธิภาพของระบบ วิธีการนี้ประกอบด้วยสองขั้นตอนหลักคือ การแทนที่ลักษณะเด่นโดยใช้ฮิสโทแกรมของแผนที่การออกเสียง และการทำนายคำโดยใช้ phonologically-structured sparse linear model (PSSLM) ด้วยคะแนนความเข้าใจ (intelligibility score) โครงสร้างของ PSSLM คล้ายกับโครงสร้างของต้นไม้ที่เกี่ยวกับลักษณะของระบบเสียงในภาษา (Phonological attributes) ในการประเมินผลทำการเปรียบเทียบวิธีการทำนายคำที่นำเสนอ คือ PSSLM กับวิธีการอื่นเช่น SLM, multiple speech feature และ GMM supervectors โดยใช้ตัวชี้วัดประสิทธิภาพของการทำนายคือ ค่าความผิดพลาดของรากที่สองของค่าเฉลี่ยกำลังสอง (root mean square error : RMSE) และค่าสหสัมพันธ์ (R) ของคำที่ทำนายได้กับคำที่พูดจริง พบว่าวิธีทำนายด้วย PSSLM ให้ค่า RMSE น้อยที่สุด และค่า R มากที่สุด

วรรณกรรม [10] นำเสนอการใช้เทคนิคการปรับค่าได้สำหรับคนที่มีความบกพร่องด้านการพูด ด้วยการสร้างโมเดลเสียงแบบปรับค่าได้ที่ผู้พูดเป็นอิสระต่อกัน บนพื้นฐานของแหล่งข้อมูลของคนปกติ (unimpaired TIMIT) และผู้ป่วย (impaired TORGO) โดยทำการเปรียบเทียบเทคนิคการปรับค่าได้สองแบบคือ maximum likelihood linear regression (MLLR) และ constrained MLLR (C-MLLR) กับแหล่งข้อมูล ด้วยการวัดอัตราความผิดพลาดของคำ (word error rate) จากการเติมคำ การแทนคำ และการลบคำ จากฐานข้อมูลที่มีอยู่ในแต่ละระดับของความรุนแรง เพื่อให้สามารถเข้าใจคำพูดของคนที่มีอาการพูดไม่เป็นความ ลักษณะเด่นที่ใช้คือ MFCC โดยมีกรอบสัญญาณแฮมมิงเท่ากับ 25 มิลลิวินาทีต่อเฟรมและช่วงห่างระหว่างเฟรมเท่ากับ 10 มิลลิวินาที และตัวจำแนกคือโมเดลมาร์คอฟซ่อนเร้น ผลการทดลองในกรณีของเทคนิคการปรับตัวได้ พบว่าวิธี

C-MLLR ให้ค่าความถูกต้องของอัตราความผิดพลาดของคำได้ดีกว่า MLLR ทั้งสามระดับความรุนแรงของโรค คือ เบา ปานกลาง และรุนแรง เมื่อใช้โมเดลคำพูดของผู้ป่วย

1.6 ขอบเขตของการวิจัย

1.6.1 วิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงในอาสาสมัครปกติและผู้ที่มีอาการพูดไม่เป็นความ

1.6.2 จำแนกพยางค์ไทยจำนวน 12 พยางค์ คือ อา อี อู คา คี คู นา นี นู มา มี และ มู สำหรับคนปกติเพื่อใช้ในการฟื้นฟูอาการพูดไม่เป็นความ

1.6.3 จำแนกพยางค์ไทยจำนวน 9 พยางค์ คือ คา คี คู นา นี นู มา มี และ มู สำหรับผู้ที่มีอาการพูดไม่เป็นความ

1.7 ระเบียบวิธีการวิจัย และขั้นตอนการดำเนินการวิจัย

1.7.1 ทบทวนวรรณกรรมที่เกี่ยวข้องกับการรู้จำคำพูด หน่วยเสียงและพยางค์ทั้งแบบที่ใช้สัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อ เพื่อให้เข้าใจถึงความรู้ทันสมัย (state of the art) ของงานวิจัยในปัจจุบัน

1.7.2 สร้างระบบเก็บสัญญาณทั้งสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อ โดยต้องเชื่อมต่อจุดเริ่มต้นและสิ้นสุดของสัญญาณทั้งสองเพื่อง่ายในการตัดแยกข้อมูล

1.7.3 เก็บข้อมูลทั้งสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อจากจำนวนอาสาสมัครทั้ง 7 คน โดยแบ่งเป็นชาย 4 คนและหญิง 3 คน ที่มีช่วงอายุระหว่าง 20–22 ปี โดยติดอิเล็กโทรดทั้งหมด 5 ช่องสัญญาณบริเวณใบหน้าช่วงล่างและบางส่วนของลำคอ

1.7.4 วิเคราะห์คุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 9 พยางค์ทั้งของคนปกติและผู้ที่มีอาการพูดไม่เป็นความ ด้วยการแทนที่ลักษณะเด่น 3 กลุ่มคือขนาดของสัญญาณ ความถี่ของสัญญาณ และค่าทางสถิติของการกระจายข้อมูล โดยเปรียบเทียบค่าความถูกต้องในการจำแนกพยางค์ของลักษณะเด่นทั้ง 3 กลุ่ม

1.7.5 เปรียบเทียบประสิทธิภาพของระบบจำแนกพยางค์โดยลดจำนวนช่องสัญญาณของอิเล็กโทรด โดยแบ่งเป็น 5 กลุ่มดังนี้ 1 ช่องสัญญาณ 2 3 4 และ 5 ช่องสัญญาณรวมกัน

1.7.6 วิเคราะห์ขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันของสัญญาณเสียงและจำนวนสัมประสิทธิ์ของ MFCC ที่เหมาะสมกับการจำแนกพยางค์

1.7.7 วิเคราะห์คุณลักษณะของสัญญาณเสียงทั้ง 12 พยางค์ด้วยการแทนที่ลักษณะเด่น 2 กลุ่ม โดยกลุ่มที่หนึ่งคือลักษณะเด่นในโดเมนเวลาที่นิยมใช้กับสัญญาณไฟฟ้ากล้ามเนื้อ และกลุ่มที่สองคือลักษณะเด่น MFCC

1.7.8 เปรียบเทียบประสิทธิภาพของระบบจำแนกพยางค์ของการรวมข้อมูลจากแหล่งที่มาเดียวกันคือสัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียว และการรวมข้อมูลจากหลายแหล่งที่มา หมายถึงสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง โดยวิธีการรวมข้อมูลอยู่ในระดับลักษณะเด่น (feature-level)

1.7.8 ทดลองฝึกสอนและทดสอบระบบแบบไม่ขึ้นกับบุคคล (independent-speaker)

1.7.9 วิเคราะห์ สรุปผล และจัดทำวิทยานิพนธ์ฉบับสมบูรณ์

1.8 ผลลัพธ์จากงานวิจัย

งานวิจัยนี้ศึกษาระบบจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ สัญญาณเสียง และการรวมข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง โดยมุ่งเน้นประสิทธิภาพของระบบจำแนกพยางค์ไทย

ในกรณีของระบบจำแนกพยางค์ไทยบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ การทดลองถูกแบ่งออกเป็น 2 ประเด็นหลักคือ

ก) ศึกษาคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อขณะออกเสียงพยางค์ไทยจำนวน 9 พยางค์ทั้งของคนปกติและผู้ที่มีอาการพูดไม่เป็นความ โดยแบ่งลักษณะเด่นออกเป็น 3 กลุ่มคือ ขนาดของสัญญาณ ความถี่ของสัญญาณ และค่าทางสถิติของการกระจายข้อมูล จากผลการทดลองพบว่าพารามิเตอร์ที่มีอิทธิพลต่อความถูกต้องในการจำแนกพยางค์คือขนาดของสัญญาณและความถี่ของสัญญาณ อย่างไรก็ตามใช้จำนวนอิเล็กโทรดทั้งหมด 5 ช่องสัญญาณส่งผลต่อการนำไปใช้ในทางปฏิบัติจริงหรือการรักษาทางคลินิกเนื่องจากจำนวนช่องสัญญาณของอิเล็กโทรดมากเกินไป

ข) เปรียบเทียบประสิทธิภาพในการจำแนกพยางค์ไทยจำนวน 9 พยางค์ด้วยเทคนิคการฉายข้อมูล 3 แบบคือ Principal Component Analysis (PCA) Linear Discriminant Analysis (LDA) และ spectral regression extreme learning machine (SRELM) โดยเทคนิคการฉายข้อมูลเป็นวิธีการลดมิติของ

ข้อมูล เพื่อสกัดลักษณะเด่นที่สามารถเพิ่มความถูกต้องในการจำแนกพยางค์และหลีกเลี่ยงการเกิด curse of dimension ผลการทดลองแสดงให้เห็นว่าเทคนิคการฉายข้อมูลแบบ SRELM มีประสิทธิภาพในการจำแนกพยางค์สูงสุดทั้งในคนปกติและผู้ที่มีอาการพูดไม่เป็นความ

ในส่วนของระบบรู้จำคำพูดบนพื้นฐานของสัญญาณเสียง จากการทบทวนวรรณกรรมด้านการนำเทคโนโลยีปัญญาประดิษฐ์มาประยุกต์ใช้กับผู้ที่มีอาการพูดไม่เป็นความพบว่างานวิจัยส่วนใหญ่เป็นการจำแนกระดับความรุนแรงของโรคเพื่อวางแผนในการรักษา โดยพบว่าลักษณะเด่นที่นิยมนำมาใช้คือ MFCC ในงานวิจัยนี้ผู้วิจัยได้แบ่งการทดลองออกเป็น 2 ประเด็นดังนี้

ก) เปรียบเทียบประสิทธิภาพในการจำแนกพยางค์ไทยจำนวน 12 พยางค์ โดยพิจารณา 2 ประเด็นย่อยคือจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน (overlapped frame size) ที่เหมาะสมในการจำแนกพยางค์ ในส่วนของจำนวนสัมประสิทธิ์ของ MFCC พบว่าวรรณกรรมส่วนใหญ่เลือกใช้จำนวนสัมประสิทธิ์เท่ากับ 13 ดังนั้นผู้วิจัยต้องการทราบแนวโน้มของความถูกต้องในการจำแนกพยางค์เมื่อจำนวนสัมประสิทธิ์ของ MFCC น้อยกว่าหรือมากกว่า 13 ในการทดลองผู้วิจัยกำหนดจำนวนสัมประสิทธิ์ของ MFCC ไว้ 3 ค่าดังนี้ 8, 13 และ 18 พบว่าเมื่อจำนวนสัมประสิทธิ์ของ MFCC เพิ่มขึ้น ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นด้วย โดยเฉพาะอย่างยิ่งเมื่อจำนวนสัมประสิทธิ์ของ MFCC เพิ่มจาก 8 เป็น 13 ในกรณีของขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันกำหนดไว้ 2 ค่าคือ ขนาดของเฟรมเท่ากับ 25 และขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 10 มิลลิวินาที เขียนแทนด้วย 25(10) และขนาดของเฟรมเท่ากับ 250 และขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 125 มิลลิวินาที เขียนแทนด้วย 250(125) เนื่องจากพบว่าขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันนิยมใช้คือ 25(10) และ 250(125) ถูกเลือกเนื่องจากผู้วิจัยต้องการนำลักษณะเด่น MFCC ไปต่อเรียง (concatenate) กับลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ โดยการรวมลักษณะเด่นมีข้อจำกัดในเรื่องของมิติของลักษณะเด่นที่จะมารวมกันคือลักษณะเด่นทั้ง 2 กลุ่มต้องมีจำนวนข้อมูล (sample) ทั้งข้อมูลสอนและทดสอบเท่ากัน รวมถึงคลาส (class) ของข้อมูลด้วย และ 250(125) เป็นค่าที่เหมาะสมสำหรับสัญญาณไฟฟ้ากล้ามเนื้อ ผลการทดลองแสดงให้เห็นว่าที่ 25(10) และจำนวนสัมประสิทธิ์เท่ากับ 18 ให้ค่าความถูกต้องสูงสุด อย่างไรก็ตามค่าความ

ถูกต้องในการจำแนกพยางค์ของ 25(10) และ 250(125) สำหรับทุกค่าของจำนวนสัมประสิทธิ์ของ MFCC มีความแตกต่างอย่างไม่มีนัยสำคัญ

ข) เปรียบเทียบค่าความถูกต้องในการจำแนกพยางค์โดยใช้ลักษณะเด่นในโดเมนเวลาจำนวน 5 ค่าที่นิยมนำไปใช้กับสัญญาณไฟฟ้ากล้ามเนื้อกับลักษณะเด่น MFCC พบว่าประสิทธิภาพในการจำแนกพยางค์เมื่อใช้ลักษณะเด่น MFCC ดีกว่าอย่างมีนัยสำคัญสำหรับอาสาสมัครปกติทั้ง 7 คน อย่างไรก็ตามข้อด้อยของสัญญาณเสียงคือง่ายต่อการถูกสัญญาณรอบข้างรบกวน ส่งผลให้ประสิทธิภาพในการจำแนกพยางค์ลดลง

สำหรับงานวิจัยในส่วนของการรวมข้อมูลแบ่งออกเป็น 2 ประเด็นคือการรวมข้อมูลจากแหล่งที่มาเดียวกันและการรวมข้อมูลจากหลายแหล่งที่มา โดยการรวมข้อมูลเป็นการนำลักษณะเด่นมาต่อเรียงกัน

ก) การรวมข้อมูลจากแหล่งที่มาเดียวกันเป็นการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อหลายช่องสัญญาณมาต่อเรียงกัน พารามิเตอร์ที่ถูกพิจารณาประกอบด้วยช่วงเวลาของการวิเคราะห์สัญญาณและจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ สำหรับช่วงเวลาของการวิเคราะห์สัญญาณมีดังนี้ 1.125, 1.75 และ 2.4 วินาที พบว่าค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นอย่างไม่มีนัยสำคัญเมื่อช่วงเวลาของการวิเคราะห์สัญญาณเพิ่มขึ้น ในส่วนของจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อถูกแบ่งออกเป็น 5 กลุ่มคือ 1 ช่องสัญญาณ 2, 3, 4 และ 5 ช่องสัญญาณรวมกัน โดยในแต่ละกลุ่มถูกแบ่งเป็นกลุ่มย่อยด้วยการจับคู่ช่องสัญญาณ จากนั้นช่องสัญญาณที่ดีที่สุดในแต่ละกลุ่มจะถูกเลือกเพื่อนำมาวิเคราะห์ในหัวข้อการรวมข้อมูลจากหลายแหล่งที่มา จากการทดลองพบว่าประสิทธิภาพของระบบจำแนกพยางค์ลดลงเมื่อจำนวนช่องสัญญาณลดลง อีกทั้งค่าเบี่ยงเบนมาตรฐานมีค่าสูงกว่าเมื่อเทียบกับการใช้ช่องสัญญาณทั้งหมด ดังนั้นประเด็นสำคัญคือมีวิธีการใดที่จะสามารถลดจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อโดยยังคงประสิทธิภาพของการจำแนกพยางค์

ข) ในส่วนของการรวมข้อมูลจากหลายแหล่งที่มาเป็นการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อมาต่อเรียงกับลักษณะเด่นของสัญญาณเสียงที่ช่วงเวลาเดียวกัน โดยพิจารณาพารามิเตอร์ 3 แบบคือช่วงเวลาของการวิเคราะห์สัญญาณ จำนวนสัมประสิทธิ์ของ MFCC และจำนวนช่องสัญญาณของ

สัญญาณไฟฟ้ากล้ามเนื้อ ผลการศึกษาสรุปได้ดังนี้ ประเด็นที่ 1 การเพิ่มช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงทำให้ประสิทธิภาพของระบบรู้จำคำพูดเพิ่มขึ้น ยกเว้นที่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาที ทั้งนี้อาจจะเป็นเพราะว่าแรงหดตัวของมัดกล้ามเนื้อทั้ง 5 ตำแหน่งของการวัดสัญญาณขณะออกเสียงในช่วงปลายค่อนข้างคงที่จึงทำให้สัญญาณมีความใกล้เคียงกันสำหรับทุกพยางค์ ประเด็นที่ 2 การเพิ่มจำนวนสัมประสิทธิ์ของ MFCC ส่งผลให้ประสิทธิภาพของระบบรู้จำคำพูดดีขึ้นอย่างมีนัยสำคัญเมื่อจำนวนสัมประสิทธิ์เพิ่มจาก 8 เป็น 13 ประเด็นที่ 3 การเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้น เมื่อจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อเพิ่มขึ้นและค่าเบี่ยงเบนมาตรฐานจะลดลง โดยเฉพาะอย่างยิ่งเมื่อจำนวนของช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อเท่ากับ 1 และ 2 ช่องสัญญาณรวมกันสามารถปรับปรุงระบบได้อย่างมีนัยสำคัญเนื่องจากค่าความถูกต้องเพิ่มขึ้นมากเมื่อเทียบกับกรณีที่ใช้สัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียว นอกจากนี้ยังพบว่าการรวมข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 5 ช่องสัญญาณมีประสิทธิภาพลดลงเมื่อเทียบกับการใช้เพียง 4 ช่องสัญญาณ

บทที่ 2

ทฤษฎีและหลักการ

บทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับงานวิจัยประกอบด้วย วิธีบำบัดการพูดอธิบายถึงกิจกรรมที่ใช้สำหรับบำบัดการพูดรวมถึงอวัยวะที่เกี่ยวข้องและการเลือกคำที่ใช้สำหรับบำบัดการพูดเบื้องต้น ตำแหน่งของมัดกล้ามเนื้อที่เกี่ยวข้องกับการพูดโดยกล่าวถึงมัดกล้ามเนื้อที่สัมพันธ์กับการพูด กระบวนการของการรู้จำคำพูดเริ่มจากการปรับปรุงสัญญาณก่อนการประมวลผล การแทนที่ลักษณะเด่น การลดมิติของข้อมูล และการจำแนกประเภท สุดท้ายกล้าวถึงวิธีการรวมข้อมูลทั้งจากแหล่งที่มาเดียวกันและจากหลายแหล่งที่มา

2.1 วิธีบำบัดการพูด

กิจกรรมที่ใช้ในการฝึกบำบัดการพูดแบ่งเป็นสองกิจกรรมหลักคือ กิจกรรมที่ใช้ในการแสดงทางสีหน้า และกิจกรรมเกี่ยวกับการพูด โดยกิจกรรมที่ใช้ในการแสดงทางสีหน้าได้แก่ การอ้าปาก การทำปากจู๋ การยิ้ม การแลบลิ้นตรง และการยกลิ้นขึ้น เป็นกิจกรรมที่ช่วยเพิ่มประสิทธิภาพและปรับปรุงการเคลื่อนไหวให้กับอวัยวะต่างๆ ที่เกี่ยวข้องกับการพูด ได้แก่ ริมฝีปาก ขากรรไกร และลิ้นซึ่งจะมีผลทำให้ผู้ป่วยสามารถเคลื่อนไหวอวัยวะดังกล่าวได้ดีขึ้น จึงมีส่วนในการช่วยแก้ไขคำพูดให้ดียิ่งขึ้น โดยพบว่าการบำบัดอวัยวะที่เกี่ยวกับการพูดโดยไม่ออกเสียง (non-speech oral motor treatments) สามารถนำมาใช้ในการพัฒนารูปแบบการเคลื่อนไหว และการจัดท่าของอวัยวะต่างๆที่เกี่ยวข้องกับการพูด [20][24] อย่างไรก็ตามในงานวิจัยนี้สนใจกิจกรรมเกี่ยวกับการพูด โดยคำที่ใช้ในการทำกิจกรรมเกี่ยวกับการพูดจะคำนึงถึงหลักการออกเสียงตามการใช้อวัยวะในการออกเสียง และความง่ายของการออกเสียง เนื่องจากผู้ป่วยมีภาวะกล้ามเนื้ออ่อนแรงบริเวณใบหน้าส่วนล่างและบางส่วนของลำคอ โดยปกติการออกเสียงของมนุษย์ประกอบด้วยสองส่วนหลักคือ ส่วนที่อยู่กับที่หรือเรียกว่าฐานกรณ์ (place of articulation) และส่วนที่เคลื่อนที่ได้ (place of manner) หรือลิ้น สำหรับพยางค์ที่ใช้ในการฝึกออกเสียงเบื้องต้นจะเลือกใช้พยัญชนะที่ออกเสียงง่ายที่สุด จากการทบทวนวรรณกรรมพบว่าการออกเสียงพยัญชนะที่ง่ายที่สุดที่เด็ก 2 ขวบสามารถทำได้คือ ค ควาย (การออกเสียงโดยการใช้โคนลิ้นกับเพดานอ่อน) น หนู (การออกเสียงโดยใช้ปลายลิ้นกับปุ่มเหงือก) และ ม ม้า (การออกเสียงโดยการใช้ริมฝีปากบนและล่าง) [30] โดยผสมกับสระสามตัวคือสระ “อา” “อี” และ “อู” ซึ่งเสียงสระดังกล่าวเป็นที่นิยมใช้ในการฝึกบำบัดการพูด เนื่องจากเป็นขีดสุดของเสียงสระในการอ้าปาก ยิ้ม และย่นริมฝีปาก หากผู้ป่วยสามารถออกเสียงสระดังกล่าวได้เสียงสระอื่นก็จะสามารถทำได้นั่นเอง ดังนั้นคำที่ใช้ในการบำบัดการพูดมีทั้งหมด 12 พยางค์ดังตารางที่ 2-1

ตารางที่ 2-1 กลุ่มของพยางค์ภาษาไทยซึ่งใช้สำหรับการบำบัดการพูดเบื้องต้น

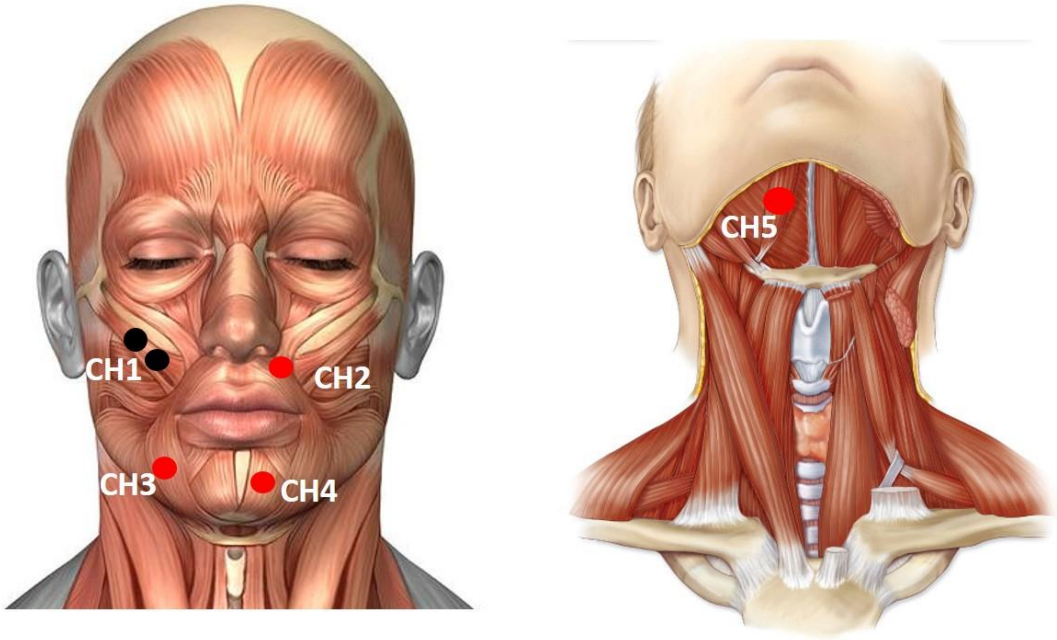
สระ / พยัญชนะ	อา	อี	อู
ค	คา	คี	คู
น	นา	นี	นู
ม	มา	มี	มู

2.2 ตำแหน่งของมัดกล้ามเนื้อที่เกี่ยวข้องกับการพูด

ตำแหน่งของอิเล็กโทรดเป็นองค์ประกอบที่มีความสำคัญและส่งผลต่อประสิทธิภาพของระบบรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ แต่ด้วยจำนวนอิเล็กโทรดที่มากเกินไป [22] ทำให้การบันทึกสัญญาณไฟฟ้ากล้ามเนื้อบนใบหน้าทำได้ยากในทางปฏิบัติและส่งผลต่อการนำไปประยุกต์ใช้งานจริง จากที่ได้กล่าวไว้ในหัวข้อการทบทวนวรรณกรรมด้านการรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อเรื่องตำแหน่งการวางอิเล็กโทรดสรุปได้ว่ามัดกล้ามเนื้อที่เกี่ยวข้องกับการพูด ได้แก่ กล้ามเนื้อ Zygomaticus major (CH1); 2) กล้ามเนื้อ Levator anguli oris (CH2); 3) กล้ามเนื้อ Depressor anguli oris (CH3); 4) Mentalis (CH4) และ 5) กล้ามเนื้อ Anterior belly of the digastrics (CH5) [44] ดังแสดงในภาพประกอบที่ 2-1 นอกจากนี้ยังมีวรรณกรรมที่ใช้มัดกล้ามเนื้ออื่นอีกเช่น กล้ามเนื้อ กล้ามเนื้อ Mylohyoid กล้ามเนื้อ Orbicularis oris และกล้ามเนื้อ Platysma [18][19] เป็นต้น จากการทบทวนวรรณกรรมพบว่ากล้ามเนื้อ Zygomaticus major มีความเกี่ยวข้องกับการยิ้ม โดยค่าเฉลี่ยของกำลังของสัญญาณไฟฟ้ากล้ามเนื้อมีค่าสูงสุดในระหว่างยิ้ม กล้ามเนื้อ Mentalis ทำงานร่วมกับกล้ามเนื้อ Depressor anguli oris มีความเกี่ยวข้องกับการทำปากจู๋ โดยค่าเฉลี่ยของกำลังของสัญญาณไฟฟ้ากล้ามเนื้อมีค่าสูงสุดในการทำปากจู๋ [24]

2.3 การปรับปรุงสัญญาณก่อนการประมวลผล

กระบวนการของการปรับปรุงสัญญาณก่อนการประมวลผล (Signal pre-processing) สามารถแบ่งเป็น 2 หัวข้อหลักคือการกำจัดสัญญาณรบกวนและการตัดแยกสัญญาณ (Segmentation) มีรายละเอียดต่างๆ ดังต่อไปนี้



ภาพประกอบที่ 2-1 ตำแหน่งของกล้ามเนื้อในการติดอิเล็กโทรด โดยจุดสีแดงคือการวัดแบบขั้วเดียว และจุดสีดำคือการวัดแบบสองขั้ว

2.3.1 การกำจัดสัญญาณรบกวน

จุดประสงค์ของการกำจัดสัญญาณรบกวนคือเพื่อเพิ่มคุณภาพของสัญญาณ ในกรณีของสัญญาณเสียงการกำจัดสัญญาณรบกวนนิยมใช้วงจรกรองบัตเตอร์เวิร์ธแบบความถี่ต่ำผ่าน (Butterworth low pass filter) อันดับ 8 ที่ความถี่ตัด (cut-off frequency) 5 กิโลเฮิรซต์ [18][19] ในส่วนของสัญญาณไฟฟ้ากล้ามเนื้อ สัญญาณรบกวนที่เกิดจากการวัดสัญญาณไฟฟ้ากล้ามเนื้อมีหลายแหล่งที่มาเช่น จากการเคลื่อนของสายนำสัญญาณ (movement artifact) จากสายส่ง (power line) จากกลุ่มของกล้ามเนื้อบริเวณใกล้เคียง (crosstalk) และจากสัญญาณไฟฟ้ากล้ามเนื้อหัวใจ (electrocardio graphic artifacts) เป็นต้น [14] วงจรกรองทั้งแบบแอนะล็อกและดิจิตัลถูกใช้ในการประมวลผลสัญญาณประกอบด้วยตัวกรองความถี่ต่ำผ่าน (low pass filter) ตัวกรองความถี่สูงผ่าน (high pass filter) นอตช์ฟิลเตอร์ (notch filter) ตัวกรองแถบความถี่ผ่าน (band pass filter) และตัวกรองแบบปรับค่าได้ (adaptive filter) นอกจากที่กล่าวมานี้ยังมีวิธีอื่นๆ ที่สามารถแยกแยะสัญญาณรบกวนออกจากสัญญาณไฟฟ้ากล้ามเนื้อเช่น วิธีเวฟเลท (wavelet transform) วิธีค่าสถิติอันดับสูง (higher order statistics) และการใช้อัลกอริทึมการวิเคราะห์องค์ประกอบอิสระ (independent component analysis) เป็นต้น จากการทบทวนวรรณกรรมพบว่าวิธีเวฟเลทสามารถกำจัดสัญญาณรบกวนได้อย่างมีประสิทธิภาพ โดยฟังก์ชันของเวฟเลทที่นิยมใช้คือ db2, db7, sym2, sym5, coif4, bior5.5 และ rbio2.2 [32] อย่างไรก็ตามจากการทบทวนวรรณกรรมด้านการ

รู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อพบว่าการกำจัดสัญญาณรบกวนนิยมใช้ตัวกรองแถบความถี่ผ่านที่ความถี่ตัดผ่านระหว่าง 0.1 ถึง 500 เฮิรซต์ เนื่องจากช่วงคลื่นความถี่ที่สำคัญของสัญญาณไฟฟ้ากล้ามเนื้ออยู่ในช่วงประมาณ 0 – 500 เฮิรซต์ สำหรับในงานวิจัยนี้ใช้วงจรกรองความถี่แบบดิจิตอลที่มีผลตอบสนองอิมพัลส์จำนวนจำกัด (finite impulse response: FIR) ซึ่งจะอธิบายในหัวข้อ 2.8

2.3.2 การตัดแยกสัญญาณ

การตัดแยกสัญญาณ (segmentation) เป็นการตรวจจับจุดเริ่มต้นและสิ้นสุดสัญญาณไฟฟ้ากล้ามเนื้อ โดยส่วนใหญ่สามารถแบ่งออกเป็นการตัดแยกด้วยมือ (manual segmentation) และการตัดแยกอัตโนมัติขึ้นอยู่กับวัตถุประสงค์ของการใช้งานเป็นหลัก ซึ่งแต่ละเทคนิคก็จะเหมาะสมและทำงานได้ดีกับสัญญาณไฟฟ้าในบางกล้ามเนื้อในแต่ละบริเวณของร่างกายเท่านั้น จากการทบทวนวรรณกรรมด้านการรู้จำคำพูดพบว่าส่วนใหญ่เป็นการตัดแยกด้วยมือ (manually segmented) [19][33]

2.4 การแทนที่ลักษณะเด่น

การแทนที่ลักษณะเด่น (feature representation) คือวิธีสกัดข้อมูลที่มีประโยชน์ซึ่งซ่อนอยู่ในสัญญาณไฟฟ้ากล้ามเนื้อและขจัดส่วนที่ไม่ต้องการและสิ่งรบกวน โดยทั่วไปลักษณะเด่นที่ใช้ในการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อสามารถแบ่งได้เป็น 3 กลุ่มหลักคือลักษณะเด่นในโดเมนเวลา โดเมนความถี่และโดนเมนเวลา-ความถี่หรือการแทนที่บนสเกลเวลาและความถี่ นอกจากนี้พบว่าค่าทางสถิติของการกระจายข้อมูลได้ถูกนำมาใช้กับสัญญาณไฟฟ้ากล้ามเนื้อเพื่อที่จะแยกสัญญาณรบกวนออกจากสัญญาณไฟฟ้ากล้ามเนื้ออีกด้วย [35] ในงานวิจัยนี้มุ่งเน้นลักษณะเด่น 3 กลุ่มคือลักษณะเด่นในโดเมนเวลา โดเมนความถี่และค่าทางสถิติของการกระจายข้อมูล คำนิยามเชิงคณิตศาสตร์ของลักษณะเด่นแสดงรายละเอียดดังนี้

2.4.1 ลักษณะเด่นในโดเมนเวลา

ในส่วนของสัญญาณไฟฟ้ากล้ามเนื้อ ลักษณะเด่น 5 ค่าถูกใช้ในการคำนวณประกอบด้วยค่าเฉลี่ยของค่าสัมบูรณ์ (MAV) ความยาวคลื่น (WL) การตัดผ่านค่าศูนย์ (ZC) การเปลี่ยนแปลงเครื่องหมายของความชัน (SSC) และสัมประสิทธ์ความถดถอยอันดับที่สี่ (fourth-order autoregressive (AR)) เนื่องจากลักษณะเด่นเหล่านี้ให้ผลดีในระบบจำแนกประเภทที่ใช้กับสัญญาณไฟฟ้ากล้ามเนื้อ [26][35] และ[36]

ค่าเฉลี่ยของค่าสัมบูรณ์เป็นลักษณะเด่นที่นิยมใช้ในการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อ โดยค่าเฉลี่ยของค่าสัมบูรณ์ของขนาดของสัญญาณไฟฟ้ากล้ามเนื้อแสดงดังสมการที่ (2-1)

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{2-1}$$

เมื่อ x_i คือสัญญาณไฟฟ้ากล้ามเนื้อในช่วงของสัญญาณย่อยที่ถูกพิจารณาและ N คือจำนวนของสัญญาณไฟฟ้ากล้ามเนื้อที่ใช้ในการคำนวณ

ความยาวคลื่นเป็นลักษณะเด่นที่ใช้วัดความซับซ้อนของสัญญาณไฟฟ้ากล้ามเนื้อ โดยความยาวคลื่นของสัญญาณไฟฟ้ากล้ามเนื้อในช่วงของสัญญาณย่อยที่ถูกพิจารณา คำนวณได้จากผลรวมของผลต่างของสัญญาณไฟฟ้ากล้ามเนื้อดังสมการที่ (2-2)

$$WL = \sum_{i=1}^{N-1}|x_{i+1} - x_i| \tag{2-2}$$

การตัดผ่านค่าศูนย์คือการวัดค่าความถี่ของข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อซึ่งแสดงจำนวนครั้งของสัญญาณไฟฟ้ากล้ามเนื้อที่ตัดผ่านค่าศูนย์ดังสมการที่ (2-3) และ (2-4) เพื่อหลีกเลี่ยงสัญญาณผันผวนที่แรงดันต่ำหรือสัญญาณรบกวนจึงมีการตั้งค่าระดับอ้างอิงซึ่งมีค่าเท่ากับ 30 มิลลิโวลท์

$$ZC = \sum_{i=1}^{N-1}[f(x_i \times x_{i+1})\ and\ |x_i - x_{i+1}| \geq 30\ mV] \tag{2-3}$$

$$f(x) = \begin{cases} 1, & if\ x < 0 \\ 0, & otherwise \end{cases} \tag{2-4}$$

การเปลี่ยนแปลงเครื่องหมายของความชันเป็นลักษณะเด่นที่แสดงถึงจำนวนครั้งของการเปลี่ยนแปลงความชันของสัญญาณไฟฟ้ากล้ามเนื้อระหว่างค่าบวกและค่าลบโดยคำนวณได้ตามสมการที่ 2-5 และ 2-6 สามารถใช้เป็นข้อมูลสนับสนุนความถี่ของสัญญาณ โดยมีค่าแรงดันอ้างอิงเท่ากับ 10 มิลลิโวลท์

$$SSC = \sum_{i=2}^{N}[s\{(x_i - x_{i-1})(x_i - x_{i+1})\} \cap \{|x_i - x_{i-1}| \geq 10\ \cup\ |x_i - x_{i+1}| \geq 10\}] \tag{2-5}$$

$$s(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & otherwise \end{cases} \tag{2-6}$$

สัมประสิทธิ์การถดถอย โมเดลประมาณค่าสัมประสิทธิ์การถดถอย (AR) คือโมเดลทำนายซึ่งประมาณค่าปัจจุบันโดยใช้ผลรวมเชิงเส้นของค่าที่สังเกตได้ก่อนหน้า (x_{i-p}) รวมกับ white Gaussian noise (w_p) แสดงโมเดลตามสมการที่ 2-7 ดังนี้

$$AR = \sum_{p=0}^{P-1} a_p x_{i-p} + w_p \tag{2-7}$$

โดยที่ P คืออันดับของโมเดลประมาณค่าสัมประสิทธิ์การถดถอย ซึ่งในงานวิจัยนี้ใช้ค่า P เท่ากับ 4

2.4.2 ลักษณะเด่นในโดเมนความถี่

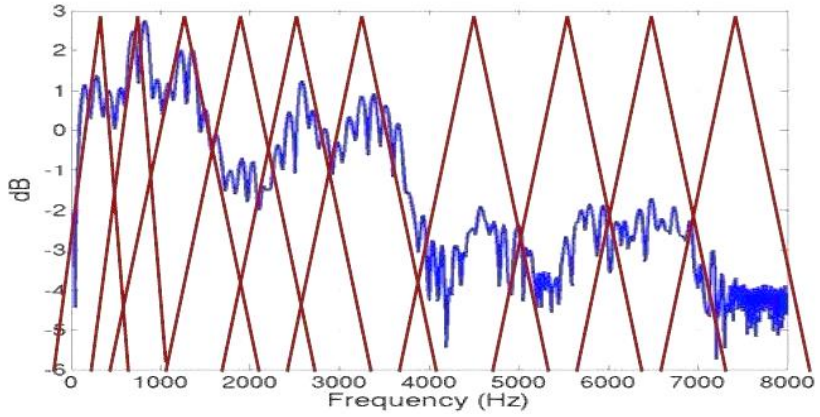
ค่าความถี่กลางคือความถี่เฉลี่ยซึ่งคำนวณจากผลรวมของผลคูณของสเปกตรัมกำลังของสัญญาณไฟฟ้ากล้ามเนื้อกับความถี่หารด้วยผลรวมทั้งหมดของสเปกตรัมกำลังดังสมการ 2-8

$$MNF = \sum_{j=1}^{M} f_j P_j \Big/ \sum_{j=1}^{M} P_j \tag{2-8}$$

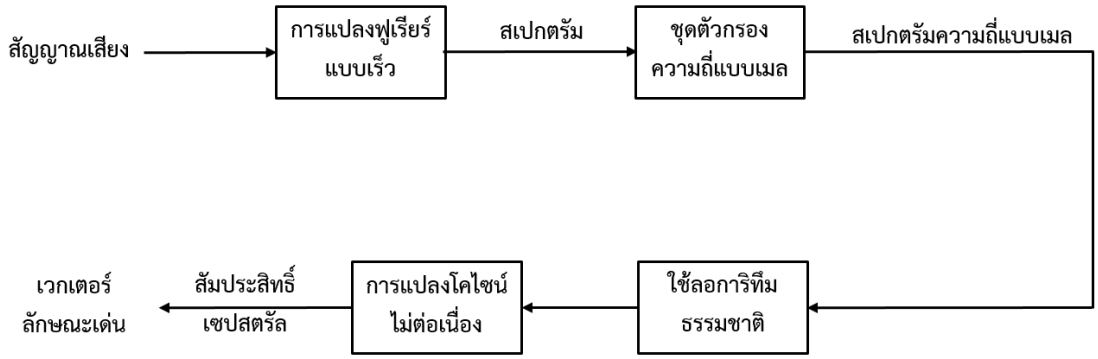
โดย f_j คือความถี่ของสเปกตรัมที่ช่วงความถี่ย่อย j P_j คือสเปกตรัมกำลังของสัญญาณไฟฟ้ากล้ามเนื้อที่ช่วงความถี่ย่อย j และ M คือจำนวนของช่วงความถี่ย่อยทั้งหมด

สัมประสิทธิ์เซปสตรัลที่คำนวณบนแกนความถี่แบบเมล

ในการวิเคราะห์สัญญาณเสียงนิยมนำลักษณะเด่น MFCC มาใช้ เนื่องจากชุดตัวกรองความถี่แบบเมลถูกออกแบบให้คล้ายกับการรับรู้ในการได้ยินเสียงของคน คุณลักษณะของชุดตัวกรองความถี่แบบเมลคือมีความหนาแน่นมากบริเวณความถี่ต่ำ (ที่หูสามารถรับรู้ได้) และมีความหนาแน่นน้อยบริเวณความถี่สูง ดังภาพประกอบที่ 2-2 และแผนภาพบล็อกของอัลกอริทึม MFCC แสดงดังภาพประกอบที่ 2-3 รายละเอียดของการคำนวณลักษณะเด่น MFCC อธิบายได้ดังนี้



ภาพประกอบที่ 2-2 รูปแบบของชุดตัวกรองความถี่แบบเมล [37]



ภาพประกอบที่ 2-3 แผนภาพบล็อกของอัลกอริทึม MFCC

ขั้นตอนที่ 1 ตัดแยกสัญญาณเสียงให้มีขนาดเล็กเรียกว่าเฟรม

ขั้นตอนที่ 2 คำนวณขนาดของสเปกตรัมโดยใช้วิธีการแปลงฟูเรียร์แบบเร็ว (fast Fourier transform: FFT)

ขั้นตอนที่ 3 ใช้ชุดตัวกรองความถี่แบบเมลกับสเปกตรัมที่ได้ในขั้นตอนที่ 2 โดยตัวกรองเป็นแบบสามเหลี่ยม การแปลงจากความถี่ซึ่งมีหน่วยเป็นเฮิรซต์เป็นสเกลของเมลถูกนำเสนอใน [38] แสดงดังสมการ (2-9) จากนั้นรวมพลังงานในแต่ละตัวกรอง

$$Mel(f) = 1125 \ln(1 + f/700) \tag{2-9}$$

ขั้นตอนที่ 4 นำฟังก์ชันลอการิทึมไปใช้กับพลังงานของชุดตัวกรองความถี่ทั้งหมด เพื่อแยกสเปกตร้มออกเป็น spectral envelop ซึ่งมีความถี่ต่ำและ spectral details ซึ่งมีความถี่สูง

ขั้นตอนที่ 5 ทำการแปลงโคไซน์แบบไม่ต่อเนื่องกับพลังงานที่ได้ในขั้นตอนที่ 4 เอาต์พุตของการแปลงโคไซน์แบบไม่ต่อเนื่องของ spectral envelop คือ MFCC ซึ่งแทนด้วยเวกเตอร์ลักษณะเด่นของสัญญาณเสียง

2.4.3 ค่าทางสถิติของการกระจายข้อมูล

ค่าความโด่ง ($L-KURT$) ของการกระจายข้อมูลคือการวัดจุดสูงสุดหรือความโด่งของการกระจายขนาดของสัญญาณไฟฟ้ากล้ามเนื้อ ถูกนิยามโดยใช้อัตราส่วนของ L-momentl อันดับ 4 ต่อ L-moment อันดับ 2 และคำนวณจากผลรวมเชิงเส้นของสถิติเชิงอันดับของตัวแปรแบบสุ่ม แสดงดังสมการที่ (2-10) [34]

$$L-KURT = \frac{20b_3 - 30b_2 + 12b_1 - b_0}{2b_1 - b_0} \tag{2-10}$$

โดยที่ b_r คือตัวประมาณค่าแบบ unbiased ของ probability-weight moment เมื่อ $r = 0, 1, 2, 3$ สมการทั่วไปของ b_r แสดงดังสมการที่ (2-11)

$$b_r = \frac{1}{N}\sum_{j=r+1}^{N} y_j \left[\frac{(j-1)(j-2)\cdots(j-r)}{(N-1)(N-2)\cdots(N-r)}\right] \quad (2\text{-}11)$$

โดยค่า y_j หาได้จากการเรียงข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อที่ผ่านการนอร์มซึ่งมีค่าเฉลี่ยเท่ากับศูนย์และค่าแปรปรวนเท่ากับหนึ่งโดยเรียงลำดับจากน้อยไปมากจากข้อมูลที่ 1 ถึง *N* ค่ามาตรฐานของ $L-KURT$ สำหรับการแจกแจงแบบปกติของเกาส์เซียนและลาปลาเซียนคือ 0.1226 และ 0.2357 ตาม ลำดับ

ค่าความเบ้ ($L-SKW$) ของการกระจายของข้อมูลคือการวัดมิติของความไม่สมมาตรของการกระจายขนาดของสัญญาณไฟฟ้ากล้ามเนื้อซึ่งอาจจะเป็นค่าบวกหรือค่าลบ สำหรับการกระจายข้อมูลแบบปกติค่า $L-SKW$ จะอยู่ในช่วง 0 ถึง 1 การคำนวณค่า $L-SKW$ ของชุดข้อมูลแสดงดังสมการที่ (2-12) [39]

$$L-SKW = \frac{6b_2 - 6b_1 + b_0}{2b_1 - b_0} \quad (2\text{-}12)$$

2.5 การลดมิติของข้อมูล

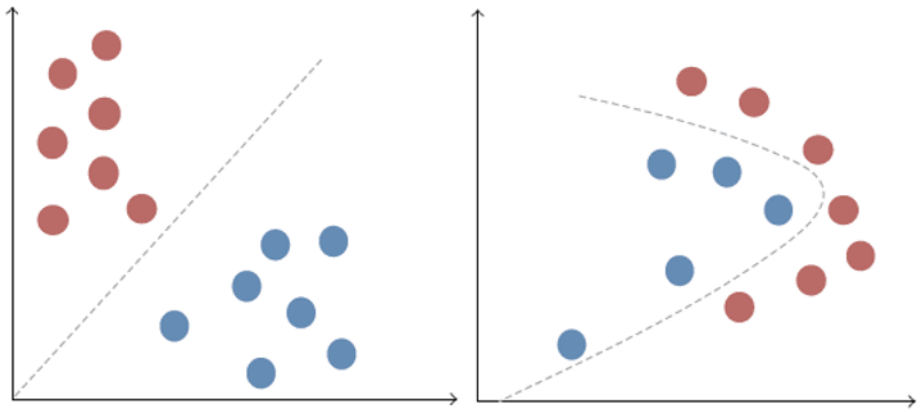
การลดมิติของข้อมูล (dimensional reduction) เป็นขั้นตอนหนึ่งที่มีความสำคัญในการคัดเลือกลักษณะเด่นที่สามารถเพิ่มความถูกต้องของการจำแนกและลดเวลาที่ใช้ในการประมวลผล นอกจากนี้ยังช่วยหลีกเลี่ยงปัญหา curse of dimension ซึ่งหมายถึงโมเดลการสอนให้ผลดีมากแต่เมื่อนำโมเดลนั้นมาใช้กับข้อมูลทดสอบพบว่าความถูกต้องของการจำแนกมีค่าน้อย โดยทั่วไปการลดมิติของข้อมูลสามารถแบ่งออกเป็น 2 แบบคือการเลือกลักษณะเด่นและการฉายลักษณะเด่น

2.5.1 การเลือกลักษณะเด่น

การเลือกลักษณะเด่น (feature selection) เป็นวิธีที่ง่ายที่สุดสำหรับการลดมิติของข้อมูลโดยการกำจัดลักษณะเด่นที่ไม่สัมพันธ์หรือมีความซ้ำซ้อนโดยไม่ปรับเปลี่ยนข้อมูลเดิม อัลกอริทึมของการเลือกลักษณะเด่นแบ่งได้เป็น 3 ประเภทคือประเภทที่ 1 คือ filters ซึ่งสกัดลักษณะเด่นจากข้อมูลที่ปราศจากการกระบวนการเรียนรู้ ประเภทที่ 2 คือ wrappers ใช้เทคนิคการเรียนรู้เพื่อประเมินลักษณะเด่นที่มีประโยชน์ และประเภทที่ 3 คือ embedded เป็นการรวมขั้นตอนของการเลือกลักษณะเด่นและการจำแนกประเภท [40]

2.5.2 การฉายลักษณะเด่น

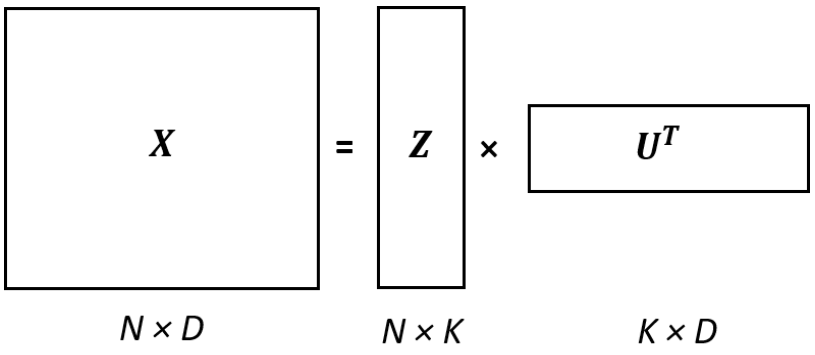
การลดมิติของข้อมูลด้วยการฉายลักษณะเด่น (feature projection) เป็นการสร้างตัวแปรใหม่ อัลกอริทึมของการฉายข้อมูลแบ่งได้เป็นแบบเชิงเส้นและไม่เป็นเชิงเส้นแสดงดังภาพประกอบที่ 2-4(ก) และ 2-4(ข) มีรายละเอียดดังนี้



ภาพประกอบที่ 2-4 การฉายลักษณะเด่น (ก) แบบเป็นเชิงเส้น (ข) แบบไม่เป็นเชิงเส้น [40]

ก) แบบเชิงเส้น

การฉายลักษณะเด่นแบบเชิงเส้นเป็นการสมมติว่าข้อมูลที่ถูกฉายแล้วตั้งอยู่บนปริภูมิย่อยเชิงเส้นที่มีมิติต่ำกว่าโดยใช้วิธีแยกตัวประกอบเมตริกซ์ กำหนดให้ชุดข้อมูล (X) มีมิติเท่ากับ $N \times D$ เมตริกซ์การฉายข้อมูล (U) มีมิติเท่ากับ $D \times K$ ข้อมูลที่ฉายแล้ว (Z) มีมิติเท่ากับ $N \times K$ โดยที่ $Z = X \cdot U$ และ $UUT = I$ (คุณสมบัติเชิงตั้งฉากของเวกเตอร์ลักษณะเฉพาะ (eigen vector)) ดังนั้นจะได้ว่า $X = Z \cdot U^T$ แสดงทางกราฟิกดังภาพประกอบที่ 2-5



ภาพประกอบที่ 2-5 การลดมิติของข้อมูลโดยใช้การแยกตัวประกอบเมตริกซ์เชิงเส้นฉายข้อมูลบนปริภูมิย่อยที่มีมิติต่ำกว่า

อัลกอริทึมการลดมิติข้อมูลที่เป็นที่รู้จักกันมากที่สุดคือการวิเคราะห์องค์ประกอบหลัก (principal component analysis: PCA) โดยใช้เมตริกซ์ความแปรปรวนร่วมและค่าลักษณะเฉพาะ (eigen value) และเวกเตอร์ลักษณะเฉพาะ (eigen vector) การวิเคราะห์องค์ประกอบหลักเป็นการหาองค์ประกอบหลักของข้อมูลซึ่งไม่สัมพันธ์กับเวกเตอร์ลักษณะเฉพาะและองค์ประกอบหลักแต่ละตัวแทนที่สัดส่วนความแปรปรวนของข้อมูล อย่างไรก็ตามในการคำนวณองค์ประกอบหลักของชุดข้อมูลไม่สามารถรับประกันได้ว่าองค์ประกอบหลักเหล่านั้นจะสัมพันธ์กับคลาสของชุดข้อมูลซึ่งเป็นข้อบกพร่องของการแยกข้อมูล ดังนั้นการวิเคราะห์จำแนกกลุ่มเชิงเส้น (linear discriminant analysis: LDA) ถูกนำเสนอ โดยนำข้อมูลของคลาสมาใช้ในการวิเคราะห์ด้วยส่งผลให้จำแนกประเภทได้ดีกว่า ความคิดพื้นฐานของการวิเคราะห์จำแนกกลุ่มเชิงเส้นคือการฉายข้อมูลไปยังปริภูมิเวกเตอร์จำแนกกลุ่มที่เหมาะสมที่สุดเพื่อที่จะบรรลุผลของการจำแนกประเภท โดยมีระยะทางระหว่างคลาสมากที่สุดและระยะทางในคลาสเดียวกันน้อยที่สุดพร้อมทั้งบีบอัดมิติปริภูมิลักษณะเด่นด้วย

ข) แบบไม่เป็นเชิงเส้น

การลดมิติแบบไม่เป็นเชิงเส้นมีลักษณะแตกต่างจากแบบเชิงเส้น ยกตัวอย่างเช่นพื้นผิวที่มีมิติต่ำสามารถถูกแปลงไปยังปริภูมิที่มีมิติสูงได้ ดังนั้นความสัมพันธ์แบบไม่เป็นเชิงเส้นระหว่างลักษณะเด่นสามารถถูกค้นพบ ในทางทฤษฎีลิฟติ้งฟังก์ชัน (lifting function) สามารถใช้แปลงลักษณะเด่นไปยังปริภูมิที่มีมิติสูงได้ ในปริภูมิที่มีจำนวนมิติที่สูงขึ้นความสัมพันธ์ระหว่างลักษณะเด่นสามารถมองเห็นเป็นแบบเชิงเส้น ดังนั้นจึงตรวจจับได้ง่ายขึ้น และเมื่อแปลงกลับไปยังปริภูมิที่มีมิติต่ำกว่าสามารถมองเห็นความสัมพันธ์แบบไม่เป็นเชิงเส้น ในทางปฏิบัติฟังก์ชันเคอร์เนลถูกออกแบบเพื่อสร้างผลลัพธ์เดียวกันโดยไม่ต้องคำนวณลิฟติ้งฟังก์ชัน ในวรรณกรรม [36] นำเสนอวิธีการฉายข้อมูลแบบไม่เป็นเชิงเส้นเรียกว่า spectral regression extreme learning Machine (SRELM) ซึ่งเป็นการรวมเทคนิคของ extreme learning machine (ELM) และ spectral regression (SR) โครงสร้างของ ELM เหมือนกับโครงข่ายประสาทเทียมแบบไปข้างหน้าชั้นเดียว โดยใช้หลักการสุ่มเพื่อทำการกำหนดค่าน้ำหนักในชั้นอินพุต และค่าเอนเอียงในชั้นซ่อน ขณะที่ค่าถ่วงน้ำหนักของชั้นเอาต์พุตถูกคำนวณด้วยค่ากำลังสองน้อยที่สุด นั่นแสดงว่าไม่มีการทำซ้ำในกระบวนการสอนข้อมูลส่งผลให้การสอนข้อมูลด้วยเทคนิค ELM เร็วกว่าเมื่อเปรียบเทียบกับโครงข่ายประสาทเทียมแบบไปข้างหน้าชั้นเดียวแบบดั้งเดิม อย่างไรก็ตามเทคนิค SRELM สามารถเพิ่มประสิทธิภาพของ ELM ด้วยการนำ SR มาใช้คำนวณค่าถ่วงน้ำหนักเอาต์พุต (output weight) ข้อดีของวิธีนี้คือสามารถนำคลาสมาพิจารณาในกระบวนการเรียนรู้ เทคนิค SR คือการวิเคราะห์เชิงสเปกตรัมของกราฟลาปลาเซียนถูกแก้ปัญหาด้วยวิธีการถดถอยกำลังสองน้อยที่สุด โดย SR สร้างเวกเตอร์

ลักษณะเฉพาะซึ่งฉายพื้นที่อินพุตไปยังพื้นที่เอาต์พุต พารามิเตอร์เพียงสองตัวถูกนำมาใช้เพื่อทำให้ประสิทธิภาพของ SRELM ดีที่สุดนั่นคือจำนวนของโหนดชั้นซ่อนซึ่งแปรผันจาก 100 ถึง 1500 โหนด โดยเพิ่มทีละ 100 โหนด และค่าแอลฟาซึ่งแปรผันจาก 10 ถึง 20 โดยเพิ่มทีละ 1

จากที่กล่าวไปแล้วข้างต้นส่วนองขั้นตอนวิธีการทำงานของ SRELM คล้ายคลึงกับโครงข่ายประสาทเทียมแบบไปข้างหน้าชั้นเดียว นั่นคือแทนที่จะทำการปรับค่าถ่วงน้ำหนักในชั้นซ่อนของโครงข่ายประสาทเทียมแบบไปข้างหน้าชั้นเดียว ELM จะกำหนดค่าถ่วงน้ำหนักในชั้นซ่อนแบบสุ่มและคำนวณค่าถ่วงน้ำหนักเอาต์พุท สำหรับตัวอย่างที่แตกต่างกันในรูปแบบ $\{(x_i, t_i)\}_{i=1}^{N}$ โดยที่ $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in R^n$ และ $t_i = (t_{i1}, t_{i2}, \dots, t_{in})^T \in R^m,\ i = 1, 2, \dots, N$ และ N คือจำนวนข้อมูลทั้งหมด ค่าเอาต์พุตของโครงข่ายประสาทเทียมแบบไปข้างหน้าชั้นเดียวที่มี K โหนดชั้นซ่อนแสดงดังสมการที่2.13

$$f(x_i) = \sum_{i=1}^{K} \beta_j\, G(a_j, b_j, x_i) = h(x_i)\beta = t_i \tag{2-13}$$

โดยที่ f คือเอาต์พุตของ ELM, G คือชั้นซ่อนของชั้นเอาต์พุต (hidden layer output), $h(x_i) \in R^{N\times K}$ คือเมตริกซ์ของชั้นซ่อนของชั้นเอาต์พุตและ $\beta \in R^{K\times m}$ คือเมตริกซ์ของค่าถ่วงน้ำหนักเอาต์พุต จากสมการที่ 2-13 เพื่อที่จะแก้ปัญหาสามารถประมาณค่าถ่วงน้ำหนักเอาต์พุต โดยวิธีคำนวณค่าน้อยที่สุดของ ELM จะลดผลรวมของกำลังสองของค่าสูญเสียของค่าผิดพลาด แสดงดังสมการที่ 2.14

$$\begin{aligned} &\text{minimize} && L_{ELM} = \tfrac{1}{2}\|\beta\|^2 + C\tfrac{1}{2}\textstyle\sum_{i=1}^{N}\|e\|^2 \\ &\text{subject to} && h(x_i)\beta = t_i^T - e_i^T,\ \ i = 1, \dots\dots, N \end{aligned} \tag{2-14}$$

จากนั้นแทนลงใน objective function จะได้สมการ2.15

$$\text{minimize} \qquad L_{ELM} = \tfrac{1}{2}\|\beta\|^2 + C\tfrac{1}{2}\textstyle\sum_{i=1}^{N}\|T - H\beta\|^2 \tag{2-15}$$

โดยที่ $H = [(h(x_1), \dots, h(x_N)^T]^T \in R^{N\times K}$ และ $T \in R^{N\times m}$ เพื่อที่จะประยุกต์ ELM สำหรับลดมิติของข้อมูล เราจะไม่พิจารณาคลาสของข้อมูล ดังนั้น objective function จะถูกปรับเปลี่ยนแสดงได้ดังสมการ 2.16

$$\begin{aligned} &\text{minimize} && L_{ELM} = \tfrac{1}{2}\left\|\beta\right\|^2 + \lambda\tfrac{1}{2}Tr(F^T L F) \\ &\text{subject to} && f_i = h(x_i)\beta \qquad i = 1, \dots\dots, N \end{aligned} \tag{2-16}$$

โดยที่ L_{ELM} คือฟังก์ชันเป้าหมาย, F คือเมตริกซ์เอาต์พุตและ L คือกราฟลาปลาเซียน โดยการแทนที่ลงใน objective function จะได้สมการดังนี้ 2.17

$$\begin{aligned} \text{minimize} \quad & L_{ELM} = \frac{1}{2}\|\beta\|^2 + \lambda\frac{1}{2}Tr(\beta^T H^T L\beta H) \\ \text{subject to} \quad & \beta^T H^T L\beta H = I_m \end{aligned} \tag{2-17}$$

วิธีแก้ปัญหาที่ดีที่สุด (optimization solution) ของสมการ (2-15) คือการแก้ปัญหาด้วยวิธีค่าลักษณะเฉพาะ (eigen value) แสดงดังสมการ

$$(I_L + \lambda H^T LH)u = \gamma H^T LHu \tag{2-18}$$

จากนั้นนำการวิเคราะห์กราฟสเปกตรัมมาใช้ซึ่งเป็นการนำ SR มาใช้คำนวณค่าถ่วงน้ำหนักเอาต์พุต โดยสมมติว่าการฉายของกราฟไปบนแกน y แบบฟังก์ชันเชิงเส้นแสดงดังสมการ 2-19

$$y = Hu \tag{2-19}$$

ผลที่ตามมาคือสมการ (2-18) ถูกจัดรูปแบบใหม่ได้สมการดังนี้

$$(I_L + \lambda H^T LH)u = \gamma H^T Ly \tag{2-20}$$

ตามทฤษฎีค่าถดถอยเชิงสเปกตรัมพบว่าค่า y ที่ดีที่สุด (optimal) หาได้โดยวิธีค่าน้อยที่สุด (minimize) แสดงดังสมการ

$$(\sum_{i,j}(y_i - y_j)\,2W_{ij} = 2y^T Ly \tag{2-21}$$

โดยที่ $L = D - W$ คือกราฟลาปลาเซียน, D คือเมตริกซ์ทแยงมุมที่มีองค์ประกอบคือ $D_{ii} = \sum_j W_{ji}$ และ W คือเมตริกซ์สมมาตรขนาด $N \times N$ ซึ่งคือความเหมือนระหว่างข้อมูล 2 จุด ดังนั้นสมการ (2-20) สามารถหาค่าที่ดีที่สุดโดยการหาค่าลักษณะเฉพาะที่มากที่สุดแสดงดังสมการ 2-22

$$Wy = \lambda Dy \tag{2-22}$$

นอกจากนี้สำหรับอัลกอริทึมค่าถดถอยเชิงสเปกตรัม (spectral regression) สามารถนำคลาสมาพิจารณาด้วย ยกตัวอย่างเช่นข้อมูลมีจำนวนคลาสเท่ากับ c หลังจากผ่านกระบวนการ SRELM แทนด้วยเมตริกซ์ u จะประกอบด้วย c-1 ค่า

กล่าวโดยภาพรวมการแก้สมการ (2-20) ทำได้ใน 2 ขั้นตอน ขั้นตอนที่หนึ่งแก้ปัญหาค่าลักษณะเฉพาะ (eigen value) ในสมการ (2-20) ขั้นตอนที่สองหาค่า $\boldsymbol{u}$ ด้วยเทคนิต SR แสดงดังสมการ

$$u = \underset{u}{\arg\min}\left(\sum_{i=1}^{N}(u^T h(x_i) - y_i)^2 + \alpha\sum_{j=1}^{L} u_j\right) \quad (2\text{-}23)$$

โดยที่ α คือพารามิเตอร์การถดถอยและ u_j คือส่วนประกอบของ u ซึ่งเป็นองค์ประกอบของ β แสดงดังสมการ

$$\beta = [u_1, u_2, \dots, u_{c-1}] \in R^L \quad (2\text{-}24)$$

2.6 การจำแนกประเภท

การออกแบบหรือการเลือกตัวจำแนกประเภทสำหรับรู้จำเป็นกระบวนการที่มีความสำคัญซึ่งควรเลือกใช้ให้เหมาะกับลักษณะของข้อมูล วรรณกรรม [42] นำเสนอตัวจำแนกประเภทที่ใช้ในการรู้จำสามารถแบ่งออกเป็น 3 วิธี คือวิธีเปรียบเทียบความเหมือน วิธีทางสถิติ และวิธีสร้างขอบเขตการตัดสินใจ (decision boundaries) ในส่วนของวิธีเปรียบเทียบความเหมือนคือการจับคู่กับแม่แบบหรือการหาระยะห่างน้อยที่สุดเช่น วิธีการแบ่งเวกเตอร์ (vector quantization) และกฎเพื่อนบ้านที่อยู่ใกล้กันมากที่สุด (k-nearest neighbor rule) ในกรณีของวิธีทางสถิติประกอบด้วยกฎการตัดสินใจของเบย์และกฎการตัดสินใจด้วยวิธีความเป็นไปได้สูงสุด กฎการตัดสินใจของเบย์เป็นการทำให้ความน่าจะเป็นหลัง (posterior probability) ที่มีค่ามากที่สุดซึ่งคำนวณได้จากความน่าจะเป็นก่อน (prior probability) และฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบมีเงื่อนไข (conditional probability density function) ของคลาส ขณะที่กฎการตัดสินใจด้วยวิธีความเป็นไปได้สูงสุดเป็นการประมาณค่าพารามิเตอร์ที่ไม่ทราบค่าจากพารามิเตอร์ที่มีอยู่เช่น ค่าเฉลี่ยและค่าความแปรปรวนเป็นต้น สำหรับวิธีสร้างขอบเขตการตัดสินใจเป็นตัวจำแนกประเภทแบบเชิงเส้นและไม่เป็นเชิงเส้นเช่น โครงข่ายประสาท (neural network) ซัพพอร์ตเวกเตอร์แมชชีน (support vector machine) และ การวิเคราะห์การจำแนกประเภทเชิงเส้น (linear discriminant analysis) เป็นต้น นอกจากนี้ตัวจำแนกประเภทชนิดโครงข่ายประสาทเทียมยังถูกพัฒนาเพื่อลดเวลาในการเรียนรู้หรือทำให้ลู่เข้าเร็วขึ้น โดยการนำเทคนิคต่างๆมาผนวกกับวิธีโครงข่ายประสาทเทียมแบบดั้งเดิมเช่น log-linearized Gaussian mixer network (LLGMN) ร่วมกับ probabilistic neural network (PNN), fuzzy mean max neural network (FMMNN) และ radial basis function artificial neural network (RBFNN) เป็นต้น ตัวชี้วัดความผิดพลาดในการเรียนรู้ถูกคำนวณในรูปแบบของความผิดพลาดของค่าเฉลี่ยกำลังสอง (mean squared error : MSE)

จากการทบทวนวรรณกรรมที่เกี่ยวกับระบบรู้จำคำพูดพบว่าตัวจำแนกที่ได้รับความนิยมสูงสุดคือแบบจำลองมาร์คอฟซ่อนเร้น (hidden markov model) [8][9][10][12][18][19] ซึ่งเป็นตัวจำแนกประเภทที่ใช้ในการรู้จำด้วยวิธีทางสถิติ เนื่องจากรูปแบบของสัญญาณการพูดเป็นลำดับของเวกเตอร์แบบสเปกตรัมที่แปรตามเวลา (time-varying spectral vector sequence) พารามิเตอร์ของแบบจำลองมาร์คอฟซ่อนเร้นได้แก่ จำนวนสถานะของแบบจำลอง จำนวนเอาต์พุตของแบบจำลอง และความน่าจะเป็นในการเปลี่ยนสถานะ ดังนั้นการกำหนดค่าพารามิเตอร์เหล่านี้จะส่งผลต่อความถูกต้องของการจำแนกคำพูด นอกจากนี้ยังมีตัวจำแนกประเภทอีกชนิดหนึ่งที่ถูกนำมาใช้ในระบบรู้จำคำพูดคือโครงข่ายประสาทเทียม [9][22] โดยนิยมใช้ในการจำแนกคำเดี่ยว ดังนั้นในงานวิจัยนี้เน้นการนำโครงข่ายประสาทเทียมแบบไปข้างหน้ามาใช้งาน

2.6.1 สถาปัตยกรรมโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนแบบหลายชั้น

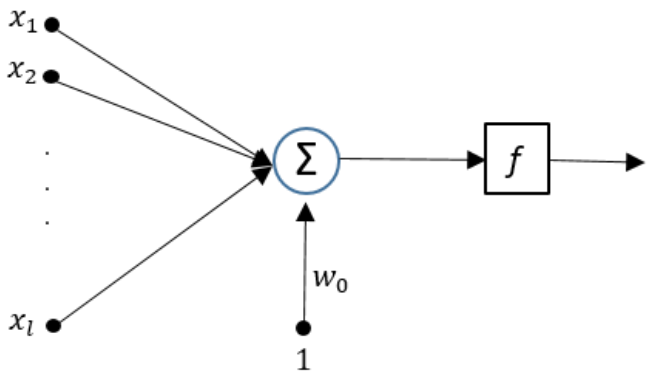
ลักษณะต้นแบบของโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนแบบหลายชั้นประกอบด้วยหนึ่งชั้นอินพุต หนึ่งชั้นเอาต์พุต และอย่างน้อยหนึ่งชั้นซ่อน ขั้นตอนวิธีเพอร์เซ็ปตรอนจะทำการคำนวณเพื่อหาเวกเตอร์ถ่วงน้ำหนัก (w) ของสมการจำแนกประเภทเชิงเส้น โดยเริ่มจากการกำหนดค่าเริ่มต้นให้กับเวกเตอร์ถ่วงน้ำหนัก จากนั้นใช้ค่าเวกเตอร์ถ่วงน้ำหนักดังกล่าวเพื่อคำนวณหาค่าเวกเตอร์ถ่วงน้ำหนักในรอบถัดไป สิ่งที่กล่าวมาจะถูกทำซ้ำเพื่อปรับปรุงเวกเตอร์ถ่วงน้ำหนักในรอบถัดไป จนกระทั่งค่าเวกเตอร์ถ่วงน้ำหนักที่ได้สามารถจำแนกประเภทของเวกเตอร์ฝึกฝนทุกตัวได้อย่างถูกต้องทั้งหมดโดยสมการเชิงเส้น การปรับปรุงค่าเวกเตอร์ถ่วงน้ำหนักดังที่ได้กล่าวไปแล้วข้างต้นสามารถแสดงดังสมการ (2-23)

$$w(t+1) = w(t) - \rho_t \sum_{x \in Y} \delta_x x \tag{2-23}$$

โดยที่ w คือเวกเตอร์ถ่วงน้ำหนักที่ทำการเพิ่ม w_0 เข้าเป็นสมาชิกตัวสุดท้ายของเวกเตอร์ถ่วงน้ำหนักแล้ว Y คือเวกเตอร์ที่รวมเอาเวกเตอร์ฝึกฝน x ทุกตัวที่ถูกจำแนกผิดประเภทเมื่อใช้เวกเตอร์ถ่วงน้ำหนัก $w(t)$ ค่าล่าสุด ρ_t คือพารามิเตอร์ซึ่งกำหนดค่าขึ้นเพื่อใช้ควบคุมความเร็วในการลู่เข้าสู่ผลเฉลยโดยสามารถกำหนดให้เป็นค่าคงที่ และ δ_x มีค่าเป็น -1 เมื่อ $x \in w_1$ และมีค่าเป็น +1 เมื่อ $x \in w_2$ โดยที่ขั้นตอนวิธีเพอร์เซ็ปตรอนจะลู่เข้าสู่ผลเฉลยเมื่อ Y เป็นเวกเตอร์ว่าง (ไม่มีเวกเตอร์ฝึกฝน x เหลืออยู่) [43] เมื่อได้เวกเตอร์ถ่วงน้ำหนัก w แล้ว เวกเตอร์ถ่วงน้ำหนักดังกล่าวจะถูกนำมาใช้เพื่อสร้างตัวประเภทเชิงเส้น จากนั้นเวกเตอร์ x ที่ใช้ในการทดสอบตัวจำแนกประเภทใดๆ จะถูกจำแนกประเภทให้เป็นประเภทใดประเภทหนึ่งจากประเภทที่เป็นไปได้ทั้งหมด โดยการพิจารณาจากค่าของฟังก์ชันแสดงดังสมการ (2-24)

$$f(w^T x) = f(w_1 x_1 + w_2 x_2 + \cdots + w_l x_l + w_0) \quad (2\text{-}24)$$

แผนภาพของแบบจำลองโครงข่ายเพอร์เซ็ปตรอนแสดงดังภาพประกอบที่ 2-6



ภาพประกอบที่ 2-6 แผนภาพของแบบจำลองโครงข่ายเพอร์เซ็ปตรอน

2.7 การรวมข้อมูล (Fusion)

การรวมข้อมูลสามารถแบ่งเป็น 2 หัวข้อใหญ่คือการรวมข้อมูลจากแหล่งที่มาเดียวกัน (unimodal fusion) และการรวมข้อมูลจากหลายแหล่งที่มา (multimodal fusion)

2.7.1 การรวมข้อมูลจากแหล่งที่มาเดียวกัน

ในที่นี้แหล่งที่มาเดียวกันหมายถึงสัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้าและบางส่วนของลำคอจำนวน 5 ช่องสัญญาณ จากการทบทวนวรรณกรรมด้านการรู้จำคำพูดบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ [12][26][27] พบว่าวรรณกรรมส่วนใหญ่ล้วนนำลักษณะเด่นของแต่ละช่องสัญญาณมาต่อเรียงกัน นั่นหมายความว่าเป็นการรวมข้อมูลในระดับลักษณะเด่น วรรณกรรม [27] รายงานว่าสัญญาณไฟฟ้ากล้ามเนื้อของกล้ามเนื้อแต่ละมัดมีลักษณะพึ่งพากัน ดังนั้นการรวมลักษณะเด่นจากหลายช่องสัญญาณสามารถเพิ่มประสิทธิภาพในการรู้จำเสียงได้ นอกจากนี้บทความ [26] รายงานว่าการเพิ่มจำนวนช่องสัญญาณไม่เพียงแต่ประสิทธิภาพของระบบเพิ่มขึ้น ค่าเบี่ยงเบนมาตรฐานยังลดลงอีกด้วย

2.7.2 การรวมข้อมูลจากหลายแหล่งที่มา

การรวมข้อมูลจากหลายแหล่งที่มานิยมแพร่หลายในงานที่เกี่ยวข้องกับการระบุตัวบุคคลโดยการตรวจวัดคุณลักษณะทางกายภาพ (physical characteristics) เช่น ลายนิ้วมือ (fingerprint) ลักษณะใบหน้า (facial recognition) ลักษณะของมือ (hand geometry) และ เรตินา (retina) ภายในดวงตา เป็นต้น ที่เป็นลักษณะเฉพาะของแต่ละคนมาใช้ในการระบุตัวบุคคลนั้นๆ แล้วนำสิ่ง

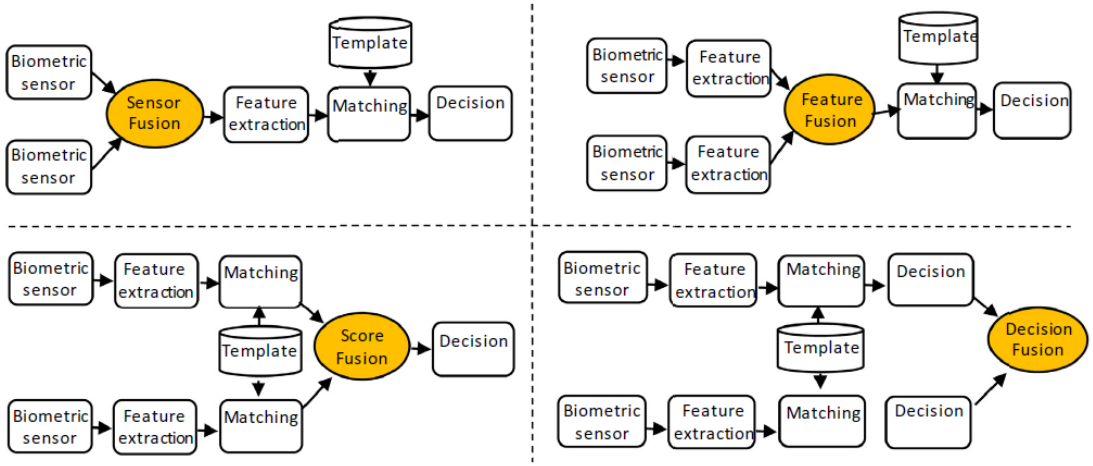
เหล่านั้นมาเปรียบเทียบกับคุณลักษณะที่ได้มีการบันทึกไว้ในฐานข้อมูลก่อนหน้านี้เพื่อใช้แยกแยะบุคคลนั้นจากบุคคลอื่นๆ ข้อดีของการรวมข้อมูลคือสามารถเพิ่มประสิทธิภาพของระบบเมื่อเทียบกับการใช้เพียงข้อมูลเดียว [28] จากการทบทวนวรรณกรรมพบว่ามีการนำเสนอเทคนิคการนำคุณลักษณะทางกายภาพมารวมกันอย่างหลากหลายเรียกว่า biometrics fusion ซึ่งสามารถแบ่งออกเป็น 4 รูปแบบคือ การรวมระดับเซนเซอร์ (sensor level) ระดับลักษณะเด่น (feature level) ระดับคะแนน (score level) และระดับการตัดสินใจ (decision level) ดังภาพประกอบที่ 2-7 ซึ่งมีรายละเอียดดังนี้

- ระดับเซนเซอร์เป็นการรวมข้อมูลดิบ (raw data) ที่ได้จากเซนเซอร์ต่างกันหรือจากระบบเก็บข้อมูลที่ต่างกันแต่ใช้เซนเซอร์ตัวเดียวกัน

- ระดับลักษณะเด่นเป็นการรวมลักษณะเด่นและเก็บในฐานข้อมูล การรวมแบบนี้จะมีประสิทธิภาพก็ต่อเมื่อลักษณะเด่นที่รวมกันมีความอิสระต่อกัน และมีมิติของข้อมูลเท่ากันโดยการใช้เทคนิคการลดมิติของข้อมูล

- ระดับคะแนนเป็นการจับคู่ (matching) ข้อมูลกับฐานข้อมูลเพื่อคำนวณหาคะแนน จากนั้นจึงนำคะแนนของเซนเซอร์แต่ละตัวมารวมกัน สุดท้ายนำคะแนนมาใช้ในการตัดสินใจ โดยคะแนนของแต่ละเซนเซอร์จะถูกนอร์ม (normalization) ก่อนแล้วจึงนำมารวมกัน

- ระดับการตัดสินใจเป็นการรวมหลังจากการจำแนกประเภทของแต่ละเซนเซอร์ โดยอาจใช้วิธีการถือเสียงข้างมากเป็นเกณฑ์ (majority vote) วิธีนี้มีความซับซ้อนน้อยและการทำงานร่วมกันระหว่างไบโอเมตริกซ์ที่ต่างกันมากที่สุด แต่มีประสิทธิภาพน้อยกว่าการรวมในระดับคะแนน เนื่องจากข้อจำกัดด้านจำนวนของข้อมูลที่มี



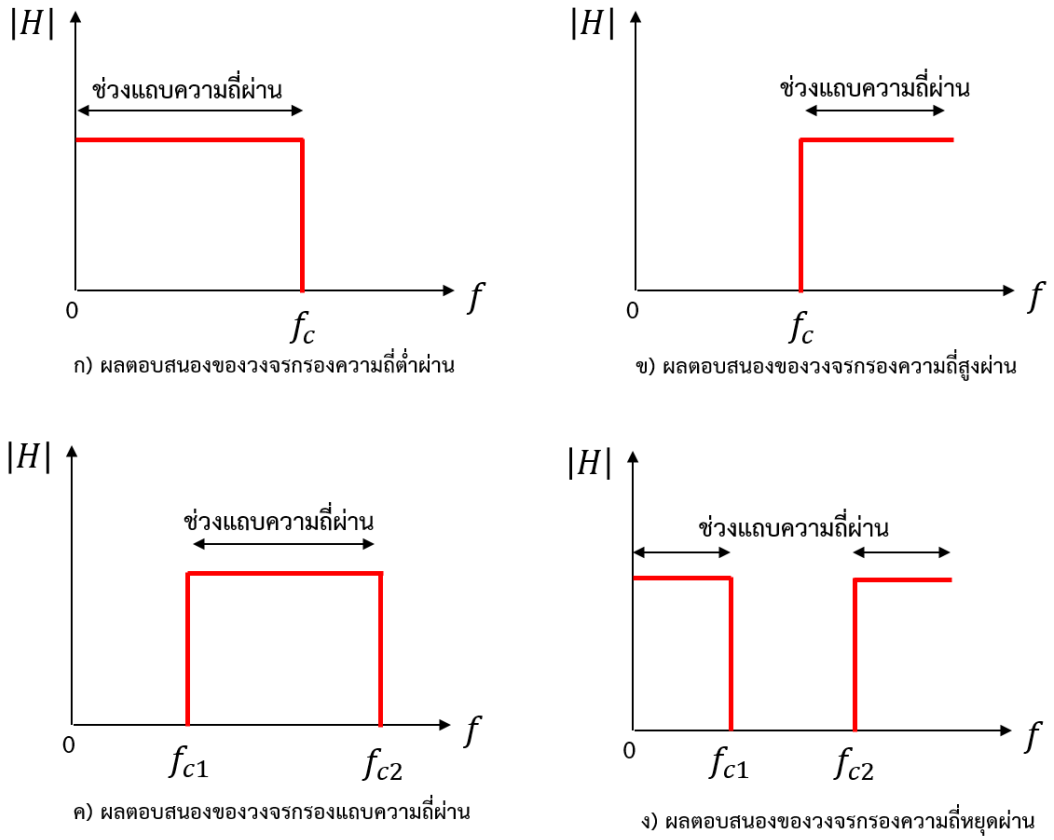
ภาพประกอบที่ 2-7 การรวมสัญญาณในระดับต่างๆ ของระบบไบโอเมตตริกซ์ [28]

2.8 การออกแบบวงจรกรองความถี่แบบดิจิตอลที่มีผลตอบสนองอิมพัลส์จำนวนจำกัด (Finite impulse response: FIR)

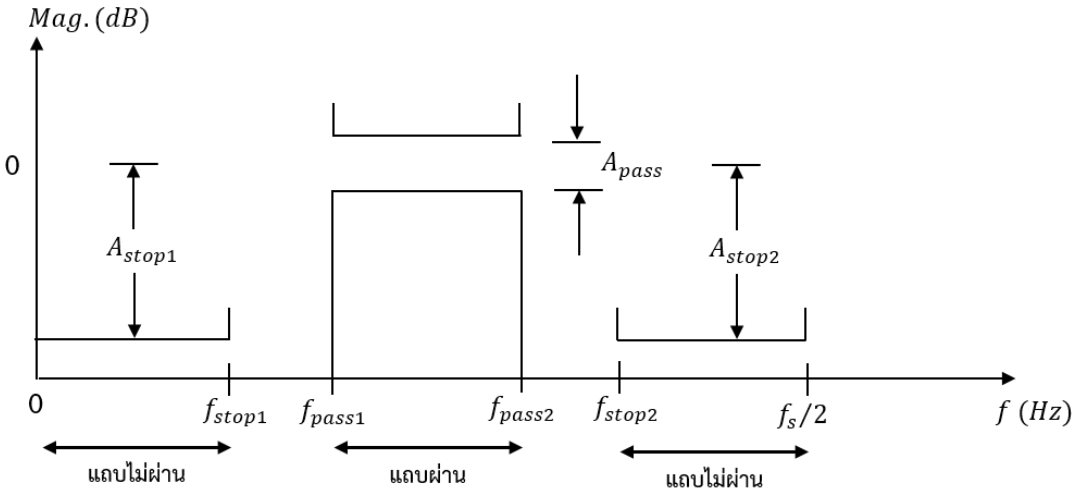
วงจรกรองความถี่สามารถแบ่งคุณลักษณะของผลตอบสนองความถี่ (frequency response) ดังภาพประกอบที่ 2.8 ได้ 4 ชนิดคือ วงจรกรองความถี่ต่ำผ่าน (low pass filter: LPF) วงจรกรองความถี่สูงผ่าน (high pass filter: HPF) วงจรกรองแถบความถี่ผ่าน (band pass filter: BPF) และวงจรกรองแถบความถี่หยุดผ่าน (band stop filter หรือ notch filter) ผลตอบสนองความถี่ของวงจรกรองแต่ละชนิดแสดงดังภาพประกอบที่ 2-8 (ก) ถึง (ง) ตามลำดับ โดยทั่วไปวงจรกรองความถี่แบบดิจิตอลสามารถแบ่งได้เป็นผลตอบสนองอิมพัลส์จำนวนจำกัด (FIR) กับผลตอบสนองอิมพัลส์แบบไม่จำกัด (infinite impulse response: IIR) ข้อดีของ FIR คือมีผลตอบสนองความถี่แบบเชิงเส้น (linear phase) เหมาะสำหรับการพัฒนาเครื่องมือวัดทางการแพทย์และในงานที่เกี่ยวข้อง

พารามิเตอร์ที่สำคัญที่ใช้ในการออกแบบวงจรกรองความถี่แบบดิจิตอล โดยยกตัวอย่างวงจรกรองแถบความถี่ผ่าน FIR ชนิด equiripple ดังภาพประกอบที่ 2.9 มีรายละเอียดดังนี้

- ความถี่ผ่าน (f_{pass}) หมายถึงจุดตัดของความถี่ที่ให้สัญญาณผ่าน
- ความถี่ไม่ผ่าน (f_{stop}) หมายถึงจุดตัดของความถี่ที่ให้ไม่ผ่านความถี่
- การลดทอนของแถบหยุด (stop-band attenuation : A_{stop}) คือจำนวนเท่าที่แถบหยุดลดทอนลง วัดค่าเป็น dB
- การลดทอนของแถบผ่าน (pass-band attenuation : A_{pass}) คือค่าสูงสุดที่ขนาดแถบผ่านแกว่งออกจากค่า 0 วัดค่าเป็น dB



ภาพประกอบที่ 2-8 ผลตอบสนองความถี่ของวงจรกรองความถี่ในอุดมคติ



ภาพประกอบที่ 2-9 คุณลักษณะเฉพาะของผลตอบสนองความถี่ของตัวกรองแบบ FIR

บทที่ 3

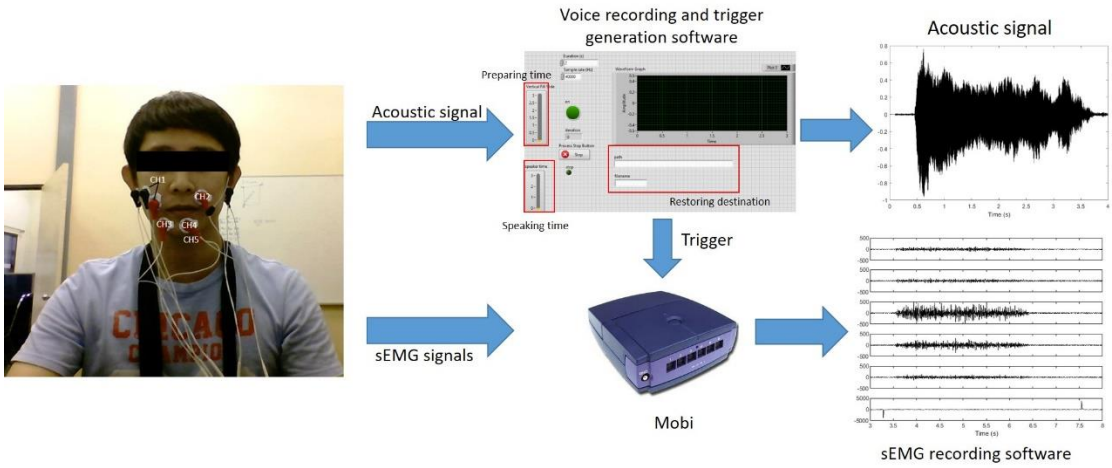
วัสดุ อุปกรณ์และระเบียบวิธีการวิจัย

บทนี้จะกล่าวถึงการออกแบบระบบการได้ข้อมูลเพื่อให้สามารถเก็บสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงพร้อมกันได้ โดยออกแบบทั้งในส่วนของอุปกรณ์และโปรแกรมที่ใช้เก็บข้อมูล จากนั้นทำการหาจุดเริ่มต้นของสัญญาณโดยคำนวณจากค่าเฉลี่ยของค่าสัมบูรณ์และค่าเบี่ยงเบนมาตรฐาน โดยอ้างอิงจากสัญญาณจุดชนวนเป็นตัวกำหนดจุดเริ่มต้นของสัญญาณ ถัดมากล่าวถึงคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อของคนปกติและผู้ที่มีอาการพูดไม่เป็นความโดยแยกตามคุณลักษณะของลักษณะเด่นเพื่อเปรียบเทียบความเหมือนหรือแตกต่างระหว่างคนปกติและคนที่มีอาการพูดไม่เป็นความ นอกจากนี้ยังมีการเปรียบเทียบประสิทธิภาพของการสกัดลักษณะเด่นแบบฉายข้อมูลอีกด้วย สิ่งที่สำคัญอีกอย่างหนึ่งคือศึกษาคุณลักษณะของสัญญาณเสียงเพื่อเลือกลักษณะเด่นที่เหมาะสมรวมถึงจำนวนสัมประสิทธิ์ของ MFCC ที่ส่งผลต่อประสิทธิภาพของระบบรู้จำคำพูดอีกด้วย ส่วนที่สำคัญที่สุดที่ขาดมิได้คือวิธีการรวมข้อมูลทั้งจากแหล่งที่มาเดียวกันและจากหลายแหล่งที่มา เพื่อเปรียบเทียบประสิทธิภาพที่เกิดขึ้น โดยมีการแปรผันทั้งช่วงเวลาของการวิเคราะห์สัญญาณ จำนวนสัมประสิทธิ์ของ MFCC และจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ ผลการทดลองในส่วนนี้จะถูกนำไปประยุกต์ใช้สำหรับการทดสอบการรู้จำคำพูดแบบไม่ขึ้นกับบุคคลต่อไป

3.1 ระบบการได้ข้อมูล

ระบบการได้ข้อมูล (data acquisition system) ประกอบด้วยระบบบันทึกสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อตามภาพประกอบที่ 3-1 ในส่วนของระบบบันทึกสัญญาณเสียงประกอบด้วยไมโครโฟนแบบสวมศีรษะยี่ห้อ SpeechWare FlexyMike Dual Ear Cardioid (DEC) โดยมีคุณลักษณะ (specification) ดังนี้ รูปแบบการรับเสียงเฉพาะด้านหน้า (unidirectional polar pattern) ช่วงความถี่ (frequency range) 50 ถึง 18,000 เฮริตซ์ ความไว (sensitivity) -65dB ± 3dB และคอมพิวเตอร์โน๊ตบุ๊ก หลักการทำงานของระบบบันทึกสัญญาณเสียงคือเมื่ออาสาสมัครออกเสียงผ่านไมโครโฟนซึ่งถูกต่อกับเครื่องคอมพิวเตอร์ ระบบบันทึกเสียงที่เขียนด้วยโปรแกรม LabVIEW จะทำการเก็บสัญญาณเสียงในโฟลเดอร์ที่สร้างไว้ โดยมีความถี่ชักตัวอย่าง (sampling frequency) ที่ 20 กิโลเฮิรซต์ ขณะที่ระบบเริ่มบันทึกสัญญาณเสียง สัญญาณจุดชนวน (trigger signal) จะถูกสร้างขึ้นและส่งไปยังช่องสัญญาณที่ 6 ของเครื่องมือวัดสัญญาณไฟฟ้ากล้ามเนื้อซึ่งเป็นเครื่องมือเชิงพาณิชย์ (commercial) ผ่าน NI USB 6009 14 bit 48 kS/s ซึ่งทำหน้าที่แปลงสัญญาณดิจิตอลเป็นสัญญาณแอนะล็อก จากนั้นสัญญาณจุดชนวนจะถูกบันทึกบนซอฟต์แวร์ของ

เครื่องมือวัดสัญญาณไฟฟ้ากล้ามเนื้อแบบอัตโนมัติ ผลที่ได้รับคือสามารถสังเกตจุดเริ่มต้นของสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อที่ถูกบันทึกทำให้การตัดแยกสัญญาณทำได้ง่าย



ภาพประกอบที่ 3-1 ระบบการได้ข้อมูลสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อ หลังจากที่อาสาสมัครออกเสียง สัญญาณเสียงถูกเก็บโดยระบบบันทึกเสียง ขณะที่สัญญาณไฟฟ้ากล้ามเนื้อถูกวัดและบันทึกโดยเครื่องมือวัดสัญญาณไฟฟ้ากล้ามเนื้อซึ่งเป็นเครื่องมือเชิงพาณิชย์

ในส่วนของการบันทึกสัญญาณไฟฟ้ากล้ามเนื้อถูกแบ่งเป็นสองส่วนคือสัญญาณของคนปกติและสัญญาณของผู้ที่มีอาการพูดไม่เป็นความ โดยมีตำแหน่งมัดกล้ามเนื้อที่เหมือนกัน นอกจากนี้เครื่องมือและวัสดุที่ใช้ในการวัดและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อที่เหมือนกันได้แก่ อิเล็กโทรดซิลเวอร์-ซิลเวอร์ คลอไรด์ (Ag/AgCl) ที่มีคุณลักษณะของอิเล็กโทรดเป็นแบบอิเล็กโทรดรูปจาน (disc-shaped) ขนาดเส้นผ่านศูนย์กลาง 5 มิลลิเมตร และเครื่องวัดและบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ วิธีการวัดสัญญาณไฟฟ้ากล้ามเนื้อเริ่มจากการติดอิเล็กโทรดบนใบหน้าทั้งหมด 5 จุดตามภาพประกอบที่ 2-1 คือ 1) กล้ามเนื้อ Zygomaticus major (CH1) 2) กล้ามเนื้อ Levator anguli oris (CH2) 3) กล้ามเนื้อ Depressor anguli oris (CH3) 4) Mentalis (CH4) และ 5) กล้ามเนื้อ Anterior belly of the digastrics (CH5) โดย CH1 เป็นการบันทึกแบบอิเล็กโทรดสองขั้ว (Bipolar configuration) เนื่องจากมัดกล้ามเนื้อมีขนาดใหญ่ และ CH2 CH3 CH4 และ CH5 เป็นการบันทึกแบบอิเล็กโทรดขั้วเดียว (monopolar configuration) เนื่องจากมัดกล้ามเนื้อมีขนาดเล็ก โดยอิเล็กโทรดอ้างอิง (reference electode) ถูกติดบริเวณใบหูทั้งด้านขวาและซ้ายเพื่อลดทอนสัญญาณรบกวนจากกล้ามเนื้อบริเวณใกล้เคียงและอิเล็กโทรดสายดิน (ground electrode) ถูกติดบริเวณข้อมือซ้าย จากนั้นอิเล็กโทรดทั้งหมดถูกต่อกับสายนำสัญญาณแบบมีฉนวนหุ้มเพื่อเชื่อมต่อกับเครื่องวัดสัญญาณไฟฟ้าแบบพกพาได้รุ่น Mobi6-6b (TMS International B.B., Netherlands) ซึ่งมีความถี่ของการสุ่มสัญญาณสูงสุด 2,048 เฮริตซ์ และ ความละเอียด 24 bits Bipolar 12.2 nV per

bit สุดท้ายสัญญาณไฟฟ้ากล้ามเนื้อจะถูกบันทึกด้วยซอฟต์แวร์สำเร็จรูปที่ติดมากับเครื่องวัดสัญญาณไฟฟ้ากล้ามเนื้อบนเครื่องคอมพิวเตอร์ โดยมีความถี่ชักตัวอย่าง (sampling rate) ที่ 1024 เฮิรซต์ สำหรับคนปกติระบบการได้ข้อมูลจะบันทึกช่วงเวลาของสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อเป็นเวลา 4 วินาทีประกอบด้วยช่วงเวลาเตรียมตัว 1 วินาที และช่วงเวลาออกเสียง 3 วินาที ในส่วนของผู้ที่มีอาการพูดไม่เป็นความระบบการได้ข้อมูลจะบันทึกช่วงเวลาของสัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียวเป็นเวลา 2 วินาที

ในการทดลองมีอาสาสมัครปกติทั้งหมด 7 คนซึ่งไม่มีความบกพร่องหรืออาการผิดปกติด้านการพูดประกอบด้วยผู้ชาย 4 คน ผู้หญิง 3 คน อายุ 20 ถึง 22 ปี ความสูง 160 ถึง 180 เซนติเมตร และน้ำหนัก 46 ถึง 75 กิโลกรัม โดยทำการทดลองที่ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ โดยผู้วิจัยเน้นศึกษาระบบรู้จำคำพูดในคนปกติเพื่อพัฒนาระบบสำหรับผู้ป่วยอาการพูดไม่เป็นความในอนาคต วิธีการคืออาสาสมัครแต่ละคนจะทำการออกเสียงพยางค์ในภาษาไทยทั้ง 12 เสียงดังตารางที่ 2-1 หัวข้อเรื่องวิธีบำบัดการพูด โดยแต่ละพยางค์จะออกเสียงซ้ำทั้งหมด 5 ครั้ง ในกรณีของผู้ที่มีอาการพูดไม่เป็นความ โดยได้รับความอนุเคราะห์ข้อมูลจากสาขาวิชาวิศวกรรมชีวการแพทย์ คณะแพทยศาสตร์ มหาวิทยาลัยสงขลา-นครินทร์ โดยทำการวัดสัญญาณไฟฟ้ากล้ามเนื้อ ณ โรงพยาบาลสงขลานครินทร์ อาสาสมัครมีอาการอ่อนแรงของมัดกล้ามเนื้อบริเวณใบหน้าเป็นส่วนใหญ่ประกอบด้วยผู้ชาย 1 คน ผู้หญิง 4 คน อายุ 39 ถึง 69 ปี ความสูง 150 ถึง 162 เซนติเมตร และน้ำหนัก 46 ถึง 66 กิโลกรัม อย่างไรก็ตามข้อมูลที่ได้รับมีเพียง 9 พยางค์ ไม่นับรวมสระทั้ง 3 เสียงคือ อา อี และอู [31]

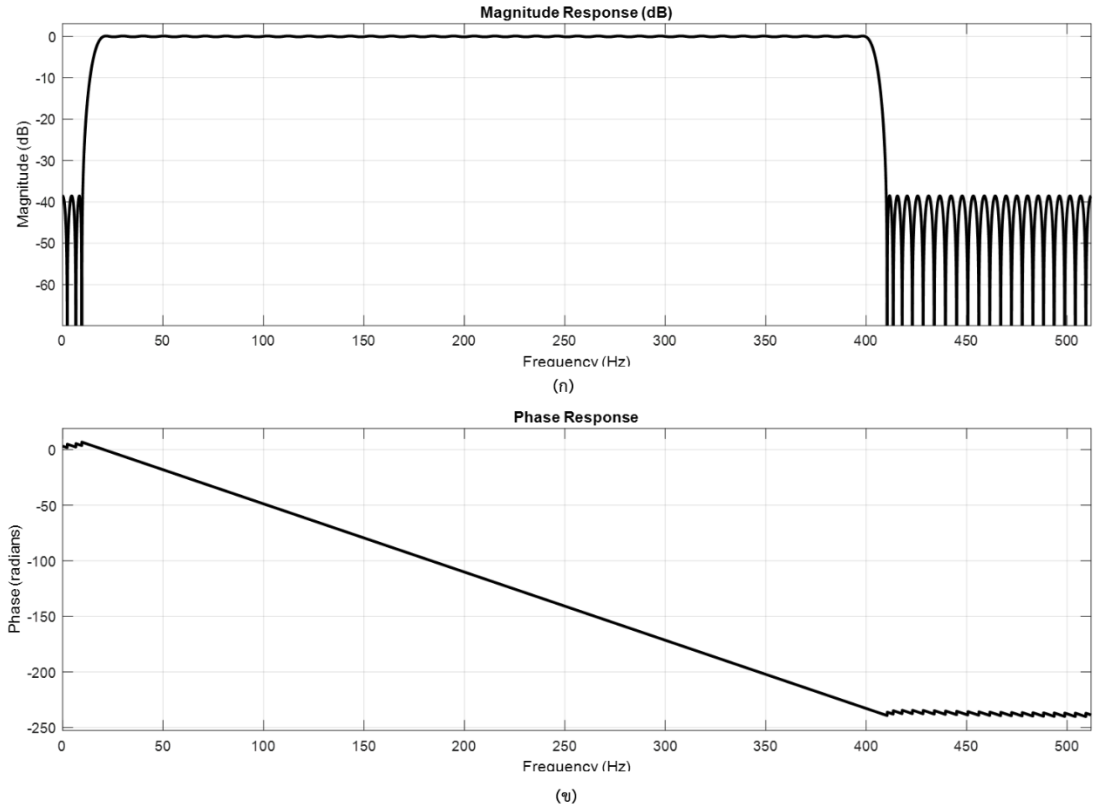
3.2 การออกแบบวงจรกรองความถี่ของสัญญาณ

จากหัวข้อการออกแบบวงจรกรองแบบดิจิตอลระบบ FIR นั้น ในงานวิจัยนี้ได้ทำการออกแบบวงจรกรอง 3 ชนิดคือวงจรกรองความถี่แถบผ่าน และนอตช์ฟิลเตอร์สำหรับสัญญาณไฟฟ้ากล้ามเนื้อ วงจรกรองความถี่ต่ำผ่านสำหรับสัญญาณเสียงโดยมีพารามิเตอร์ในการออกแบบดังนี้

3.2.1 วงจรกรองความถี่แถบผ่าน

จากภาพประกอบที่ 2-9 ได้กำหนดพารามิเตอร์ต่างๆ ซึ่งสามารถแบ่งออกเป็น 2 หัวข้อคือคุณลักษณะของความถี่ (frequency specifications) และคุณลักษณะของความขนาด (magnitude specifications) โดยให้แถบความถี่ผ่านในช่วง 20 ถึง 400 เฮริตซ์ ดังนี้ อัตราการสุ่มสัญญาณ (f_s) เท่ากับ 1,024 เฮริตซ์ ความถี่ผ่าน (f_{pass1}, f_{pass2}) เท่ากับ 20 และ 400 เฮริตซ์ ความถี่ไม่ผ่าน (f_{stop1}, f_{stop2}) เท่ากับ 15 และ 405 เฮริตซ์ ในส่วนของคุณลักษณะของขนาดมีหน่วยเป็น dB ประกอบด้วย การลดทอนของแถบผ่าน (A_{pass}) เท่ากับ 1 และ การลดทอนของแถบ

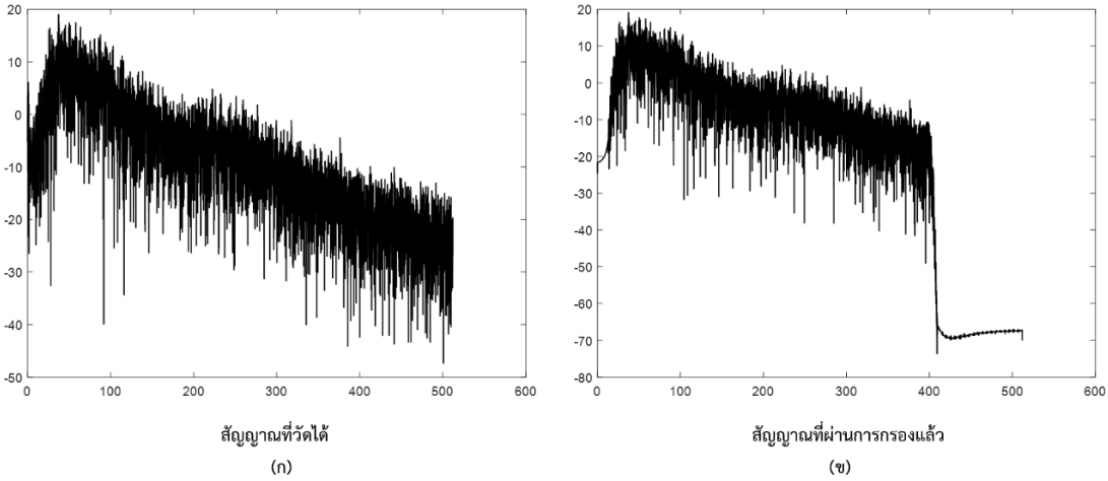
หยุด (A_{stop1}, A_{stop2}) เท่ากับ 25 ผลที่ได้แสดงดังภาพประกอบที่ 3-2 โดยวงจรกรองที่ได้มีอันดับ (order) เท่ากับ 200 จากนั้นนำวงจรกรองที่ออกแบบไปใช้กับสัญญาณไฟฟ้ากล้ามเนื้อแสดงดังภาพประกอบที่ 3-3



ภาพประกอบที่ 3-2 ลักษณะของวงจรกรองแถบความถี่ผ่านที่อันดับ 200 (ก) ผลตอบสนองความถี่ (ข) ผลตอบสนองทางเฟส

3.2.2 วงจรกรองแบบนอตซ์ฟิลเตอร์ (Notch filter)

การออกแบบวงจรกรองแบบนอตซ์ฟิลเตอร์คล้ายคลึงกับการออกแบบวงจรกรองแถบความถี่ผ่าน กำหนดให้แถบความถี่หยุดที่ 50 เฮริตซ์ มีพารามิเตอร์ดังนี้ อัตราการสุ่มสัญญาณ (f_s) เท่ากับ 1,024 เฮริตซ์ ความถี่ผ่าน (f_{pass1}, f_{pass2}) เท่ากับ 50 และ 55 เฮริตซ์ ความถี่ไม่ผ่าน (f_{stop1}, f_{stop2}) เท่ากับ 45 และ 60 เฮริตซ์ ในส่วนของคุณลักษณะของขนาดมีหน่วยเป็น dB ประกอบด้วย การลดทอนของแถบผ่าน (A_{pass1}, A_{pass2}) เท่ากับ 5 และ 1 และ การลดทอนของแถบหยุด (A_{stop}) เท่ากับ 25 ผลที่ได้แสดงดังภาพประกอบที่ 3-4 โดยวงจรกรองที่ได้มีอันดับ (order) เท่ากับ 336 จากนั้นนำวงจรกรองที่ออกแบบไปใช้กับสัญญาณไฟฟ้ากล้ามเนื้อแสดงดังภาพประกอบที่ 3-5



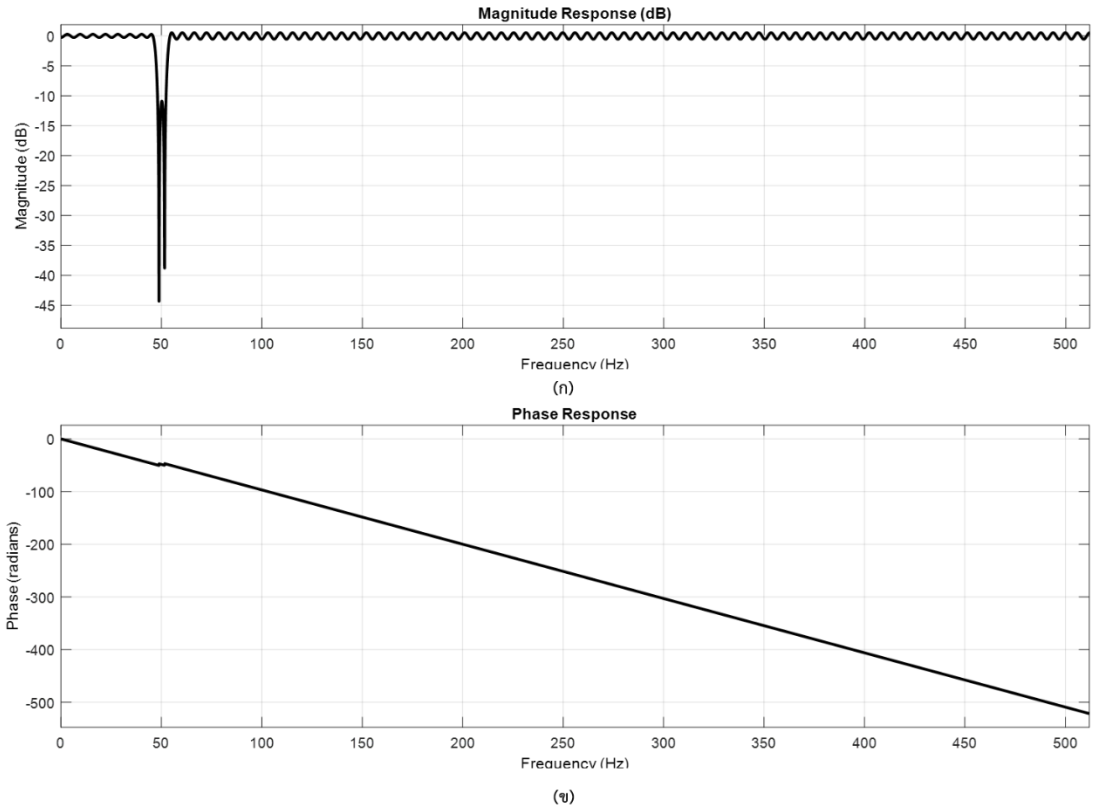
ภาพประกอบที่ 3-3 สัญญาณไฟฟ้ากล้ามเนื้อของช่องสัญญาณที่ 3 (ก) สัญญาณที่วัดได้ (ข) สัญญาณที่ผ่านการกรองด้วยวงจรกรองความถี่แถบผ่าน

3.2.3 วงจรกรองแบบความถี่ต่ำผ่าน

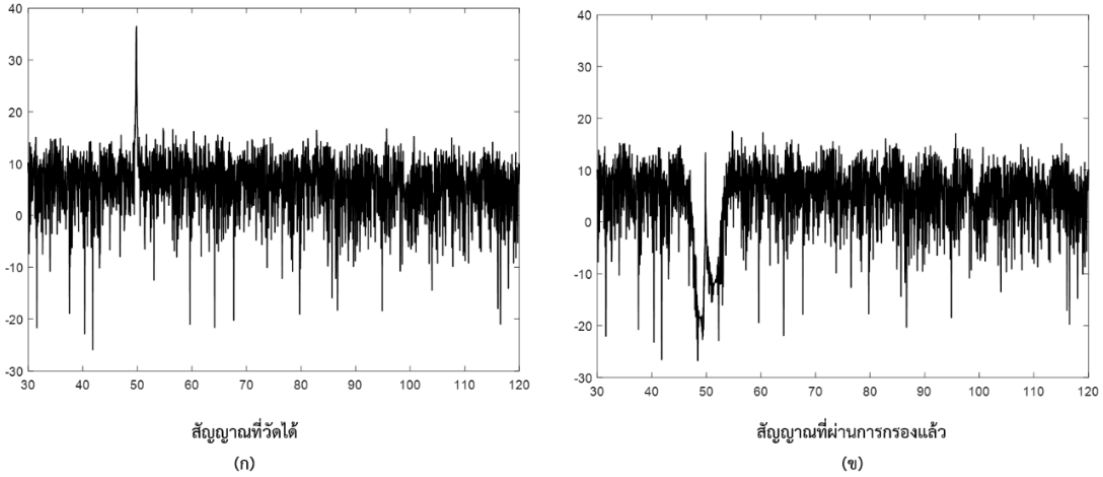
การออกแบบวงจรกรองแบบความถี่ต่ำผ่านมีพารามิเตอร์คล้ายคลึงกับการออกแบบวงจรกรองที่กล่าวไปแล้วข้างต้น กำหนดให้แถบความถี่หยุดที่ 5 กิโลเฮริตซ์ มีพารามิเตอร์ดังนี้ ดังนี้ อัตราการสุ่มสัญญาณ (f_s) เท่ากับ 20,000 เฮริตซ์ ความถี่ผ่าน (f_{pass1}) เท่ากับ 5 กิโลเฮริตซ์ ความถี่ไม่ผ่าน (f_{stop}) เท่ากับ 5.1 กิโลเฮริตซ์ ในส่วนของคุณลักษณะของขนาดมีหน่วยเป็น dB ประกอบด้วย การลดทอนของแถบผ่าน (A_{pass}) เท่ากับ 1 และ การลดทอนของแถบหยุด (A_{stop}) เท่ากับ 25 ผลที่ได้แสดงดังภาพประกอบที่ 3-6 โดยวงจรกรองที่ได้มีอันดับ (order) เท่ากับ 200 จากนั้นนำวงจรกรองที่ออกแบบไปใช้กับสัญญาณไฟฟ้ากล้ามเนื้อแสดงดังภาพประกอบที่ 3-7

3.3 การหาจุดเริ่มต้นของสัญญาณ

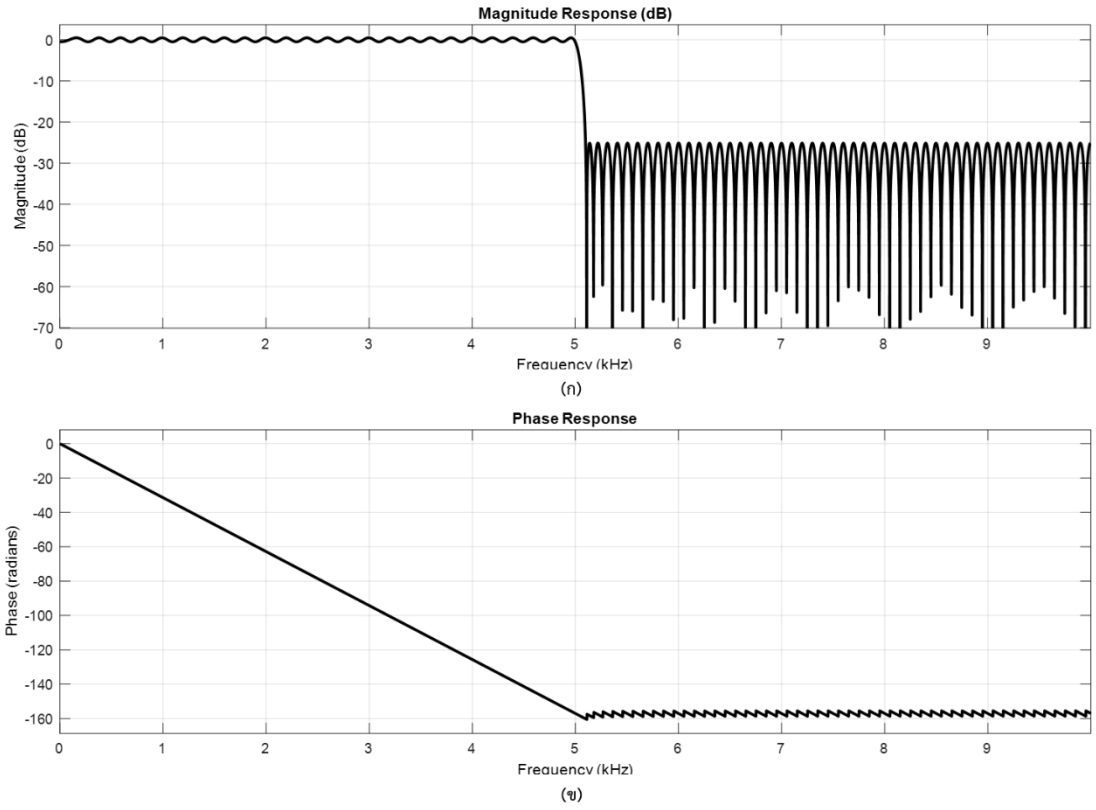
การหาจุดเริ่มต้นของสัญญาณสามารถแบ่งเป็นการหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อ และสัญญาณเสียง ในส่วนของการหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อเป็นแบบกึ่งอัตโนมัติ ก่อนที่จะหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อต้องทำการกำจัดสัญญาณรบกวนเสียก่อนโดยใช้วงจรกรองสัญญาณความถี่แถบผ่านที่มีความถี่ตัดระหว่าง 20 ถึง 450 เฮิรซต์ กระบวนการหาจุดเริ่มต้นของสัญญาณประกอบด้วย 5 ขั้นตอน มีรายละเอียดดังนี้



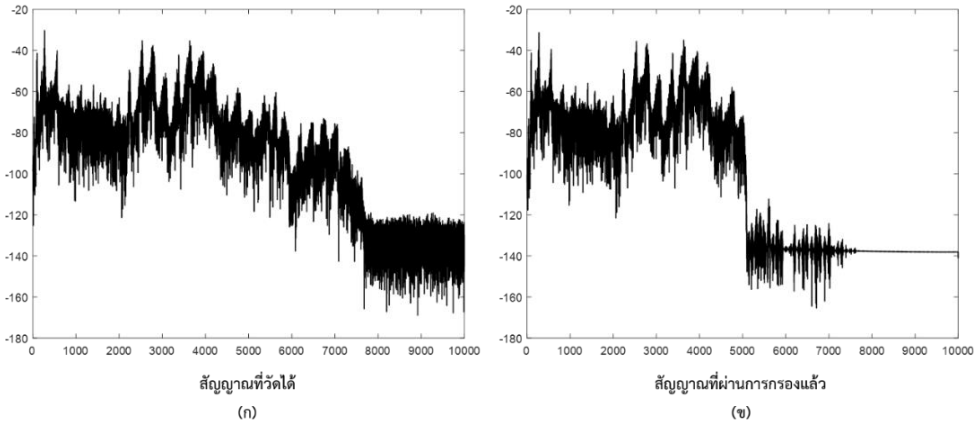
ภาพประกอบที่ 3-4 ลักษณะของวงจรกรองแบบนอตซ์ฟิลเตอร์ที่อันดับ 336 (ก) ผลตอบสนองความถี่ (ข) ผลตอบสนองทางเฟส



ภาพประกอบที่ 3-5 สัญญาณไฟฟ้ากล้ามเนื้อของช่องสัญญาณที่ 3 (ก) สัญญาณที่วัดได้ (ข) สัญญาณที่ผ่านการกรองด้วยวงจรกรองความถี่แถบผ่าน

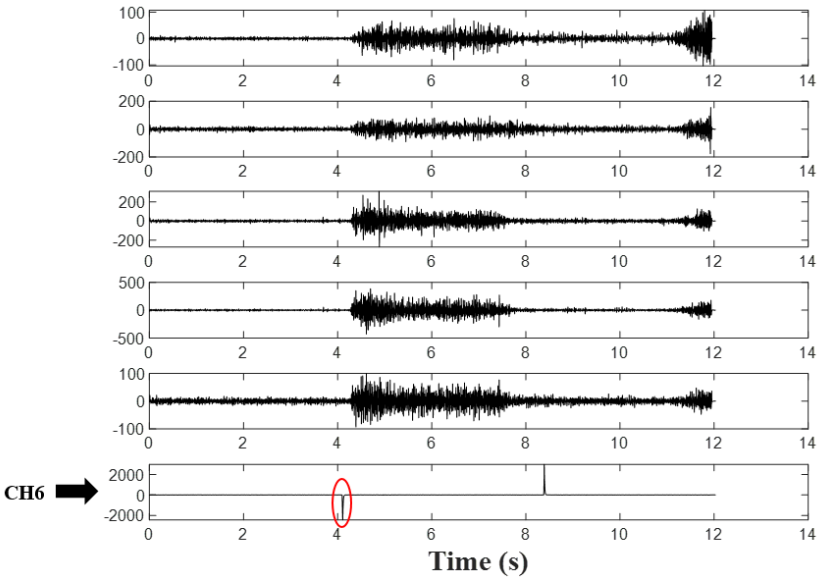


ภาพประกอบที่ 3-6 ลักษณะของวงจรกรองแถบความถี่ต่ำผ่านที่อันดับ 200 (ก) ผลตอบสนองความถี่ (ข) ผลตอบสนองทางเฟส



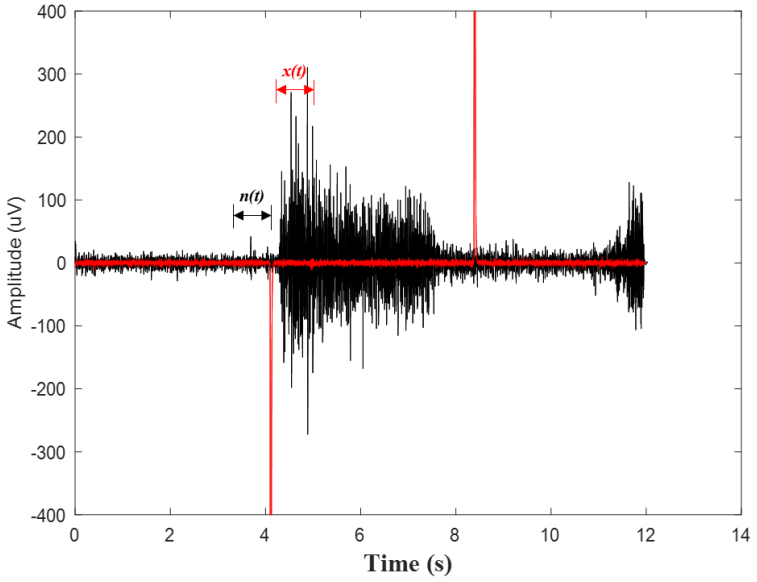
ภาพประกอบที่ 3-7 สัญญาณไฟฟ้าเสียง (ก) สัญญาณที่วัดได้ (ข) สัญญาณที่ผ่านการกรองด้วยวงจรกรองความถี่แถบผ่าน

ขั้นตอนที่ 1 หาตำแหน่งของสัญญาณจุดชนวนในช่องสัญญาณที่ 6 ดังภาพประกอบที่ 3-8



ภาพประกอบที่ 3-8 วงกลมสีแดงคือตำแหน่งของสัญญาณจุดชนวนในช่องสัญญาณที่ 6

ขั้นตอนที่ 2 กำหนดให้ตำแหน่งของสัญญาณรบกวน $(n(t))$ นำหน้าสัญญาณจุดชนวนเป็นระยะทาง 500 มิลลิวินาที ดังภาพประกอบที่ 3-9 และคำนวณขนาดของสัญญาณรบกวนตามหลักสถิติดังสมการ $A = \mu_n + 2\sigma_n$ โดยที่ μ_n และ σ_n คือค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าสัมบูรณ์ของสัญญาณรบกวน $|n(t)|$ ตามลำดับ เหตุผลที่เลือกใช้ค่า $2\sigma_n$ ได้มาจากการทดลอง ซึ่งเป็นค่าขีดจำกัด (threshold) ของพลังงานของสัญญาณรบกวนที่เหมาะสม และไม่สูงกว่าพลังงานของสัญญาณไฟฟ้ากล้ามเนื้อ



ภาพประกอบที่ 3-9 ตำแหน่งของสัญญาณรบกวน $(n(t))$ และสัญญาณหลังสัญญาณจุดชนวน $(x(t))$ โดยสีแดงคือสัญญาณจุดชนวนและสีดำคือสัญญาณของพยางค์ “อี” ของช่องสัญญาณที่สาม

ขั้นตอนที่ 3 กำหนดให้สัญญาณหลังสัญญาณจุดชนวน $(x(t))$ หรือเรียกว่าเฟรม (frame) ดังภาพประกอบที่ 3-9 มีขนาด 32 มิลลิวินาที และคำนวณขนาดของสัญญาณตามสมการ $B = \mu_x + 2\sigma_x$ โดยที่ μ_x และ σ_x คือค่าเฉลี่ยและค่าเบียงเบนมาตรฐานของค่าสัมบูรณ์ของสัญญาณ $|x(t)|$ ตามลำดับ

ขั้นตอนที่ 4 เปรียบเทียบค่าของ B และ $8A$ ถ้าค่าของ B มากกว่า $8A$ แสดงว่าจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อคือจุดกึ่งกลางของ $x(t)$ โดยทำการแปรผันจำนวนเท่าของค่า A พบว่า 8 เท่าของค่า A ให้ค่าใกล้เคียงกับจุดเริ่มต้นของสัญญาณมากที่สุด กระบวนการนี้ถูกทำซ้ำหากเงื่อนไขไม่เป็นตามที่กำหนด โดยเลื่อนเฟรมไปทีละ 32 มิลลิวินาที และเปรียบเทียบค่าระหว่าง B และ A จนกระทั่งค่าของ B มากกว่า $8A$ และจุดกึ่งกลางของเฟรมคือจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อ

ขั้นตอนที่ 5 คำนวณจุดสิ้นสุดของสัญญาณไฟฟ้าโดยบวกระยะทางจากจุดเริ่มต้นไป 2.4 วินาที

ในส่วนของการหาจุดเริ่มต้นของสัญญาณเสียง ก่อนที่จะหาจุดเริ่มต้นของสัญญาณเสียงต้องทำการกำจัดสัญญาณรบกวนเสียก่อนโดยใช้วงจรกรองความถี่ต่ำผ่านที่มีความถี่ตัด 5 กิโลเฮิรซต์ กระบวนการหาจุดเริ่มต้นของสัญญาณเสียงจะอ้างอิงจากตำแหน่งเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อเช่นเดียวกับจุดสิ้นสุดของสัญญาณเสียง

3.4 คุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อของคนปกติและผู้ที่มีอาการพูดไม่เป็นความ

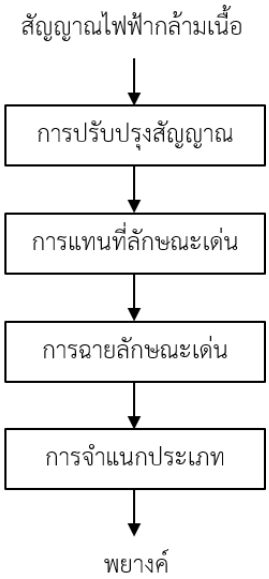
ในหัวข้อนี้เป็นการจำแนกพยางค์ไทยจำนวน 9 เสียง คือ คา คี คู นา นี นู มา มี และ มู โดยเปรียบเทียบสัญญาณไฟฟ้ากล้ามเนื้อระหว่างคนปกติและผู้ที่มีอาการพูดไม่เป็นความ เนื่องจากข้อจำกัดด้านข้อมูลของผู้ที่มีอาการพูดไม่เป็นความที่ได้รับความอนุเคราะห์จากสาขาวิชาวิศวกรรมชีวการแพทย์ คณะแพทย์ศาสตร์ มหาวิทยาลัยสงขลานครินทร์ การทดลองถูกแบ่งเป็นสองหัวข้อย่อยคือการอธิบายคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อด้วยลักษณะเด่น 3 กลุ่ม และเปรียบเทียบประสิทธิภาพของการลดมิติข้อมูลแบบฉายข้อมูลจำนวน 3 ประเภท ดังรายละเอียดต่อไปนี้

3.4.1 คุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อแยกตามประเภทของลักษณะเด่น

เนื่องจากผู้วิจัยต้องการศึกษาคุณลักษณะของลักษณะเด่นที่ส่งผลต่อการจำแนกพยางค์ของคนปกติและผู้ป่วยที่มีอาการพูดไม่เป็นความ ดังนั้นผู้วิจัยจึงแบ่งลักษณะเด่นออกเป็น 3 กลุ่มคือขนาดของสัญญาณ (ABF) ความถี่ของสัญญาณ (FBF) และค่าสถิติของการกระจายข้อมูล (SBF) โดยแต่ละกลุ่มประกอบลักษณะเด่น 2 ค่าดังนี้ กลุ่มขนาดของสัญญาณ ประกอบด้วย MAV และ WL กลุ่มความถี่ของสัญญาณประกอบด้วย ZC และ MNF กลุ่มค่าทางสถิติของการกระจายข้อมูลประกอบด้วย $L-KURT$ และ $L-SKW$ กระบวนการจำแนกพยางค์ไทยประกอบด้วย 4

ขั้นตอนหลักคือการปรับปรุงสัญญาณ การแทนที่ลักษณะเด่น การฉายลักษณะเด่นและการจำแนกประเภท แสดงดังภาพประกอบที่ 3-10

ขั้นตอนที่ 1 การปรับปรุงสัญญาณของคนปกติจำนวน 7 คนและผู้ที่มีอาการพูดไม่เป็นความจำนวน 5 คน ในกรณีของการกำจัดสัญญาณรบกวนวงจรกรองความถี่แถบผ่านซึ่งได้อธิบายไว้แล้วในหัวข้อ 3.2.1 ถูกนำมาประยุกต์ใช้ อย่างไรก็ตามพบว่ามีสัญญาณรบกวนที่มากับสายส่ง (power line noise) ที่ความถี่ 50 เฮิรซต์ปะปนมากับสัญญาณของผู้ที่มีอาการพูดไม่เป็นความด้วย ดังนั้นผู้วิจัยจึงเพิ่มวงจรกรองแบบนอตซ์ฟิลเตอร์ (notch filter) ที่ความถี่ 50 เฮิรซต์ เพื่อลดทอนสัญญาณรบกวนที่เกิดขึ้นซึ่งได้ออกแบบไว้แล้วในหัวข้อ 3.2.2 มาใช้กับสัญญาณของผู้ที่มีอาการพูดไม่เป็นความ ในส่วนของการหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อสำหรับข้อมูลของคนปกติได้อธิบายไว้แล้วในหัวข้อ 3.3 และการหาจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อในผู้ที่มีอาการพูดไม่เป็นความใช้วิธีการตัดด้วยมือ เนื่องจากขนาดของสัญญาณไฟฟ้ากล้ามเนื้อค่อนข้างต่ำ ทั้งนี้ระยะจากจุดเริ่มต้นถึงจุดสิ้นสุดของสัญญาณทั้งของคนปกติขนาดเท่ากับ 2.4 วินาทีและผู้ที่มีอาการพูดไม่เป็นความมีขนาดเท่ากับ 1.6 วินาที จำนวนของสัญญาณไฟฟ้ากล้ามเนื้อในแต่ละช่องสัญญาณของคนปกติและผู้ที่มีอาการพูดไม่เป็นความเท่ากับ 45 (9 พยางค์ × 5 การทำซ้ำต่อหนึ่งพยางค์)



ภาพประกอบที่ 3-10 แผนภาพบล็อกของระบบจำแนกพยางค์บนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ

ขั้นตอนที่ 2 การแทนที่ลักษณะเด่น สัญญาณไฟฟ้ากล้ามเนื้อที่ผ่านกระบวนการปรับปรุงสัญญาณแล้วจะถูกแบ่งย่อยเป็นเฟรม โดยแต่ละเฟรมมีขนาดเท่ากับ 250 มิลลิวินาที เฟรมที่

สองจะเลื่อนไปข้างหน้าโดยมีส่วนที่คาบเกี่ยวกัน 50 เปอร์เซ็นต์ (125 มิลลิวินาที) กระบวนการนี้จะทำซ้ำจนถึงจุดสิ้นสุดของสัญญาณ ผลที่ได้คือจำนวนเฟรมของแต่ละสัญญาณไฟฟ้ากล้ามเนื้อในคนปกติและผู้ที่มีอาการพูดไม่เป็นความเท่ากับ 22 และ 11 ตามลำดับ ดังนั้นจำนวนเฟรมของสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 9 พยางค์ของคนปกติและผู้ที่มีอาการพูดไม่เป็นความคือ 990 (22 เฟรม × 5 การทำซ้ำต่อหนึ่งพยางค์ × 9 พยางค์) และ 495 (11 เฟรม × 5 การทำซ้ำต่อหนึ่งพยางค์ × 9 พยางค์) ตามลำดับ ในแต่ละเฟรมจะมีการคำนวณลักษณะเด่นเรียกว่าการแทนที่ลักษณะเด่นรวมทั้งสิ้น 30 ลักษณะเด่น ทั้งนี้มาจากสัญญาณไฟฟ้ากล้ามเนื้อทั้ง 5 ช่องสัญญาณ และลักษณะเด่นทั้ง 6 แบบที่ได้กล่าวไว้ข้างต้น (5 ช่องสัญญาณ × 6 ลักษณะเด่นต่อช่องสัญญาณ) ดังนั้นในแต่ละเฟรมขนาดของเวคเตอร์ลักษณะเด่นเท่ากับ 30 จากนั้นลักษณะเด่นแต่ละตัวจะถูกนอร์ม (normalization) ด้วยอัลกอริทึ่ม min-max โดยค่าที่ได้จะอยู่ในช่วง -1 ถึง 1

ขั้นตอนที่ 3 การฉายลักษณะเด่น ลักษณะเด่นที่ได้จากขั้นตอนที่ 2 จะถูกลดมิติของข้อมูล โดยใช้เทคนิค SRELM ซึ่งจะทำให้มิติของเวคเตอร์ลักษณะเด่นลดลงจาก 30 เหลือ c-1 โดย c คือจำนวนคลาส สำหรับในการทดลองนี้ c มีค่าเท่ากับ 9 ดังนั้นมิติของเวคเตอร์ลักษณะเด่นหลังจากการฉายลักษณะเด่นจะถูกลดเหลือ 8 ลักษณะเด่น

ขั้นตอนที่ 4 การจำแนกประเภท เวกเตอร์ลักษณะเด่นที่ถูกลดมิติ (8 โหนด) แล้วจะถูกป้อนเข้าโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า โครงสร้างของโครงข่ายประสาทเทียมประกอบด้วย 3 ชั้นคือชั้นอินพุตจำนวน 8 โหนด ชั้นซ่อนและชั้นเอาต์พุตจำนวน 9 โหนด เนื่องจากจำนวนคลาสมีทั้งสิ้น 9 พยางค์ โดยใช้ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ซิกมอยด์ (hyberbolic tangent sigmoid) เป็นฟังก์ชันถ่ายโอน (transfer function) ทั้งในชั้นซ่อนและชั้นเอาต์พุต จากการทดลองปรับจำนวนโหนดในชั้นซ่อนจาก 15 ถึง 25 โหนดพบว่าจำนวน 20 โหนดให้ค่าความถูกต้องมากที่สุด

ในส่วนของการประเมินประสิทธิภาพของขั้นตอนวิธี ผู้วิจัยใช้วิธีการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) โดยนำข้อมูล 4 ส่วนมาใช้ในการสอนโครงข่ายประสาทเทียม และส่วนที่เหลือมาใช้ทดสอบประสิทธิภาพของโครงข่ายประสาทเทียม โดยข้อมูลที่ถูกสอนและทดสอบจะต้องไม่ใช่ข้อมูลเดียวกัน กระบวนการนี้จะทำซ้ำ 5 ครั้ง นั่นหมายความว่าข้อมูลทั้ง 5 ส่วนจะต้องถูกทดสอบ ประสิทธิภาพของการจำแนกคำนวณจากค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของทั้ง 5 ชุดทดสอบ ยิ่งไปกว่านั้นระบบรู้จำคำพูดที่ใช้ในการทดลองนี้เป็นระบบที่ขึ้นกับผู้พูด (speaker-dependent criteria) นั่นหมายความว่าข้อมูลที่ถูกสอนและทดสอบเป็นข้อมูลที่มาจากคนเดียวกัน

3.4.2 การเปรียบเทียบประสิทธิภาพของการสกัดลักษณะเด่นแบบฉายข้อมูล

ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของตัวฉายลักษณะเด่น 3 แบบคือ SRELM การวิเคราะห์องค์ประกอบหลัก (principal component analysis: PCA) และการวิเคราะห์จำแนกกลุ่มเชิงเส้น (linear discriminant analysis: LDA) โดยแทนที่สัญญาณไฟฟ้ากล้ามเนื้อด้วยลักษณะเด่นทั้ง 6 แบบ (ACF) คือ MAV, WL, ZC, MNF, $L-KURT$ และ $L-SKW$ สำหรับกระบวนการในการจำแนกพยางค์เหมือนกับที่กล่าวไปแล้วในหัวข้อ 3.4.1

3.5 คุณลักษณะของสัญญาณเสียง

จากการทบทวนวรรณกรรมทางด้านการรู้จำคำพูดบนพื้นฐานของสัญญาณเสียงในหัวข้อ 1.5 พบว่าลักษณะเด่น MFCC ได้รับความนิยมในการนำมาใช้มากที่สุด อย่างไรก็ตามประเด็นที่น่าสนใจคือจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันที่เหมาะสมสำหรับใช้ในการจำแนกพยางค์ นอกจากนี้ผู้วิจัยได้นำลักษณะเด่นที่นิยมใช้กับสัญญาณไฟฟ้ากล้ามเนื้อมาประยุกต์ใช้กับสัญญาณเสียงอีกด้วย ดังนั้นในการออกแบบการทดลองผู้วิจัยได้แบ่งการทดลองเป็น 2 ส่วนหลักคือการเปรียบเทียบความถูกต้องของการจำแนกพยางค์โดยแปรผันจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน และการเปรียบเทียบประสิทธิภาพของการจำแนกพยางค์โดยใช้ลักษณะเด่น 2 กลุ่ม

ในกรณีการทดลองที่หนึ่ง ทำการวิเคราะห์สัญญาณเสียงทั้ง 12 พยางค์แสดงดังตารางที่ 2-1 โดยแปรผันจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน ในการทดลองนี้ตั้งค่าจำนวนสัมประสิทธิ์ต่อเฟรมของลักษณะเด่น MFCC ไว้ 3 ค่าคือ 8 13 และ 18 ตามลำดับ ในส่วนของขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันกำหนดไว้ 2 แบบคือขนาดของเฟรมเท่ากับ 25 มิลลิวินาที โดยมีขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 10 มิลลิวินาที เขียนแทนด้วย 25 (10) และขนาดของเฟรมเท่ากับ 250 มิลลิวินาที โดยมีขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 125 มิลลิวินาที เขียนแทนด้วย 250 (125) สำหรับเหตุผลในการเลือกค่าเหล่านี้ได้อธิบายไว้แล้วในหัวข้อ 1.8 ในส่วนของระบบรู้จำเสียงพูดบนพื้นฐานของสัญญาณเสียง กระบวนการในการจำแนกพยางค์เริ่มจากการนำสัญญาณเสียงมาผ่านการปรับปรุงสัญญาณโดยการกำจัดสัญญาณรบกวนด้วยวงจรกรองแบบความถี่ต่ำผ่านซึ่งได้อธิบายในหัวข้อ 3.2.3 และการตัดแยกสัญญาณโดยอ้างอิงจากจุดเริ่มต้นของสัญญาณไฟฟ้ากล้ามเนื้อ จากนั้นนำสัญญาณที่ได้มาแบ่งย่อยเป็นเฟรมโดยมีขนาดตามที่กำหนดได้ข้างต้นคือ 25 (10) และ 250 (125) ถัดมานำเฟรมที่ได้มาคำนวณลักษณะเด่น MFCC ซึ่งมีจำนวนสัมประสิทธิ์เท่ากับ 8, 13 และ 18 เนื่องจากเวกเตอร์ลักษณะเด่นมีขนาดเล็กดังนั้นจึงไม่จำเป็นต้องลดมิติของข้อมูล ดังนั้นมิติของเวคเตอร์ลักษณะเด่นจึงเท่ากับ 8, 13 และ 18 ตามลำดับ สุดท้ายจำแนกพยางค์ด้วยโครงข่ายประสาทเทียมแบบไปข้างหน้า โดยที่จำนวนโหนดของ

แต่ละชั้นแสดงดังนี้ จำนวนโหนดของชั้นอินพุตแปรผันตามมิติของเวคเตอร์ลักษณะเด่น จำนวนโหนดของชั้นซ่อนเท่ากับ 20 และจำนวนโหนดของชั้นเอาต์พุตเท่ากับ 12 โหนด เนื่องจากจำนวนคลาสมีทั้งสิ้น 12 พยางค์ พารามิเตอร์ที่สำคัญในการคำนวณค่าสัมประสิทธิ์ของ MFCC มีดังนี้จำนวนตัวกรอง (filter bank) เท่ากับ 20 และจำนวน cepstral sine lifter เท่ากับ 22 กระบวนการในการคำนวณค่า MFCC แสดงดังภาพประกอบที่ 3-11 มีรายละเอียดนี้

ขั้นตอนที่ 1 ตัดแยกสัญญาณเสียงให้มีขนาดความยาว 2.4 วินาที

ขั้นตอนที่ 2 กำจัดสัญญาณรบกวนโดยใช้วงจรความถี่ต่ำผ่านที่ความถี่ตัด (cut-off frequency) 5 กิโลเฮิรซต์

ขั้นตอนที่ 3 เพิ่มอัตราส่วนของสัญญาณที่ต้องการต่อสัญญาณรบกวน โดยประมวลผลสัญญาณที่ได้จากขั้นตอนที่ 2 ด้วยตัวกรองการเน้นล่วงหน้า (pre-emphasis) ที่มีค่าสัมประสิทธิ์ของตัวกรอง 0.97

ขั้นตอนที่ 4 แบ่งสัญญาณจากขั้นตอนที่ 3 ออกเป็นเฟรมขนาด 25 มิลลิวินาที ที่มีการเลื่อนเฟรมไปข้างหน้าโดยมีส่วนที่ซ้อนทับกันขนาด 10 มิลลิวินาที เขียนแทนด้วย 25(10) และเฟรมขนาด 250 มิลลิวินาที ที่มีการเลื่อนเฟรมไปข้างหน้าโดยมีส่วนที่ซ้อนทับกันขนาด 125 มิลลิวินาที เขียนแทนด้วย 250(125) ในแต่ละเฟรมประกอบด้วยเวกเตอร์ลักษณะเด่นจำนวน 8, 13 และ 18 ตามลำดับ

ขั้นตอนที่ 5 คูณแต่ละเฟรมที่ได้จากขั้นตอนที่ 4 ด้วยสัญญาณหน้าต่างแบบแฮมมิง (Hamming windowing) เพื่อลดทอนความไม่ต่อเนื่องที่ขอบของเฟรม

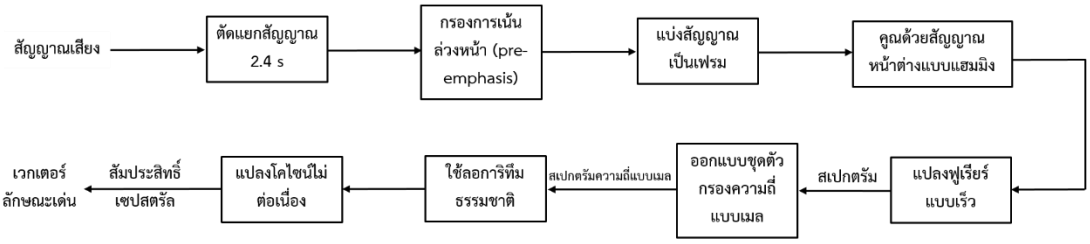
ขั้นตอนที่ 6 คำนวณขนาดของสเปคตรัมของแต่ละเฟรมที่ได้จากขั้นตอนที่ 5 โดยใช้วิธีผลการแปลงฟูเรียร์แบบเร็ว (fast Fourier transform: FFT)

ขั้นตอนที่ 7 ออกแบบตัวกรองที่มีลักษณะเป็นสามเหลี่ยมที่มีพื้นที่เท่ากันจำนวน 20 ตัวกรอง บนสเกลของเมลระหว่าง 15.99 ถึง 2363.50 หรือบนความถี่ระหว่าง 10 ถึง 5000 เฮิรซต์ ซึ่งครอบคลุมพลังงานส่วนใหญ่ของสัญญาณเสียง การแปลงจากความถี่ซึ่งมีหน่วยเป็นเฮิรซต์เป็นสเกลของเมลแสดงดังสมการ 2-7

ขั้นตอนที่ 8 คูณขนาดของสเปคตรัมของแต่ละเฟรมที่ได้จากขั้นตอนที่ 6 กับตัวกรองที่ได้ออกแบบในขั้นตอนที่ 7 จากนั้นหาผลรวมของผลคูณดังกล่าว ผลลัพธ์สุดท้ายจากขั้นตอนนี้จะได้ผลรวมของพลังงานในแต่ละย่านความถี่จำนวน 20 ค่า.

ขั้นตอนที่ 9 หาค่าลอการิทึมธรรมชาติของผลรวมของพลังงานในแต่ละย่านความถี่จำนวน 20 ค่าจากขั้นตอนที่ 8 จากนั้นแปลงโคไซน์ไม่ต่อเนื่อง (discrete cosine transform) เพื่อให้ได้จำนวนสัมประสิทธิ์ MFCC ตามที่ต้องการ คือ 8 13 หรือ 18

ขั้นตอนที่ 10 ประยุกต์ใช้เทคนิคลิฟเตอร์ริง (liftering) ที่มีค่าพารามิเตอร์ 22 กับค่าสัมประสิทธิ์ MFCC ที่ได้จากขั้นตอนที่ 9 เพื่อลดปัญหาเรื่องการแปรผันอย่างมีนัยสำคัญของค่าขนาดสัมประสิทธิ์ MFCC ลำดับต่ำและค่าขนาดสัมประสิทธิ์ MFCC ลำดับสูง



ภาพประกอบที่ 3-11 แผนภาพบล็อกของการคำนวณค่า MFCC

สุดท้ายนี้จะได้เวกเตอร์ลักษณะเด่นของสัญญาณเสียงโดยจัดรูปแบบ MFCC ที่อันดับเดียวกัน ผลที่ตามมาคือสัญญาณเสียง 1 สัญญาณถูกแปลงเป็นเวกเตอร์ลักษณะเด่นที่มีจำนวนสัมประสิทธิ์เท่ากับ 8, 13 และ18 ตามลำดับ

ในส่วนของการทดลองที่สอง ลักษณะเด่นออกถูกแบ่งออกเป็น 2 กลุ่มหลักดังนี้ (1) ลักษณะเด่นในโดเมนเวลาที่นิยมประยุกต์ใช้กับสัญญาณไฟฟ้ากล้ามเนื้อดังได้กล่าวไปแล้วข้างต้น จำนวน 5 ลักษณะเด่นดังนี้ MAV, WL, ZC, SSC และ AR ลำดับที่ 4 รวมเป็น 8 ลักษณะเด่น (2) ลักษณะเด่นที่นิยมใช้ในการจำแนกเสียงคือ MFCC ที่มีจำนวนสัมประสิทธิ์ต่อเฟรมเท่ากับ 13 โดยลักษณะเด่นในกลุ่มที่หนึ่งกำหนดขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 250 มิลลิวินาที และ 125 มิลลิวินาที ตามลำดับ ขณะที่ลักษณะเด่นกลุ่มที่กำหนดขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 25 มิลลิวินาที และ 10 มิลลิวินาที ตามลำดับ สุดท้ายจำแนกพยางค์ด้วยโครงข่ายประสาทเทียมแบบไปข้างหน้า โดยที่จำนวนโหนดของแต่ละชั้นแสดงดังนี้ จำนวนโหนดของชั้นอินพุตเท่ากับ 8 สำหรับลักษณะเด่นในกลุ่มที่หนึ่งและเท่ากับ 13 สำหรับลักษณะเด่นในกลุ่มที่สอง จำนวนโหนดของชั้นซ่อนเท่ากับ 20 และจำนวนโหนดของชั้นเอาต์พุตเท่ากับ 12 โหนด

3.6 การรวมข้อมูล

การรวมข้อมูลประกอบด้วยการรวมข้อมูลจากแหล่งที่มาเดียวกัน (unimodal fusion) และจากหลายแหล่งที่มา (multimodal fusion) มีรายละเอียดดังนี้

3.6.1 การรวมข้อมูลจากแหล่งที่มาเดียวกัน

การรวมข้อมูลจากแหล่งที่มาเดียวกันเป็นการรวมลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อจากหลายช่องสัญญาณ ในการทดลองนี้ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของการจำแนกพยางค์แยกตามกลุ่มของช่องสัญญาณดังนี้ สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณเดียวและการรวมสัญญาณไฟฟ้ากล้ามเนื้อหลายช่องสัญญาณประกอบด้วย 2 ช่องสัญญาณ (10 กลุ่มย่อย) 3 ช่องสัญญาณ (10 กลุ่มย่อย) 4 ช่องสัญญาณ (5 กลุ่มย่อย) และ 5 ช่องสัญญาณดังตารางที่ 3-1 โดยลักษณะเด่นที่ใช้ประกอบด้วย MAV, WL, ZC, SSC และ AR ลำดับที่ 4 รวมทั้งสิ้น 8 ลักษณะเด่น ดังนั้นความยาวของเวกเตอร์ลักษณะเด่นของ 1 2 3 4 และ 5 ช่องสัญญาณคือ 8, 16, 24, 32 และ 40 ตามลำดับ ในส่วนของกระบวนการจำแนกพยางค์ไทยได้กล่าวไปแล้วในหัวข้อ 3.4 ดังภาพประกอบที่ 3-10 หลังจากกำจัดสัญญาณรบกวนแล้วสัญญาณไฟฟ้ากล้ามเนื้อจะถูกตัดแยกโดยความยาวของสัญญาณไฟฟ้ากล้ามเนื้อจากจุดเริ่มต้นถึงจุดสิ้นสุดเท่ากับ 2.4 วินาที ดังได้กล่าวไปแล้วในหัวข้อ 3.3 หลังจากนั้นลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อในแต่ละช่องสัญญาณจะถูกคำนวณถัดมากระบวนการลดมิติข้อมูลด้วยเทคนิค SRELM ถูกนำมาใช้กับการรวมสัญญาณไฟฟ้ากล้ามเนื้อหลายช่องสัญญาณส่งผลให้ความยาวของเวกเตอร์ลักษณะเด่นของการรวมสัญญาณแบบ 2 3 4 และ 5 ช่องสัญญาณรวมกันถูกลดเหลือเพียง 11 เท่านั้น ยกเว้นลักษณะเด่นของช่องสัญญาณเดียวที่ไม่ผ่านกระบวนการลดมิติข้อมูลเนื่องจากความยาวของเวกเตอร์ลักษณะเด่นเท่ากับ 8 ซึ่งน้อยกว่าเวกเตอร์เอาท์พุตของ SRELM ซึ่งมีขนาดเท่ากับ 11 (จำนวนคลาส - 1) กระบวนการสุดท้ายคือการจำแนกพยางค์โดยใช้โครงข่ายประสาทเทียมแบบไปข้างหน้า โดยที่จำนวนโหนดของแต่ละชั้นแสดงดังนี้ จำนวนโหนดของชั้นอินพุตแปรผันตามมิติของเวคเตอร์ลักษณะเด่น จำนวนโหนดของชั้นซ่อนเท่ากับ 20 และจำนวนโหนดของชั้นเอาต์พุตเท่ากับ 12 การรวมข้อมูลจากหลายแหล่งที่มา

3.6.2 การรวมข้อมูลจากหลายแหล่งที่มา

การรวมข้อมูลจากหลายแหล่งที่มาเป็นการรวมลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง สำหรับการทดลองในหัวข้อนี้ผู้วิจัยได้ทำการทดสอบประสิทธิภาพของการจำแนกพยางค์ไทยโดยการรวมข้อมูลในระดับลักษณะเด่นซึ่งเกิดขึ้นก่อนการฉายข้อมูล วิธีการรวมข้อมูลคือการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงมาเชื่อมต่อกัน โดยผู้วิจัยกำหนดพารามิเตอร์ที่สำคัญ 3 ค่าคือช่วงเวลาของการวิเคราะห์สัญญาณของสัญญาณไฟฟ้ากล้ามและสัญญาณเสียงซึ่งมีขนาดเท่ากัน จำนวนสัมประสิทธิ์ของ MFCC และจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ ในส่วนของช่วงเวลาของการวิเคราะห์สัญญาณกำหนดไว้ 3 ค่า คือ 1.125 วินาที (8 เฟรมต่อหนึ่งสัญญาณ) 1.75 วินาที (13 เฟรมต่อหนึ่งสัญญาณ) และ 2.4 วินาที (18 เฟรมต่อหนึ่งสัญญาณ) ตามลำดับ ผลที่ตามมาคือจำนวนข้อมูลของการสอนและทดสอบเปลี่ยนแปลงตาม

ช่วงเวลาของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงดังนี้ 480 (8 เฟรม × 60 สัญญาณ) 780 (13 เฟรม × 60 สัญญาณ) และ 1080 (18 เฟรม × 60 สัญญาณ) ตามลำดับ ในกรณีของจำนวนสัมประสิทธิ์ของ MFCC กำหนดไว้ 3 ค่าเช่นกันดังได้กล่าวไปแล้วในหัวข้อ 3.5 คือ 8 13 และ 18 ตามลำดับ สำหรับจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้ออ้างอิงตามการรวมข้อมูลจากแหล่งที่มาเดียวกันคือ 1 ช่องสัญญาณ 2, 3, 4 และ 5 ช่องสัญญาณรวมกัน กระบวนการของการรวมข้อมูลจากหลายแหล่งที่มาแสดงดังภาพประกอบที่ 3-12 อธิบายได้ดังนี้

ขั้นตอนที่ 1 เก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงพร้อมกัน

ขั้นตอนที่ 2 ลดทอนสัญญาณรบกวนด้วยวงจรกรองในกระบวนการปรับปรุงสัญญาณที่ได้กล่าวไปแล้วข้างต้น โดยแยกทำทีละสัญญาณ

ขั้นตอนที่ 3 คำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงดังได้อธิบายไปแล้วข้างต้น สำหรับสัญญาณไฟฟ้ากล้ามเนื้อได้คัดเลือกช่องสัญญาณที่มีประสิทธิภาพสูงสุดในการจำแนกเป็นตัวแทนของ 1, 2, 3 และ 4 ช่องสัญญาณที่ได้จากการรวมสัญญาณแบบแหล่งที่มาเดียวกัน

ขั้นตอนที่ 4 นอร์มลักษณะเด่นของแต่ละสัญญาณเพื่อให้เป็นบรรทัดฐานเดียวกัน โดยใช้อัลกอริทึม min-max ค่าที่ได้อยู่ในช่วง -1 ถึง 1

ขั้นตอนที่ 5 เชื่อมต่อเวกเตอร์ลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง ยกตัวอย่างเช่นเวกเตอร์ลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อของ 1 ช่องสัญญาณเท่ากับ 8 เมื่อเชื่อมกับเวกเตอร์ลักษณะเด่นของสัญญาณเสียงซึ่งเท่ากับ 13 ผลที่ได้คือเวกเตอร์ลักษณะเด่นของการรวมสัญญาณเท่ากับ 21 (8 ลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ + 13 ลักษณะเด่นของสัญญาณเสียง) เช่นเดียวกับเวกเตอร์ลักษณะเด่นของการรวม 2, 3, 4 และ 5 ช่องสัญญาณกับสัญญาณเสียงมีขนาดเป็น 24, 32, 40 และ 48 ตามลำดับ

ขั้นตอนที่ 6 ลดมิติของข้อมูลโดยการฉายข้อมูลด้วยอัลกอริทึม SRELM ทำให้ขนาดของเวกเตอร์ลักษณะเด่นลดลงเหลือ 11 ค่า ซึ่งเป็นอินพุตเวกเตอร์ของโครงข่ายประสาทเทียม

ขั้นตอนที่ 7 จำแนกพยางค์โดยใช้โครงข่ายประสาทเทียมแบบไปข้างหน้าโดยวิธีการกำหนดค่าพารามิเตอร์ของโครงข่ายประสาทเทียมได้อธิบายแล้วข้างต้น โดยจำนวนโหนดของชั้นอินพุตแปรผันตามมิติของเวคเตอร์ลักษณะเด่น

ตารางที่ 3-1 การจับคู่ของช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ

จำนวนของช่องสัญญาณ	การจับคู่ช่องสัญญาณ				
1 (5 กลุ่ม)	1	2	3	4	5
2 (10 กลุ่ม)	1-2	1-3	1-4	1-5	2-3
	2-4	2-5	3-4	3-5	4-5
3 (10 กลุ่ม)	1-2-3	1-2-4	1-2-5	1-3-4	1-3-5
	1-4-5	2-3-4	2-3-5	2-4-5	3-4-5
4 (5 กลุ่ม)	1-2-3-4	1-2-3-5	1-2-4-5	1-3-4-5	2-3-4-5
5 (1 กลุ่ม)			1-2-3-4-5		

3.7 การรู้จำคำพูดแบบไม่ขึ้นกับบุคคล

การรู้จำคำพูดแบบไม่ขึ้นกับบุคคล (independent-speaker) คือการนำระบบที่สร้างขึ้นจากข้อมูลของกลุ่มบุคคลหนึ่งไปทดสอบกับข้อมูลจากบุคคลอื่นที่ไม่เคยได้รับการฝึกสอนมาก่อน การทดลองนี้ผู้วิจัยแบ่งข้อมูลเป็น 2 ส่วนคือข้อมูลสอนและข้อมูลทดสอบ ทั้งนี้ได้นำข้อมูลของอาสาสมัครปกติ 5 คนมาใช้เป็นข้อมูลฝึกสอน และข้อมูลของอาสาสมัครปกติ 2 คนที่เหลือเป็นข้อมูลทดสอบ โดยนำช่วงเวลาของการวิเคราะห์สัญญาณและจำนวนสัมประสิทธิ์ของ MFCC ที่ดีที่สุดของการรวมจากหลายแหล่งที่มาเมื่อจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อเท่ากับ 2 ที่ได้จากการทดลองในหัวข้อ 3.6.2 มาประยุกต์ใช้ และกระบวนการในการจำแนกพยางค์คล้ายคลึงกับที่กล่าวไปแล้วในหัวข้อ 3.6.2 สิ่งที่แตกต่างคือค่าน้อยสุดและมากสุดที่ใช้ในการนอร์ม เนื่องจากใช้อัลกอริทึม min-max วิธีการหาค่าน้อยสุดและมากสุดสามารถทำได้ดังนี้

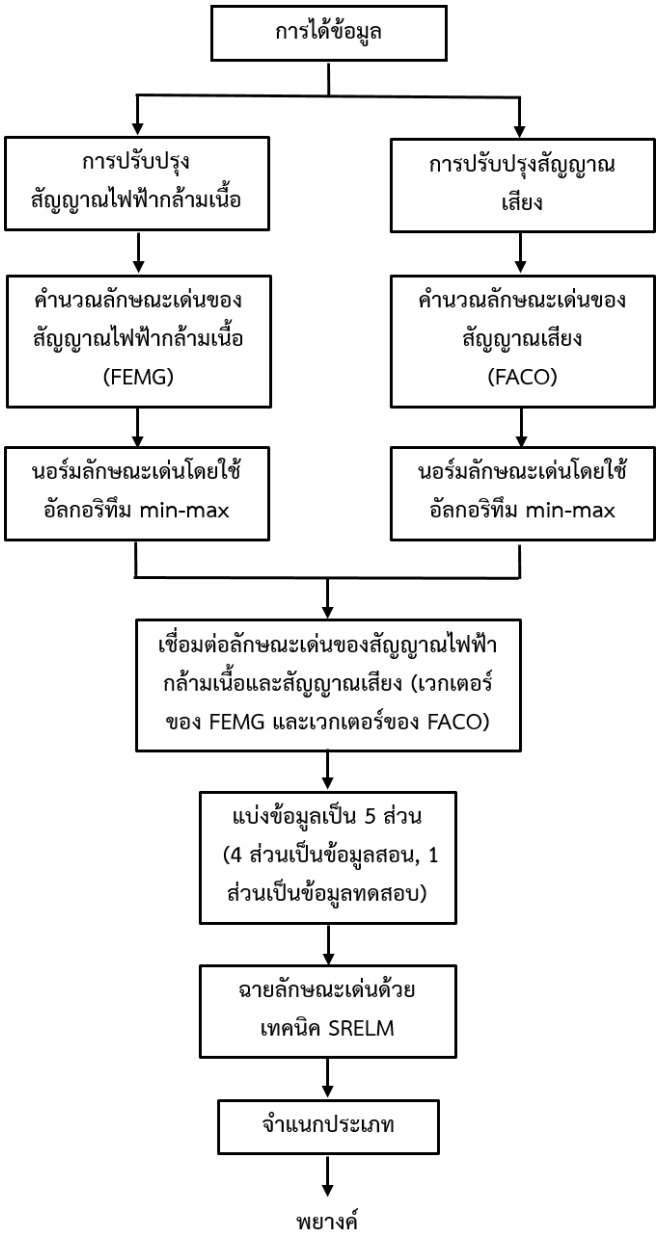
ขั้นตอนที่ 1 หาค่าน้อยสุดและมากสุดของอาสาสมัครแต่ละคน พร้อมบันทึกไว้

ขั้นตอนที่ 2 หาค่าเฉลี่ยของค่าน้อยสุดและมากสุดของอาสาสมัครทั้งหมด พร้อมบันทึกไว้

ขั้นตอนที่ 3 นำค่าที่ได้จากขั้นตอนที่ 2 มาใช้นอร์มลักษณะเด่นของแต่ละสัญญาณ

ในส่วนของการเรียงข้อมูลคือการนำข้อมูลของอาสาสมัครทั้ง 5 คนมาต่อกันเพื่อใช้เป็นข้อมูลฝึกสอน ดังนั้นจำนวนเฟรมทั้งหมดเท่ากับ 2400 (8 เฟรม × 60 สัญญาณ × 5 คน) ในส่วน

ของข้อมูลทดสอบก็เช่นกันเป็นการนำข้อมูลของอาสาสมัคร 2 คนมาต่อกัน ดังนั้นจำนวนเฟรมทั้งหมดเท่ากับ 960 (8 เฟรม × 60 สัญญาณ × 2 คน) โดยผู้วิจัยได้จับกลุ่มข้อมูลฝึกสอนและข้อมูลทดสอบยกตัวอย่างเช่นครั้งแรกใช้ข้อมูลของอาสาสมัครคนที่ 1 ถึง 5 เป็นข้อมูลฝึกสอน และข้อมูลของอาสาสมัครคนที่ 6 และ 7 เป็นข้อมูลทดสอบ จากนั้นทำการลดมิติของข้อมูลวิธี SRELM และจำแนกพยางค์โดยใช้โครงข่ายประสาทเทียมแบบไปข้างหน้า สำหรับค่าพารามิเตอร์ต่างๆ ได้อธิบายไปแล้วข้างต้น ครั้งถัดไปจะทำการสลับสับเปลี่ยนข้อมูลโดยใช้ข้อมูลของอาสาสมัครคนที่ 2 ถึง 6 เป็นข้อมูลสอน และข้อมูลของอาสาสมัครคนที่ 1 และ 7 เป็นข้อมูลทดสอบ โดยชุดข้อมูลจะถูกสลับสับเปลี่ยนจนกว่าจะครบทุกเงื่อนไข เพื่อเฉลี่ยค่าความถูกต้องของการจำแนกพยางค์



ภาพประกอบที่ 3-12 แผนภาพบล็อกของการรวมลักษณะเด่นก่อนการฉายข้อมูล

บทที่ 4

ผลการทดลองและการอภิปรายผล

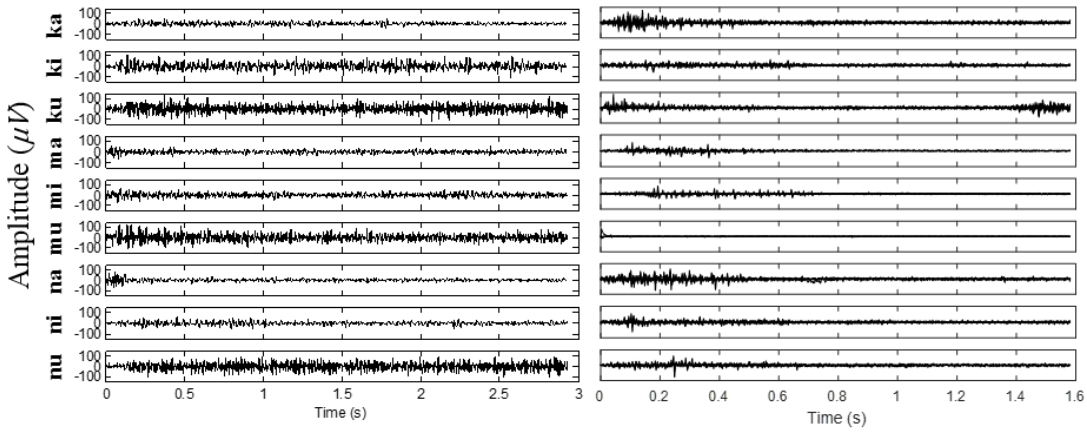
บทนี้กล่าวถึงผลที่ได้จากการทดลองซึ่งได้อธิบายวิธีการทดลองแล้วในบทที่ 3 แยกตามหัวข้อของการทดลองประกอบด้วยผลจากการศึกษาคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อของคนปกติและผู้ที่มีอาการพูดไม่เป็นความ ประสิทธิภาพของการจำแนกพยางค์ของสัญญาณเสียง ประสิทธิภาพของการรวมข้อมูลและประสิทธิภาพของการรู้จำพยางค์แบบไม่ขึ้นกับบุคคล

4.1 คุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อของคนปกติและผู้ที่มีอาการพูดไม่เป็นความ

ผลการทดลองสามารถแบ่งเป็น 2 หัวข้อย่อยคือประสิทธิภาพของการจำแนกพยางค์ของสัญญาณไฟฟ้ากล้ามเนื้อตามประเภทของกลุ่มลักษณะเด่น และประสิทธิภาพของการสกัดลักษณะเด่นแบบฉายข้อมูลด้วยเทคนิค PCA, LDA และ SRELM แสดงรายละเอียดดังนี้

4.1.1 ประสิทธิภาพของการจำแนกพยางค์ของสัญญาณไฟฟ้ากล้ามเนื้อตามประเภทของกลุ่มลักษณะเด่น

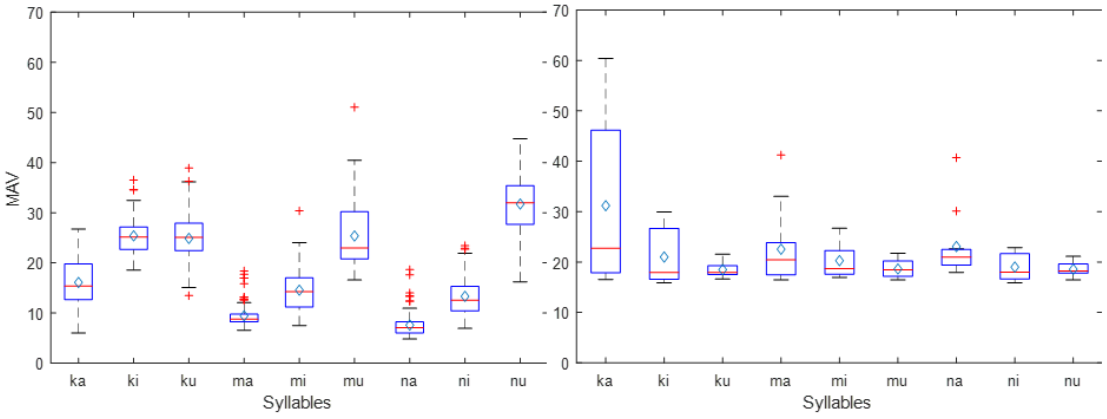
ภาพประกอบที่ 4-1(ก) และ (ข) แสดงตัวอย่างของสัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 เมื่อออกเสียงทั้ง 9 พยางค์ทั้งของคนปกติและผู้ที่มีอาการพูดไม่เป็นความตามลำดับ ผลการทดลองแสดงให้เห็นว่าขนาดของสัญญาณไฟฟ้ากล้ามเนื้อของคนปกติมีค่าสูงกว่าของผู้ป่วยอาการพูดไม่เป็นความทั้ง 9 พยางค์เนื่องจากคนปกติมีการหดตัวของมัดกล้ามเนื้อที่แข็งแรงกว่า



(ก) (ข)

ภาพประกอบที่ 4-1 ตัวอย่างของสัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 เมื่อออกเสียงทั้ง 9 พยางค์ (ก) คนปกติ (ข) ผู้ที่มีอาการพูดไม่เป็นความ

ภาพประกอบที่ 4-2 (ก) และ (ข) แสดงแผนภาพกล่องของ MAV ซึ่งถูกกำหนดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 ของคนปกติและผู้ที่มีอาการพูดไม่เป็นความตามลำดับ สำหรับแต่ละกล่องแสดงข้อมูลทั้งหมด 3 ควอไทล์โดยมีการจัดเรียงอันดับของข้อมูลคือ ข้อมูล 25 เปอร์เซ็นต์ แรกจากค่าต่ำขึ้นมา (Q1) ข้อมูลที่มากกว่า 25 เปอร์เซ็นต์จนถึง 75 เปอร์เซ็นต์ (Q2) และข้อมูลที่มากกว่า 75เปอร์เซ็นต์ขึ้นไป (Q3) โดยค่า Q1 และ Q3 จะเขียนแทนด้วยเส้นตรง (ขีดล่างและขีดบนของกล่อง) ส่วนค่า Q2 แสดงด้วยกล่องสี่เหลี่ยมผืนผ้า และสี่เหลี่ยมข้าวหลามตัดสีน้ำเงินคือค่าเฉลี่ยของข้อมูล นอกจากนี้ยังแสดงค่ากลางของข้อมูล (เส้นตรงสีแดง) และ ค่าพิสัยควอร์ไทล์ (IRQ = Q3-Q1) ของแต่ละพยางค์ จากแผนภาพกล่องสามารถสังเกตได้ว่าความแตกต่างของความแปรผันระหว่างกลุ่ม (มา มี และมู) มีความชัดเจนในคนปกติ ขณะที่ความแปรผันภายในกลุ่ม (คา มา นา) มีค่าน้อย ในทางตรงข้ามความแตกต่างของความแปรผันระหว่างกลุ่มของผู้ป่วยค่อนข้างคลุมเครือ โดยมีค่าใกล้เคียงกันทุกพยางค์ ขณะที่ความแปรผันของการออกเสียง “คา” มีค่ามาก อย่างไรก็ตามเพื่อที่จะตรวจสอบนัยสำคัญทางสถิติของผลจากแผนภาพกล่อง การวิเคราะห์ความแปรปรวนทางเดียว (One-way Analysis of Variance (ANOVA)) ถูกนำมาใช้และกำหนดระดับนัยสำคัญเท่ากับ 0.05 ผลการทดลองบ่งชี้ว่าความแตกต่างของความแปรผันระหว่างกลุ่มในคนปกติมีความชัดเจนขณะที่ความแตกต่างของความแปรผันระหว่างกลุ่มในผู้ป่วยไม่แตกต่างอย่างมีนัยสำคัญ ดังนั้นผลการทดสอบทางสถิติเป็นไปทางเดียวกันกับผลการทดลองจากการสังเกต

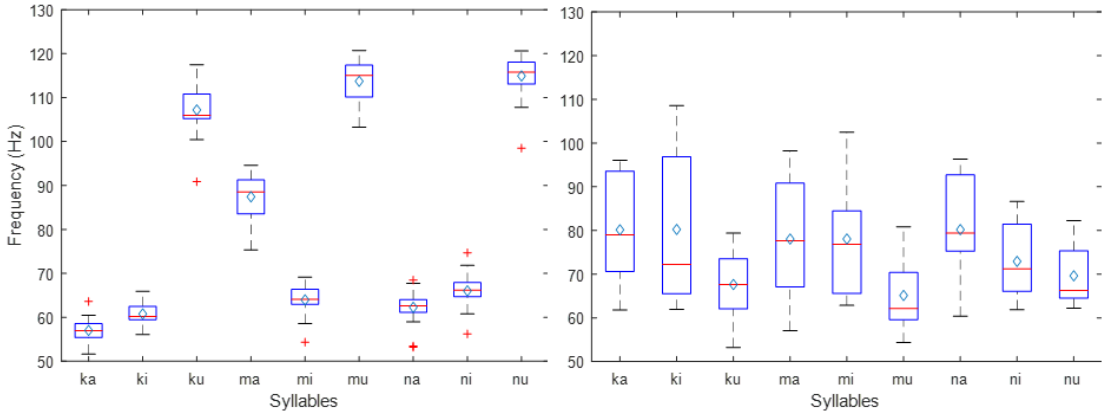


(ก) (ข)

ภาพประกอบที่ 4-2 แผนภาพกล่องของค่า MAV ซึ่งถูกกำหนดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 (ก) คนปกติ (ข) ผู้ที่มีอาการพูดไม่เป็นความ

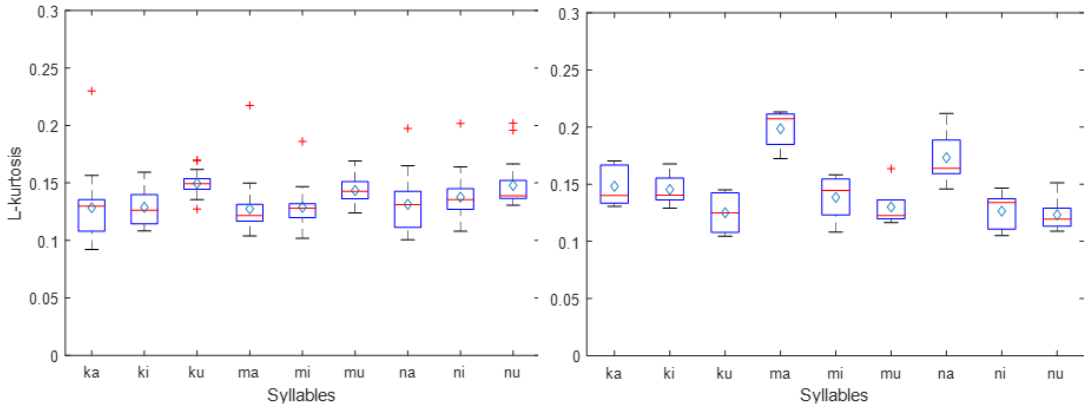
ภาพประกอบที่ 4-3(ก) แสดงแผนภาพกล่องของ MNF ของคนปกติ พบว่ามีแนวโน้มคล้ายคลึงกับ MAV คือความแปรผันระหว่างกลุ่มมีความชัดเจน โดยเฉพาะอย่างยิ่งเมื่อนำพยัญชนะทั้ง 3 มาผสมกับสระอู ทำให้มีความถี่ค่อนข้างสูงแยกจากสระอื่นอย่างชัดเจน ขณะที่ความ

แปรผันภายในกลุ่ม (คา มา นา) มีค่าน้อย ในทางตรงข้ามค่าความถี่กลางของผู้ที่มีอาการพูดไม่เป็นความของแต่ละพยางค์ค่อนข้างเหลื่อมล้ำกันแสดงดังภาพประกอบที่ 4-3(ข) ผลการทดสอบจากการวิเคราะห์ความแปรปรวนทางเดียวมีความสอดคล้องกับการสังเกตแผนภาพกล่องที่ได้อธิบายแล้วในข้างต้น



(ก) (ข)

ภาพประกอบที่ 4-3 แผนภาพกล่องของ MNF ซึ่งถูกกำหนดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 (ก) คนปกติ (ข) ผู้ที่มีอาการพูดไม่เป็นความ



(ก) (ข)

ภาพประกอบที่ 4-4 แผนภาพกล่องของค่า $L-KURT$ ซึ่งถูกกำหนดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 (ก) คนปกติ (ข) ผู้ที่มีอาการพูดไม่เป็นความ

ภาพประกอบที่ 4-4 (ก) และ (ข) แสดงแผนภาพกล่องของค่า $L-KURT$ ซึ่งถูกกำหนดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากช่องสัญญาณที่ 2 ของคนปกติและผู้ที่มีอาการพูดไม่เป็นความตามลำดับ จากการสังเกตพบว่าความแตกต่างของความแปรผันระหว่างกลุ่มและความแปรผัน

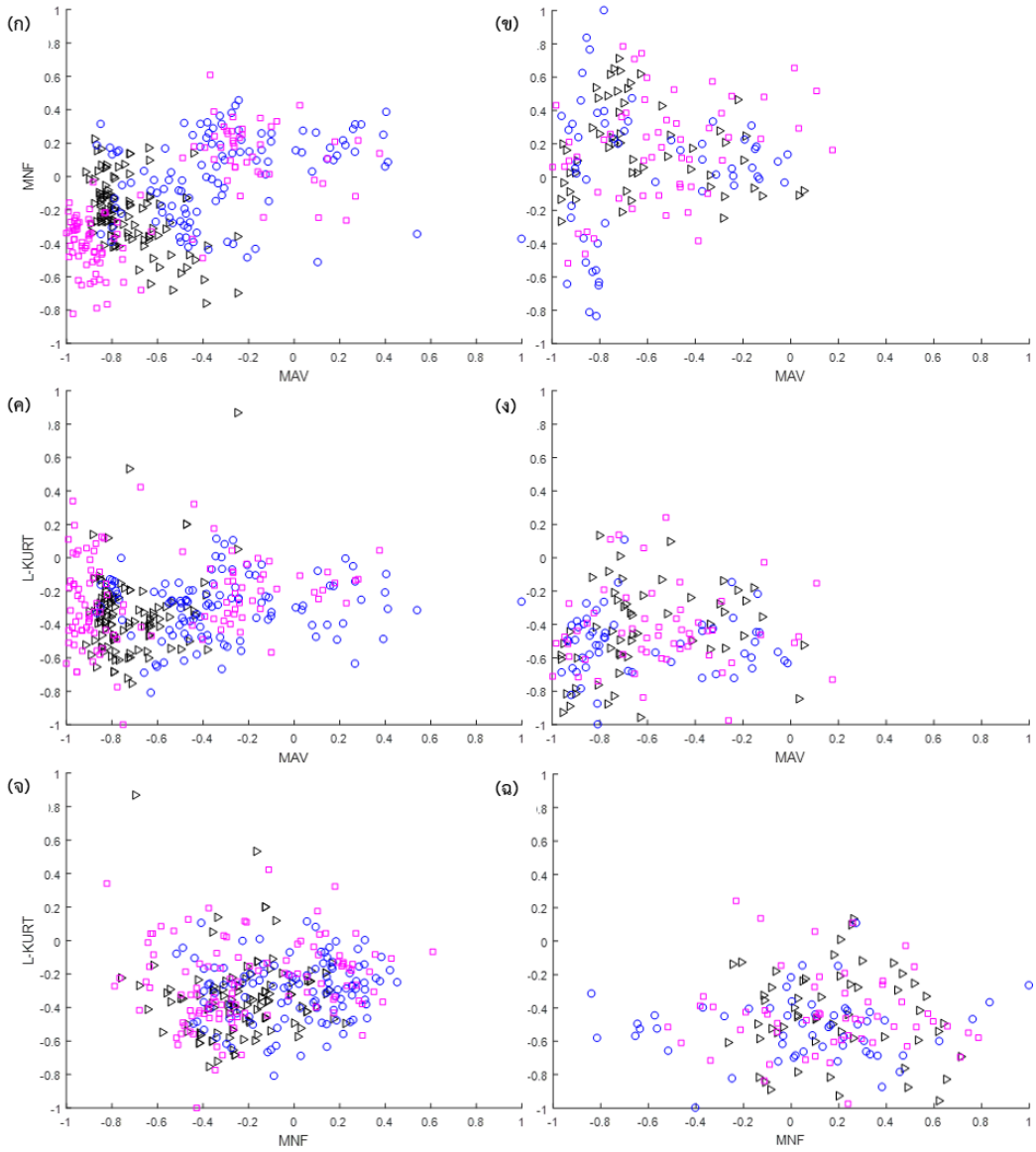
ภายในกลุ่มมีค่าใกล้เคียงกันทั้งในคนปกติและผู้ที่มีอาการพูดไม่เป็นความ อย่างไรก็ตามผลจากการวิเคราะห์ค่าความแปรปรวนทางเดียวบ่งชี้ว่าความแตกต่างของความแปรผันภายในกลุ่มของผู้ที่มีอาการพูดไม่เป็นความต่ำกว่าของคนปกติ ผลการทดลองเหล่านี้แสดงให้เห็นว่าการจำแนกเสียงทั้ง 9 พยางค์ด้วยลักษณะเด่นเพียงกลุ่มเดียวทำได้ค่อนข้างยาก

ดังภาพประกอบที่ 4-5 แสดงแผนภาพการกระจายซึ่งแสดงความสามารถของคู่ลักษณะเด่นที่สามารถแยกสัญญาณไฟฟ้ากล้ามเนื้อของ 3 พยางค์ได้แก่ มา มี มู ภาพประกอบที่ 4-5 (ก) (ค) และ (จ) แสดงแผนภาพการกระจายของลักษณะเด่น 3 คู่ คือ (ก) MAV และ MNF (ค) MAV และ $L-KURT$ และ (จ) MNF และ $L-KURT$ ซึ่งผ่านการนอร์มข้อมูลแล้วก่อนที่จะเข้ากระบวนการฉายลักษณะเด่นสำหรับคนปกติ เช่นเดียวกับแผนภาพการกระจายของผู้ป่วยแสดงดังภาพประกอบที่ 4-5 (ข) (ง) และ (ฉ) อย่างไรก็ตามพบว่าแผนภาพการกระจายยังคลุมเครือไม่สามารถแยกออกเป็นกลุ่มได้อย่างชัดเจน ดังนั้นเพื่อที่จะเพิ่มความถูกต้องในการจำแนกพยางค์ ลักษณะเด่นทั้ง 4 กลุ่มคือ ABF, FBF, SBF, และการรวมลักษณะเด่นจากทุกกลุ่มที่กล่าวมาแล้วข้างต้น (ACF) ถูกนำมาทดสอบโดยใช้เทคนิค SRELM

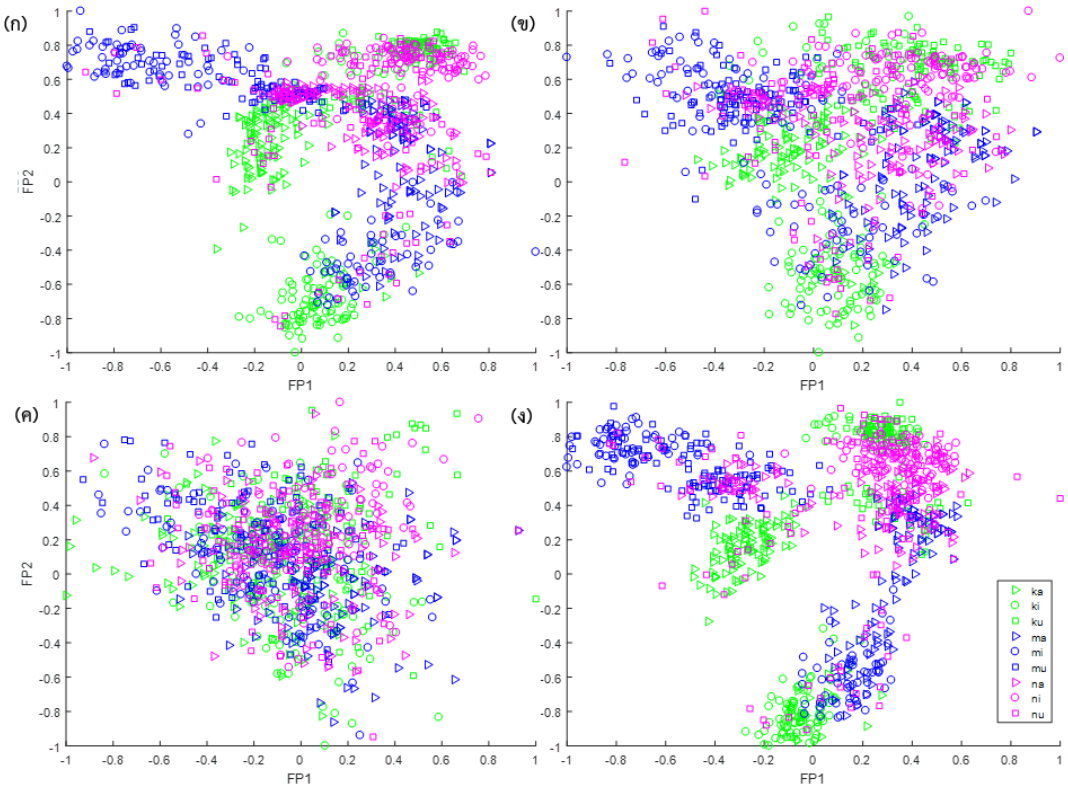
ภาพประกอบที่ 4-6 แสดงแผนภาพการกระจายของลักษณะเด่นสองตัวแรกที่ผ่านการฉายข้อมูลด้วยวิธี SRELM แล้วของคนปกติโดยภาพประกอบที่ 4-6 (ก) ถึง (ง) แสดงแผนภาพการกระจายของลักษณะเด่นสองตัวแรกที่ผ่านการฉายข้อมูลจากกลุ่มลักษณะเด่น ABF, FBF, SBF และ ACF ตามลำดับ ผลการทดลองแสดงให้เห็นว่าลักษณะเด่นหลังการฉายข้อมูลของ ACF ให้ระดับการแบ่งแยกสูงที่สุด โดยสังเกตจากกลุ่มของสี เช่นเดียวกับแผนภาพการกระจายของลักษณะเด่นของผู้ที่มีอาการพูดไม่เป็นความดังภาพประกอบที่ 4-7 อย่างไรก็ตามระดับการแบ่งแยกของ ACF ในผู้ที่มีอาการพูดไม่เป็นความน้อยกว่าในคนปกติ นั่นอาจเป็นเพราะว่าความอ่อนแอของการหดตัวของมัดกล้ามเนื้อขณะออกเสียงในผู้ที่มีอาการพูดไม่เป็นความ นอกจากนี้ลักษณะเด่นกลุ่ม ABF แสดงระดับการแบ่งแยกสูงกว่าลักษณะเด่นกลุ่ม FBF และ SBF

ภาพประกอบที่ 4-8 แสดงค่าความถูกต้องเฉลี่ยในการจำแนกเสียงทั้ง 9 พยางค์ที่ได้จากลักษณะเด่นทั้ง 4 กลุ่มจากอาสาสมัครที่เป็นคนปกติทั้ง 7 คน สำหรับอาสาสมัครแต่ละคนค่าความถูกต้องเฉลี่ยสูงสุดได้จากลักษณะเด่นกลุ่ม ACF เช่นเดียวกลับผลการทดลองของผู้ที่มีอาการพูดไม่เป็นความแสดงดังภาพประกอบที่ 4-9 ผลการทดลองเหล่านี้สอดคล้องกับแผนภาพกระจายของลักษณะเด่นสองตัวแรกที่ผ่านการฉายข้อมูลด้วยวิธี SRELM นอกจากนี้จากผลการทดลองสามารถชี้ให้เห็นว่าระบบที่นำเสนอซึ่งประกอบด้วย (1) ลักษณะเด่นหลายตัวจาก ABF, FBF และ SBF (2) การฉายข้อมูลด้วยเทคนิค SRELM เหมาะสมกับการจำแนกเสียงทั้ง 9 พยางค์ จากสัญญาณไฟฟ้ากล้ามเนื้อ ตารางที่ 4-1 แสดงค่าทางสถิติของค่าความถูกต้องในการจำแนกพบว่าลักษณะเด่นกลุ่ม

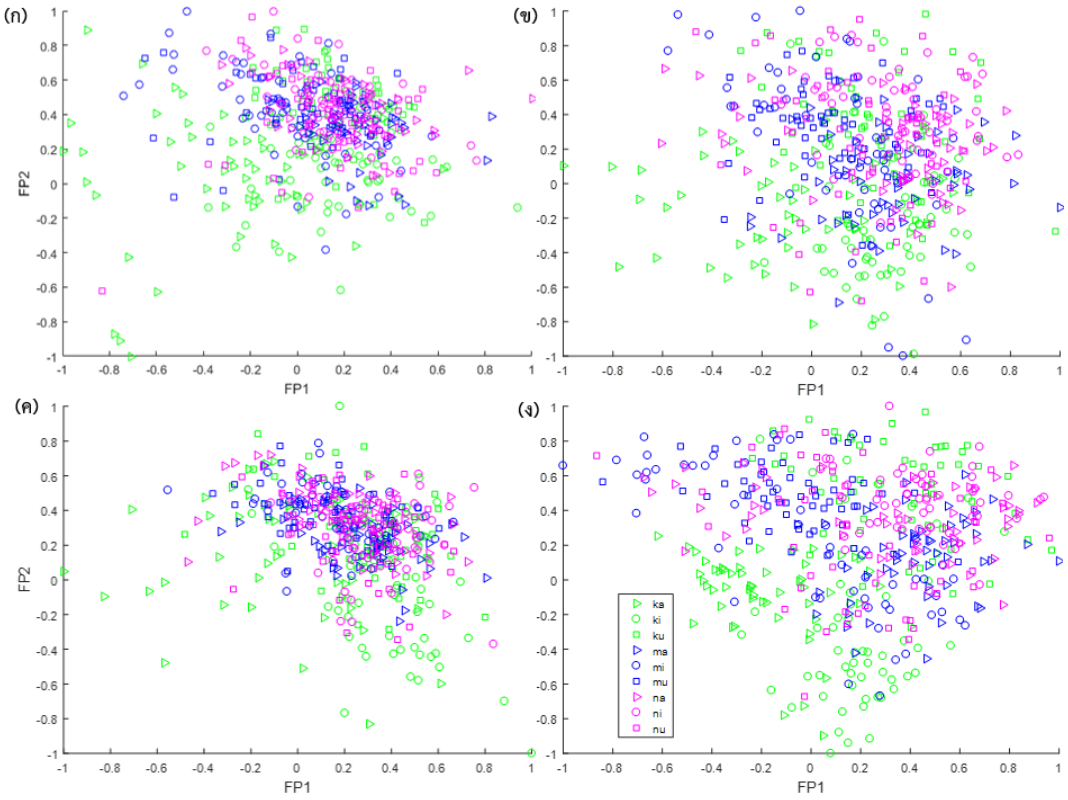
ACF ให้ค่าความถูกต้องเฉลี่ยมากที่สุดคือ 94.5±0.5 และ 89.4±1.2 เปอร์เซ็นต์ ในคนปกติและผู้ที่มีอาการพูดไม่เป็นความ ตามลำดับ นอกจากนี้ค่าความถูกต้องสูงสุดของกลุ่มอาสาสมัครทั้งคนปกติและผู้ที่มีอาการพูดไม่เป็นความสามารถเทียบเคียงกันได้โดยมีค่าอยู่ที่ 96.9 และ 94.1 เปอร์เซ็นต์ ตามลำดับ อย่างไรก็ตามค่าความถูกต้องต่ำสุดของผู้ที่มีอาการพูดไม่เป็นความคนที่ 4 (D4) เท่ากับ 86.6 เปอร์เซ็นต์ ซึ่งมีค่าค่อนข้างต่ำเมื่อเทียบกับคนปกติคนที่ 3 (H3) ที่มีค่าความถูกต้องอยู่ที่ 92.0 เปอร์เซ็นต์ ที่เป็นเช่นนี้อาจมีสาเหตุมาจากความแตกต่างของระดับความรุนแรงของโรค



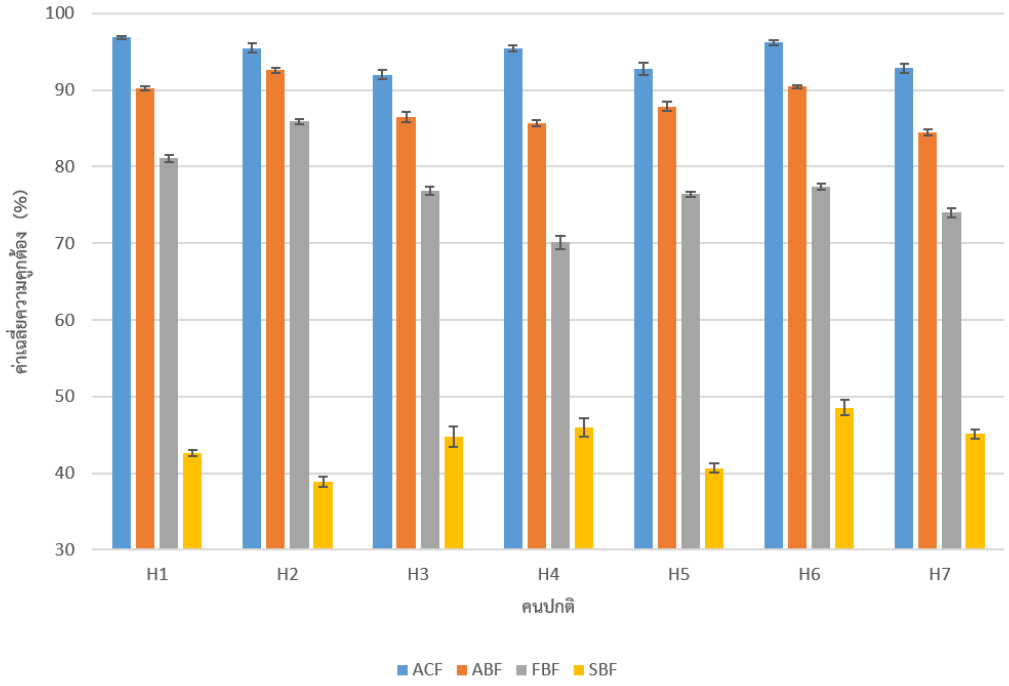
ภาพประกอบที่ 4-5 แผนภาพกระจายของลักษณเด่นที่ผ่านการนอร์มก่อนที่จะฉายข้อมูลโดยสัญลักษณ์สามเหลี่ยมสีดำคือ “มา” วงกลมสีน้ำเงินคือ “มี” และสี่เหลี่ยมสีม่วงคือ “มู” ฝั่งซ้ายมือเป็นของคนปกติและฝั่งขวามือเป็นของผู้ที่มีอาการพูดไม่เป็นความ



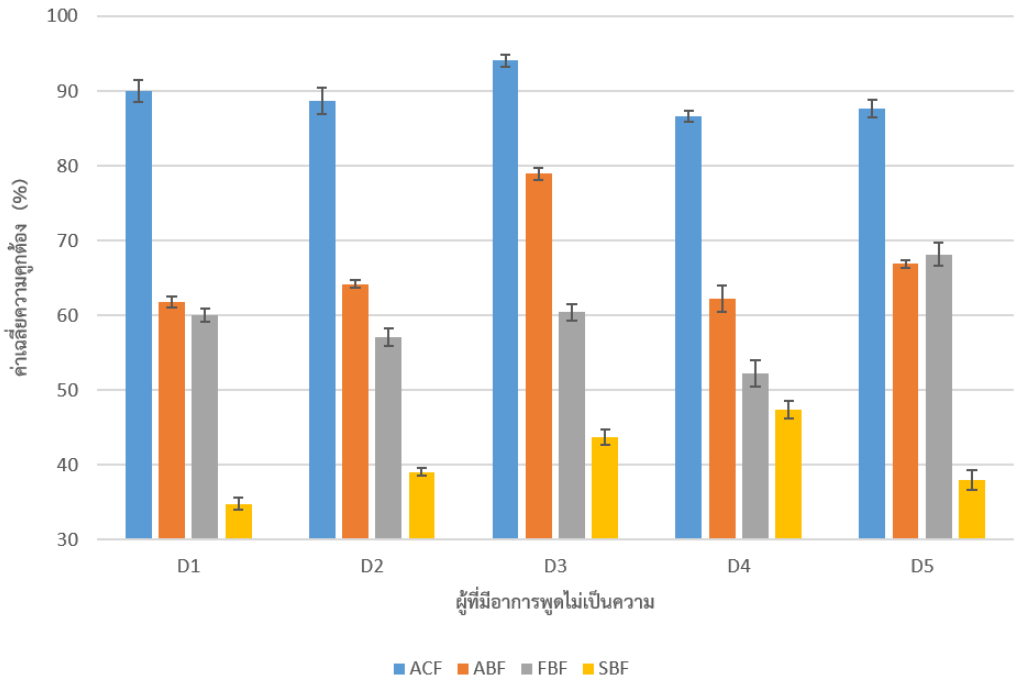
ภาพประกอบที่ 4-6 แผนภาพกระจายของลักษณะเด่นที่ผ่านการนอร์มหลังจากฉายข้อมูลแล้วของคนปกติดังนี้ (ก) ABF; (ข) FBF; (ค) SBF; (ง) ACF



ภาพประกอบที่ 4-7 แผนภาพกระจายของลักษณะเด่นที่ผ่านการนอร์มหลังจากฉายข้อมูลแล้วของผู้ที่มีอาการพูดไม่เป็นความดังนี้ (ก) ABF; (ข) FBF; (ค) SBF; (ง) ACF



ภาพประกอบที่ 4-8 ค่าความถูกต้องเฉลี่ยในการจำแนกเสียงทั้ง 9 พยางค์ของลักษณะเด่นทั้ง 4 กลุ่มจากคนปกติ



ภาพประกอบที่ 4-9 ค่าความถูกต้องเฉลี่ยในการจำแนกเสียงทั้ง 9 พยางค์ของลักษณะเด่นทั้ง 4 กลุ่มจากผู้ที่มีอาการพูดไม่เป็นความ

ตารางที่ 4-1 ค่าทางสถิติของความถูกต้องในการจำแนกจาก 4 กลุ่มลักษณะเด่นเมื่อผ่านการฉายข้อมูลด้วยเทคนิค SRELM

กลุ่มของลักษณะเด่น	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
		คนปกติ		
ABF	88.2	0.4	84.5	92.6
FBF	77.4	0.5	74.0	85.9
SBF	43.8	0.8	38.9	48.6
ACF	94.5	0.5	92.0	96.9
		คนที่มีอาการพูดไม่เป็นความ		
ABF	66.8	0.9	61.8	79.0
FBF	59.6	1.3	52.2	68.1
SBF	40.6	1.0	34.8	47.4
ACF	89.4	1.2	86.6	94.1

4.1.2 ประสิทธิภาพของการสกัดลักษณะเด่นแบบฉายข้อมูลด้วยเทคนิค PCA, LDA และ SRELM

ตารางที่ 4-2 แสดงค่าทางสถิติของค่าความถูกต้องในการจำแนกโดยเปรียบเทียบประสิทธิภาพของวิธีการฉายข้อมูลแบบ SRELM, PCA และ LDA เมื่อใช้กลุ่มลักษณะเด่น ACF ค่าความถูกต้องของการจำแนกเฉลี่ยบ่งชี้ได้ว่าการฉายข้อมูลด้วยเทคนิค SRELM ให้ผลดีกว่า LDA และ PCA โดยค่าความถูกต้องเฉลี่ยของ SRELM เท่ากับ 94.5 เปอร์เซ็นต์ ในคนปกติซึ่งมากกว่า LDA (88.4 เปอร์เซ็นต์) และ PCA (78.5 เปอร์เซ็นต์) เช่นเดียวกับค่าความถูกต้องเฉลี่ยของผู้ที่มีอาการพูดไม่เป็นความพบว่าค่าความถูกต้องของการจำแนกเฉลี่ยเมื่อใช้เทคนิคการฉายข้อมูลแบบ SRELM (89.4 เปอร์เซ็นต์) มีค่าสูงกว่าแบบ LDA (74.0 เปอร์เซ็นต์) และ PCA (60.6 เปอร์เซ็นต์) ผลการทดลองทั้งหมดที่ได้กล่าวไปแล้วนั้นแสดงให้เห็นว่าแนวโน้มของค่าความถูกต้องคล้ายคลึงกับการจำแนกการเคลื่อนไหวของนิ้วมือบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ [45] โดยพบว่าการจับคู่ระหว่าง SRELM และโครงข่ายประสาทเทียมให้ค่าความถูกต้องของการจำแนกดีกว่าการจับคู่แบบอื่น

ตารางที่ 4-2 ค่าทางสถิติของความถูกต้องในการจำแนกจากการฉายข้อมูลด้วยเทคนิค PCA, LDA และ SRELM ของลักษณะเด่นกลุ่ม ACF

ชนิดของการฉายข้อมูล	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
		คนปกติ		
PCA	78.5	1.7	72.1	89.7
LDA	88.4	1.0	81.9	92.4
SRELM	94.5	0.5	92.0	96.9
		คนที่มีอาการพูดไม่เป็นความ		
PCA	60.6	3.0	50.7	73.5
LDA	74.0	1.6	67.7	82.6
SRELM	89.4	1.2	86.6	94.1

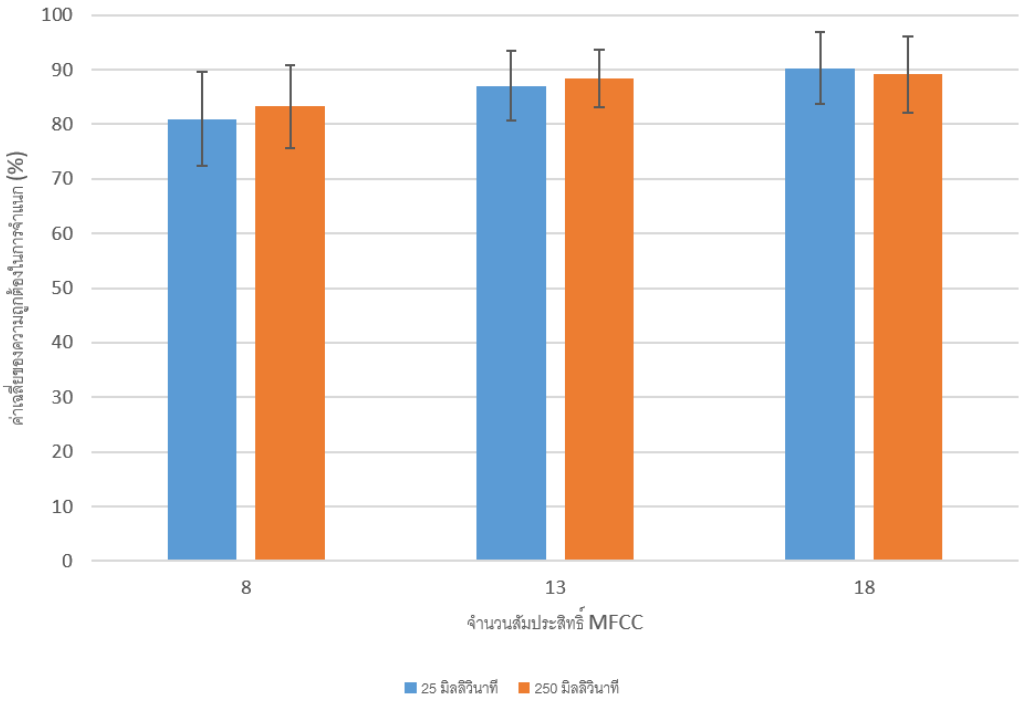
4.2 ประสิทธิภาพของการจำแนกพยางค์ด้วยสัญญาณเสียง

ผลการทดลองถูกแบ่งออกเป็น 2 หัวข้อหลักคือเปรียบเทียบความถูกต้องของการจำแนกพยางค์โดยการแปรผันจำนวนสัมประสิทธิ์ต่อเฟรมของลักษณะเด่น MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน และเปรียบเทียบความถูกต้องของการจำแนกพยางค์เมื่อใช้ลักษณะเด่น 2 กลุ่ม ผลการทดลองแสดงดังนี้

4.2.1 ลักษณะเด่นสัมประสิทธิ์เซปสตรัลที่คำนวณบนแกนความถี่แบบเมล และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน

ผลทดลองของหัวข้อที่ 1 แสดงดังภาพประกอบที่ 4-10 พารามิเตอร์ที่สำคัญและส่งผลต่อประสิทธิภาพในการรู้จำคำพูดคือจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน เมื่อพิจารณาในส่วนของจำนวนสัมประสิทธิ์ต่อเฟรมของลักษณะเด่น MFCC ซึ่งถูกกำหนดไว้ 3 ค่าคือ 8 13 และ 18 ตามลำดับ พบว่าค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์แปรผันตามจำนวนสัมประสิทธิ์ของ MFCC โดยจำนวนสัมประสิทธิ์เท่ากับ 18 ให้ค่าความถูกต้องสูงที่สุด รองลงมาคือจำนวนสัมประสิทธิ์เท่ากับ 13 และจำนวนสัมประสิทธิ์เท่ากับ 8 ให้ค่าความถูกต้องน้อยที่สุด เมื่อพิจารณาในส่วนของขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกัน

พบว่าที่จำนวนสัมประสิทธิ์เท่ากับ 8 และ 13 ขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 250(125) ให้ค่าความถูกต้องในการจำแนกพยางค์สูงกว่าที่ 25(10) และค่าเบี่ยงเบนมาตรฐานต่ำกว่า จากตารางที่ 4-3 ค่าความถูกต้องในการจำแนกพยางค์ที่ 250(125) และจำนวนสัมประสิทธิ์เท่ากับ 8 และ 13 มีค่าเท่ากับ 83.3±7.6 และ 88.4±5.2 เปอร์เซ็นต์ ตามลำดับ เมื่อเทียบกับที่ 25(10) ที่มีค่าความถูกต้องในการจำแนกพยางค์เท่ากับ 81.0±8.6 เปอร์เซ็นต์ และ 87.1±6.4 เปอร์เซ็นต์ ตามลำดับ ในทางกลับกันถ้าจำนวนสัมประสิทธิ์เท่ากับ 18 ขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 25(10) ให้ค่าความถูกต้องในการจำแนกพยางค์สูงกว่าและค่าเบี่ยงเบนมาตรฐานต่ำกว่าโดยมีค่าเท่ากับ 90.3±6.5 และ 89.1±7.0 เปอร์เซ็นต์ ตามลำดับ อย่างไรก็ตามพบว่าเมื่อเพิ่มจำนวนสัมประสิทธิ์ของ MFCC จาก 13 เป็น 18 ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นอย่างไม่มีนัยสำคัญทั้งที่ขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 25(10) และ 250(125)



ภาพประกอบที่ 4-10 ค่าความถูกต้องเฉลี่ยของการจำแนกพยางค์โดยการแปรผันขนาดของเฟรมและจำนวนสัมประสิทธิ์ของ MFCC ของกลุ่มของอาสาสมัครทั้ง 7 คน

ตารางที่ 4-3 เปอร์เซ็นต์ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าความถูกต้องในการจำแนกพยางค์โดยแปรผันขนาดของเฟรมและจำนวนสัมประสิทธิ์ของ MFCC

ขนาดของเฟรม	จำนวนสัมประสิทธิ์ของ MFCC		
	8	13	18
25(10)	81.0±8.6	87.1±6.4	90.3±6.5
250(125)	83.3±7.6	88.4±5.2	89.1±7.0

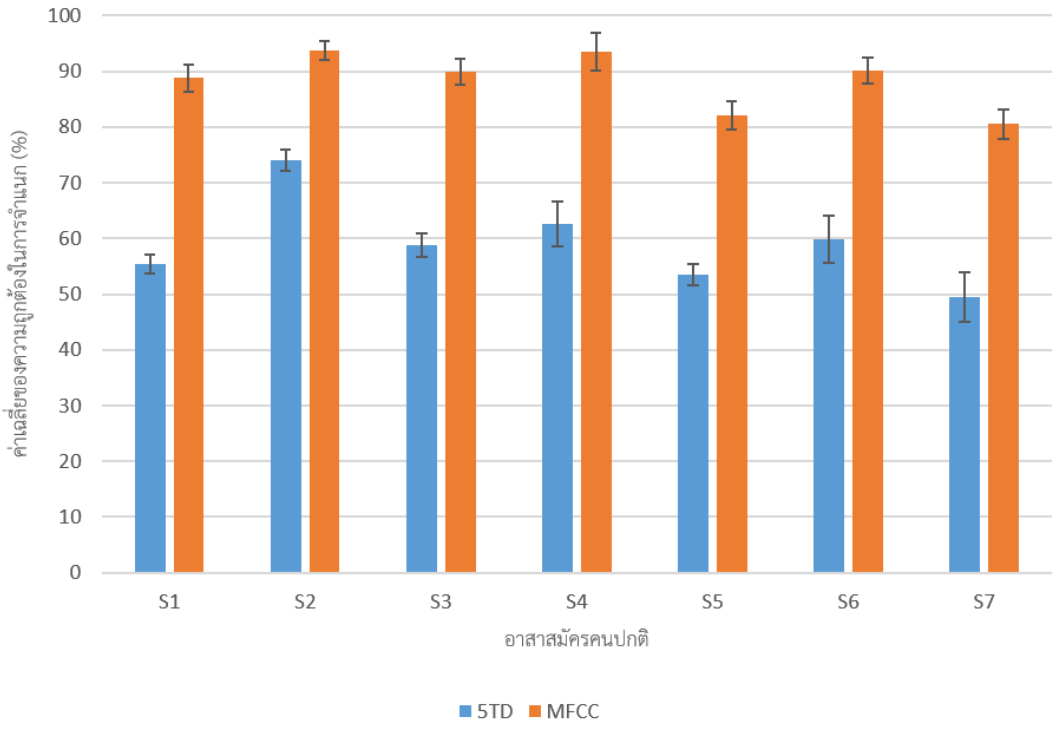
หมายเหตุ 25(10) หมายถึงขนาดของเฟรมเท่ากับ 25 มิลลิวินาที โดยมีส่วนที่คาบเกี่ยวกัน 10 มิลลิวินาที เช่นเดียวกับ 250(125) ซึ่งมีความหมายเหมือนกัน

4.2.2 ลักษณะเด่นในโดเมนเวลา

ภาพประกอบที่ 4-11 แสดงค่าความถูกต้องเฉลี่ยของการจำแนกพยางค์โดยเปรียบเทียบลักษณะเด่นจำนวน 2 กลุ่มของอาสาสมัครทั้ง 7 คนคือลักษณะเด่นในโดเมนเวลาที่นิยมใช้ในการจำแนกประเภทบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อรวมทั้งสิ้น 8 ลักษณะเด่น โดยมีขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 250(125) และลักษณะเด่น MFCC ที่มีจำนวนสัมประสิทธิ์ต่อเฟรมเท่ากับ 13 โดยขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 25(10) พบว่าประสิทธิภาพของการจำแนกพยางค์ด้วยลักษณะเด่น MFCC สูงกว่าลักษณะเด่นในโดเมนเวลาอย่างมีนัยสำคัญสำหรับอาสาสมัครปกติทั้งหมด นั่นแสดงว่าลักษณะเด่น MFCC เหมาะกับสัญญาณเสียง โดยค่าเฉลี่ยความถูกต้องของการจำแนกพยางค์ของลักษณะเด่น MFCC มากกว่าลักษณะเด่นในโดเมนเวลาถึง 29 เปอร์เซ็นต์ และค่าเบี่ยงเบนมาตรฐานลดลงประมาณ 2.7 เปอร์เซ็นต์

4.3 ประสิทธิภาพของการรวมข้อมูล

ประสิทธิภาพของการรวมข้อมูลถูกแบ่งเป็น 2 หัวข้อหลักคือประสิทธิภาพของการรวมข้อมูลจากแหล่งที่มาเดียวกันและประสิทธิภาพของการรวมข้อมูลจากหลายแหล่งที่มา โดยแสดงเป็นค่าทางสถิติของความถูกต้องในการจำแนกพยางค์คือค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานมีรายละเอียดดังนี้



ภาพประกอบที่ 4-11 ค่าความถูกต้องเฉลี่ยของการจำแนกพยางค์โดยเปรียบเทียบลักษณะเด่น MFCC และลักษณะเด่นในโดเมนเวลาของอาสาสมัครทั้ง 7 คน

4.3.1 ประสิทธิภาพของการรวมข้อมูลจากแหล่งที่มาเดียวกัน

ตารางที่ 4-4 แสดงค่าทางสถิติของความถูกต้องในการจำแนกพยางค์ (เปอร์เซ็นต์) เมื่อประยุกต์ใช้การรวมสัญญาณจากแหล่งที่มาเดียวกัน โดยมีช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงที่ใช้ในการคำนวณลักษณะเด่นดังนี้ 1.125 วินาที 1.75 วินาที และ 2.4 วินาที ตามลำดับ กำหนดขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 250(125) จากที่ได้อธิบายวิธีการทดลองไว้ในหัวข้อ 3.6.1 โดยคอลัมน์แรกของตารางแสดงช่องสัญญาณที่ให้ค่าความถูกต้องในการจำแนกพยางค์สูงที่สุดในแต่ละกลุ่มย่อย ยกตัวอย่างเช่น ช่องสัญญาณเดียวที่ดีที่สุดคือช่องสัญญาณที่ 2 หรือ 2 ช่องสัญญาณรวมกันที่ดีที่สุดคือช่องสัญญาณที่ 1 และ 4 เป็นต้น ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์ของคนปกติทั้ง 7 คนเพิ่มขึ้นจาก 44.9 เป็น 77.1 เปอร์เซ็นต์ เมื่อจำนวนของช่องสัญญาณของการรวมสัญญาณไฟฟ้ากล้ามเนื้อเพิ่มขึ้นจาก 1 ช่องสัญญาณเป็น 5 ช่องสัญญาณรวมกัน และค่าเบี่ยงเบนมาตรฐานลดลงจาก 7.6 เป็น 3.5 เปอร์เซ็นต์ ที่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 วินาที เช่นเดียวกับที่ช่วงเวลาของการวิเคราะห์สัญญาณ 1.75 วินาที และ 2.4 วินาที ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์ของคนปกติทั้ง 7 คนมีแนวโน้มเพิ่มขึ้นและค่าเบี่ยงเบนมาตรฐานลดลงเมื่อช่องสัญญาณเพิ่มขึ้นคือ 46.9±7.6 เป็น 78.8±4.3 เปอร์เซ็นต์ และ 47.41±9.0 เป็น 79.5±4.2 เปอร์เซ็นต์ ตามลำดับ ผลการทดลองแสดงให้

เห็นว่าการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อสามารถทำให้ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้น อย่างไรก็ตามส่งผลให้ค่าใช้จ่ายเพิ่มขึ้นเช่นเดียวกันและการทำเวชศาสตร์ฟื้นฟูโดยผู้ป่วยนำไปใช้ยากขึ้นด้วย

นอกจากนี้เมื่อพิจารณาค่าเฉลี่ยความถูกต้องและค่าเบี่ยงเบนมาตรฐานในการจำแนกพยางค์เมื่อช่วงเวลาของการวิเคราะห์สัญญาณเพิ่มขึ้น พบว่าค่าเฉลี่ยความถูกต้องของการจำแนกพยางค์เพิ่มขึ้นอย่างไม่มีนัยสำคัญ ยกตัวอย่างเช่นที่ 1 ช่องสัญญาณ ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นเมื่อช่วงเวลาของสัญญาณเพิ่มขึ้นคือ 44.9, 46.9 และ 47.4 เปอร์เซ็นต์ตามลำดับ ขณะที่ช่องสัญญาณ 2 ช่องรวมกันพบว่าเมื่อช่วงเวลาของการวิเคราะห์สัญญาณเพิ่มขึ้นจาก 1.125 เป็น 1.75 วินาที ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นจาก 66.7 เป็น 67.1 เปอร์เซ็นต์ แต่กลับลดลงเมื่อช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาทีคือ 66.4 เปอร์เซ็นต์ อย่างไรก็ตามค่าที่แสดงมีความแตกต่างกันเพียงเล็กน้อย ดังนั้นอาจจะสรุปได้ว่าการเพิ่มช่วงเวลาในการวิเคราะห์สัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อไม่ได้ส่งผลต่อประสิทธิภาพในการจำแนกพยางค์ซึ่งแตกต่างกับสัญญาณเสียง พบว่าเมื่อช่วงเวลาในการวิเคราะห์สัญญาณของสัญญาณเสียงเพิ่มขึ้น ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นและค่าเบี่ยงเบนมาตรฐานลดลงดังนี้ 83.0±7.6, 87.0±6.6 และ 91.7±6.2 เปอร์เซ็นต์ ตามลำดับ ดังแสดงในแถวสุดท้ายของตาราง

ตารางที่ 4-4 เปอร์เซ็นต์ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์ เมื่อประยุกต์ใช้การรวมสัญญาณแบบแหล่งที่มาเดียวกัน

ช่องสัญญาณ	ช่วงเวลาของการวิเคราะห์สัญญาณ (วินาที)		
	1.125	1.75	2.4
CH2	44.9±7.6	46.9±7.6	47.4±9.0
CH1+CH4	66.7±6.7	67.1±8.0	66.4±7.0
CH1+CH2+CH4	73.0±4.9	73.2±5.2	74.0±4.6
CH1+CH2+CH3+CH4	77.0±3.8	76.6±4.4	77.8±4.1
CH1+CH2+CH3+CH4+CH5	77.1±3.5	78.8±4.3	79.5±4.2
FACO	83.3±7.6	87.0±6.6	91.7±6.2

4.3.2 ประสิทธิภาพของการรวมข้อมูลจากหลายแหล่งที่มา

ตารางที่ 4-5 แสดงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์ (เปอร์เซ็นต์) เมื่อประยุกต์ใช้การรวมข้อมูลจากหลายแหล่งที่มา กำหนดให้ช่วงเวลาในการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงมีค่าเท่ากับ 1.125 วินาที (ประมาณ 47 เปอร์เซ็นต์ของสัญญาณทั้งหมด) วิธีการรวมข้อมูลคือการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากการทดลองในหัวข้อ 4.3.1 มาต่อเรียงกับลักษณะเด่น MFCC ซึ่งมีจำนวนสัมประสิทธ์ต่อเฟรมเท่ากับ 8 13 และ 18 ตามลำดับ ผลการทดลองสามารถแบ่งได้ 2 ประเด็นคือประสิทธิภาพที่ได้จากการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อและการเพิ่มจำนวนสัมประสิทธิ์ MFCC ในกรณีของการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อพบว่าการรวมลักษณะเด่นของสัญญาณเสียงกับสัญญาณไฟฟ้ากล้ามเนื้อสามารถพัฒนาประสิทธิภาพของระบบให้ดีขึ้นอย่างมีนัยสำคัญสำหรับทุกค่าของจำนวนสัมประสิทธิ์ต่อเฟรมของ MFCC ยกตัวอย่างเช่นค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์จำนวน 1 ช่องสัญญาณเรียงตามจำนวนสัมประสิทธิ์ MFCC คือ 89.9±4.0, 94.6±1.9 และ 96.1±2.3 เปอร์เซ็นต์ ตามลำดับ ซึ่งมีค่าเฉลี่ยเพิ่มขึ้นจากการใช้สัญญาณไฟฟ้ากล้ามเนื้อเพียงอย่างเดียวถึง 45.0, 49.8 และ 51.2 เปอร์เซ็นต์ตามลำดับ นอกจากนี้ค่าเบี่ยงเบนมาตรฐานของระบบลดลงประมาณ 3.6, 5.7 และ 5.3 เปอร์เซ็นต์ตามลำดับ นั่นแสดงให้เห็นถึงประสิทธิภาพของความทนทานต่อสัญญาณรบกวนรอบข้าง (robustness)

เช่นเดียวกับการรวมลักษณะเด่นของสัญญาณเสียงกับสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 2, 3, 4, และ 5 ช่องสัญญาณพบว่าค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นอย่างมีนัยสำคัญ โดยมีค่าเฉลี่ยสูงสุดเมื่อจำนวนสัมประสิทธิ์ MFCC เท่ากับ 18 คือ 97.2, 97.4, 97.7 และ 97.0 เปอร์เซ็นต์ ตามลำดับ ดังแสดงในคอลัมน์สุดท้ายของตารางที่ 4-5 นอกจากนี้สิ่งที่น่าสังเกตคือค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์ของการรวมลักษณะเด่นของสัญญาณเสียงกับสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 2 ช่องสัญญาณรวมกันและ 4 ช่องสัญญาณรวมกันมีค่าใกล้เคียงกันคือ 97.2 และ 97.7 เปอร์เซ็นต์ ตามลำดับ ขณะที่จำนวนอิเล็กโทรดลดลงครึ่งนึง ผลที่ตามมาคือสามารถประหยัดค่าใช้จ่ายและง่ายต่อการใช้งาน

จากการทดลองสามารถสรุปได้ว่าการเพิ่มลักษณะเด่นของสัญญาณเสียงทำให้ประสิทธิภาพของระบบดีขึ้น อย่างไรตามพบว่าการเพิ่มค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์ไม่เป็นสัดส่วนกับจำนวนช่องสัญญาณของสัญญาณไฟฟ้าคือ ยกตัวอย่างเช่นเมื่อพิจารณาที่จำนวนสัมประสิทธิ์ MFCC เท่ากับ 18 พบว่าจาก 1 ช่องสัญญาณเป็น 2 ช่องสัญญาณรวมกัน ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้น 20 เปอร์เซ็นต์ จาก 2 เป็น 3 ช่องสัญญาณรวมกัน ค่าความถูกต้อง

ในการจำแนกพยางค์เพิ่มขึ้น 6 เปอร์เซ็นต์ และ จาก 3 เป็น 4 ช่องสัญญาณรวมกัน ค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้น 4 เปอร์เซ็นต์ นอกจากนี้ยังพบว่าประสิทธิภาพของการรวมลักษณะเด่นของสัญญาณเสียงกับสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 4 ช่องสัญญาณรวมกันดีกว่า 5 ช่องสัญญาณรวมกันสำหรับทุกค่าของจำนวนสัมประสิทธิ์ของ MFCC ในเชิงค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์และค่าความเบี่ยงเบนมาตรฐาน

เมื่อพิจารณาเรื่องจำนวนสัมประสิทธิ์ของ MFCC พบว่าค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นเมื่อจำนวนสัมประสิทธิ์เพิ่มขึ้น โดยเฉพาะอย่างยิ่งเมื่อจำนวนสัมประสิทธิ์เปลี่ยนจาก 8 เป็น 13 ค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นอย่างมีนัยสำคัญ ขณะที่จำนวนสัมประสิทธิ์เปลี่ยนจาก 13 เป็น 18 ค่าเฉลี่ยเพิ่มขึ้นอย่างไม่มีนัยสำคัญ ยกตัวอย่างเช่นค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์ของการรวมข้อมูลของสัญญาณเสียงและสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 2 ช่องสัญญาณรวมกันเรียงตามจำนวนสัมประสิทธิ์ของ MFCC คือ 92.8, 96.1 และ 97.2 เปอร์เซ็นต์ตามลำดับ จากที่กล่าวไปแล้วข้างต้นพบว่าการเชื่อมต่อลักษณะเด่นสามารถพัฒนาประสิทธิภาพของการจำแนกพยางค์ไทยได้ ทั้งๆ ที่วิธีการรวมข้อมูลจากหลายแหล่งที่มาค่อนข้างเรียบง่ายคือการเชื่อมต่อลักษณะเด่นแต่ผลที่ได้กลับดีขึ้นเป็นอย่างมาก เนื่องจากสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงเป็นสัญญาณที่ได้จากการออกเสียงและมีช่วงเวลาเดียวกัน ดังนั้นจึงสามารถรวมกันได้ดีในระดับลักษณะเด่น

ตารางที่ 4-5 เปอร์เซ็นต์ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์ เมื่อประยุกต์ใช้การรวมสัญญาณหลายแหล่งที่มา กำหนดช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 วินาที (ประมาณ 47 เปอร์เซ็นต์ของสัญญาณทั้งหมด) โดยนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อมาต่อเรียงกับจำนวนสัมประสิทธิ์ของ MFCC ทั้ง 3 ค่าคือ 8 13 และ 18 ตามลำดับ

ช่องสัญญาณ	**จำนวนสัมประสิทธิ์ของ MFCC**		
	8	**13**	**18**
FACO+ CH2	89.9±4.0	94.6±1.9	96.1±2.3
FACO+ CH1+CH4	92.8±3.0	96.1±2.0	97.2±2.3
FACO+ CH1+CH2+CH4	93.7±2.7	96.5±2.3	97.4±2.0
FACO+CH1+CH2+CH3+CH4	93.9±2.3	97.0±1.5	97.7±1.9
FACO+ CH1+CH2+CH3+CH4+CH5	93.4±2.6	96.5±2.1	97.0±2.1

ตารางที่ 4-6 เปอร์เซ็นต์ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องของการจำแนกพยางค์เมื่อประยุกต์ใช้การรวมสัญญาณหลายแหล่งที่มา กำหนดช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.75 วินาที (ประมาณ 73 เปอร์เซ็นต์ของสัญญาณทั้งหมด) โดยนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อมาต่อเรียงกับจำนวนสัมประสิทธิ์ของ MFCC ทั้ง 3 ค่าคือ 8 13 และ 18 ตามลำดับ

ช่องสัญญาณ	จำนวนสัมประสิทธิ์ของ MFCC		
	8	13	18
FACO+ CH2	90.6±3.5	94.8±2.3	96.9±2.1
FACO+ CH1+CH4	93.3±3.3	96.1±2.2	97.1±2.3
FACO+ CH1+CH2+CH4	94.1±2.0	96.7±1.8	97.6±2.0
FACO+CH1+CH2+CH3+CH4	94.4±2.4	96.5±1.9	97.4±1.7
FACO+ CH1+CH2+CH3+CH4+CH5	93.8±2.3	96.3±1.9	97.4±2.0

ตารางที่ 4-7 เปอร์เซ็นต์ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องของการจำแนกพยางค์เมื่อประยุกต์ใช้การรวมสัญญาณหลายแหล่งที่มา กำหนดช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาที (ประมาณ 100 เปอร์เซ็นต์ของสัญญาณทั้งหมด) โดยนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อมาต่อเรียงกับจำนวนสัมประสิทธิ์ของ MFCC ทั้ง 3 ค่าคือ 8 13 และ 18 ตามลำดับ

ช่องสัญญาณ	จำนวนสัมประสิทธิ์ของ MFCC		
	8	13	18
FACO+ CH2	88.1±4.4	93.5±2.9	95.6±2.9
FACO+ CH1+CH4	92.0±3.5	94.9±2.3	96.2±2.7
FACO+ CH1+CH2+CH4	93.5±2.4	95.8±1.9	96.9±2.1
FACO+CH1+CH2+CH3+CH4	93.8±2.1	95.8±1.7	96.9±2.1
FACO+ CH1+CH2+CH3+CH4+CH5	93.5±2.0	95.8±1.8	96.7±2.1

ตารางที่ 4-6 และ 4-7 แสดงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์เมื่อประยุกต์ใช้การรวมข้อมูลจากหลายแหล่งที่มาเช่นเดียวกับตารางที่ 4-5 โดย

กำหนดให้ช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงเท่ากับ 1.75 (ประมาณ 73 เปอร์เซ็นต์ของสัญญาณทั้งหมด) และ 2.4 วินาที (ประมาณ 100 เปอร์เซ็นต์ของสัญญาณทั้งหมด) ตามลำดับ ผลการทดลองคล้ายคลึงกับสิ่งที่ได้อธิบายไปแล้วตามตารางที่ 4-5 นั่นคือประสิทธิภาพของระบบดีขึ้นเมื่อมีการประยุกต์ใช้วิธีรวมข้อมูลจากหลายแหล่งที่มา ทั้งในแง่ของความถูกต้องในการจำแนกที่เพิ่มขึ้นและค่าเบี่ยงเบนมาตรฐานที่ลดลง เมื่อพิจารณาถึงประเด็นการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อและการเพิ่มจำนวนสัมประสิทธิ์ของ MFCC

ในส่วนของช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง พบว่าประสิทธิภาพของการจำแนกพยางค์เมื่อช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 วินาที และ 1.75 วินาทีมีค่าใกล้เคียง อย่างไรก็ตามประสิทธิภาพในการจำแนกพยางค์เมื่อช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาทีกลับมีค่าลดลง ยกตัวอย่างเช่นค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องในการจำแนกพยางค์จำนวน 2 ช่องสัญญาณรวมกัน เรียงตามจำนวนสัมประสิทธิ์ของ MFCC เมื่อช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงเท่ากับ 1.75 วินาทีคือ 93.3±3.3, 96.1±2.2 และ 97.1±2.3 เปอร์เซ็นต์ ตามลำดับ ขณะที่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาทีคือ 92.0±3.5, 94.9±2.3 และ 96.2±2.7 เปอร์เซ็นต์ ตามลำดับ พบว่าประสิทธิภาพของการจำแนกพยางค์ลดลงสำหรับทุกค่าของสัมประสิทธิ์ของ MFCC ในแง่ของการเพิ่มจำนวนสัมประสิทธิ์ของ MFCC พบว่าประสิทธิภาพในการจำแนกพยางค์เพิ่มขึ้นอย่างมีนัยสำคัญเมื่อจำนวนสัมประสิทธิ์ของ MFCC เปลี่ยนจาก 8 เป็น 13 และมีค่าใกล้เคียงกันเมื่อเมื่อจำนวนสัมประสิทธิ์ของ MFCC เปลี่ยนจาก 13 เป็น 18

ดังนั้นจากตารางที่ 4-5, 4-6 และ 4-7 สามารถสรุปได้เป็น 3 ประเด็นคือการเพิ่มช่วงเวลาของการวิเคราะห์สัญญาณ การเพิ่มจำนวนสัมประสิทธิ์ของ MFCC และการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อดังนี้

ก) ในส่วนของการเพิ่มช่วงเวลาของการวิเคราะห์สัญญาณ ผลการทดลองแสดงให้เห็นว่าประสิทธิภาพของการจำแนกพยางค์ไม่ต่างกันอย่างมีนัยสำคัญ โดยพบว่าค่าเฉลี่ยของความถูกต้องในการจำแนกพยางค์ที่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 และ 1.75 วินาทีมีความใกล้เคียงกัน แต่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 2.4 วินาทีกลับมีค่าลดลง นอกจากนี้ยังพบว่าค่าเบี่ยงเบนมาตรฐานของ 1 ช่องสัญญาณ และ 2 ช่องสัญญาณรวมกันมีค่ามากกว่า 3, 4 และ 5 ช่องสัญญาณรวมกันสำหรับทุกช่วงเวลาของการวิเคราะห์สัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง

ข) ในส่วนของการเพิ่มจำนวนสัมประสิทธิ์ของ MFCC พบว่าประสิทธิภาพของการจำแนกพยางค์เพิ่มขึ้นอย่างมีนัยสำคัญเมื่อจำนวนสัมประสิทธิ์เปลี่ยนจาก 8 เป็น 13 อย่างไรก็ตามหากจำนวนสัมประสิทธิ์เปลี่ยนจาก 13 เป็น 18 พบว่าประสิทธิภาพของการจำแนกพยางค์มีค่าใกล้เคียงกันสำหรับทุกช่วงเวลาของการวิเคราะห์สัญญาณแสดงดังภาพประกอบที่ 4-12

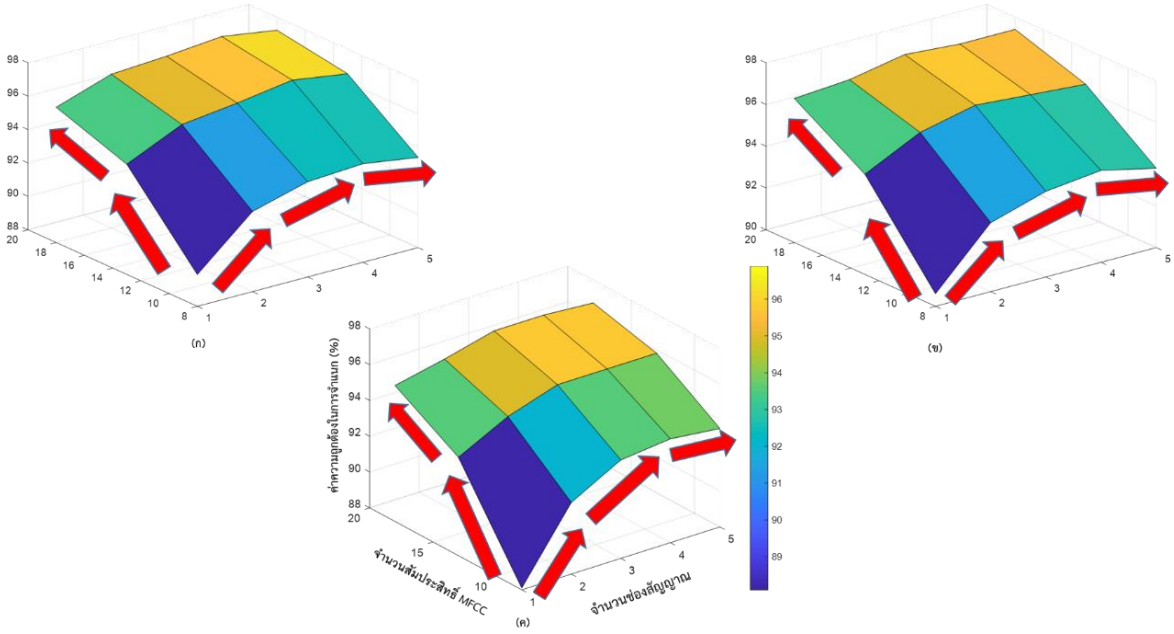
ค) ในกรณีของการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ จากผลการทดลองในหัวข้อ 4.3.1 พบว่าการเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อสามารถเพิ่มประสิทธิภาพของการจำแนกพยางค์อย่างมีนัยสำคัญ เช่นเดียวกับการรวมข้อมูลจากหลายแหล่งที่มาทำให้ประสิทธิภาพดีขึ้น โดยเฉพาะอย่างยิ่งเมื่อจำนวนของช่องสัญญาณเพิ่มจาก 1 ช่องสัญญาณเป็น 2 ช่องสัญญาณรวมกันแสดงดังภาพประกอบที่ 4-12 สำหรับทุกช่วงเวลาของการวิเคราะห์สัญญาณ อย่างไรก็ตามพบว่าการรวมข้อมูลโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจำนวน 4 ช่องสัญญาณรวมกันกับสัญญาณเสียงให้ประสิทธิภาพดีกว่าการใช้สัญญาณไฟฟ้ากล้ามเนื้อทุกช่องสัญญาณ ทั้งนี้อาจจะเกิดจากปัญหาของมิติข้อมูล (curse of dimension) ที่มากเกินไป

นอกจากนี้จากตารางที่ 4-5 4-6 และ 4-7 เมื่อพิจารณาการรวมข้อมูลระหว่างช่องสัญญาณไฟฟ้ากล้ามเนื้อ 2 ช่องสัญญาณรวมกันและสัญญาณเสียง พบว่าที่ช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 วินาที และจำนวนสัมประสิทธิ์ของ MFCC เท่ากับ 18 ให้ค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์ดีที่สุดคือ 97.2 เปอร์เซ็นต์ ดังนั้นผู้วิจัยนำค่าดังกล่าวซึ่งหมายถึงช่วงเวลาของการวิเคราะห์สัญญาณและจำนวนสัมประสิทธิ์ของ MFCC ไปพัฒนาต่อในหัวข้อ 4.4 เรื่องประสิทธิภาพของการรู้จำพยางค์แบบไม่ขึ้นกับบุคคล

4.4 ประสิทธิภาพของการรู้จำพยางค์แบบไม่ขึ้นกับบุคคล

วิธีการรู้จำพยางค์แบบไม่ขึ้นกับบุคคลเริ่มจากเลือกพารามิเตอร์ที่ทำให้ค่าความถูกต้องเฉลี่ยมากที่สุดจากหัวข้อ 4.3.2 ประกอบด้วยช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ 1.125 วินาที จำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อเท่ากับ 2 ช่องสัญญาณรวมกัน และจำนวนสัมประสิทธิ์ของ MFCC เท่ากับ 18 ดังนั้นขนาดของเวกเตอร์ลักษณะเด่นของการรวมข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงเท่ากับ 34 (16 ลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อ + 18 ลักษณะเด่นของสัญญาณเสียง) ผลจากการทดลองพบว่าค่าเฉลี่ยความถูกต้องในการจำแนกพยางค์ค่อนข้างต่ำโดยมีค่าอยู่ในช่วง 14.7 และ 36.4 เปอร์เซ็นต์ ผลการทดลองแสดงดังตารางที่ 4-8 ทั้งนี้เนื่องจากค่าน้อยสุดและมากสุดของเวกเตอร์ลักษณะเด่นของอาสาสมัครแต่ละคนค่อนข้าง

แตกต่างกัน แสดงดังภาพประกอบที่ 4-13 พบว่าแผนภาพแบบกล่องของลักษณะเด่นที่ผ่านการฉายข้อมูลแล้วของการรู้จำพยางค์แบบไม่ขึ้นกับบุคคล ข้อมูลที่ใช้สอนและข้อมูลทดสอบมีความแตกต่างกันค่อนข้างมากเช่นในคอลัมน์ที่ 8, 9 และ 11 ซึ่งแตกต่างกับแผนภาพแบบกล่องของการรู้จำพยางค์แบบขึ้นกับบุคคลซึ่งมีความใกล้เคียงกันแสดงดังภาพประกอบที่ 4-14 ส่งผลให้ประสิทธิภาพในการจำแนกพยางค์ลดลง



ภาพประกอบที่ 4-12 ค่าความถูกต้องเฉลี่ยในการจำแนกแปรผันตามจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อและจำนวนสัมประสิทธิ์ของ MFCC เมื่อช่วงเวลาของการวิเคราะห์สัญญาณเท่ากับ (ก) 1.125 วินาที (ข) 1.75 วินาที (ค) 2.4 วินาที

ตารางที่ 4-8 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องของการจำแนกพยางค์ (เปอร์เซ็นต์)

กลุ่ม	ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องของการจำแนกพยางค์ (เปอร์เซ็นต์)
1	25.1±1.8
2	14.7±2.3
3	23.2±1.5
4	29.2±3.1
5	31.0±1.3

กลุ่ม	ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถูกต้องของการจำแนกพยางค์ (เปอร์เซ็นต์)
6	36.4±3.1
7	27.7±1.6
8	32.6±3.3

หมายเหตุ

กลุ่มที่ 1 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 2 3 4 และ 5 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 6 และ 7

กลุ่มที่ 2 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 2 3 4 และ 7 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 5 และ 6

กลุ่มที่ 3 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 3 4 5 และ 6 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 2 และ 7

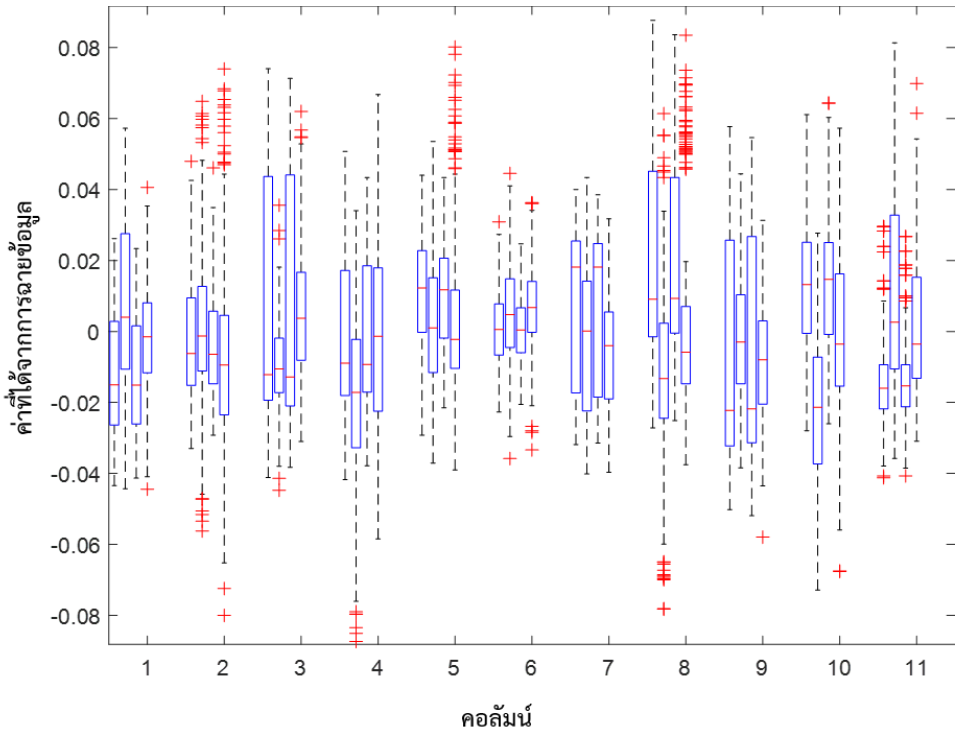
กลุ่มที่ 4 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 3 4 5 และ 7 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 2 และ 6

กลุ่มที่ 5 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 4 5 6 และ 7 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 2 และ 3

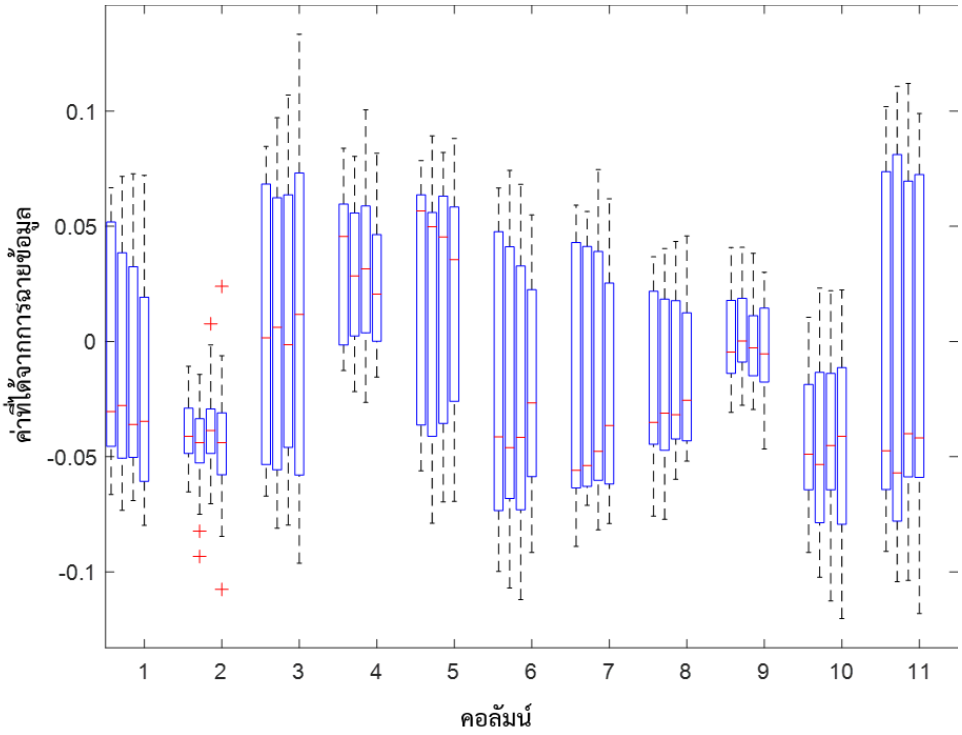
กลุ่มที่ 6 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 2 3 4 5 และ 6 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 และ 7

กลุ่มที่ 7 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 2 3 4 5 และ 7 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 และ 6

กลุ่มที่ 8 หมายถึงข้อมูลสอนประกอบด้วยข้อมูลของอาสาสมัครคนที่ 3 4 5 6 และ 7 และข้อมูลทดสอบประกอบด้วยข้อมูลของอาสาสมัครคนที่ 1 และ 2



ภาพประกอบที่ 4-13 แผนภาพแบบกล่องของลักษณะเด่นที่ผ่านการฉายข้อมูลแล้วของการรู้จำพยางค์แบบไม่ขึ้นกับบุคคลประกอบด้วยข้อมูลที่ใช้สอนและข้อมูลทดสอบ



ภาพประกอบที่ 4-14 แผนภาพแบบกล่องของลักษณะเด่นที่ผ่านการฉายข้อมูลแล้วของการรู้จำพยางค์แบบขึ้นกับบุคคลประกอบด้วยข้อมูลที่ใช้สอนและข้อมูลทดสอบ

บทที่ 5

สรุปผลการวิจัย ปัญหาและข้อเสนอแนะ

บทนี้กล่าวถึงการสรุปผลการวิจัย ปัญหาและข้อเสนอแนะของงานวิจัย โดยงานวิจัยนี้ได้ศึกษาการจำแนกพยางค์ไทยที่ใช้ในการฟื้นฟูอาการพูดไม่เป็นความบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อ สัญญาณเสียง และการรวมข้อมูลจากสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง

5.1 สรุปผลการวิจัย

งานวิจัยนี้เน้นศึกษาระบบจำแนกพยางค์ไทยจำนวน 12 พยางค์ซึ่งใช้สำหรับฟื้นฟูสมรรถนะในการออกเสียงเบื้องต้นในผู้ที่มีอาการพูดไม่เป็นความ อันเนื่องมาจากข้อจำกัดบางประการของทั้งสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง งานวิจัยจึงถูกแบ่งออกเป็น 3 ส่วนแยกตามแหล่งที่มาของข้อมูลคือ สัญญาณไฟฟ้ากล้ามเนื้อบริเวณใบหน้าและบางส่วนของลำคอ สัญญาณเสียง และการรวมข้อมูลจากสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียง มีรายละเอียดสรุปได้ดังต่อไปนี้

5.1.1 สัญญาณไฟฟ้ากล้ามเนื้อ

ในส่วนของการจำแนกพยางค์ไทยบนพื้นฐานของสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 5 ช่องสัญญาณมุ่งเน้นศึกษาคุณลักษณะของสัญญาณไฟฟ้ากล้ามเนื้อในการออกเสียงภาษาไทยจำนวน 9 พยางค์ (อันเนื่องมาจากข้อจำกัดด้านข้อมูลของผู้ที่มีอาการพูดไม่เป็นความ) ทั้งของอาสาสมัครคนปกติและผู้ที่มีอาการพูดไม่เป็นความ แยกตามประเภทของลักษณะเด่น 3 กลุ่มคือขนาดของสัญญาณ ความถี่ของสัญญาณ และค่าทางสถิติของการกระจายข้อมูล พบว่าการใช้ลักษณะเด่นทั้ง 3 กลุ่มร่วมกันก่อให้เกิดประสิทธิภาพสูงสุดทั้งในอาสาสมัครคนปกติและผู้ที่มีอาการพูดไม่เป็นความ โดยค่าเฉลี่ยความถูกต้องของการจำแนกพยางค์ของอาสาสมัครคนปกติและผู้ที่มีอาการพูดไม่เป็นความคือ 94.5 และ 89.4 เปอร์เซ็นต์ ตามลำดับ อย่างไรก็ตามเมื่อพิจารณาลักษณะเด่นแต่ละกลุ่มพบว่าลักษณะเด่นกลุ่มขนาดและความถี่ของสัญญาณส่งผลต่อประสิทธิภาพในการจำแนกพยางค์ดีกว่าอีกกลุ่มหนึ่ง ในส่วนของวิธีสกัดลักษณะเด่นแบบฉายข้อมูล 3 แบบคือ PCA, LDA และ SRELM พบว่าการฉายข้อมูลด้วยเทคนิค SRELM ให้ประสิทธิภาพสูงสุดเมื่อเทียบกับ PCA และ LDA

5.1.2 สัญญาณเสียง

ในส่วนของการจำแนกพยางค์ไทยบนพื้นฐานของสัญญาณเสียงมุ่งเน้นศึกษาลักษณะเด่น MFCC โดยพิจารณาในเรื่องของจำนวนสัมประสิทธิ์ของ MFCC และขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเพื่อเปรียบเทียบประสิทธิภาพในการจำแนกพยางค์ นอกจากนี้ยังมีการใช้ลักษณะเด่นในโดเมนเวลาซึ่งนิยมใช้กับสัญญาณไฟฟ้ากล้ามเนื้อมาเปรียบเทียบประสิทธิภาพในการ

จำแนกพยางค์กับลักษณะเด่น MFCC อีกด้วย ผลการทดลองพบว่าจำนวนสัมประสิทธิ์ของ MFCC เท่ากับ 18 ให้ค่าความถูกต้องมากที่สุดทั้งขนาดของเฟรมและขนาดของเฟรมที่คาบเกี่ยวกันเท่ากับ 25(10) และลักษณะเด่น MFCC เหมาะที่จะใช้กับสัญญาณเสียงมากกว่าลักษณะเด่นในโดเมนเวลา โดยมีประสิทธิภาพในการจำแนกสูงกว่าอย่างมีนัยสำคัญ นอกจากนี้ยังได้เปรียบเทียบประสิทธิภาพในการจำแนกพยางค์ของสัญญาณเสียงที่ไม่ผ่านการกรองและผ่านการกรองด้วยวงจรกรองแถบความถี่ต่ำ พบว่าค่าเฉลี่ยความถูกต้องของสัญญาณเสียงที่ผ่านการกรองสูงกว่าสัญญาณเสียงที่ไม่ผ่านการกรองเล็กน้อย

5.1.3 การรวมข้อมูล

การรวมข้อมูลแบ่งได้เป็น 2 แบบคือการรวมข้อมูลจากแหล่งที่มาเดียวกันและจากหลายแหล่งที่มา โดยมีพารามิเตอร์ที่สำคัญคือช่วงเวลาของการวิเคราะห์สัญญาณ จำนวนสัมประสิทธิ์ของ MFCC และจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ สรุปผลการวิจัยได้ดังนี้

ก) การรวมข้อมูลจากแหล่งที่มาเดียวกันหรือกล่าวอีกนัยหนึ่งเป็นการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อในแต่ละช่องสัญญาณมาต่อเรียงกัน โดยใช้ลักษณะเด่นในโดเมนเวลาจำนวน 5 ค่าคือ MAV, WL, ZC, SSC และ AR อันดับที่ 4 ในการทดลองนี้จำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อและช่วงเวลาของการวิเคราะห์สัญญาณถูกพิจารณา โดยประเมินประสิทธิภาพในการจำแนกพยางค์ ในกรณีของจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อประกอบด้วย 1 ช่องสัญญาณ 2, 3, 4 และ 5 ช่องสัญญาณรวมกัน จากผลการทดลองพบว่าค่าความถูกต้องในการจำแนกพยางค์เพิ่มขึ้น เมื่อจำนวนช่องของสัญญาณไฟฟ้ากล้ามเนื้อเพิ่มขึ้น ในทางกลับกันค่าความเบี่ยงเบนมาตรฐานจะลดลง อย่างไรก็ตามการใช้ช่องสัญญาณจำนวนมากไม่เหมาะจะนำไปใช้ในทางปฏิบัติซึ่งเป็นข้อด้อยของการนำสัญญาณไฟฟ้ากล้ามเนื้อไปประยุกต์ใช้ ในส่วนของช่วงเวลาของการวิเคราะห์สัญญาณ พบว่าการเพิ่มช่วงเวลาของการวิเคราะห์สัญญาณไม่ได้ทำให้ประสิทธิภาพของการจำแนกพยางค์เพิ่มขึ้นอย่างไม่มีนัยสำคัญ

ข) การรวมข้อมูลจากหลายแหล่งที่มาหรือกล่าวอีกนัยหนึ่งเป็นการนำลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงมาต่อเรียงกัน โดยพิจารณาทั้งในเรื่องของช่วงเวลาของการวิเคราะห์สัญญาณ จำนวนสัมประสิทธิ์ของ MFCC และจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อ จากผลการทดลองสรุปได้เป็น 3 ประเด็นคือ ประเด็นที่ 1 การเพิ่มช่วงเวลาในการวิเคราะห์สัญญาณไม่ได้

ทำให้ความถูกต้องในการจำแนกพยางค์เพิ่มขึ้นอย่างมีนัยสำคัญ ประเด็นที่ 2 การเพิ่มจำนวนสัมประสิทธิ์ของ MFCC พบว่าถ้าจำนวนสัมประสิทธิ์เพิ่มจาก 8 เป็น 13 ค่าความถูกต้องเพิ่มขึ้นอย่างมีนัยสำคัญ ประเด็นที่ 3 การเพิ่มจำนวนช่องสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อส่งผลต่อการเพิ่มความถูกต้องในการจำแนกพยางค์อย่างมีนัยสำคัญเมื่อเพิ่มจาก 1 ช่องสัญญาณและ 2 ช่องสัญญาณรวมกัน

5.2 ปัญหา

5.2.1 ข้อมูลของอาสาสมัครคนปกติค่อนข้างหลากหลายเนื่องจากวิธีออกเสียงโดยธรรมชาติของแต่ละคนแตกต่างกันตามความเคยชิน ทำให้หาวิธีนอร์มข้อมูลค่อนข้างยาก ส่งผลต่อระบบรู้จำพยางค์แบบไม่ขึ้นกับบุคคล เมื่อนำข้อมูลมารวมกันพบว่าประสิทธิภาพในการรู้จำพยางค์ค่อนข้างต่ำ

5.2.2 การรวมข้อมูลในระดับลักษณะเด่นมีข้อจำกัดในเรื่องของมิติของข้อมูลที่จะนำมารวมกันต้องเท่ากัน ดังนั้นหากข้อมูลที่นำมารวมกันมีมิติแตกต่างกันไม่สามารถใช้วิธีนี้ได้

5.2.3 ระบบการได้ข้อมูลยังต้องแยกปลายทางการจัดเก็บเนื่องจากการจัดเก็บสัญญาณไฟฟ้ากล้ามเนื้อต้องใช้โปรแกรมที่มาพร้อมอุปกรณ์ในการจัดเก็บ นอกจากนี้ความถี่ของการสุ่มสัญญาณของสัญญาณไฟฟ้ากล้ามเนื้อและสัญญาณเสียงยังแตกต่างกันอีกด้วย ทำให้ยากต่อการนำไปประยุกต์ใช้จริง

5.3 ข้อเสนอแนะ

5.3.1 ค้นหาวิธีการนอร์มแบบใหม่ที่สามารถลดความหลากหลายของข้อมูลได้

5.3.2 ศึกษาการรวมข้อมูลในระดับอื่นๆ เช่นระดับคะแนนหรือระดับการตัดสินใจ รวมทั้งพัฒนาระบบให้สามารถปรับปรุงตัวเองได้ว่าจะเลือกใช้ข้อมูลเดียวหรือการรวมข้อมูลในกรณีที่ประสิทธิภาพของการใช้ข้อมูลเดียวดีกว่า

5.3.3 พัฒนาระบบการได้ข้อมูลให้ทันสมัยกับสถานการณ์การใช้งานในปัจจุบัน เช่น ทำเป็นอุปกรณ์ที่สามารถพกพาได้

5.3.4 อาจนำกระบวนการที่นำเสนอไปประยุกต์ใช้ในการวัดระดับความรุนแรงของอาการพูดไม่เป็นความได้ หรือสามารถนำไปใช้เป็นเครื่องมือสื่อสารสำหรับให้ผู้ป่วยใช้ชีวิตได้ง่ายขึ้น

บรรณานุกรม

[1] B. H. Dobkin, “Rehabilitation after stroke,” *N. Engl. J. Med.*, vol. 352, no. 16, pp. 1677–1684, 2005.

[2] N. Venketasubramanian, B. W. Yoon, J. Pandian, and J. C. Navarro, “Stroke epidemiology in South, East, and South-East Asia: a review,” *J. stroke*, vol. 19, no. 3, pp. 286–294, 2017.

[3] A. B. Kain, J.-P. Hosom, X. Niu, J. P. Van Santen, M. Fried-Oken, and J. Staehely, “Improving the intelligibility of dysarthric speech,” *Speech Commun.*, vol. 49, no. 9, pp. 743–759, 2007.

[4] C. H. Adler and J. E. Ahlskog, "Parkinson's disease and movement disorders. Diagnosis and treatment guidelines for the practicing physician," *Humana Press*, pp. 35-53, 2000.

[5] Mayo foundation for medical education and research (2005, March 10).12 cranial nerves [Online].

Available: https://12cranialnerves.wordpress.com/cranial-nerve-7-facial-nerve/

[6] R. Palmer and P. Enderby, “Methods of speech therapy treatment for stable dysarthria: A review,” *Adv. Speech Lang. Pathol.*, vol. 9, no. 2, pp. 140–153, 2007.

[7] จันทร์ชัย เจรียงประเสริฐ, "การฝึกพูดสำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีปัญหาด้านการพูด," *วิชัยยุทธจุลสาร*, ฉบับที่ 32, หน้าที่ 35-43, 2548.

[8] M. J. Kim, Y. Kim and H. Kim, "Automatic Intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model," *IEEE/ACM Trans. Audio, Speech and Language Process.*, vol. 23, no. 4, April 2015.

[9] P. Kayasith and T. Theeramunkong, “Speech confusion index (Ø): A confusion-based speech quality indicator and recognition rate prediction for dysarthria,” *Comput. Math. Appl.*, vol. 58, no. 8, pp. 1534–1549, 2009.

[10] M. B. Mustafa, S. S. Salim, N. Mohamed, B. Al-Qatab and C. E. Siong, "Severity-based adaptation with limited data for ASR to aid dysarthric speakers," *Plos one*, vol. 9, no. 1, pp. 1-11, January 2014.

[11] A. B. Dobrucki, P. Pruchnicki, P. Plaskota, P. Staroniewicz, S. Brachmanski, and M. Walczynski, "Silent speech recognition by surface electromyography," *New Trends and Developments in Metrology*, pp. 81-97, 2016.

[12] L. Maier-Hein, F. Metze, T. Schultz, and A. Waibel, "Session independent non-audible speech recognition using surface electromyography," *in Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2005, pp. 331–336.

[13] X. Xi, M. Tang, and Z. Luo, "Feature-level fusion of surface electromyography for activity monitoring," *Sensors*, vol. 18, no. 2, pp. 1-14, 2018.

[14] B. Cesqui, P. Tropea, S. Micera, and H. I. Krebs, "EMG-based pattern recognition approach in post stroke robot-aided rehabilitation: a feasibility study," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, pp. 1-15, 2013.

[15] D. Borms, I. Ackerman, P. Smets, G. Van den Berge, and A. M. Cools, "Biceps disorder rehabilitation for the athlete: a continuum of moderate to high-load exercises," *Am. J. Sports Med.*, vol. 45, no. 3, pp. 642–650, 2017.

[16] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan, and T. Chang, "Surface electromyography signal processing and classification techniques," *Sensors*, vol. 13, no. 9, pp. 431–466, 2013.

[17] E. Lopez-Larraz, O. M. Mozos, J. M. Antelis, and J. Minguez, "Syllable based speech recognition using EMG," *in Proc. of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2010, pp. 4699–4702.

[18] A. D. Chan, K. B. Englehart, B. Hudgins, and D. F. Lovely, "Multiexpert automatic speech recognition using acoustic and myoelectric signals," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 676–685, 2006.

[19] 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, 2007.

[20] P. Tammarote, S. Chatpun, P. Phukpattaranont, and D. Vongjandaeng, “The optimum myography feature for oral muscle movements,” in Proc. of the 6th Biomedical Engineering International Conference (BMEiCON). IEEE, Thailand, 2013, pp. 139–143.

[21] พรทิพา ทองสว่าง, “ระบบเสียง”in ภาษาและภาษาศาสตร์, พิมพ์ครั้งที่ 2 ดียู ศรีนราวัฒน์ บรรณาธิการ. กรุงเทพ: สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์, หน้า 67-85, 2559.

[22] T. Tsuji, N. Bu, J. Arita and M. Ohga, "A speech synthesizer using facial EMG signals," *Int. J. Comput. Intell. and Appl*., vol.7, no. 1, pp. 1-15, 2008.

[23] B. G. Lapatki, D. F. Stegeman, and I. E. Jonas, “A surface EMG electrode for the simultaneous observation of multiple facial muscles,” *J. Neurosci. Methods*, vol. 123, no. 2, pp. 117–128, 2003.

[24] N. P. Schumann, K. Bongers, O. Guntinas-Lichius, and H. C. Scholle, “Facial muscle activation patterns in healthy male humans: A multichannel surface EMG study,” *J. Neurosci. Methods*, vol. 187, no. 1, pp. 120–128, 2010.

[25] C. E. Stepp, “Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation,” *J. Speech, Lang. Hear. Res.*, vol. 55, pp. 1232–1246, 2012.

[26] N. Sae Jong and P. Phukpattaranont, “A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A thai syllable study,” *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 234–245, 2019.

[27] K. S. Lee, "EMG-based speech recognition using hidden Markov models with global control variables," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 930-940, March 2008.

[28] A. Lumini and L. Nanni, “Overview of the combination of biometric matchers,” *Inf. Fusion*, vol. 33, pp. 71–85, 2017.

[29] Y. Deng, R. Patel, J. T. Heaton, G. Colby, L. D. Gilmore, J. Cabrera, S. H. Roy, C.J.D. Luca and G.S. Meltzner, "Disordered speech recognition using acoustic and sEMG signals," *in Proc. of the 10th Annual Conference of the International Speech Communication Association (ISCA)*, Interspeech, United Kingdom, 2009, pp. 644-647.

[30] P. Boonyathitisuk, "Articulatory characteristics of kindergarten children aged three to four years eleven months in Bangkok," *M.S. thesis*, Dept. Communication Disorders, Mahidol Univ., Bangkok, Thailand, 1982.

[31] T. Pothirat, " Characterization and analysis of neck and face surface electromyography for speech rehabilitation training," *M.S. thesis*, Dept. Sci. in Biomed. Eng., Prince of Songkla Univ., Songkhla, Thailand, 2014.

[32] A. Phinyomark, P. Phukpattaranont and C. Limsakul, "The usefulness of wavelet transform to reduce noise in the SEMG signal" *in EMG Methods for Evaluating Muscle and Nerve Function*, M. Schwartz, Ed. Rijeka: InTech, pp. 107-132, 2012.

[33] Q. Zhou, N. Jiang, K. Englehart, and B. Hudgins, “Improved Phoneme-based myoelectric speech recognition,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2016-2023, March 2009.

[34] S. Thongpanja, A. Phinyomark, F. Quaine, Y. Laurillau, C. Limsakul, and P. Phukpattaranont, “Probability density functions of stationary surface EMG signals in noisy environments,” *IEEE Trans. Instrum. Meas.*, vol. 65, pp. 1547-1557, 2016.

[35] Q. Ai, Y. Zhang, W. Qi, Q. Liu, and K. Chen, “Research on lower limb motion recognition based on fusion of semg and accelerometer signals,” *Symmetry*, vol. 9, no. 8, pp. 147, 2017.

[36] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, “Classification of finger movements for the dexterous hand prosthesis control with surface electromyography,” *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 608–618, 2013.

[37] F. Liu, J. Zhang and N. Pai (2012). Automatic sign language detection [Online].

Available: https://sites.google.com/site/autosignlan/algorithms-used/mfcc

[38] F. Sohrab, “Railway vehicle detection from audio recordings using one-class classification M.S. thesis, Dept. Electron. Eng., Sabanci Univ., Istanbul, Turkey, 2016.

[39] A. Sankarasubramanian and K. Srinivasan, “Investigation and comparison of sampling properties of L-moments and conventional moments,” *J. Hydrol*, vol. 218, pp.13–34, 1999.

[40] Z. M. Hira and D. F. Gillies," A review of feature selection and feature extraction methods applied on microarray data," *Adv Bioinformatics*, pp. 1-13, 2015

[41] K. Anam and A. Al-Jumaily, “A novel extreme learning machine for dimensionality reduction on finger movement classification using sEMG,” ," *in Proc. of the 7th Annual International IEEE EMBS Conference on Neural Engineering*, France, 2015, pp. 824-7.

[42] A. K. Jain, R. 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.

[43] จาตุรงค์ ตันติบัณฑิต, “การรู้จำรูปแบบ,” สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2555, หน้า 79-81.

[44] N. Sae Jong, M. Kiatweerasakul, P. Phukpattaranont, “Channel Reduction in Speech Recognition System based on Surface Electromyography,” *in Proc. of the* 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE, 2018, Chiang Rai, Thailand, pp. 184-187.

[45] P. Phukpattaranont, S. Thongpanja, K. Anam, A. Al-Jumaily, and C. Limsakul. “Evaluation of feature extraction techniques and classifiers for finger movement recognition using surface electromyography signal,” *Med. Biol. Eng. Comput.*, 56(12), pp. 2259–71, 2018.

ภาคผนวก ก

การตีพิมพ์เผยแพร่ผลงาน

1. N. Sae Jong and P. Phukpattaranont, “A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A thai syllable study,” Biocybernetics and Biomedical Engineering, vol. 39, no. 1, pp. 234–245, 2019.

2. N. Sae Jong, M. Kiatweerasakul, P. Phukpattaranont, “Channel reduction in speech recognition system based on surface electromyography,” 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 184-187, 18-21 July 2018, Chiang Rai, Thailand.

N. Sae Jong and P. Phukpattaranont, “A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A thai syllable study,” Biocybernetics and Biomedical Engineering, vol. 39, no. 1, pp. 234–245, 2019.

BIOCYBERNETICS AND BIOMEDICAL ENGINEERING 39 (2019) 234–245

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A Thai syllable study

Nida Sae Jong, Pornchai Phukpattaranont *

Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai, Songkhla, Thailand

ARTICLE INFO

Article history:
Received 5 April 2018
Received in revised form
20 September 2018
Accepted 25 November 2018
Available online 3 December 2018

Keywords:
Dysarthria
Surface electromyography (sEMG)
Feature
Classification
Speech recognition system

ABSTRACT

The objective of this study is to develop a speech recognition system for classifying nine Thai syllables, which is used for the rehabilitation of dysarthric patients, based on five channels of surface electromyography (sEMG) signals from the human articulatory muscles. After the sEMG signal from each channel was collected, it was processed by a band-pass filter from 20–450 Hz for noise removal. Then, six features from three feature categories were determined and analyzed, namely, mean absolute value (MAV) and wavelength (WL) from amplitude based features (ABF), zero crossing (ZC) and mean frequency (MNF) from frequency based features (FBF), and L-kurtosis (L-KURT) and L-skewness (L-SKW) from statistics based features (SBF). Subsequently, a spectral regression extreme learning machine (SRELM) was used as the feature projection technique to reduce the dimension of feature vector from 30 to 8. Finally, the projected features were classified using a feed forward neural network (NN) classifier with 5-fold cross-validation. The proposed system was evaluated with the sEMG signals from seven healthy volunteers and five dysarthric volunteers. The results show that the proposed system can recognize the sEMG signals from both healthy and dysarthric volunteers. The average classification accuracies obtained from all six features in the healthy and dysarthric volunteers were 94.5% and 89.4%, respectively.

1. Introduction

Stroke is a major health issue worldwide. One of the possible effects from a stroke is flaccid dysarthria or speech impediment. Dysarthria reflects neuromuscular disturbances of strength, speed, tone, steadiness, or accuracy of the movements that underlie the execution of speech. The main characteristic of dysarthria is speech with constantly abnormal pitch and rhythm and other characteristics including slurred, choppy, or mumbled speech that may be difficult to understand. In order to alleviate these problems, a speech recognition system based on an acoustic signal can be used to improve communication [1] and as an

* *Corresponding author at*: Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University, Hat Yai, Songkhla, Thailand.
E-mail addresses: 5710130014@email.psu.ac.th (N. Sae Jong), pornchai.p@psu.ac.th (P. Phukpattaranont).
https://doi.org/10.1016/j.bbe.2018.11.010

index for measuring the severity level of dysarthria [2–7]. In addition to speech recognition systems based on acoustic signals, which are widely proposed in the literatures, systems implemented on surface electromyography (sEMG) signals are currently of interest [8]. The sEMG signal is an electrical potential that is measured from the human articulatory muscles on the face and the neck by surface electrodes. Three advantages of using sEMG signals include (1) capability in recognizing speech in the environment with very loud ambient noise [9], (2) capability to be speech interfaces in rehabilitation applications for individuals with speech disabilities [10], and (3) combination with the acoustic signal for performance improvement [11]. The sEMG-based speech recognition has been proposed in both silent and audible modes. This research focused on sEMG-based audible speech recognition. In other words, the sEMG signals were collected when the volunteers pronounced the words with voicing.

The objective of most research on sEMG-based audible speech recognition is to improve classification accuracy by accurately identifying words or phonemes. To this end, the whole system is divided into four parts consisting of preprocessing, feature extraction, dimensional reduction, and classification. Research on feature extraction and classification is important. In 2008, Lee acquired the sEMG signals from three articulatory muscles to classify 60 isolated Korean words by extracting mel frequency cepstral coefficients and their first temporal derivatives called delta coefficients from 100 ms sEMG frames at 20 ms intervals [12]. These feature parameters were then combined and modeled in a hidden Markov model (HMM) framework. An accuracy of 87.1% was obtained by the word recognition system with the features based on HMM parameters and the maximum likelihood classifier [12]. In the study of Tsuji and his colleagues [13], three channels of sEMG signals were recorded from facial muscles using a monopolar configuration. The process of phoneme recognition was divided into three parts: calculating the differential sEMG signals between the monopolar electrodes attached to different muscles, extracting features using a filter bank, and classifying the phoneme using a probability neural network known as a reduced-dimensional log-linearized Gaussian mixture model, which is capable of combining feature extraction and classification into one step [13]. Further, Kubo and his colleagues collected sEMG signals from an electrode grid arranged in an array of eight rows by eight columns employing 64 pins electrodes and five Japanese vowels were classified [14]. They employed cepstral coefficients, their delta and delta-delta coefficients as features using a window length of 25 ms while the window period was set to 12.5 ms. Dimension reduction was performed using sparse discriminant analysis (SDA) and compared with linear discriminant analysis (LDA). An HMM was used as a classifier and the results showed that LDA gave the highest recognition accuracy using 2520 selected features. Although SDA gave 15% lower recognition accuracy compared to LDA, it was able to reduce the number of selected features to one twentieth of the total features [14].

The problem associated with sEMG-based audible speech recognition is how to build a relationship between contextual information and the corresponding sEMG signal. Feature extraction maps a given sequence of the sEMG signal into a sequence of context words (or phonemes). Feature extraction is also an important part of pattern recognition, which has a significant influence on the recognition results [15]. Based on a review of the literature relating to speech recognition using sEMG signals, features in the time domain including root mean square, zero crossing, and auto-regressive [16] and features in the frequency domain such as the mel-frequency cepstral coefficients and a filter bank [17] have been used for word classification.

The study reported in this paper focused on finding appropriate features that achieve the maximum classification accuracy for a speech recognition system based on electromyography for the rehabilitation of dysarthric patients. Three feature categories were studied consisting of amplitude, frequency and statistical features. The performance of the features from each category was evaluated using boxplots and scatter plots. Moreover, the classification accuracies of these three feature categories and their combination were reported when a spectral regression extreme learning machine (SRELM), which was an efficient feature projection technique proposed by Anam and Al-Jumaily [18], was used. In addition, a neural network (NN) was used as the classifier because its pairing with SRELM feature projection showed better accuracy than other combinations for finger movement recognition using the sEMG signals [19].

2. Feature extraction

Feature extraction is an important process that affects the classification performance of a recognition system and employs a mathematical model that can reduce the dimensionality of the raw sEMG signal by mapping it into a feature vector. Additionally, by selecting suitable features, important information can be extracted and undesirable components such as noise from the sEMG signal can be removed. In this research, three categories of features were studied including amplitude based features (ABF), frequency based features (FBF), and statistics based features (SBF). Two features from each category, which have been successfully used in the sEMG recognition systems from research reported in previous publications were selected, namely, mean absolute value (MAV) and wavelength (WL) from ABF, zero crossing (ZC) and mean frequency (MNF) from FBF, and L-kurtosis (L-KURT) and L-skewness (L-SKW) from SBF. Details of each feature are as follows:

MAV is one of the most popular features used in sEMG signal analysis. MAV is an average of the absolute values of sEMG signal amplitudes in a sampled segment, which can be defined as [20]

$$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{1}$$

where x_i represents the sEMG amplitude at sample i and N denotes the length of the sEMG signal.

WL is a measure of the complexity of the sEMG signal. It is defined as the cumulative length of the sEMG waveform over a time segment. It can be calculated by [21]:

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{2}$$

ZC is a measure of the frequency information of the sEMG signal. It represents the number of times that the amplitude value of the sEMG signal crosses the zero amplitude level. To avoid low voltage fluctuations or background noises, a threshold condition is implemented. ZC is defined as [21]:

$$ZC = \sum_{i=1}^{N-1} [f(x_i \times x_{i+1}) \text{ and } |x_i - x_{i+1}| \geq 10] \tag{3}$$

$$f(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

MNF is the average frequency, which is calculated as the sum of the product of the sEMG power spectrum and the frequency divided by the total sum of the spectrum intensity. The MNF feature is also known as the central frequency or the spectral center of gravity [22]. It can be calculated as [23]:

$$MNF = \sum_{j=1}^{M} f_j P_j / \sum_{j=1}^{M} P_j \tag{5}$$

where f_j is the frequency of the spectrum at frequency bin j, P_j is the sEMG power spectrum at frequency bin j, and M is the number of the frequency bins.

L-KURT is a measurement of the peakedness or kurtosis of the sEMG amplitude distribution. It is determined using the ratio of the fourth L-moment to the second L-moment, and is calculated from linear combinations of order statistics of a random variable, which can be expressed as [24]:

$$\text{L-KURT} = \frac{20b_3 - 30b_2 + 12b_1 - b_0}{2b_1 - b_0} \tag{6}$$

where b_r is an unbiased estimator of the probability-weight moment. The general expression for b_r is:

$$b_r = \frac{1}{N} \sum_{j=r+1}^{N} y_j \left[\frac{(j-1)(j-2)\cdots(j-r)}{(N-1)(N-2)\cdots(N-r)} \right] \tag{7}$$

where y_j is obtained by sorting the normalized EMG data that have zero mean and unit variance to ascending order from 1 to N. The standard L-KURT values for standard Gaussian and Laplacian distributions are 0.1226 and 0.2357, respectively.

L-SKW is a dimensionless measure of the asymmetry of the sEMG amplitude distribution, which may take on positive or negative values. For a distribution or sample data, L-SKW is in the range of 0–1. The relative magnitude of L-SKW for a dataset or distribution is given by [25]:

$$\text{L-SKW} = \frac{6b_2 - 6b_1 + b_0}{2b_1 - b_0} \tag{8}$$

3. Materials and methods

3.1. Data acquisition

The sEMG signals resulting from speech are generated by the facial muscles. Normally, pursing the lips, lifting the corner of the mouth, and opening the jaw are the postures used in rehabilitation for a speech impediment or speech disorders. Additionally, the sEMG signal appears in the extrinsic muscles of the tongue, which are responsible for moving the tongue up, forward or backward. Fig. 1 shows the five electrode placement positions used for sEMG data acquisition in this paper. The electrodes were placed on the zygomaticus major, levator anguli oris, depressor anguli oris, mentalis and anterior belly of the digastrics, which were shown in Fig. 1 as CH1, CH2, CH3, CH4, and CH5, respectively. The selection of electrode placement positions followed reports in a previous publication [26]. Moreover, the preliminary results showed that the sEMG signal acquired from these muscles were appropriate for rehabilitation using Thai word classification [27].

The sEMG signals were measured from seven healthy volunteers (three females and four males) aged between 20 and 22 years with no speech impediments or disorders, in order to test reproducibility. A commercial sEMG measurement system (Mobi6-6b, TMS International B.V., Oldenzaal, the Netherlands) with sintered disc shaped sEMG electrodes (5 mm diameter, Ag/AgCl) and shielded cables was used for acquiring the sEMG signals. The signals measured were digitized at a sampling frequency of 1024 Hz. The Mobi6-6b had six channels for data acquisition. The sEMG signals were recorded both in unipolar (CH2, CH3, CH4 and CH5) and bipolar (CH1) configurations using the first five channels. The sixth channel was connected to a multifunction DAQ USB device (NI USB-6009, National Instruments Corporation, USA) that received the trigger signal from program to mark the start point of syllable. The reference electrode was placed on the earlobe and the ground electrode was placed on the left wrist.

For the preliminary training, syllables were chosen based on those most easily articulated by children. The volunteers were requested to articulate isolated syllables from the Thai language composed of a consonant followed by a vowel. The syllables were divided into three groups according to their place of articulation: velar, alveolar and bilabial. In order to have a representative set of syllables for the system applied for flaccid dysarthria speaker, one syllable was selected from a representative consonant from each group combined with three vowels. The vowels "a", "i" and "u" were selected for training because they were the vowels that maximize the opening of the mouth, smiling and the pursing of the lips. Thus, the final set was composed of nine syllables as shown in Table 1. The volunteer was instructed to articulate each syllable for 3 s. This was controlled using our developed interface software consisting of an LED and a bar graph on a computer screen. When the LED was on, the volunteer started to articulate a syllable. The bar graph, which was used as a timer, showed the elapsed time. When 3 s was reached, the

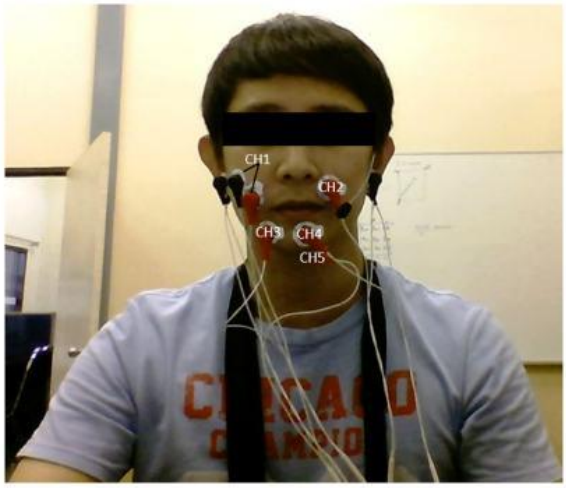


Fig. 1 – Electrode placement of 5-channel sEMG data acquisition.

Table 1 – Set of syllables used in rehabilitation.

Velars	Labials	Alveolars
/ka/	/ma/	/na/
/ki/	/mi/	/ni/
/ku/	/mu/	/nu/

LED was off and the volunteer stopped to articulate the syllable. Moreover, each syllable was repeated five times in order to test the robustness of the algorithm. As a result, the total number of sEMG signals from each volunteer was thus 225 (9 syllables × 5 channels × 5 trials). Fig. 2(a) shows an example of the 5-channel sEMG signals from the syllable "ki" of a healthy volunteer.

Five dysarthric volunteers (four females and one male) aged 52 ± 12.6 years who had neurological facial problems causing motor speech disorders were recruited for this study. The same data acquisition and sampling rate as in the healthy volunteer were used. However, the articulatory time was reduced to 2 s for the dysarthric volunteer because it is quite difficult for the dysarthric volunteer to perform 3-s articulation. Fig. 2(b) shows an example of the 5-channel sEMG signals from the syllable "ki" of a dysarthric volunteer.

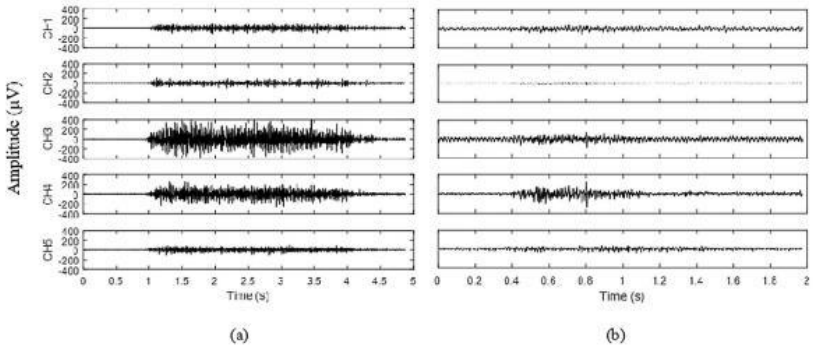


Fig. 2 – Example of the 5-channel sEMG signals from the syllable "ki": (a) Healthy volunteer; (b) Dysarthric volunteer.

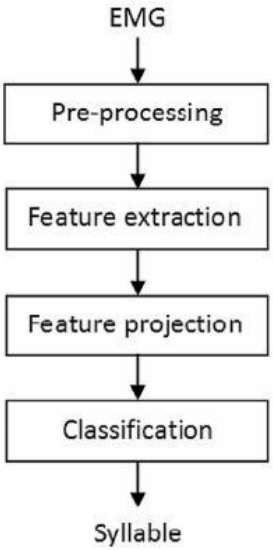


Fig. 3 – Block diagram of the proposed syllable recognition system based on electromyography.

3.2. *Proposed method*

Fig. 3 shows a block diagram of the proposed syllable recognition system based on electromyography consisting of four modules: pre-processing, feature extraction, feature projection, and classification. Details of each module are given as follows.

Pre-processing: After the sEMG signal was collected, a band-pass filter was applied with cutoff frequencies of 20 and 450 Hz for noise removal. After that, the start point of the syllable was detected. Details of the detection algorithm consisting of five steps are as follows.

Step 1. Determine the position of the trigger signal.

Step 2. Define the 500-ms signal preceding the trigger signal as the noise $n(t)$ and calculate the noise level, $A = \mu_n + 2\sigma_n$, where μ_n and σ_n are the mean and standard deviation of $|n(t)|$, respectively.

Step 3. Define the 32-ms signal succeeding the trigger signal as $x(t)$ and calculate the signal level, $B = \mu_x + 2\sigma_x$, where μ_x and σ_x are the mean and standard deviation of $|x(t)|$, respectively.

Step 4. Compare B to 8A. If B is greater than 8A, the start point is the midpoint of $x(t)$ segment. Otherwise, define the next 32-ms segment of the sEMG signal and calculate B until B is greater than 8A and the midpoint of $x(t)$ segment is the start point of the syllable.

Step 5. Calculate the end point of the syllable by adding the start point with 2.5 s for the healthy volunteer and 1.6 s for the dysarthric volunteer. Therefore, the total number of the sEMG signals for each channel is 45 (9 syllables × 5 trials).

Feature extraction: Each of 45 sEMG signals from the pre-processing module was segmented into frames using a window length of 256 samples (250 ms) with 50% overlap. As a result, the numbers of frames for the healthy and dysarthric volunteers were 22 and 11, respectively. The total numbers of frames from each healthy and dysarthric volunteers were 990 (22 frames × 5 trials × 9 syllables) and 495 (11 frames × 5 trials × 9 syllables), respectively. Thirty features (5 channels × 6 features/channel) for each EMG frame were extracted resulting in the feature vector with a dimension of 30 for each syllable. Subsequently, each feature from each channel of all frames was normalized so that its value was in a range of −1 to 1.

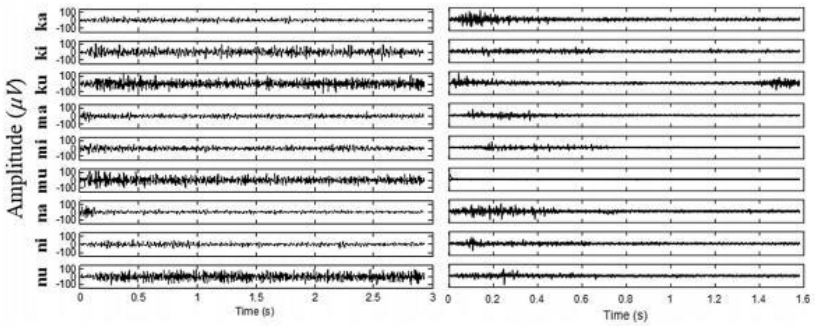


Fig. 4 – Example of the sEMG signals from CH2 when articulating the nine syllables. (Left) Healthy volunteer. (Right) Dysarthric volunteer.

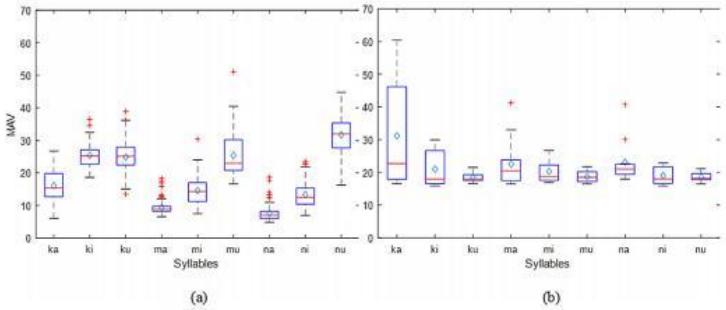


Fig. 5 – Boxplots of MAV determined using the sEMG signals from CH2: (a) Healthy volunteer; (b) Dysarthric volunteer.

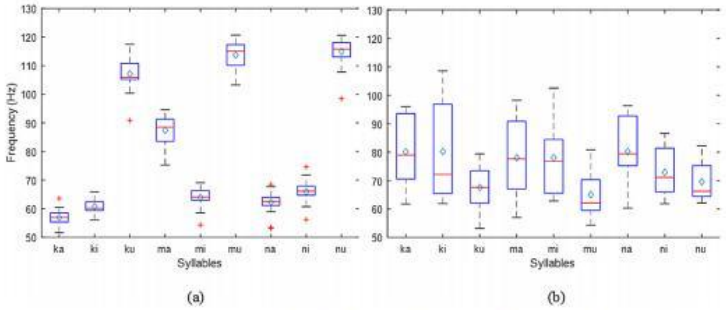


Fig. 6 – Boxplots of MNF determined using the sEMG signals from CH2: (a) Healthy volunteer; (b) Dysarthric volunteer.

Feature projection: The dimension of the normalized features was reduced using SRELM. In brief, SRELM is a technique combining an extreme learning machine (ELM) and spectral regression (SR). It utilizes the eigenvector obtained to project the hidden layer output to the output layer. The hidden layer weights are estimated randomly while the output weight is computed by SR which uses the least squares method to obtain the best projection direction. Two parameters were employed to optimize SRELM performance: the number of hidden nodes and alpha. In order to select the optimal parameters, the

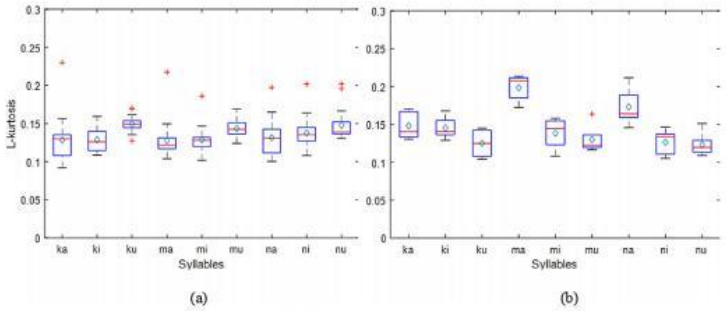


Fig. 7 – Boxplots of L-KURT determined using the sEMG signals from CH2: (a) Healthy volunteer; (b) Dysarthric volunteer.

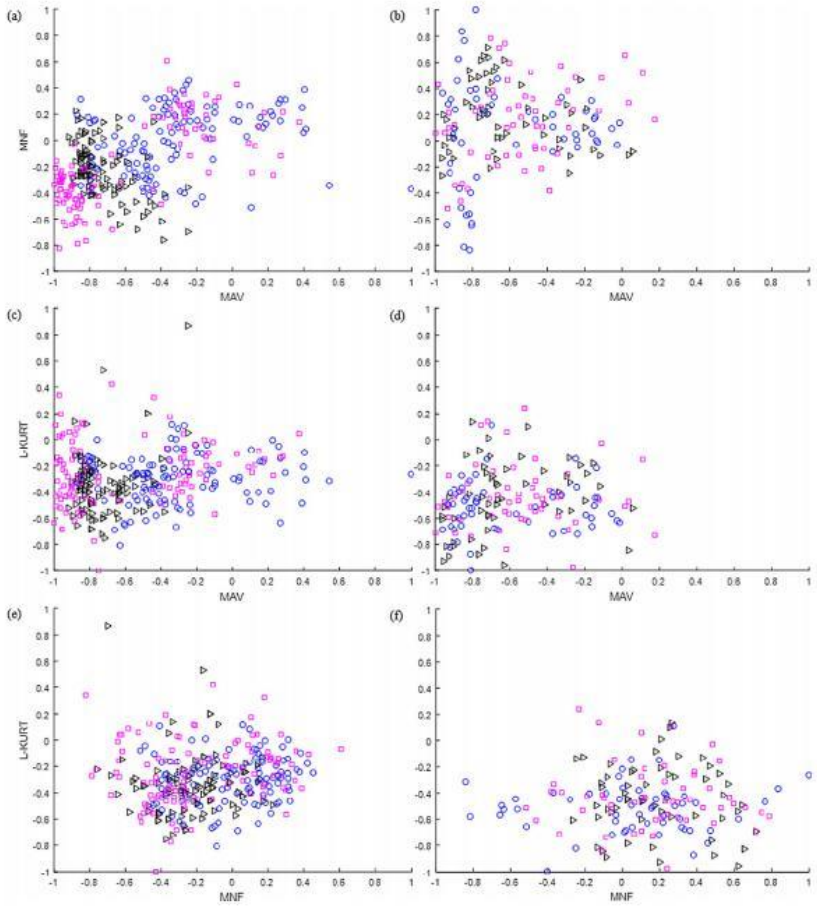


Fig. 8 – Scatter plots of the normalized features before feature projection from "ma" in black, "mi" in blue, and "mu" in magenta. (Left) Healthy volunteer. (Right) Dysarthric volunteer.

number of hidden nodes was varied from 100 to 1500 nodes with an increment of 100 nodes and alpha was varied from 1 to 20 with an increment of 1. SRELM reduced the length of the feature vector from 30 to $c - 1$, where c was the number of syllables ($c = 9$ in this paper).

Classification: The projected features from SRELM were classified using a feed forward NN classifier. The structure of the NN consisted of three layers: an input layer (8 nodes), hidden layers, and an output layer (9 nodes). The hyberbolic tangent sigmoid transfer function was used in both hidden and output layers. There were 20 nodes in the hidden layer, which gave the maximum accuracy when the number of nodes from 15 to 25 was tested.

3.3. *Performance evaluation*

We evaluated the performance of proposed algorithm using 5-fold cross-validation. In other words, the normalized features were randomly partitioned into 5 subsets. The SRELM feature projection and NN classifier were trained using 4 subsets. After training, the remaining subset, which was used as the test dataset, was applied with the SRELM and NN from the training step. This process was repeated 5 times such that each of the 5 subsets was used as the test dataset. The performance was evaluated and compared using the mean and standard deviation of classification accuracies from 5 test dataset. In addition, we used speaker-dependent criteria in this paper. In

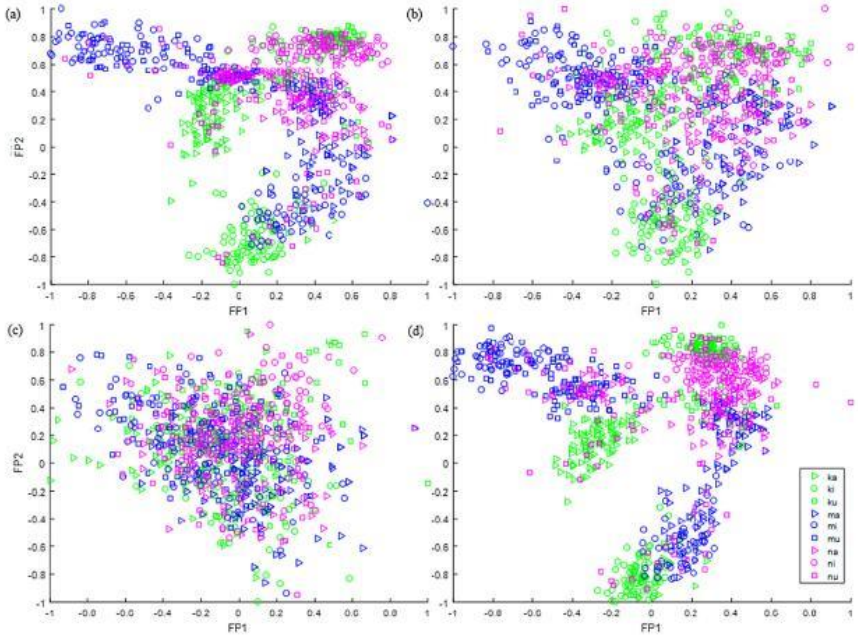


Fig. 9 – Scatter plots of the normalized features after feature projection from the healthy volunteer: (a) ABF; (b) FBF; (c) PBF; (d) ACF.

other words, the training process was performed using the sEMG dataset from each individual volunteer and the testing process was done using the different sEMG dataset from the same volunteer. Moreover, we compared the classification accuracies from SRELM with those from principal component analysis (PCA) and linear discriminant analysis (LDA), which were commonly used feature projection techniques. While LDA represented a supervised feature projection with consideration of the class label, PCA was a representative of an unsupervised feature projection.

4. Results and discussion

4.1. *Surface EMG signals and feature characteristics*

Fig. 4(a) and (b) shows an example of the sEMG signals from CH2 when articulating the nine syllables of the healthy and dysarthric volunteers, respectively. The results show that the magnitude of the sEMG from the healthy volunteer is higher than that from the dysarthric volunteer for all the syllables because the healthy volunteer has stronger muscle contractions. The characteristics of MAV, MNF, and L-KURT, which are representatives from the ABF, FBF, and SBF feature categories, are given using boxplots and scatter plots. Fig. 5(a) and (b) shows the boxplots of MAV determined using the sEMG signal from CH2 of the healthy and dysarthric volunteers, respectively. Each box shows the lower (Q1), median (Q2) and upper (Q3) quartile and the interquartile range (IRQ = Q3 – Q1) of each syllable. Additionally, the average value is shown using a blue diamond marker. It can be observed that the differences in the between-groups variation of the healthy volunteer are significant while the within-group variation is quite low. On the contrary, the differences in the between-group variation of the dysarthric volunteer is vague while the within-group variation is quite high. However, in order to check the statistical significance of the boxplot results, a one-way analysis of variance (ANOVA) was conducted and a significance level of 0.05 was utilized. The results indicate that the differences in the between-group variation for the healthy volunteer are significant but they are not significant for the dysarthric volunteer. Therefore, the statistical test agrees well with the observation results.

Fig. 6(a) shows the boxplot of the MNF from the healthy volunteer. The results show that the MNF values from each syllable are well separated. In contrast, the MNF values from each syllable of the dysarthric volunteer overlap as shown in Fig. 6(b). The results from the ANOVA are in agreement with these observations. Moreover, these trends are similar to those of the MAV feature.

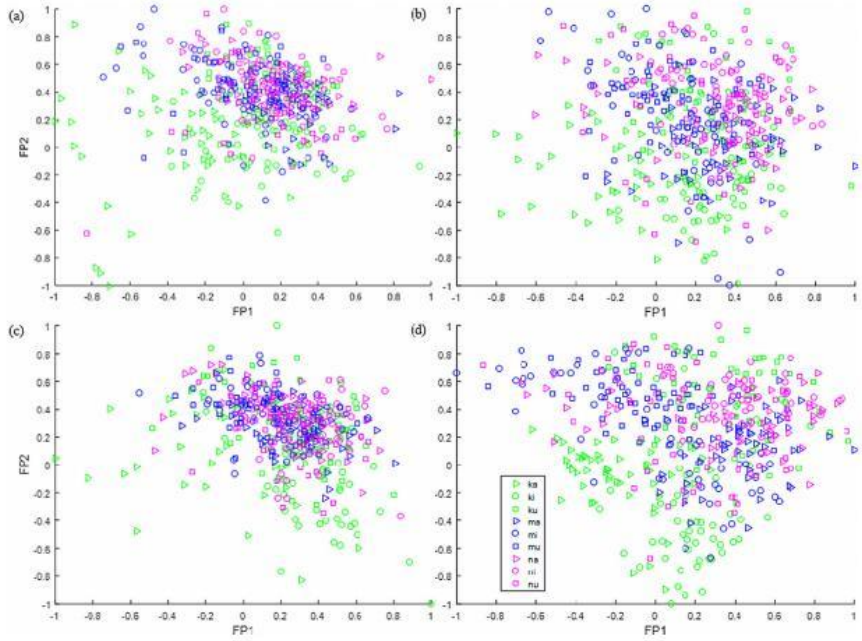


Fig. 10 – Scatter plots of the normalized features after feature projection from the dysarthric volunteer: (a) ABF; (b) FBF; (c) PBF; (d) ACF.

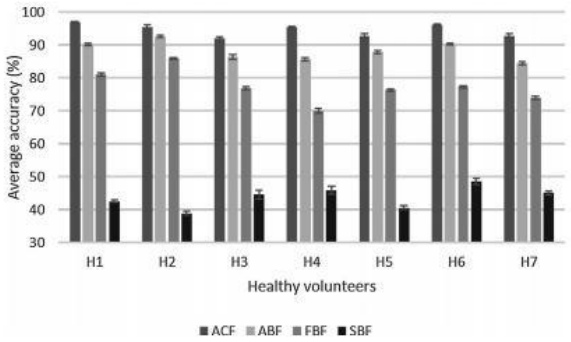


Fig. 11 – Classification accuracies for nine syllables obtained with four feature groups from the healthy volunteers.

Fig. 7(a) and (b) shows the boxplots for L-KURT determined using the sEMG signal from CH2 of the healthy and dysarthric volunteers, respectively. It can be observed that the differences in the between-group variation and the differences in the within-group variation are quite similar for the healthy and dysarthric volunteers. However, the results from the ANOVA indicate that the differences in the within-group variation of the dysarthric volunteer are lower than those from the healthy volunteer. These results show that it is quite difficult to classify the nine syllables by using the features from only a single feature category.

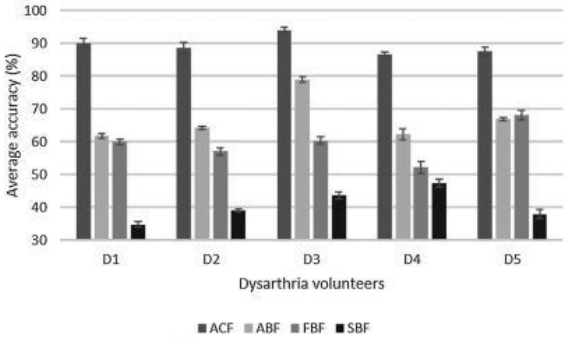


Fig. 12 – Classification accuracies for nine syllables obtained with four feature groups from the dysarthric volunteers.

Table 2 – Statistical values of classification accuracies from 4 feature categories when the SRELM feature projection is used.

Feature category	Classification accuracy (%)			
	Mean	SD	Min	Max
Healthy volunteer				
ABF	88.2	0.4	84.5	92.6
FBF	77.4	0.5	74.0	85.9
SBF	43.8	0.8	39.0	48.6
ACF	94.5	0.5	92.0	96.9
Dysarthric volunteer				
ABF	66.8	0.9	62.0	79.0
FBF	59.6	1.3	52.2	68.1
SBF	40.6	1.0	34.9	47.4
ACF	89.4	1.2	86.6	94.1

Note. SD = standard deviation; Min = minimum accuracy; Max = - maximum accuracy.

Table 3 – Statistical values of classification accuracies from 3 feature projection techniques when the ACF feature category is used.

Feature projection	Classification accuracy (%)			
	Mean	SD	Min	Max
Healthy volunteer				
PCA	78.5	1.7	72.1	89.7
LDA	88.4	1.0	81.9	92.4
SRELM	94.5	0.5	92.0	96.9
Dysarthric volunteer				
PCA	60.6	3.0	50.7	73.5
LDA	74.0	1.6	67.7	82.6
SRELM	89.4	1.2	86.6	94.1

Note. SD = standard deviation; Min = minimum accuracy; Max = - maximum accuracy.

Scatter plots are employed to show the ability of feature pairs to separate the sEMG signals from three syllables, "ma", "mi" and "mu". Fig. 8(a), (c), and (e) shows the scatter plots of three pairs of normalized features before the feature projection for the healthy volunteer. Similar scatter plots from the dysarthric volunteer are shown in Fig. 8(b), (d), and (f). All scatter plots show overlapping features, which make them difficult to classify syllables. Therefore, in order to achieve better classification accuracy, the features from four categories, ABF, FBF, SBF, and the combined features from all categories (ACF), are tested using the feature projection technique based on SRELM. Fig. 9 shows the scatter plots of two projected features after feature projection with SRELM from the healthy volunteer. Fig. 9(a)–(d) shows the scatter plots of two projected features from the ABF, FBF, SBF and ACF feature categories, respectively. The results show that the projected features from the ACF give the maximum degree of separation. Similar scatter plots are obtained from the features of the dysarthric volunteer as shown in Fig. 10. However, the degree of separation from the ACF in the dysarthric volunteer is slightly lower than that from the ACF in the healthy volunteer. This may be caused by the weakness of the muscle contractions for articulation in the dysarthric volunteer. Moreover, the ABF shows a higher degree of separation than the FBF and SBF.

4.2. *Classification accuracy*

Fig. 11 shows the average classification accuracies for nine syllables obtained with four feature categories from seven healthy volunteers. For each volunteer, the maximum classification accuracy is obtained from the ACF. Similar results from the dysarthric volunteers are shown in Fig. 12. These results are consistent with those from the scatter plots. Moreover, the results indicate that the proposed system consisting of: (1) multiple features from ABF, FBF, and SBF; and (2) SRELM feature projection; are appropriate for classifying nine syllables from sEMG signals. Table 2 shows the statistical significance of the classification accuracies from ACF. The mean ± SD values of the classification accuracies from ACF in the healthy and dysarthric volunteers are 94.5 ± 0.5% and 89.4 ± 1.2%, respectively. The maximum accuracy from both groups of volunteers is comparable at the level of 94.1–96.9%. However, the minimum accuracy for the dysarthric volunteer D4 is 86.6%, which is quite low compared to that

from the healthy volunteer H3 (92.0%). This may be caused by variation in the severity level of the dysarthric volunteers.

Table 3 shows statistical values of classification accuracies from SRELM compared to those from PCA and LDA when the ACF feature category is used. The mean classification accuracies indicate that SRELM outperforms LDA and PCA. On the one hand, the mean classification accuracy from SRELM in the healthy volunteer is 94.5%, which is better than those from both LDA (88.4%) and PCA (78.5%). On the other hand, the mean classification accuracy from SRELM in the dysarthric volunteer (89.4%) is also better than those from both LDA (74.0%) and PCA (60.6%). Results obtained from this paper show similar trend with the accuracies in [19] where the pairing of SRELM feature projection technique and NN classifier gives better classification accuracy than other combinations for finger movement recognition using the sEMG signals.

5. Conclusions

This paper proposes a speech recognition system based on a 5-channel sEMG signal for classifying nine Thai syllables used for rehabilitation in dysarthric patients. The system consists of four modules, preprocessing, feature extraction, dimensional reduction, and classification. The sEMG signal in each channel was processed with a band-pass filter from 20–450 Hz for noise removal in the preprocessing module. Then, six features from four feature categories were determined and analyzed: (1) MAV and WL from ABF, (2) ZC and MNF from FBF, (3) L-KURT and L-SKW from SBF, and (4) all six features from all three categories. Subsequently, SRELM was used as the feature projection technique to reduce the length of the feature vector from 30 to 8. Finally, the projected features were classified using a feed forward NN classifier with 5-fold cross-validation. The proposed method was evaluated with the sEMG signals from seven healthy volunteers and five dysarthric volunteers. The results show that the best performance is achieved when all six features are used. The average classification accuracies obtained from the healthy and dysarthric volunteers are 94.5% and 89.4%, respectively. These results suggest that the proposed method has the potential to be used for classifying the syllables for rehabilitation in dysarthric patients. Based on the success of the proposed system in syllable classification from both healthy and dysarthric volunteers, possible future research directions include improving the accuracy of measuring the severity level of dysarthria using a combination of sEMG and acoustic signals in the recognition system as well as improving dysarthria subtype identification. Moreover, the combination of sEMG and acoustic signals can be employed to improve the recognition system used for communication in dysarthric patients.

Acknowledgments

This work was jointly funded by Prince of Songkla University and by the Office of the Higher Education Commission, Ministry of Education, Thailand (Grant No. 006/2558). Moreover, the authors would like to thank Duangmon Vongjandaeng from the Department of Otolaryngology, Hat Yai Hospital, Songkhla, Thailand for her suggestions on the syllables selected based on their use for the rehabilitation of dysarthric speakers, and her training on syllable pronunciation for the volunteers.

REFERENCES

[1] Mustafa MB, Salim SS, Mohamed N, Al-Qatab B, Siong CE. Severity-based adaptation with limited data for ASR to aid dysarthric speakers. PLoS One 2014;9(5):1–11.
[2] Kayasith P, Theeramunkong T. Speech confusion index (Φ): a confusion-based speech quality indicator and recognition rate prediction for dysarthria. Comput Math Appl 2009;58:1534–49.
[3] Lansford KL, Liss JM. Vowel acoustics in dysarthria: speech disorder diagnosis and classification. J Speech Lang Hear Res 2014;57(1):57–67.
[4] Kim MJ, Kim Y, Kim H. Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model. IEEE/ACM Trans Audio Speech Lang Process 2015;23(4):694–704.
[5] Kim YJ, Weismer G, Kent RD, Duffy JR. Statistical models of F2 slope in relation to severity of dysarthria. Folia Phoniatr Logo 2009;61(6):329–35.
[6] Kim YJ, Kent RD, Weismer G. An acoustic study of the relationships among neurologic disease, dysarthria type and severity of dysarthria. J Speech Lang Hear Res 2011;54 (2):417–29.
[7] Darley F, Aronson A, Brown J. Differential diagnostic patterns of dysarthria. J Speech Lang Hear Res 1969;12 (2):246–69.
[8] Srisuwan N, Phukpattaranont P, Limsakul C. Comparison of feature evaluation criteria for speech recognition based on electromyography. Med Biol Eng Comput 2018;56(6):1041–51.
[9] Janke M, Diener L. EMG-to-Speech: direct generation of speech from facial electromyographic signals. IEEE/ACM Trans Audio Speech Lang Process 2017;25(12):2375–85.
[10] Schultz T, Wand M, Hueber T, Krusienski DJ, Herff C, Brumberg JS. Biosignal-based spoken communication: a survey. IEEE/ACM Trans Audio Speech Lang Process 2017;25 (12):2257–71.
[11] Chan ADC, Englehart KB, Hudgins B, Lovely DF. Myoelectric signals to augment speech recognition. Med Biol Eng Comput 2001;39:500–4.
[12] Lee KS. EMG-based speech recognition using hidden Markov models with global control variables. IEEE Trans Biomed Eng 2008;55(3):930–40.
[13] Tsuji T, Bu N, Arita J, Ohga M. A speech synthesizer using facial EMG signals. Int J Comput Intell Appl 2008;7(1):1–15.
[14] Kubo T, Yoshida M, Hattori T, Ikeda K. Towards excluding redundancy in electrode grid for automatic speech recognition based on surface EMG. Neurocomputing 2014;134:15–9.
[15] Bunderson NE, Kuiken TA. Quantification of feature space changes with experience during electromyogram pattern recognition control. IEEE Trans Neural Syst Rehabil Eng 2012;20(3):239–46.
[16] Chan ADC, Englehart K, Hudgins B, Lovely DF. Hidden Markov model classification of myoelectric signals in speech. IEEE Eng Med Biol Mag 2002;21(5):143–6.
[17] Jou SCS, Schultz T. Automatic speech recognition based on electromyographic biosignals. International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC). 2008. pp. 305–20.
[18] Anam K, Al-Jumaily A. A novel extreme learning machine for dimensionality reduction on finger movement classification using sEMG. Proceedings of the Seventh

International IEEE/EMBS Conference on Neural Engineering (NER); 2015. p. 824–7.
[19] Phukpattaranont P, Thongpanja S, Anam K, Al-Jumaily A, Limsakul C. Evaluation of feature extraction techniques and classifiers for finger movement recognition using surface electromyography signal. Med Biol Eng Comput 2018;56(12):2259–71.
[20] Englehart K, Hudgins B. A robust, real-time control scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 2003;50(7):848–54.
[21] Du YC, Lin CH, Shyu LY, Chen T. Portable hand motion classifier for multi-channel surface electromyography recognition using grey relation analysis. Expert Syst Appl 2010;37(6):4283–91.
[22] Phinyomark A, Phukpattaranont P, Limsakul C. Feature reduction and selection for EMG signal classification. Expert Syst Appl 2012;39(8):7420–31.
[23] Du S, Vuskovic M. Temporal vs. spectral approach to feature extraction from prehensile EMG signals. Proceedings of IEEE International Conference on Information Reuse and Integration (IRI); 2004. p. 344–50.
[24] Thongpanja S, Phinyomark A, Quaine F, Laurillau Y, Limsakul C, Phukpattaranont P. Probability density functions of stationary surface EMG signals in noisy environments. IEEE Trans Instrum Meas 2016;65(7):1547–57.
[25] Sankarasubramanian A, Srinivasan K. Investigation and comparison of sampling properties of L-moments and conventional moments. J Hydrol 1999;218(1–2):13–34.
[26] Lapatki BG, Stegeman DF, Jonas IE. A surface EMG electrode for the simultaneous observation of multiple facial muscles. J Neurosci Methods 2003;123(2):117–28.
[27] Pothirat T. Characterization and analysis of neck and face surface electromyography for speech rehabilitation training.[Master thesis] Prince of Songkla University; 2014.

N. Sae Jong, M. Kiatweerasakul, P. Phukpattaranont, “Channel Reduction in Speech Recognition System based on Surface Electromyography,” 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 184-187, 18-21 July 2018, Chiang Rai, Thailand.

Channel reduction in speech recognition system based on surface electromyography

Nida Sae Jong, Dr. Monthep Kiatweerasakul, and Associate Professor Dr. Pornchai Phukpattaranont
Department of Electrical Engineering, Faculty of Engineering, Prince of Songkla University
Hat Yai, Songkhla, 90112, Thailand
Email: jern.eng@gmail.com, montep.k@psu.ac.th, pornchai.p@psu.ac.th

***Abstract*— The disadvantage of surface electromyography (sEMG) application is the number of electrodes that affects to undesirable use in practical. So, this research focuses on studying the performance of the proposed sEMG-based speech recognition in classifying 12 Thai syllables by reducing the number of electrodes. The sEMG signals were acquired from the 5 facial muscles of 7 volunteers. Each sEMG signal was filtered using a band-pass filter with a cutoff frequency 20-450 Hz. Then, feature representations were extracted including feature in time domain, frequency domain and statistical domain. Next, the spectral regression extreme learning machine (SRELM) was used to reduce the dimension of data and retain most of the relevant information. Finally, feed forward neural network (NN) was applied as a classifier. The number of electrodes was reduced and its performance was tested. Results show that the average classification accuracies of 2, 3, 4 and 5 channel electrode combination were 87.05%, 94.77%, 97.81%, and 99.01%, respectively.**

Keywords—surface electromyography (sEMG); feature; classification; speech recognition.

I. Introduction

Nowadays, surface electromyography (sEMG) based automatic speech recognition (ASR) researches both in a silent mode [1], [2] and an audible mode [3], [4] is of interest. The various applications of the sEMG-based ASR include (1) its utility in inappropriate environments, such as in a firefighter suit, a hazardous materials suit, and an inspiratory mask; (2) its ability as a communication tool for the patients who have the speech disorder due to the larygectomy, the cerebral palsy, and the tracheotomy; and (3) its employment in rehabilitation robotics for motor recovery. Moreover, the sEMG signals captured from eight facial and neck muscles were used in improving coordination for people with stroke [4].

One of the drawback of sEMG application is the number of electrodes that is fairly large. To overcome this limitation, Tsuji and colleagues [3] proposed an sEMG acquisition method based on a monopolar configuration, which was different from a conventional bipolar configuration. In other words, only one electrode was attached to each muscle in the monopolar configuration. Subsequently, differential signals between every two electrodes can be derived and used as input channels for classification. Colby and colleagues [5] presented the speaker-dependent automatic speech recognition accuracy for isolated words using eleven sEMG sensors in fixed recording locations on the face and neck. They reduce the number of sensors needed and find the best combination of sensor locations to achieve word recognition rates comparable to the full set. They evaluated each of the different possible subsets by the average of word recognition rate across nine speakers using hidden markov model (HMM) modeling of mel-frequency cepstral coefficient (MFCC) and co-activation features derived from the subset of sensor signals. The result shows that five sensors are sufficient to achieve a high recognition rate compared to that obtainable from the full set of sensors.

This paper aims to investigate the performance of the speech recognition based on sEMG signals when reducing the number of electrodes. The rest of this paper is organized as follows. Section II describes an sEMG acquisition and the proposed speech recognition system based on the sEMG signals. Section III gives the classification accuracies resulting from the reduction in the number of electrodes. Finally, conclusions are drawn in Section IV.

II. Materials and Methods

Fig. 1 shows a block diagram of the proposed sEMG-based Thai speech recognition system. Details of each block consisting of data acquisition, preprocessing, feature extraction, feature projection, and classification are given as follows.

A. Data acquisition

Fig. 2 shows five channels of electrode placement for EMG data acquisition in this paper. The electrodes were placed on the zygomaticus major (ZYG), levator anguli oris (LAO), depressor anguli oris (DAO), mentalis (M) and anterior belly of the digastrics (ABD), which were shown in Fig. 2 as CH1, CH2, CH3, CH4, and CH5, respectively. The selection of electrode placement positions was suggested by the previous publication [6]. The preliminary results showed that the EMG signals acquired from these muscles were appropriate for Thai rehabilitation word classification.

We measured the sEMG signals from seven healthy volunteers (three females and four males) aged in range 20 to 22 years old with non-speech impediment or disorders to test reproducibility. A Porti-system from TMSi (Oldenzaal, the

978-1-5386-3555-1/18/$31.00 ©2018 IEEE

Nethelands) with sintered disc shape sEMG electrode (5 mm diameter, Ag/AgCl) and shielded cables were used for acquiring the sEMG signals. The measured signals were digitized at a sampling frequency of 1024 Hz. Moreover, we recorded the signals both in unipolar (CH2, CH3, CH4 and CH5) and bipolar (CH1) configurations. The reference electrodes were placed on the earlobe and the ground electrode was placed on the left wrist.

For the preliminary training, the syllables were chosen based on the easiest articulation for children. We requested our volunteers to articulate isolated syllables from the Thai language that were composed of a consonant followed by a vowel. All syllables were divided into 4 groups according to place of articulation: laryngeal, velar, alveolar and bilabial. The syllable was selected from one representative consonant of each group combined with the three vowels. So, the final set is composed of 12 syllables as shown in Table 1.

B. Pre-processing

After the sEMG signal was collected, we applied a band-pass filter with a cut-off frequency 20 Hz - 450 Hz for noise removal. Subsequently, the start point of the sEMG signal was detected and the sEMG signal was segmented into frames using a window length of 256 samples (250 ms) with 50% overlap. As a result, twenty-two frames per sEMG channel were obtained for each healthy volunteers.

C. Feature generation

For each frame of sEMG data, feature values were calculated including (1) time domain features consisting of mean absolute value, waveform length, zero crossing, slope sign change, and the forth-order autoregressive (4 values); (2) frequency domain features consisting of mean frequency; and (3) statistical features consisting of L-kurtosis and L-skewness. Details of each feature were shown in Table 2. The feature vector with a length of 11 was determined from each frame of the sEMG signal. Then, to prepare the feature values to be used as the input of feature projection, they are normalized by the min-max normalization technique, which can be expressed as

$$\text{norm}(x) = -1 + \left\{ \frac{2 \times \left[x_i - \min(x) \right]}{\max(x) - \min(x)} \right\} \quad (1)$$

As a result, the normalized features were in the range of value from -1 to 1. The dimension of normalized data was 1320 (22×12×5) rows and 55 (11×5) columns for each features when all 5-channel EMG signals were used.

D. Feature projection

The dimension of the normalized features was reduced using spectral regression extreme learning machine (SRELM), which is an efficient feature projection technique as proposed by Anam and college [13]. In brief, SRELM is a technique combining extreme learning machine (ELM) and spectral regression (SR). It utilizes the obtained eigenvector to project the hidden layer output to the output layer. The hidden layer weights are estimated randomly while the output weight is

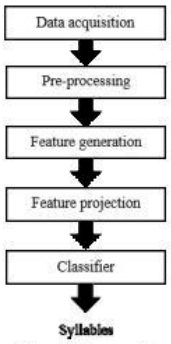


Fig. 1. Block diagram of the syllable recognition system based on electromyography.

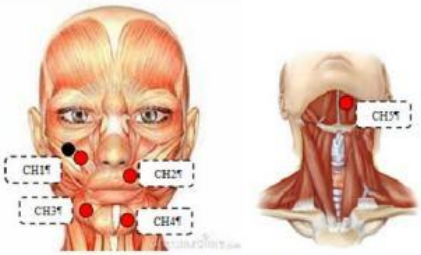


Fig. 2. Electrode placement of 5-channel EMG data acquisition.

Table 1. Set of 12 syllables used in rehabilitation.

Vowel	Laryngeal	Velar	Alveolar	Bilabial
/a/	/a/	/ka/	/na/	/ma/
/i/	/i/	/ki/	/ni/	/mi/
/u/	/u/	/ku/	/nu/	/mu/

computed by SR that uses the least squares method to obtain the best projection direction. There are two parameters to optimize SRELM performance: number of hidden nodes and alpha. In order to select the optimal parameters, the number of hidden nodes was varied from 100 to 1,500 nodes with an increment of 100 nodes and alpha was varied from 1 to 20 with an increment of 1. The dimensionality reduction techniques reduce the features dimensions from 1320 rows × 55 columns (5-channel) to 1320 rows × c-1 columns, where c is the number of syllables, which is 12 in this paper.

E. Classifier

The projected features from SRELM were classified using a feed forward neural network (NN) classifier with a 10-fold cross-validation. The projected feature values were randomly partitioned into 10 subsets. The NN classifier training was performed using 9 subsets and the remaining subset was used

for classifier testing. This process was repeated 10 times such that each of the 10 subsets was used as the testing data. The structure of NN consisted of three layers: input layer (11 nodes), hidden layer, and output layer (12 nodes). There are 20 nodes in the hidden layer, which gives the maximum accuracy when the number of nodes from 15 to 25 are tested. The transfer functions in both hidden and output layers are TANSIG.

III. Results and Discussion

Fig. 3(a) shows the average accuracies for 12 syllable classification from all healthy volunteers obtained with all possible 2-channel combination. For example, 1-2 means that all features used in the algorithm shown Fig. 1 are determined using the sEMG signal from CH1 and CH2 only. The top 3 maximum values in the classification accuracy from 2-channel combination are 1-3, 2-3 and 1-5. Moreover, results show that the channel combination consisting of CH4 and CH5 give low classification accuracy when compared with other channel combination. These may be caused by the redundancy in the features from CH4 and CH5. Fig. 3(b) and (c) show the average accuracies for 12 syllable classification from all healthy volunteers obtained with all possible 3-channel combination and 4-channel combination, respectively. The top 3 maximum values in the classification accuracy from 3-channel combination are 2-3-5, 1-3-5 and 2-3-4. Additionally, the results from the classification accuracy of 4-channel combination shows that 1-2-3-5 gives the maximum value.

Table 3 shows the statistical values of classification accuracy from 2, 3, 4 and 5 channel combinations. The mean ± standard deviation (SD) values of classification accuracies from 2, 3, 4 and 5 channel combinations are 87.05±1.04 %, 94.77±0.71 %, 97.81±0.50 % and 99.01±0.44 %, respectively. From the observation, we found that the SD values decreased when the number of channel combinations increased.

IV. Conclusions

This paper studied the performance of speech recognition system based on the sEMG signals recorded in the five articulatory muscles when reducing the number of electrodes. The approach focused on 12 Thai syllables. The system consisted of 5 steps including data acquisition, pre-processing, feature generation, feature projection and classifier. The sEMG signals from 7 healthy volunteers were collected. Next, the sEMG signals in each channel was filtered by a band-pass filter with cut-off frequency 20-450 Hz. Then, eleven features are calculated from feature in time domain, frequency domain, and statistical domain. Subsequently, the min-max normalization is determined for each feature vector. After that the dimension of data are reduced using SRELM that is one of feature projection. Finally, the projected data are classified using feed forward NN with 10-fold cross validation. Results show that the best performance can be achieved when 5-channel data were used. However, the classification accuracies from 2-, 3- and 4-channel combinations approximately reduced 12%, 4% and 1% compared with the classification accuracy from 5-channel combination. Based on the results of the proposed system on syllable classification, the possible future research direction includes the accuracy improvement in the recognition system aiming to apply with rehabilitation in dysarthric speakers by using the confusion method between the sEMG signal and the acoustic signal.

ACKNOWLEDGMENT

This work is jointly funded by the Prince of Songkla University and by the Office of the Higher Education Commission (Grant No. 006/2558). Moreover, the authority would like to thank Ms. Duangmon Vongjandaeng from Department of Otolaryngology, Hatyai hospital, Songkhla, Thailand for her suggestion in the selected syllybles that can improve the dysarthric speaker in preliminary rehabilitation.

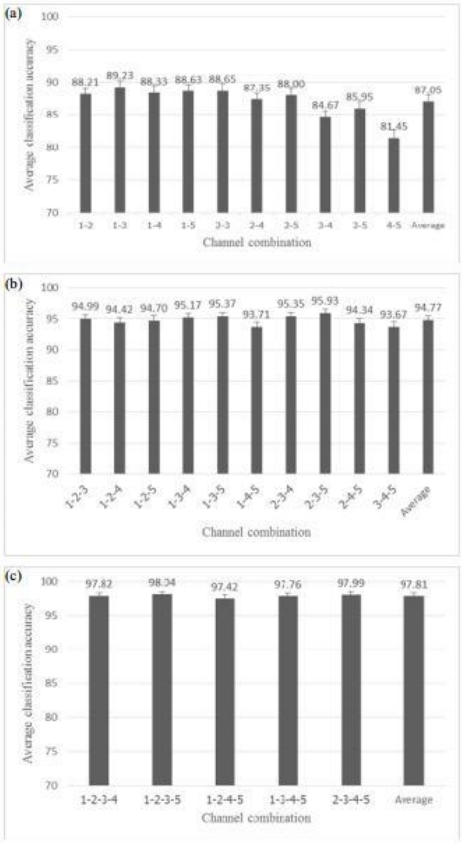


Fig. 3. Average classification accuracies for twelve syllables obtained with (a) 2-channel combination, (b) 3-channel combination, and (c) 4-channel combination.

Table 3. Statistical values of classification accuracies from various number of channel combinations.

The number of channels	Classification accuracy (%)			
	Mean	*SD*	*Min*	*Max*
2	87.05	1.04	81.45	89.23
3	94.77	0.71	93.67	95.93
4	97.81	0.50	97.42	98.04
5	99.01	0.44	98.54	99.65

REFERENCES

[1] E. L. Larraz, O. M. Mozos, J. M. Antelis, and J. Minguez. "Syllable-based speech recognition using EMG," in *Proceedings of the Annual Conference of the IEEE / EMBC*, Buenos Aires, Argentina, Nov., 2010.

[2] N. Srisuwan, P. Phukpattaranont., and C. Limsakul,. "Comparison of feature evaluation criteria for speech recognition based on electromyography," *Medical and Biological Engineering and Computating*, pp. 1-11, 2017.

[3] T. Tsuji, N. Bu, J. Arita, and M. Ohga, "A speech synthesizer using facial EMG signals," *Inter. J. Computational Intelligence and Application*, vol.7, no. 1, pp. 1-15, 2008.

[4] M. Lyu, C. Xiong, and Q. Zhang. "Electromyography (EMG)-based Chinese voice command recognition," in *Proceedings of IEEE International Conference on Information and Automation*, pp. 926–931, Jul. 2014.

[5] G. Colby, J. T. Heaton, L. D. Gilmore, J. Sroka, Y. Deng, J. Cabrera, S. Roy, C. J. De Luca, and G. S. Meltzner. "Sensor subset selection for surface electromyography based speech recognition." in *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 473-476, 2009.

[6] T. Pothirat, " Characterization and analysis of neck and face surface electromyography for speech rehabilitation training,"M.S. thesis, Dept. Sci. in Biomed. Eng., Prince of Songkla Univ.,Songkhla, Thailand, 2014.

[7] K. Englehart, and B. Hudgins. "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol.50, no.7, pp. 848–854, 2003.

[8] Y. C. Du, C. H. Lin, L. Y. Shyu, and T. Chen, "Portable hand motion classifier for multi-channel surface electromyography recognition using grey relation analysis," *Expert Syst Appl*, vol.37, pp. 4283-4291, 2010.

[9] A. D. C. Chan, K. Englehart, B. Hudgins, and D. F. Lovely, "Hidden Markov model classification of myoelectric signals in speech," *IEEE Eng. in Med. Bio. Mag.*, pp. 143-146, 2002.

[10] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst Appl*, vol.39, pp. 7420-7431, 2012.

[11] S. Thongpanja, A. Phinyomark, F. Quaine, Y. Laurillau, C. Limsakul, and P. Phukpattaranont, "Probability density functions of stationary surface EMG signals in noisy environments. *IEEE Trans. Instru. Meas.*, vol.65, pp. 1547-1557, 2016.

[12] A. Sankarasubramanian, and K. Srinivasan, "Investigation and comparison of sampling properties of L-moments and conventional moments," *J. Hydrology*, vol.218, pp. 13-34, 1999.

[13] K. Anam, and A. Al-Jumaily, "A novel extreme learning machine for dimensionality reduction on finger movement classification using sEMG," in *Proceeding of the 7th IEEE conference in Neural Engineering*, pp. 824-827, 2015.

Table 2. The mathematical definitions of features.

Feature	Definition
Mean absolute value (MAV) [7]	$\mathrm{MAV} = \sum_{i=1}^{N} \lvert x_i \rvert$
Waveform length (WL) [8]	$\mathrm{WL} = \sum_{i=1}^{N-1} \lvert x_{i+1} - x_i \rvert$
Zero crossing (ZC) [8]	$\mathrm{ZC} = \sum \left[\mathrm{sgn}(x_i \times x_{i+1}) \cap \lvert x_i - x_{i+1} \rvert \geq \mathrm{threshold} \right]$ $\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq \mathrm{threshold} \\ 0, & \text{other} \end{cases}$, threshold = 10 μV
Slope sign change (SSC) [8]	$\mathrm{SSC} = \sum_{i=2}^{N-1} \left[f\left[(x_i - x_{i-1}) \times (x_i - x_{i+1}) \right] \right]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \mathrm{threshold} \\ 0, & \text{otherwise} \end{cases}$, threshold = 16 μV
Autoregressive coefficient (AR) [9]	$x_i = \sum_{p=1}^{P} a_p x_{i-p} + w_p$, order = 4
Mean frequency (MNF) [10]	$\mathrm{MNF} = \sum_{j=1}^{M} f_j P_j \Big/ \sum_{j=1}^{M} P_j$
L-kurtosis (L-KURT) [11]	$\mathrm{L-KURT} = \frac{20b_3 - 30b_2 + 12b_1 - b_0}{2b_1 - b_0}$
Skewness (SKW) [12]	$\mathrm{L-SKW} = \frac{60b_2 - 6b_1 + b_0}{2b_1 - b_0}$

The variables of feature: x_i is the i th sEMG amplitude and N denotes the length of the sEMG data. w_p is the weighted window function. P is the order of the AR model. a_p is the coefficients of the AR model. f_j is the frequency of spectrum at frequency bin j. P_j is the sEMG power spectrum at frequency bin j. M is the number of total power spectrum. b_r is an unbiased estimator of the probability weight moment.

ประวัติผู้เขียน

ชื่อ สกุล นางสาวนิดา แซ่จอง

รหัสประจำตัวนักศึกษา 5710130014

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตร์บัณฑิต (สาขาวิศวกรรมไฟฟ้า)	มหาวิทยาลัยสงขลานครินทร์	2544
วิศวกรรมศาสตร์มหาบัณฑิต (สาขาวิศวกรรมไฟฟ้า)	มหาวิทยาลัยสงขลานครินทร์	2553

ทุนการศึกษา

ทุนอุดหนุนการวิจัยเพื่อวิทยานิพนธ์ บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์

โครงการพัฒนาอาจารย์และบุคลากรสำหรับสถาบันอุดมศึกษาในเขตพัฒนาเฉพาะกิจจังหวัดชายแดนภาคใต้ สำนักงานคณะกรรมการอุดมศึกษา ที่กรุณาสนับสนุนทุน การศึกษาและการทำวิจัย

การตีพิมพ์เผยแพร่ผลงาน

N. Sae Jong and P. Phukpattaranont, “A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A thai syllable study,” Biocybernetics and Biomedical Engineering, vol. 39, no. 1, pp. 234–245, 2019.

N. Sae Jong, M. Kiatweerasakul, P. Phukpattaranont, “Channel reduction in speech recognition system based on surface electromyography,” 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 184-187, 18-21 July 2018, Chiang Rai, Thailand.