

การจำแนกวรรณยุกต์ไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ

นิวัติ ศรีสุวรรณ¹ พรชัย พฤษภัทรานนท์² และ ชูศักดิ์ ลิ้มสกุล³

มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ต.คอหงส์ อ.หาดใหญ่ จ.สงขลา 90110

บทคัดย่อ

งานวิจัยนี้นำเสนอเทคนิคการจำแนกวรรณยุกต์ไทย 5 เสียง สำหรับระบบรู้จำการพูดภาษาไทยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อจากกล้ามเนื้อคอเป็นอินพุตให้กับระบบ มีการเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้งหมด 10 คน โดยออกเสียงพูดคำโดดภาษาไทย 105 คำ ซึ่งครอบคลุมการออกเสียงวรรณยุกต์ทั้ง 5 เสียง สัญญาณที่ได้จะถูกนำมาผ่านกระบวนการเตรียมสัญญาณเบื้องต้น และนำไปคำนวณค่าลักษณะเด่นของสัญญาณไฟฟ้ากล้ามเนื้อจำนวน 160 ค่า จากนั้นทำการประเมินลักษณะเด่นดังกล่าวโดยใช้ตัวประเมินอาร์อีเอส (RES) และเลือกลักษณะเด่นจากผลการประเมินเพื่อเป็นอินพุตให้กับตัวจำแนก โดยปรับจำนวนอินพุตเป็น 3 5 7 9 12 15 18 21 24 และ 27 ตามลำดับ ในการฝึกสอนและทดสอบระบบจะใช้เทคนิคการจำแนกสองรูปแบบ คือ การจำแนกแบบสามขั้นตอนโดยใช้โครงข่ายประสาทเทียมสามโครงข่าย และการจำแนกแบบขั้นตอนเดียวโดยใช้โครงข่ายประสาทเทียมหนึ่งโครงข่าย จากผลการทดลองพบว่ารูปแบบการจำแนกแบบสามขั้นตอนมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว โดยให้ความแม่นยำสูงสุด 95.56% จากสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S04 เมื่อใช้จำนวนลักษณะเด่น 21 ตัว

คำสำคัญ : ระบบรู้จำการพูด / สัญญาณไฟฟ้ากล้ามเนื้อ / วรรณยุกต์ไทย

* Corresponding Author : chusak.l@psu.ac.th

¹ นักศึกษาปริญญาเอก ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

² รองศาสตราจารย์ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

³ รองศาสตราจารย์ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

Classification of Thai Tones using Electromyography Signals

Niyawadee Srisuwan¹ Pornchai Phukpattaranont² and Chusak Limsakul^{3*}

Prince of Songkla University, Hat Yai Campus, Karnjanavanit Road, Kho Hong, Hat Yai, Songkhla 90110

Abstract

This paper proposes the technique for classifying five Thai tones using Electromyography (EMG) signals. Two channels of EMG signals were captured from the neck muscles while a subject spoke 105 Thai isolated words including five tones. The EMG signals were collected from a total number of 10 subjects. After EMG capturing and pre-processing, 160 features, including time domain and frequency domain features were calculated. These features were evaluated by RES (ratio of Euclidean and Standard deviation) criterion. The features with high RES values were selected. They were used as an input of Artificial Neural Network (ANN) classifier, in which the number of inputs were adjusted to 3, 5, 7, 9, 12, 15, 18, 21, 24, and 27. For training and testing processes, we used two pattern classification techniques. One was the single-step classification using one ANN. The other was the three-step classification using three ANNs. Results showed that the three-step classification exhibited a better performance. The maximal classification accuracy was 95.56% from subject S04 when 21 features were used.

Keywords : Speech Recognition / Electromyography / Thai Tone

* Corresponding Author : chusak.l@psu.ac.th

¹ Doctoral Candidate, Department of Electrical Engineering, Faculty of Engineering.

² Associate Professor, Department of Electrical Engineering, Faculty of Engineering.

³ Associate Professor, Department of Electrical Engineering, Faculty of Engineering.

1. บทนำ

ระบบรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อถูกพัฒนามาจากระบบรู้จำการพูดแบบเดิม (conventional automatic speech recognition, ASR) ซึ่งใช้สัญญาณเสียงพูดเป็นข้อมูลเข้าหรืออินพุต (input) ของระบบ โดยระบบใหม่นี้จะใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ (Electromyography, EMG) ที่เกี่ยวข้องกับการออกเสียงพูดมาเป็นอินพุตให้กับระบบแทน หรืออาจจะใช้ทั้งสองสัญญาณร่วมกัน ซึ่งจากการทบทวนวรรณกรรมพบว่าสัญญาณไฟฟ้ากล้ามเนื้อดังกล่าวสามารถใช้เป็นตัวแทนของสัญญาณเสียงพูดได้

สาเหตุที่ระบบรู้จำการพูดแบบใหม่นี้ถูกพัฒนาขึ้นมา มีสาเหตุหลักอยู่ 4 ประการ ซึ่งในส่วนของเหตุผล 3 ประการแรกเป็นปัญหาที่เกิดขึ้นจากการใช้งานระบบรู้จำการพูดแบบเดิม ปัญหาแรกคือ ปัญหาการใช้งานระบบขณะที่มีเสียงรบกวนอย่างมาก ซึ่งทำให้สัญญาณเสียงที่ป้อนเข้าสู่ระบบถูกรบกวนด้วยสัญญาณรบกวนดังกล่าว ดังนั้นอินพุตของระบบจะผิดเพี้ยนและทำให้ระบบจำแนกข้อมูลผิดพลาด ปัญหาที่สองคือ เมื่อต้องการสื่อสารด้วยข้อมูลที่เป็นส่วนตัวหรือเป็นความลับ เช่น การประยุกต์ใช้งานทางด้านทหาร ธนาคาร เป็นต้น การใช้งานด้วยระบบรู้จำการพูดแบบเดิมจะทำให้ข้อมูลอาจถูกเปิดเผย ซึ่งจะส่งผลกระทบต่อความปลอดภัยของข้อมูลและผู้ใช้งานได้ สำหรับปัญหาที่สามของการใช้งานระบบรู้จำการพูดแบบเดิม คือ ระบบจะไม่สามารถใช้งานได้เมื่อถูกใช้งานในสถานการณ์ที่การสื่อสารด้วยเสียงพูดเป็นไปได้อย่าง เช่น ขณะที่ผู้พูดสวมชุดซึ่งมีหน้ากากปกปิดใบหน้า ได้แก่ ชุดดับเพลิง ชุดอวกาศ ชุดป้องกันสารเคมี เป็นต้น นอกจากนี้ระบบรู้จำการพูดแบบใหม่ที่ถูกพัฒนาขึ้นยังสามารถนำมาประยุกต์ใช้เป็นเครื่องมือช่วยในการสื่อสารสำหรับผู้ป่วยหรือผู้สูงอายุที่มีปัญหาเกี่ยวกับการออกเสียงพูด เช่น การได้รับการผ่าตัดมะเร็งกล่องเสียง การได้รับความกระทบกระเทือนทางสมองจนเกิดปัญหากับระบบประสาทที่เกี่ยวข้องกับกลไกการพูด เป็นต้น ผู้ป่วยเหล่านี้ไม่สามารถพูดได้อย่างปกติเหมือนคนทั่วไป ถึงแม้ว่าจะมีอุปกรณ์หรือเทคนิคที่จะช่วยให้ผู้ป่วยสามารถพูดได้ เช่น วิธีการพูดโดยใช้อากาศจากกระเพาะอาหาร (esophageal speech) การใช้เครื่องช่วยพูดอิเล็กทรอนิกส์ (electrolarynx)

เป็นต้น แต่อุปกรณ์และเทคนิคเหล่านี้ยังไม่มีประสิทธิภาพเพียงพอ

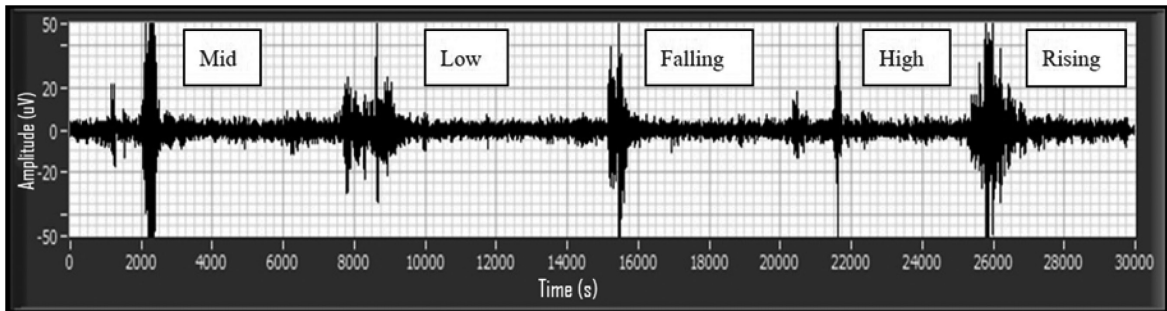
ระบบรู้จำการพูดแบบใหม่นี้ถูกพัฒนาขึ้นโดยนักวิจัยหลายกลุ่ม เริ่มต้นเมื่อปี ค.ศ. 1985 นักวิจัยชาวญี่ปุ่น Sugie และ Tsunoda [1] ได้สังเกตเห็นถึงปัญหาของการเพิ่มขึ้นของผู้ป่วยที่มีความบกพร่องของระบบการทำงานด้านการสร้างเสียงพูด จึงได้พัฒนาอุปกรณ์การสร้างเสียงเทียมให้กับผู้ป่วยโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อซึ่งได้จากการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อรอบปาก เพื่อใช้ในการจำแนกสระภาษาญี่ปุ่น 5 เสียง ซึ่งกล่าวได้ว่างานวิจัยนี้เป็นต้นแบบของแนวคิดการพัฒนากระบวนรู้จำการพูดโดยใช้สัญญาณไฟฟ้ากล้ามเนื้อ ในเวลาต่อมางานวิจัยดังกล่าวถูกพัฒนาอย่างต่อเนื่องโดยหลากหลายกลุ่มนักวิจัย โดยในช่วงเริ่มแรกของงานวิจัย เป็นการรู้จำสระ [1-3] คำพูดโดด [4] และ พยางค์ [5] ซึ่งเป็นการรู้จำจำนวนคำไม่มากนัก ดังนั้น ข้อจำกัดของงานวิจัยดังกล่าว คือ การประยุกต์ใช้งานถูกจำกัดอยู่ในวงแคบ เพื่อให้สามารถนำระบบนี้ไปใช้งานได้หลากหลายมากขึ้น ระบบดังกล่าวจึงถูกพัฒนาไปเป็นการรู้จำคำพูดต่อเนื่องซึ่งสามารถรู้จำคำพูดได้มากขึ้น โดยเริ่มขึ้นเป็นครั้งแรกเมื่อปี ค.ศ. 2006 โดย Jou และคณะ [6] งานวิจัยนี้สามารถสร้างระบบรู้จำคำพูดต่อเนื่องที่สามารถรู้จำคำศัพท์ภาษาอังกฤษได้ 108 คำและได้มีการพัฒนาอย่างต่อเนื่องจนถึงปัจจุบัน เมื่อเร็ว ๆ นี้ งานวิจัยของ Deng และคณะ [7] ก็เริ่มพัฒนาระบบให้เป็นการรู้จำคำพูดต่อเนื่องเช่นเดียวกัน ซึ่งสามารถรู้จำคำศัพท์ภาษาอังกฤษได้ถึง 202 คำ

งานวิจัยดังกล่าวมามีข้อจำกัดในด้านภาษาของผู้ใช้งาน กล่าวคือ ระบบถูกพัฒนาขึ้นสำหรับผู้ใช้งานสำหรับแต่ละภาษาโดยเฉพาะ ซึ่งความแตกต่างของภาษาที่หลากหลายมีผลทำให้ระบบรู้จำการพูดแบบใหม่นี้ถูกพัฒนาขึ้นในหลากหลายภาษา ได้แก่ ภาษาอังกฤษ [7-12] ภาษาญี่ปุ่น [2], [13-14] ภาษาจีน [15-16] ภาษาเกาหลี [17] ภาษาโปรตุเกส [18] ภาษาสเปน [5] ภาษาอารบิก [3] ภาษาเซอร์เบีย [19] เป็นต้น เช่นเดียวกับในบทความนี้ได้พัฒนาระบบดังกล่าวให้สามารถใช้ได้กับผู้ที่ใช้ภาษาไทยในการสื่อสาร โดยเฉพาะอย่างยิ่งเพื่อสามารถนำไปใช้งานกับผู้ป่วยหรือผู้สูงอายุที่มีปัญหาตามที่กล่าวไว้แล้วข้างต้น ซึ่งอาจจะไม่สามารถพูดภาษาอื่นได้นอกจากภาษาไทย

ภาษาไทยเป็นภาษาที่มีลักษณะเฉพาะซึ่งแตกต่างไปจากภาษาอื่น กล่าวคือ เป็นภาษาที่มีวรรณยุกต์ ซึ่งวรรณยุกต์คือ การเปลี่ยนแปลงของระดับเสียง คำที่มีวรรณยุกต์แตกต่างกันจะทำให้ความหมายของคำเปลี่ยนไปอย่างสิ้นเชิง ดังนั้นการออกแบบระบบเพื่อให้สามารถรู้จำวรรณยุกต์ได้จึงเป็นประเด็นสำคัญที่ถูกพิจารณาในขั้นแรกสำหรับงานวิจัยนี้

งานวิจัยที่ผ่านมา [20-21] ได้ทำการออกแบบระบบการ

จำแนกวรรณยุกต์ไทย 5 เสียง โดยได้ออกแบบตำแหน่งอิเล็กทรอนิกส์ที่เหมาะสมในการตรวจจับสัญญาณไฟฟ้ากล่ามเนื้อของเสียงวรรณยุกต์ดังกล่าว ซึ่งเป็นตำแหน่งของกล่ามเนื้อบริเวณข้างกล่องเสียง และกล่ามเนื้อบริเวณส่วนล่างสุดของลำคอ จากการพิจารณาลักษณะของสัญญาณไฟฟ้ากล่ามเนื้อของการพูดคำพูดที่มีวรรณยุกต์ทั้ง 5 เสียง พบว่ามีความแตกต่างกันอย่างเห็นได้ชัดแสดงดังรูปที่ 1



รูปที่ 1 รูปแบบของสัญญาณไฟฟ้ากล่ามเนื้อของการออกเสียงพูดคำพูดที่มีวรรณยุกต์ทั้ง 5 เสียง คือ สามัญ (mid tone) เอก (low tone) โท (falling tone) ตรี (high tone) และ จัตวา (rising tone) ซึ่งเป็นการออกเสียงคำว่า “ดู, ดู, ดู, ดู, ดู”

ซึ่งความแตกต่างของสัญญาณดังกล่าวทำให้ง่ายต่อการจำแนกเสียงวรรณยุกต์ไทย สำหรับงานวิจัยในภาษาอื่นๆ นั้นพบว่าไม่จำเป็นต้องใช้สัญญาณไฟฟ้ากล่ามเนื้อจากตำแหน่งดังกล่าว ซึ่งโดยทั่วไปจะใช้สัญญาณไฟฟ้าบริเวณกล่ามเนื้อหน้าและใต้คางก็เพียงพอแล้ว เนื่องจากกล่ามเนื้อบริเวณนี้ครอบคลุมการออกเสียงหน่วยเสียงทั้งหมดแล้ว (ยกเว้นเสียงวรรณยุกต์) อย่างไรก็ตาม สำหรับภาษาอื่นๆ เช่น ภาษาอังกฤษ เป็นภาษาที่ไม่มีเสียงวรรณยุกต์ แต่จะมีลักษณะการออกเสียงที่เน้นเสียงสูงเสียงต่ำซึ่งเกิดจากการเปลี่ยนแปลงของระดับเสียง เช่นเดียวกับเสียงวรรณยุกต์ อย่างไรก็ตามลักษณะของการเน้นเสียงสูงต่ำดังกล่าวไม่ได้ทำให้ความหมายของคำเปลี่ยนแปลง ดังนั้นจึงไม่จำเป็นต้องจำแนกในส่วนนี้

เมื่อได้ตำแหน่งที่เหมาะสมในการจำแนกวรรณยุกต์แล้ว สำหรับในกระบวนการจำแนก มีการออกแบบรูปแบบการจำแนกของตัวจำแนกโดยใช้เทคนิคการจำแนกแบบสามขั้นตอนซึ่งใช้โครงข่ายประสาทเทียมสามโครงข่ายซึ่งพบว่าระบบดังกล่าวสามารถจำแนกเสียงวรรณยุกต์

ได้ทั้ง 5 เสียง โดยที่มีความแม่นยำในการจำแนก 80% อย่างไรก็ตาม เทคนิคดังกล่าวถูกออกแบบโดยใช้ข้อมูลของสัญญาณไฟฟ้ากล่ามเนื้อจากอาสาสมัครเพียง 1 คน ดังนั้นเพื่อให้สามารถนำเทคนิคที่ออกแบบไปใช้ได้กับข้อมูลของสัญญาณไฟฟ้ากล่ามเนื้อจากอาสาสมัครที่ครอบคลุมมากขึ้น งานวิจัยนี้จึงนำเสนอการปรับปรุงเทคนิคการจำแนกวรรณยุกต์ โดยการออกแบบจะอ้างอิงข้อมูลของสัญญาณไฟฟ้ากล่ามเนื้อจากอาสาสมัคร 10 คน (S01-S10) เพื่อปรับปรุงรูปแบบการจำแนกแบบสามขั้นตอนให้สามารถนำไปใช้งานได้กับสัญญาณไฟฟ้ากล่ามเนื้ออื่นๆ โดยจะปรับปรุงในส่วนของการคัดเลือกลักษณะเด่น การปรับปรุงประสิทธิภาพในการจำแนกโดยเพิ่มจำนวนลักษณะเด่น การเลือกจำนวนลักษณะเด่นที่เหมาะสม นอกจากนี้ จะทำการเปรียบเทียบความแม่นยำที่ได้จากการจำแนกด้วยเทคนิคที่ได้ออกแบบ (การจำแนกแบบสามขั้นตอน) กับการจำแนกแบบดั้งเดิมซึ่งเป็นการจำแนกแบบขั้นตอนเดียวโดยใช้โครงข่ายประสาทเทียมหนึ่งโครงข่าย

2. วัสดุ อุปกรณ์ และ วิธีการวิจัย

2.1 วัสดุ อุปกรณ์

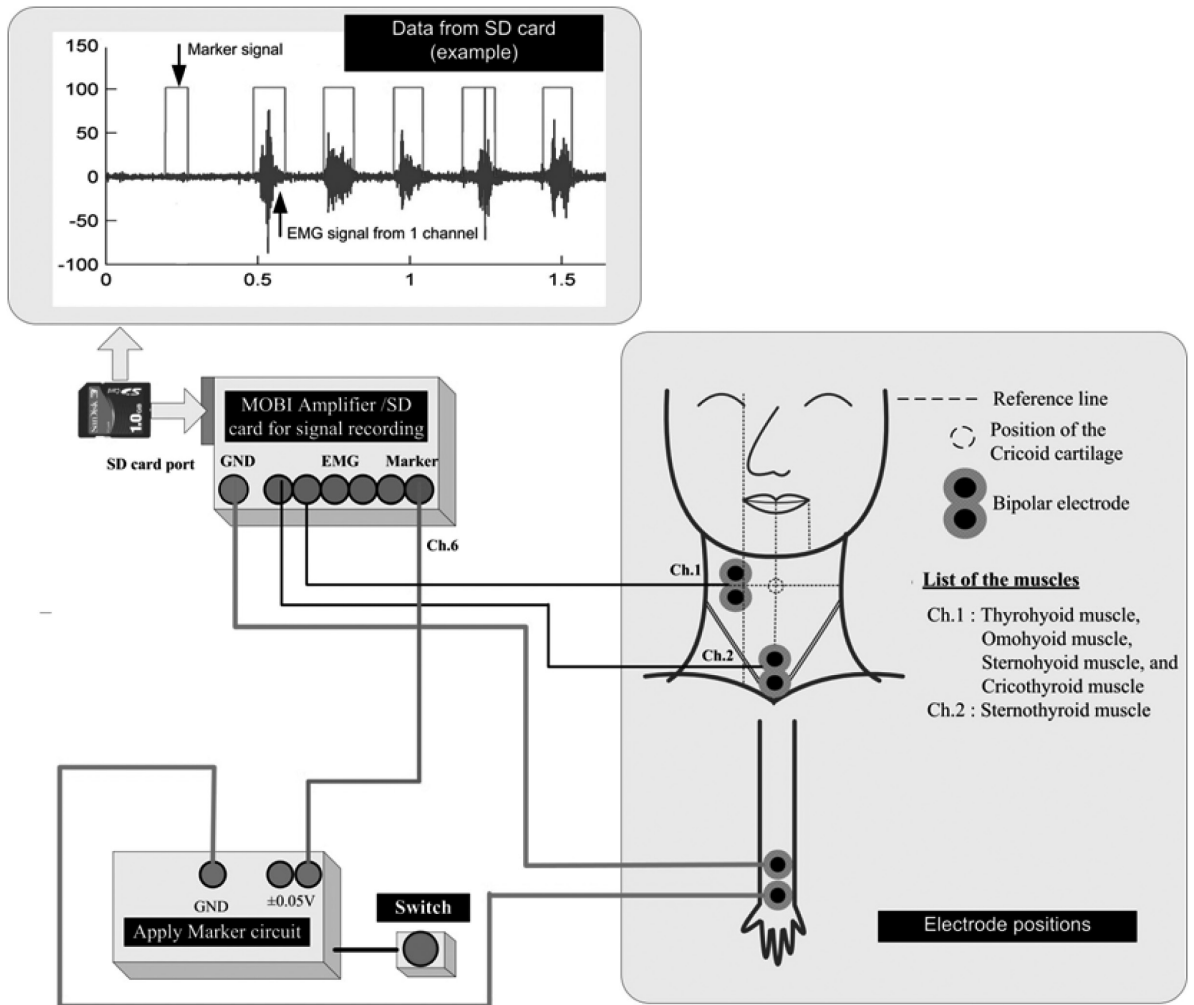
วัสดุและอุปกรณ์ที่ใช้ในการทดลองประกอบด้วย เครื่องวัดสัญญาณ MOBIO6 ซึ่งผลิตโดยบริษัท TMS international ประเทศเนเธอร์แลนด์ ใช้สำหรับการบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ อิเล็กโทรดแบบผิวหน้าชนิด Ag/AgCl ยี่ห้อ ARBO มีขนาดเส้นผ่านศูนย์กลางภายนอก 2.4 เซนติเมตร และขนาดเส้นผ่านศูนย์กลางภายใน (ส่วนของเจล) 1 เซนติเมตร ซึ่งผลิตโดยบริษัท Tyco healthcare Group LP ประเทศเยอรมัน ใช้สำหรับตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ และอุปกรณ์สำหรับประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อ ซึ่งก็คือ คอมพิวเตอร์และโปรแกรม ซึ่งโปรแกรมที่ใช้ในการตรวจวัดสัญญาณไฟฟ้ากล้ามเนื้อ คือ Portilab และ โปรแกรมที่ใช้ในการประมวลผลสัญญาณไฟฟ้ากล้ามเนื้อ คือ MATLAB นอกจากนี้ยังมีวงจรสร้างสัญญาณการกวดสวิตช์เพื่อสร้างสัญญาณสำหรับการกำหนดขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด ซึ่งสัญญาณส่วนนี้จะถูกต่อเข้ากับเครื่องวัดสัญญาณ MOBIO6

2.2 วิธีการวิจัย

2.2.1 การเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อ

รูปที่ 2 แสดงแผนภาพการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อโดยเริ่มจากการติดอิเล็กโทรดลงบนตำแหน่งของผิวหนังที่ใช้ในการตรวจจับสัญญาณ

ไฟฟ้ากล้ามเนื้อของเสียงพูดวรรณยุกต์ ซึ่งเป็นตำแหน่งของกล้ามเนื้อคอ 2 ตำแหน่ง (2 ช่องสัญญาณ) ที่ได้ถูกออกแบบไว้แล้ว [20-21] ทำการตรวจวัดสัญญาณดังกล่าวขณะที่อาสาสมัครออกเสียงพูด โดยใช้เครื่องวัดสัญญาณ MOBIO6 ซึ่งมีทั้งหมด 6 ช่องสัญญาณ สำหรับในการทดลองนี้จะใช้เพียง 3 ช่องสัญญาณ คือ ช่องสัญญาณที่ 1 (Ch.1) และ ช่องสัญญาณที่ 2 (Ch.2) จะถูกใช้ในการตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อ และช่องสัญญาณที่ 6 (Ch.6) จะถูกใช้สำหรับตรวจวัดสัญญาณการกวดสวิตช์ (Marker) ซึ่งสัญญาณการกวดสวิตช์จะถูกใช้สำหรับการกำหนดจุดเริ่มต้นของการบันทึกสัญญาณไฟฟ้ากล้ามเนื้อ และการกำหนดขอบเขตของสัญญาณไฟฟ้ากล้ามเนื้อของคำพูดแต่ละเสียง จะมีการกวดสวิตช์เมื่อเริ่มพูดและปล่อยสวิตช์เมื่อพูดคำพูดในแต่ละคำจบ ซึ่งทำให้สามารถระบุได้ว่าสัญญาณที่อยู่ในช่วงการกวดสวิตช์เป็นสัญญาณที่เกิดจากการพูดจริง ไม่ใช่สัญญาณอื่นๆ เช่น สัญญาณที่เกิดจากการไอ จาม กลืนน้ำลาย หรือการขยับกล้ามเนื้อบริเวณที่จับสัญญาณโดยไม่ได้ตั้งใจ เป็นต้น สัญญาณการกวดสวิตช์จะถูกบันทึกไปพร้อมกับสัญญาณไฟฟ้ากล้ามเนื้ออีก 2 ช่องสัญญาณ ซึ่งทั้งสองสัญญาณจะเป็นสัญญาณที่ขนานกัน โดยความถี่ในการซักรับตัวอย่างของสัญญาณทั้ง 3 ช่องสัญญาณกำหนดไว้ 1024 เฮิร์ตซ์ โดยข้อมูลของสัญญาณทั้งหมดจะถูกบันทึกลง SD card ภายใน Mobio6 หลังจากนั้นจะทำการดึงข้อมูลของสัญญาณโดยใช้โปรแกรม Portilab



รูปที่ 2 แสดงลักษณะการเก็บข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อ และแสดงตำแหน่งอิเล็กโทรดสำหรับตรวจจับสัญญาณไฟฟ้ากล้ามเนื้อสำหรับการทดลองนี้

ในการทดลองนี้อาสาสมัครจะออกเสียงพูดคำโดดภาษาไทยทั้งหมด 105 คำ ซึ่งประกอบด้วยคำพูดที่มีโครงสร้างของพยัญชนะและสระเหมือนกัน 21 คำ โดย

แต่ละคำจะออกเสียงวรรณยุกต์ครบทั้ง 5 เสียง แสดงดังตารางที่ 1

ตารางที่ 1 คำพูดที่ใช้ในการฝึกสอนและทดสอบระบบ 21 ชุดคำพูด (105 คำ)

วรรณยุกต์ คำ	สามัญ	เอก	โท	ตรี	จัตวา	หมายเหตุ
1	กา	ก่า	ก้า	ก๊า	ก๋า	อักษรกลาง
2	โจ	โจ่	โจ้	โจ๊	โจ๋า	
3	ดู	ดู๋	ดู๊	ดู็	ดู๋า	
4	เบ	เบ่	เบ้	เบ๊	เบ๋า	
5	ออ	ออ่	ออ้	ออ๊	ออ๋า	
6	ตี	ตี๋	ตี้	ตี๊	ตี๋า	
7	เปอ	เป่อ	เป้อ	เป็อ	เป๋า	
8	เพอ	เพ่อ	เพ้อ	เพ็อ	เพ๋า	อักษรต่ำ
9	ฟา	ฟา่	ฟา้	ฟา๊	ฟา๋า	
10	ทอ	ทอ่	ทอ้	ทอ๊	ทอ๋า	
11	คู	คู๋	คู้	คู็	คู๋า	
12	เชื่อ	เชื่อ่	เชื่อ้	เชื่อ๊	เชื่อ๋า	
13	โฮ	โฮ่	โฮ้	โฮ๊	โฮ๋า	
14	แซร์	แซร์	แซร์	แซร์	แซร์	
15	งู	งู๋	งู้	งู๊	งู๋า	อักษรสูง
16	ยา	ยา่	ยา้	ยา๊	ยา๋า	
17	นี่	นี่๋	นี่้	นี่๊	นี่๋า	
18	เรือ	เรือ่	เรือ้	เรือ๊	เรือ๋า	
19	ว้าว	ว้าว	ว้าว	ว้าว	ว้าว	
20	มือ	มือ	มือ	มือ	มือ	
21	ล่อ	ล่อ	ล่อ	ล่อ	ล่อ	

แต่ละคำพูดจะถูกพูดซ้ำ 9 ครั้ง โดยทำการเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้งหมด 10 คน อายุระหว่าง 18-50 ปี โดยมีจำนวนเพศชายและเพศหญิงเท่ากัน ดังนั้นจะได้ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด 945 ข้อมูล (105 คำ x 9 ครั้ง) สำหรับอาสาสมัครแต่ละคน

2.2.2 การเตรียมสัญญาณเบื้องต้น

- การกำจัดสัญญาณรบกวน
สัญญาณไฟฟ้ากล้ามเนื้อที่ได้จากกระบวนการเก็บข้อมูลมีสัญญาณรบกวนปนอยู่มาก ได้แก่ สัญญาณรบกวน 50 เฮิรตซ์ สัญญาณรบกวนจากการเคลื่อนที่ (motion artifact) และ สัญญาณคลื่นไฟฟ้าหัวใจ (Electrocardiography, ECG) โดยเทคนิคที่ใช้ในการกำจัด

สัญญาณรบกวน จะใช้ตัวกรองความถี่แถบผ่าน (band pass filter) ซึ่งมีความถี่ตัดผ่าน (cut off frequency) อยู่ในช่วง 30-350 เฮิรตซ์ ซึ่งสามารถกำจัดสัญญาณรบกวนจากการเคลื่อนที่ และลดทอนสัญญาณรบกวนคลื่นไฟฟ้าหัวใจได้ โดยตัวกรองความถี่แถบผ่านที่ใช้ คือ ตัวกรองชนิดบัตเตอร์เวิร์ธ (Butterworth) อันดับที่ 5 ซึ่งสามารถกำจัดสัญญาณรบกวนคลื่นไฟฟ้าหัวใจได้มากที่สุด นอกจากนี้ ตัวกรองนอตช์ (Notch filter) จะถูกใช้สำหรับกำจัดสัญญาณรบกวน 50 เฮิรตซ์

- การตรวจจับจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณ

หลังจากผ่านการกรองสัญญาณรบกวนแล้ว สัญญาณไฟฟ้ากล้ามเนื้อจะถูกนำมาทำการวิเคราะห์หาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อขณะพูด โดยเริ่มจากการตัดสัญญาณไฟฟ้ากล้ามเนื้อของ

แต่ละคำพูดออกมาก่อน ซึ่งไม่ได้ตัดมาจากจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อที่แท้จริง แต่จะตัดมาจากการใช้สัญญาณจากการกดสวิทช์เป็นตัวกำหนดขอบเขต โดยจะดึงสัญญาณมา 2500 ตัวอย่าง (samples) จากขอบเขตดังกล่าว หลังจากนั้นจะนำสัญญาณที่ถูกตัดไปวิเคราะห์หาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อที่แท้จริงของคำพูดนั้นๆ โดยใช้วิธีที่เรียกว่า Three standard deviation threshold [22] ซึ่งเป็นการคำนวณหาค่าขีดแบ่ง (threshold) เพื่อใช้ในการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณไฟฟ้ากล้ามเนื้อที่แท้จริง โดยคำนวณจากค่าผลรวมของค่าเฉลี่ยและสามเท่าของค่าเบี่ยงเบนมาตรฐานของสัญญาณอ้างอิง ซึ่งสัญญาณอ้างอิงคือ สัญญาณสัญญาณไฟฟ้ากล้ามเนื้อในขณะที่ไม่มีการออกเสียงพูดหรือขณะเงียบ



รูปที่ 3 ภาพจริงขณะทำการทดลองเก็บข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร ซึ่งบทความนี้จะใช้สัญญาณไฟฟ้ากล้ามเนื้อจากตำแหน่งกล้ามเนื้อคอ 2 ตำแหน่งเท่านั้น สำหรับสัญญาณจากตำแหน่งอื่นๆ จะไม่ถูกใช้ในบทความนี้

2.2.3 การคำนวณค่าลักษณะเด่นของสัญญาณ

สัญญาณไฟฟ้ากล่อมเนื้อของคำพูดแต่ละคำ หลังจากผ่านกระบวนการเตรียมสัญญาณเบื้องต้น จะถูกนำมาคำนวณค่าลักษณะเด่นของสัญญาณ (feature) เป็นการสกัดลักษณะสำคัญของสัญญาณ (feature extraction) เพื่อใช้เป็นตัวแทนของคำพูดนั้นๆ ในงานวิจัยนี้ลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อจะถูกคำนวณโดยใช้สมการที่ (1)-(15) ซึ่งถูกนำมาใช้กันอย่างแพร่หลายในงานวิจัยทางด้านนี้ และมีความครอบคลุมลักษณะเด่นทั้งในแกนเวลาและในแกนความถี่

กำหนดให้ x_n คือ ข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อที่คอมพิวเตอร์ทำการซักรตัวอย่าง (sampling) เข้ามาครั้งที่ n โดยที่ N คือ จำนวนข้อมูลทั้งหมด และกำหนดให้ P_i คือ กำลังไฟฟ้าสเปกตรัม (power spectrum) ค่าที่ i และ f_i คือ ความถี่ที่ตำแหน่ง i และ M คือ $N/2-1$ การคำนวณลักษณะเด่นมีรายละเอียดดังต่อไปนี้

- ค่ารากเฉลี่ยกำลังสอง (Root Mean Square, RMS)

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

- ค่าความแปรปรวน (Variance, VAR)

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

- ค่าความยาวรูปร่างคลื่น (Waveform Length, WL)

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \quad (3)$$

- ค่าแอมพลิจูดของวิลลิสัน (Willison Amplitude, WAMP)

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

$$f(x) = 1 \text{ if } x \geq TH \text{ (threshold),} \quad (4)$$

$$f(x) = 0 \text{ otherwise}$$

- ค่าจุดตัดศูนย์ (Zero Crossing, ZC)

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

$$\text{sgn}(x) = 1 \text{ if } x \geq 0, \text{sgn}(x) = 0 \text{ otherwise} \quad (5)$$

- ค่าการเปลี่ยนแปลงความชัน (Slope Sign Change, SSC)

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

$$f(x) = 1 \text{ if } x \geq TH, f(x) = 0 \text{ otherwise} \quad (6)$$

- ค่าผลรวมของค่าสัมบูรณ์ (Integrated Absolute Value, IAV)

$$IAV = \sum_{n=1}^N |x_n| \quad (7)$$

- ค่าเฉลี่ยของค่าสัมบูรณ์ (Mean Absolute Value, MAV)

$$MAV = \frac{1}{N} \sum_{n=1}^N |x_n| \quad (8)$$

- ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 1 (Modified Mean Absolute Value, MMAV1)

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

$$w_n = 1 \text{ if } 0.25N \leq n \leq 0.75N, \quad (9)$$

$$w_n = 0.5 \text{ otherwise}$$

- ค่าเฉลี่ยของค่าสัมบูรณ์แบบให้น้ำหนักแบบที่ 2 (Modified Mean Absolute Value, MMAV2)

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

$$w_n = 1 \text{ if } 0.25N \leq n \leq 0.75N, \quad (10)$$

$$w_n = 4n/N \text{ if } 0.25N > n,$$

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

- ค่าเฉลี่ยของสัญญาณ (Mean, MN)

$$MN = \frac{1}{N} \sum_{n=1}^N x_n \quad (11)$$

- ค่าความชันของค่าเฉลี่ยค่าสัมบูรณ์ (Mean Absolute Value Slope, MAVS)

$$MAVS_i = MAV_{i+1} - MAV_i, \quad i = 1, 2, 3, \dots, s-1 \text{ เมื่อ } s = 6 \quad (12)$$

โดยที่ S คือ จำนวนการแบ่งสัญญาณ-1

- ค่าความถี่กลาง (Median Frequency, MDF)

$$\sum_{i=1}^{MDF} P_i = \sum_{i=MDF}^M P_i = \frac{1}{2} \sum_{i=1}^M P_i \quad (13)$$

- ค่าเฉลี่ยของความถี่ (Mean Frequency, MNF)

$$MNF = \frac{\sum_{i=1}^M f_i P_i}{\sum_{i=1}^M P_i} \quad (14)$$

- ค่าสเปกตรัลโมเมนต์ (Spectral Moment, SM)

$$SM_n = \sum_{i=1}^M f_i^n P_i \quad (15)$$

$$n = 1, 2, 3, 5$$

เมื่อ n คืออันดับของสเปกตรัลโมเมนต์

จากสมการดังกล่าว การคำนวณลักษณะเด่นแต่ละชนิดจะให้ค่าพารามิเตอร์ 1 ตัว ยกเว้นลักษณะเด่นชนิด MAVS จากสมการ (12) จะให้ค่าพารามิเตอร์ 5 ตัว ซึ่งถูกเรียกว่า MAVS_1 MAVS_2 MAVS_3 MAVS_4 และ MAVS_5 สำหรับลักษณะเด่นชนิด SM จากสมการที่ (15) จะเป็นการคำนวณสเปกตรัลโมเมนต์ 4 อันดับ คือ 1 2 3 และ 5 ซึ่งจะได้ลักษณะเด่น 4 ตัว คือ SM1 SM2 SM3 และ SM5

รูปที่ 4 แสดงการคำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อ ซึ่งแบ่งออกเป็น 2 รูปแบบ คือ

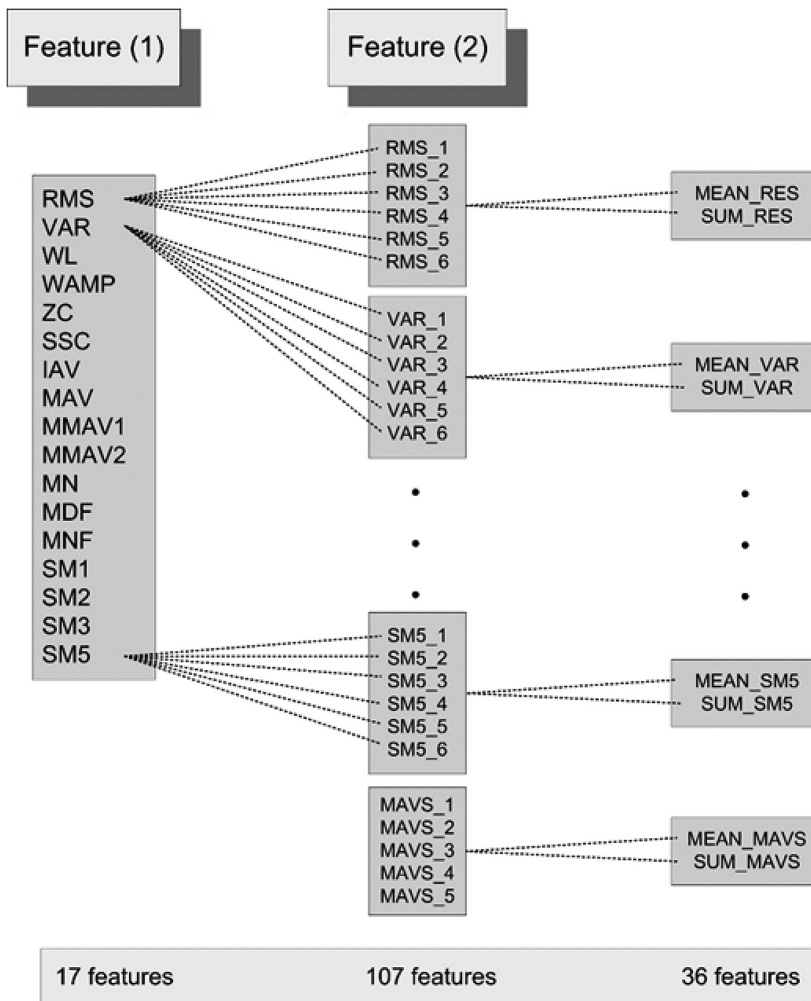
รูปแบบที่ 1) คำนวณโดยใช้ข้อมูลของสัญญาณทั้งหมด แต่ละสัญญาณจะมีความยาวเท่ากับ N โดยใช้สมการที่ (1)-(15) ยกเว้นสมการที่ (12) ซึ่งจะทำให้ได้ลักษณะเด่น 17 ตัว

รูปแบบที่ 2) คำนวณโดยแบ่งสัญญาณไฟฟ้ากล่อมเนื้อออกเป็น 6 ส่วนเท่าๆ กัน ซึ่งแต่ละส่วนจะมีความยาวของสัญญาณเท่ากับ $N/6$ แล้วนำสัญญาณในแต่ละส่วนมาคำนวณโดยใช้สมการที่ (1)-(15) (เทคนิคที่ประยุกต์จากการคำนวณลักษณะเด่นชนิด MAVS) ค่าลักษณะเด่นในแต่ละส่วนจะถูกเรียกใหม่ โดยใช้ตัวเลขซึ่งอยู่ด้านหลังเครื่องหมายสัญญาณประกาศ (underscore) เป็นตัวบ่งชี้ส่วนของสัญญาณที่นำมาคำนวณ เช่น RMS_1 คือ ลักษณะเด่นชนิด RMS ของสัญญาณส่วนที่หนึ่ง เป็นต้น จากการคำนวณสมการ (12) สำหรับทั้ง 6 ส่วนของสัญญาณ จะได้ลักษณะเด่นตัวใหม่ 5 ตัว สำหรับสมการอื่นๆ จะได้ลักษณะเด่นชนิดละ 6 ตัว ($6 \times 17 = 102$) ในส่วนนี้จะได้ลักษณะเด่นทั้งหมด 107 ตัว

นอกจากนั้น ลักษณะเด่นที่ได้จากการคำนวณในรูปแบบที่ 2) จะถูกนำไปหาค่าผลรวมและค่าเฉลี่ยของลักษณะเด่นจากทั้ง 6 ส่วนของสัญญาณ ซึ่งจะทำให้ได้ลักษณะเด่นตัวใหม่ โดยในการเรียกชื่อลักษณะเด่นชนิดใหม่นี้จะใช้คำว่า “SUM” และ “MEAN” ตามด้วยเครื่องหมายสัญญาณประกาศหน้า เช่น SUM_RMS คือ ผลรวมของ RMS_1 ถึง RMS_6 เป็นต้น ในส่วนนี้จะได้ลักษณะเด่นทั้งหมด 36 ตัว

เทคนิคการคำนวณลักษณะเด่นโดยแบ่งสัญญาณออกเป็นส่วนๆ แล้วคำนวณลักษณะเด่นในแต่ละส่วนของสัญญาณ จะทำให้สามารถวิเคราะห์ได้ว่า สัญญาณตรงส่วนใดมีผลในการจำแนกวรรณยุกต์ หรือทุกส่วนของสัญญาณมีผลในการจำแนกวรรณยุกต์

จากที่กล่าวมาข้างต้นในส่วนของการคำนวณลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อทั้งสองรูปแบบ จะได้จำนวนลักษณะเด่นทั้งหมด 160 ตัว



รูปที่ 4 ลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื่องจากการคำนวณทั้งสองรูปแบบ (Feature(1) และ Feature(2))

2.2.4 การประเมินลักษณะเด่น

ลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมดที่ถูกคำนวณจากหัวข้อ 2.2.3 จะถูกนำมาประเมินเพื่อวัดความสามารถในการแบ่งกลุ่มข้อมูลของลักษณะเด่นแต่ละตัว โดยใช้ตัวประเมินอาร์อีเอส (RES) ซึ่งแสดงดังสมการที่ (16) ซึ่งพบว่ามีประสิทธิภาพในการประเมินลักษณะเด่นของสัญญาณไฟฟ้ากล่อมเนื้อ ในทิศทางเดียวกับความแม่นยำในการจำแนกของตัวจำแนก [23]

$$RES = \frac{\overline{ED}}{\sigma} \quad (16)$$

โดยที่ \overline{ED} คือ ค่าเฉลี่ยของระยะทางแบบยูคลิด (Euclidean distance) ซึ่งเป็นการวัดระยะทางระหว่างกลุ่มข้อมูลโดยวัดจากจุดศูนย์กลาง (centroid) ของข้อมูลแต่ละกลุ่ม เพื่อวัดการกระจายของข้อมูลระหว่างกลุ่ม แล้วนำมาหาค่าเฉลี่ย

σ คือ ค่าเฉลี่ยของส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ซึ่งวัดการกระจายภายในกลุ่มข้อมูลแล้วนำมาหาค่าเฉลี่ย รายละเอียดการคำนวณถูกแสดงใน [23]

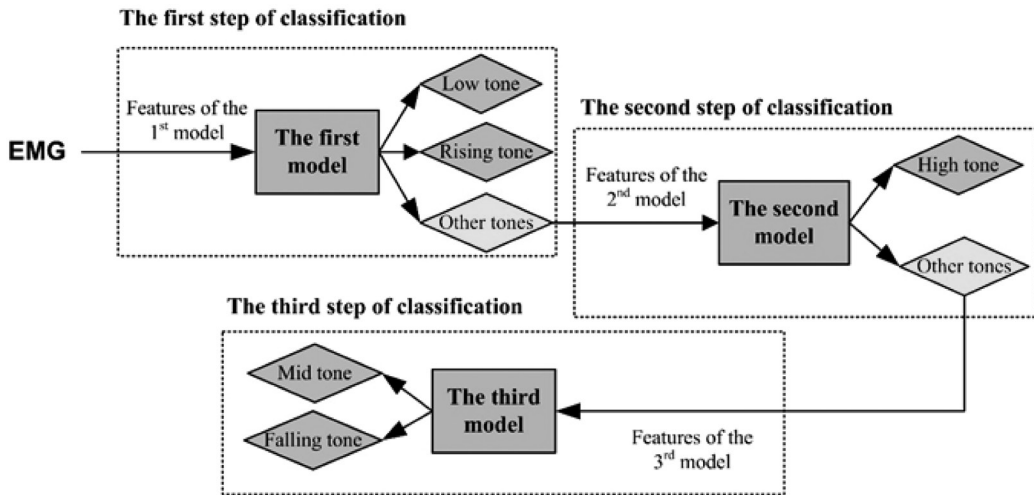
การประเมินลักษณะเด่นโดยใช้ตัวประเมินอาร์อีเอส ถูกแบ่งออกเป็น 2 รูปแบบ ดังนี้

รูปแบบที่ 1) การประเมินลักษณะเด่นสำหรับใช้ในระบบจำแนกแบบขั้นตอนเดียว (single step of classification) โดยประเมินความสามารถในการแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่ม ซึ่งเป็นกลุ่มของเสียงวรรณยุกต์ทั้ง 5 เสียง รูปแบบนี้จะจำแนกวรรณยุกต์ออกมา 5 เสียงภายในขั้นตอนเดียว

รูปแบบที่ 2) การประเมินลักษณะเด่นสำหรับใช้ในระบบจำแนกแบบสามขั้นตอน (three steps of classification) ดังแสดงในรูปที่ 5 รูปแบบนี้จะจำแนกวรรณยุกต์ออกมาทีละขั้นตอน โดยขั้นตอนแรกจะจำแนกวรรณยุกต์ออกมาสามกลุ่ม (โดยพิจารณาวรรณยุกต์ทั้ง 5 เสียง) คือ กลุ่มวรรณยุกต์เอก กลุ่มวรรณยุกต์จัตวา

และ กลุ่มวรรณยุกต์อื่นๆ (สามัญ โท ตรี) ขั้นตอนที่สองเป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่ม (โดยพิจารณาเฉพาะวรรณยุกต์สามัญ โท และ ตรี) คือ กลุ่มวรรณยุกต์ตรี และ กลุ่มวรรณยุกต์อื่นๆ (สามัญ โท) ขั้นตอนที่สามเป็นการจำแนกวรรณยุกต์ออกเป็น 2 กลุ่มเช่นเดียวกัน (โดยพิจารณาเฉพาะวรรณยุกต์สามัญ และ โท) คือ กลุ่มวรรณยุกต์สามัญ และ กลุ่มวรรณยุกต์โท

ดังนั้นในรูปแบบนี้จะแบ่งการประเมินออกเป็น 3 กลุ่มย่อยตามขั้นตอนของการจำแนก ซึ่งจะประเมินความสามารถในการแบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่ม 2 กลุ่ม และ 2 กลุ่ม ตามลำดับ



รูปที่ 5 แผนภาพแสดงรูปแบบการจำแนกแบบสามขั้นตอนในการจำแนกวรรณยุกต์ไทย 5 เสียง คือ สามัญ (mid tone) เอก (low tone) โท (falling tone) ตรี (high tone) และ จัตวา (rising tone)

2.2.5 การเลือกลักษณะเด่น

ขั้นตอนนี้เป็นการเลือกลักษณะเด่นเพื่อใช้เป็นอินพุตให้กับตัวจำแนก โดยทำการเลือกลักษณะเด่นในแต่ละรูปแบบการประเมิน โดยมีขั้นตอนในการเลือกลักษณะเด่น ดังต่อไปนี้

- 1) เลือกลักษณะเด่นที่มีค่าอาร์อีเอสสูงที่สุด 10 อันดับแรกสำหรับสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน (10 คน) ดังนั้นในขั้นตอนนี้จะได้ลักษณะเด่นทั้งหมด 100 ตัว
- 2) พิจารณาลักษณะเด่นทั้งหมด (100 ตัว) ที่

ถูกเลือกจากหัวข้อที่ 1) จากนั้น หาความถี่ของลักษณะเด่นแต่ละตัว ซึ่งความถี่สูงสุดเท่ากับ 10 นั่นคือ ข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครทั้ง 10 คน มีลักษณะเด่นตัวนั้นปรากฏอยู่ในกลุ่มของลักษณะเด่นที่มีอาร์อีเอสสูงที่สุด 10 อันดับแรกจากหัวข้อที่ 1)

3) เรียงลำดับลักษณะเด่นจากความถี่สูงสุดไปหาความถี่ต่ำสุด อย่างไรก็ตาม ในกรณีที่ความถี่ของลักษณะเด่นเท่ากัน จะเรียงลำดับโดยพิจารณาค่าอาร์อีเอสเฉลี่ยซึ่งคำนวณจากค่าเฉลี่ยของอาร์อีเอสของสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครทั้งหมด

การเลือกลักษณะเด่นจากวิธีดังกล่าว จะทำให้ได้ลักษณะเด่นเพียงชุดเดียว ที่สามารถนำไปใช้ได้กับสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้งหมด เนื่องจากระบวนการเลือกลักษณะเด่นข้างต้น เลือกจากการพิจารณาข้อมูลของสัญญาณไฟฟ้ากล่อมเนื้อจากอาสาสมัครทั้งหมด

2.2.6 กระบวนการจำแนก

ในกระบวนการนี้จะทำการฝึกสอน (training) และทดสอบระบบ (testing) โดยใช้โครงข่ายประสาทเทียมเป็นตัวจำแนก เพื่อจำแนกวรรณยุกต์ 5 เสียง โดยนำลักษณะเด่นที่ได้จากหัวข้อ 2.2.5 มาเป็นอินพุตให้กับตัวจำแนก โดยทำการทดลองปรับจำนวนอินพุตของโครงข่ายประสาทเทียม โดยใช้จำนวนลักษณะเด่น 3 5 7 9 12 15 18 21 24 และ 27 ตามลำดับ โดยเรียงลำดับลักษณะเด่นที่ดีที่สุดจากกระบวนการเลือกลักษณะเด่นในหัวข้อ 2.2.5

ในการทดลองนี้ จะทดสอบประสิทธิภาพของรูปแบบการจำแนกแบบสามชั้นตอนโดยใช้โครงข่ายประสาทเทียมสามโครงข่าย ซึ่งเป็นเทคนิคที่ได้นำเสนอในงานวิจัยที่ผ่านมา [21] และได้รับการปรับปรุงในบทความนี้ โดยจะเปรียบเทียบกับกรจำแนกแบบชั้นตอนเดียวซึ่งใช้โครงข่ายประสาทเทียมหนึ่งโครงข่าย (รูปแบบทั่วไป) ซึ่งได้กล่าวไว้เบื้องต้นในหัวข้อ 2.2.4

รูปที่ 5 แสดงรูปแบบการจำแนกแบบสามชั้นตอน โดยใช้โครงข่ายประสาทเทียมสามโครงข่าย ในแต่ละโครงข่ายจะใช้ลักษณะเด่นที่แตกต่างกัน โดยโครงข่ายที่หนึ่งจะใช้ในการจำแนกวรรณยุกต์เอกและจัตวา ในกรณีที่ผลการจำแนกออกมาเป็นวรรณยุกต์อื่นๆ ระบบจะส่งต่อไปที่โครงข่ายที่สองซึ่งใช้ในการจำแนกวรรณยุกต์ตรี ในกรณีที่ผลการจำแนกออกมาเป็นวรรณยุกต์อื่นๆ ระบบก็จะส่งต่อไปที่โครงข่ายที่สามซึ่งใช้ในการจำแนกวรรณยุกต์สามัญและโท รายละเอียดการทำงานของระบบถูกแสดงใน [21]

สัญญาณไฟฟ้ากล่อมเนื้อที่ใช้ในการทดสอบระบบ (testing) จะใช้ข้อมูลของชุดคำพูดที่ 1 2 8 9 และ 15 ในตารางที่ 1 สำหรับข้อมูลที่เหลือจะถูกใช้สำหรับกระบวนการฝึกสอนระบบ เพื่อเป็นการทดสอบว่าระบบที่ออกแบบสามารถจำแนกวรรณยุกต์ได้ กล่าวคือ ถึงแม้ว่า

จะออกเสียงคำพูดที่มีองค์ประกอบของพยัญชนะและสระต่างกัน แต่วรรณยุกต์เดียวกัน ระบบจะสามารถจัดกลุ่มคำนั้นให้อยู่ในกลุ่มวรรณยุกต์เดียวกันได้

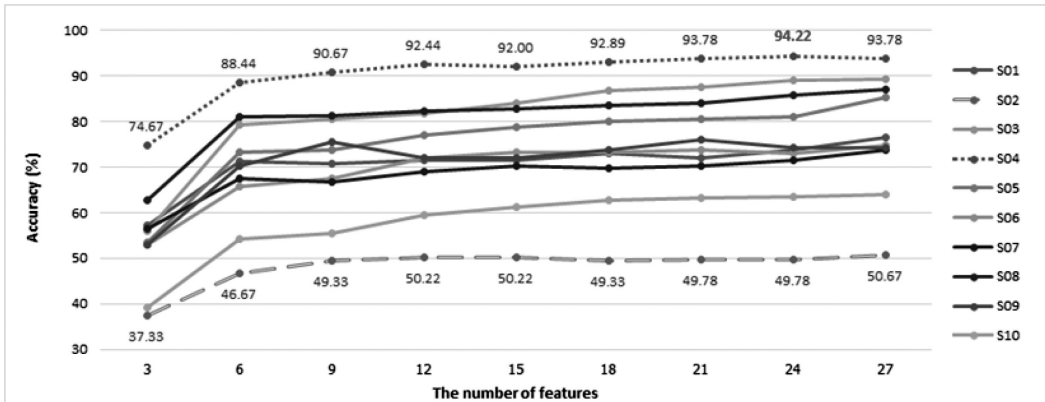
3. ผลการทดลอง

รูปที่ 6 แสดงผลความแม่นยำในการจำแนกวรรณยุกต์ 5 เสียง สำหรับรูปแบบการจำแนกทั้งสองรูปแบบ คือ (ก) การจำแนกแบบชั้นตอนเดียว และ (ข) การจำแนกแบบสามชั้นตอน โดยเปรียบเทียบความแม่นยำในการจำแนกของสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) ขณะที่มีการเพิ่มจำนวนลักษณะเด่นจาก 3 ตัว จนถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว จะเห็นได้ว่าสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครเกือบทั้งหมดจากทั้งสองรูปแบบการจำแนก ให้ผลไปในทิศทางเดียวกัน กล่าวคือ เมื่อจำนวนลักษณะเด่นเพิ่มขึ้น ความแม่นยำในการจำแนกจะเพิ่มขึ้นด้วย โดยความแม่นยำจะเพิ่มขึ้นอย่างมากในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 และ ความแม่นยำจะเพิ่มขึ้นเล็กน้อยและค่อยๆ คงที่ในช่วงการปรับจำนวนลักษณะเด่นตั้งแต่ 12 ตัวเป็นต้นไป

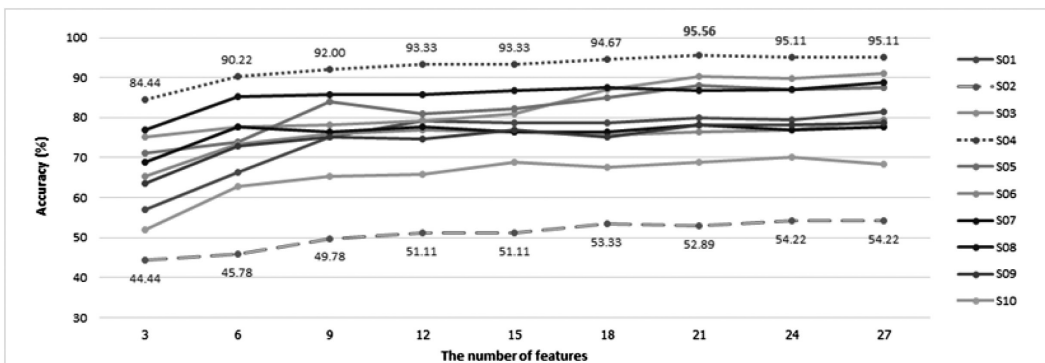
เมื่อพิจารณาในแต่ละช่วงของการปรับจำนวนลักษณะเด่น จะเห็นได้ว่าสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครแต่ละคนให้ผลความแม่นยำที่แตกต่างกัน ซึ่งมีความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุดอย่างเห็นได้ชัด พิจารณาผลการจำแนกแบบชั้นตอนเดียวซึ่งแสดงในรูปที่ 6(ก) ณ ตำแหน่งการใช้จำนวนลักษณะเด่น 3 ตัว จะเห็นได้ว่า ความแม่นยำสูงสุด คือ 76.67% เป็นผลความแม่นยำของ S04 ในขณะที่ความแม่นยำต่ำสุดคือ 37.33% ได้จากสัญญาณของ S02 จะเห็นได้ว่ามีความแตกต่างกัน 39.34% และเมื่อพิจารณา ณ ตำแหน่งการใช้จำนวนลักษณะเด่นที่ให้ความแม่นยำสูงสุดจากรูปแบบการจำแนกแบบชั้นตอนเดียว พบว่า การใช้จำนวนลักษณะเด่น 24 ตัว ให้ความแม่นยำสูงสุดซึ่งได้รับจากสัญญาณ S04 เช่นเดียวกัน โดยความแม่นยำที่ได้ คือ 94.22% ในขณะที่ความแม่นยำต่ำสุด คือ 49.78% ซึ่งเป็นสัญญาณของ S02 เช่นเดียวกัน โดยที่ความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุด ณ ตำแหน่งการใช้ลักษณะเด่น 24 ตัว คือ 44.44% ดังนั้นแสดงให้เห็นว่าสัญญาณของ S04 ให้ผลความแม่นยำสูงสุด และ S02

ให้ผลความแม่นยำต่ำสุด โดยที่มีความแตกต่างระหว่างความแม่นยำสูงสุดและต่ำสุดมาก ซึ่งให้ผลเช่นเดียวกับรูปแบบการจำแนกแบบสามขั้นตอนในรูปที่ 6 (ข) นอกจากนี้เมื่อพิจารณาการจำแนกแบบสามขั้นตอน จะเห็นได้ว่า S02 ให้ผลความแม่นยำที่แตกต่างกับสัญญาณอื่นๆ อย่าง

เห็นได้ชัด โดยให้ความแม่นยำที่ต่ำมาก อย่างไรก็ตามการเพิ่มจำนวนลักษณะเด่น สามารถเพิ่มความแม่นยำในการจำแนก โดยให้ผลการเปลี่ยนแปลงไปในทิศทางเดียวกันสำหรับข้อมูลสัญญาณทั้งหมด (S01-S10)



(ก)



(ข)

รูปที่ 6 กราฟแสดงผลความแม่นยำในการจำแนกรรณยุกต์ 5 เสียง ซึ่งเปรียบเทียบความแม่นยำของข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอน

พิจารณาเฉพาะสัญญาณ S04 ซึ่งให้ผลการจำแนกที่ดีที่สุดจากทั้งสองรูปแบบการจำแนก พบว่า สำหรับการจำแนกแบบขั้นตอนเดียว การเพิ่มจำนวนลักษณะเด่น 24 ตัว ให้ความแม่นยำสูงสุด (94.22%) ซึ่งสามารถเพิ่มความแม่นยำ 19.55% เมื่อเปรียบเทียบกับการใช้จำนวนลักษณะเด่นเพียง 3 ตัว ในขณะที่การจำแนกแบบสามขั้นตอน การเพิ่มจำนวนลักษณะเด่น 21 ตัว ให้ความ

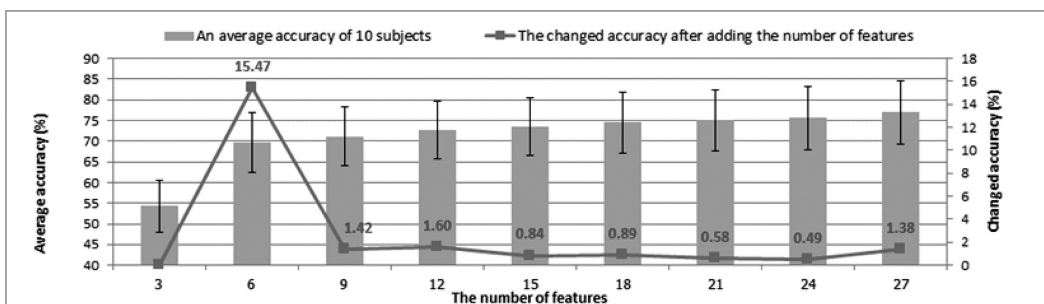
แม่นยำสูงสุด (95.56%) ซึ่งความแม่นยำเพิ่มขึ้น 10.67% เมื่อเปรียบเทียบกับการใช้จำนวนลักษณะเด่นเพียง 3 ตัว ดังนั้นจะเห็นได้ว่ารูปแบบการจำแนกแบบสามขั้นตอน ให้ความแม่นยำที่ดีกว่าการจำแนกแบบขั้นตอนเดียว และใช้จำนวนลักษณะเด่นน้อยกว่า

อย่างไรก็ตาม จากผลการทดลองส่วนนี้ยังไม่สามารถสรุปจำนวนลักษณะเด่นที่เหมาะสมในงานวิจัยนี้ได้อย่าง

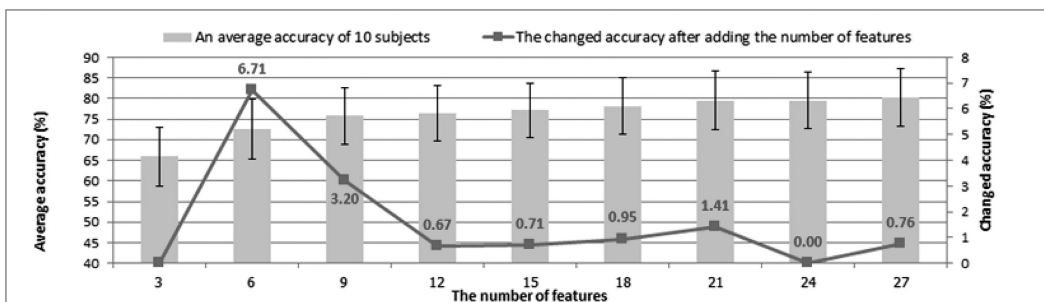
ชัดเจน เนื่องจากความแม่นยำสูงสุดที่เกิดขึ้นในแต่ละสัญญาณ เกิดจากการใช้จำนวนลักษณะเด่นที่แตกต่างกัน ดังนั้นเพื่อให้สามารถวิเคราะห์จำนวนลักษณะเด่นที่เหมาะสม และระบุอัตราการเปลี่ยนแปลงของความแม่นยำในขณะที่มีการปรับจำนวนลักษณะเด่นได้อย่างชัดเจน จึงนำผลความแม่นยำในการจำแนกรรณยุกต์ 5 เสียงของสัญญาณไฟฟ้ากล่อมเนื้อทั้งหมด (S01-S10) มาคำนวณหาค่าเฉลี่ย

รูปที่ 7 เป็นผลความแม่นยำเฉลี่ยซึ่งถูกแสดงในรูปแบบของแผนภูมิแท่ง โดยที่กราฟเส้นตรงแสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น พิจารณาการจำแนกแบบขั้นตอนเดียว รูปที่ 7(ก) จะเห็นว่าในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 ความแม่นยำเพิ่มขึ้นมากที่สุด โดยเพิ่มขึ้น

15.47% หลังจากนั้นความแม่นยำจะเพิ่มขึ้นเพียงเล็กน้อย โดยที่ความแม่นยำสูงสุดได้จากการใช้จำนวนลักษณะเด่น 27 ตัว (76.93%) สำหรับการจำแนกแบบสามขั้นตอน รูปที่ 7(ข) จะเห็นว่าในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 ความแม่นยำเพิ่มขึ้นมากที่สุดเช่นเดียวกัน โดยเพิ่มขึ้น 6.71% หลังจากนั้นความแม่นยำจะเพิ่มขึ้นเพียงเล็กน้อย โดยที่ความแม่นยำสูงสุดได้จากการใช้จำนวนลักษณะเด่น 27 ตัว (80.27%) อย่างไรก็ตามถึงแม้ว่าอัตราการเพิ่มขึ้นของความแม่นยำเฉลี่ยของการจำแนกแบบสามขั้นตอนในช่วงการปรับจำนวนลักษณะเด่นจาก 3 เป็น 6 จะต่ำกว่าการจำแนกแบบขั้นตอนเดียว แต่ความแม่นยำเฉลี่ยของการจำแนกแบบสามขั้นตอนก็ยังคงสูงกว่าการจำแนกแบบขั้นตอนเดียว



(ก)

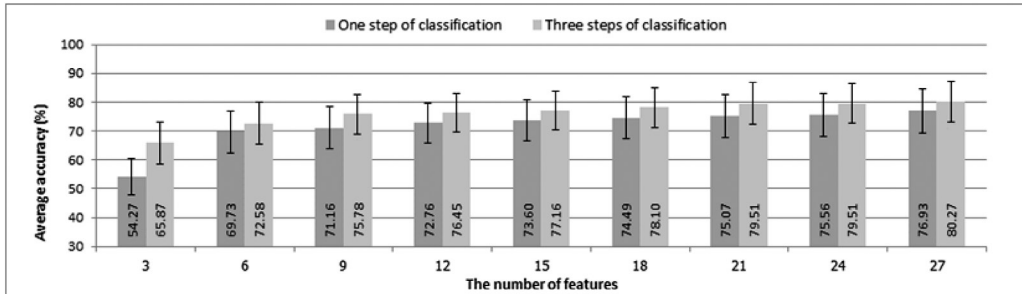


(ข)

รูปที่ 7 กราฟแกนปฐมภูมิ (ซ้าย) แสดงผลความแม่นยำเฉลี่ยในการจำแนกรรณยุกต์ 5 เสียง ซึ่งคำนวณจากข้อมูลสัญญาณไฟฟ้ากล่อมเนื้อของอาสาสมัครทั้ง 10 คน (S01-S10) เมื่อมีการเพิ่มจำนวนลักษณะเด่นจาก 3 ถึง 27 ตัว โดยเพิ่มครั้งละ 3 ตัว กราฟแกนทุติยภูมิ (ขวา) แสดงค่าการเปลี่ยนแปลงความแม่นยำในการจำแนกเมื่อจำนวนลักษณะเด่นเพิ่มขึ้น (ก) การจำแนกแบบขั้นตอนเดียว (ข) การจำแนกแบบสามขั้นตอนหมายเหตุ : แถบค่าคลาดเคลื่อน (error bar) แสดงช่วงความเชื่อมั่น (confidence interval) 95%

รูปที่ 8 แสดงการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนกรรณยุกต์ 5 เสียง ระหว่างรูปแบบการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอน จะเห็นได้ว่าการจำแนกแบบสามขั้นตอนมีความแม่นยำสูงกว่าการจำแนกแบบขั้นตอนเดียว ซึ่งจะแตกต่างอย่างเห็นได้ชัด

เมื่อพิจารณา ณ ตำแหน่งที่มีการใช้จำนวนลักษณะเด่นเพียง 3 ตัว โดยที่การจำแนกแบบสามขั้นตอนมีความแม่นยำเฉลี่ยสูงกว่า 11.6% เมื่อเพิ่มจำนวนลักษณะเด่นมากขึ้น จะเห็นได้ว่ามีความแตกต่างเพียงเล็กน้อยระหว่างความแม่นยำเฉลี่ยของทั้งสองรูปแบบ



รูปที่ 8 แผนภูมิแท่งแสดงการเปรียบเทียบความแม่นยำเฉลี่ยในการจำแนก ระหว่างรูปแบบการจำแนกแบบขั้นตอนเดียวและการจำแนกแบบสามขั้นตอนหมายเหตุ : แถบค่าคลาดเคลื่อน (error bar) แสดงช่วงความเชื่อมั่น (confidence interval) 95%

4. วิเคราะห์ผลการทดลอง

ในส่วนของการวิเคราะห์ผลการทดลอง จะวิเคราะห์ใน 4 ประเด็นหลัก ได้แก่ ประสิทธิภาพของระบบเมื่อเพิ่มจำนวนลักษณะเด่นให้กับระบบ ประสิทธิภาพในการจำแนกของสัญญาณจากอาสาสมัครแต่ละคน ประสิทธิภาพของรูปแบบการจำแนกสองรูปแบบ และคุณลักษณะของลักษณะเด่นที่ถูกเลือก

ประเด็นแรก พิจารณาประสิทธิภาพของระบบเมื่อเพิ่มจำนวนลักษณะเด่นให้กับระบบ จากผลการทดลองจะเห็นได้ว่าการเพิ่มจำนวนลักษณะเด่นสามารถเพิ่มความแม่นยำให้กับระบบ โดยที่การเพิ่มลักษณะเด่น 27 ตัว ให้ความแม่นยำเฉลี่ยสูงสุดสำหรับทั้งสองรูปแบบการจำแนก อย่างไรก็ตาม การเพิ่มจำนวนอินพุตให้กับตัวจำแนก ยิ่งเพิ่มจำนวนอินพุตมากขึ้นจะทำให้เวลาในการประมวลผลสัญญาณช้าลง ดังนั้นจำนวนลักษณะเด่นที่เหมาะสมในงานวิจัยนี้ จึงต้องพิจารณาทั้งในแง่ของความแม่นยำและเวลาในการประมวลผล ซึ่งจากรูปที่ 7(ก) และ รูปที่ 7 (ข) แสดงให้เห็นว่าจำนวนลักษณะเด่นที่เหมาะสมสำหรับการจำแนกแบบขั้นตอนเดียวและสามขั้นตอน คือ 12 ตัว และ 9 ตัว ตามลำดับ ทั้งนี้เนื่องจากการเพิ่มจำนวนลักษณะเด่นหลังจากจำนวนดังกล่าวทำให้ความแม่นยำเพิ่มขึ้นเพียงเล็กน้อยเท่านั้น กล่าวคือ สำหรับ

การจำแนกแบบขั้นตอนเดียว ความแม่นยำที่เพิ่มขึ้นหลังจากเพิ่มจำนวนลักษณะเด่น 12 ตัวเป็นต้นไป มีอัตราการเพิ่มขึ้นอยู่ในช่วงระหว่าง 0.49-1.38% สำหรับการจำแนกแบบสามขั้นตอน ความแม่นยำที่เพิ่มขึ้นหลังจากเพิ่มจำนวนลักษณะเด่น 9 ตัวเป็นต้นไป มีอัตราการเพิ่มขึ้นอยู่ในช่วงระหว่าง 0-1.41% การที่ความแม่นยำเพิ่มขึ้นเพียงเล็กน้อยในช่วงหลังของการเพิ่มจำนวนลักษณะเด่น อาจจะเป็นเพราะว่าลักษณะเด่นที่เพิ่มเข้าไปเป็นลักษณะเด่นที่มีคุณลักษณะใกล้เคียงกับลักษณะเด่นก่อนหน้า หรือกล่าวได้ว่าเป็นลักษณะเด่นที่ความซ้ำซ้อนกัน ทำให้ความแม่นยำในการจำแนกเพิ่มขึ้นเพียงเล็กน้อยหรือไม่เพิ่มขึ้นเลย

ประเด็นที่สอง พิจารณาประสิทธิภาพในการจำแนกของสัญญาณจากอาสาสมัครแต่ละคน จากการทดลองจะเห็นได้ว่า สัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน ให้ผลความแม่นยำที่แตกต่างกันออกไป ซึ่งในการทดลองนี้พบว่า สัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S04 เป็นสัญญาณที่ให้ผลการจำแนกที่ดีที่สุด และ สัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร S02 เป็นสัญญาณที่ให้ผลการจำแนกต่ำที่สุด ซึ่งสาเหตุที่ทำให้เกิดความแตกต่างของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัครแต่ละคน น่าจะเกิดจากส่วนของกระบวนการเก็บข้อมูลของสัญญาณ

ไฟฟ้ากล่อมเนื้อ กล่าวคือ ธรรมชาติในการออกเสียงพูดของอาสาสมัครแต่ละคนแตกต่างกัน ถึงแม้ว่าเสียงพูดที่ออกมาจะมีลักษณะของเสียงที่ถูกต้องและชัดเจน แต่วิธีการขยับปากเพื่อออกเสียงพูดของแต่ละคนแตกต่างกัน ซึ่งอาสาสมัครที่มีการขยับปากถูกต้องตามหลักการออกเสียงเพื่อเปล่งเสียงออกมา จะได้สัญญาณไฟฟ้ากล่อมเนื้อที่มีแอมพลิจูดชัดเจน และทำให้มีความแตกต่างของสัญญาณไฟฟ้ากล่อมเนื้อของคำพูดแต่ละคำ ทำให้ง่ายต่อการจำแนกคำพูด ในทางตรงกันข้าม อาสาสมัครบางคนไม่ค่อยมีการขยับปากขณะออกเสียง กล่าวคือมีการเคลื่อนไหวของอวัยวะในการออกเสียงพูดน้อยซึ่งมีผลทำให้เกิดการทำงานของสัญญาณไฟฟ้ากล่อมเนื้อน้อย ดังนั้นสัญญาณไฟฟ้ากล่อมเนื้อที่ได้จะมีแอมพลิจูดต่ำ และมีความแตกต่างของสัญญาณไฟฟ้ากล่อมเนื้อของคำพูดแต่ละคำน้อย ดังนั้นการควบคุมลักษณะการออกเสียงพูดของอาสาสมัครให้คงที่เป็นปัจจัยสำคัญที่จะช่วยเพิ่มประสิทธิภาพและความน่าเชื่อถือให้กับระบบได้

ประเด็นต่อมา พิจารณาประสิทธิภาพของรูปแบบการจำแนกสองรูปแบบ คือ รูปแบบการจำแนกแบบขั้นตอนเดียวโดยใช้โครงข่ายประสาทเทียมหนึ่งโครงข่าย และรูปแบบการจำแนกแบบสามขั้นตอนโดยใช้โครงข่ายประสาทเทียมสามโครงข่าย จากผลการทดลอง จะเห็นได้ว่า การจำแนกแบบสามขั้นตอนโดยใช้โครงข่ายประสาทเทียมสามโครงข่ายให้ผลความแม่นยำเฉลี่ยสูงกว่า แต่อย่างไรก็ตามการจำแนกแบบสามขั้นตอนมีความซับซ้อนและใช้เวลาในการประมวลผลมากกว่าการจำแนกแบบขั้นตอนเดียวถึงสามเท่า ดังนั้นถ้าระบบต้องการความเร็วในการประมวลผลการจำแนกแบบขั้นตอนเดียวน่าจะเป็น

ทางเลือกที่ดีกว่า เนื่องจากความแม่นยำในการจำแนกต่ำกว่าการจำแนกแบบสามขั้นตอนเพียงเล็กน้อย (1.78% เมื่อพิจารณา ณ ตำแหน่งที่ให้ความแม่นยำสูงสุด การใช้จำนวนลักษณะเด่น 21 ตัว) อย่างไรก็ตาม การจำแนกแบบสามขั้นตอนจะให้ความแม่นยำเฉลี่ยที่สูงกว่าการจำแนกแบบขั้นตอนเดียวอย่างเห็นได้ชัด (9.77%) เมื่อมีการใช้จำนวนลักษณะเด่นเพียงเล็กน้อย เช่น การใช้ลักษณะเด่นเพียง 3 ตัว เป็นต้น

ประเด็นสุดท้าย พิจารณาคุณลักษณะของลักษณะเด่นที่ถูกเลือก ซึ่งจะพิจารณาลักษณะเด่นที่ถูกใช้สำหรับการจำแนกแบบสามขั้นตอน เนื่องจากผลการทดลองแสดงให้เห็นว่า การจำแนกแบบสามขั้นตอนโดยใช้โครงข่ายประสาทเทียมสามโครงข่ายมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว โดยให้ความแม่นยำสูงสุดเมื่อเพิ่มจำนวนลักษณะเด่น 27 ตัว อย่างไรก็ตามจะพิจารณาลักษณะเด่นเพียง 9 ตัว เนื่องจาก การเพิ่มจำนวนลักษณะเด่นมากกว่า 9 ตัวมีผลในการเพิ่มความแม่นยำเพียงเล็กน้อย

ตารางที่ 2 แสดงลักษณะเด่นที่ถูกใช้สำหรับการจำแนกแบบสามขั้นตอน ซึ่งจะแบ่งเป็นลักษณะเด่นย่อย 3 ชุด ในแต่ละขั้นตอนจะใช้ลักษณะเด่นแตกต่างกัน จะเห็นว่าลักษณะเด่นที่ถูกเลือกส่วนใหญ่เป็นลักษณะเด่นในรูปแบบที่สอง ซึ่งมาจากการคำนวณโดยแบ่งสัญญาณไฟฟ้ากล่อมเนื้อออกเป็น 6 ส่วนเท่าๆ กัน ดังนั้นจะเห็นได้ว่าลักษณะเด่นในรูปแบบนี้มีประสิทธิภาพในการจำแนกวรรณยุกต์ได้ดี เนื่องจากเป็นลักษณะเด่นที่คำนวณมาจากบางส่วนของสัญญาณ ซึ่งในแต่ละส่วนของสัญญาณมีผลในการจำแนกวรรณยุกต์แตกต่างกัน

ตารางที่ 2 ลักษณะเด่นที่ถูกเลือกจากหัวข้อ 2.2.4 และ 2.2.5 สำหรับใช้ในรูปแบบการจำแนกแบบสามขั้นตอน

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

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

จากผลการทดลองทั้งหมด เมื่อเปรียบเทียบกับงานวิจัยก่อนหน้า [21] ซึ่งได้นำเสนอเทคนิคการจำแนกแบบสามขั้นตอน โดยออกแบบจากสัญญาณไฟฟ้ากล้ามเนื้อจากอาสาสมัครเพียง 1 คน ซึ่งพบว่าเทคนิคนี้สามารถจำแนกวรรณยุกต์ 5 เสียงได้ โดยมีความแม่นยำในการจำแนก 80% อย่างไรก็ตาม การทดลองดังกล่าวยังไม่สามารถยืนยันได้ว่า เทคนิคนี้จะมีประสิทธิภาพเพียงพอเมื่อนำไปใช้กับสัญญาณไฟฟ้ากล้ามเนื้ออื่นๆ จากผลการทดลองนี้ จะเห็นได้ว่าเทคนิคการจำแนกแบบสามขั้นตอนที่ได้นำเสนอมีประสิทธิภาพเพียงพอในการจำแนกวรรณยุกต์ 5 เสียง ซึ่งได้ปรับปรุงเทคนิคดังกล่าวในส่วนของ การคัดเลือกลักษณะเด่น และ การเพิ่มจำนวนลักษณะเด่นให้กับระบบ

5. สรุปผลการทดลอง

งานวิจัยนี้เป็นการปรับปรุงเทคนิคการจำแนกแบบสามขั้นตอนซึ่งใช้โครงข่ายประสาทเทียมสามโครงข่ายในการจำแนกวรรณยุกต์ไทย 5 เสียง โดยการใช้ข้อมูลของสัญญาณไฟฟ้ากล้ามเนื้อของอาสาสมัคร 10 คนในการออกแบบ โดยได้ปรับปรุงในส่วนของการคำนวณลักษณะเด่น คัดเลือกลักษณะเด่น และเพิ่มจำนวนลักษณะ

เด่นเพื่อเป็นอินพุตให้กับโครงข่ายประสาทเทียม นอกจากนี้ทำการเปรียบเทียบรูปแบบการจำแนกสองรูปแบบ คือ การจำแนกแบบสามขั้นตอน และการจำแนกแบบขั้นตอนเดียว จากผลการทดลองพบว่า การจำแนกแบบสามขั้นตอนมีประสิทธิภาพดีกว่าการจำแนกแบบขั้นตอนเดียว โดยให้ความแม่นยำสูงสุด 95.56% จากข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อ S04 และความแม่นยำเฉลี่ย 80.27% จากข้อมูลสัญญาณไฟฟ้ากล้ามเนื้อทั้งหมด นอกจากนี้ เมื่อพิจารณาจากความแม่นยำเฉลี่ย การใช้จำนวนลักษณะเด่น 27 ตัว ให้ความแม่นยำสูงสุด แต่อย่างไรก็ตาม จำนวนลักษณะเด่นที่เหมาะสม คือ 9 ตัว เนื่องจากการเพิ่มจำนวนลักษณะเด่นมากกว่า 9 ตัว จะเพิ่มประสิทธิภาพในการจำแนกเพียงเล็กน้อย

อย่างไรก็ตาม การทดลองนี้เป็นการคัดเลือกลักษณะเด่นโดยการจัดอันดับลักษณะเด่นตามผลการประเมินด้วยตัวประเมินแบบอาร์อีเอส ซึ่งวิธีนี้ ทำให้ลักษณะเด่นที่ใช้ อาจมีความสัมพันธ์กัน นั่นก็แสดงว่า ข้อมูลมีความซ้ำซ้อนซึ่งมีผลทำให้ความแม่นยำในการจำแนกไม่เพิ่มขึ้น แต่จะไปเพิ่มพารามิเตอร์เข้าไปในระบบ ทำให้การประมวลผลช้าลง นอกจากนี้ ในการเพิ่มจำนวนลักษณะเด่นครั้งละ 3 ตัว ทำให้ไม่สามารถระบุได้ชัดว่าลักษณะเด่นตัวใดมีผลในการเพิ่มหรือลดประสิทธิภาพของระบบ ดังนั้น งานต่อไปจะปรับปรุงในส่วนของการเลือกลักษณะเด่น โดยใช้เทคนิคการค้นหา (Search stagey based feature selection) เพื่อให้ได้ลักษณะเด่นที่เหมาะสม

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนโดยโครงการร่วมให้ทุนปริญญาเอกกาญจนาภิเษก (คปก.) ระหว่างสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) และมหาวิทยาลัยสงขลานครินทร์ (สัญญาณเลขที่ PHD/0164/2552)

7. เอกสารอ้างอิง

1. Sugie, N. and Tsunoda, K., 1985, "A Speech Prosthesis Employing a Speech Synthesizer Vowel Discrimination from Perioral Muscle Activities and Vowel Production," *IEEE Transactions on Biomedical Engineering*, 32 (7), pp. 485-490.

2. Kubo, T., Yoshida, M., Hattori, T. and Ikeda, K., 2014, "Towards Excluding Redundancy in Electrode Grid for Automatic Speech Recognition based on Surface EMG," *Neurocomputing*, 134, pp. 15-19.
3. Fraiwan, L., Lweesy, K., Al-Nemrawi, A., Addabass, S. and Saifan, R., 2011, "Voiceless Arabic vowels recognition using facial EMG," *Medical & Biological Engineering & Computing*, 49 (7), pp. 811-818.
4. Chan, A.D., Englehart, K.B., Hudgins, B. and Lovely, D.F., 2006, "Multiexpert Automatic Speech Recognition using Acoustic and Myoelectric Signals," *IEEE Transactions on Biomedical Engineering*, 53 (4), pp. 676-685.
5. Lopez-Larraz, E., Mozos, O.M., Antelis, J.M. and Minguez, J., 2010, "Syllable-based speech recognition using EMG," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, EMBC, pp. 4699-4702 .
6. Jou, S.C.S., Schultz, T., Walliczek, M., Kraft, F. and Waibel, A., 2006, "Towards Continuous Speech Recognition using Surface Electromyography," *INTERSPEECH 2006*, ICSLP.
7. Deng, Y., Colby, G., Heaton, J.T. and Meltzner, G.S., 2012, "Signal Processing Advances for the MUTE sEMG-based Silent Speech Recognition System," *IEEE Military Communications Conference*, pp. 1-6.
8. Wand, M., Janke, M. and Schultz, T., 2014, "Tackling Speaking Mode Varieties in EMG-based Speech Recognition," *IEEE Transactions on Biomedical Engineering*, 61 (10), pp. 2515-2526.
9. Betts, B.J. and Jorgensen, C., 2005, "Small Vocabulary Recognition using Surface Electromyography in an Acoustically Harsh Environment," *NASA TM-2005-21347*.
10. Zhou, Q., Jiang, N., Englehart, K. and Hudgins, B., 2009, "Improved Phoneme-based Myoelectric Speech Recognition," *IEEE Transactions on Biomedical Engineering*, 58 (8), pp. 2016-2023.
11. Arjunan, S.P., Kumar, D.K., Yau, W.C. and Weghorn, H., 2006, "Unspoken Vowel Recognition using Facial Electromyogram," *Proceedings of IEEE Engineering in Medicine and Biology Society*, pp. 2191-2194.
12. Morse, M.S., Day, S.H., Trull, B. and Morse, H., 1989, "Use of Myoelectric Signals to Recognize Speech," *Proceedings of IEEE Engineering in Medicine and Biology Society*, pp. 1793-1794.
13. Manabe, H. and Zhang, Z., 2004, "Multi-stream HMM for EMG-based Speech Recognition," *Proceedings of IEEE Engineering in Medicine and Biology Society*, 2, pp. 4389-4392.
14. Bu, N., Tsuji, T., Arita, J. and Ohga, M., 2005, "Phoneme Classification for Speech Synthesizer using Differential EMG Signals between Muscles," *Proceedings of IEEE Engineering in Medicine and Biology Society*, pp. 5962-5966.
15. Jia, X., Wang, X., Li, J., Yang, D. and Song, Y., 2006, "Unvoiced Chinese Digital Recognition based on Facial Myoelectric Signal," *International Conference on Communications, Circuits and Systems*, 1, pp. 598-601
16. Lyu, M., Xiong, C. and Zhang, Q., 2014, "Electromyography (EMG)-based Chinese Voice Command Recognition," *IEEE International Conference on Information and Automation*, pp. 926-931.
17. Lee, K.S., 2008, "EMG-based Speech Recognition using Hidden Markov Models with Global Control Variables," *IEEE Transactions on Biomedical Engineering*, 55 (3), pp. 930-940.
18. Freitas, J., Teixeira, A. and Dias, M. S., 2012, "Towards a Silent Speech Interface for Portuguese," *Biosignals*, pp. 91-100.

19. Topalović, M., Damnjanović, Đ., Peulić, A., Blagojević, M. and Filipović, N., 2015, "Syllable-based Speech Recognition using Electromyography and Decision Set Classifier," *Biomedical Engineering: Applications, Basis and Communications*, 27 (02).
20. Srisuwan, N., Phukpattaranont, P. and Limsakul, L., 2011, "Feature Selection for Thai Tone Classification based on Surface EMG," *Procedia Engineering*, 32, pp. 253-259.
21. Srisuwan, N., Phukpattaranont, P. and Limsakul, L., 2013, "Three Steps of Neural Network Classification for EMG-based Thai Tones Speech Recognition," *Proceeding of International Conference on Electrical Engineering/Electronics* *Computer Telecommunications and Information Technology (ECTI-CON)*, pp. 1-6.
22. Di Fabio, R.P., 1987, "Reliability of Computerized Surface Electromyography for Determining the Onset of Muscle Activity," *Physical Therapy*, 67 (1), pp. 43-48.
23. Phinyomark, A., Hirunviriyaya, S., Limsakul, C. and Phukpattaranont, P., 2010, "Evaluation of EMG Feature Extraction for Hand Movement Recognition based on Euclidean Distance and Standard Deviation," *International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON)*, pp. 856-860.