การพัฒนาซอฟต์แวร์คอมพิวเตอร์สำหรับวิเคราะห์หน่วยเสียงของผู้พูดภาษาไทย

นลินี ชนะมูล [1] เบญญาดา อุณหเลขจิตร [2] และ มนตรี โพธิโสโนทัย [3*]
มหาวิทยาลัยบูรพา อ.เมือง จ.ชลบุรี 20131

บทคัดย่อ

การใช้ภาษาไทยของคนในสังคมพบว่ายังมีปัญหาเกี่ยวกับการอ่านและการเขียน โดยเฉพาะการออกเสียงภาษาไทย ซึ่งเป็นภาษาที่มีเอกลักษณ์เฉพาะ ยากแก่การออกเสียถ้าขาดการฝึกฝน บทความนี้จึงเกิดแนวความคิดในการพัฒนาซอฟต์แวร์สำหรับวิเคราะห์การออกเสียงภาษาไทยด้วยการประมวลผลสัญญาณเสียงพูด ซึ่งซอฟต์แวร์นี้ถูกพัฒนาด้วยภาษาเดลไฟ (Delphi) เพื่อช่วยให้ผู้ฝึกออกเสียงสามารถฝึกออกเสียงภาษาไทยและประเมินผลการออกเสียงได้ด้วยตนเองโดยผ่านการวิเคราะห์ของซอฟต์แวร์ วิธีการวิเคราะห์ประกอบด้วย การรับเสียงด้วยไมโครโฟน จากนั้นนำสัญญาณเสียงที่บันทึกได้มาประมวลผลบนพื้นฐานของการประมวลผลสัญญาณ ซึ่งมีขั้นตอนดังนี้ (1) กระบวนการประมวลผลเบื้องต้น คือปรับค่าต่างๆ ของสัญญาณเสียงให้มีความเหมาะสมสำหรับการวิเคราะห์ (2) กระบวนการหาลักษณะเด่นเพื่อเก็บค่าพารามิเตอร์สำหรับวิเคราะห์ (ค่าอัตราการตัดผ่านศูนย์ และ ค่าเวลาในการเกิดเสียง) และ (3) การเปรียบเทียบเพื่อแบ่งเกรด การทดลองบทความนี้ได้เลือกกลุ่มคำจำนวน 12 คำ กลุ่มตัวอย่างในช่วงอายุ 20–25 ปี เพศชาย 30 คน และเพศหญิง 30 คน ผลการทดลองพบว่าซอฟต์แวร์ประเมินผลค่าความถูกต้องเฉลี่ยอยู่ในช่วงร้อยละ 71.24 ถึง 92.47

คำสำคัญ : การออกเสียงภาษาไทย / การวิเคราะห์สัญญาณเสียงพูด / การประมวลผลสัญญาณ / หน่วยเสียง

*Corresponding author : E-mail : montrip@buu.ac.th, montrip@gmail.com

1, 2 นักศึกษาปริญญาตรี ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

3 อาจารย์ วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา

Development of Computer Software for Pathological Analyzing of Thai Language Speakers

Nalinee Chanamool [1], Benyada Unhalekajit [2], and Montri Phothisonothai [3*]

Burapha University, Muang, Chonburi 20131

Abstract

The main problems of Thai language learners are on reading and writing issues. Especially, Thai language is unique comparing with other languages due to the different part of consonant phonemes from the vowel. It plays key role in pronouncing Thai language especially in the specific tones, segmentation of each consonant phoneme from the vowel is required for analyzing Thai speech signal. The conventional assessment method was on Thai language clinic. This paper, therefore, proposes the development of computer software for pathological analyzing based on the speech processing techniques, which developed in Delphi language. Our software can be also to support not only Thai speakers, but also foreigners practice themselves pronunciation of specific Thai words. There are three main processes of the proposed method as follows: 1) preprocessing, 2) template enrollment and 3) assessment. To obtain the featured templates, we have selected two parameters that are zero-crossing rate (ZCR) and time duration (TD) of phonemes. The speech signal of correct pronouncing is recorded and then ZCR and TD of each word will be categorized as the template. Finally, the final grade of assessment is determined by comparing input parameters with stored templates in the databases. In the experiment, this paper used 12 words, subjects (30 male and 30 female 20-25 years old) participated to test the software. The results show that the method can archive average accuracy rates of 71.24% to 92.47%.

Keywords : Thai Speech Pronouncing / Speech Signal Analysis / Speech Processing / Phonemes

* *Corresponding author : E-mail : montrip@buu.ac.th, montrip@gmail.com*

1, 2 *Undergraduate Students, Department of Electrical Engineering, Faculty of Engineering.*

3 *Lecturer, College of Research Methodology and Cognitive Science (RMCS).*

1. บทนำ

เนื่องจากในปัจจุบันการใช้ภาษาไทยในชีวิตประจำวันของคนในสังคมพบว่ายังมีปัญหาที่เกิดขึ้นกับการใช้ภาษาไทย ทั้งการพูด อ่าน เขียน โดยเฉพาะการออกเสียงเป็นปัญหาเบื้องต้นที่นำไปสู่การใช้ภาษาไทยที่ไม่ถูกต้อง ปัญหาการออกเสียงภาษาไทยพบมากในกลุ่มเยาวชนของไทย รวมไปถึงชาวต่างชาติที่มีความสนใจศึกษาภาษาไทย โจทย์ปัญหาคือทำอย่างไรให้กลุ่มคนเหล่านี้ออกเสียงภาษาไทยได้อย่างถูกต้อง จากเดิมการออกเสียงภาษาไทยได้ถูกต้องจำเป็นต้องอาศัยแพทย์ผู้เชี่ยวชาญทางด้านสัทศาสตร์และภาษาไทยโดยเฉพาะ ในปัจจุบันมีโครงการหมอภาษาพัฒนาเยาวชน สังกัดสำนักงานคณะกรรมการวัฒนธรรมแห่งชาติ ได้บัญญัติการออกเสียงภาษาไทยผิดถือว่าเป็นโรค [1] โดยโรคที่วินิจฉัยพบมาก แบ่งตามลักษณะปัญหาการออกเสียงภาษาไทยที่พบบ่อย ตัวอย่าง เช่น โรค /ร/ /ล/ /ส/ /ซ/ /ถ/ /ท/ /ฝ/ /ฟ/ /จ/ /ง/ /ด/ /ล/ เป็นต้น ดังนั้นจึงเกิดโครงการต่างๆ มากมายที่มุ่งเน้นการแก้ปัญหาการออกเสียงภาษาไทยที่มีวิธีการรักษาเหมือนกับการพบแพทย์ที่โรงพยาบาล มีการทำประวัติ วินิจฉัยโรค บำบัดรักษาและจ่ายยา

จากการค้นคว้าบทความวิจัยนี้ได้รวบรวมภูมิหลังงานวิจัยที่เกี่ยวข้องโดยแบ่งได้เป็น 2 กลุ่มหลัก คือ กลุ่มการวิเคราะห์เชิงความถี่ (Frequency domain analysis) และกลุ่มการวิเคราะห์เชิงเวลา (Time domain analysis) [2] มีงานวิจัยที่ประเมินค่าหน่วยเสียงภาษาไทย โดย J. Jitsup และคณะ [3] งานวิจัยนี้นำเสนอวิธีการแยกพยางค์ของสัญญาณเสียงพูดภาษาไทยในกรณีที่เสียงพูดติดกันอันเนื่องมาจากเสียงของผู้พูดมีลักษณะเสียงใหญ่ หรือมีการพูดที่ค่อนข้างเร็ว จึงทำให้ปลายพยางค์ของเสียงข้างเคียงเกิดการซ้อนทับกัน (Overlap) ซึ่งในงานวิจัยนี้นำขบวนการแปลงเวฟเลต (Wavelet transform) หนึ่งระดับเข้ามาช่วยในการแยกพยางค์ ซึ่งการแปลงเวฟเลตเป็นการแยกสัญญาณเสียงออกเป็น 2 ส่วนคือ ส่วนของสัญญาณความถี่ต่ำและส่วนของสัญญาณความถี่สูง นำเฉพาะสัญญาณความถี่สูงมาแปลงเวฟเลตกลับ (Inverse wavelet transform: IWT) จากนั้นนำสัญญาณเสียงนี้ไปใช้เพื่อช่วยในการแยกพยางค์ของสัญญาณเสียง K. Chamnongthai [4] ได้เสนอวิธีการคัดแยกตัวสะกดในเสียงพูดภาษาไทย โดยได้นิยามเสียงตัวสะกดดังแสดงตาม S = CVTF โดยที่พยางค์ (Syllable: S) ประกอบด้วยตัวสะกด (Consonant: C) สระ (Vowel: V) ระดับเสียง (Tone: T) และตัวสะกดสุดท้าย (Final consonant: F) แต่ละส่วนชี้ตำแหน่งเสียงตัวสะกดของพยางค์ เมื่อไม่นานมานี้ โกวิทย์ พันธุ์ครู และคณะ [5] ได้เสนอระบบการวิเคราะห์การออกเสียง ร และ ล เพื่อวินิจฉัยความผิดปกติของอวัยวะการออกเสียง อย่างไรก็ตามระบบดังกล่าวสามารถแยกแยะการออกเสียงเฉพาะเสียง ร และ ล เท่านั้น นอกจากนี้ได้มีงานวิจัยที่ได้นำเสนอวิธีการประเมินการออกเสียงในภาษาต่างประเทศ เช่น ภาษาอังกฤษ และภาษาจีน [6]–[8] อย่างไรก็ตาม พบว่าภาษาไทยมีการออกเสียงที่แตกต่างจากภาษาอื่นและมีเอกลักษณ์เฉพาะ จากปัญหาข้างต้นบทความนี้จึงเกิดแนวความคิดในการพัฒนาซอฟต์แวร์คอมพิวเตอร์สำเร็จรูปสำหรับวิเคราะห์และประเมินผลการออกเสียงภาษาไทยด้วยเทคนิคการประมวลผลสัญญาณเสียงพูด ซึ่งช่วยให้ผู้เข้ารับการรักษาสามารถฝึกออกเสียงได้ด้วยตนเองรวมถึงประเมินผลการออกเสียงเบื้องต้นเพื่อช่วยประหยัดเวลาในการเข้าพบแพทย์

2. แนวความคิด

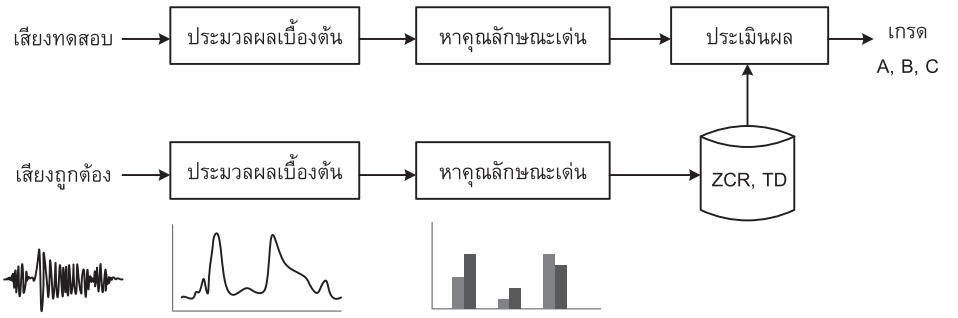
งานวิจัยนี้มีแนวความคิดพื้นฐานสำหรับใช้วิเคราะห์และประเมินผลหน่วยเสียงของผู้พูดภาษาไทยดังนี้ (1) การแบ่งแยกช่วงสัญญาณเสียงและช่วงที่ไม่ใช่สัญญาณเสียงได้ใช้สมมติฐานที่ว่า ช่วงที่เป็นเสียงควรมีค่าความแปรปรวนมากกว่าช่วงที่ไม่ใช่เสียง (2) การวิเคราะห์สัญญาณเสียงแบ่งออกตามหน่วยเสียงที่เกิดขึ้นเนื่องจากพบว่าสัญญาณเสียงที่ได้เกิดจากหน่วยเสียงย่อยประกอบกัน และ (3) การหาคุณลักษณะเด่นของสัญญาณใช้ค่าพารามิเตอร์ 2 ค่า คือ อัตราการตัดผ่านศูนย์ (Zero-Crossing Rate : ZCR) และ ช่วงเวลาที่เกิดเสียงของแต่ละช่วงหน่วยเสียง (Time Duration : TD) ซึ่งค่าพารามิเตอร์ทั้งสองของหน่วยเสียงมีความแตกต่างกัน

3. วิธีการ

ระเบียบวิธีการของงานวิจัยนี้แสดงดังรูปที่ 1 ซึ่งแบ่งออกเป็น 4 ขั้นตอนหลักๆ ดังนี้

3.1 การรับสัญญาณเสียงพูด

สัญญาณเสียงพูดที่ได้บันทึกจากกลุ่มตัวอย่างถูกบันทึกผ่านไมโครโฟน และจากนั้นสัญญาณเสียงถูกแปลงให้อยู่ในรูปสัญญาณดิจิตอลผ่านการ์ดเสียงความละเอียดขนาด 16 บิต ที่อัตราการสุ่มข้อมูลที่ 44,100 เฮิรตซ์ สัญญาณดิจิตอลเข้าสู่กระบวนการต่อไปเพื่อทำการประมวลผล



รูปที่ 1 ขั้นตอนการวิเคราะห์และประเมินเสียงพูดของงานวิจัยนี้

3.2 การประมวลผลเบื้องต้น (Preprocessing)

ขั้นตอนนี้เป็นขั้นตอนการปรับแต่งสัญญาณเบื้องต้นก่อนนำไปหาลักษณะเด่น เนื่องจากสัญญาณเสียงอินพุตที่ได้บันทึกจากกลุ่มตัวอย่างมีความแตกต่างกันในด้าน ระดับความดังของเสียงและช่วงเวลาในการออกเสียง โดยมีขั้นตอนย่อยดังต่อไปนี้

3.2.1 การทำค่าเฉลี่ยเป็นศูนย์ (Zero mean) : เป็นขั้นตอนการทำให้สัญญาณเสียงมีค่าเฉลี่ยเป็นศูนย์ เพื่อเหมาะแก่การประมวลผลเนื่องจากระดับสัญญาณเสียงของผู้ทดสอบแต่ละคนนั้นแตกต่างกันจึงต้องมีการปรับให้สัญญาณเสียงมีค่าเฉลี่ยเป็นศูนย์

3.2.2 การทำรูปแบบบรรทัดฐาน (Normalization) : ขั้นตอนนี้ทำการปรับระดับของสัญญาณให้อยู่ในช่วง -1 ถึง 1 เพื่อให้สัญญาณเสียงอยู่ในบรรทัดฐานเดียวกันก่อนนำไปประมวลผล

3.2.3 กำหนดความยาวของข้อมูลเสียงพูด (Fixed-length signal) : จากการทดสอบบันทึกเสียงแต่ละคำพบว่าความยาวของสัญญาณเสียงอยู่ในช่วงไม่เกิน 1.5 วินาที ดังนั้น งานวิจัยนี้จึงได้กำหนดความยาวของข้อมูลเสียงพูดถูกปรับความยาวของข้อมูลให้อยู่ในช่วง 2^{16} จุด (65,536 จุด) เพื่อให้ครอบคลุมกลุ่มคำที่ทำการวิเคราะห์ การกำหนดความยาวข้อมูลกระทำสองแบบ คือ กรณีที่ความยาวของสัญญาณอินพุตเกินช่วงความยาวที่กำหนดช่วงที่เกินถูกตัดทิ้งไป และกรณีที่ความยาวข้อมูลไม่ถึง 65,536 จุด ทำการแทรกค่าศูนย์เข้า (zero padding) เพื่อให้ได้ความยาวที่ต้องการ

3.2.4 การหาค่าพลังงาน (Energy signal) : เป็นขั้นตอนการหาค่าพลังงานของสัญญาณเสียง สามารถทำได้โดยการคำนวณค่ากำลังสองของสัญญาณเสียง เพื่อแยกความแตกต่างของพลังงานที่เกิดขึ้นระหว่างระดับสัญญาณของส่วนที่เป็นเสียงและไม่ใช่สัญญาณเสียง

3.2.5 การหาค่าเบี่ยงเบนมาตรฐานด้วยวิธีหน้าต่าง (Windowing-based standard deviation) : เป็นขั้นตอนการหาค่าเบี่ยงเบนมาตรฐานในช่วงความยาวของข้อมูลที่กำหนดด้วยเทคนิควิธีหน้าต่าง (windowing technique) งานวิจัยนี้ได้เลือกจำนวนข้อมูลสำหรับประมวลผลแบบหน้าต่างขนาด 256 จุด โดยแต่ละช่วงของข้อมูลถูกนำมาหาค่าเบี่ยงเบนมาตรฐาน ตัวอย่างสัญญาณที่ได้หลังจากการหาค่าเบี่ยงเบนมาตรฐาน พบว่าได้ข้อมูลทั้งหมด 256 จุด (65,536/256) นอกจากนี้จากรูปสัญญาณที่ได้ยังพบว่าช่วงของสัญญาณเสียงให้ค่าเบี่ยงเบนมาตรฐานสูงกว่าสัญญาณที่ไม่ใช่เสียง ซึ่งทำให้สามารถนำไปใช้คัดแยกเฉพาะส่วนที่เป็นสัญญาณเสียงออกมาได้

3.2.6 การกำหนดค่าตัดสินใจ (Thresholding) : เป็นขั้นตอนการกำหนดค่าตัดสินใจ เพื่อแยกส่วนที่เป็นเสียงและไม่ใช่เสียงออกงานวิจัยนี้เลือกใช้ค่าตัดสินใจในช่วง 0.001 ถึง 0.1 และทำการปรับค่าเองอัตโนมัติ (adaptive

threshold) ซึ่งค่าตัดสินใจนี้ปรับขึ้นครั้งละ 0.001 จนกระทั่งเงื่อนไขการแบ่งช่วงหน่วยเสียงตรงกับคำที่ทดสอบ โดยที่ค่าเบี่ยงเบนมาตรฐานต่ำกว่าค่าตัดสินใจกำหนดให้เป็นส่วนที่ไม่ใช่เสียงพูด และส่วนที่มากกว่าค่าตัดสินใจกำหนดให้เป็นส่วนที่เป็นเสียงพูด

3.2.7 การแยกสัญญาณเสียงพูด (Speech segmentation) : การแยกส่วนที่เป็นสัญญาณเสียงพูดออกจากส่วนที่ไม่ใช่สัญญาณเสียงใช้ค่าตัดสินใจจากขั้นตอนที่ 6 ถ้าค่าเบี่ยงเบนมาตรฐานสูงกว่าค่าตัดสินใจกำหนดให้เป็นสัญญาณเสียง (หรือค่า 1) และถ้าต่ำกว่าค่าตัดสินใจกำหนดให้ไม่ใช่สัญญาณเสียง (หรือค่า 0) เมื่อได้สัญญาณเสียงพูดที่แบ่งแล้วในขั้นตอนนี้ทำการขยายขนาดของหน้าต่างกลับคืนเพื่อให้ได้ขนาดของสัญญาณตรงกับสัญญาณอินพุต

3.3 การหาคุณลักษณะเด่น (Feature extraction)

ขั้นตอนนี้เป็นการหาคุณลักษณะเด่นที่เกิดขึ้นของสัญญาณเสียงแต่ละกลุ่มคำ โดยงานวิจัยนี้เลือกใช้ค่าพารามิเตอร์ 2 ค่าดังนี้

3.3.1 กราฟแสดงช่วงเวลา (Time duration) : การตรวจสอบเวลาของการเกิดในแต่ละช่วงหน่วยเสียงย่อยใช้เพื่อหากราฟแสดงช่วงเวลา หรือเรียกว่าค่า T_D ของแต่ละหน่วยเสียงย่อย หาได้จากสมการ

$$T_D = \frac{T_{n+1} - T_n}{f_s} \quad \textbf{(1)}$$

โดยที่ T_D คือ ช่วงเวลาที่เกิดสัญญาณเสียง มีหน่วยเป็นวินาที, T_n คือ ช่วงเวลาการเกิดสัญญาณเสียงลำดับที่ n, และ f_s คือ ความถี่ในการสุ่มข้อมูล (ในงานวิจัยนี้ใช้อัตราการสุ่มข้อมูลที่ 44,100 เฮิรตซ์)

3.3.2 อัตราการตัดผ่านศูนย์ (Zero-crossing rate) : ใช้เพื่อหากราฟแสดงอัตราการตัดผ่านศูนย์ หรือเรียกว่าค่า ZCR เพื่อใช้สำหรับหาค่าอัตราตัดผ่านศูนย์ของแต่ละหน่วยเสียงย่อย สามารถคำนวณได้จากสมการ

$$Z_s = \sum_{n=0}^{T-1} I\{S_n S_{n-1} < 0\} \quad \textbf{(2)}$$

เมื่อ S คือ สัญญาณเสียงที่ผ่านขั้นตอนการประมวลผลเบื้องต้นแล้ว, T คือคาบเวลาหรือเท่ากับ 256 และฟังก์ชัน $I\{\mathrm{A}\}$ เท่ากับ 1 ถ้า A เป็นจริง และเท่ากับ 0 ถ้าไม่เป็นจริง ซึ่งตัวอย่างรูปสัญญาณเสียงของคำว่า “ฉุกเฉิน” ที่ผ่านขั้นตอนการประมวลผลเบื้องต้นแสดงดังรูปที่ 2 หลังจากผ่านขั้นตอนการแยกสัญญาณส่วนที่เป็นเสียงและไม่ใช่เสียงออกแล้วพบว่าสัญญาณเสียงที่ได้เกิดจากหน่วยเสียงประกอบกัน

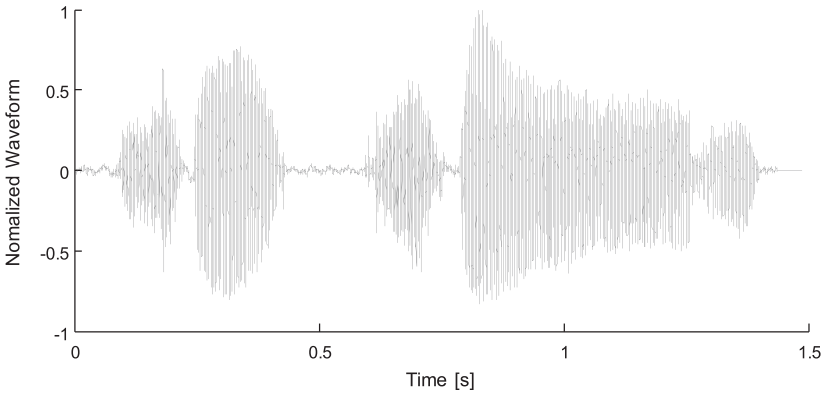
จากรูปที่ 3 เป็นแสดงระดับความเข้มของเสียงด้วยกราฟสเปกโตรแกรมพบว่าสัญญาณเสียงคำว่า “ฉุกเฉิน” ประกอบด้วยหน่วยเสียง 5 ส่วนคือ (1) “/ฉะ/” (2) “/อุก/” (3) “/ฉะ/” (4) “/อา/” และ (5) “/น/” โดยทั้ง 5 ส่วนถูกแยกเพื่อวิเคราะห์ค่าพารามิเตอร์ค่าเวลาของการเกิด (TD) และค่าอัตราการตัดผ่านศูนย์ (ZCR) ของแต่ละส่วน ซึ่งผลลัพธ์แสดงดังรูปที่ 4 จากรูปพบว่าค่าพารามิเตอร์ของส่วนที่ 1 และ 3 มีค่าใกล้เคียงกัน เนื่องจากมีเสียงตัวสะกดเดียวกัน และส่วนที่ 4 มีค่าเวลาของการเกิดสูงสุด

3.4 การประเมินผล (Assessment)

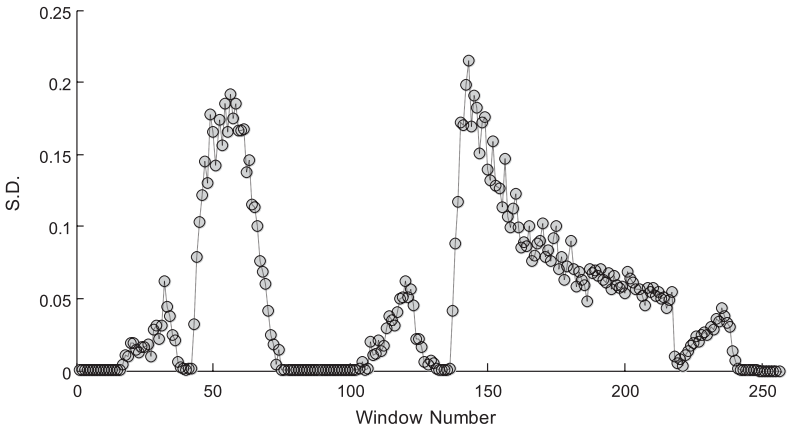
การประเมินผลงานวิจัยนี้ได้เปรียบเทียบกับเสียงต้นแบบ (Template) ที่ถูกต้อง ซึ่งได้จากการบันทึกเสียงผู้พูดในสภาพแวดล้อมที่มีการควบคุมสัญญาณรบกวนและประเมินด้วยผู้เชี่ยวชาญด้านการออกเสียงภาษาไทย โดยนำค่า ZCR และค่า TD ของเสียงที่ถูกต้องจัดเก็บเป็นฐานข้อมูลใช้สำหรับเปรียบเทียบกับสัญญาณเสียงพูดของผู้ทดสอบ โดยงานวิจัยนี้ได้แบ่งเกรดของการวินิจฉัยดังตารางที่ 1

ตารางที่ 1 เกณฑ์การประเมินผลการออกเสียงภาษาไทย

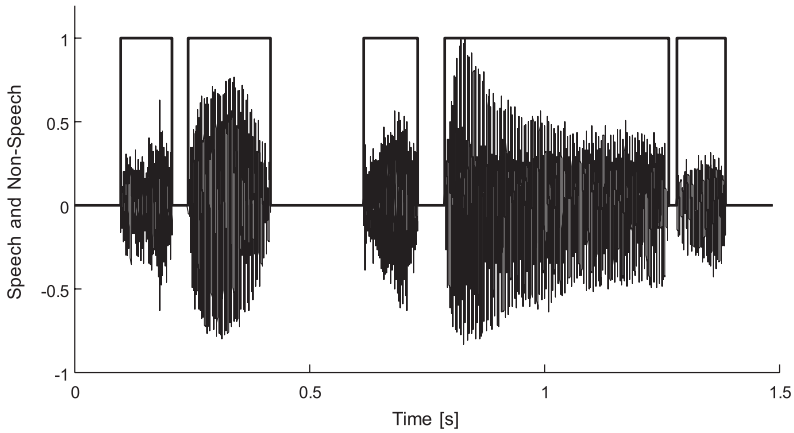
เกรด	ความหมาย
A	ออกเสียงได้ถูกต้อง ผ่านการทดสอบ
B	ออกเสียงได้ถูกบางส่วน ฝึกออกเสียงได้ด้วยตนเอง
C	ออกเสียงไม่ถูกต้อง เข้ารับการบำบัดจากแพทย์



(ก) สัญญาณเสียงพูดที่ผ่านขั้นตอนการทำรูปแบบบรรทัดฐาน

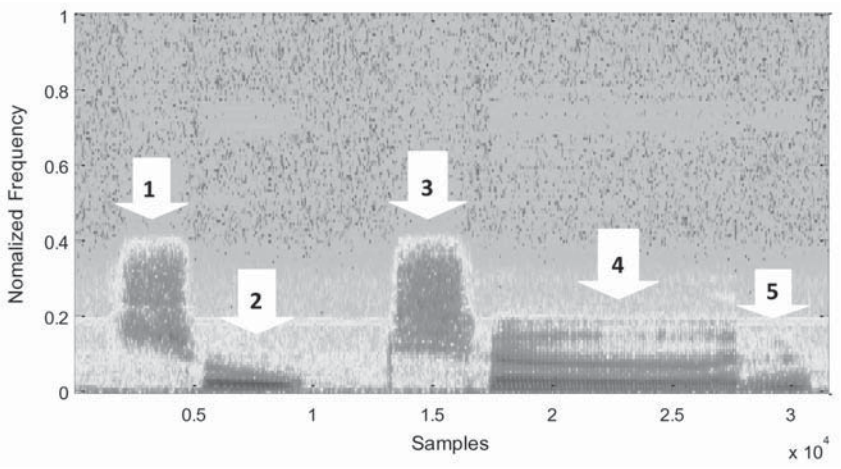


(ข) ค่าเบี่ยงเบนมาตรฐานด้วยวิธีหน้าต่างขนาด 256 จุด

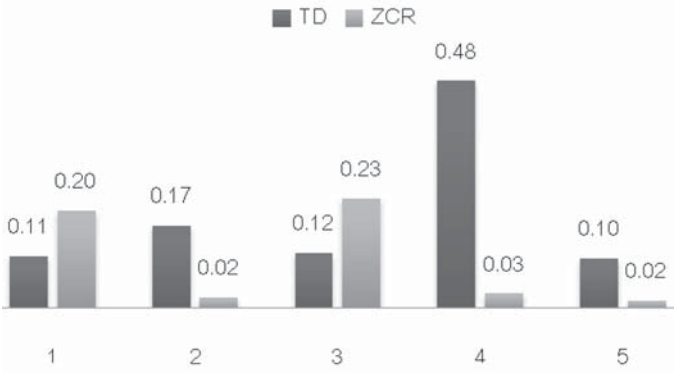


(ค) สัญญาณเสียงที่ถูกแบ่งออกเป็นหน่วยเสียงย่อยทั้ง 5 ส่วน

รูปที่ 2 ตัวอย่างรูปสัญญาณที่ผ่านขั้นตอนการประมวลผลเพื่อแยกหน่วยเสียงย่อย



รูปที่ 3 สเปกโตรแกรม (Spectrogram) ของระดับความเข้มของเสียง แบ่งตามหน่วยเสียงย่อยทั้ง 5 ส่วน



รูปที่ 4 กราฟเปรียบเทียบค่าพารามิเตอร์ TD (วินาที) และ ZCR ของหน่วยเสียงย่อยทั้ง 5 ส่วน

4. ผลการทดลอง

งานวิจัยนี้ได้เลือกทดสอบกับกลุ่มตัวอย่างเพศหญิง จำนวน 30 คน อายุระหว่าง 20-25 ปี และกลุ่มตัวอย่างเพศชาย จำนวน 30 คน อายุระหว่าง 20 – 25 ปี ชุดคำที่ใช้จำนวน 12 คำ ได้แก่ ฉุกเฉิน, เชื่องช้า, ซ้อน, เสริม, ที่, ถ่วง, ลอดรั้ว, ลงรอย, ง่วง, หงุดหงิด, เจือจาง, และฟุ่มเฟือย ผู้วิจัยได้พัฒนาซอฟต์แวร์ขึ้นเพื่อใช้สำหรับวิเคราะห์หน่วยเสียงและประเมินผลโดยงานวิจัยนี้ได้พัฒนาวิธีการทั้งหมดด้วยภาษาเดลไฟ (Delphi) และตั้งชื่อซอฟต์แวร์นี้ว่า “PATHAI (Pathological Analyzer for THAI language)” ดังแสดงในรูปที่ 5 การทดลองได้บันทึกเสียงจากกลุ่มโดยที่กำหนดให้ออกเสียง 12 คำ คำละ 10 ครั้ง เงื่อนไขการประเมินผลพิจารณาค่าพารามิเตอร์ของเสียงทดสอบเทียบกับต้นแบบที่เก็บไว้ งานวิจัยนี้แบ่งเกรดการประเมินทั้งหมด 3 ระดับ โดยมีเงื่อนไขดังตารางที่ 2

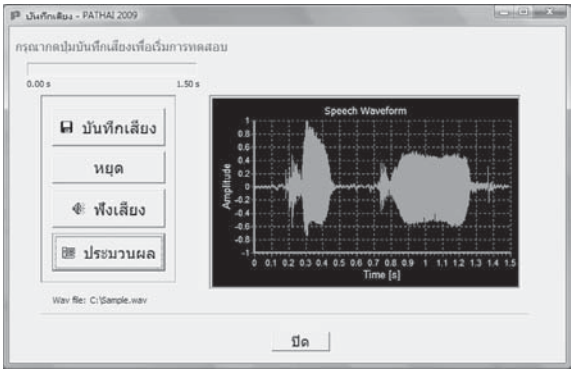
ตารางที่ 2 เงื่อนไขสำหรับใช้ประเมินด้วยค่าพารามิเตอร์

เกรด	ค่าพารามิเตอร์		
	จำนวนหน่วยเสียง	ZCR	TD
A	○	○	○
B	○	×	×
C	×	×	×

หมายเหตุ : ○ = ถูกต้องตามเงื่อนไข, × = ไม่ถูกต้องตามเงื่อนไข

(ก) หน้าต่างหลักของซอฟต์แวร์

(ข) หน้าต่างบันทึกประวัติส่วนตัวของผู้ทดสอบ



(ค) หน้าต่างสำหรับบันทึกเสียงและประมวลผล

(ง) หน้าต่างสรุปผลการวิเคราะห์และผลการประเมิน

รูปที่ 5 ซอฟต์แวร์ PATHAI ที่งานวิจัยนี้ได้พัฒนาขึ้น

ผลการทดลองแสดงดังตารางที่ 3 โดยใช้การประเมินค่าความถูกต้องของการออกเสียงใช้เกณฑ์การให้เกรด A เท่ากับ 3 คะแนน, เกรด B เท่ากับ 2 คะแนน, และเกรด C เท่ากับ 1 คะแนน ตามลำดับ จากนั้นนำคะแนนที่ได้มารวมเพื่อประเมินผล

ตารางที่ 3 ผลการทดลองใช้ซอฟต์แวร์ PATHAI กับกลุ่มตัวอย่าง

คำ	หญิง	ชาย	เฉลี่ย
	20 – 25 ปี	20 – 25 ปี	
ฉุกเฉิน	83.74%	75.93%	79.84%
เชื่องช้า	72.50%	72.45%	72.48%
ซ้อน	81.09%	80.87%	80.98%
เสริม	85.93%	84.10%	85.02%
ที่	78.33%	77.41%	77.87%
ถ่วง	95.37%	89.56%	92.47%

5. วิจารณ์ผล

จากผลการทดลองงานวิจัยนี้พบว่า หน่วยเสียงของผู้พูดภาษาไทยสามารถแยกแยะได้ด้วยวิธีการที่ได้นำเสนอ โดยที่ ตัวสะกด ระดับเสียง และตัวสะกดสุดท้ายถูกนำมาวิเคราะห์ด้วยค่าพารามิเตอร์ ZCR และ TD ซึ่งแต่ละคำที่ใช้ทดสอบมีระดับที่แตกต่างกันที่สามารถ

ตารางที่ 3 (ต่อ) ผลการทดลองใช้ซอฟต์แวร์ PATHAI กับกลุ่มตัวอย่าง

คำ	หญิง	ชาย	เฉลี่ย
	20 – 25 ปี	20 – 25 ปี	
ลอตรั้ว	70.87%	71.60%	71.24%
ลงรอย	73.54%	75.86%	74.70%
ง่วง	91.32%	92.43%	91.88%
หงุดหงิด	80.21%	78.06%	79.14%
เจือจาง	79.89%	77.92%	78.91%
ฟุ่มเฟือย	80.30%	82.65%	81.48%

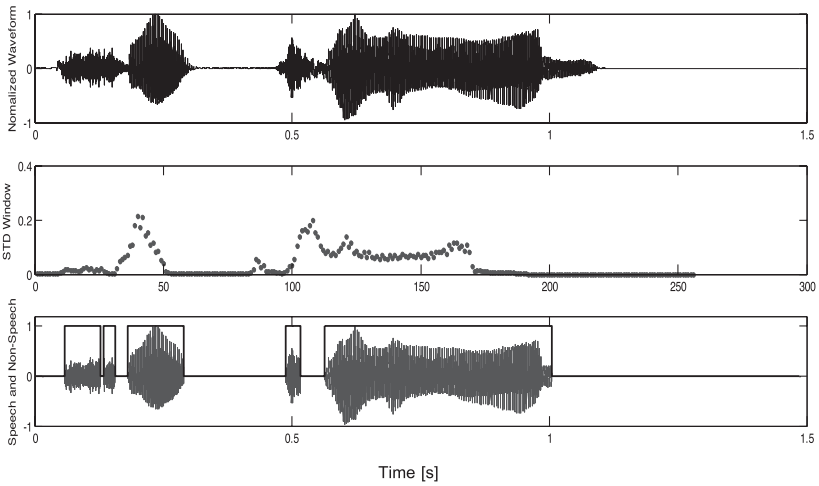
ตารางที่ 4 ผลการทดลองโดยแยกตามจำนวนพยางค์

จำนวนพยางค์	หญิง	ชาย	เฉลี่ย
	20 – 25 ปี	20 – 25 ปี	
หนึ่งพยางค์ *	86.41%	84.87%	85.64%
สองพยางค์ **	77.29%	76.35%	76.83%

(*p < 0.05, **p < 0.1)

จัดกลุ่มเพื่อวินิจฉัยเสียงพูดได้ จากตารางที่ 4 พบว่าผลการประเมินเฉลี่ยของจำนวนหนึ่งพยางค์มีค่าร้อยละ 85.64 สองพยางค์มีค่าร้อยละ 76.83 เนื่องจากหนึ่งพยางค์ใช้พารามิเตอร์ในการประเมิน 1 ชุด (TD และ ZCR) เมื่อเปรียบเทียบกับสองพยางค์ต้องใช้พารามิเตอร์ 2 ชุด ทำให้การเปรียบเทียบเพื่อประเมินผลทำได้ซับซ้อนกว่า เพื่อทดสอบความเชื่อมั่นประสิทธิภาพของซอฟต์แวร์เปรียบเทียบกับผลการประเมินของผู้เชี่ยวชาญด้านการออกเสียงด้วยการวิเคราะห์ผลทางสถิติแบบ T-test การทดสอบที่ระดับความเชื่อมั่นร้อยละ 95 โดยมีค่ามากกว่าค่ามาตรฐานอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95 ($p < 0.05$) และในขณะเดียวกันที่ระดับความเชื่อมั่นร้อยละ 90 ($p < 0.1$) สรุปได้ว่าไม่แตกต่างจากค่ามาตรฐาน อย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95 และที่ระดับความเชื่อมั่นร้อยละ 90 สำหรับจำนวนหนึ่งพยางค์และจำนวนสองพยางค์ตามลำดับ สาเหตุของค่าคลาดเคลื่อนที่เกิดมาจากการนำช่วงหน่วยเสียงย่อยที่แบ่งได้มาเปรียบเทียบกับเสียงต้นแบบ พบว่าจำนวนของช่วงหน่วยเสียงย่อยของจำนวนสองพยางค์การแบ่งหน่วยเสียงทำได้ยากกว่าแบบหนึ่งพยางค์ ดังรูปที่ 6 แต่เมื่อนำเอาค่าพารามิเตอร์ในการวิเคราะห์มาเปรียบเทียบพบว่าค่าที่ได้จากการวิเคราะห์เป็นค่าที่ไม่ถูกต้อง งานวิจัยนี้สรุปปัญหาและแนวทางพัฒนาที่เป็นปัจจัยสำคัญดังนี้

1. ขณะที่กดปุ่มบันทึกเสียงในช่วงแรกมีสัญญาณรบกวนเกิดขึ้นซึ่งเกิดจากไมโครโฟนที่มีคุณภาพต่ำ แนวทางการแก้ไข การเลือกใช้ไมโครโฟนต้องเลือกไมโครโฟนที่มีความไวต่อเสียงและมีความไวต่อการรับสัญญาณเสียงในพื้นที่ที่จำกัด



รูปที่ 6 ค่าคลาดเคลื่อนที่เกิดขึ้นระหว่างการแบ่งช่วงหน่วยเสียง

2. เสียงรบกวนจากบริเวณใกล้เคียงที่ทำการทดสอบ ซึ่งอาจส่งผลให้การทดสอบผิดพลาดได้ แนวทางการแก้ไขควรเลือกสถานที่ใช้ในการทดสอบให้มีความดังของเสียงรบกวนจากบริเวณใกล้เคียงมีค่าความดังไม่เกิน 10 เดซิเบล และควรเพิ่มขั้นตอนการลดสัญญาณเสียงรบกวนเพื่อปรับคุณภาพเสียงให้ดีขึ้นก่อนทำการประมวลผล

6. สรุป

งานวิจัยนี้ได้พัฒนาซอฟต์แวร์ PATHAI สามารถวิเคราะห์การออกเสียงภาษาไทยด้วยการประมวลผลสัญญาณเสียงพูดสามารถใช้ทดสอบกับกลุ่มผู้ทดสอบได้ทั้งผู้หญิงและผู้ชายในระดับที่น่าพอใจ ซอฟต์แวร์สำหรับวิเคราะห์การออกเสียงภาษาไทยด้วยการประมวลผลสัญญาณเสียงพูดสามารถใช้ทดสอบกับกลุ่มผู้ทดสอบได้ทั้งผู้หญิงและผู้ชาย ที่ต้องการทดสอบความถูกต้องของการออกเสียง สำหรับวิเคราะห์การออกเสียงภาษาไทยด้วยการประมวลผลสัญญาณเสียงพูดประเมินผลการทดสอบเสียงออกมาเป็นเกรด ซึ่งแบ่งเกรดที่ใช้ออกเป็น 3 เกรด เนื่องจากเงื่อนไขการทดลองอยู่ภายใต้สภาวะที่ควบคุมเสียงรบกวนจากภายนอก งานวิจัยในอนาคตคือการปรับปรุงประสิทธิภาพของการประเมินให้สูงขึ้นและมีความคงทนต่อสภาพแวดล้อมที่หลากหลายขณะใช้งานจริง

7. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ คณาจารย์ภาควิชาวิศวกรรมไฟฟ้าและวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพาที่ให้การสนับสนุนงานวิจัยนี้ และขอขอบคุณ อาจารย์อภิรักษ์ ชัยปัญหา คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยบูรพา ที่ให้การช่วยเหลือเกี่ยวกับความรู้เรื่องการออกเสียงภาษาไทย รวมถึงนิสิตวิศวกรรมไฟฟ้าที่ให้ความร่วมมือในการบันทึกเสียง

8. เอกสารอ้างอิง

1. ศิริวรรณ์ ฉายะเกษตริน, 2543, “พบผู้นำโครงการหมอภาษาพัฒนาเยาวชน”, *สกุลไทย*, ฉบับที่ 2389 ปีที่ 46.

2. วรรณชัย วรรณสวัสดิ์ และ ยืน ภู่วรวรรณ, 2548, “การเปรียบเทียบการออกเสียงพยัญชนะภาษาไทยระหว่างเสียงต้นฉบับกับเสียงที่สังเคราะห์”, *การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 1.*

3. Jitsup, J., Sritheeravirojana, U., and Udomhunsakul, S., 2007, “Syllable Segmentation of Thai Human Speech using Stationary Wavelet Transform”, *Asia-Pacific Conference on Communications*, pp. 29–32.

4. Chamnongthai, K., Pichitwong, W., and Navaratana Na Ayutthaya, P., 2005, “Final Consonant Segmentation for Thai Syllable by Using Vowel Characteristics and Wavelet Packet Transform”, *ECTI Transaction on Computer and Information Theory*, Vol. 1, No. 1, pp. 50–62.

5. โกวิทย์ พันธุ์ครู, โกสินทร์ จำนงไทย และ เรืองรอง สุลีสถิร, 2550, “ระบบการวิเคราะห์การออกเสียง ร และ ล เพื่อวินิจฉัยความผิดปกติของอวัยวะการออกเสียง”, *การประชุมวิชาการวิศวกรรมการแพทย์ไทย ครั้งที่ 1.*

6. Wallen, E.J. and Hansen, J.H.L., 1996, “A Screening Test for Speech Pathology Assessment using Objective Quality Measures”, *Int. Conf. Spoken Lang. Pro. (ICSLP)*, Vol. 2, pp. 776–779.

7. Xu, W., Zhiyan, H., and Jian W., 2008, “Pathological Speech Deformation Degree Assessment based on Integrating Feature and Neural Network”, *The 27th Chinese Control Conference (CCC)*, pp. 441–444.

8. Zhiyan, H., Xu, W., and Jian W., 2008, “Pathological Speech Deformation Degree Assessment based on Dynamic and Static Feature Integration”, *The 2nd Int. Conf. Bioinfo. and Biomed. Eng.* (ICBBE), pp. 2036–2039.