การแยกพยางค์เสียงพูดภาษาไทยโดยใช้การแปลงเวฟเล็ต

จักรพันธ์ จิตรทรัพย์* อุทัย ศรีธีระวิโรจน์* และ สมเกียรติ อุดมมหรรษากุล*

บทคัดย่อ

งานวิจัยนี้นำเสนอวิธีการแยกพยางค์ของสัญญาณเสียงพูดภาษาไทยในกรณีที่เสียงพูดนั้นติดกัน อันเนื่องมาจากเสียงของผู้พูดมีลักษณะเสียงใหญ่ หรือ มีการพูดที่ค่อนข้างเร็ว จึงทำให้ปลายพยางค์ของเสียงข้างเคียงเกิดการซ้อนทับกัน หรือ เกยกัน (Overlap) ซึ่งในงานวิจัยนี้จะนำขบวนการแปลงเวฟเล็ต (Wavelet transform) 1 ระดับเข้ามาช่วยในการแยกพยางค์ ซึ่งการแปลงเวฟเล็ตจะเป็นการแยกสัญญาณเสียงออกเป็น 2 ส่วน คือ ส่วนของสัญญาณความถี่ต่ำ และ ส่วนของสัญญาณความถี่สูง จากนั้นทำการทิ้งสัญญาณความถี่ต่ำ และนำเฉพาะสัญญาณความถี่สูง (High frequency) มาแปลงเวฟเล็ตกลับ (IDWT) จากนั้นนำสัญญาณเสียงนี้ไปใช้เพื่อช่วยในการแยกพยางค์ของสัญญาณเสียง ซึ่งผลของการนำเอาสัญญาณเสียงความถี่สูงมาวิเคราะห์จะทำให้การแยกพยางค์มีความถูกต้องถึงร้อยละ 98.12%

คำสำคัญ: การแยกพยางค์ การแปลงเวฟเล็ต การหาขอบเขตพยางค์

1. บทนำ

การวิเคราะห์สัญญาณเสียงพูด (Speech) ในบางงานอาจจะสามารถประมวลผลสัญญาณเสียงพร้อมกันทีเดียวหลายพยางค์ แต่ในบางครั้งอาจจะต้องวิเคราะห์สัญญาณเสียงทีละหนึ่งพยางค์ ซึ่งในงานวิจัยนี้จะทำการแยกพยางค์เสียงพูดภาษาไทยที่มีลักษณะสัญญาณเสียงของพยางค์ที่เกยกัน (Overlap) หรือซ้อนทับกันบางส่วน ซึ่งอาจจะเกิดจากลักษณะเสียงของผู้พูดนั้นเป็นเสียงที่ใหญ่ หรือ เสียงพูดเร็วในบางพยางค์

การนำเสียงพูดมาแยกวิเคราะห์ทีละส่วนจะทำให้การวิเคราะห์สัญญาณเสียงนั้นง่ายขึ้น เนื่องจากเป็นการลดความซับซ้อนของการประมวลผลลง แต่เนื่องจากความไม่แน่นอนของลักษณะเสียงพูดที่รับเข้ามาในแต่ละครั้ง รวมถึงลักษณะของสัญญาณเสียงพูดที่มีความสั้นยาวที่ไม่เท่ากัน จึงส่งผลทำให้ยากแก่การหาจุดแบ่งพยางค์ ในงานวิจัยที่ผ่านมาจะใช้วิธีการหาค่าพลังงานของสัญญาณเสียง และ หาค่าการตัดสินใจ (Threshold) เพื่อใช้ในการแยกพยางค์ดังงานวิจัยของ ปิยสวัสดิ์ [1] และชาญชัย [2] แต่วิธีการนี้มีข้อจำกัดคือ ไม่เหมาะกับเสียงพูดที่มีพยางค์ติดกัน เนื่องจากเมื่อพูดเร็วจะทำให้พลังงานของเสียงตรงบริเวณพยางค์ที่พูดติดกันจะรวมกันเป็นกลุ่มข้อมูลเดียวกัน และ ทำให้รอยต่อของพยางค์ไม่ตัดที่แกนเวลา ทำให้เมื่อผ่านการแยกพยางค์ด้วยค่าการตัดสินใจ (Threshold) แล้วอาจจะได้พยางค์ที่ไม่ถูกต้อง เนื่องจากการกำหนดค่าการตัดสินใจมักจะกำหนดให้ต่ำที่สุดเท่าที่จะทำได้ เพื่อให้ได้เนื้อเสียงครบถ้วนที่สุด ในงานวิจัยของ ประณิต [3] เสนอวิธีการแยกพยางค์โดยใช้วิธีการหาค่าพลังงานของสัญญาณเสียงร่วมกับการหาค่าอัตราการตัดศูนย์ (Zero crossing) ของสัญญาณเสียง ซึ่งวิธีนี้จะใช้หลักการหาตำแหน่งของสัญญาณที่ตัดผ่านแกนเวลา ซึ่งอาจจะมีความผิดพลาดในกรณีที่พยางค์นั้นติดกัน ซึ่งจะส่งผลให้รอยต่อของพยางค์ที่ได้จากการหาค่าพลังงานนั้นไม่ตัดผ่านแกนเวลา

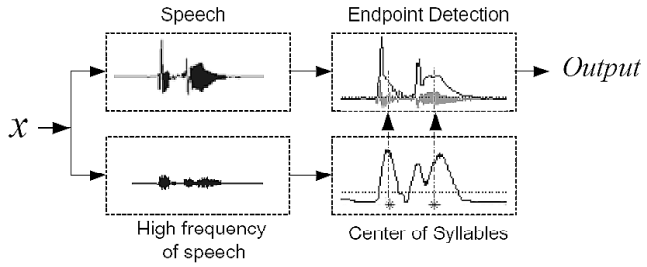
จากงานวิจัยที่ผ่านมาข้างต้นนั้นยังไม่สามารถแยกสัญญาณเสียงออกเป็นพยางค์ได้ในกรณีที่สัญญาณเสียงนั้นมีลักษณะที่ซ้อนทับกัน ซึ่งบทความนี้จะเสนอวิธีการแยกพยางค์เสียงพูดไทยที่เป็นคำประสม (Compound word) โดยใช้การแปลงเวฟเล็ต (Wavelet) เข้ามาช่วยวิเคราะห์สัญญาณเสียงพูดเพื่อที่จะสามารถเห็นกลุ่มพยางค์ได้อย่างชัดเจนมากยิ่งขึ้น และจะส่งผลให้การแยกพยางค์เสียงพูดไทยได้อย่างถูกต้องมากขึ้นตามไปด้วย เพื่อที่จะนำพยางค์ที่ได้แยกแล้วเหล่านี้ไปใช้ในการวิเคราะห์ในรูปแบบอื่นๆ ต่อไป

2. แนวความคิด

จากงานวิจัยที่ผ่านมาได้ใช้วิธีการแยกพยางค์ด้วยบล็อกไดอะแกรมการทำงานเฉพาะส่วนบน ดังแสดงในภาพที่ 1 โดยค่าการตัดสินใจ (Threshold) อาจจะกำหนดไว้ที่ค่าๆ หนึ่ง หรืออาจจะกำหนดเป็นอัตราส่วนตามค่าพลังงานสูงสุดก็ได้ แต่

* *ภาควิชาวิศวกรรมสารสนเทศ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง*

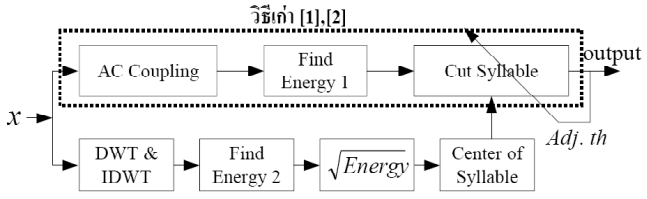
อย่างไรก็ตาม ค่าการตัดสินใจจะต้องพยายามให้มีค่าน้อยที่สุดเท่าที่จะเป็นไปได้ เพราะว่าถ้าค่าการตัดสินใจยิ่งมีค่าสูงเท่าไหร่ก็จะทำให้ส่วนต้น และปลายพยางค์นั้นถูกตัดทิ้งไปมากเท่านั้น ในงานวิจัยนี้จึงได้เพิ่มวิธีการในส่วนล่างของภาพที่ 1 โดยใช้การแปลงเวฟเล็ต (DWT) 1 ระดับ จากนั้นจะทิ้งข้อมูลของสัญญาณความถี่ต่ำ และ ทำการแปลงกลับ (IDWT) ซึ่งผลที่ได้จะทำให้ได้เฉพาะสัญญาณความถี่สูงของสัญญาณเสียงออกมา จากนั้นจะผ่านขบวนการหาค่าพลังงาน และ หาจุดกึ่งกลางของพยางค์ ซึ่งเมื่อถึงตรงจุดนี้จะทำให้ได้จำนวนพยางค์ และตำแหน่งของพยางค์ออกมา เพื่อนำไปใช้ปรับค่าการตัดสินใจที่อยู่ในบล็อก “Cut Syllable” อีกครั้งหนึ่ง ซึ่งเริ่มแรกค่าการตัดสินใจนี้จะถูกกำหนดไว้ที่ 0.1 จากนั้นค่าการตัดสินใจนี้จะเพิ่มขึ้นไปเรื่อยๆ จนกว่าจะแยกพยางค์ได้ตามจำนวนพยางค์ที่ระบุไว้จากบล็อก “Center of Syllable” และ ตำแหน่งของพยางค์ที่แยกได้ต้องตรงกัน



ภาพที่ 1 หลักการของขบวนการแยกพยางค์

3. การออกแบบระบบ

การออกแบบวิธีการแยกพยางค์นี้จะนำวิธีของปิยสวัสดิ์ [1] และ ชาญชัย [2] ซึ่งมีวิธีการแยกพยางค์ที่เหมือนกัน มาปรับปรุงให้มีการแยกพยางค์ที่แม่นยำมากขึ้น ดังแสดงในภาพที่ 2



ภาพที่ 2 บล็อกไดอะแกรมการแยกพยางค์

บล็อกไดอะแกรมในภาพที่ 2 แสดงวิธีการแยกพยางค์ ซึ่งจุดสำคัญจะอยู่ที่บล็อก “Cut Syllable” ซึ่งจะนำผลจากบล็อก “Center of Syllable” มาใช้ในการปรับระดับค่าการตัดสินใจ

3.1 ตัวกรองสัญญาณ AC ผ่าน

สัญญาณเสียงที่รับเข้ามาทางอินพุตจะถูกปรับระดับสัญญาณเฉลี่ยให้อยู่ที่ศูนย์ [2] ซึ่งจะใช้วิธีหาค่าเฉลี่ยของสัญญาณเพื่อทำหน้าที่เป็นตัวกรองสัญญาณ AC ผ่าน (AC Coupling) และ ทำหน้าที่ตัดองค์ประกอบของสัญญาณไฟตรง (DC) ทิ้งไป

$$x(i) = x(i) - \frac{1}{n}\sum_{j=1}^{n} x(j) \qquad (1)$$

โดยที่ $x(i)$ เป็นสัญญาณเสียงอินพุต แล้วนำไปลบกับผลของการหาค่าเฉลี่ยของสัญญาณอินพุตเดิม

3.2 การแปลงเวฟเล็ต และ การแปลงกลับ

ในงานวิจัยนี้จะใช้การแปลงเวฟเล็ต 1 ระดับ และ ใช้ Haar เป็นเวฟเล็ตแม่ (Mother wavelet) ซึ่งผลจะทำให้ได้สัญญาณออกมา 2 ส่วน คือ ส่วนสัญญาณความถี่ต่ำ (Approximation coefficients) และ สัญญาณความถี่สูง (Detail coefficients) จากนั้นทำการทิ้งข้อมูลในส่วนความถี่ต่ำ และ ทำการแปลงกลับ (IDWT) ซึ่งจะทำให้ได้สัญญาณเฉพาะความถี่สูงออกมา เพื่อนำไปใช้ในการหาจำนวนพยางค์ และ ตำแหน่งของพยางค์ในบล็อกการทำงาน “Center of Syllable” ต่อไป การแปลงเวฟเล็ต (DWT) สามารถเขียนเป็นสมการได้ดังนี้

$$y(n) = A \cdot x(n) \qquad (2)$$

เมื่อ A คือ Analysis matrix ซึ่งมีขนาด N×N และ N คือ ความยาวของข้อมูลอินพุต $x(n)$

$$A = \left[\frac{Lp}{Hp} \right] \qquad (3)$$

$$A = \left[\begin{array}{cccccccc} L_1 & L_2 & \mathbf{L} & L_k & 0 & 0 & 0 & \mathbf{L} \\ 0 & 0 & L_1 & L_2 & \mathbf{L} & L_k & 0 & \mathbf{L} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \\ \hline H_1 & H_2 & \mathbf{L} & H_k & 0 & 0 & 0 & \mathbf{L} \\ 0 & 0 & H_1 & H_2 & \mathbf{L} & H_k & 0 & \mathbf{L} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \end{array}\right] \qquad (4)$$

ค่า $y(n)$ ที่ได้ในสมการที่ 2 จะเป็นค่าสัมประสิทธิ์จากการแปลงเวฟเล็ต ซึ่งมีขนาด N×1 โดยข้อมูลครึ่งหนึ่งทางด้านบนของ $y(n)$ จะเป็นส่วนของสัญญาณความถี่ต่ำ และ อีกครึ่งหนึ่งทางด้านล่างจะเป็นส่วนของความถี่สูง จากนั้นทำการทิ้งข้อมูลความถี่ต่ำโดยกำหนดให้เป็นศูนย์ จากนั้นทำการแปลงกลับเวฟเล็ต (IDWT) ดังสมการต่อไปนี้

$$x(n) = \overline{S} \cdot y(n) \quad (5)$$

เมื่อ ตัวแปร $x(n)$ คือ สัญญาณหลังจากผ่านการแปลงกลับเวฟเล็ต และSynthesis matrix, $\overline{S} = A^{-1}$

3.3 การหาค่าพลังงานของสัญญาณเสียง

การหาค่าพลังงานของสัญญาณเสียงนั้นมีวัตถุประสงค์เพื่อที่จะนำไปใช้ในการตรวจหาขอบเขต (Endpoint detection) ของสัญญาณเสียงพูด เมื่อเสียงพูด คือ เสียงที่คนเปล่งออกมาเพื่อใช้ในการสื่อสาร โดยเสียงที่คนเปล่งออกมาอย่างมีจุดหมายจะเรียกว่า เสียง (Voice) และ เสียงที่เกิดจากสภาพแวดล้อมอื่นๆ จะเรียกว่า ไม่ใช่เสียง (Unvoice) หรือสัญญาณรบกวน (Noise) ซึ่งโดยปกติสัญญาณที่ไม่ใช่เสียง จะมีขนาดสัญญาณเสียงเฉลี่ยที่ต่ำ เมื่อเทียบกับสัญญาณเสียง และเนื่องจากสัญญาณเสียงมีการเปลี่ยนแปลงที่แปรตามเวลา และมีการเปลี่ยนแปลงที่ไม่แน่นอน ดังนั้น จึงใช้การหาค่าพลังงานช่วงสั้น (Short time Energy) [2],[4] ในการหาขอบเขตของพยางค์

การวิเคราะห์เพื่อหาขอบเริ่มต้น (Beginning point) และขอบปลาย (Ending point) [5],[6],[7] ของสัญญาณเสียงในงานวิจัยนี้จะใช้หลักการหาค่าพลังงานของสัญญาณด้วยการเอารูปคลื่นสัญญาณเสียงอินพุตไปยกกำลังสอง จากนั้นนำไปผ่านการคอนโวลูชั่น (Convolution) [8] เพื่อแสดงให้เห็นพลังงานของสัญญาณ ดังแสดงในสมการที่ 6 และ 7

$$E(n) = \sum_{m=-\infty}^{\infty} \left[x(m)^2 \cdot w(n-m)^2 \right] \quad (6)$$

เมื่อในทางปฏิบัติขนาดของหน้าต่าง (Window) มีค่าไม่เป็นอนันต์ ดังนั้น

$$E(n) = \sum_{m=0}^{N-1} \left[x(m)^2 \cdot w(n-m)^2 \right] \quad (7)$$

โดยขนาดของ Window ที่ใช้ควรจะมีขนาดที่เหมาะสม คือ ขนาดของ Window ถ้าเลือกขนาดยาวมากพอจะทำให้พลังงานของสัญญาณเสียงมีความราบเรียบ [8] ซึ่งผลจะทำให้เห็นความแตกต่างระหว่างสัญญาณที่เป็นเสียง (Voice) และ ส่วนของสัญญาณที่ไม่ใช่เสียง (Unvoice) ชัดเจนมากยิ่งขึ้น แต่ก็ต้องไม่ยาวเกินไปจนทำให้คุณสมบัติของเสียงที่กำลังพิจารณาอยู่นั้นเปลี่ยนแปลงไป ซึ่งในงานวิจัยนี้จะใช้ขนาดของหน้าต่าง (Window size) เท่ากับ 220 ตัว สำหรับบล็อกการทำงาน “Find Energy1" และ ขนาด 150 ใช้กับบล็อกการทำงาน “Find Energy2” โดยเลือกหน้าต่างเป็นแบบสี่เหลี่ยม (Rectangular window) ดังแสดงในสมการที่ 8

$$w(n) = \begin{cases} 1 & ;0 \le n \le N-1 \\ 0 & ;Otherwise \end{cases} \quad (8)$$

เมื่อ N คือ ขนาดหน้าต่าง (window)

3.4 การหาค่ากึ่งกลางของพยางค์

การหาค่ากึ่งกลางของพยางค์ทำได้โดยการนำเอาพลังงานของสัญญาณเสียงความถี่สูงที่ได้ผ่านการหารากอันดับที่ 2 (Square root) มาแยกพยางค์ด้วยค่าการตัดสินใจ (Threshold) ซึ่งจะกำหนดไว้ที่ 20% ของค่าพลังงานสูงสุดดังแสดงในสมการที่ 9 เมื่อตัดแยกพยางค์แล้ว ณ ตำแหน่งเส้นการตัดสินใจ (Threshold) ลากผ่านจะนำมาหาค่ากึ่งกลางของพยางค์ด้วยสมการที่ 10 โดยขนาดของพยางค์นี้จะต้องมีไม่น้อยกว่า 300 samples และ ระยะห่างระหว่างพยางค์จะต้องมีไม่น้อยกว่า 620 samples

$$th = E_{\max} \times ratio \quad (9)$$

เมื่อกำหนดให้ Ratio = 0.2 หรือ 20% ของค่าพลังงานสูงสุด

$$Cen = \frac{StartPoint + EndPoint}{2} \quad (10)$$

เมื่อตัวแป StartPoint คี จุดเริ่มพยางค์ ณ จุดตัดของค่าการตัดสินใจ (Threshold) แล EndPoint คือ จุดปลายพยางค์ และตั วแปร Cen คี ตำแหน่งกึ่งกลางพยางค์ ซึ่งคือ เครื่องหมาย “*” ดังแสดงในภาพที่ 3

3.5 การแยกพยางค์

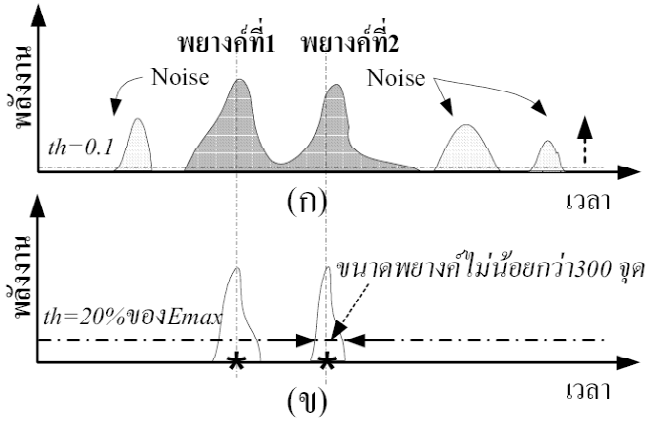
การแยกสัญญาณเสียงออกเป็นพยางค์จะทำในบล็อก “Cut Syllable” โดยจะมีตัวแปรที่สำคัญคือ th ซึ่งจะสามารถปรับเปลี่ยนค่าได้ตามข้อมูลที่ได้จากภาค “Center of Syllable” ซึ่งการแยกพยางค์ มีสมการดังนี้ คือ [2]

$$y(n) = \left[\frac{E(n) - th}{E(n) - th} + 1 \right] \times \left(\frac{1}{2} \right) \quad (11)$$

$$CutOut = y(n) \times x(n) \quad (12)$$

จากสมการที่ 11 ผลลัพธ์ที่ได้จากสมการจะมีอยู่ 2 ค่า คือ “0” หมายถึง ค่าพลังงานที่น้อยกว่าค่า th ซึ่งไม่ใช่พยางค์ และ “1” หมายถึง ค่าพลังงานที่มากกว่าค่า th ซึ่งจะเป็นส่วนของพยางค์ที่ต้องการ ดังนั้น เมื่อนำค่าใน $y(n)$ มาคูณกับ $x(n)$ จะได้พยางค์ที่ต้องการ ซึ่งจะถูกเก็บอยู่ในตัวแปร CutOut ดังแสดงให้เห็นในสมการที่ 12

พยางค์ที่อยู่ในตัวแปร CutOut จะมีข้อมูลบางส่วนที่ไม่ใช่พยางค์ที่แท้จริง ซึ่งเป็นผลอันเนื่องมาจากพลังงานของสัญญาณรบกวนขนาดใหญ่บางส่วนที่อยู่สูงกว่าค่าการตัดสินใจ th ดังนั้น พยางค์เหล่านี้จะถูกตรวจสอบกับค่าตำแหน่งของพยางค์ที่ได้จากบล็อก “Center of Syllable” อีกครั้งหนึ่ง เพื่อให้ได้พยางค์พูดที่ถูกต้องที่สุด ดังแสดงในภาพที่ 3

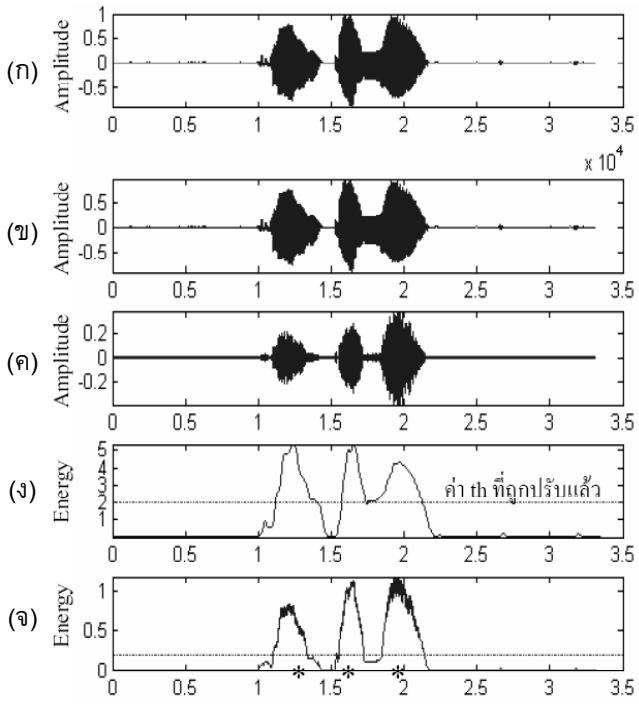


ภาพที่ 3 การแยกพยางค์ออกจากสัญญาณรบกวน [5],[7],[8]

กรณีเสียงพูดที่พิจารณามีลักษณะเสียงใหญ่ หรือ พูดเร็วมากเกินไป อาจก่อให้เกิดปัญหาทำให้การหาค่าพลังงานแล้วปลายของสองพยางค์นั้นเกย (Overlap) กัน จึงทำให้ปลายพยางค์ ไม่ตัดผ่านเส้นการตัดสินใจ (Threshold) ดังแสดงในภาพที่ 3 (ก) เป็นการแสดงให้เห็นถึงพลังงานจากบล็อก “Find Energy1” ซึ่งปลายพยางค์ที่ 1 กับปลายพยางค์ที่ 2 นั้นซ้อนทับกันจนทำให้ปลายพยางค์ทั้ง 2 นั้นอยู่สูงกว่าค่าการตัดสินใจ ซึ่งโดยปกติค่านี้จะกำหนดไว้ให้ต่ำที่สุดเท่าที่จะทำได้เพื่อให้ได้เนื้อพยางค์ครบถ้วนที่สุด โดยในที่นี้กำหนดให้เริ่มต้นที่ 0.1 เนื่องด้วยปัญหานี้ การออกแบบจึงทำให้ค่าการตัดสินใจสามารถปรับเปลี่ยนค่าได้จนกว่าการแยกพยางค์จะได้จำนวนพยางค์ตามค่าที่ได้จากบล็อก “Center of Syllable” โดยค่าการตัดสินใจนี้จะเลื่อนขึ้นไปเรื่อยๆ ส่วนในภาพที่ 3 (ข) แสดงพลังงานของสัญญาณส่วนความถี่สูงที่ได้จากบล็อก “Find Energy2” ซึ่งจะเห็นว่าพลังงานที่ได้จากส่วนนี้ ปลายของแต่ละพยางค์นั้นอยู่ต่ำหรือเห็นขอบเขตของพยางค์ได้ชัดเจนมากขึ้น

4. ผลการทดลอง

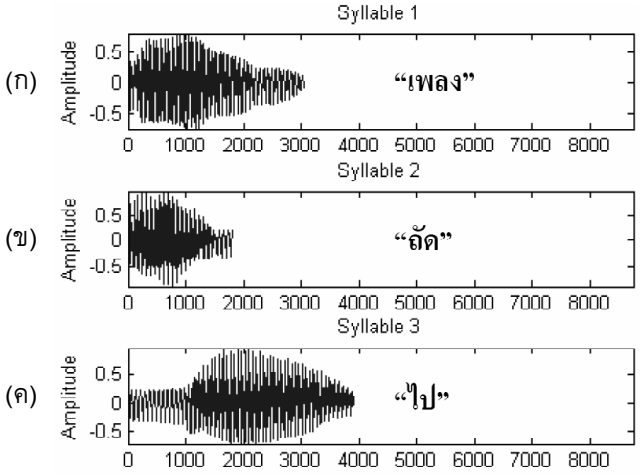
การทดสอบความสามารถของการแยกพยางค์ในงานวิจัยนี้ได้ทดสอบกับคำประสม (Compound word) จำนวน 8 คำ บันทึกเสียงคำละ 3 ครั้ง กับชาย 5 คน และ ผู้หญิง 5 คน มีอายุระหว่าง 19 ถึง 25 ปีรวมเป็น 240 คำ โดยคำที่ใช้ทดสอบจะมีทั้งคำละ 2 พยางค์ และ 3 พยางค์ ซึ่งได้แก่คำว่า “เปิดเครื่อง” “ปิดเครื่อง” “เพลงก่อนหน้า” “เพลงถัดไป” “เพิ่มเสียง” “ลดเสียง” “ปิดเสียง” และ “หยุดเพลง” ด้วยอัตราการสุ่มสัญญาณ (Sampling rate) เท่ากับ 11,025Hz รับเสียงเข้ามาครั้งละ 3 วินาที



ภาพที่ 4 ผลลัพธ์ในแต่ละขั้นตอนของการแยกพยางค์ “เพลงถัดไป”

สัญญาณเสียงของการทำงานในแต่ละขั้นตอนแสดงให้เห็นในภาพที่ 4 โดยภาพที่ 4(ก) เป็นการแสดงสัญญาณเสียงต้นแบบของคำว่า “เพลงถัดไป” ภาพที่ 4 (ข) แสดงสัญญาณเสียงที่ผ่านตัวกรองไฟ AC ผ่าน เพื่อปรับให้ค่าเฉลี่ยของสัญญาณอยู่ที่ระดับศูนย์ ภาพที่ 4(ค) แสดงสัญญาณเสียงความถี่สูงที่ได้จากการแปลงเวฟเล็ตหนึ่งระดับ ซึ่งจะเห็นว่าขอบเขตระหว่างพยางค์ที่ 2 และ ที่ 3 นั้นถูกแยกได้ชัดเจน

มากกว่าภาพที่ 4(ข) ภาพที่ 4(ง) แสดงการหาค่าพลังงานของสัญญาณเสียงที่ได้จากภาค "AC Coupling" ซึ่งจะเห็นว่าพลังงานของปลายพยางค์ที่ 2 และ 3 นั้นอยู่สูงมาก ภาพที่ 4(จ) แสดงการหาค่าพลังงานของสัญญาณความถี่สูงที่ได้จากการแปลงเวฟเล็ต ซึ่งจะเห็นว่าขอบเขตของพยางค์ชัดเจนกว่าภาพที่ 4(ง) จากนั้นจะผ่านกระบวน การหาจุดศูนย์กลางของพยางค์ ซึ่งจะทำให้ได้จำนวนพยางค์ และ ตำแหน่งของพยางค์ออกมาแสดงด้วยเครื่องหมาย "*"ในภาพที่ 4(จ)



ภาพที่ 5 ผลลัพธ์ของการแยกพยางค์ของคำว่า "เพลงถัดไป"

ภาพที่ 5 แสดงผลลัพธ์ของการแยกพยางค์ของคำว่า "เพลงถัดไป" ซึ่งจากการทดลองพบว่าวิธีที่นำเสนอมีความถูกต้องถึงร้อยละ 98.12% ในขณะที่วิธีเก่า[1],[2] ที่ใช้ในการทดสอบมีความถูกต้องเพียง 93.13% ซึ่งผลของการทดลองแยกพยางค์ของแต่ละคำ ดังแสดงไว้ในตารางที่ 1

ตารางที่ 1 ผลลัพธ์ที่ได้จากการทดลองวิธีใหม่ เปรียบเทียบกับวิธีเก่า

คำพูด	ชาย		หญิง	
	วิธีใหม่	วิธีเก่า	วิธีใหม่	วิธีเก่า
"เปิดเครื่อง"	100%	93.33%	93.33%	93.33%
"ปิดเครื่อง"	100%	100%	93.33%	93.33%
"เพลงก่อนหน้า"	93.33%	60%	100%	53.33%
"เพลงถัดไป"	100%	66.67%	100%	73.33%
"เพิ่มเสียง"	86.67%	100%	100%	100%
"ลดเสียง"	93.33%	80%	100%	93.33%
"ปิดเสียง"	100%	100%	93.33%	80%
"หยุดเพลง"	93.33%	93.33%	100%	93.33%

จากการทดลองมีบางคำสำหรับวิธีที่นำเสนอที่ให้เปอร์เซ็นต์ไม่ดีนัก คือ คำว่า "เพิ่มเสียง" ซึ่งผลที่ไม่ดีนี้เกิดจากเสียงผู้พูดเพียงคนเดียว ซึ่งหมายความว่าโทนความถี่เสียงของผู้พูดคนนี้เมื่อผ่านการทิ้งข้อมูลความถี่ต่ำ ด้วยการแปลงเวฟเล็ตแล้วเสียงแต่ละพยางค์ยังไม่แยกกันอย่างชัดเจน จึงทำให้การแบ่งพยางค์นั้นผิดพลาด แต่สำหรับคำพูดที่มี 3 พยางค์วิธีที่นำเสนอนี้ทำได้ดีกว่า และเมื่อพิจารณาเปอร์เซ็นต์รวมทั้งหมดจะเห็นว่าวิธีที่นำเสนอนี้มีประสิทธิภาพสูงกว่า

5. สรุป

การแยกพยางค์สัญญาณเสียงพูดภาษาไทยโดยใช้การแปลงเวฟเล็ตเข้ามาช่วยในการแยกพยางค์นั้น ทำให้ความสามารถในการแยกพยางค์มีความแม่นยำมากยิ่งขึ้น ดังที่ได้แสดงให้เห็นแล้วในตารางที่ 1 ซึ่งวิธีการแยกพยางค์นี้สามารถนำไปใช้กับงานเปลี่ยนเสียงไปเป็นข้อความ (Speech to text) ได้ หรือ จะนำไปใช้กับการรู้จำเสียง (Speech recognition) ก็ได้เช่นกัน

เอกสารอ้างอิง

[1] ปิยะสวัสดิ์ นวรัตน์ ณ อยุธยา, 2542, "การดึงคุณลักษณะและการลดขนาดข้อมูลโดยอาศัยแบบโสตรับเสียง และเวฟเล็ต สำหรับการรู้จำเสียงภาษาไทย", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

[2] ชาญชัย นุชนารถ, ปณิธาน จันทรอุดร และวิทยา ฤกษ์สำเร็จ,2544, "การทำฮาร์ดแวร์ระบบรู้จำเสียงภาษาไทย ระยะที่1", โครงงานวิศวกรรมศาสตรบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.

[3] ประณิต อ่อนไสว และ พรชัย ธรรมานุรัตพันธุ์ "การจดจำเสียงพูด" ปริญญานิพนธ์ ภาควิชาวิศวกรรมโทรคมนาคม สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2542.

[4] Navaratana, P., Khawparisuth, D. and Chamnongthai, K., "Speaker-Independent Isolated Thai Consonant Recognition by using Simulated Auditory system", IEEE International Symposium on Intelligent Signal Processing and Communication Systems, 1999, pp.797-800.

[5] Ahkuputra, V., Jitapunkul, S., Pornsukchandra, W.,

"A Speaker-Independent Thai Polysyllabic Word Recognition Using Hidden Markov Model" 1997 IEEE Pacific Rim Conference, pp. 593 – 599.

[6] Thomas W.Parsons, "Voice and Speech Processing", McGraw-Hill Book Company, 1987.

[7] Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing : Processing and Perception of Speech and Music", John Wiley & Sons, Inc, 2000.

[8] Sophocles J. Orfanidis, "Introduction to Signal Processing", Prentice Hall, Inc, 1996.

[9] Pensiri, R., Jitapunkul, S., "Speaker-Independent Thai Numerical Voice Recognition by using Dynamic Time Warping", 18th Electrical Engineering Conference, 22-24 November, 1995, pp.977-981.