บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

โดยปกติมนุษย์มักจะสื่อสารกันด้วยเสียงพูด และการส่งเสียงพูดเป็นลักษณะเฉพาะของมนุษย์เท่านั้น โดยจังหวะการหายใจของมนุษย์จะสั้นยาวเปลี่ยนแปลงไปตามจำนวนพยางค์ และตามความต่อเนื่องของพยางค์ ซึ่งเอื้ออำนวยให้มนุษย์สามารถพูดติดต่อกันได้เป็นเวลานาน และการที่มนุษย์สามารถพูดต่อเนื่องกันได้เป็นเวลานาน จึงทำให้การติดต่อสื่อสารด้วยเสียงพูดสะดวกและรวดเร็ว แต่เมื่อต้องการนำเสียงพูดมาติดต่อสื่อสารระหว่างมนุษย์กับคอมพิวเตอร์ การพูดต่อเนื่องจึงเป็นปัญหาสำคัญในการรู้จำเสียงพูด (Speech Recognition) สำหรับคอมพิวเตอร์ ดังนั้นในระบบรู้จำเสียงพูดจึงจำเป็นต้องมีขั้นตอนการแยกเสียงพูดออกเป็นหน่วยภาษาต่าง ๆ ตามระบบของการรู้จำเสียงพูด เช่น

1. หน่วยภาษาที่เปล่งออกมาทั้งชุด (Word-Based) หน่วยภาษานี้อาจจะเป็นพยางค์ คำ กลุ่มคำ วลี หรือประโยค ระบบรู้จำเสียงพูดที่ใช้หน่วยภาษานี้ จะหลีกเลี่ยงผลกระทบอันเนื่องมาจากฐานเสียงภายในคำ หรือกลุ่มคำนั้น
2. หน่วยเสียง (Phoneme-Based) จะพิจารณาลักษณะของหน่วยเสียงที่เล็กลงไป เช่น หน่วยเสียงพยัญชนะ สระ หรือวรรณยุกต์ ซึ่งจะทำให้ระบบรู้จำเสียงพูดสามารถรู้จำเสียงพูดได้เป็นจำนวนมาก

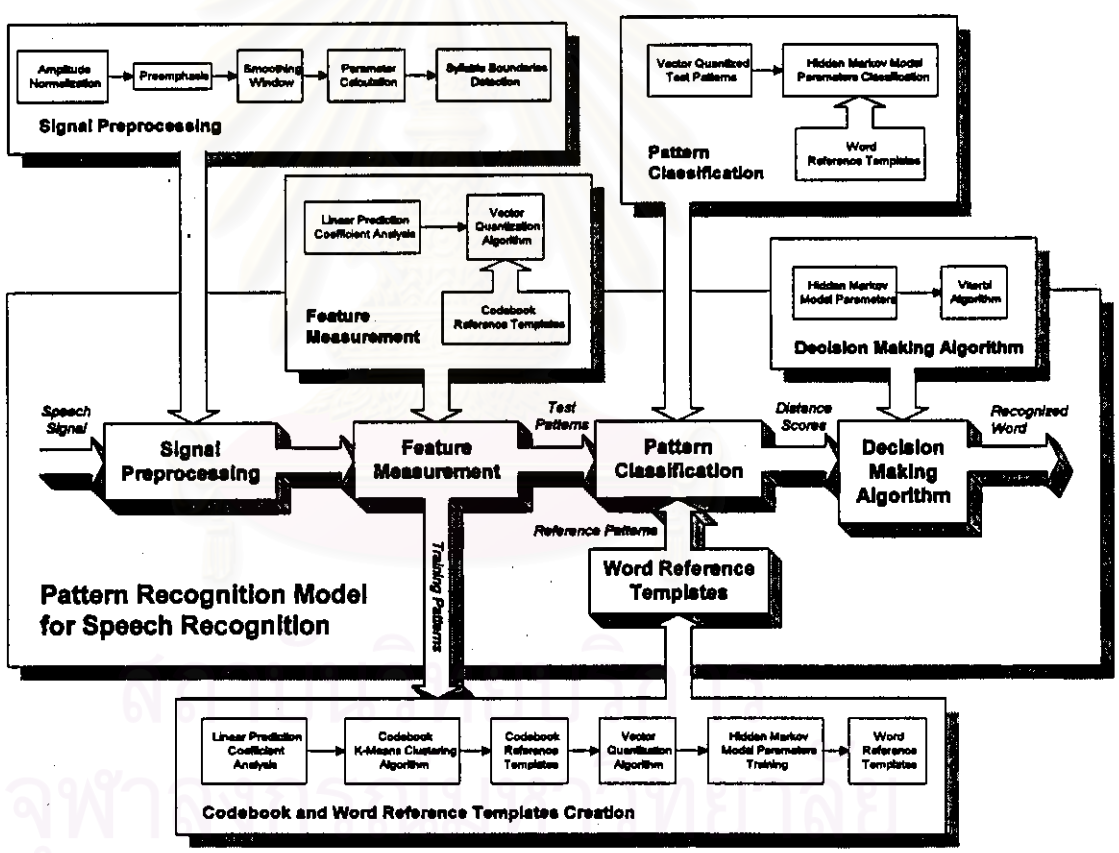
ในการเปล่งเสียงพูด ผู้พูดไม่ได้เปล่งเสียงพยัญชนะ สระ และวรรณยุกต์ แยกกัน แต่จะเปล่งเป็นพยางค์ (Syllable) ต่อเนื่องกันไป ซึ่งพยางค์จะประกอบด้วย สระเป็นใจกลางของพยางค์ (Nucleus) และพยัญชนะเป็นส่วนขอบของพยางค์ (Marginal Sound) ดังนั้นถ้าเราสามารถกำหนดขอบเขตพยางค์ในเสียงพูดต่อเนื่องได้ ก็จะให้คอมพิวเตอร์สามารถรู้จำคำพูดต่อเนื่องได้ง่ายขึ้น (ทวี ประทุมทาน, 2530)

กรรมวิธีการหาจุดต้นและจุดปลายของคำ (Endpoint Detection Algorithm) ในงานวิจัยระบบรู้จำคำเดี่ยว (Isolate Word Recognition) (ไพศาล ธรรมโพธิทอง, 2533) และระบบรู้จำคำไทยหลายพยางค์ (Polysyllabic Word) แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (วิศรุต อาขุบุตร, 2539) ใช้คุณสมบัติของเสียงพูดคือ พลังงาน (Energy Parameter) และระดับพลังงาน (Energy Level) ของสัญญาณเสียงพูด ในการตรวจหาจุดต้นและจุดปลายของคำ ส่วนในระบบรู้จำคำพูดต่อเนื่อง (Connected Speech) ใช้คุณสมบัติของพลังงาน ความถี่มูลฐาน (Fundamental Frequency) และค่าระยะเวลา (Duration) ของสัญญาณเสียงพูด ในการหาขอบเขตพยางค์ (Boundaries Detection) (ทวี ประทุมทาน, 2530)

สำหรับในงานวิจัยนี้ได้เล็งเห็นว่า อัตราความถูกต้องในการรู้จำของระบบรู้จำคำพูดต่อเนื่อง จะขึ้นอยู่กับประสิทธิภาพของการหาขอบเขตพยางค์ หรือการหาขอบเขตพยางค์ได้อย่างถูก

ต้อง จะทำให้ระบบรู้จำคำพูดต่อเนื่องมีความถูกต้องและรวดเร็วมากยิ่งขึ้น ดังรูปที่ 1 แสดงแบบจำลองของระบบรู้จำเสียงพูดภาษาไทย ซึ่งแสดงให้เห็นความสัมพันธ์กันระหว่างระบบรู้จำเสียงพูดกับการหาขอบเขตพยางค์ กรรมวิธีการหาขอบเขตพยางค์จะอยู่ในขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) ของระบบรู้จำเสียงพูด ซึ่งในกรรมวิธีการหาขอบเขตพยางค์ประกอบด้วยขั้นตอนต่าง ๆ ดังนี้คือ

1. การปรับบรรทัดฐานแอมปลิจูด (Amplitude Normalization)
2. การเน้นล่วงหน้า (Preemphasis)
3. การวางกรอบหน้าต่าง (Windowing)
4. การคำนวณค่าพารามิเตอร์ (Parameter Calculation)
5. การหาขอบเขตพยางค์ (Syllable Boundaries Detection)



รูปที่ 1.1. แบบจำลองระบบรู้จำเสียงพูดภาษาไทย (วิศรุต อาขุบุตร, 2539)

เบรน แมค (Brain MAK, 1992) เสนอว่ากรรมวิธีการหาขอบเขตพยางค์ที่ดี ต้องมีคุณสมบัติดังนี้คือ

1. เป็นวิธีที่ง่าย (Simplicity) และใช้เวลาในการคำนวณน้อย (Real Time) เนื่องจากขั้นตอนการหาขอบเขตพยางค์เป็นขั้นตอนแรกของระบบรู้จำเสียงพูด วิธีการหาขอบเขตพยางค์ที่ง่าย และใช้เวลาในการคำนวณน้อย จะส่งผลให้เวลาที่ใช้ทดสอบระบบรู้จำเสียงพูดน้อยลงด้วย
2. ต้องมีความทนต่อสัญญาณรบกวน (Robustness) และไม่ทราบชนิดของสัญญาณรบกวนล่วงหน้า (No Prior Knowledge of Noise) ในสัญญาณเสียงพูด จะมีสัญญาณรบกวนปนอยู่ด้วย ซึ่งอาจเกิดจากสภาพแวดล้อม เช่น เสียงคนเดิน เสียงประตูปิด-เปิด เสียงฝนตก เป็นต้น หรืออาจเกิดจากผู้พูด เช่น เสียงริมฝีปาก และลิ้น ที่เกิดขึ้นจากการเตรียมการพูด หรือเสียงลมหายใจ เป็นต้น
3. มีความเชื่อถือได้ (Reliability) เนื่องจากสัญญาณเสียงพูดเป็นสัญญาณที่ถูกทำให้ผิดเพี้ยน (distort) ได้ง่าย จึงทำให้สัญญาณเสียงพูดแตกต่างกันไป ขึ้นอยู่กับการเน้นเสียง (Stress) ความดัง (Loundness) ความเร็วในการพูด (Tempo) ระดับเสียง (Pitch) จังหวะและลีลาการพูดของแต่ละคน ดังนั้นวิธีการหาขอบเขตพยางค์ที่มีประสิทธิภาพจึงควรจะสามารถใช้งานได้ดี แม้ว่าสัญญาณเสียงพูดมีการเปลี่ยนไปตามลักษณะดังกล่าวข้างต้น

ในงานวิจัยนี้ทำการศึกษากรรมวิธีต่าง ๆ ที่ใช้ในการหาขอบเขตพยางค์ โดยยึดหลักคุณสมบัติในด้านที่เป็นวิธีที่ง่ายและใช้เวลาในการคำนวณน้อย โดยใช้คุณลักษณะ (Feature) พลังงานของสัญญาณเสียงพูด ดังต่อไปนี้ คือ

- พลังงานสัมบูรณ์ (Absolute Energy) (วิศรุต อาชุบุตร, 2539)
- พลังงานเฉลี่ย (Root Mean Square Energy) (Ying, G.S., 1993)
- พลังงานความถี่และเวลา (Frequency-Time Energy) (Jean-Claude Junqua, 1994)
- พลังงานกำลังสอง (Square Energy) (Phillips, W.J., 1993)
- พลังงาน Teager (Teager Energy) (Ying, G.S., 1993)
- พลังงานจากการแปลงแบบ Walsh (Walsh Transform Energy) (John Huang, 1991)

ส่วนคุณสมบัติที่ดีในการหาขอบเขตพยางค์ด้านที่มีความทนต่อสัญญาณรบกวนและมีความเชื่อถือได้ จะใช้คุณลักษณะพลังงานร่วมกับอัตราการตัดผ่านระดับกำหนด (Band Crossing Rate, Taboada T., 1994) และความถี่มูลฐาน (Fundamental Frequency)

1.2 วัตถุประสงค์

1. เพื่อพัฒนากรรมวิธีที่มีประสิทธิภาพในการหาขอบเขตพยางค์ของเสียงพูดต่อเนื่องภาษาไทย

2. เพื่อเป็นแนวทางในการวิจัยค้นคว้าวิธีการหาขอบเขตพยางค์วิธีอื่นต่อไป

1.3 เป้าหมายและขอบเขต

ได้กรรมวิธีการหาขอบเขตพยางค์ ที่มีความถูกต้องไม่น้อยกว่า ร้อยละ 90

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เรียนรู้และเข้าใจสมรรถนะของกรรมวิธีแบบต่าง ๆ ที่ใช้ในการหาขอบเขตพยางค์
2. ได้กรรมวิธีในการหาขอบเขตพยางค์ที่มีประสิทธิภาพ ในงานรู้จำเสียงพูดต่อเนื่องภาษาไทย

1.5 ขั้นตอนและวิธีดำเนินการ

1. ศึกษาคุณลักษณะของเสียงพูด โครงสร้างพยางค์ และปัญหาในการกำหนดขอบเขตพยางค์
2. ค้นคว้าและเก็บรวบรวมข้อมูลรายละเอียดที่เกี่ยวข้องกับวิธีการหาค่าคุณลักษณะที่ใช้ในการหาขอบเขตพยางค์
3. เก็บข้อมูลเสียงพูดภาษาไทยของกลุ่มตัวอย่าง ด้วยวิธีขออาสาสมัครโดยการปิดประกาศ ได้กลุ่มตัวอย่างจำนวน 28 คน แบ่งเป็นชาย 15 คน และหญิง 13 คน มีอายุตั้งแต่ 15 – 50 ปี
4. วิเคราะห์และพัฒนาโปรแกรมการหาขอบเขตพยางค์
5. ทดสอบความถูกต้องในการหาขอบเขตพยางค์
6. ปรับปรุงแก้ไขโปรแกรม เพื่อเพิ่มประสิทธิภาพในการหาขอบเขตพยางค์
7. สรุปรวบรวมผลการวิจัยทั้งหมด พร้อมทั้งจัดทำเอกสารเกี่ยวกับวิทยานิพนธ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย