



## บทที่ 1

### บทนำ

#### ความเป็นมาและความสำคัญของปัญหา

หน่วยเสียง คือ ส่วนย่อยของเสียงพูดที่มีขนาดเล็กที่สุดที่ทำให้เสียงของคำในภาษามีความหลากหลายแตกต่างกัน และขอบเขตของหน่วยเสียงก็คือตำแหน่งบอกเวลาเริ่มต้นและสิ้นสุดของหน่วยเสียงนั้นๆ โดยส่วนใหญ่ระบบรู้จำเสียงพูดจะแยกความแตกต่างระหว่างเสียงของคำในภาษาด้วยการพิจารณารูปแบบการประกอบกันของหน่วยเสียง เช่นเดียวกันกับระบบสังเคราะห์เสียง ที่นำหน่วยเสียงย่อยๆ มาประกอบกลับเป็นเสียงของคำ ด้วยเหตุนี้สิ่งที่จำเป็นต่อการสร้างระบบการรู้จำและการสังเคราะห์เสียงในภาษา คือการมีข้อมูลเสียงที่มีการกำกับขอบเขตของหน่วยเสียง หรือจัดวางตำแหน่งทางเวลาของแต่ละหน่วยเสียงไว้ก่อนแล้ว

ในการรู้จำเสียงพูดแบบอาศัยเซกเมนต์ (Segment-based Speech Recognition) [1] จะพิจารณาหน่วยเสียงให้เป็นเซกเมนต์ โดยขั้นตอนการทำงานจะประกอบไปด้วยขั้นตอนย่อยสองขั้นตอนที่ทำงานต่อเนื่องกัน คือ ขั้นตอนการแบ่งเสียงพูดเป็นเซกเมนต์ และขั้นตอนการรู้จำเสียงพูด จุดประสงค์ของขั้นตอนการแบ่งเสียงพูดเป็นเซกเมนต์ ก็เพื่อสันนิษฐานหาขอบเขตของเซกเมนต์ แล้วนำมาประกอบกันเป็นกราฟของเซกเมนต์ โดยในขั้นตอนการรู้จำเสียงพูด กราฟนี้จะถูกใช้เป็นข้อมูลอินพุตเพื่อค้นหาลำดับของหน่วยเสียงที่ดีที่สุดออกมา ด้วยเหตุนี้สิ่งที่จำเป็นและส่งผลโดยตรงต่อประสิทธิภาพในการรู้จำเสียงพูดแบบอาศัยเซกเมนต์ คือประสิทธิภาพของกระบวนการแบ่งเสียงพูดเป็นเซกเมนต์ ซึ่งจะต้องมีความถูกต้องแม่นยำและทำงานได้อย่างรวดเร็ว

การแบ่งเสียงพูดเป็นเซกเมนต์สามารถกระทำได้ด้วยคนหรือด้วยวิธีแบบอัตโนมัติ แม้ว่า การกระทำด้วยคนจะได้ผลลัพธ์ที่มีความแม่นยำสูงกว่า แต่ก็อาศัยเวลามากเมื่อมีข้อมูลเป็นจำนวนมาก โดยในการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยคนจะใช้เวลาประมาณ 11 ถึง 30 วินาทีต่อหนึ่งหน่วยเสียง [2] ในระหว่างที่การแบ่งเสียงพูดเป็นเซกเมนต์ด้วยกระบวนการแบบอัตโนมัติใช้เวลาอยู่ระหว่าง 0.001 ถึง 0.009 วินาทีต่อหนึ่งหน่วยเสียง บนเครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผลกลางเพนเทียมเอ็ม 1.4 กิกะเฮิร์ต ความแตกต่างนี้จะมากยิ่งขึ้นเมื่อประสิทธิภาพและความเร็วในการประมวลผลบนคอมพิวเตอร์ได้รับการพัฒนามากขึ้น ในขณะที่ความเร็วของคนในการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยมือจะยังคงเท่าเดิม

วิธีการแบ่งเสียงพูดเป็นเซกเมนต์ในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ในปัจจุบันยังมีประสิทธิภาพไม่ดีพอ โดยระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ SUMMIT ของ MIT [3] ซึ่งใช้วิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด มีประสิทธิภาพต่ำกว่าวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยคนอยู่ประมาณ 20 ถึง 30% เมื่อทดลองกับฐานข้อมูลเสียง TIMIT [4]

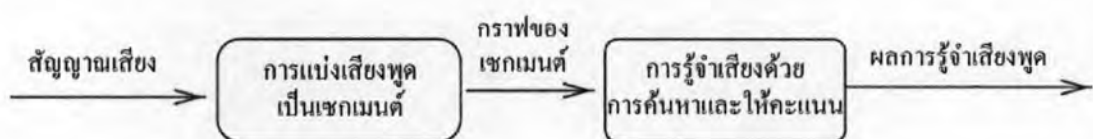
วิทยานิพนธ์นี้เสนอวิธีการแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดแบบอาศัยเซกเมนต์แบบใหม่ โดยนำสารสนเทศสวนศาสตร์ (Acoustic-Phonetic Information) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) มาประยุกต์ใช้เพื่อให้ประสิทธิภาพทั้งในด้านคุณภาพกราฟของเซกเมนต์และความเร็วในการแบ่งเสียงพูดเป็นเซกเมนต์ สูงกว่าวิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด

### วัตถุประสงค์ของการวิจัย

เป้าหมายของวิทยานิพนธ์นี้ ต้องการที่จะพัฒนาวิธีการแบ่งเสียงพูดเป็นเซกเมนต์เพื่อนำไปใช้งานในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ โดยอาศัยเทคนิคการเรียนรู้ของเครื่องแบบซัพพอร์ตเวกเตอร์แมชชีน และการสกัดลักษณะสำคัญโดยใช้สารสนเทศสวนศาสตร์เพื่อเพิ่มความแม่นยำในหาขอบเขตของหน่วยเสียง ซึ่งส่งผลให้ได้กราฟของเซกเมนต์ที่มีคุณภาพดีกว่า รวมถึงสามารถทำงานได้รวดเร็วกว่าวิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด

### ขอบเขตของการวิจัย

ในขอบเขตของงานวิจัยนี้จะพัฒนาเฉพาะส่วนของขั้นตอนการแบ่งเสียงพูดเป็นเซกเมนต์เพียงอย่างเดียวเท่านั้น ไม่ได้พัฒนาในส่วนของการรู้จำเสียงพูดด้วยการค้นหาและให้คะแนนในขอบข่ายงานของการรู้จำเสียงพูดแบบอาศัยเซกเมนต์ ส่วนการวัดประสิทธิภาพจะทำการทดลองเปรียบเทียบโดยใช้ฐานข้อมูลเสียงที่มีข้อมูลเวลากำกับขอบเขตของหน่วยเสียงไว้ก่อนแล้ว



รูปที่ 1.1 แผนภาพส่วนประกอบในระบบการรู้จำเสียงพูดแบบอาศัยเซกเมนต์

### ขั้นตอนการวิจัย

#### 1. ขั้นตอนเตรียมตัว

1. ศึกษาข้อมูลและทฤษฎีต่างๆเกี่ยวกับงานวิจัย เช่น ทฤษฎีทางภาษาศาสตร์ การวิเคราะห์เสียงพูด แบบจำลองทางภาษา การรู้จำเสียงแบบอาศัยเซกเมนต์ การตรวจหาขอบเขตของหน่วยเสียงในเสียงพูด และความรู้อื่นๆ

2. ศึกษาเกี่ยวกับสารสนเทศสวนศัทศาสตร์
3. ศึกษางานวิจัยที่เกี่ยวข้อง
2. ขั้นตอนออกแบบวิธีการในการแบ่งเสียงพูดเป็นเซกเมนต์
  1. ศึกษาคุณสมบัติของหน่วยเสียงและออกแบบเขตของค่าลักษณะสำคัญที่เหมาะสมต่อการนำไปใช้ในกระบวนการตรวจหาขอบเขตของหน่วยเสียงในเสียงพูด
  2. ออกแบบขั้นตอนในการแบ่งเสียงพูดเป็นเซกเมนต์
3. ขั้นตอนพัฒนาและทดสอบระบบการแบ่งเสียงพูดเป็นเซกเมนต์
  1. จัดทำโปรแกรมในแต่ละส่วนที่ได้ออกแบบไว้สำหรับใช้ทำการแบ่งเสียงพูดเป็นเซกเมนต์
  2. ทดสอบประสิทธิภาพของการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยการทดสอบกับฐานข้อมูลเสียงภาษาไทย
4. ขั้นตอนการวิเคราะห์และสรุปผล
  1. วิเคราะห์และสรุปผลการทดลอง
  2. เรียบเรียงวิทยานิพนธ์

### ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถปรับปรุงและพัฒนาวิธีการแบ่งเสียงพูดเป็นเซกเมนต์บนโดเมนภาษาไทยไปใช้สร้างระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ที่มีประสิทธิภาพได้
2. นำไปใช้ประโยชน์ในการเตรียมข้อมูลเสียงสำหรับระบบสังเคราะห์เสียงภาษาไทยและประยุกต์ใช้กับเทคโนโลยีเสียงพูดอื่นๆ ได้ด้วย

### ผลงานที่ตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตอบรับให้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “Locating Phone Boundaries from Acoustic Discontinuities using a Two-staged Approach” โดย ไพโรจน์ ลีลาภทกรกิจ อติวงศ์ สุชาโต และ โปรดปราน บุญยพุกกณะ ในงานประชุมวิชาการ “The Ninth International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)” ณ เมืองฟิตส์เบิร์ก รัฐเพนซิลเวเนีย ประเทศสหรัฐอเมริกา ในระหว่างวันที่ 17-21 กันยายน 2549