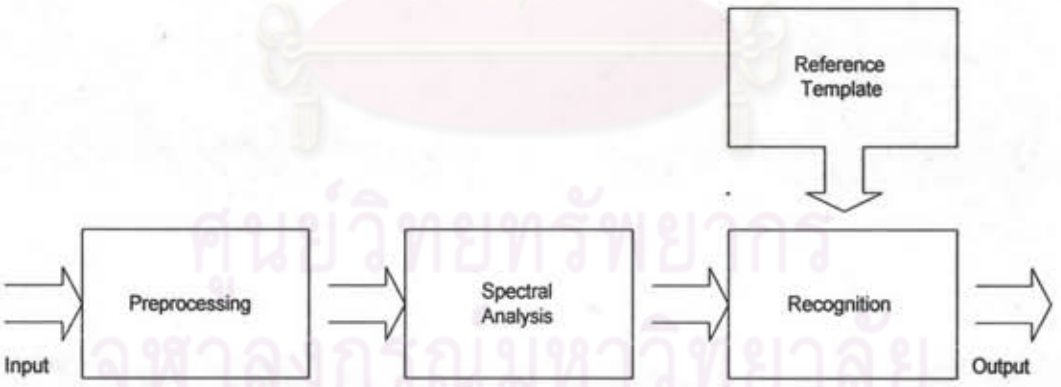
บทที่ 1

บทนำ

ความเป็นมาของปัญหา

ในปัจจุบันได้มีความพยายามในการพัฒนาวิธีการใช้คอมพิวเตอร์ให้ง่ายและสะดวกยิ่งขึ้น ไม่จำเป็นที่ผู้ใช้จะต้องจดจำคำสั่งมากมายเหมือนเมื่อก่อน โดยมีระบบติดต่อกับผู้ใช้แบบรูปภาพ (Graphic User Interface) เข้ามาแทนระบบการสั่งงานแบบบรรทัด (Command Line) ผู้ใช้เพียงแต่ใช้ตัวชี้ชี้ไปยังสัญญลักษณ์หรือรูปภาพ แทนการสั่งงานของคำสั่งต่างๆ ทำให้ลดความสลับซับซ้อนและความยุ่งยากในการทำงานลง

นอกจากระบบการติดต่อกับผู้ใช้แบบรูปภาพแล้ว การสั่งงานด้วยเสียงคงจะเป็นระบบมาตรฐานของการสั่งงานบนระบบคอมพิวเตอร์ในระยะเวลาอันใกล้นี้ เนื่องจากการ์ดเสียงซึ่งเป็นอุปกรณ์ที่สามารถติดตั้งเพิ่มเติมบนเครื่องคอมพิวเตอร์แบบ พีซี มีราคาถูกลงอย่างมาก จนเกือบจะกลายเป็นอุปกรณ์มาตรฐานของเครื่องคอมพิวเตอร์แบบ พีซี ไปแล้ว ดังนั้นจึงเป็นการเหมาะสมสำหรับการพัฒนาระบบการรู้จำเสียงพูดบนเครื่องคอมพิวเตอร์แบบ พีซี เพื่อเป็นพื้นฐานของการศึกษาและพัฒนาระบบสั่งงานด้วยเสียงพูดต่อไป



รูปที่ 1.1 บล็อกไดอะแกรมของระบบรู้จำเสียงพูด

ระบบการรู้จำเสียงพูดโดยทั่วไปจะสามารถแบ่งออกเป็นส่วนต่าง ๆ ตามบล็อกไดอะแกรมในรูปที่ 1.1 ซึ่งประกอบไปด้วยส่วนพรีโพรเซสซิง ส่วนการวิเคราะห์สเปคตรัม ส่วนแบบอ้างอิง และส่วนการรู้จำเสียงพูด สำหรับในส่วนของพรีโพรเซสซิง จะเป็นขั้นตอนการลดสัญญาณรบกวน การปรับข้อมูลจากอินพุตให้เหมาะสม ก่อนจะส่งต่อไปในส่วนการวิเคราะห์สเปคตรัม ซึ่งวิธีที่ใช้วิเคราะห์สเปคตรัมของสัญญาณมีด้วยกันหลายวิธี เช่น Short-time Spectrum, Cepstrum, Linear Predictive Coding และวิธีวิเคราะห์สเปคตรัมของสัญญาณด้วยวิธี Short-time Spectrum โดยใช้ฟูเรียร์

ทรานส์ฟอร์ม เป็นวิธีการที่ทำได้ง่ายและรวดเร็ว [16] ซึ่งอาศัยการแปลงสัญญาณเสียงในเชิงเวลาช่วงสั้น ๆ ไปในเชิงความถี่ แล้วทำการวิเคราะห์พารามิเตอร์ต่าง ๆ ของสเปคตรัมเสียง เพื่อใช้ตัดสินใจในขั้นตอนการรู้จำเสียงพูด ในขั้นตอนการรู้จำเสียงพูดนี้จะใช้การวัดระยะห่าง (Distance) ของสเปคตรัมเสียงของแบบอ้างอิงกับแบบทดสอบ และส่วนใหญ่จะนำเอาเทคนิคไดนามิคไทม์วาร์ปปิงมาประยุกต์ใช้ด้วย เพื่อลดความแตกต่างของสัญญาณเสียงในเชิงเวลาอย่างไม่เชิงเส้นที่เกิดจากการพูด และสำหรับในส่วนของแบบอ้างอิงนั้นจะเก็บ template เสียงของคำศัพท์ต่าง ๆ เอาไว้ ซึ่งในคำศัพท์แต่ละคำนั้นอาจมี template หลาย template เพื่อสามารถรู้จำคำ ๆ นั้นได้หลายลักษณะ หรือรู้จำเสียงพูดได้หลายบุคคล

จากการวิเคราะห์เสียงพูดของคนเรานั้น ในการเปล่งเสียงแต่ละพยางค์ สามารถแยกออกเป็นส่วนย่อยๆ ได้เป็น 3 ส่วน [3] คือ

1. เสียงที่เกิดในส่วนหน้าของพยางค์
2. เสียงที่เกิดในส่วนกลางของพยางค์
3. เสียงที่เกิดในส่วนท้ายของพยางค์

ในเสียงที่เกิดในส่วนหน้าของพยางค์นั้นจะเป็นเสียงพยัญชนะต้น เสียงที่เกิดในส่วนกลางของพยางค์จะเป็นเสียงสระ สำหรับเสียงที่เกิดในส่วนท้ายของพยางค์จะเป็นเสียงตัวสะกด และในพยางค์ในภาษาไทยจะสามารถผันเสียงวรรณยุกต์ได้สูงถึง 5 เสียง แต่ไม่ใช่ทุกพยางค์ที่ผันได้ครบทั้ง 5 เสียง ซึ่งสามารถแบ่งพยางค์ออกเป็นโครงสร้าง (Luksaneeyanawin Sudaporn, 1983 อ้างใน ทวี ปทุมทาน, 2530) ได้ดังนี้

$$S:= C(C)V(:)(C)^{T}$$

S = พยางค์
C = พยัญชนะ
V = สระ
: = สระเสียงยาว
T = วรรณยุกต์
โดย h = เสียงวรรณยุกต์ตรี
m = เสียงวรรณยุกต์สามัญ
l = เสียงวรรณยุกต์เอก
f = เสียงวรรณยุกต์โท
r = เสียงวรรณยุกต์จัตวา

การวิเคราะห์เสียงโดยการแยกแยะหน่วยเสียงย่อยนี้ จะมีข้อดีสำหรับนำไปใช้กับระบบรู้จำที่มีจำนวนคำศัพท์มาก ๆ โดยใช้แบบอ้างอิงจำนวนน้อย [2], [19] ซึ่งจากงานวิจัยของ สุดาพร ลักษณียนาวิน โดยใช้เสียงพยัญชนะเดี่ยว 21 หน่วย, พยัญชนะผสม 12 หน่วย, สระเดี่ยวเสียงสั้นและยาว 18 หน่วย, สระผสมเสียงสั้น และยาว 6 หน่วย, พยัญชนะท้าย 8 หน่วย และวรรณยุกต์ 5 หน่วย มาสังเคราะห์เป็นเสียงพยางค์ที่คนไทยสามารถเปล่งเสียงได้มีทั้งสิ้น 30,096 พยางค์ และเมื่อตัดพยางค์ที่ไม่สามารถเนื่องจากข้อจำกัดในการออกเสียงของคน จะเหลือจำนวนพยางค์ทั้ง

สิ้น 26,928 พยางค์ [6], [7] ในทั้งหมดนี้ มีปรากฏเป็นคำ และส่วนของคำ 5,912 พยางค์ ซึ่งสร้างจากหน่วยเสียงจำนวนดังกล่าว

เนื่องจากลักษณะการพูด, ลักษณะของเสียงของแต่ละบุคคล จะมีลักษณะแตกต่างกันออกไป ซึ่งเกิดได้จากหลาย ๆ ปัจจัยด้วยกัน เช่น เพศ, อวัยวะที่ใช้ในการออกเสียงซึ่งขึ้นกับสรีระของแต่ละบุคคล, สภาพร่างกายในขณะนั้น, ถิ่นที่อยู่อาศัย, อายุของผู้พูด (Labov, 1972d อ้างถึงใน อรุณี อรุณเรือง, 2533) ซึ่งลักษณะการพูดของแต่ละคนไม่เหมือนกัน จึงทำให้เสียงที่พูดออกมาแตกต่างกัน แม้ว่าจะเป็นคำ ๆ เดียวกันก็ตาม ดังนั้นการรู้จำโดยไม่ขึ้นกับบุคคลส่วนมากจะใช้จำนวนแบบอ้างอิงหลายแบบต่อคำศัพท์หนึ่งคำ เพื่อให้สามารถรู้จำคำพูดจากบุคคลต่าง ๆ ได้มากขึ้น ดังนั้นการใช้อ้างอิงมากกว่าหนึ่งแบบต่อหนึ่งคำจึงถูกนำมาใช้กับวิธีการรู้จำแบบไม่ขึ้นกับบุคคล เนื่องจากเสียงที่นำมาทดสอบมีโอกาสคล้ายกับแบบอ้างอิงแบบใดแบบหนึ่งในจำนวนทั้งหมด

ในวิทยานิพนธ์ฉบับนี้ ทำการศึกษาการรู้จำเสียงพูดสระภาษาไทยโดด ๆ ไม่ขึ้นกับผู้พูด (Speaker Independent) เนื่องจากเสียงสระเป็นองค์ประกอบหลักของพยางค์ หากไม่มีเสียงสระประกอบอยู่ด้วย จะทำให้ไม่สามารถเปล่งเสียงออกมาได้ และการเปลี่ยนแปลงความถี่ของเสียงสระในช่วงเวลาต่าง ๆ ของพยางค์ทำให้เกิดวรรณยุกต์ขึ้น แต่เสียงพยัญชนะจะมีอิทธิพลต่อเสียงสระและวรรณยุกต์ ขึ้นอยู่กับพยัญชนะนั้น ๆ การศึกษาเสียงสระและวรรณยุกต์ทำได้ยากขึ้น

วัตถุประสงค์ของวิทยานิพนธ์

1. เพื่อศึกษาและพัฒนาการรู้จำเสียงพูดสระภาษาไทยโดดๆ ไม่ขึ้นกับผู้พูดโดยใช้ไดนามิคไทม์วาร์ปปิง
2. เพื่อศึกษาและพัฒนาการรู้จำเสียงพูดที่ใช้แบบอ้างอิงหนึ่งแบบต่อเสียงพูดหนึ่งเสียง
3. เพื่อศึกษาและพัฒนาการจำแนกเสียงวรรณยุกต์ภาษาไทย

ขอบเขตของวิทยานิพนธ์

1. ทำการทดลองการรู้จำเสียงพูดโดยวิธีการวัดค่าระยะห่างของเสียงพูดบุคคลในกลุ่มที่นำมาเป็นแบบอ้างอิง กับเสียงของบุคคลนอกกลุ่มให้มีความถูกต้องไม่ต่ำกว่าร้อยละ 80.0 โดยใช้แบบอ้างอิง 1 แบบต่อ 1 เสียง
2. สามารถแยกแยะเสียงวรรณยุกต์ภาษาไทยได้ไม่ต่ำกว่าร้อยละ 80.0

ขั้นตอนและวิธีดำเนินการวิจัย

1. ศึกษาวิชาภาษาศาสตร์และวิชาสัทศาสตร์เกี่ยวกับหน่วยเสียงภาษาไทย
2. ศึกษาและนำเทคนิคต่าง ๆ มาประยุกต์ใช้กับการรู้จำสระภาษาไทยโดด ๆ ไม่ขึ้นกับบุคคล รวมทั้ง การจำแนกเสียงวรรณยุกต์ภาษาไทยด้วย
3. ออกแบบโปรแกรมที่ใช้ทดสอบการรู้จำของเสียงพูดเสียงสระ, โปรแกรมสร้างเสียงอ้างอิง, โปรแกรมแบ่งกลุ่มของเสียงสระ, โปรแกรมจำแนกเสียงวรรณยุกต์ภาษาไทย
4. เก็บตัวอย่างเสียงสระภาษาไทย และวรรณยุกต์ โดยบันทึกผ่านการ์ดเสียงบนเครื่องคอมพิวเตอร์แบบ พีซี ซึ่งจะถูกเก็บอยู่ในรูปแบบไฟล์ของ Sound Blaster และทำการทดสอบการรู้จำกับเสียงที่ได้บันทึกไว้

5. สรุปผลการวิจัยและข้อเสนอแนะ

ประโยชน์ที่จะได้รับจากการวิจัยนี้

1. เป็นแนวทางการสร้างระบบติดต่อของคอมพิวเตอร์กับผู้ใช้ที่รับคำสั่งด้วยเสียง และประยุกต์ใช้กับระบบอื่น เช่น คอมพิวเตอร์สั่งงานด้วยเสียง, ใช้ทำหน้าที่แทนพนักงานสลับสายโทรศัพท์ เป็นต้น
2. เป็นแนวทางการพัฒนาระบบรู้จำเสียงพูดแบบทั้งคำ โดยการพิจารณาหน่วยเสียงย่อย ๆ ของพยางค์ร่วมกัน