

บทที่ 3

วิธีการวิจัย

3.1 ฐานข้อมูลเสียง

ภายในงานวิจัยนี้ได้ทำการทดลองกับข้อมูลเสียงคำศัพท์ทั่วไป เป็นเสียงจากผู้พูด 55 คน ช่วงอายุตั้งแต่ 20-25 ปี เป็นชาย 31 คน เป็นหญิง 24 คน พูดคนละ 20 ประโยค ซึ่งข้อความภายในประโยคเหล่านี้ได้มาจากโครงการวิจัยฐานข้อมูลเสียงภาษาไทยของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (Somlertlamvanich, V. et al., 2000) บันทึกเสียงในห้องบันทึกเสียงที่ไม่มีสัญญาณรบกวน ใช้ความถี่ในการสุ่มตัวอย่างสัญญาณ 16 กิโลเฮิร์ต และความละเอียดของข้อมูล 16 บิต

ข้อมูลเสียงได้แบ่งออกเป็น 2 ส่วนคือ

1. ข้อมูลฝึกฝน (training set) จากผู้พูด 45 คน เป็นชาย 25 คน เป็นหญิง 20 คน
2. ข้อมูลทดสอบ (testing set) จากผู้พูด 10 คน เป็นชาย 6 คน เป็นหญิง 4 คน

3.2 อุปกรณ์/เครื่องมือ/ซอฟต์แวร์

ประกอบด้วยซอฟต์แวร์และฮาร์ดแวร์ดังต่อไปนี้

- 1) ชุดโปรแกรมระบบรู้จำเสียงพูด HTK เวอร์ชัน 3.0
- 2) โปรแกรมสำหรับพัฒนาภาษาจาวา Netbeans เวอร์ชัน 3.3.1
- 3) โปรแกรมคอมไพเลอร์ภาษาจาวา (Java Compiler, j2sdk1.4.2)
- 4) เครื่องคอมพิวเตอร์ แบบตั้งโต๊ะ 1 เครื่อง (ใช้ในการพัฒนาและทดสอบระบบรู้จำ)
- 5) เครื่องคอมพิวเตอร์ แบบพกพา 1 เครื่อง (ใช้ในการทดสอบระบบรู้จำในสภาพแวดล้อมทั่วไป)

3.3 วิธีดำเนินการ

ในการพัฒนาระบบรู้จำเสียงพูดจะต้องมีการปรับค่าพารามิเตอร์ให้เหมาะสมกับภาษาไทยดังนี้

3.3.1 คัดเลือกค่าลักษณะเด่น

ค่าลักษณะเด่นเป็นการค้นหาลักษณะพิเศษของเสียงพูดแต่ละเสียงเพื่อใช้ในการจดจำ จึงต้องมีการคัดเลือกค่าลักษณะเด่นที่เหมาะสม โดยการสร้างแบบจำลองเสียงที่ใช้ค่าลักษณะเด่นที่ต้องการเปรียบเทียบ แล้วนำแบบจำลองเสียงเหล่านั้นมาหาค่าความแม่นยำ ด้วยการจำแนกหน่วยเสียง แล้วจึงเลือกค่าลักษณะเด่นที่ให้ความแม่นยำสูงสุด ค่าลักษณะเด่นที่นำมาเปรียบเทียบได้แก่ MFCC และ LPC ที่เพิ่มค่าอนุพันธ์อันดับที่ 1 (+D) และ 2 (+DA) เนื่องจากค่าลักษณะเด่นทั้งสองมีการใช้กันอย่างแพร่หลายในการพัฒนาระบบรู้จำเสียงพูด โดยเฉพาะค่าลักษณะเด่นแบบ MFCC+DA ที่งานวิจัยส่วนใหญ่ (Gowdy, J.N. and Tufekci, Z., 2000) เลือกใช้

3.3.2 คัดเลือกโครงสร้าง HMM (Choosing HMM Structure)

ขั้นตอนนี้เป็น การคัดเลือกโครงสร้าง HMM ที่เหมาะสม โดยการสร้างแบบจำลองเสียงที่มีโครงสร้าง HMM แบบต่างๆ แล้วนำมาหาค่าความแม่นยำโดยวิธีการจำแนกหน่วยเสียง จากนั้นจึงเลือกโครงสร้าง HMM ที่ทำให้แบบจำลองเสียงให้มีค่าความแม่นยำสูงสุด

โครงสร้าง HMM ที่นำมาคัดเลือกนี้มีจำนวนสเตตตั้งแต่ 2-5 และจำนวนมิชเจอร์ตั้งแต่ 3-12 โดยปกติแล้วระบบรู้จำส่วนใหญ่เลือกใช้โครงสร้าง HMM แบบ 5 สเตต 3 มิชเจอร์ สำหรับหน่วยเสียงภาษาไทยความยาวช่วงเสียงสั้นกว่าคำศัพท์ปกติมากจำนวนสเตตจึงไม่ควรเกิน 5 สำหรับค่ามิชเจอร์ที่บ่งบอกถึงความละเอียดของเสียงว่ามีอย่างน้อยเพียงใด ซึ่งค่ามิชเจอร์ที่มากจะเหมาะสมกับการจำแนกหน่วยเสียงจำนวนมาก สำหรับชุดหน่วยเสียงภาษาไทยที่มีตั้งแต่ 170 หน่วยเสียง จึงได้เลือกจำนวนมิชเจอร์อยู่ในช่วงที่กว้างคือ 3-12

3.3.3 คัดเลือกชุดหน่วยเสียง

ชุดหน่วยเสียงเป็นเสียงย่อยสุดที่ระบบรู้จำจดจำได้ ชุดหน่วยเสียงที่ดีจะต้องมีหน่วยเสียงครอบคลุมหน่วยเสียงในภาษาไทยทั้งหมด จึงต้องมีการกำหนดชุดหน่วยเสียงที่เหมาะสม ได้นำชุดหน่วยเสียง $C_i+V_i+C_r$ และชุดหน่วยเสียง $C_i+V_i+C_n$ มาสร้างแบบจำลองเสียง 2 แบบ แล้วนำมาหาค่าความแม่นยำโดยวิธีการจำแนกหน่วยเสียง เพื่อเลือกชุดหน่วยเสียงที่มีความแม่นยำสูงสุดมาใช้ในระบบ

3.3.4 สร้างและหาค่าความแม่นยำของแบบจำลองเสียงต้นแบบ

ขั้นตอนนี้เป็น การสร้างแบบจำลองเสียงด้วยค่าพารามิเตอร์ที่คัดเลือกจากขั้นตอนที่ผ่านมา ซึ่งจะได้แบบจำลองเสียงต้นแบบ นำมาหาค่าความแม่นยำด้วยวิธีการจำแนกหน่วยเสียง ชุดหน่วยเสียงที่ใช้ในการจำแนกนั้นแบ่งออกเป็น 4 กลุ่มคือหน่วยเสียงทั้งหมด หน่วยเสียงกลุ่มพยัญชนะ หน่วยเสียงกลุ่มสระ หน่วยเสียงกลุ่มตัวสะกด ค่าความแม่นยำที่ได้ทั้ง 4 ค่าจะชี้ให้เห็นถึงจุดอ่อน และจุดแข็งของแบบจำลองเสียงในการจำแนกเสียงแต่ละกลุ่ม เพื่อนำไปปรับปรุงแก้ไขได้อย่างถูกต้อง รวมถึงเป็นคำอ้างอิงในการพัฒนาในขั้นต่อไป

3.3.5 วิเคราะห์ผลการรู้จำหน่วยเสียง พยางค์ และคำ

การทดลองนี้จะนำแบบจำลองเสียงต้นแบบมาหาค่าความแม่นยำด้วยวิธีการรู้จำเสียงพูด โดยกำหนดรูปแบบการรู้จำ 3 แบบได้แก่ ได้แก่ การรู้จำหน่วยเสียง การรู้จำพยางค์ การรู้จำคำศัพท์ โดยใช้เทคนิคการปรับปรุงพจนานุกรมตามรูปแบบการรู้จำ ซึ่งค่าความแม่นยำที่ได้จะเป็นตัวบ่งชี้ว่าแบบจำลองเสียงนี้เหมาะกับการนำไปใช้รู้จำเสียงพูดแต่ละแบบได้ดีเพียงใด และยังเป็นคำอ้างอิงในการพัฒนาระบบรู้จำเสียงพูดในขั้นต่อไปได้

3.3.6 พัฒนาไวยากรณ์ภาษาเบื้องต้น

ขั้นตอนนี้เป็น การกำหนดรูปแบบการรู้จำหรือไวยากรณ์ภาษาที่เหมาะสมกับการรู้จำทั่วไป โดยออกแบบไวยากรณ์ภาษาเป็นรูปแบบต่างๆ แล้วนำแบบจำลองเสียงต้นแบบ และไวยากรณ์เหล่านั้นมาใช้ในการรู้จำเสียงพูด ค่าความแม่นยำที่ได้จากไวยากรณ์ที่ต่างกันจะแสดงถึงความเหมาะสมระหว่างไวยากรณ์กับแบบจำลองเสียงได้

ไวยากรณ์ที่ใช้ในการคัดเลือกนี้มีด้วยกัน 3 แบบคือ ไวยากรณ์ประโยคบอกเล่าพื้นฐาน ประโยคบอกเล่าซับซ้อน และไวยากรณ์คำศัพท์ทั่วไป ใช้ชุดคำศัพท์ 30 คำ ในขั้นตอนนี้ทำให้ทราบว่าไวยากรณ์ใดที่เหมาะสมกับแบบจำลองเสียงต้นแบบ เนื่องจากแบบจำลองไม่มีความแม่นยำเท่าที่ควร การออกแบบไวยากรณ์จึงได้จำกัดขอบเขตการรู้จำให้แคบที่สุดเพื่อให้ผลการรู้จำมีความแม่นยำสูงสุด และสามารถนำไปใช้งานต่อไปได้

3.3.7 พัฒนาโปรแกรมคอมพิวเตอร์สำหรับรู้จำเสียงพูด

เมื่อสรุปแนวทางในการรู้จำเสียงพูดเรียบร้อยแล้วต่อมาจะเป็นการพัฒนาโปรแกรมรู้จำเสียงพูดสำหรับใช้งานจริง ซึ่งเป็นการพัฒนาโปรแกรมควบคุมการทำงานของชุดโปรแกรม HTK อีกต่อหนึ่ง โดยมีปุ่มคอยรับคำสั่งจากผู้ใช้และมีหน้าจอแสดงผลการรู้จำ ขั้นตอนนี้นอกจากการพัฒนาโปรแกรมรู้จำเสียงพูดแล้วยังรวมถึงการทดสอบการทำงานของโปรแกรม เพื่อเปรียบเทียบผลการรู้จำกับชุดโปรแกรม HTK โดยใช้แบบจำลองเสียงต้นแบบ และไวยากรณ์ที่เหมาะสม

3.3.8 ทดสอบระบบรู้จำในสภาพแวดล้อมภายนอกห้องบันทึกเสียง

ขั้นตอนสุดท้ายของงานวิจัยชิ้นนี้คือการแสดงถึงประสิทธิภาพของระบบรู้จำเสียงพูดต้นแบบที่พัฒนาขึ้น โดยอาศัยแบบจำลองเสียงต้นแบบ ไวยากรณ์ที่เหมาะสม และโปรแกรมรู้จำเสียงพูด เพื่อรู้จำเสียงพูดในสภาพแวดล้อมจริงโดยมีเสียงรบกวนจากภายนอก และใช้เครื่องคอมพิวเตอร์ต่างจากที่ทำการทดลอง ค่าความแม่นยำที่ได้จะชี้ให้เห็นถึง ปัญหาและอุปสรรคในกระบวนการรู้จำเสียงพูดทั้งระบบ ได้อย่างชัดเจน