

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

งานวิจัยทางด้านการรู้จำเสียงพูดในปัจจุบันนี้มีความก้าวหน้าไปมากและเริ่มนำมาประยุกต์ใช้งานมากขึ้น โดยเฉพาะในระบบสื่อสารโทรคมนาคม เช่น โทรศัพท์มือถือที่สามารถส่งโทรออกโดยใช้เสียง จุดประสงค์หลักของการรู้จำเสียงพูดคือการเพิ่มความสามารถให้อุปกรณ์ต่างๆ สามารถรับรู้และโต้ตอบกับมนุษย์ได้มากขึ้น ซึ่งการใช้เสียงพูดในการสื่อสารควบคุมสั่งการนี้ถือได้ว่าเป็นวิธีที่เป็นธรรมชาติมากที่สุดของมนุษย์

อย่างไรก็ตาม วิธีการต่างๆ ของการรู้จำเสียงพูดที่ได้มีการวิจัยกันมา [8,9,11,12,13] อยู่บนพื้นฐานของภาษาต่างประเทศ ถ้านำมาปรับใช้กับเสียงพูดภาษาไทยโดยตรง ย่อมจะได้ประสิทธิภาพที่ด้อยลงไม่มากนัก [6] เนื่องจากลักษณะการออกเสียงและโครงสร้างของภาษาไทยมีความแตกต่างไปจากภาษาของประเทศตะวันตกมาก อย่างเช่น ภาษาไทยจะมีเสียงวรรณยุกต์ “สามัญ เอก โท ตรี จัตวา” ในการบ่งบอกความหมายของคำซึ่งไม่มีในภาษาของประเทศตะวันตก งานวิจัยนี้จึงได้เน้นการศึกษาค้นคว้าวิธีการรู้จำหน่วยเสียงที่น่าจะเหมาะสมกับเสียงพูดภาษาไทย เพื่อที่จะให้ได้ความถูกต้องในการรู้จำเพียงพอที่จะนำไปประยุกต์ใช้งานได้

ในปัจจุบันการรู้จำเสียงพูดนั้นได้มีการศึกษาวิจัยกันอย่างแพร่หลาย สามารถจำแนกออกไปได้หลายด้านตามวัตถุประสงค์ของงานวิจัยนั้นๆ ซึ่งสามารถแบ่งออกเป็นแนวทางหลักๆ ได้เป็น การศึกษาวิจัยที่เน้นกลุ่มผู้พูด ลักษณะการพูด ลักษณะของเสียงพูด และกรรมวิธีหลักในการรู้จำเสียงพูด [6]

1.1.1 การศึกษาวิจัยที่เน้นตามกลุ่มผู้พูด แบ่งย่อยได้เป็น

- แบบขึ้นกับผู้พูด (Speaker-Dependent)
- แบบไม่ขึ้นกับผู้พูด (Speaker-Independent)
- แบบตรวจรู้ผู้พูด (Speaker-Identification)

งานวิจัยส่วนใหญ่จะเน้นไปที่แบบไม่ขึ้นกับผู้พูด เพราะสามารถนำไปประยุกต์ใช้ได้อย่างกว้างขวางแต่ผลที่ได้ก็ยังไม่มีความถูกต้องไม่เพียงพอที่จะนำมาประยุกต์ใช้งานจริง ส่วนแบบขึ้นกับผู้พูดนั้นสามารถพัฒนาได้เร็วกว่าเพราะสามารถควบคุมความแปรปรวนของกลุ่มตัวอย่างได้ดีกว่า ส่งผลให้ได้ความถูกต้องที่สูงกว่าและเพียงพอที่จะนำมาประยุกต์ใช้งานจริง ซึ่งปัจจุบันมี

ผลิตภัณฑ์ทางการรู้จำเสียงพูดภาษาต่างประเทศ เช่น ภาษาอังกฤษ สเปน อิตาลี เป็นต้น แต่ก็ เป็นแบบขึ้นกับผู้พูดหรืออยู่ในแนวทางดังกล่าว ผู้ที่จะใช้ผลิตภัณฑ์นี้จะต้องทำการฝึกฝนเป็นระยะ เวลาหนึ่ง ก่อนที่จะใช้ผลิตภัณฑ์ได้ ในขณะที่แทบจะไม่ปรากฏงานวิจัยแบบตรวจรู้ผู้พูด เนื่องจาก ยังไม่สามารถสร้างแบบจำลองกลไกการพูดของมนุษย์ที่สมบูรณ์ได้

1.1.2 การศึกษาวิจัยที่เน้นตามลักษณะการพูด แบ่งย่อยได้เป็น

- แบบคำเดี่ยว (Isolated Word) เป็นการศึกษาวิจัยที่เริ่มก่อนแบบอื่น เพราะรูปแบบ ของสัญญาณง่ายต่อการวิเคราะห์และวัดหาลักษณะสำคัญ [14]
- แบบคำเชื่อม (Connected Word) เป็นแบบที่มีการศึกษาวิจัยต่อจากแบบแรก คำ เชื่อมคือเสียงพูดคำเดี่ยวต่อกันโดยมีการเว้นจังหวะระหว่างพยางค์เล็กน้อย ซึ่งการพูดแบบนี้จะช้า และชัดเจนกว่าเสียงพูดปกติเล็กน้อย [15]
- แบบเสียงพูดต่อเนื่อง (Continuous Speech) เป็นแบบที่มนุษย์ใช้กันอยู่โดยปกติ จึง มีการศึกษาวิจัยกันมากในปัจจุบัน

1.1.3 การศึกษาวิจัยที่เน้นตามลักษณะเสียงพูด

- แบบจำกัดเฉพาะเสียงตัวเลข
- แบบเสียงคำสั่งเฉพาะงาน มีจำนวนคำไม่มาก
- แบบเสียงคำทั่วไปที่มีคำศัพท์จำนวนมากพอประมาณ
- แบบเสียงคำทั่วไปที่มีคำศัพท์จำนวนมาก

1.1.4 การศึกษาวิจัยที่เน้นตามกรรมวิธีหลัก แบ่งได้เป็น 4 กรรมวิธี คือ

- การเข้าคู่ต้นแบบ (Template Matching)
วิธีนี้เป็นที่นิยมกันมากคือ Dynamic Time Warping (DTW) ซึ่งมีความง่ายต่อ การพัฒนาและใช้เวลาในการฝึกฝนน้อย แต่ประสิทธิภาพจะต่ำลงเมื่อมีจำนวนคำศัพท์เพิ่มมากขึ้น [3]
- ระบบตามกฎเกณฑ์ (Rule-Based System)
จะอาศัยเงื่อนไขในการตัดสินใจ ซึ่งเมื่อระบบมีขนาดใหญ่และซับซ้อนมากขึ้นจะทำให้ การตัดสินใจผิดพลาดได้ง่ายตั้งแต่ขั้นตอนแรก [14]
- แบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Model, HMM)
วิธีนี้เป็นที่นิยมกันมากที่สุดของงานวิจัยในปัจจุบันเพราะมีข้อดีหลายประการ เช่น

ความยืดหยุ่นในการปรับให้ใช้กับการรู้จำเสียงพูดที่มีลักษณะการพูดต่าง ๆ กัน สมรรถนะในการรู้จำสูง ความสามารถรองรับคำศัพท์จำนวนมาก แต่มีข้อเสียคือ ต้องใช้เวลาในการฝึกฝนมาก และเมื่อมีคำศัพท์เพิ่มขึ้นจะต้องเริ่มกระบวนการฝึกฝนใหม่ทุกครั้งไป [10]

- โครงข่ายประสาทเทียม (Neural Network)

โครงข่ายประสาทเทียม หรือนิวรอลเน็ตเวิร์ก เป็นวิธีการเลียนแบบโครงข่ายประสาทของมนุษย์ที่มีประสิทธิภาพในการรู้จำเหนือกว่าวิธีอื่น ข้อเสียของวิธีนี้จะเหมือนกับวิธีฮิดเดน มาร์คอฟตรงที่ต้องการเวลาในการฝึกฝนมาก และต้องเริ่มการฝึกฝนใหม่เมื่อมีการเพิ่มคำศัพท์ [8]

จะเห็นได้ว่าวิธีการต่างๆ ที่ได้กล่าวมาข้างต้นยังมีข้อเสีย เช่น แบบจำลองฮิดเดน มาร์คอฟ และโครงข่ายประสาท ที่จะต้องมีการฝึกฝนใหม่ทุกครั้งเมื่อมีการเพิ่มคำศัพท์ใหม่ๆ เข้ามา ในวิธีระบบตามกฎเกณฑ์ ก็จะมีข้อจำกัดมากขึ้นเมื่อมีคำศัพท์มาก ส่งผลให้เกิดความผิดพลาดได้ง่าย ส่วนวิธีการเข้าคู่ต้นแบบประสิทธิภาพจะต่ำลงเมื่อมีจำนวนคำศัพท์มากขึ้นเพราะจะต้องสร้างรูปแบบอ้างอิงของคำศัพท์ทุกคำ

สำหรับงานวิจัยนี้จะแยกหน่วยเสียงออกมาก่อน แล้วจึงให้ระบบจำหน่วยเสียงที่แยกออกมาทั้งหมดด้วยโครงข่ายประสาท เนื่องจากพยางค์หนึ่งพยางค์จะต้องประกอบไปด้วยหน่วยเสียงพยัญชนะต้น, สระ, วรรณยุกต์และตัวสะกด ดังนั้นเมื่อนำผลจากการรู้จำหน่วยเสียงแต่ละส่วนมาประกอบกันก็สามารถสร้างเป็นคำอ่านต่างๆ ได้ ด้วยเหตุนี้การเพิ่มคำศัพท์จึงไม่มีผลกระทบต่อระบบ กล่าวคือไม่ต้องมีการฝึกฝนใหม่

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อศึกษาการรู้จำเสียงพูดที่มีอยู่ในปัจจุบัน
- 1.2.2 เพื่อหาวิธีการรู้จำเสียงพูดภาษาไทยที่เหมาะสม
- 1.2.3 เพื่อเพิ่มความสามารถให้คอมพิวเตอร์ในการรับรู้คำพูดของมนุษย์ได้

1.3 ขอบเขตของการทำวิจัย

สร้างวิธีรู้จำเสียงพูดภาษาไทยพยางค์เดี่ยว แบบขึ้นกับผู้พูด (ไม่รวมคำควบกล้ำ) ด้วยวิธีการสกัดลักษณะเด่นของหน่วยเสียงสระ, วรรณยุกต์, พยัญชนะต้นและตัวสะกด แล้วใช้โครงข่ายประสาทในการการแยกแยะหน่วยเสียง โดยสามารถแสดงคำอ่านภาษาไทยตรงตามเสียงพูดได้

1.4 ประโยชน์ที่จะได้รับจากการศึกษาเชิงทฤษฎีและเชิงประยุกต์

- 1.4.1 ทราบถึงวิธีการรู้จำเสียงพูดภาษาไทยที่มีอยู่ในปัจจุบัน
- 1.4.2 ทราบถึงลักษณะทางภาษาศาสตร์ของเสียงพูดภาษาไทย
- 1.4.3 เป็นการเพิ่มความสามารถให้เครื่องคอมพิวเตอร์ในการรู้จำเสียงพูดภาษาไทย
- 1.4.4 เป็นแนวทางในการนำไปพัฒนาต่อจากคำอ่านมาเป็นคำเขียนที่ถูกต้องได้

1.5 ขั้นตอนการทำวิจัย

- 1.5.1 ศึกษาทฤษฎีและผลงานวิจัยที่เกี่ยวข้องกับโครงการวิจัยนี้
- 1.5.2 ศึกษาและค้นคว้าขั้นตอนที่สำคัญ
 - การประมวลสัญญาณเบื้องต้น (Signal Preprocessing)
 - การวัดค่าลักษณะสำคัญ (Feature Measurement)
 - การวัดหาความคล้ายคลึงกันของรูปแบบ (Pattern Similarity Determination)
 - ขั้นตอนและวิธีการตัดสินใจ (Decision Algorithm)
- 1.5.3 กำหนดกลุ่มคำที่อ่านออกเสียงแล้วจะประกอบไปด้วยหน่วยเสียงภาษาไทย ทั้งหมด เพื่อใช้ในการสร้างหน่วยเสียงอ้างอิง และกลุ่มคำที่ใช้ในการทดสอบ
- 1.5.4 วิเคราะห์ลักษณะเด่นของหน่วยเสียง และทำการจำแนกหน่วยเสียงออกเป็นกลุ่ม หน่วยเสียงที่มีระยะเวลาการออกเสียงสั้นและยาว เพื่อความสะดวกและรวดเร็วในการค้นหา
- 1.5.5 วิเคราะห์และเขียนโปรแกรมในแต่ละส่วน
- 1.5.6 ทำการทดสอบความถูกต้องในการรู้จำเสียงพูด
 - แบ่งกลุ่มผู้ทดสอบออกเป็น 2 กลุ่ม คือกลุ่มผู้ชาย 5 คน และ กลุ่มผู้หญิง 5 คน อายุ ระหว่าง 20-25 ปี
 - ให้ผู้ทดสอบทั้งหมดอ่านกลุ่มคำที่กำหนดในข้อ 1.5.3 เพื่อสร้างหน่วยเสียงอ้างอิง ทั้งหมด และเก็บเป็นฐานข้อมูลของแต่ละบุคคล
 - ให้ผู้ทดสอบทั้งหมดอ่านคำที่อยู่ในกลุ่มคำที่ใช้สำหรับการทดสอบ เพื่อทดสอบโปรแกรม
 - วัดอัตราความถูกต้องจากอักขระแทนหน่วยเสียงที่แสดงออกมาตรงกับเสียงพูด จาก หน่วยเสียงทั้งหมดที่มีอยู่ในเสียงพูด และวัดอัตราความถูกต้องในการรู้จำพยัญชนะ ต้น, สระ, วรรณยุกต์และตัวสะกด โดยกลุ่มคำที่นำมาทดสอบจะประกอบด้วยคำ 135 คำ

1.5.7 ปรับปรุงแก้ไขโปรแกรมเพื่อเพิ่มความถูกต้องในการรู้จำ

1.5.8 วิจารณ์ สรุปผลและเสนอรายงานวิทยานิพนธ์

1.6 ผลงานวิจัยอื่นๆ ที่เกี่ยวข้อง

ระพีพัฒน์ เพ็ญศิริ [3] ได้เสนอเรื่อง “การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยการ ใช้ไดนามิกส์โทมวาร์ปิง” เป็นงานวิจัยที่ใช้การแปลงฮาร์ตเลย์ (Hartley Transform) ในการ วิเคราะห์หาพารามิเตอร์เพื่อใช้ในการรู้จำเสียงพูดตัวเลขไทย โดยที่เลือกใช้คั่นแบบ (template) แบบ คำเพราะมีคำจำนวนไม่มาก

วุฒิพงษ์ พิชิตวงศ์ และคณะ [4] ได้เสนอเรื่อง “วิธีการแบ่งเสียงตัวสะกดในภาษาไทย พยางค์เดียวโดยใช้คุณสมบัติเสียงสระและการแปลงเวฟเลตแพกเกตส์” เป็นงานวิจัยที่นำเอาการ แปลงเวฟเลตแพกเกตส์ มาช่วยในการวิเคราะห์หาตำแหน่งสิ้นสุดของหน่วยเสียงสระซึ่งจุดเดียวกัน นี้ก็จะเป็นจุดเริ่มต้นของตัวสะกด ข้อดีของการแปลงเวฟเลตแพกเกตส์ก็คือการปรับเปลี่ยนแกน เวลาให้เหมาะสมกับสัญญาณที่มีความถี่ไม่แน่นอน ทำให้ได้ข้อมูลเวลาที่ละเอียดขึ้นที่ความถี่สูงๆ เนื่องจากหน่วยเสียงสระจะมีความถี่สูงกว่าหน่วยเสียงอื่น

สมชาย จิตะพันธ์กุล [6] ได้เสนอเรื่อง “การรู้จำเสียงพูดคำไทยใดๆโดยไม่ขึ้นกับผู้พูด” เป็นการรวบรวมงานวิจัยที่เกี่ยวข้องกับการรู้จำด้วยวิธีที่เป็นที่นิยม ได้แก่ วิธี Dynamic Time Warping (DTW), Hidden Markov Model (HMM) และ Neural Network (NN)

Jun *et al.* [9] ได้เสนอเรื่อง “An Approach to Smooth Fundamental Frequencies in Tone Recognition” เป็นการเสนอวิธีการหาเส้นทางของระดับเสียง (pitch contour) ซึ่งใช้ในการวิเคราะห์ เสียงวรรณยุกต์ที่ง่ายและมีประสิทธิภาพเทียบเท่ากับวิธีประมาณส่วนโค้ง (curve fitting), การทำ ให้เรียบเชิงเส้น (linear smoothing) และ การทำให้เรียบเชิงมัธยฐาน (median smoothing) และยังสามารถทำงานได้ในกรณีที่มีความถี่มูลฐานหลายความถี่ สำหรับงานวิจัยนี้จะเลือกใช้วิธีการ ประมาณส่วนโค้ง เนื่องจากการวิเคราะห์เสียงวรรณยุกต์นั้นจะพิจารณาเพียงความถี่มูลฐานเดียว ณ เวลาต่างๆ

McInnes *et al.* [11] ได้เสนอเรื่อง “Template Adaptation in an Isolated Word-Recognition System” เป็นงานวิจัยที่ใช้คั่นแบบ (template) แบบคำ ในการรู้จำคำพูดที่มีจำนวนคำ ศัพท์ไม่มาก แต่มีข้อดีที่เป็นคั่นแบบที่สามารถเปลี่ยนแปลงไปตามเสียงของผู้ใช้ที่เปลี่ยนไปได้