บทที่ 1
บทนำ

แนวเหตุผล

เทคโนโลยีทางด้านการประมวลผลสัญญาณเชิงเลข ถูกนำมาช่วยในการติดต่อสื่อสารอย่างกว้างขวาง โดยมีจุดมุ่งหมายเพื่ออำนวยความสะดวกแก่มนุษย์ การติดต่อสื่อสารพื้นฐานของมนุษย์คือการพูด ดังนั้นจึงมีงานวิจัยมากมายที่เกี่ยวข้องกับการประมวลผลทางด้านเสียงพูด และนำเทคโนโลยีเหล่านี้ใช้ประโยชน์ในแง่ต่างๆ กัน เช่น เครื่องแปลภาษา (Machine Translator) เครื่องบริการให้ข้อมูลเฉพาะเรื่องโดยใช้เสียง เช่น การให้ข้อมูลสภาพอากาศทั่วโลกผ่านทางโทรศัพท์ และอุปกรณ์ช่วยเหลือคนพิการบางประเภท เช่น เครื่องแปลงเสียงเป็นตัวหนังสือ (Speech-to-Text Converter)

ส่วนประกอบที่สำคัญ และมีความยุ่งยากในการจัดการมากที่สุดในอุปกรณ์ต่างๆ ที่กล่าวมา คือ ส่วนการรู้จำเสียงพูด (Speech Recognition) ซึ่งมีหน้าที่ในการจดจำเพื่อแยกแยะให้ได้ว่าเสียงพูดที่เข้ามาในระบบนั้นเป็นเสียงของคำใดบ้าง การรู้จำเสียงพูดแบ่งตามหน่วยที่ใช้ในการเทียบได้เป็น 2 ประเภท (Furui, 1989)

1. การรู้จำโดยใช้ลักษณะทางเสียงของคำ (Word-Based Recognition)
2. การรู้จำโดยใช้ลักษณะทางเสียงของหน่วยเสียงย่อย (Phoneme-Based Recognition)

วิธีแรกอาจจะใช้พยางค์ คำ หรือวลีเป็นหน่วยในการเทียบเคียง ซึ่งเป็นการหลีกเลี่ยงผลกระทบอันเกิดจากฐานของเสียงภายในคำ หรือกลุ่มคำนั้นๆ ในขณะที่วิธีที่สองมักจะใช้หน่วยเสียงที่ย่อยลงไป เช่น เสียงพยัญชนะ และเสียงสระ ซึ่งเป็นวิธีที่เหมาะสมในการรู้จำคำจำนวนมากๆ ได้

นอกจากนี้ยังสามารถแบ่งการรู้จำเสียงตามระดับของการรู้จำได้เป็น 2 ระดับดังนี้

1. การรู้จำคำเดี่ยว (Isolated Word Recognition)
2. การรู้จำคำพูดต่อเนื่อง (Sentence or Continuous Speech Recognition)

สำหรับรายละเอียดของการรู้จำเสียงแต่ละชนิด จะกล่าวถึงในบทต่อไป วิธีการที่ถูกนำมาใช้ในการรู้จำเสียงพูดในภาษาไทยมีอยู่มากมายหลายวิธี อาทิเช่น วิธีไดนามิกทามวาร์ปปิง (Dynamic Time Warping, DTW) (ระพีพัฒน์ เพ็ญศิริ 2538) การใช้นิวรอลเน็ตเวอร์ก (Artificial Neural Network, ANN) (Pornsukchantra, 1996) การใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model, HMM) (วิศรุต อาชุบุตร 2539) การใช้แบบจำลองฮิดเดนมาร์คอฟร่วมกับนิวรอลเน็ตเวอร์ก (Hybrid HMM-ANN) (Bourlard and Morgan, 1993) และการใช้เทคนิคแบบฟัซซี (Fuzzy) (Pal and Mitra, 1992) (Carlos and Wyllis, 1994) (Yingyong and Bobby, 1993) นอกจากนี้

ยังได้มีงานวิจัยอีกหลายงานที่พัฒนากรรมวิธีการพิจารณาลักษณะบ่งความต่างของเสียง (Distinctive Feature) เข้ามาช่วยเพิ่มประสิทธิภาพของการรู้จำ เช่น การตรวจสอบจำนวนพยางค์ (Syllable Detection) (Zhou and Imai, 1996) (วิศรุต อาชุบุตร 2539) และการตรวจสอบระดับเสียง (Tone Recognition) (ณัฐกร ทับทอง 2538) (Lee, Ching, Chan, Cheng and Mak, 1995)

สำหรับงานวิจัยนี้ เป็นระบบรู้จำเสียงคำไทยหลายพยางค์ โดยอาศัยคำเป็นหน่วยเสียงที่ใช้เทียบเคียง วิธีการในการรู้จำใช้นิวรอลเน็ตเวอร์ก และใช้เทคนิคทางฟัซซีมาช่วยเพิ่มประสิทธิภาพในการรู้จำ สาเหตุที่เลือกใช้วิธีการดังกล่าวคือ

1. ประสิทธิภาพของระบบรู้จำเสียงโดยใช้นิวรอลเน็ตเวอร์ก เทียบเท่า หรืออาจจะดีกว่าวิธีการที่ใช้อยู่ทั่วไป เช่น DTW หรือ HMM (Pornsukchantra, 1996) ในขณะที่ใช้เวลาในการทดสอบการรู้จำน้อยกว่าวิธีอื่นๆ และมีขั้นตอนในการเรียนรู้ที่ง่าย นอกจากนี้ ในปัจจุบันยังได้มีการวิจัยค้นคว้าเทคนิคใหม่ๆ อีกมากมาย เพื่อเพิ่มประสิทธิภาพของนิวรอลเน็ตเวอร์ก ทำให้เป็นวิธีที่สามารถขยายขีดความสามารถต่อไปได้อีกในอนาคต

2. ธรรมชาติของเสียงมักเกิดความกำกวมขึ้นเสมอ อันเนื่องมาจากความคล้ายกันของหน่วยเสียง เช่นเสียงคำว่า เจ็ด และเสียงคำว่าเป็ด เสียงคำว่า สอง และเสียงคำว่า สาม เป็นผลให้การรู้จำผิดพลาดอยู่เสมอ โดยเฉพาะคำที่มีเสียงสระสั้น แต่มนุษย์จะมีสามารถในการจดจำเสียงที่กำกวมเหล่านั้นได้โดยอาศัยการตัดสินใจที่ละเอียดอ่อนขึ้น ความสามารถดังกล่าวนี้ สามารถนำมารวมในระบบรู้จำเสียงได้โดยการใช้เทคนิคแบบฟัซซี ดังนั้นจึงมีงานวิจัยมากมายที่อาศัยเทคนิคนี้ในการรู้จำ เช่น การใช้ค่าทางฟัซซีแทนค่าความต่างในการเปรียบเทียบโดยตรง (Liu, Li and Shi, 1995) ใช้ฟัซซีร่วมกับวิธีการควอนไตซ์เวกเตอร์ (FVQ) (Lin and Anthony, 1994) ใช้ร่วมกับนิวรอลเน็ตเวอร์ก และแบบจำลองฮิดเดนมาร์คอฟ (Tsuboka and Nakahashi, 1994) แต่การใช้เทคนิคแบบฟัซซีจะส่งผลให้เวลาที่ต้องใช้ในการเรียนรู้ และการรู้จำมากขึ้นอย่างมาก ดังนั้นถ้าใช้ร่วมกับนิวรอลเน็ตเวอร์กซึ่งมีความสามารถในการคำนวณแบบขนานกันจำนวนมากทำให้ใช้เวลาน้อยลงได้มาก ดังที่เห็นในงานวิจัยมากมาย (Pal and Mitra, 1992) (Carlos and Wyllis, 1994) (Yingyong and Bobby, 1993) (Lin and Anthony, 1994) (Amano, Aritsuka, Hataoka and Ichikawa, 1989)

นอกจากนี้ในงานวิจัยนี้ยังได้อาศัยการพิจารณาลักษณะบ่งความต่างบางอย่าง มาช่วยเพิ่มประสิทธิภาพของการรู้จำ เช่น การตรวจสอบจำนวนพยางค์

วัตถุประสงค์

1. เพื่อศึกษาและพัฒนาการรู้จำเสียงพูดคำไทยหลายพยางค์โดยใช้นิวรอลเน็ตเวอร์ก
2. เพื่อศึกษาถึงแนวทางในการนำเทคนิคแบบฟัซซีมา
   ช่วยเพิ่มประสิทธิภาพของระบบรู้จำเสียง
3. พัฒนาการรู้จำเสียงพูดคำภาษาไทยหลายพยางค์
   โดยใช้เทคนิคผสมระหว่างฟัซซี และนิวรอลเน็ตเวอร์ก

เป้าหมายและขอบเขต

1. สร้างระบบรู้จำเสียงพูดคำภาษาไทยได้ถึง 3 พยางค์ ตรงตามชุดคำศัพท์ที่กำหนด โดยใช้เทคนิคแบบฟัซซี และนิวรอลเน็ตเวอร์ก

2. แสดงให้เห็นว่าระบบรู้จำเสียงที่อาศัยเทคนิคแบบฟัซซีร่วมกับนิวรอลเน็ตเวิร์ค ให้ประสิทธิภาพในการรู้จำดีกว่าระบบที่ไม่ใช้เทคนิคแบบฟัซซี

3. อัตราการรู้จำของระบบไม่ต่ำกว่า 90 %

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ทราบถึงแนวทางที่เหมาะสมในการพัฒนาเพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูด
2. เพิ่มเติมความสามารถทางด้านความแม่นยำของระบบรู้จำเสียงพูดคำภาษาไทย
3. สามารถวางแนวทางที่เหมาะสมในการพัฒนาระบบรู้จำเสียงพูดคำภาษาไทยต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย