# บทที่ 1

# บทนำ

## 1.1 แนวเหตุผล

การพัฒนาระบบสื่อสารระหว่างมนุษย์และเครื่องจักร (Human-Machine Communication) ต้องอาศัยระบบการรู้จำคำพูด (Speech Recognition) ซึ่งได้รับการพัฒนาสำหรับงานต่างๆ (Lee, Soong and Paliwan, 1996) ตั้งแต่การรู้จำคำสำคัญ (Keyword Recognition) ที่มีจำนวนคำศัพท์ขนาดเล็ก, ระบบที่ใช้เสียงสั่งงานและควบคุมคอมพิวเตอร์ส่วนบุคคล (Voice Interactive Command and Control System on PC) ที่มีจำนวนคำศัพท์ขนาดกลางและระบบแปลภาษา (Speech Translation) ที่มีจำนวนคำศัพท์มาก การรู้จำเสียงพูดต้องอาศัยความรู้หลายแขนง (Rabiner and Juang, 1993) เช่นการประมวลผลสัญญาณ (Signal Processing) กายภาพของเสียง (Acoustics Physics) การรู้จำแบบรูป (Pattern Recognition) การสื่อสารและทฤษฎีสารสนเทศ (Communication and Information Theory) ภาษาศาสตร์ (Linguistics) สรีรศาสตร์ (Physiology) วิทยาศาสตร์คอมพิวเตอร์ (Computer Science) และจิตวิทยา (Psychology)

แบบจำลองที่ใช้กับระบบรู้จำเสียงพูดมีหลายแบบเช่น การเข้าคู่ต้นแบบ (Template Matching) ระบบตามกฎเกณฑ์ (Rule-Based System) แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model, HMM) และ โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) .ในงานวิจัยนี้ได้เลือกวิธีการของโครงข่ายประสาทเทียมซึ่งมีข้อได้เปรียบในด้านของเวลาในการฝึกฝนและทดสอบการรู้จำที่ต่ำกว่าวิธีการอื่นๆ สามารถนำไปประยุกต์ใช้งานแบบเวลาจริงได้ นอกจากนี้โครงข่ายประสาทเทียมยังสามารถเรียนรู้คำศัพท์เพิ่มเติมได้อีกด้วย

โครงข่ายประสาทเทียมเป็นแบบจำลองการทำงานของสมองมนุษย์ที่ถูกนำไปใช้ในศาสตร์แขนงต่างๆ เช่น การแพทย์ วิทยาศาสตร์คอมพิวเตอร์ จิตวิทยาและวิศวกรรม เป็นต้น คุณสมบัติที่เหมาะสมของโครงข่ายประสาทเทียมในการใช้รู้จำรูปแบบได้แก่ (Tebelskils, 1995)

- สามารถฝึกฝนได้ (Trainability) โครงข่ายประสาทเทียมสามารถถูกสอนให้แยกแยะรูปแบบที่มีความแตกต่างกันได้
- รู้จำรูปแบบทั่วไป (Generalization) โครงข่ายประสาทเทียมจะต้องไม่จำรูปแบบที่ฝึกฝนแต่จะต้องเรียนรู้จากข้อมูลฝึก คุณสมบัติข้อนี้มีความสำคัญมากในการรู้จำเสียงพูดเนื่องจากเสียงพูดแต่ละครั้งไม่มีความเหมือนกันโดยสิ้นเชิง
- ความไม่เป็นเชิงเส้น (Nonlinearity) โครงข่ายประสาทเทียมสามารถคำนวณฟังก์ชันที่มีลักษณะไม่เชิงเส้น คือสามารถแปลงข้อมูลที่มีความซับซ้อนได้ คุณสมบัติข้อนี้เป็นประโยชน์ในการรู้จำเสียงพูดเนื่องจากเสียงพูดมีความไม่เป็นเชิงเส้นสูง

- ทนทานต่อสัญญาณรบกวน (Robustness) โครงข่ายประสาทเทียมมีความทนทานต่อข้อมูลที่มีสัญญาณรบกวนได้ดี คุณสมบัติข้อนี้เป็นประโยชน์ในการรู้จำเสียงพูดเนื่องจากเสียงพูดมีสัญญาณรบกวนปนอยู่ด้วย
- ความเท่าเทียมกัน (Uniformity) โครงข่ายประสาทเทียมสามารถคำนวณข้อมูลเข้าที่มีความแตกต่างกัน เช่น ค่า cepstrum และอนุพันธ์ของ cepstrum ในการรู้จำเสียงพูด
- การทำงานแบบขนาน (Parallelism) โครงข่ายประสาทเทียมเป็นกระบวนการทำงานแบบขนานโดยธรรมชาติ ดังนั้นจึงมีความเหมาะสมที่จะนำมาประยุกต์ใช้งานสร้างฮาร์ดแวร์แบบขนานซึ่งมีความเร็วสูงในการประมวลผลเสียงพูด

ตัวอย่างของโครงข่ายประสาทเทียมที่ใช้รู้จำหน่วยเสียงได้แก่ Time Delay Neural Networks (TDNN) (Waibel, 1989) Hybrid NN-HMM (Cerf, Ma, and Compernolle, 1994) Linked Predictive Neural Networks (LPNN) (Waibel and Tebelskis, 1990) Learning Vector Quantization (LVQ) (Deller Jr., Proakis, and Hansen, 1993) และ Hidden Control Neural Networks (HCNN) (Petek and Tebelskis, 1992) เป็นต้น จากงานวิจัยของต่างประเทศข้างต้นแสดงให้เห็นว่ามีความเป็นไปได้ที่จะนำโครงข่ายประสาทเทียมมาใช้รู้จำหน่วยเสียงภาษาไทย

สำหรับงานวิจัยเกี่ยวกับระบบรู้จำเสียงพูดภาษาไทยที่ผ่านมาดังแสดงในตารางที่ 1.1 อยู่บนพื้นฐานของการรู้จำเป็นคำโดด (Word Model) ดังแสดงในรูปที่ 1.1 ในงานวิจัยระบบรู้จำเสียงพูดภาษาไทยที่ใช้โครงข่ายประสาทเทียม (วุฒิพงษ์ พรสุขจันทรา, 2539) จำเป็นจะต้องแบ่งโครงข่ายออกเป็นหลายส่วนเพื่อที่จะรู้จำคำโดดที่มีจำนวนพยางค์ต่างกันเนื่องจากคำโดด พยางค์เดียว สองพยางค์และสามพยางค์มีความยาวต่างกันแต่โนดเข้า (Input Node) ของโครงข่ายประสาทเทียมมีจำนวนคงที่ ถ้านำวิธีการรู้จำทั้งคำดังกล่าวมาใช้สำหรับรู้จำคำต่อเนื่อง (Connected Word) จะต้องแยกโครงข่ายออกเป็นหลายส่วนเพื่อที่จะรู้จำคำศัพท์หลายพยางค์และจำเป็นต้องอาศัยกระบวนการตัดแบ่งพยางค์ (Syllable Segmentation) ที่มีความเที่ยงตรงสูงเพื่อที่จะส่งต่อให้โครงข่ายของโครงข่ายประสาทเทียมได้อย่างถูกต้อง จุดอ่อนของการรู้จำแบบคำโดดอีกประการหนึ่งคือเสียงที่มีความคล้ายกันจะมีความผิดพลาดในการรู้จำมากเช่นเสียงเจ็ด /cet1/ และ เป็ด /pet1/ (ชัย วุฒิวิวัฒน์ชัย, 2540)

การรู้จำหน่วยเสียง (Phonemes Recognition) จะอยู่ในขั้นตอนการจำแนกรูปแบบ (Pattern Classification) ซึ่งจะต้องอาศัยความรู้ทางด้านสัทศาสตร์และภาษาศาสตร์มาสร้างแบบจำลองของหน่วยเสียง ซึ่งแบบจำลองของหน่วยเสียงนี้จะถูกนำมาสร้างเป็นคำและประโยคโดยอาศัยหลักไวยากรณ์ (Grammar) ดังนั้นการรู้จำหน่วยเสียงสามารถเพิ่มอัตราการรู้จำได้สูงขึ้นเนื่องจากอาศัยความรู้ทางด้านภาษาศาสตร์มาช่วยวิเคราะห์และสามารถนำไปประยุกต์ใช้งานกับระบบรู้จำคำต่อเนื่องและระบบรู้จำเสียงพูดต่อเนื่อง (Continuous Speech) ได้อย่างมีประสิทธิภาพ ข้อดีของการใช้แบบจำลองของหน่วยเสียงคือมีจำนวนหน่วยเสียงที่จำกัดทำให้ประหยัดการจัดเก็บต้นแบบของหน่วยเสียงและสามารถเพิ่มเติมจำนวนคำศัพท์จากหน่วยเสียงที่มีอยู่ได้ นอกจากนี้การฝึกฝนเพื่อเพิ่มจำนวนคำศัพท์ของแบบจำลองของคำโดดจะต้องทำใหม่ทั้งหมด แต่สำหรับแบบจำลองของหน่วยเสียงไม่ต้องทำการฝึกฝนใหม่ทั้งหมดเพียงแต่เพิ่มรูปแบบการประกอบคำของ

หน่วยเสียง ความแตกต่างของการรู้จำเสียงพูดที่ใช้แบบจำลองของคำโดดและแบบจำลองของหน่วยเสียง แสดงในรูปที่ 1.2 ซึ่งในวิทยานิพนธ์นี้จะมีขอบเขตดังแสดงในกรอบของรูปที่ 1.2

**1.2 ปัญหาของการรู้จำเสียงพูดภาษาไทยที่ผ่านมา**

1. การใช้แบบจำลองของคำโดด การใช้แบบจำลองของคำโดดจะไม่สามารถแยกแยะเสียงพูดที่มีความคล้ายคลึงกันออกจากกันอย่างมีประสิทธิภาพ การรู้จำหน่วยเสียงสามารถที่จะแก้ไขปัญหาดังกล่าวได้ นอกจากนี้การใช้หน่วยเสียงซึ่งมีจำนวนจำกัดสามารถประหยัดการจัดเก็บฐานข้อมูลได้และสามารถที่จะเพิ่มจำนวนคำศัพท์เข้าไปได้โดยไม่ต้องฝึกฝนระบบใหม่ทั้งหมด
2. ข้อมูลที่แทนเสียงพูด ประเภทของข้อมูลที่แทนเสียงพูดมีนัยสำคัญที่จะแทนลักษณะสำคัญของเสียงพูดได้เพียงพอหรือไม่ นอกจากนี้จำนวนอันดับของข้อมูลที่แทนเสียงพูดมีความสำคัญต่อเวลาในการประมวลผลสัญญาณเสียง
3. ระบบรู้จำเสียงพูด จะต้องมีความเร็วในการประมวลสูงเพื่อที่จะสามารถทำงานแบบเวลาจริงได้และต้องมีอัตราการรู้จำที่สูงเพื่อประยุกต์ใช้งานได้

**1.3 วัตถุประสงค์**

1. เพื่อนำกรรมวิธีโครงข่ายประสาทเทียมมาใช้รู้จำหน่วยเสียงสระภาษาไทย
2. เพื่อพัฒนากรรมวิธีที่เหมาะสมในการรู้จำหน่วยเสียงภาษาไทย
3. เพื่อเป็นแนวทางในการพัฒนาระบบรู้จำเสียงพูดภาษาไทยให้มีประสิทธิภาพที่ดีขึ้น

**1.4 เป้าหมายและขอบเขตของงานวิจัย**

1. สามารถรู้จำหน่วยเสียงสระเดี่ยวแท้ภาษาไทยได้ครบ
2. อัตราการรู้จำมากกว่า 90 % ขึ้นไป
3. กลุ่มตัวอย่างเสียงจำนวน 50 คน (กลุ่มฝึกฝน 30 คน กลุ่มทดสอบ 20 คน)

**1.5 ขั้นตอนและวิธีการดำเนินการ**

1. ศึกษา ค้นคว้าและเก็บรวบรวมข้อมูลรายละเอียดดังต่อไปนี้
   - คุณลักษณะของเสียงพูด แบบจำลองการเปล่งเสียงพูดและหน่วยเสียงสระภาษาไทย
   - การวิเคราะห์และวัดค่าลักษณะสำคัญ
   - การประยุกต์ใช้โครงข่ายประสาทเทียมในการรู้จำหน่วยเสียง
2. การวิเคราะห์หน่วยเสียงสระภาษาไทย
   - เก็บข้อมูลเสียงพูด
   - พัฒนาโปรแกรมในการวิเคราะห์หน่วยเสียงภาษาไทย

- พิจารณาเลือกค่าลักษณะสำคัญที่เหมาะสม

3. สร้างระบบรู้จำหน่วยเสียงสระภาษาไทย
   - จำแนกกลุ่มหน่วยเสียงสระเพื่อแบ่งโครงข่ายประสาทเทียมออกเป็นโครงข่ายย่อยให้มีความสอดคล้องกับกลุ่มของหน่วยเสียง
   - พัฒนาโปรแกรมในการฝึกฝนโครงข่ายประสาทเทียม
4. ทำการทดสอบ แก้ไขและประเมินผลระบบ
5. สรุปรวบรวมผลการวิจัยทั้งหมด พร้อมทั้งจัดทำเอกสารเกี่ยวกับวิทยานิพนธ์

**1.6 ประโยชน์ที่คาดว่าจะได้รับ**

1. ทราบถึงวิธีการรู้จำคำพูดภาษาไทยและลักษณะทางภาษาศาสตร์ของเสียงสระคำพูดภาษาไทย
2. ทราบถึงกรรมวิธีที่เหมาะสมในการรู้จำเสียงพูดภาษาไทย
3. เป็นแนวทางในการพัฒนาระบบรู้จำเสียงพูดภาษาไทยแบบคำต่อเนื่องและคำพูดต่อเนื่อง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1.1 งานวิจัยเกี่ยวกับระบบรู้จำเสียงพูดภาษาไทย

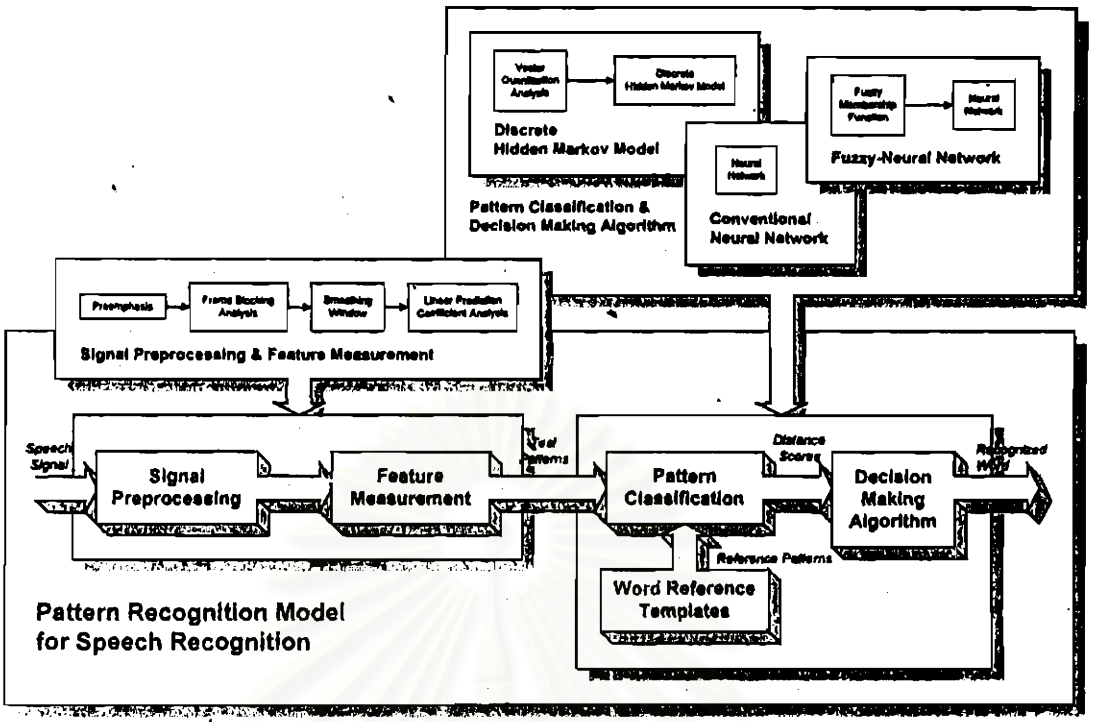
| กรรมวิธี | ค่าลักษณะสำคัญ | ประเภทคำศัพท์ | จำนวนคำศัพท์ | อัตราการรู้จำ |
|---|---|---|---|---|
| การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นกับผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปปิง (ระพีพัฒน์ เพ็ญศิริ, 2538) | Normalized DHT Parameters | ตัวเลขไทย | 10 | 79.25 % |
| ระบบรู้จำเสียงสระภาษาไทยคำเดี่ยวแบบต่างบุคคลโดยการวัดสเปกตรัมดิสแตนซและไดนามิกไทม์วาร์ปปิง (ธีระ ภัทราพรนันท์, 2538) | FFT and Log Energy | เสียงสระภาษาไทย<br>เสียงวรรณยุกต์ภาษาไทย | 24<br>15 | 86.17 %<br>81.00 % |
| การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดนมาร์คอฟโมเดลและเวกเตอร์ควอนไตซ์เซชั่น (เสาวลักษณ์ อารีพงศา, 2538) | 10 LP Coefficients<br>64-vector codebook | ตัวเลขไทย | 10 | 82.00 % |
| ระบบรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (วิศรุต อาชุบุตร, 2539) | 10 LP Coefficients<br>128 and 256 vectors | คำพยางค์เดียว สองพยางค์และสามพยางค์ | 70 | 89.91 % |
| การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซีและโครงข่ายประสาทเทียมแบบแบ็กพรอพาเกชัน (วุฒิพงษ์ พรสุขจันทรา, 2539) | 10 LP Coefficients | ตัวเลขไทย | 10 | 89.40 % |

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

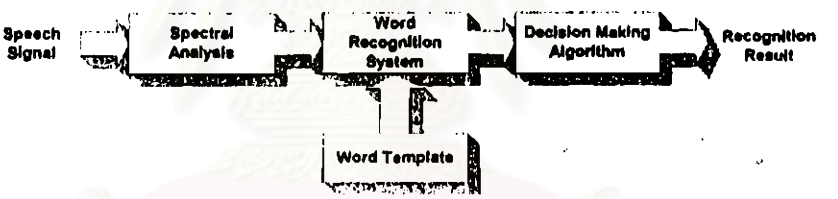
ตารางที่ 1.1 (ต่อ) งานวิจัยเกี่ยวกับระบบรู้จำเสียงพูดภาษาไทย

| กรรมวิธี | ค่าลักษณะสำคัญ | ประเภทคำศัพท์ | จำนวนคำศัพท์ | อัตราการรู้จำ |
|---|---|---|---|---|
| การรู้จำเสียงพูดคำไทยหลายพยางค์แบบไม่ขึ้นต่อผู้พูดโดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวอร์ค (ชัย วุฒิวิวัฒน์ชัย, 2540) | 10 LP Coefficients | คำพยางค์เดียว สองพยางค์และสามพยางค์ | 70 | 91.90 % |
| Tone Recognition for Thai (Tungthangthum, 1998) | ความถี่มูลฐาน | สระภาษาไทย | 10 | 91.00 % |
| การรู้จำเสียงพูดตัวเลขไทยแบบหลายผู้พูดด้วยนิวรัลเน็ตเวอร์ค (ไชยันต์ สุวรรณชีวะศิริ, 2541) | กลุ่มความถี่ฟอร์แมนท์ที่หนึ่งและสอง สัมประสิทธิ์พหุนามของคาบเวลาพิตช์ | ตัวเลขไทย | 10 | 91.60 % |

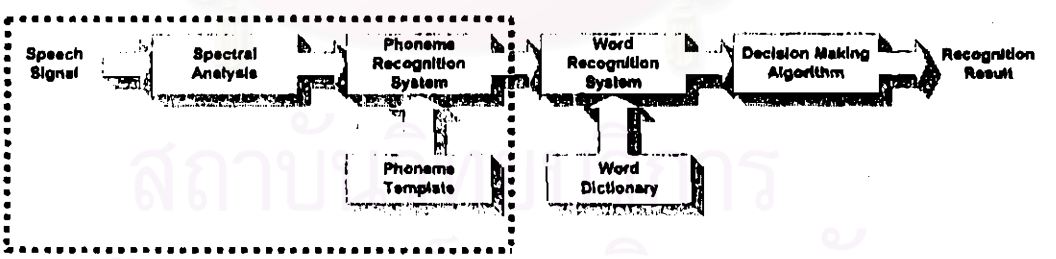
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย




รูปที่ 1.1 แบบจำลองระบบรู้จำเสียงพูดภาษาไทย (วิศรุต อาชุบุตร, 2539)




(ก)




(ข)

รูปที่ 1.2 ระบบรู้จำเสียงพูดที่ใช้แบบจำลองของคำโดดและหน่วยเสียง (ชัย วุฒิวิวัฒน์ชัย, 2540)