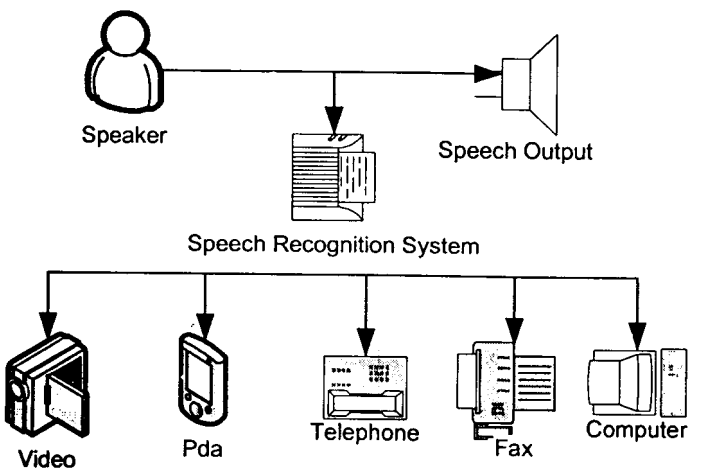
บทที่ 1

บทนำ

1.1 บทนำต้นเรื่อง

ระบบรู้จำเสียงพูด (speech recognition system) หมายถึง ระบบที่เรียนรู้และเข้าใจเสียงพูด เมื่อได้ยินเสียงพูดใดๆก็จะแสดงผลการรู้จำเป็นกลุ่มคำศัพท์ หรือ ประโยค ที่มีการออกเสียงตรงกัน ระบบนี้ได้มีการนำไปประยุกต์ใช้กับงานต่างๆ เพื่ออำนวยความสะดวกเช่น การพิมพ์เอกสารด้วยเสียง การควบคุมเครื่องใช้ไฟฟ้าด้วยเสียง (เสมือนรีโมตคอนโทรล) การตอบรับโทรศัพท์อัตโนมัติแบบคุยโต้ตอบได้ การโทรออกด้วยเสียงแทนการกดปุ่ม ดังภาพประกอบ 1.1 ในปัจจุบันมีซอฟต์แวร์ที่เป็นระบบรู้จำเสียงพูดมากมายในท้องตลาดทั้งที่เป็นฟรีซอฟต์แวร์ และ ซอฟต์แวร์สำหรับจำหน่าย ฟรีซอฟต์แวร์ได้แก่ Hidden Markov Model Tool Kit: HTK (Young, S., et al., 2000) ซอฟต์แวร์สำหรับจำหน่ายได้แก่ IBM ViaVoice (IBM. ,2000) แต่การใช้งานยังไม่แพร่หลายเนื่องจากปัญหาหลายด้าน เช่น ระดับความแม่นยำไม่สูงพอ ภาษาที่สามารถนำไปใช้งานได้ยังจำกัดอยู่ที่ภาษาอังกฤษ สภาพแวดล้อมไม่เหมาะสม เนื่องจากสัญญาณรบกวน ฯลฯ



ภาพประกอบ 1.1 ระบบรู้จำเสียงพูดกับการประยุกต์ใช้ในงานต่าง ๆ

ปัจจัยหนึ่งที่ทำให้ระบบรู้จำยังไม่เป็นที่นิยมคือ ภาษาที่เหมาะสมกับระบบรู้จำเพราะระบบส่วนใหญ่ก็มีวิธีหรือเทคนิคในการรู้จำภาษาที่เหมาะกับภาษาใดภาษาหนึ่งโดยเฉพาะภาษาอังกฤษ

สำหรับภาษาไทยมีโครงสร้างภาษาที่แตกต่างออกไปก็จะมีวิธีและเทคนิคในการรู้จำที่แตกต่างกัน ซึ่งจะเห็นได้จากงานวิจัยที่เกี่ยวข้องกับระบบรู้จำเสียงพูดภาษาไทย ผลงานวิจัยเหล่านี้จะแสดงถึงเทคนิคที่ใช้ในกระบวนการรู้จำเสียงพูด และเป็นแนวทางในการพัฒนา แต่การนำงานวิจัยเหล่านี้ไปใช้งานกับเสียงพูดทั่วไปยังไม่สามารถทำได้เนื่องจาก ข้อมูลเสียงที่นำมาใช้มีจำนวนน้อย ขอบเขตของคำศัพท์ที่ใช้ยังอยู่ในวงแคบคือ 10 – 100 คำที่ใช้บ่อย เช่น ตัวเลข ราคาสินค้า ประโยคบอกเล่าทั่วไป คำสั่งควบคุมเครื่องใช้ไฟฟ้า ซึ่งไม่มีคุณสมบัติพอที่จะนำไปใช้กับการรู้จำเสียงพูดทั่วไปได้ นอกจากนี้ยังไวยากรณ์ของภาษาไทยก็ยังไม่ได้นำมาใช้อย่างเหมาะสม

ปัญหาส่วนใหญ่ของระบบรู้จำภาษาไทยมาจากการพัฒนาแบบกระจัดกระจาย ส่วนหนึ่งเน้นการรู้จำคำบางคำ เน้นการรู้จำตัวเลข เน้นการคำนวณคุณลักษณะเด่น (feature extraction) แบบใหม่ เน้นวิธีการประมวลผลการรู้จำแบบใหม่ แต่ไม่สามารถนำมาพัฒนาร่วมกันเพราะใช้ปัจจัยแวดล้อมที่ต่างกัน นอกจากนี้ตัวระบบรู้จำยังไม่มีการกำหนดข้อมูลให้เป็นมาตรฐานที่ชัดเจน เช่นข้อมูลเสียงควรมีการจัดกลุ่มหมวดหมู่ไว้สำหรับนักพัฒนารุ่นต่อไป เพื่อให้ผู้ทำการวิจัยใช้ข้อมูลดังกล่าวให้เป็นมาตรฐานเดียวกัน ในขณะที่ภาษาอังกฤษมีฐานข้อมูลกลางสำหรับใช้งานในระบบรู้จำเสียงพูด (Cole, R., Muthusamy, Y. and Fanty, M., 1990)

จากปัญหาที่กล่าวมาผู้เขียนจึงได้คิดวิธีการแก้ปัญหาโดยการสร้างระบบรู้จำเสียงพูดทั้งระบบที่มีการกำหนด ชุดข้อมูลเสียง (speech database) โครงสร้างระบบรู้จำ (HMM structure) ชุดหน่วยเสียงที่ใช้ (phonetic set) วิธีการคำนวณความแม่นยำ ที่เป็นมาตรฐานในการพัฒนา ทำให้เกิดความสะดวกในการเปรียบเทียบและการใช้งานร่วมกับเทคนิคอื่นๆได้ ซึ่งระบบรู้จำเสียงพูดต้นแบบนี้จะใช้เป็นเกณฑ์และมาตรฐานสำหรับการพัฒนาระบบรู้จำเสียงพูดต่อไปในอนาคต

1.2 การตรวจเอกสาร

งานวิจัยส่วนใหญ่เน้นการค้นคว้าเทคนิคและวิธีการต่างๆ เพื่อให้ระบบรู้จำมีค่าความแม่นยำสูงสุด เช่น "ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้แบบจำลองฮิดเดนมาร์คอฟ" (วิศรุต อาชุบุตร, สมชาย จิตะพันธ์สกุล และ สุดาพร ลักษณียนาวิน, 2539) เป็นการพัฒนาระบบรู้จำเสียงพูดโดยใช้วิธีการคำนวณตามหลักการของฮิดเดนมาร์คอฟ (hidden Markov model) เพื่อพิสูจน์ว่าวิธีดังกล่าวสามารถนำมาใช้ในการรู้จำเสียงพูดได้ แต่ข้อมูลเสียงและจำนวนคำศัพท์ที่ยังมีน้อย เพราะเป็นการพัฒนาในช่วงแรก หรือ "การรู้จำตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซี และ นิวรอลเน็ตเวิร์คแบบแบ็กพรอพาเกชัน" (วุฒิพงษ์ พรสุขจันทรา และ สมชาย จิตะพันธ์สกุล, 2539) เป็นการใช้คุณลักษณะเด่นแบบ แอลพีซี (LPC, linear prediction coefficient) บน

ระบบการรู้จำแบบนิวรอลเน็ตเวิร์ค (neural network) ซึ่งคำศัพท์ที่ใช้ยังมีจำนวนน้อย และ "การรู้จำเสียงพูดโดยใช้ลักษณะแบ่งความต่าง" (ชินภัทร นันทจิวากรชัย. และคณะ. 2536.) เป็นหลักการอีกอย่างหนึ่งที่ใช้ในการรู้จำเสียงพูดซึ่งจะใช้วิธีการคำนวณที่แตกต่างออกไป จะเห็นได้ว่างานวิจัยทั้งสามนี้เป็นการพัฒนาระบบรู้จำเสียงพูดเหมือนกัน แต่ผลงานวิจัยทั้ง 3 มีปัจจัยแวดล้อมที่ต่างกันได้แก่ ข้อมูลเสียงที่ใช้ในการสอน ข้อมูลเสียงที่ใช้ในการทดสอบ ชุดคำศัพท์ที่รู้จำได้ โครงสร้างของระบบรู้จำ ทำให้การพัฒนาระบบรู้จำเสียงพูดยังไม่มีความก้าวหน้าไปในทิศทางเดียวกัน

ในงานวิจัยบางกลุ่มได้กล่าวถึงหลักการพื้นฐานสำหรับการพัฒนาระบบรู้จำเสียงพูดภาษาไทย เช่น หน่วยเสียงภาษาไทยทั้งหมดสำหรับการรู้จำ "Syllable Structure Base Phonetic Units for Context-Dependent Continuous Thai Speech Recognition" (Kanokphara, S., 2003) เป็นงานวิจัยที่แสดงความสามารถในการรู้จำของหน่วยเสียงชุดต่างๆ ที่ใช้ในระบบรู้จำเสียงพูดประกอบกับการใช้เทคนิคต่างๆ เพื่อพัฒนาการรู้จำให้ดีขึ้น รายละเอียดของงานวิจัยนี้เป็นการรู้จำเสียงพูดของหน่วยเสียงภาษาไทยทั้งหมด จำนวน 70 หน่วยเสียงไม่รวมวรรณยุกต์ และใช้แบบจำลอง ฮิดเดนมาร์คอฟ วิธีการคำนวณคุณลักษณะเด่นแบบสัมประสิทธิ์เคปสตรัมความถี่เมล (MFCC, Mel scale frequency cepstrum coefficient) ผลงานวิจัยชิ้นนี้สรุปได้ว่าการทำระบบให้รู้จำหน่วยเสียงจำนวนมากได้นั้นจำเป็นต้องมีตัวอย่างเสียงจำนวนมากพอ

ในการรู้จำวรรณยุกต์ภาษาไทยได้มีงานวิจัย Pitch Detection Algorithm: Auto Correlation Method (Tan, L. and Karnjanadecha, M., 2003) เป็นการพัฒนาคุณสมบัติเด่นของเสียงเพื่อใช้ในการจำแนกระดับวรรณยุกต์ โดยใช้หลักการคำนวณค่าพิตช์ บนระบบโครงข่ายประสาทเทียม (neural network) แต่ระบบนี้จำแนกวรรณยุกต์ของเสียงตัวเลขเท่านั้นไม่ได้ครอบคลุมคำศัพท์ทั้งหมดในภาษาไทย

1.3 วัตถุประสงค์

1) เพื่อศึกษาและพัฒนาเทคนิคที่เหมาะสมสำหรับการรู้จำเสียงพูดภาษาไทยแบบคำต่อเนื่อง
2) เพื่อพัฒนาระบบรู้จำเสียงพูดโดยใช้หลักการรู้จำตามโครงสร้างของภาษาไทย
3) เพื่อศึกษาปัญหาและสาเหตุของปัญหาในการพัฒนาระบบรู้จำเสียงพูดทั้งระบบ

1.4 ขอบเขตของการวิจัย

1) เป็นระบบรู้จำที่สามารถรู้จำเสียงพูดภาษาไทยสำเนียงภาคกลางแบบคำต่อเนื่องโดยไม่จำกัดผู้พูด จากฐานข้อมูลเสียงพูดที่มีอยู่ (Thongprasert, R., 2002)
2) เป็นแบบจำลองเสียงที่พัฒนาโดยใช้หลักการของฮิดเดนมาร์คอฟ ด้วยชุดหน่วยเสียงจากงานวิจัย (Thongprasert, R., 2002) โดยใช้ชุดโปรแกรม HTK ในการพัฒนา
3) เป็นระบบรู้จำเสียงพูดต้นแบบที่รับเสียงพูดของผู้ใช้ผ่านทางไมโครโฟนและแสดงผลการรู้จำที่จอคอมพิวเตอร์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้แบบจำลองหน่วยเสียง (phonetic model) สำหรับระบบรู้จำเสียงพูด
2) ได้ไวยากรณ์ภาษา (language model) ที่เหมาะสมกับระบบรู้จำเสียงพูดภาษาไทย
3) ได้ต้นแบบระบบรู้จำเสียงพูดที่สามารถนำไปพัฒนาต่อได้