หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

ระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยนี้เป็นระบบที่ทำการสร้างเสียงพูดออกมาโดยใช้หลักการของการสังเคราะห์เชิงวิเคราะห์ (analysis synthesis) ซึ่งให้คุณภาพในการฟังออกค่อนข้างดี ระบบทั้งระบบจำลองลงบนเครื่องไมโครคอมพิวเตอร์และเป็นซอฟแวร์ทั้งหมดยกเว้นส่วนที่ทำหน้าที่แปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิตอลที่เป็นฮาร์ดแวร์ ขั้นตอนแรกจะต้องนำเสียงพูดต้นแบบมาทำการวิเคราะห์ก่อนด้วยเทคนิคการทำนายแบบเชิงเส้น (LPC) ออร์เดอร์ 10 และหาคาบของพิทช์ด้วยเทคนิคเซนเตอร์คลิปปิง ได้ชุดพารามิเตอร์ที่ใช้แทนสัญญาณเสียงพูดต้นแบบรวมกันออกมาเก็บเอาไว้เพื่อทำการสังเคราะห์กลับมาเป็นเสียงพูดอีกที โดยระบบนี้ใช้พยางค์เป็นหน่วยย่อยของคำพูดในการวิเคราะห์ จากนั้นก็จะแก้ไขปรับปรุงค่าพารามิเตอร์บางตัวเพื่อให้เสียงแต่ละพยางค์ที่สังเคราะห์ออกมาได้มีความถูกต้องแล้วเก็บเข้าไว้ในพจนานุกรมข้อมูลต่อไป จำนวนคำที่ระบบสามารถสร้างขึ้นมาได้ก็ขึ้นอยู่กับจำนวนพยางค์ทั้งหมดที่ผ่านการวิเคราะห์แล้วเก็บเอาไว้ในพจนานุกรมข้อมูล ขั้นตอนต่อไปก็จะเป็นการสร้างเสียงพูดขึ้นมาตามข้อความภาษาไทยที่ป้อนเข้ามาเป็นอินพุทสู่ระบบ ความสามารถในด้านของการประมวลผลข้อมูล (text processing) อย่างเต็มรูปแบบยังไม่ได้รวมเข้าไว้ในระบบอันนี้ นั่นคือยังไม่สามารถรับอินพุทเข้ามาเป็นตัวอักษรในภาษาเขียนทั่วๆไปได้ หรือยังไม่สามารถทำการแยกแยะวิเคราะห์ภาษาเขียนออกมาเป็นเสียงอ่านที่ถูกต้องทั้งหมดได้ การป้อนข้อความอินพุทจะป้อนเป็นรูปของคำอ่านซึ่งจะต้องมีอยู่ในพจนานุกรมข้อมูลประกอบกับสัญญลักษณ์พิเศษที่ใช้กำหนดลักษณะของเสียงที่จะให้เปล่งออกมา นั่นคือสามารถที่จะกำหนดช่วงว่างหรือจังหวะระหว่างพยางค์หรือระหว่างคำได้ และสามารถเปลี่ยนแปลงลักษณะทางด้านความถี่ของเสียงในแต่ละพยางค์ได้โดยการใส่สัญญลักษณ์กำหนดการแก้ไขค่าคาบของพิทช์ได้ตามต้องการ เมื่อระบบได้รับข้อความอินพุทแล้วก็จะไปดึงชุดพารามิเตอร์ของพยางค์นั้นๆจากพจนานุกรมข้อมูลมาผ่านการสังเคราะห์เสียงพูดโดยใช้เทคนิคของตัวกรองแลททิส (lattice filter) ได้สัญญาณเสียงสังเคราะห์ออกมาตามต้องการ

จากผลการทดสอบโดยการนำเอาผู้ทดสอบจำนวน 20 คน มาทดลองฟังเสียงที่สังเคราะห์ขึ้นมาเพื่อวัดผลด้านการรับรู้จากการฟัง เสียงที่สังเคราะห์ขึ้นมาทำการทดสอบแบ่ง

เป็น 2 ลักษณะคือคำพูดพยางค์เดี่ยวจำนวน 22 คำ และคำพูดหลายพยางค์จำนวน 22 วลี ผลการทดสอบเฉลี่ยวัดออกมาได้เป็นร้อยละที่รับรู้ได้ถูกต้องสำหรับคำพูดพยางค์เดี่ยวเท่ากับร้อยละ 99.10 และสำหรับคำพูดหลายพยางค์เท่ากับร้อยละ 98.86

การวิจัยเกี่ยวกับการแปลงอักขระภาษาไทยให้ออกมาเป็นเสียงพูดนั้นยังคงเป็นสิ่งที่ใหม่เป็นสิ่งที่เพิ่งจะเริ่มต้นเท่านั้นสำหรับในบ้านเรา ถึงแม้ว่าสำหรับภาษาอังกฤษซึ่งเป็นภาษาที่มีการบุกเบิกมาก่อนในด้านนี้จะมีการทำสำเร็จถึงขั้นประยุกต์ออกมาในรูปแบบต่างๆในลักษณะการค้ามานานแล้วก็ตาม นับจากปี ค.ศ. 1976 ที่เริ่มจะมีการกล่าวขวัญกันอย่างแพร่หลายในเรื่องทฤษฎีของการสังเคราะห์เสียงพูดจนถึงปัจจุบันก็นับเป็นเวลากว่า 14 ปีแล้วที่มีการวิจัยค้นคว้ากันทางด้านการสังเคราะห์เสียงพูดออกมา วิธีใหม่ๆก็เริ่มมีการประกาศตามออกมาเพื่อปรับปรุงประสิทธิภาพให้ดีขึ้นกว่าเดิมชดเชยส่วนที่ไม่ดีของวิธีเก่าๆ แต่สำหรับการประยุกต์ใช้กับภาษาไทยนั้นยังเพิ่งจะมีการเริ่มต้นกันอย่างจริงจังเมื่อไม่นานมานี้ อาจจะเนื่องจากติดขัดปัญหาในด้านต่างๆ ทั้งนี้ก็คงเพราะว่าการวิจัยในลักษณะนี้จำเป็นที่จะต้องใช้ความรู้และความร่วมมืออย่างมากจากทางด้านสัทศาสตร์และภาษาศาสตร์เป็นหลัก เพื่อทำการรวบรวมข้อกำหนดกฎเกณฑ์ต่างๆเกี่ยวกับการออกเสียงพูดทั้งหมดในภาษาไทยซึ่งเป็นภาษาที่มีลักษณะเฉพาะตัวต่างไปจากภาษาอื่นๆ เช่นภาษาอังกฤษ แล้วนำมาประยุกต์เข้ากับทฤษฎีและอัลกอริทึมทางด้านการวิเคราะห์และสังเคราะห์เสียงและการนำมาประยุกต์ลงบนระบบคอมพิวเตอร์อีกทีเพื่อให้ได้บทสรุปออกมา

งานวิจัยนี้ได้เริ่มต้นจากทางด้านทฤษฎีการวิเคราะห์และสังเคราะห์เสียงพูดและการนำมาประยุกต์ลงบนระบบไมโครคอมพิวเตอร์ เพื่อมุ่งให้เกิดความสะดวกคล่องตัวกว่าระบบคอมพิวเตอร์ที่ใหญ่กว่า แต่ก็ทำให้อยู่ภายใต้ข้อจำกัดบางอย่างของระบบไมโครคอมพิวเตอร์เช่น ความเร็วในการคำนวณ ปริมาณของหน่วยความจำ ฯลฯ

งานวิจัยในด้านการสังเคราะห์เสียงพูดจากข้อความนี้ก็สามารถที่จะทำได้เช่นเดียวกับภาษาอังกฤษ และยิ่งกว่านั้นเราสามารถที่จะเน้นถึงลักษณะเฉพาะของภาษาไทยอันหนึ่งในด้านของเสียงวรรณยุกต์ซึ่งไม่มีในภาษาอังกฤษออกมาได้ ดังนั้นผลงานวิจัยนี้ผู้วิจัยจึงคิดว่าได้ทำให้เป็นจุดเริ่มต้นจุดหนึ่งแล้วสำหรับหัวข้อการสังเคราะห์เสียงพูดจากข้อความภาษาไทย แต่ก็ยังต้องการการวิจัยเพื่อพัฒนาต่อเนื่องไปสู่ผลงานที่สมบูรณ์แบบกว่านี้ ซึ่งก็ยังต้องการความร่วมมือในการพัฒนาโดยเฉพาะทางด้านสัทศาสตร์และภาษาศาสตร์อีกมาก จึงหวังว่าจากจุดเริ่มต้นจุดนี้เป็นต้นไปคงจะมีการร่วมมือกันเพื่อทำการพัฒนาผลงานวิจัยในด้านนี้ให้มีความก้าวหน้าและผลงานอันสมบูรณ์อันจะเป็นการส่งเสริมการวิจัยที่ลึกซึ้งในด้านนี้ต่อไป

5.2 ข้อเสนอแนะและแนวทางในการวิจัยต่อ

1) เนื่องจากขั้นตอนของการสังเคราะห์เสียงพูดหลังจากที่รับอักขระภาษาไทยมาแล้ว ในงานวิจัยชิ้นนี้ยังคงไม่สามารถที่จะทำให้เป็นไปในลักษณะประมวลผลเวลาจริงได้ จากข้อจำกัดอันนี้มีแนวทางที่แก้ไขให้ระบบสามารถประมวลผลเวลาจริงได้อยู่หลายแนวทางดังนี้

- แก้ไขที่ข้อจำกัดของเครื่องไมโครคอมพิวเตอร์ ในการวิจัยนี้ใช้เครื่องไมโครคอมพิวเตอร์รุ่น XT ซึ่งเป็นเครื่องที่มีประสิทธิภาพค่อนข้างจะต่ำมากในปัจจุบัน ถึงแม้จะมีตัวประมวลผลช่วยทางคณิตศาสตร์ 8087 อยู่ด้วยก็ตาม ดังนั้นถ้าเปลี่ยนไปใช้เครื่องไมโครคอมพิวเตอร์ที่มีประสิทธิภาพสูงขึ้นกว่านี้แทนอย่างเช่น ไมโครคอมพิวเตอร์รุ่น AT ที่ดำเนินการที่สัญญาณนาฬิกาขนาด 10,12,16,20 MHz หรือใช้ไมโครคอมพิวเตอร์ที่ใช้ตัวประมวลผลเบอร์ 80386 ที่ดำเนินการที่สัญญาณนาฬิกาขนาด 20,25,33 MHz เป็นต้น และควรจะมีตัวประมวลผลช่วยทางคณิตศาสตร์ประกอบอยู่ด้วย ก็จะทำให้ความเร็วในการประมวลผลเร็วขึ้นตามลำดับจนถึงขั้นเวลาจริงได้ แต่แนวทางนี้ก็มีข้อจำกัดอยู่ที่ค่าใช้จ่ายที่ยังนับว่าสูงอยู่จะต้องพิจารณาดูว่าคุ้มกันหรือไม่

- แก้ไขที่รูทีนการสังเคราะห์เสียง เนื่องจากเรามีไมโครโปรเซสเซอร์ความเร็วสูง TMS32010 อยู่แล้วบน TSB บอร์ด ดังนั้นก็อาจที่จะแปลงรูทีนในการคำนวณที่ใช้เวลามากเมื่อดำเนินการโดยไมโครคอมพิวเตอร์เป็นภาษาแอสเซมบลี้ของ TMS32010 ซึ่งจะให้ความเร็วสูงกว่ามากเพราะคุณสมบัติพิเศษของมัน ก็จะทำให้มีการประมวลผลแบบเวลาจริงได้ แต่ก็จะต้องมีการแก้ไขโปรแกรมเดิมเพื่อให้สามารถทำการสลับสับเปลี่ยนหน้าที่ในการคำนวณและส่งถ่ายข้อมูลในช่วงเวลาที่เหมาะสมระหว่างไมโครคอมพิวเตอร์และ TMS32010

2) หาทางลดปริมาณข้อมูลหรือลดขนาดของพจนานุกรมข้อมูลให้ใช้เนื้อที่น้อยลงไปอีก โดยอาจจะหาเทคนิคในการลดขนาดของข้อมูลลงแต่ต้องไม่ทำให้คุณภาพของเสียงเสียไปด้วยจากเหตุผลของการประยุกต์ไปใช้งานหลายๆด้าน ถ้ายิ่งมีการลดขนาดของพจนานุกรมข้อมูลได้มากเท่าไรก็จะยิ่งมีประโยชน์มากขึ้นเท่านั้น เหตุผลดังกล่าว เช่น

- การเก็บหน่วยพยางค์ภาษาไทยให้ครบทุกพยางค์ เพื่อจะทำให้ระบบสามารถสังเคราะห์เสียงภาษาไทยได้ครบถ้วนทุกพยางค์ โดยจำนวนพยางค์ทั้งหมดในภาษาไทยมีอยู่ 30,096 พยางค์ แบ่งเป็นที่ใช้จริงๆอยู่ในภาษาไทย 4,104 พยางค์ ที่เหลืออีก 25,992 พยางค์ เป็นพยางค์ที่ใช้เปล่งออกมาเป็นเสียงได้แต่ไม่ค่อยปรากฏอยู่ในภาษาไทย (Sudaporn Luksaneeyanawin, 1989) ซึ่งจะทำให้พจนานุกรมข้อมูลมีขนาดใหญ่โตมากไม่สะดวกในการใช้งานทำงานช้า หรือสิ้นเปลืองค่าใช้จ่ายมาก

- การประยุกต์ในด้านการสื่อสารข้อมูลดิจิตอล ในกรณีที่จะต้องมีการส่งถ่ายข้อมูลของพจนานุกรมข้อมูลเพื่อนำไปสังเคราะห์เป็นเสียงในที่อื่นที่ไกลออกไป ก็จำเป็นที่จะต้องส่งข้อมูลที่เป็นค่าพารามิเตอร์ต่างๆไปตามสายสื่อสาร ซึ่งค่าใช้จ่ายและอาจจะรวมถึงเวลาในการส่งข้อมูลก็จะขึ้นกับปริมาณของข้อมูลที่ส่งไป ดังนั้นถ้าปริมาณข้อมูลยิ่งน้อยลงเท่าไรก็จะทำให้ส่ง

ข้อมูลได้เร็วเท่านั้นและประหยัดค่าใช้จ่ายด้วย

3) ปรับปรุงคุณภาพของเสียงที่ได้จากการสังเคราะห์ซึ่งมีแนวทางทำได้หลายขั้นตอนคือ

- เพิ่มความละเอียดของภาค A/D และ D/A เช่นอาจจะเพิ่มความละเอียดจาก 12 บิท เป็น 16 บิทหรือมากกว่า

- อุปกรณ์ ภาคขยายเสียง ไมโครโฟน ลำโพง วงจรกรองสัญญาณความถี่ต่ำผ่าน อาจจะเปลี่ยนไปใช้อุปกรณ์ที่มีประสิทธิภาพที่สูงกว่านี้ มีคุณภาพดีและมีความเชื่อถือได้มากกว่าที่ใช้อยู่

- หาทางลดระดับของสัญญาณรบกวนที่เกิดขึ้นในระบบเพื่อให้ค่าอัตราส่วนระหว่างสัญญาณเสียงกับสัญญาณรบกวนมีค่ามากขึ้น จะทำให้ได้เสียงต้นฉบับที่มีคุณภาพดีกว่านี้ซึ่งก็หมายถึงคุณภาพของเสียงสังเคราะห์จะดีขึ้นตามไปด้วย

- เพิ่มอุปกรณ์ที่ใช้ในการปรับแต่งเสียงหรือที่เรียกว่า อิควอไลซ์เซอร์ (equalizer) แบบหลายแถบความถี่ เพื่อนำมาใช้ในการปรับแต่งความถี่ของเสียงสังเคราะห์ที่ได้ออกมา เช่น กดความถี่ที่เป็นช่วงของสัญญาณรบกวนลง หรือขยายความถี่ช่วงกลางๆเพื่อเน้นความชัดเจนเฉพาะเสียงสังเคราะห์เท่านั้น

- เพิ่มอันดับในการคำนวณการทำนายแบบเชิงเส้น จากเดิมใช้ค่า 10 อาจจะเพิ่มเป็น 12 หรือ 15 ก็ได้ แต่ก็จะทำให้ขนาดของพจนานุกรมข้อมูลใหญ่ขึ้นกว่าเดิมด้วย

- เลือกเสียงต้นแบบที่พูดได้ชัดเจนฟังได้ง่ายกว่านี้

- เปลี่ยนขนาดของเฟรมในการวิเคราะห์ให้เหมาะสมกับเสียงของพยางค์ เช่น เสียงพยางค์ที่สั้นๆ จะมีการเปลี่ยนแปลงเร็วอาจจะต้องใช้จำนวนจุดของข้อมูลต่อเฟรมให้น้อยกว่า 200 จุดก็ได้เพื่อเพิ่มความละเอียดในการพิจารณาวิเคราะห์สัญญาณเสียง

- เปลี่ยนค่าคาบของพิทช์ให้สามารถผันเสียงตามเสียงวรรณยุกต์ได้ชัดขึ้น

4) แก้ไขให้เสียงที่สังเคราะห์ออกมามีลักษณะเป็นธรรมชาติมากขึ้นกว่านี้ เช่นมีการเชื่อมต่อระหว่างพยางค์ที่ราบเรียบกว่านี้ เน้นเสียงได้ กำหนดจังหวะของเสียงแต่ละพยางค์แต่ละคำ ทำให้เสียงพูดมีสำเนียงที่ถูกต้อง

5) ปรับปรุงภาครับอักขระให้มีความเป็นกันเองกับผู้ใช้มากกว่านี้ อาจจะให้ผู้ใช้พิมพ์ข้อความเป็นประโยคภาษาไทยตามภาษาเขียนเข้าไป หรือใช้แฟ้มข้อมูลที่เป็นข้อความหรือบทความต่างๆเป็นข้อมูลอินพุทได้โดยตรง นั่นคือต้องเพิ่มเติมส่วนของการวิเคราะห์ข้อความเหล่านี้ให้แตกออกมาเป็นพยางค์ด้วยก่อนที่จะนำไปสังเคราะห์เป็นเสียงต่อไป ซึ่งอาจจะต้องใช้กฎเกณฑ์ต่างๆมากำหนดเพื่อให้สามารถครอบคลุมการแปลงภาษาเขียนเป็นภาษาพูดได้ทั้งหมด จนอาจจะถึงขั้นต้องใช้ความรู้ทางด้านปัญญาประดิษฐ์มาช่วย

6) ศึกษาข้อดีข้อเสียของการใช้พยางค์เป็นหน่วยย่อยของคำ หรืออาจจะเปลี่ยนหน่วยย่อยของคำจากพยางค์มาเป็นหน่วยย่อยของคำที่เล็กลงไปอีก เช่น เดมิ-ซิลลาเบิล (demi-syllable) หรือ โฟนีม (phoneme) หรือ อโลโฟน (allophone) เป็นต้น เนื่องจากหน่วย

ย่อยของคำยิ่งแตกย่อยเล็กลงก็จะยิ่งสามารถตัดส่วนย่อยที่ซ้ำซ้อนกันออกไปได้ ทำให้เหลือจำนวนหรือปริมาณของหน่วยย่อยที่จะใช้ประกอบเป็นคำขึ้นมานั้นน้อยลงไปเป็นอย่างมาก จะได้ลดขนาดของพจนานุกรมข้อมูลลงหรือเป็นการลดหน่วยความจำที่จำเป็นต้องใช้น้อยลง แต่จะต้องศึกษาวิธีการเชื่อมโยงหน่วยย่อยของเสียงเหล่านั้นเพิ่มขึ้น

7) ประยุกต์การใช้งานในลักษณะของฟังก์ชั่นที่ฝังตัวอยู่ในหน่วยความจำ โปรแกรมประยุกต์อื่นๆ สามารถจะเรียกใช้งานได้โดยตรง เมื่อต้องการให้เครื่องไมโครคอมพิวเตอร์เปล่งเสียงข้อความที่ต้องการออกมา

8) ศึกษาเทคนิคในการวิเคราะห์และสังเคราะห์เสียงพูดที่มีการคิดค้นขึ้นมาใหม่ๆนำมาประยุกต์ใช้แทนเทคนิคเก่าๆ ซึ่งอาจจะทำให้ระบบมีความสามารถและประสิทธิภาพเพิ่มมากขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย