บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

การสังเคราะห์เสียงสระภาษาไทยตามเสียงวรรณยุกต์ที่ได้ดำเนินการวิจัยในครั้งนี้อาศัยแบบจำลองของอวัยวะที่ใช้ในการเปล่งเสียง (articulatory synthesis) โดยระบบทั้งหมดได้ทำการจำลองลงบนไมโครคอมพิวเตอร์ ซึ่งจำลองอวัยวะและปรากฏการณ์ของการเกิดเสียงพูดด้วยแบบจำลองทางคณิตศาสตร์ เริ่มต้นด้วยการหาพื้นที่หน้าตัดของช่องทางเดินเสียง ซึ่งได้นำช่องทางเดินเสียงมาจากสระภาษาอังกฤษซึ่งมีสัทอักษรที่ตรงกันกับภาษาไทยจำนวน 6 สระ คือ อี, เอ, แอ, เออ, อา และอู ส่วนอีก 3 สระ คือ อือ, โอ และออ สามารถหาได้ด้วยการทดลองบนพื้นความรู้ด้านภาษาศาสตร์ในส่วนของลักษณะทางกายภาพของการออกเสียงสระ ต่อมาได้นำช่องทางเดินเสียงของสระเดี่ยวมาเป็นค่าตั้งต้นและค่าสิ้นสุดของแบบจำลองเพื่อที่จะสังเคราะห์ออกมาเป็นสระประสม 3 เสียงคือ เอีย (อี - อา), เอือ (อือ - อา) และอัว (อู - อา) ด้วยการเพิ่มลดขนาดช่องทางเดินเสียงจากสระตั้งต้นลงไปสระสิ้นสุด 3 รูปแบบคือแบบเชิงเส้น แบบลอกการิธึม และแบบเลขชี้กำลัง พบว่าการเพิ่มลดขนาดช่องทางเดินเสียงแบบเลขชี้กำลังให้ผลการสังเคราะห์ดีที่สุดกับสระเอียอย่างชัดเจน ส่วนสระเอือและและอูให้ผลดีกับการเพิ่มลดขนาดช่องทางเดินเสียงแบบเชิงเส้นและแบบลอกการิธึม แต่ได้กำหนดให้เป็นแบบลอกการิธึมเพราะให้เสียงดีกว่าเล็กน้อย

ทั้งนี้สระเสียงสั้นสามารถสังเคราะห์ขึ้นได้ด้วยช่องทางเดินเสียงเดียวกันกับคู่สระเสียงสั้นยาว โดยการหาพารามิเตอร์ในช่วงเวลาของการเกิดเสียงจากเสียงต้นแบบคือค่าของความถี่หลักมูลของเสียงด้วยวิธีออโตคอรีเลชัน รวมถึงการหากรอบคลื่นแอมปลิจูดเพื่อให้ได้เสียงสังเคราะห์ที่มีพลังงานเหมือนกับเสียงต้นแบบและมีความเป็นธรรมชาติมากยิ่งขึ้น และด้วยวิธีการเดียวกันนี้ได้นำไปหาพารามิเตอร์ต่าง ๆ เพื่อที่จะสังเคราะห์ให้ได้เป็นเสียงวรรณยุกต์ โดยใช้เสียงต้นแบบเป็นเสียงสระ อะ สำหรับเสียงสั้น และสระอาสำหรับเสียงยาว เพื่อนำไปเป็นพารามิเตอร์ให้กับการสังเคราะห์เสียงอื่นๆได้ด้วยเพราะค่าพลังงานและความถี่หลักมูลไม่ได้แตกต่างกันมากนักและเสียงที่สังเคราะห์ได้ก็ให้เสียงที่มีคุณภาพดีพอ ๆ กับที่สังเคราะห์เสียงสระอะ และสระอา

ได้ประเมินคุณภาพเสียงสังเคราะห์ด้วยวิธีทางความถี่ได้เปอร์เซ็นต์ความแตกต่างจากค่ามาตรฐานของฟอร์แมนท์ 1 เป็น 9.89 % และฟอร์แมนท์ 2 เป็น 7.44 % เทียบกับเสียงที่ได้บันทึกได้เปอร์เซ็นต์ความแตกต่างจากเสียงบันทึกของฟอร์แมนท์ 1 เป็น 16.40 % และฟอร์แมนท์ 2 เป็น 13.99 %

ระบบที่ได้มีข้อดีตรงที่มีความยืดหยุ่นสูง การปรับพารามิเตอร์ต่าง ๆ เพื่อสังเคราะห์ให้ได้ออกมาเป็นเสียงที่มีคุณภาพที่ดีทำได้ง่าย เช่น การสังเคราะห์เพื่อให้ได้ค่าพลังงานเหมือนต้นแบบสามารถทำได้โดยตรงด้วยการนำค่าของกรอบคลื่นแอมปลิจูดของเสียงมาใช้เป็นค่าของอัตราขยายได้เลย และโดยเฉพาะอย่างยิ่งกับการสังเคราะห์เสียงภาษาไทยซึ่งเป็นภาษาที่มีเสียงวรรณยุกต์บ่งบอกความแตกต่างของคำก็สามารถทำได้โดยตรงจากค่าของความถี่ของเสียงวรรณยุกต์นั้น ๆ โดยไม่ทำให้ฟอร์แมนท์ของเสียงเปลี่ยนแปลงไปจากเดิม นอกจากนั้นเสียงที่สังเคราะห์ได้จะมีความต่อเนื่องตลอดช่วงที่ทำการสังเคราะห์ ตัวอย่างเช่นการสังเคราะห์เสียงสระประสม ซึ่งอันที่จริงประกอบด้วย 2 หน่วยเสียงประกอบกัน พบว่าเสียงที่ได้มีความต่อเนื่องทั้งในโดเมนเวลา และโดเมนความถี่ เพราะฉะนั้นถ้านำข้อดีของแบบจำลองตามที่กล่าวมานี้ไปประยุกต์ใช้งานแล้วจะทำให้สามารถพัฒนาระบบสังเคราะห์เสียงที่มีความเป็นธรรมชาติได้ในอนาคต เช่น การเน้นอารมณ์

การสังเคราะห์เสียงจากแบบจำลองในงานชิ้นนี้สามารถทำได้ไม่จำกัด ซึ่งนอกจากการปรับความถี่หลักมูลของเสียง และกรอบคลื่นแอมปลิจูดเพื่อให้ได้ท่วงทำนองของเสียงที่เปลี่ยนไปตามลักษณะของเสียงที่ได้บันทึกของแต่ละคนแล้ว ยังสามารถที่จะปรับช่องทางเดินเสียงได้เพื่อให้ได้ลักษณะเฉพาะของเสียงแต่ละเสียงได้ แต่ต้องอาศัยความรู้ความเข้าใจด้านภาษาศาสตร์ในส่วนของลักษณะทางกายภาพของการเกิดเสียงด้วย

4.2 ข้อเสนอแนะและแนวทางการวิจัย

ข้อเสนอแนะและแนวทางการวิจัยของระบบสังเคราะห์เสียงพูดด้วยแบบจำลองการเปลี่ยนรูปร่างช่องทางเดินเสียงมีดังนี้

1. ควรที่จะหาความถี่หลักมูลและกรอบคลื่นแอมปลิจูดของทุกวรรณยุกต์ในแต่ละเสียงสระเพราะการออกเสียงของแต่ละคนแม้ว่าเป็นวรรณยุกต์เดียวกันแต่เป็นสระที่ต่างกันก็จะได้ค่าที่ต่างกัน สังเกตได้จากการฟังเสียงสระที่ออกเสียงโดยผู้พูดคนเดียวกันจะมีท่วงทำนองและการลงน้ำหนักที่ต่างกัน

2. อาจจะวิจัยในเรื่องของการเน้นอารมณ์ว่าการออกเสียงสระแต่ละครั้งอารมณ์ของผู้พูดมีส่วนต่อค่าความถี่หลักมูลและพลังงานอย่างไรเพื่อเป็นประโยชน์ในการพัฒนาการออกเสียงของคอมพิวเตอร์ให้มีความเป็นธรรมชาติมากยิ่งขึ้น

3. แม้ว่าการสังเคราะห์เสียงสระในงานชิ้นนี้จะให้ผลออกมาทำงานได้ตามเวลาจริง (real time) แต่อย่างไรก็ตามพอที่จะสังเกตได้ว่ายิ่งโปรแกรมมีการคำนวณมากครั้งการทำงานจะช้าลงกว่าเดิม ตัวอย่างเช่น การสังเคราะห์เสียงสระประสมซึ่งต้องคำนวณมากกว่าสระเดี่ยวเป็นจำนวนเท่าของลูกคลื่นที่ต้องการสังเคราะห์โดยประมาณแล้วสระเสียงยาวจะคำนวณมากกว่า 60 เท่า และสระเสียง

สั้นประมาณ 30 เท่า ซึ่งถ้าจะนำไปใช้งานควรที่จะทำงานได้ทันในหน่วยเสียงของพยางค์ได้ เพราะฉะนั้นแนวทางการปรับปรุงคือ

3.1 โปรแกรมที่พัฒนาขึ้นทำงานบนโปรแกรม MATLAB ซึ่งอาจจะใช้เวลาในการทำงานมากเพราะฉะนั้นถ้าเปลี่ยนเป็นโปรแกรมที่มีการทำงานที่เร็วขึ้นเช่นโปรแกรมภาษาซี จะทำให้สามารถทำงานได้เร็วขึ้นและพัฒนาได้ไม่จำกัด

3.2 ศึกษาความเป็นไปได้ในส่วนของการเพิ่มลดขนาดช่องทางเดินเสียงให้มีจำนวนครั้งในการเพิ่มลดน้อยลงกว่าเดิมซึ่งเป็นไปได้แน่นอนเพียงแต่ต้องหาค่าที่เหมาะสมซึ่งทำให้เสียงไม่เปลี่ยนแปลงไปจากเดิมมากนัก

4. การหาพื้นที่หน้าตัดควรได้มาจากการทำ MRI และทดสอบเสียงโดยอาศัยเสียงจากต้นแบบของช่องทางเดินเสียงที่ได้มาจากการทำ MRI จะทำให้สามารถวิเคราะห์พารามิเตอร์ต่าง ๆ ได้ง่ายขึ้นและได้เสียงออกมาใกล้เคียงมากยิ่งขึ้น เช่นการเพิ่มลดจำนวนเซกเมนต์เพื่อกำหนดจำนวนเซกเมนต์ที่เหมาะสมสำหรับสังเคราะห์เสียงพูดภาษาไทยโดยการวิจัยจะต้องดำเนินการควบคู่ไปกับการวิเคราะห์ความเร็วในของการทำงาน การใช้หน่วยความจำและคุณภาพของเสียงประกอบกันให้ได้ค่าที่ดีที่สุดออกมก ทั้งยังสามารถที่จะพัฒนาให้สามารถสังเคราะห์เสียงพยัญชนะได้โดยไม่ยาก ซึ่งแบบจำลองในงานวิจัยชิ้นนี้สามารถรองรับได้ในส่วนของเสียง พยัญชนะที่ไม่ใช่เสียงที่มีเสียงนาสิกเป็นองค์ประกอบ ซึ่งเสียงพยัญชนะในภาษาไทยมีทั้งหมด 21 เสียงเป็นเสียงนาสิกเพียง 3 เสียง คือ เสียง "ม", "น" และ "ง"

5. การสังเคราะห์เสียงในงานวิจัยชิ้นนี้ใช้ต้นแบบเป็นเสียงของผู้วิจัยเพียงเสียงเดียวเท่านั้น ดังนั้นควรใช้ตัวอย่างของเสียงที่ได้มาจากการบันทึกเพิ่มเติม จะนำไปสู่การสรุปในส่วนของพารามิเตอร์ที่ใช้กำหนดเสียงวรรณยุกต์ และการเพิ่มลดขนาดช่องทางเดินเสียงในระหว่างการสังเคราะห์เสียงสระประสมที่ดียิ่งขึ้น โดยการเลือกกลุ่มตัวอย่างควรครอบคลุมทั้งเสียงผู้ชายเสียงผู้หญิงรวมถึงควบคุมอายุของกลุ่มตัวอย่างให้กระจายทุกกลุ่ม ซึ่งตัวโปรแกรมเองสามารถรองรับงานที่ได้เสนอแนะทั้งหมด