บทที่ 6

บทสรุป

ผลการทดลองวิเคราะห์และสังเคราะห์เสียงพูด โดยกำหนดจำนวนออร์เดอร์ของฟิลเตอร์เท่ากับ 10 และแซมเปิลในหนึ่งเฟรมเท่ากับ 200 เสียงที่ได้มีคุณภาพใช้งานได้เมื่อพิจารณาจากการฟัง การลดทอนข้อมูล สามารถลดทอนได้ประมาณ 15 เท่า (จากตาราง 5.1) เสียงพูดที่ได้จากการสังเคราะห์ภายในเครื่องมาสเตอร์คอมพิวเตอร์ จะมีคุณภาพดีกว่าเสียงพูดที่ได้จากการคำนวณในภาคประมวลผลสัญญาณ เนื่องจากการคำนวณด้วยมาสเตอร์คอมพิวเตอร์ใช้ตัวเลขชนิด Floating Point ซึ่งมีความละเอียดสูงกว่าการคำนวณในภาคประมวลผลสัญญาณซึ่งใช้ตัวเลขชนิด Integer ขนาด 16 บิท ปัญหาเกี่ยวกับคุณภาพของเสียงกับข้อจำกัดของความละเอียดในการคำนวณ (Finite Word Length) เป็นสิ่งหลีกเลี่ยงไม่ได้สำหรับการคำนวณสัญญาณในลักษณะ Real-time เพราะความเร็วในการคำนวณเป็นสิ่งสำคัญ ถ้าคำนวณแบบ Double Word จะไม่สามารถทำงานในลักษณะ Real-time ได้ อย่างไรก็ตามข้อมูลขนาด 16 บิทก็มีความละเอียดสูงเพียงพอ รวมทั้งการคำนวณภายในภาคประมวลผลสัญญาณมีความละเอียด 32 บิท ซึ่งมีคุณภาพเสียงที่ได้ก็ลดลงไม่มากนัก

สำหรับการทดลองในหัวข้อ 5.1 เป็นการทดสอบคุณภาพเสียงโดยการเปลี่ยนแปลงจำนวนออร์เดอร์ของฟิลเตอร์ ซึ่งอาศัยค่าผิดพลาดนอร์มัลไลซ์เป็นแนวทางร่วมกับการทดลองวิเคราะห์และสังเคราะห์เสียงพูดคำว่า "กา"

สรุปได้ว่า จำนวนออร์เดอร์เท่ากับ 10 เป็นจำนวนออร์เดอร์ที่น้อยที่สุด โดยเสียงที่ได้จากการสังเคราะห์มีคุณภาพใช้งานได้เมื่อพิจารณาจากการฟัง

อย่างไรก็ตามการทดลองโดยการวิเคราะห์และสังเคราะห์เสียงคำศัพท์อื่นๆ จะเป็นสิ่งสนับสนุนข้อสรุปดังกล่าว ส่วนการทดลองหัวข้อ 5.2 ในเรื่องจำนวนแซมเปิลในหนึ่งเฟรม ที่สรุปว่าจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 200 ถึง 250 แซมเปิลนับว่าเหมาะสมนั้น จากการทดลองวิเคราะห์และสังเคราะห์เสียงพูดคำว่า "หนึ่ง" ถึง "สิบ" พบว่าจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 250 ใช้ไม่ได้สำหรับคำที่มีเสียงสั้น เช่น คำว่า "หก" "เจ็ด" และ "สิบ" ซึ่งมีความยาวของเสียงประมาณ 200 ms อย่างไรก็ตามการทดลองในหัวข้อ 5.3 ได้เลือกใช้จำนวนแซมเปิลในหนึ่งเฟรมเท่ากันหมดคือ 200 แซมเปิล ผลก็คือคำที่มีเสียงสั้นยังคงมีปัญหาอยู่บ้าง คือคำว่า "เจ็ด" ซึ่งหลังจากมีการเปลี่ยนแปลงจำนวนแซมเปิลในหนึ่งเฟรมเป็น 160 จะได้คุณภาพเสียงดีขึ้น อย่างที่กล่าวไว้ในบทที่ 5 ว่าจำนวนแซมเปิลในหนึ่งเฟรมที่เหมาะสม จะขึ้นอยู่กับการเปลี่ยนแปลงคุณสมบัติตามเวลาของเสียงคำนั้น ๆ ซึ่งต้องอาศัยการทดลองเป็นเฉพาะคำ ๆ ไป ในเรื่องจำนวนออร์เดอร์ของ

ฟิลเตอร์ก็เช่นกัน แม้จะไม่ขึ้นต่อการเปลี่ยนแปลงคุณสมบัติตามเวลามากนัก แต่ก็ขึ้นกับคุณสมบัติทางองค์ประกอบความถี่ของเสียงพูดคำนั้น ๆ โดยตรง จึงทำให้คุณภาพของเสียงพูดที่ได้จากการสังเคราะห์แต่ละคำแตกต่างกันดังเช่นผลการทดลองในหัวข้อ 5.3 ข้อเสนอแนะคือ ทำการทดลองวิเคราะห์และสังเคราะห์เป็นคำ ๆ เพื่อหาองค์ประกอบที่เหมาะสมสำหรับสังเคราะห์เสียงแต่ละคำ

ข้อจำกัดที่เกิดขึ้นในระบบวิเคราะห์และสังเคราะห์เสียงสามารถสรุปได้ คือ

1. การลดทอนข้อมูลยังมีขีดจำกัดเนื่องจากจุดประสงค์ของงานไม่ได้รวมถึงการทำ Optimum Parameter Quantization แต่การลดทอนในขั้นแรกที่ทำได้คือ ลดทอนจำนวนข้อมูลให้เหลือน้อยที่สุด แต่ถ้าคิดเป็นจำนวนบิทข้อมูลจะยังคงมีจำนวนมากเพราะใช้ข้อมูลขนาด 16 บิท อัตราข้อมูลของเสียงพูดที่ได้จากการทดลองคือ 10,400 bits/s ข้อเสนอแนะคือการทำ Optimum Parameter Quantization ซึ่งคาดว่าสามารถลดทอนจำนวนข้อมูลลงได้อีก 30% [21]

2. การคำนวณในภาคประมวลผลสัญญาณในช่วงต่อระหว่างเฟรมกินเวลามากกว่า 100 μs เนื่องจากต้องมีการอ่านข้อมูลของเฟรม ในกรณีที่เลือกจำนวนออร์เดอร์เท่ากับ 10 ระยะเวลาระหว่างแซมเปิลสุดท้ายในเฟรมที่แล้วกับแซมเปิลแรกในเฟรมปัจจุบันจะห่างกันประมาณ 120 μs แต่ผลกระทบต่อเสียงที่ออกมาไม่สามารถบอกได้ เนื่องจากการผิดพลาดเพียงหนึ่งแซมเปิลเทียบกับจำนวนแซมเปิลทั้งหมดในเฟรมซึ่งเท่ากับ 200 แซมเปิล ยังไม่ส่งผลกระทบ

3. เนื่องจากต้องการความยืดหยุ่นในการทดลองสังเคราะห์เสียง จึงอาศัยวิธีบรรจุข้อมูลเสียงลงไปในหน่วยความจำของภาคประมวลผลสัญญาณ ทำให้คำศัพท์มีจำนวนจำกัด ข้อเสนอแนะคือ ถ้าได้กำหนดจำนวนออร์เดอร์ที่แน่นอนและจำนวนแซมเปิลในหนึ่งเฟรมที่แน่นอน จะสามารถกำหนดให้โปรแกรมที่ทำการสังเคราะห์เสียงมีการรับข้อมูลเสียงจากมาสเตอร์คอมพิวเตอร์เพื่อได้คำศัพท์ที่ต้องการ

<u>แนวทางในการวิจัยขั้นต่อไป</u>

1. การศึกษาคุณสมบัติด้านวรรณยุกต์ของเสียงพูด เนื่องจากระหว่างการทดลองวิเคราะห์และสังเคราะห์เสียงพูดเป็นคำๆ พบว่าการเปลี่ยนแปลงค่าคาบของเสียงให้เป็นไปตามวรรณยุกต์ต่างๆ สามารถให้เสียงที่สื่อความหมายได้โดยไม่ต้องเปลี่ยนข้อมูลพารามิเตอร์ของฟิลเตอร์ ตัวอย่างเช่น นำข้อมูลคาบของเสียงพูดคำว่า "กา" มาดัดแปลงให้มีค่าเพิ่มขึ้นเท่าๆกัน จะได้เสียงของวรรณยุกต์เอกซึ่งเสียงจะกลายเป็นคำว่า "ก่า" ในทำนองเดียวกันถ้าเปลี่ยนข้อมูลคาบของเสียงให้เป็นไปตามวรรณยุกต์อื่น คือ วรรณยุกต์ โท ตรี และจัตวา เสียงพูดจะกลายเป็นคำว่า "ก้า" "ก๊า" "ก๋า" ตามลำดับ อย่างไรก็ตามเสียงที่ได้จะไม่เหมือนกับคำพูดตามวรรณยุกต์นั้นๆ แต่ให้ผลดีเพียงพอสำหรับสื่อความหมาย วิธีการนี้เป็นประโยชน์สำหรับการลดทอนข้อมูล ซึ่งหน่วยเสียงสระเดียวกันสามารถผลิตเสียงวรรณยุกต์ต่างๆ จนครบ

2. การนำข้อมูลพารามิเตอร์ของฟิลเตอร์มาผ่านกระบวนการลดทอนข้อมูลโดยอาศัยกระบวนการ Optimum Parameter Quantization เนื่องจากค่าของสัมประสิทธิ์พาร์คอร์มีการกระจายไม่สม่ำเสมอ เกิดจากความไวต่อการเปลี่ยนแปลงส่วนประกอบทางความถี่ของสัญญาณเสียงสังเคราะห์ หรือ Spectral Sensitivity แปรผันไม่เป็นเชิงเส้นกับค่าสัมประสิทธิ์ ดังนั้นการนำค่าสัมประสิทธิ์พาร์คอร์มาเข้ารหัสโดยมีการจัดสรรบิท (Bit Allocation) อย่างเหมาะสมจะสามารถลดทอนข้อมูลลงได้อีก กระบวนการดังกล่าวสามารถทำได้หลายวิธี ตัวอย่างเช่น วิธีของ Itakura และ Saito [21] สามารถลดทอนข้อมูลลงได้อีก 30% หรือ วิธีของ Gray และ Markel [22] แสดงให้เห็นว่า ระบบสังเคราะห์เสียงที่มีจำนวนออร์เดอร์เท่ากับ 10 และอัตราของเฟรมเท่ากับ 50 Hz เมื่อผ่านการทำ Optimum Parameter Quantization สามารถลดข้อมูลเสียงลงเหลือ 2,800 bits/S

3. นำระบบวิเคราะห์-สังเคราะห์ไปทำการวิเคราะห์เสียงพูดในลักษณะเป็นหน่วยเสียงของภาษาไทย ซึ่งสามารถพัฒนาเพื่อสร้างระบบ Text to Speech ในภาษาไทยได้โดยไม่จำกัดคำศัพย์

<u>ข้อเสนอแนะเกี่ยวกับแนวทางในการประยุกต์ใช้งาน</u>

1. สร้างหน่วยสังเคราะห์เสียงเอนกประสงค์สำหรับใช้งานกับอุปกรณ์ต่างๆ เช่น เครื่อง ATM ระบบตอบรับทางโทรศัพท์ ระบบเตือนภัยสำหรับศูนย์กลางควบคุมในโรงงานอุตสาหกรรม เครื่องมือวัด เป็นต้น ซึ่งถ้าระบบที่กล่าวมาเหล่านี้สามารถเปล่งเสียงเป็นคำพูดได้จะก่อประโยชน์เป็นอย่างมาก อุปกรณ์สังเคราะห์เสียงเอนกประสงค์สามารถพัฒนาขึ้นได้โดยใช้ ซีพียู TMS 32010 เพียงตัวเดียวร่วมกับหน่วยความจำสำหรับบันทึกคำศัพท์ที่ต้องการ

2. ส่งข่าวสารเป็นเสียงพูดแทรกไปตามระบบต่างๆ เช่น ระบบเรียกตัวอัตโนมัติ (Auto Paging)

3. ส่งข่าวสารทางเสียงโดยป้องกันการดักฟัง ใช้ในระบบโทรศัพท์หรือวิทยุสื่อสารโดยแทนที่จะส่งสัญญาณเสียงโดยตรงก็เปลี่ยนเป็นส่งข้อมูลเสียงที่ผ่านการวิเคราะห์ และที่ปลายทางจะมีระบบสังเคราะห์เสียงรับข้อมูลมาผลิตเสียง