บทที่ 6

บทสรุปและข้อเสนอแนะ

บทสรุป

จุดประสงค์ของวิทยานิพนธ์ฉบับนี้คือ การศึกษาการประยุกต์ใช้เวฟเลตร่วมกับการบีบย่อของกระบวนการบีบย่อสัญญาณเสียงที่ใช้กันอยู่เดิม คือวิธี CELP (Code-Excited Linear Prediction) โดยได้นำเสนอวิธีการใหม่ในการบีบย่อสัญญาณเสียงพูด โดยการแปลงเวฟเลตแพคเกตกับสัญญาณเศษเหลือ (residual signal) หลังจากนั้นจึงเลือกบีบย่อในชั้นของความละเอียดที่สามารถให้อัตราการบีบย่อสูงสุด ซึ่งจากผลการวิจัย พบว่าการบีบย่อสัญญาณเศษเหลือโดยใช้การแปลงเวฟเลตแพคเกตนี้ โดยเลือกค่า parameters ที่เหมาะสม สามารถเพิ่มอัตราการบีบย่อของวิธี CELP ได้ถึง 1.5 เท่า โดยที่ยังรักษาคุณภาพเสียงเดิมไว้ได้

สำหรับ parameters ที่เป็นเงื่อนไขในการแปลงนั้นพบว่า order ของ mother wavelet มีความสำคัญต่อคุณภาพเสียงที่ได้ โดยถ้าใช้ mother wavelet ที่มี order สูงขึ้นจะให้คุณภาพเสียงที่ดีขึ้น ในขณะที่ขนาดของ excitation codebook และ ค่า threshold ที่ใช้ในการพิจารณาตัดส่วนประกอบของสัญญาณที่มีความสำคัญน้อยนั้น มีความสำคัญทั้งกับ อัตราการบีบย่อ และ คุณภาพเสียงที่ได้ด้วย โดยถ้าใช้ codebook ขนาดใหญ่ขึ้น จะได้คุณภาพเสียงที่ดีขึ้นแต่ อัตราการบีบย่อจะลดลง ขณะเดียวกัน เมื่อเราเพิ่มค่า threshold ให้สูงขึ้น คุณภาพเสียงที่ได้จะต่ำลง ซึ่งเมื่อเลือกใช้ค่า threshold น้อย คือ 0.5 และ 2 คุณภาพเสียงที่ได้นั้นก็อยู่ในเกณฑ์ดี ที่อัตราการบีบย่อข้อมูลประมาณ 12.8 : 1 ถึง 9.5 : 1 แต่ถ้าเลือกใช้ค่า threshold มาก คือ 4 คุณภาพเสียงที่ได้จะต่ำลง แต่จะได้อัตราการบีบย่อที่สูงขึ้น คือประมาณ 7.6 : 1 ซึ่งหากไม่ได้นำไปใช้ในการสื่อสาร ก็ยังใช้การได้เนื่องจากยังสามารถสื่อความหมายได้อยู่ การจะเลือกใช้จึงต้องพิจารณาถึงความจำเป็น ทั้งในด้านการเก็บข้อมูล และคุณภาพเสียงที่ได้ด้วย

ปัญหาที่เกิดขึ้น

1) เสียงรบกวนที่เกิดขึ้นเนื่องจากการบีบย่อสัญญาณเสียง เกิดจาก quantization noise ในการทำ VQ สัญญาณกระตุ้น ซึ่งในวิทยานิพนธ์นี้จะแนะนำให้ใช้ excitation codebook ขนาด 32 หรือทำ VQ ที่ 1 bpn แต่ว่าจากวิธีอื่นๆที่มีผู้ทำขึ้น และใช้การทำ VQ เหมือนกัน และอ้างว่าสามารถทำได้โดยให้คุณภาพดีมาก ใช้การทำ VQ ที่อัตรา bpn ต่ำกว่านี้มาก คือการใช้ ACELP ทำ VQ ที่ 0.4 bits/sample (Galand, 1992) และ การใช้ VSELP ทำ VQ ที่ 0.25 bits/sample (Deller, 1993) ซึ่งก็อาจจะเป็นเพราะว่า วิธีการเหล่านี้ จะ update ค่าสัญญาณการกระตุ้นทุกๆ block ที่ได้ทำ VQ โดย update ทุกๆ 2.5 ms และ 5 ms ตามลำดับ ซึ่งวิธีที่ใช้ในวิทยานิพนธ์นี้ update ทุกๆ 10 ms ซึ่งก็ไม่สามารถลดเวลานี้ลงได้เนื่องจากข้อจำกัดของการแปลงเวฟเลตแพคเกตเอง

2) Delay time ที่ค่อนข้างยาวคือ 20 ms (คิดจากช่วงเวลาที่ต้องทำ buffer สัญญาณไว้ก่อนที่จะทำการคำนวณในแต่ละ frame) ซึ่งอาจจะลดลงได้เป็น 10 ms แต่ก็จะให้อัตราการบีบย่อข้อมูลที่ต่ำลง ซึ่งในทางปฏิบัติแล้ว 10 ms นี้ก็ยังมากเกินไป โดยถ้านำไปใช้ในระบบสื่อสารจริงจะต้องใช้อุปกรณ์พิเศษคือ echo canceller ในการช่วยตัดเสียงก้องที่เกิดจากความล่าช้าของสัญญาณ

3) สำหรับสัญญาณเสียงที่จะนำมาทำการบีบย่อ ถ้าเป็นเสียงพูดที่ดังมาก จนทำให้เกิดการ clip ส่วนยอดของสัญญาณ จะทำให้คุณภาพเสียงที่ได้จากการบีบย่อไม่ดี เนื่องจากสัญญาณช่วงที่ถูก clip ไปนั้นทำให้ผลตอบเชิงความถี่ของสัญญาณเปลี่ยนไปมาก ในขณะที่การบีบย่อที่ใช้ในวิทยานิพนธ์นี้ ใช้ LPC ซึ่งเป็นการวิเคราะห์ความถี่ของสัญญาณ ซึ่งจะรับผลกระทบจากการ clip นี้โดยตรง

ข้อเสนอแนะ

1) จากปัญหาในการทำวิทยานิพนธ์นี้ จะเห็นได้ว่าปัญหาที่สำคัญในแง่การนำไปใช้งานจริง (ปัญหาข้อที่ 1 และ ข้อที่ 2) อาจแก้ได้โดยการปรับปรุงวิธีการที่ใช้ในการแปลงเวฟเลตให้สามารถทำการแปลงแยกเป็นหลายๆ block โดยสามารถนำมาต่อกันได้ ซึ่งจะเป็นผลดีในการทำให้การทำ VQ สัญญาณกระตุ้นสามารถ update ข้อมูลได้บ่อยขึ้น ซึ่งจะทำให้คุณภาพของสัญญาณที่สร้างจากการทำ VQ ดีขึ้น และยังสามารถดัดแปลงให้ระบบมี delay time ที่ต่ำลงได้ด้วย

2) การทำ VQ กับ LPC parameters ที่ใช้ในวิทยานิพนธ์นี้ใช้การวัด distortion โดยเงื่อนไข squared error distortion ซึ่งอาจให้ผลที่ไม่ดีที่สุด โดยอาจใช้การวัด Itakura-Saito

distortion ที่ให้ผลดีกับการทำ VQ LPC parameters มากกว่า (Bristow, 1984; Gray, 1984; Hussain, 1991; Deller, 1993) แต่ในการนำไปใช้งานแล้ว นิยมใช้การทำ scalar quantization LPC parameters (Galand, 1992; ITU, 1992) โดยจะทำ scalar quantization กับ LPC parameters order แรกๆด้วยจำนวน bit มากกว่า order หลังๆ เนื่องจาก order แรกๆมีความสำคัญมากกว่า (Bristow, 1984) การทำเช่นนี้มีข้อเสียคือจะต้องใช้ จำนวน bit ในการเข้ารหัสในแต่ละ frame มากขึ้นกว่าเดิม

3) การสร้าง codebook อาจสร้างโดยการใช้การสุ่มแบบ stochastic ซึ่งสามารถลดเวลาในการสร้าง codebook ลงได้ (Deller, 1992; Galand, 1992) เนื่องจากสัญญาณกระตุ้นมีลักษณะที่เป็นสัญญาณสุ่ม (random signal) โดยจะเห็นได้ว่าการบีบย่อสัญญาณเสียงที่อยู่นอก training sequence ให้ผลไม่แตกต่างไปจากการบีบย่อสัญญาณเสียงใน training sequence

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย