บทที่ 5

การทดลองบีบย่อสัญญาณเสียงพูดด้วยเวฟเลต

อุปกรณ์และวิธีการเก็บข้อมูล

ในการทดลองบีบย่อสัญญาณเสียงพูดนี้จะใช้โปรแกรมที่พัฒนาขึ้น โดยใช้ตัวแปลภาษา Borland C++ version 3.0 ทำงานบนเครื่องคอมพิวเตอร์ IBM compatible ซึ่งใช้ไมโครโปรเซสเซอร์ 486DX2-66 สำหรับการบันทึกเสียง จะเก็บโดยใช้ A/D convertor ที่อยู่บน Sound Blaster Card ด้วย sampling rate 8000 Hz ขนาดข้อมูลตัวอย่างละ 8 bits โดยจะเก็บเสียงในรูปแบบของ Creative Voice File ซึ่งมีนามสกุลของไฟล์เป็น .VOC สำหรับการเล่นเสียงกลับ จะทำผ่าน D/A convertor ซึ่งอยู่ใน Sound Blaster Card เช่นกัน

เนื่องจากเสียงที่บันทึกมาได้ จะมีสัญญาณรบกวนจากการบันทึกเสียง ดังนั้นก่อนจะนำเสียงไปใช้จะนำเสียงที่บันทึกมาผ่าน band-pass filter ช่วง 300Hz-3200Hz ก่อน โดยใช้ filter แบบ IIR ที่ออกแบบโดยใช้โปรแกรม Digital Filter Design Version 1.4 ที่พัฒนาขึ้นโดย Digital Signal Processing Research Laboratory คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยมี specification ในการออกแบบดังนี้

Type - Elliptic

Band - Bandpass

Spec. - Sampling Frequency - 8000 Hz

Low Passband Frequency - 300 Hz

High Passband Frequency - 3200 Hz

Low Stopband Frequency - 200 Hz

High Stopband Frequency - 3400 Hz

Passband Ripple - 0.96

Stopband Ripple - 0.03

ซึ่ง parameter ของ filter ที่ได้เป็นดังนี้

sec1 : b[0] = 0.4662, b[1] = -0.0188, b[2] = -0.8832, b[3] = -0.0188, b[4] = 0.4662

a[0] = 1.0000, a[1] = -0.6630, a[2] = -0.4317, a[3] = 0.0386, a[4] = 0.2472

sec2 : b[0] = 0.8121, b[1] = -0.1379, b[2] = -1.2640, b[3] = -0.1379, b[4] = 0.8121

a[0] = 1.0000, a[1] = -0.3548, a[2] = -1.1086, a[3] = -0.2029, a[4] = 0.8370

โดยอยู่ในรูปแบบที่ sec 1 และ sec 2 ต่อกันแบบ cascade ซึ่งแต่ละ sec อยู่ในรูปดังนี้

$$L(z) = \frac{\sum_{i=0}^{4} b_i z^{-i}}{\sum_{i=1}^{4} a_i z^{-i}} \qquad (5\text{-}1)$$

สัญญาณที่เก็บได้นี้จะมีส่วนที่เงียบในช่วงต้นและท้าย เพื่อตัดให้เหลือแต่ส่วนที่เป็นสัญญาณเสียงพูด จะวัดพลังงานของสัญญาณในช่วงต้นและท้าย แล้วตัดส่วนที่ยังมีค่าไม่เกิน threshold

และเนื่องจากว่าการบีบย่อข้อมูลในวิทยานิพนธ์นี้จะแบ่งสัญญาณเป็น frame ขนาด 160 ตัวอย่าง ดังนั้นก่อนนำเสียงที่บันทึกได้ไปทำการทดลองจะทำการปรับขนาดให้เป็นจำนวนเท่าของ 160 ก่อน

<u>การออกแบบสร้าง codebook</u>

ในการบีบย่อสัญญาณเสียงพูดในวิทยานิพนธ์นี้ จะต้องทำการสร้าง VQ codebook ขึ้นจำนวน 2 codebook ซึ่งได้แก่ LPC codebook และ Excitation codebook ซึ่งจะทำการ train เพื่อสร้าง codebook ทั้งสองนี้ โดยใช้ตัวอย่างเสียงพูดที่เก็บจากผู้พูด 6 คน ซึ่งเป็น ชาย 3 คน และหญิง 3 คน โดยแต่ละคนจะพูดประโยคตัวอย่าง 12 ประโยค ดังต่อไปนี้

- มหาวิทยาลัยแห่งแรกคือจุฬา
- ชาวนาใช้เคียวเกี่ยวข้าว
- ผัดถั่วงอกอร่อยเหลือเกิน
- สมเสร็จเป็นสัตว์ป่า

- ระบบป้องกันไฟต้องดี
- ข้าศึกอยู่บนเกาะเยอะแยะ
- หุบลึกมักมีเสือ
- เนื้อหมูเน่าเฟะมีหนอน
- มีธุระด่วนกรุณาติดต่อกลับด้วย
- คำทดสอบมาตรฐานภาษาไทย
- กรุงเทพมหานครอมรรัตนโกสินทร์
- ปริญญาวิศวกรรมศาสตรมหาบัณฑิต

รวมได้ตัวอย่างเสียงความยาวประมาณ 2 นาที 20 วินาที (ตัดช่วงเสียงเงียบออกแล้ว) ตัวอย่างเสียงที่ได้นี้ จะถูกแบ่งเป็น block block ละ 160 ตัวอย่าง (20 ms) นำไปหาค่า LPC parameters โดยใช้ order เป็น 10 ซึ่ง parameters ที่ได้นี้จะนำไป train เพื่อสร้าง LPC Vector Quantization Codebook ขนาดต่างๆดังนี้คือ 64, 128, 256, 512 และ 1024 โดยใช้ Lloyd Algorithm ซึ่งมีขั้นตอนการทำดังนี้ (Gray, 1984)

1) กำหนดสัญญาณที่จะนำมาทำการ train เพื่อสร้าง codebook และกำหนด initial codebook
2) เข้ารหัสสัญญาณโดยใช้ codebook โดยให้มี distortion ต่ำสุด ซึ่งถ้า average distortion มีค่าน้อยเพียงพอ ให้เลิกทำ
3) แทน codebook เดิมโดยใช้ centroid ของ training vector ในแต่ละ codeword แล้วไปทำข้อที่ 1 ใหม่

จากตัวอย่างเสียงนี้ นำมาทำการเข้ารหัสโดยละเว้นการทำ block *Excitation CBM* เก็บสัญญาณการกระตุ้นในช่วงที่จะเข้า block *Excitation CBM* แล้วแบ่งเป็น block block ละ 5 ตัวอย่าง ทำ normalize ตามสมการ 4-8 และ 4-9 แล้วจึงนำมาใช้ train เพื่อสร้าง Excitation VQ codebook ขนาดต่างๆ ได้แก่ 32, 64, 128 และ 256 โดยใช้ Lloyd algorithm เช่นเดียวกับในขั้นตอนที่ 2

การวัดคุณภาพของสัญญาณเสียงพูด

การวัดคุณภาพของเสียงพูดที่ใช้ในวิทยานิพนธ์นี้ จะทำโดยการวัดคุณภาพของเสียงพูดทั้งแบบ subjective และแบบ objective โดยการวัดคุณภาพเสียงพูดแบบ sebjective ซึ่งจะให้ผู้ฟังเป็นคนตัดสิน โดยจะให้ผู้ฟังฟังเสียงต้นฉบับ และฟังเสียงที่ได้จากการเข้ารหัสและถอดรหัสออกมา แล้วให้ผู้ฟังให้คะแนนระดับความผิดเพี้ยนของสัญญาณ จะใช้ MOS (Mean Opinion Score) ซึ่งมีระดับการให้คะแนนดังในตาราง 5-1 (Deller et al., 1993) ส่วนการวัดคุณภาพเสียงพูดในแบบ objective ซึ่งจะวัดคุณภาพสัญญาณเสียงพูดโดยตรง จะใช้การวัด SNR โดยเปรียบเทียบกับเสียงพูดต้นฉบับ อนึ่ง วิธีการวัดคุณภาพเสียงพูดที่เป็นที่ยอมรับกันทั่วไปคือแบบ subjective ดังนั้นการวัดผลของคุณภาพเสียงพูด จะพิจารณาจากผลของ subjectie test เป็นหลัก

คะแนน	คุณภาพเสียง	ระดับของการผิดเพี้ยน
5	ดีมาก	ไม่มี
4	ดี	พอมี แต่ไม่น่ารำคาญ
3	พอใช้	มี และน่ารำคาญเล็กน้อย
2	แย่	น่ารำคาญ แต่ยังพอยอมรับได้
1	รับไม่ได้	น่ารำคาญมาก และยอมรับไม่ได้

ตาราง 5-1 แสดงระดับการให้คะแนนของ MOS (Mean Opinion Score)

การทดลองบีบย่อสัญญาณเสียงพูดภายใต้เงื่อนไขต่างๆ

เนื่องจากวิธีการบีบย่อสัญญาณเสียงพูดที่ได้เสนอในวิทยานิพนธ์นี้ มีเงื่อนไข parameters อยู่หลายอย่างที่มีผลต่ออัตราการบีบย่อ และคุณภาพเสียงที่ได้ ซึ่งได้แก่ ขนาดของ LPC VQ codebook, ขนาดของ Excitation codebook, order ของ wavelet ที่ใช้, threshold ที่ใช้ในการตัดข้อมูลที่มีความสำคัญน้อย โดยจะทำการทดลองโดยการสุ่มตัวอย่างเสียงพูด 10 ประโยคขึ้นมาจากตัวอย่างเสียงพูดที่ใช้ในการสร้าง codebook เพื่อนำมาทดลองบีบย่อและให้อาสาสมัครจำนวน 8 คน ซึ่งเป็นชาย 5 คน และ หญิง 3 คน ทดลองฟัง และให้คะแนนความผิดเพี้ยน(MOS) ดังต่อไปนี้

1. ทดลองการบีบย่อสัญญาณเสียงพูด โดยปรับเปลี่ยนขนาดของ LPC VQ codebook เป็น 32, 64, 128 และ 256 โดยใช้ขนาดของ Excitation codebook เป็น 64 mother wavelet เป็น Daubechies wavelet order 10 และใช้ค่า threshold 0.5 ได้ผลดังตารางที่ 5-2

ขนาดของ LPC codebook	32	64	128	256
ผู้ฟัง #1	4.7	4.6	4.7	4.5
ผู้ฟัง #2	4.4	4.5	4.4	4.4
ผู้ฟัง #3	4.6	4.4	4.5	4.5
ผู้ฟัง #4	4.6	4.5	4.6	4.7
ผู้ฟัง #5	4.5	4.6	4.4	4.4
ผู้ฟัง #6	4.4	4.5	4.4	4.4
ผู้ฟัง #7	4.6	4.5	4.7	4.5
ผู้ฟัง #8	4.4	4.6	4.5	4.4
เฉลี่ย	4.53	4.53	4.53	4.48
Bit-rate (kbps)	12.526	12.558	12.592	12.627
SNR (dB)	5.90	6.41	7.16	8.10

ตารางที่ 5-2 แสดงผลของการบีบย่อสัญญาณเสียงพูดโดยใช้ขนาดของ LPC codebook ต่างๆกัน

จากตารางที่ 5-2 จะเห็นได้ว่า การใช้ขนาด LPC codebook ที่เล็กลงไม่ทำให้คุณภาพเสียงที่ได้จากการบีบย่อลดลงไป แต่สามารถลด bit-rate ลงได้เล็กน้อย

2. ทดลองการบีบย่อสัญญาณเสียงพูด โดยปรับเปลี่ยนขนาดของ Excitation VQ codebook เป็น 32, 64, 128 และ 256 โดยใช้ขนาดของ LPC codebook เป็น 32 mother wavelet เป็น Daubechies wavelet order 10 และใช้ค่า threshold 0.5 ได้ผลดังตารางที่ 5-3

จากตารางที่ 5-3 จะเห็นได้ว่า การใช้ขนาด excitation codebook ที่เล็กลงทำให้คุณภาพเสียงที่ได้จากการบีบย่อลดลงไป แต่ก็สามารถลด bit-rate ลงได้ด้วย

3. ทดลองการบีบย่อสัญญาณเสียงพูด โดยปรับเปลี่ยน order ของ mother wavelet เป็น 4, 6, 8 และ 10 โดยใช้ขนาดของ LPC codebook เป็น 32 ขนาดของ excitation codebook เป็น 32 และใช้ค่า threshold 0.5 ได้ผลดังตารางที่ 5-4

ขนาด excitation codebook	32	64	128	256
ผู้ฟัง #1	4.7	4.7	4.8	4.9
ผู้ฟัง #2	4.4	4.4	4.7	4.9
ผู้ฟัง #3	4.5	4.6	4.8	4.8
ผู้ฟัง #4	4.5	4.6	4.8	5.0
ผู้ฟัง #5	4.6	4.5	4.7	4.8
ผู้ฟัง #6	4.5	4.4	4.9	5.0
ผู้ฟัง #7	4.4	4.6	4.7	4.8
ผู้ฟัง #8	4.5	4.4	4.8	4.9
เฉลี่ย	4.51	4.53	4.78	4.89
Bit-rate (kbps)	11.336	12.526	14.332	16.342
SNR (dB)	7.56	8.10	8.87	9.89

ตารางที่ 5-3 แสดงผลการบีบย่อสัญญาณเสียงโดยใช้ขนาดของ excitation codebook ต่างๆกัน

จากตารางที่ 5-4 จะเห็นได้ว่า การใช้ order ของ mother wavelet ที่ต่ำลงทำให้คุณภาพเสียงที่ได้จากการบีบย่อลดลงไป โดยที่ bit-rate มีค่าใกล้เคียงกัน

4. ทดลองการบีบย่อสัญญาณเสียงพูด โดยปรับเปลี่ยน ค่า threshold ที่ใช้ในการพิจารณาตัดองค์ประกอบของสัญญาณที่มีความสำคัญน้อยออก โดยให้ค่า threshold เป็น 0.5, 2 และ 4 โดยใช้ขนาดของ LPC codebook เป็น 32 ขนาดของ excitation codebook เป็น 32 และใช้ค่า mother wavelet order 10 ได้ผลดังตารางที่ 5-5

order ของ mother wavelet	4	6	8	10
ผู้ฟัง #1	4.5	4.6	4.6	4.7
ผู้ฟัง #2	4.2	4.3	4.3	4.4
ผู้ฟัง #3	4.4	4.3	4.5	4.5
ผู้ฟัง #4	4.3	4.3	4.4	4.5
ผู้ฟัง #5	4.2	4.3	4.5	4.6
ผู้ฟัง #6	4.3	4.4	4.5	4.5
ผู้ฟัง #7	4.4	4.3	4.5	4.4
ผู้ฟัง #8	4.2	4.3	4.4	4.5
เฉลี่ย	4.31	4.35	4.46	4.51
Bit-rate (kbps)	11.392	11.336	11.439	11.336
SNR (dB)	7.85	7.56	7.39	7.69

ตารางที่ 5-4 แสดงผลการบีบย่อสัญญาณเสียงโดยใช้ order ของ mother wavelet ต่างๆกัน

threshold	0.5	2	4
ผู้ฟัง #1	4.7	4.6	3.7
ผู้ฟัง #2	4.4	4.3	3.7
ผู้ฟัง #3	4.5	4.4	3.8
ผู้ฟัง #4	4.5	4.3	3.6
ผู้ฟัง #5	4.6	4.5	3.7
ผู้ฟัง #6	4.5	4.5	3.8
ผู้ฟัง #7	4.4	4.3	3.5
ผู้ฟัง #8	4.5	4.4	3.7
เฉลี่ย	4.51	4.41	3.69
Bit-rate (kbps)	11.336	9.338	7.614
SNR (dB)	8.10	7.56	5.45

ตารางที่ 5-5 แสดงผลการบีบย่อสัญญาณเสียงโดยใช้ค่า threshold ต่างๆกัน

จากตารางที่ 5-5 จะเห็นได้ว่า การใช้ค่า threshold ที่ต่ำลงทำให้คุณภาพเสียงที่ได้จากการบีบย่อลดลงไป โดยเฉพาะเมื่อใช้ค่า threshold เป็น 4 โดยที่ bit-rate ก็ลดลงไปมากเช่นเดียวกัน ดังนั้นการเลือกใช้ค่า threshold นี้ จะขึ้นอยู่กับวัตถุประสงค์ของการนำไปใช้งานว่า ต้องการคุณภาพเสียงที่ดี หรือ อัตราการบีบย่อที่สูง

จากผลการทดลองบีบย่อสัญญาณเสียงพูดจากตัวอย่างเสียงที่ใช้สร้าง codebook สรุปเงื่อนไขในการบีบย่อที่เหมาะสมได้ดังนี้คือ

ขนาดของ LPC VQ codebook - 32

ขนาดของ excitation codebook - 32

order ของ mother wavelet - 10

threshold - ขึ้นอยู่กับวัตถุประสงค์การนำไปใช้งาน โดยถ้าใช้ threshold เป็น 2 จะให้คุณภาพเสียงที่ค่อนข้างดี ที่อัตราการบีบย่อที่ค่อนข้างสูงเช่นเดียวกัน

<u>การทดลองบีบย่อสัญญาณเสียงพูดทั่วไป</u>

เนื่องจาก LPC VQ codebook และ excitation VQ codebook ถูกสร้างขึ้นโดยตัวอย่างเสียงพูดจำนวนหนึ่งเท่านั้น การบีบย่อสัญญาณเสียงทั่วไปนั้นอาจจะให้ผลที่แตกต่างกันไปก็ได้ ซึ่งจะทำการทดลองบีบย่อสัญญาณเสียงพูดที่อยู่นอก training sequence โดยจะบันทึกสัญญาณเสียงพูด 10 ชุด แล้วทดลองบีบย่อสัญญาณเสียงพูดโดยใช้ ขนาด LPC codebook เป็น 32 ขนาด excitation codebook เป็น 32 mother wavelet order 10 และค่า threshold เป็น 2 แล้วให้อาสาสมัครทดลองฟังและให้คะแนนความผิดเพี้ยนจากต้นฉบับ (MOS) แล้วเปรียบเทียบกับการบีบย่อสัญญาณเสียงพูดใน training sequence ที่บีบย่อโดยใช้เงื่อนไขการบีบย่อเหมือนกัน ซึ่งให้ผลดังตารางที่ 5-6

จากตารางที่ 5-6 จะเห็นได้ว่า การบีบย่อสัญญาณเสียงพูดที่อยู่ในและนอก training sequence ให้ผลที่อาจจะกล่าวได้ว่า ไม่แตกต่างกันดังนั้น ระบบการบีบย่อข้อมูลนี้สามารถใช้ได้กับสัญญาณเสียงพูดทั่วไป

	เสียงพูดใน training sequence	เสียงพูดนอก training sequence
ผู้ฟัง #1	4.6	4.6
ผู้ฟัง #2	4.3	4.4
ผู้ฟัง #3	4.4	4.3
ผู้ฟัง #4	4.3	4.4
ผู้ฟัง #5	4.5	4.3
ผู้ฟัง #6	4.5	4.4
ผู้ฟัง #7	4.3	4.3
ผู้ฟัง #8	4.4	4.5
เฉลี่ย	4.41	4.40
Bit-rate (kbps)	9.338	9.437
SNR (dB)	7.56	7.61

ตารางที่ 5-6 แสดงผลการบีบย่อสัญญาณเสียงใน training sequence และนอก training sequence

<u>การเปรียบเทียบอัตราการบีบย่อเมื่อใช้การแปลงเวฟเลตแพคเกตและไม่ใช้</u>

สำหรับ bit-rate ของระบบบีบย่อเสียงพูดโดย CELP อย่างเดียว ไม่ใช้การแปลงเวฟเลตแพคเกตที่ใช้นั้น สามารถคำนวณได้ดังนี้

สำหรับเสียงพูด 1 frame (20 msec)

LPC parameters VQ index	A	bits
Long-term predictive parameters		
pitch period	2x7	bits
pitch gain	2x3	bits
excitation gain	32x3	bits
excitation VQ index	32xB	bits
total	116+A+32xB	bits

โดยค่า A ที่ใช้มีค่าอยู่ในช่วง 5 ถึง 8 และ B ที่ใช้มีค่าอยู่ในช่วง 5 ถึง 8 เช่นกัน ดังนั้น จะได้ bit-rate ระหว่าง 14,050 bps และ 19,000 bps โดยที่เงื่อนไขการบีบย่อที่เหมาะสม (ซึ่งกล่าวถึงไว้ก่อนหน้านี้) ได้ bit-rate ที่ 14,050 bps

ซึ่งอาจกล่าวได้ว่า ที่เงื่อนไขที่เหมาะสมนี้ การใช้การแปลงเวฟเลตแพคเกตสามารถเพิ่มอัตราการบีบย่อเสียงพูดของวิธี CELP ได้ถึง 14,050/9,400 = 1.5 เท่า