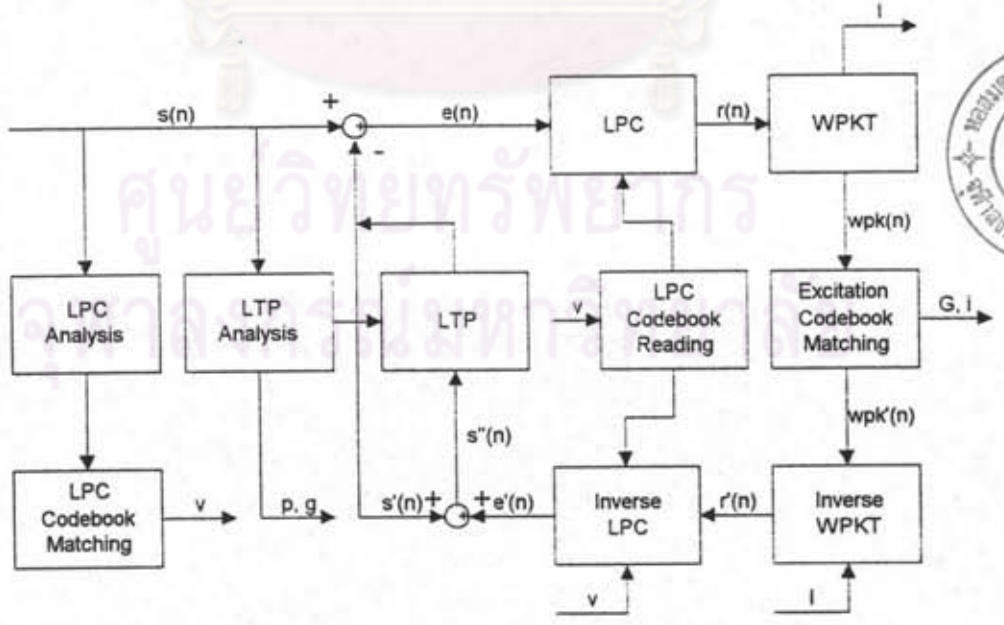
บทที่ 4

การบีบย่อสัญญาณเสียงพูดด้วยเวฟเลต

การบีบย่อสัญญาณเสียงพูดโดยใช้เวฟเลต

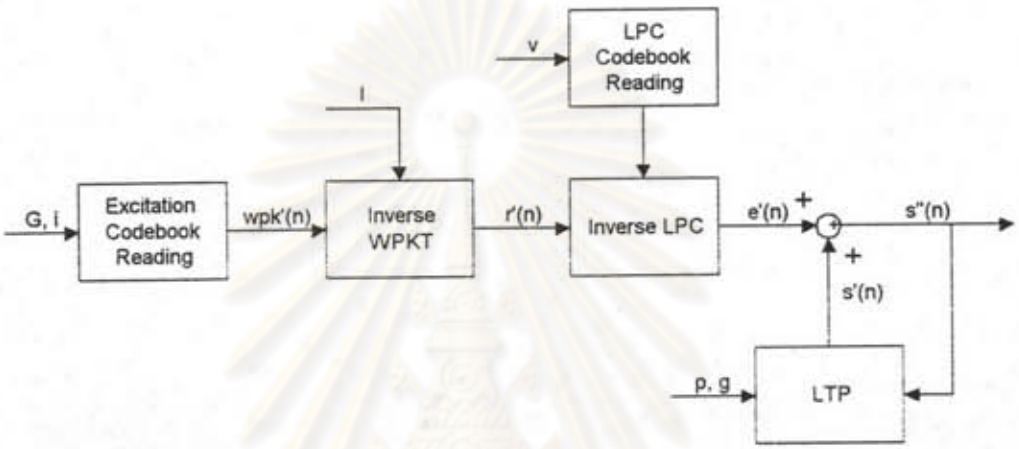
การบีบย่อสัญญาณเสียงพูดโดยใช้เวฟเลตนั้นได้มีผู้ทดลองทำไว้หลายวิธี ในวิทยานิพนธ์นี้จะเสนอแนวความคิดใหม่ในการบีบย่อสัญญาณเสียงพูดโดยใช้เวฟเลตร่วมกับวิธี CELP โดยจะทำการแปลงเวฟเลตกับสัญญาณเศษเหลือ (residual signal) ในการบีบย่อสัญญาณเสียงพูดโดยวิธี CELP แล้วเลือก level ที่มีจำนวน block ที่มีค่าน้อยกว่าค่า threshold มาก block ที่สุด เพื่อให้ได้อัตราการบีบย่อที่สูงที่สุด block diagram ของตัวเข้ารหัส (encoder) และตัวถอดรหัส (decoder) ได้ดัดแปลงมาจาก block diagram ของการบีบย่อสัญญาณเสียงพูดโดยวิธี CELP ในรูป 3-1 โดยตัดบาง block ออกเพื่อลดความซับซ้อนของกระบวนการ และเพิ่ม block ที่ใช้ในการทำ Wavelet Packet Transform และ Inverse Wavelet Packet Transform ดังแสดงในรูปที่ 4-1 และรูปที่ 4-2



รูปที่ 4-1 แสดง block diagram ของตัวเข้ารหัส (encoder)

การทำงานของ block ต่างๆ ของตัวเข้ารหัส สามารถอธิบายได้ดังต่อไปนี้

จากรูปที่ 4-1 สัญญาณเสียงพูด s(n) จะถูกทำ windowing เพื่อแบ่งเป็น frame โดยที่แต่ละ frame มีความกว้าง 20 ms และมี 160 ตัวอย่าง ดังนี้



รูปที่ 4-2 แสดง block diagram ของตัวถอดรหัส (decoder)

$$x(n) = \begin{cases} w(n)s(n), & 0 \le n < 160 \\ 0, & otherwise \end{cases} \qquad (4\text{-}1)$$

ที่ block *LPC Analysis* เพื่อลดผลของขอบ frame (Bristow, 1984) ในการทำ LPC + Analysis จะใช้ Hamming Window ดังนี้ (Oppenhiem, 1989)

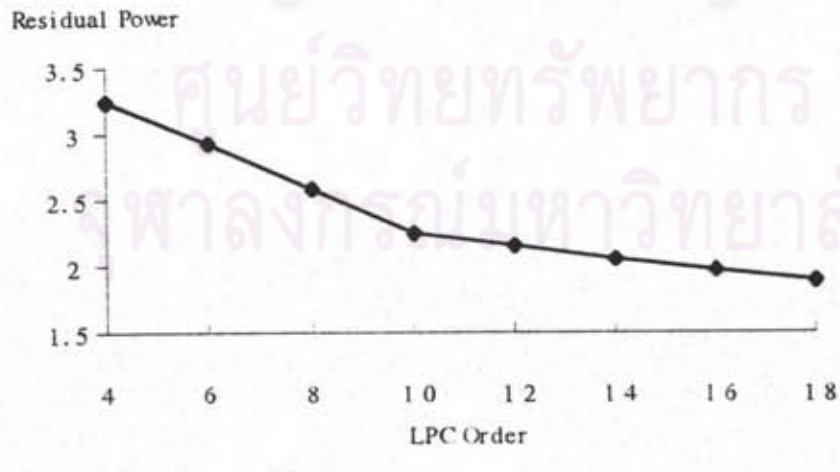
$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n / 159), & 0 \le n < 160, \\ 0, & otherwise \end{cases} \qquad (4\text{-}2)$$

หลังจากนั้นสัญญาณ x(n) จะถูกใช้ในการวิเคราะห์หาค่า LPC parameters โดย Levinson-Durbin recursive method ซึ่งแสดงไว้ในบทที่ 3 ซึ่งจากการทดลองทำ LPC กับสัญญาณตัวอย่าง โดยใช้ LPC order ตั้งแต่ 4 ถึง 18 โดยเพิ่มค่าขั้นละ 2 แล้ววัดกำลังเฉลี่ยของสัญญาณเศษเหลือ (average residual power) ได้ดังตารางที่ 4-1

LPC Order	Residual Power
4	3.24
6	2.93
8	2.58
10	2.24
12	2.15
14	2.05
16	1.97
18	1.89

ตารางที่ 4-1 แสดงความสัมพันธ์ระหว่าง
LPC Order และ Residual Power

นำเอาข้อมูลในตารางที่ 4-1 มาเขียนกราฟจะได้ดังรูปที่ 4-3 ซึ่งจะเห็นได้ว่า residual power เริ่มลดลงน้อยที่ LPC Order เป็น 10 ในขณะที่ใช้ในการคำนวณหา LPC parameters ที่ order ต่างๆนั้น ใกล้เคียงกัน เนื่องจากเวลาส่วนมากถูกใช้ในการคำนวณหา auto-correlation ดังนั้นจะเลือกใช้ LPC Order เป็น 10 ในการทำ LPC ในวิทยานิพนธ์นี้



รูปที่ 4-3 กราฟแสดงความสัมพันธ์ระหว่าง LPC Order และ Residual Power

LPC parameters ที่ได้มาจะถูกนำไปทำ Vector Quantization ที่ block *LPC CBM* (LPC CodeBook Matching) โดยเปรียบเทียบ LPC parameters ขาเข้า กับ vector ใน codebook โดยพิจารณาจากเงื่อนไขของ squared error distortion ดังนี้ (Gray, 1984)

$$d(x,\hat{x}) = \|x-\hat{x}\|^2 = \sum_{i=0}^{k-1}(x_i-\hat{x}_i)^2 \qquad (4\text{-}3)$$

โดยที่ k คือ dimension ของ vector

หมายเลขของ vector ใน codebook ที่มี distortion น้อยที่สุดเมื่อเทียบกับ LPC parameters ขาเข้า (v) จะถูกส่งออกเป็น output ของตัวเข้ารหัสนี้ โดยที่ถ้า หมายเลข vector เป็น 0 จะหมายความว่า สัญญาณเสียง frame นี้ เป็นเสียงเงียบ ซึ่งจะงดเว้นการทำ block อื่นๆ ใน frame นี้

สัญญาณ s(n) นี้จะถูกนำไปวิเคราะห์หา long-term predictive parameters โดย block *LTP Analysis* (Long-Term Predictive Analysis) โดยจะทำการวิเคราะห์การซ้ำกันเป็นคาบของสัญญาณเสียงพูด s(n) ทุกๆ frame ย่อย ขนาด 5 ms (40 ตัวอย่าง) โดยที่ LTP parametere นี้ จะประกอบด้วย pitch period (p) คือคาบของการซ้ำของรูปคลื่นสัญญาณซ้ำกัน และ pitch gain (g) ซึ่งก็คืออัตราส่วนของขนาดสัญญาณใน frame ย่อยปัจจุบัน ต่อขนาดของสัญญาณใน frame ย่อยก่อนหน้า ซึ่งสามารถคำนวณหาค่า p และ g ได้ดังนี้ (Hussain et al., 1991)

$$p = \arg_{m=p+1}\max\{r(m)\} \qquad (4\text{-}4)$$

$$\text{โดยที่ } r(m) = \sum_{n=0}^{79} x(n)x(n+m) \qquad (4\text{-}5)$$

$$\text{และ } g = r(p)/r(0) \qquad (4\text{-}6)$$

p และ g นี้ จะเป็น output ของ block โดย p จะเป็น output ขนาด 7 bits โดยมีค่าช่วง 0-127 แทนค่าช่วง pitch period เป็น 10-137 และ g จะเป็น output ขนาด 3 bits โดยค่า g นี้จะถูกทำ scalar quantization ดังตาราง 4-2 โดยถ้า pitch period มีค่าอยู่นอกช่วงดังกล่าว หรือ g มีค่าน้อยกว่า 0.6 หรือ มากกว่า 1.3 ซึ่งจะตีความว่า g มีค่าเป็น 0 จะถือว่าสัญญาณเสียงพูดใน frame

ย่อยนี้ไม่มีความเกี่ยวพันกับ สัญญาณเสียงพูดในช่วงเวลาก่อนๆ ค่า g เท่ากับ 0 และจะไม่ส่ง p เป็น output

Index	g
0	0
1	0.7
2	0.8
3	0.9
4	1.0
5	1.1
6	1.2
7	1.3

ตารางที่ 4-2 แสดงการทำ scalar quantization ของ pitch gain (g)

จากนั้น สัญญาณเสียงพูด s(n) จะถูกนำไปเปรียบเทียบกับสัญญาณเสียงพูดที่ได้จากการทำนาย $s'(n)$ ผลต่างระหว่าง s(n) และ $s'(n)$ นี้จะเรียกว่า error signal e(n) ซึ่งจะถูกนำไปผ่าน linear perdictive filter ที่ block *LPC* เพื่อสร้างสัญญาณเศษเหลือ (residual) r(n) ดังสมการต่อไปนี้ (Bristow, 1984)

$$r(n) = e(n) + \sum_{k=1}^{p} a(k)e(n-k) \qquad (4\text{-}7)$$

สำหรับ LPC parameters นี้ได้จาก block *LPC CBR* (LPC Codebook Reading) ซึ่งจะนำค่า vector index (v) จาก block *LPC CBM* มาเปิด codebook

สัญญาณเศษเหลือ r(n) นี้จะถูกแบ่งเป็น frame ขนาด 10 ms (80 ตัวอย่าง) เพื่อนำไปทำ Wavelet Packet Transform โดย block *WPKT* (Wavelet Packet Transform) โดยใช้ Daubechies wavelet (Daubechies, 1992) (parameters ของ Daubechies wavelet แสดงอยู่ในภาคผนวก) โดยจะทำ wavelet packet transform 3 ขั้น แล้วแบ่งข้อมูลเป็น block block ละ 5 ตัวอย่าง แล้วจึงเลือก level ที่มีจำนวน block ที่มีขนาดของแต่ละตัวอย่างไม่เกินค่า threshold น้อยที่สุด level (l) นี้จะเป็น output ขนาด 2 bits โดยที่ถ้า l เป็น 0 หมายถึงเลือก error signal r(n) ที่ไม่ได้

ทำ wavelet packet transform เป็น ผลลัพท์ นอกจากนี้แล้วถือเอา ผลลัพท์ของการทำ wavelet packet transform level 1 เป็น wpk(n) ทั้งนี้ในบทที่ 2 ได้กล่าวถึงวิธีการทำ wavelet packet transform แล้ว

สัญญาณ wpk(n) นี้จะถูกทำ VQ โดย block *Excitation CBM* (Excitation Codebook Matching) จะแบ่งสัญญาณ wpk(n) ออกเป็น block โดยมีขนาด block เท่ากับขนาด block ที่ใช้ในการเลือก best level ใน block *WPKT* แล้วทำ vector quantization กับสัญญาณแต่ละ block โดย ก่อนที่จะทำ VQ จะทำการ normalize สัญญาณแต่ละ block ก่อน โดย

$$G = \sqrt{\sum_{l=1}^{5} wpk^2(l)} \tag{4-8}$$

$$wpk''(n) = \frac{wpk(n)}{G}, \qquad 1 \le n \le 5 \tag{4-9}$$

$wpk''(n)$ จะถูกทำ VQ โดย excitation codebook โดยการวัดหา vector ใน codebook ที่ให้ distorttion น้อยที่สุด โดยเงื่อนไข squared error distortion ดังสมการ 4-3 โดยหมายเลขของ vector ใน codebook นี้ (i) จะเป็น output ของตัวเข้ารหัส ส่วน Gain (G) จะถูกทำ scalar quantization ขนาด 3 bits ดังตารางที่ 4-3

Index	G
0	0
1	2.0
2	3.2
3	5.12
4	8.19
5	13.11
6	20.97
7	33.55

ตารางที่ 4-3 แสดงการทำ scalar quantization ของ gain (G)

$$(G = 1.25 \cdot 1.6^{index}, \quad index \neq 0)$$

ถ้าไม่มีสมาชิกใดใน block หนึ่ง ที่มีค่าเกิน threshold เลย จะให้ G เท่ากับ 0 ซึ่งถ้า output G (index) มีค่าเป็น 0 แล้วจะไม่ส่ง หมายเลข vector (i) เป็น output โดย ค่าใน vector (i) นี้จะถูกนำไปคูณกับค่า Gain (G) เพื่อส่งออกเป็นสัญญาณ wpk"(n) ด้วย

สัญญาณ wpk"(n) นี้จะถูกนำไปทำ inverse wavelet packet transform ทุกๆ 10 ms (20 ตัวอย่าง) โดย block *Inverse WPKT* (Inverse Wavelet Packet Transform) จำนวน 1 level (โดย 1 คือ best level ที่ได้จาก block *WPKT*) ทั้งนี้ในบทที่ 2 ได้กล่าวถึงวิธีการทำ Inverse wavelet packet transform ไว้แล้วเช่นกัน ผลลัพท์ที่ได้คือสัญญาณเศษเหลือ r'(n)

สัญญาณเศษเหลือ r'(n) นี้ จะถูกนำไปทำ inverse linear predictive coding ที่ block *Inverse LPC* (Inverse Linear Predictive Coding) โดยใช้ LPC parmeters ที่ได้จาก block *LPC CBR* โดยจะทำทุกๆ 10 ms (80 ตัวอย่าง) จะได้ผลลัพท์จะเป็น error signal e'(n) โดยสมการการสร้าง e'(n) เป็นดังนี้ (Bristow, 1984)

$$e'(n) = -\sum_{k=1}^{p} a(k)e'(n-k) + r'(n) \qquad (4\text{-}10)$$

error signal e'(n) นี้จะถูกนำไปรวมกับสัญญาณเสียงพูดที่ได้จากการทำนาย s'(n) ได้เป็นสัญญาณเสียงพูดที่ถูกสร้างขึ้นใหม่ s"(n) เพื่อนำไปทำนายสัญญาณเสียงพูดใน frame ต่อไป โดยผ่าน block *LTP* (Long-Term Prediction) ซึ่ง block นี้จะทำ long-term prediction กับ s"(n) ดังสมการต่อไปนี้ (Deller et al., 1993)

$$s'(n) = g \cdot s''(n-p) \qquad (4\text{-}11)$$

โดย p และ g คือ pitch period และ pitch gain จาก block *LTP Analysis*

สำหรับตัวถอดรหัส จาก block diagram ในรูปที่ 4-1 และ 4-2 จะสังเกตเห็น ตัวถอดรหัสนั้นเป็นส่วนหนึ่งของตัวเข้ารหัส โดย block *Excitation CBR* (Excitation Codebook Reading) จะรับค่า input G และ i แล้วแปลงเป็นสัญญาณ wpk'(n) โดยจะอ่านค่า vector หมายเลข i จาก codebook แล้วคูณแต่ละตัวด้วย gain (G) ซึ่งอ่านได้จากตาราง 4-2 อนึ่งถ้าค่า G ที่อ่านมาได้เป็น 0 หมายถึงให้ wpk'(n) ของ block นั้น เป็น 0 ทุกตัว

โดยในรูปที่ 4-4 แนวตั้งคือขนาดของสัญญาณ สัญญลักษณ์ x4 ที่ตามหลังชื่อสัญญาณ แสดงว่า สัญญาณนั้นถูกแสดงด้วยขนาดเป็น 4 เท่าของสัญญาณอื่น