บทที่ 3

การเข้ารหัสเสียงตามมาตรฐาน ITU-T G.729 CS-ACELP

3.1 หลักการโดยรวมของตัวเข้ารหัส CS-ACELP

การเข้ารหัสเสียงพูดโดยวิธี CS-ACELP เป็นมาตรฐานล่าสุดของ ITU ถูกออกแบบมาเพื่อประมวลผลสัญญาณดิจิตอลที่ได้จากการกรองสัญญาณในช่วงแบนด์วิดท์ (Band width) ของระบบโทรศัพท์มีสัญญาณอินพุตเป็นสัญญาณแอนะลอก (ITU-T G.712) ที่อัตราการสุ่ม 8000 เฮิรตซ์ แล้วเปลี่ยนเป็นรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิต แล้วจึงป้อนให้กับอินพุตของตัวเข้ารหัสนี้ สัญญาณเอาต์พุตของตัวถอดรหัสจะถูกเปลี่ยนกลับมาเป็นสัญญาณในลักษณะตรงข้าม คุณสมบัติอื่นๆ ของขาเข้าและขาออกนั้นจะเป็นไปตามข้อกำหนดในมาตรฐาน ITU-T G.711 ของรหัส PCM ที่อัตรา 64 kbits/s ซึ่งจะถูกเปลี่ยนเป็นรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิตก่อนการเข้ารหัส และจากรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิต ก็จะถูกเปลี่ยนเป็นรหัสที่เหมาะสมต่อไป หลังจากการถอดรหัสแล้วซึ่งกระแสบิตที่ได้จากการเข้ารหัสที่จะส่งไปถอดรหัสนั้นจะถูกกำหนดให้เป็นไปตามมาตรฐานอัตราการเข้ารหัสอยู่ที่ 8 kbits/s

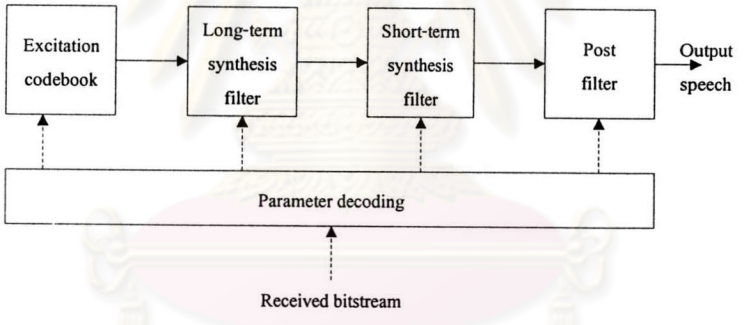
ตัวเข้ารหัสสัญญาณเสียงแบบ CS-ACELP ใช้หลักการเข้ารหัสสัญญาณเสียงแบบ CELP (Code Excited Linear Predictive) โดยตัวเข้ารหัสจะแบ่งสัญญาณเสียงออกเป็นเฟรมๆ ละ 10 มิลลิวินาที ในแต่ละเฟรมจะมีตัวอย่างที่ถูกสุ่ม 80 ตัวอย่าง โดยอัตราการสุ่มจะเท่ากับ 8000 ตัวอย่างต่อวินาที สัญญาณเสียงในทุกๆ เฟรมจะถูกวิเคราะห์เพื่อหาพารามิเตอร์ต่างๆ ตามหลักการของ CELP (ได้แก่ สัมประสิทธิ์ของวงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียง adaptive-codebook fixed-codebook และอัตราขยาย) พารามิเตอร์เหล่านี้จะถูกนำไปเก็บไว้ในหน่วยความจำถ้าในกรณีสำหรับเครื่องบันทึกเสียงหรือถ้าในกรณีของการสื่อสาร พารามิเตอร์เหล่านี้จะถูกเข้ารหัสแล้วส่งไปยังปลายทาง รายละเอียดของบิตต่างๆ ที่ได้จากการเข้ารหัสแสดงในตารางที่ 3.1 ที่ตัวถอดรหัสพารามิเตอร์เหล่านี้จะถูกใช้ในการสร้างเอ็กไซเทชัน และสัมประสิทธิ์ของวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์สัญญาณเสียง สัญญาณเสียงจะถูกสังเคราะห์โดยการกรองสัญญาณเอ็กไซเทชันผ่านวงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียงแบบ Short- term ดังรายละเอียดแสดงในรูปที่ 3.1

วงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียงแบบ Short-term นั้นจะเป็นวงจรกรองสัญญาณที่ใช้ทำนายสัญญาณแบบเชิงเส้น (Linear Prediction:LP) ที่มีอันดับเท่ากับ 10 สำหรับ

วงจรกรองสัญญาณแบบ Long-term ที่ใช้ในการสังเคราะห์พิตช์ (Pitch) จะถูกสร้างโดย Adaptive-codebook สัญญาณเสียงที่ได้จากการสังเคราะห์จะถูกปรับแต่งโดยวงจรกรอง Postfilter

ตารางที่ 3.1 การกำหนดบิตการเข้ารหัสแบบ CS-ACELP ตามมาตรฐาน G.729

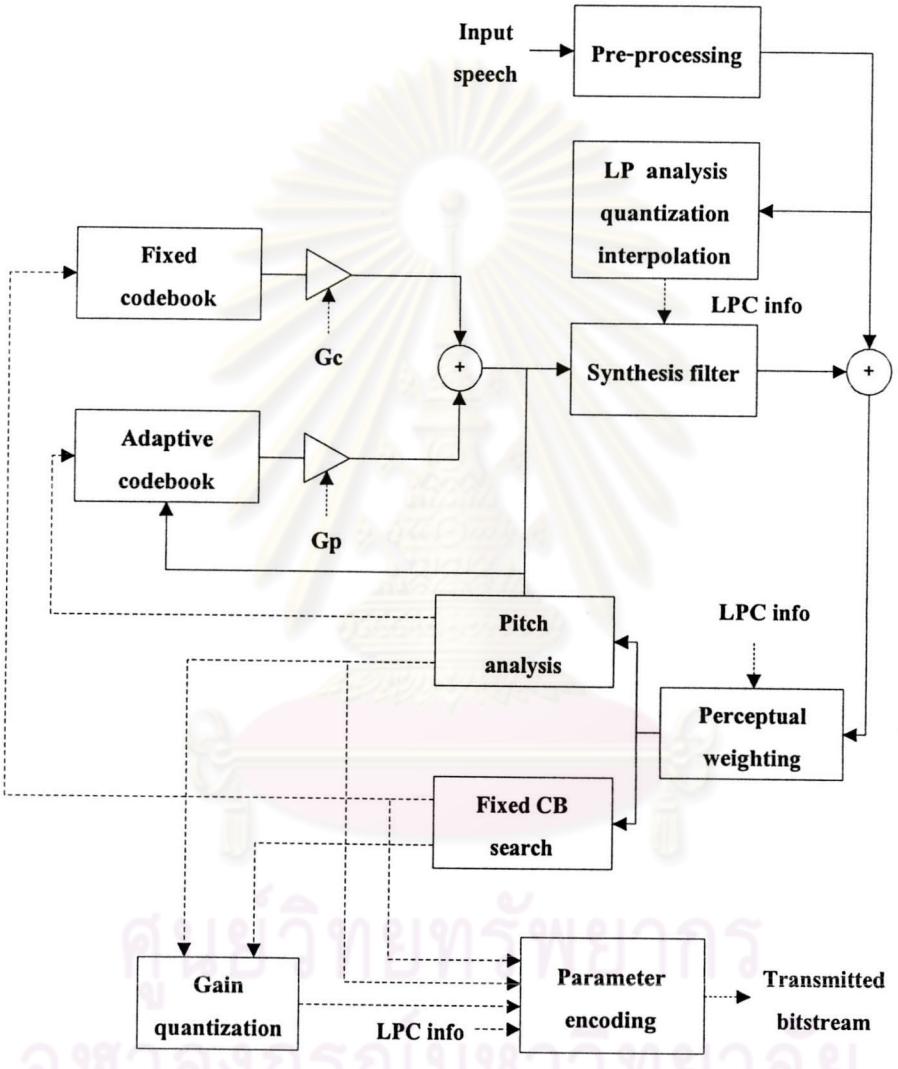
พารามิเตอร์	คำย่อ	เฟรมย่อย 1	เฟรมย่อย 2	บิตต่อเฟรม
Line spectrum pairs	L0, L1, L2, L3			18
Adaptive-codebook delay	P1,P2	8	5	13
Pitch-delay parity	P0	1		1
Fixed-codebook index	C1, C2	13	13	26
Fixed-codebook sign	S1, S2	4	4	8
Codebook gains(stage 1)	GA1, GA2	3	3	6
Codebook gains(stage 2)	GB1, GB2	4	4	8
Total				80



รูปที่ 3.1 บล็อกไดอะแกรมการทำงานของการสังเคราะห์สัญญาณเสียงด้วยวิธี CELP

หลักการของการเข้ารหัสแสดงดังรูปที่ 3.2 เริ่มจากสัญญาณอินพุตจะถูกกรองแบบความถี่ผ่านสูงและถูกปรับลดขนาดโดย pre-processing สัญญาณที่ได้จาก pre-processing นี้จะเป็นสัญญาณอินพุตให้กับทุกๆ วงจรที่เหลือเพื่อใช้ในการวิเคราะห์ต่อไป การวิเคราะห์ LP จะแบ่งสัญญาณเสียงเป็นเฟรมๆ ละ 10 มิลลิวินาที หรือ 80 ตัวอย่าง เพื่อใช้ในการคำนวณหาสัมประสิทธิ์ของวงจรกรองสัญญาณ LP สัมประสิทธิ์เหล่านี้จะถูกแปลงเป็นค่า line spectrum pairs (LSP) และทำการควอนไทซ์โดยใช้วิธีการเวกเตอร์ควอนไทซ์แบบ 2 Stage ที่มีความละเอียด 18 บิต สัญญาณเอ็กไซเทชัน จะถูกคำนวณหาจากการใช้วิธีการ analysis-by-synthesis ในการค้นหาเพื่อให้ค่าความผิดพลาดระหว่างสัญญาณเสียงดั้งเดิมกับสัญญาณเสียงที่สังเคราะห์ขึ้นมาได้มีค่า

น้อยที่สุดโดยใช้วิธีการวัดด้วย perceptual weighted distortion โดยใช้การกรองสัญญาณความผิดพลาดด้วยวงจรกรองสัญญาณแบบ perceptual weighting ที่มีสัมประสิทธิ์ที่ได้มาจากสัมประสิทธิ์ของวงจรกรองสัญญาณ LP ที่ยังไม่ควอนไทซ์ จำนวนของการทำ perceptual weighting นั้นสามารถปรับได้เพื่อให้สมรรถนะการทำงานที่ได้มีผลตอบสนองต่อสัญญาณเข้าเท่ากันทุกๆ ความถี่ (flat frequency response)



รูปที่ 3.2 หลักการเข้ารหัสของตัวเข้ารหัสแบบ CS-ACELP

พารามิเตอร์ของเอ็กไซเทชัน (พารามิเตอร์ของ fixed-codebook และ adaptive-codebook) จะถูกคำนวณหาทุกๆ เฟรมย่อย โดยแต่ละเฟรมย่อยนี้จะมีขนาด 5 มิลลิวินาที

(40 ตัวอย่างสัญญาณที่อัตราสุ่ม 8000 เฮิรตซ์) สัมประสิทธิ์ของวงจรกรองสัญญาณ LP ทั้งที่ควอนไทซ์และไม่ควอนไทซ์จะถูกใช้ในเฟรมย่อยที่ 2 ส่วนในเฟรมย่อยที่ 1 นั้นจะใช้วิธีการหาสัมประสิทธิ์ของวงจรกรองสัญญาณ LP โดยการประมาณค่าในช่วง (ทั้งที่ควอนไทซ์และไม่ควอนไทซ์)

สำหรับ open-loop pitch delay จะถูกประมาณหนึ่งครั้งในแต่ละเฟรมจากสัญญาณที่ได้จากวงจรกรองสัญญาณ perceptual weighting สำหรับกระบวนการที่จะกล่าวต่อไปนี้จะกระทำทุกๆ เฟรมย่อย สัญญาณเป้าหมาย $x(n)$ จะถูกคำนวณโดยการกรองสัญญาณ residual ของ LP ผ่านวงจรกรองสัญญาณ weighted synthesis $W(z)/A(z)$ โดยเมื่อเริ่มการทำงานวงจรกรองสัญญาณเหล่านี้จะถูกปรับให้ทันกาลจากการกรองค่าความผิดพลาดระหว่าง LP residual กับเอ็กไซเทชัน นั่นหมายถึงการหักล้างด้วยวงจรกรองสัญญาณแบบ weighted synthesis ที่ zero-input response ออกจากสัมประสิทธิ์ของสัญญาณเสียง การตอบสนองอิมพัลส์ $h(n)$ ของวงจร กรองสัญญาณ weighted synthesis จะถูกคำนวณหา ส่วนการวิเคราะห์ closed-loop pitch (เพื่อหาค่าประวิงเวลาและอัตราขยายของ adaptive-codebook) จะใช้สัญญาณเป้าหมาย $x(n)$ และการตอบสนองอิมพัลส์ $h(n)$ โดยการค้นหารอบค่า open-loop pitch delay ทั้งค่าที่เป็นจำนวนเต็มและค่าเศษส่วนด้วยความละเอียด 1/3 ค่า pitch delay นี้จะถูกเข้ารหัสด้วยความละเอียด 8 บิต ในเฟรมย่อยแรก ส่วนที่เหลือจะเข้ารหัสด้วยความละเอียด 5 บิต ในเฟรมย่อยที่สอง สัญญาณของเป้าหมาย $x(n)$ จะถูกปรับค่าโดยการนำค่าจาก adaptive codebook มาหักล้างและค่าสัญญาณของเป้าหมายที่ได้ใหม่ $x'(n)$ นี้จะถูกใช้ในการค้นหา fixed-codebook เพื่อให้ได้สัญญาณการกระตุ้นที่ถูกต้องที่สุดต่อไป

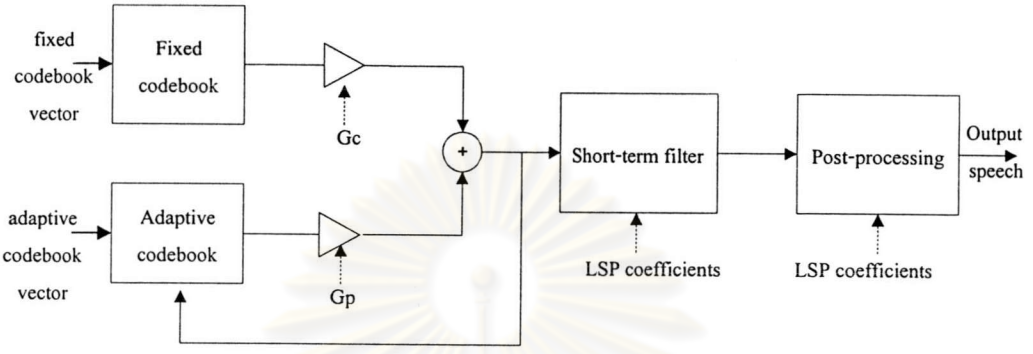
ค่าโครงสร้างของ fixed-codebook ที่นี้จะเป็นแบบพีชคณิต (Algebraic) ที่มีความละเอียด 17 บิต อัตราขยายของ adaptive-codebook และ fixed-codebook จะถูกควอนไทซ์ที่ความละเอียด 7 บิต (โดยจะใช้ตัวทำนายสัญญาณแบบ Moving average (MA) กับอัตราขยายของ fixed-codebook) สุดท้ายค่าต่างๆ ของวงจรกรองสัญญาณจะถูกปรับค่าตามสัญญาณเอ็กไซเทชันที่หาได้

3.2 หลักการโดยรวมของตัวถอดรหัส CS-ACELP

หลักการทำงานของตัวถอดรหัสแสดงดังรูปที่ 3.3 อันดับแรกพารามิเตอร์ที่รับมาได้ในรูปแบบของบิตจะถูกแยกแยะออกตามเฟรมๆ ละ 10 มิลลิวินาที และถอดรหัสออกมาเป็นค่าพารามิเตอร์ต่างๆ ได้แก่ ค่าสัมประสิทธิ์ LSP fractional pitch delay 2 ค่า fixed-codebook 2 ชุด และอัตราขยายของ fixed-codebook และ adaptive-codebook อย่างละ 2 ชุด สัมประสิทธิ์ LSP นั้นจะถูกดำเนินการประมาณค่าในช่วง แล้วเปลี่ยนเป็นสัมประสิทธิ์ของวงจรกรองสัญญาณ LP

ของแต่ละเฟรมย่อย โดยแต่ละเฟรมย่อยที่มีขนาด 5 มิลลิวินาที (40 ตัวอย่าง) นั้นจะมีขั้นตอนการทำงานดังนี้

1. สัญญาณเอ็กไซเทชันจะถูกสร้างขึ้นโดยการรวม adaptive-codebook และ fixed-codebook เข้าด้วยกันตามอัตราขยายของแต่ละตัว



รูปที่ 3.3 บล็อกการทำงานของตัวถอดรหัส

2. สัญญาณเสียงจะถูกสังเคราะห์ขึ้น โดยนำสัญญาณเอ็กไซเทชันที่ได้มากรองด้วยวงจรกรองสัญญาณสังเคราะห์เสียง LP
3. สัญญาณเสียงที่สังเคราะห์ได้จะนำไปผ่านวงจร Post-processing ซึ่งประกอบด้วย วงจร Adaptive post-filter ที่สร้างจากวงจร กรองสัญญาณแบบ Long-term และแบบ Short-term วงจรกรองสัญญาณความถี่สูงผ่าน และวงจรปรับขนาดสัญญาณ

3.3 รายละเอียดการทำงานของตัวเข้ารหัส CS-ACELP

บล็อกไดอะแกรมการทำงานของตัวเข้ารหัสแบบ CS-ACELP แสดงดังรูปที่ 3.4 สามารถอธิบายการทำงานได้ดังนี้

3.3.1 Pre-Processing

สัญญาณอินพุตกำหนดให้เป็นสัญญาณ PCM ที่มีความละเอียด 16 บิต โดยฟังก์ชันทั้งสองส่วนใน Pre-processing นั้นจะทำหน้าที่ 2 ประการ คือ การปรับระดับสัญญาณ และการกรองความถี่ผ่านสูงโดยใช้วงจรสัญญาณที่มีอันดับเท่ากับ 2 ที่ความถี่ตัด 140 Hz เป็นไปตามสมการ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$H_{hl}(z) = \frac{0.46363718 - 0.92724705z^{-1} + 0.46363718z^{-2}}{1 - 1.9059465z^{-1} + 0.9114024z^{-2}} \quad (3.1)$$

3.3.2 **การวิเคราะห์การทำนายเชิงเส้นและการควอนไทซ์** (Linear prediction analysis and quantization)

วงจรกรองสัญญาณที่ใช้ในการสังเคราะห์จะเป็นวงจรกรองทำนายเชิงเส้นที่มีอันดับเท่ากับ 10 ซึ่งฟังก์ชันการส่งผ่านเป็นดังนี้

$$\frac{1}{\hat{A}(z)} = \frac{1}{1 + \sum_{i=1}^{10} \hat{a}_i z^{-i}} \quad (3.2)$$

เมื่อ $\hat{a}_i$, i = 1,...10 คือสัมประสิทธิ์ (ที่ถูกควอนไทซ์แล้ว) ของตัวทำนายเชิงเส้น การวิเคราะห์สัมประสิทธิ์นี้จะกระทำหนึ่งครั้งต่อเฟรม โดยใช้อัลกอริทึมของ Levison-Durbin บนพื้นฐานของกรรมวิธีอัตสหสัมพันธ์ (autocorrelation method) จากข้อมูลแบบไม่สมมาตรขนาด 30 มิลลิวินาที ของตัวอย่างของเสียงขาเข้าจำนวน 80 ตัวอย่าง (10 มิลลิวินาที) หลังจากได้สัมประสิทธิ์แล้วจะเปลี่ยนเป็นค่า LSP เพื่อนำไปควอนไทซ์และใช้ในการประมาณค่าในช่วง ค่าที่ได้นี้จะถูกเปลี่ยนกลับเป็นสัมประสิทธิ์เพื่อใช้ในการสังเคราะห์และส่งให้วงจรกรองถ่วงน้ำหนัก (weighting filter) ใช้ในเฟรมย่อยแต่ละเฟรมต่อไป

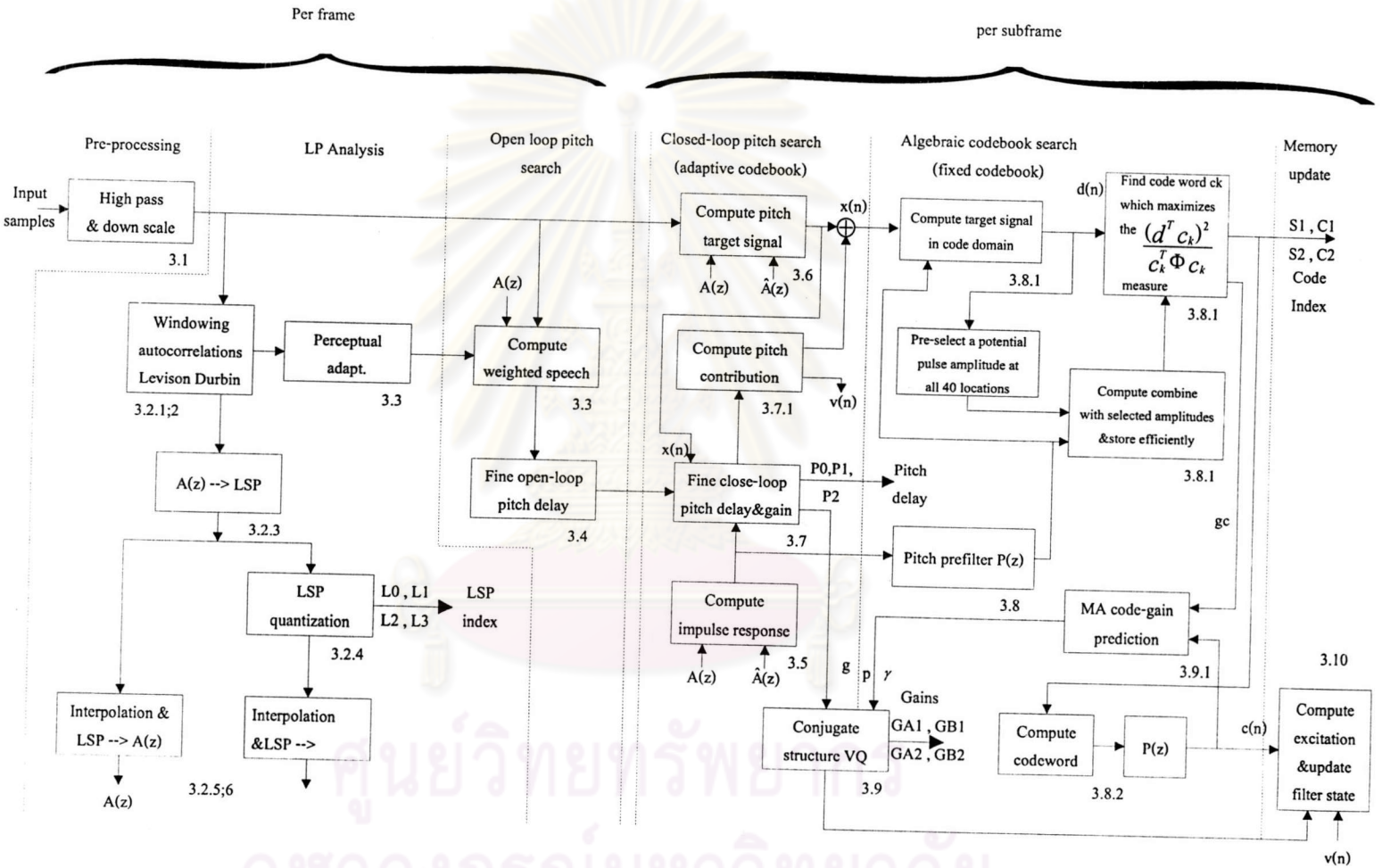
3.3.2.1 **วิธีการหน้าต่างและการคำนวณอัตสหสัมพันธ์** (Windowing and autocorrelation computation)

วิธีการหน้าต่างที่ใช้ในการวิเคราะห์ตัวทำนายเชิงเส้นประกอบด้วย 2 ส่วน ส่วนแรกคือหน้าต่าง Hamming และส่วนที่สองคือ ฟังก์ชันของโคไซน์

$$w_{lp}(n) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{399}) & n = 0,...,199 \\ \cos(\frac{2\pi(n-200)}{159}) & n = 200,...,239 \end{cases} \quad (3.3)$$

ดังนั้นชุดของสัญญาณสุ่มขาเข้าคือ

$$s'(n) = w_{lp}(n)s(n) \qquad n = 0,...239 \quad (3.4)$$



รูปที่ 3.4 บล็อกไดอะแกรมการทำงานของตัวเข้ารหัสแบบ CS-ACELP

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

และสัมประสิทธิ์ของอัตสหสัมพันธ์จะมีค่า

$$r(k) = \sum_{n=k}^{239} s'(n)s'(n-k) \qquad k = 1,...,10 \tag{3.5}$$

เพื่อหลีกเลี่ยงปัญหาการคำนวณเนื่องจากสัญญาณเข้าที่มีขนาดต่ำจะกำหนดให้ r(0) มีค่าต่ำสุดคือเป็น 1.0 และกำหนดแบนด์วิดท์ 60 Hz โดยการคูณสัมประสิทธิ์ของอัตสหสัมพันธ์ที่ได้ด้วย

$$W_{lag}(k) = \exp\left[-\frac{1}{2}\left(\frac{2\pi f_0 k}{f_s}\right)^2\right] \qquad k = 1,...,10 \tag{3.6}$$

นอกจากนี้ r(0) จะถูกคูณด้วยค่า 1.0001 ซึ่งเป็น white-noise factor ดังนั้นสัมประสิทธิ์ของอัตสหสัมพันธ์ใหม่จะมีค่าเป็น

$$\begin{aligned} r'(0) &= 1.0001r(0) \\ r'(k) &= w_{lag}(k)r(k) \qquad k = 1,...,10 \end{aligned} \tag{3.7}$$

3.3.2.2 อัลกอริทึมของ Levison-Durbin

ค่าสัมประสิทธิ์ของอัตสหสัมพันธ์ใหม่ $r'(k)$ จะถูกนำมาใช้หาสัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้นคือ a_i เมื่อ i = 1,...,10 โดยหาได้จากสมการนี้

$$\sum_{i=1}^{10} a_i r'(|i-k|) = -r'(k) \qquad k = 1,...,10 \tag{3.8}$$

เราสามารถหาคำตอบของสมการ (8) ได้โดยใช้อัลกอริทึม Levison-Durbin เป็นแบบวนซ้ำ โดยคำตอบสุดท้ายคือ $a_j = a_j^{[10]}$ เมื่อ j = 1,...10 และ $a_0 = 1.0$

3.3.2.3 การเปลี่ยนสัมประสิทธิ์ LP เป็น LSP

สัมประสิทธิ์ LSP ของวงจรกรองสัญญาณที่มีอันดับเท่ากับ 10 สามารถกำหนดได้ตามสมการผลรวมและผลลบของโพลิโนเมียลตามลำดับดังนี้

$$F_1'(z) = A(z) + z^{-11}A(z^{-1}) \tag{3.9}$$

$$F_2'(z) = A(z) - z^{-11}A(z^{-1}) \tag{3.10}$$

โดย $F_1'(z)$ จะมีรากที่ z = -1 ($\omega = \pi$) และ $F_2'(z)$ จะมีรากที่ z = 1 ($\omega = 0$) ทำให้ได้โพลิโนเมียลชุดใหม่คือ

$$F_1'(z) = F_1'(z)/(1+z^{-1}) \tag{3.11}$$

$$F_2'(z) = F_2'(z)/(1-z^{-1}) \tag{3.12}$$

โพลิโนเมียลแต่ละชุดจะมีรากที่เป็นสังยุคจำนวน 5 รากที่อยู่บนวงกลมหนึ่งหน่วย($e^{\pm j\omega}$) และเขียนได้ในรูปสมการดังนี้

$$F_1(z) = \prod_{i=1,3,...,9}(1 - 2q_i z^{-1} + z^{-2}) \tag{3.13}$$

$$F_2(z) = \prod_{i=2,4,...,10}(1 - 2q_i z^{-1} + z^{-2}) \tag{3.14}$$

เมื่อ $q_i = \cos(\omega_i)$ ซึ่งสัมประสิทธิ์ ω_i นี้เรียกว่า Line spectral frequency(LSF) ส่วนสัมประสิทธิ์ q_i นั้นคือสัมประสิทธิ์ LSP ที่อยู่ในโดเมนของโคไซน์

ทั้งโพลิโนเมียล $F_1(z)$ และ $F_2(z)$ นั้นจะสมมาตรเฉพาะสัมประสิทธิ์ 5 ตัวแรกเท่านั้น ซึ่งจะต้องทำการคำนวณหาสัมประสิทธิ์ของโพลิโนเมียลจากวิธีการรีเคอร์ซีฟดังนี้

$$f_1(i+1) = a_{i+1} + a_{10-i} - f_1(i) \qquad i = 0,...4 \tag{3.15}$$

$$f_2(i+1) = a_{i+1} - a_{10-i} + f_2(i) \qquad i = 0,...4$$

เมื่อ $f_1(0) = f_2(0) = 1.0$ และโพลิโนเมียลแบบ Chebyshev จะถูกใช้หาค่าของ $F_1(z)$ และ $F_2(z)$ วิธีนี้จะให้รากที่อยู่ในโดเมนของโคไซน์โดยตรง โพลิโนเมียล $F_1(z)$ และ $F_2(z)$ นี้จะถูกแทน $z = e^{j\omega}$ ได้ดังนี้

$$F(\omega) = 2e^{-j5\omega}C(x) \tag{3.16}$$

เมื่อ

$$C(x) = T_5(x) + f(1)T_4(x) + f(2)T_3(x) + f(3)T_2(x) + f(4)T_1(x) + f(5)/2 \tag{3.17}$$

เมื่อ $T_m(x) = \cos(m\omega)$ คือโพลิโนเมียลแบบ Chebyshev อันดับที่ m และ $f(i), i = 1,...5$ คือสัมประสิทธิ์ของ $F_1(z)$ และ $F_2(z)$ โพลิโนเมียล C(x) จะหาได้จาก $x = \cos(\omega)$ โดยวิธีการแบบรีเคอร์ซีฟดังนี้

For k = 4 down to 1

$b_k = 2xb_{k+1} - b_{k+2} + f(5-k)$

End

$C(x) = xb_1 - b_2 + f(5)/2$

โดยที่ค่าเริ่มต้น $b_5=1$ และ $b_6=0$

3.3.2.4 การควอนไทซ์สัมประสิทธิ์ LSP

สัมประสิทธิ์ LSP จะถูกควอนไทซ์ด้วยค่า LSF แทนด้วยค่า ω_i โดยจะทำการทำให้เป็นบรรทัดฐานความถี่ในช่วง $[0, \pi]$

$$\omega_i = \arccos(q_i) \qquad i = 1,...,10 \tag{3.18}$$

ตัวทำนายแบบ MA ที่มีอันดับ 4 ถูกใช้ในการทำนายค่าสัมประสิทธิ์ LSF ของเฟรมปัจจุบัน ค่าความแตกต่างระหว่างค่าสัมประสิทธิ์ที่คำนวณได้กับค่าที่ทำนายได้จะได้ถูกคอวนไทซ์ด้วยตัวควอนไทซ์แบบเวกเตอร์ (VQ) แบบ 2 ตอน โดยตอนแรกจะเป็น VQ ที่มีขนาด 1x10 ใช้ Codebook L1 สำหรับข้อมูลจำนวน 128 รูปแบบ(7บิต) ส่วนในสเตจที่สองเป็น VQ แบบ 10 บิตที่ถูกสร้างจากชุด VQ ที่มีขนาด 1x5 จำนวน 2 ตัว โดยมี Codebook เป็น L2 และ L3 สำหรับข้อมูล 32 รูปแบบ(5 บิต) ในแต่ละ VQ โดยแต่ละสัมประสิทธิ์ที่ได้จากการรวมกันของ Codebook ทั้งสองคือ

$$\hat{l}_i = \begin{cases} L1_i(L1) + L2_i(L2) & i = 1,...5 \\ L1_i(L1) + L3_{i-5}(L3) & i = 6,...10 \end{cases} \tag{3.19}$$

เมื่อ L1 , L2 และ L3 คือตัวชี้ Codebook เพื่อหลีกเลี่ยงการเกิด Sharp resonance ในวงจรกรองตัวทำนายเชิงเส้นที่ควอนไทซ์แล้ว สัมประสิทธิ์ $\hat{l}$ จะต้องจัดเรียงให้ค่าสัมประสิทธิ์ของตัวข้างเคียงให้มีค่าระยะทางต่ำ J ต่ำสุด ดังนี้

$$
\begin{aligned}
&for\ i = 2,...10 \\
&\quad if(\hat{l}_{i-1} > \hat{l}_i - J) \\
&\qquad \hat{l}_{i-1} = (\hat{l}_i + \hat{l}_{i-1} - J)/2 \\
&\qquad \hat{l} = (\hat{l}_i + \hat{l}_{i-1} + J)/2 \\
&\qquad end \\
&end
\end{aligned}
$$

หลังจากนี้เราก็จะได้สัมประสิทธิ์ LSF ที่ผ่านการควอนไทซ์แล้ว $\hat{\omega}_i^{(m)}$ สำหรับเฟรมที่ m จากการรวมกันของขาออกของตัวควอนไทซ์ก่อนหน้า $\hat{l}_i^{(m-k)}$ และขาออกของตัวควอนไทซ์ในปัจจุบัน $\hat{l}_i^{(m)}$ ดังนี้

$$\hat{\omega}_i^{(m)} = \left(1 - \sum_{k=1}^{4} \hat{p}_{i,k}\right)\hat{l}_i^{(m)} + \sum_{k=1}^{4} \hat{p}_{i,k}\hat{l}_i^{(m-k)} \qquad i = 1,...10 \qquad (3.20)$$

เมื่อ $\hat{p}_{i,k}$ คือสัมประสิทธิ์ของตัวทำนายสัญญาณ MA ซึ่งตัวทำนายสัญญาณ MA ที่ใช้นี้จะถูกกำหนดโดยบิต L0 ค่าเริ่มต้นของ $\hat{l}_i^k$ คือ $\hat{l}_i = i\pi/11$ สำหรับทุกค่า k < 0 หลังจากการคำนวณหา $\hat{\omega}_i$ วงจรกรองจะตรวจสอบความมีเสถียรภาพดังนี้

1. ทำการเรียงลำดับสัมประสิทธิ์ $\hat{\omega}_i$ จากน้อยไปหามาก
2. ถ้า $\hat{\omega}_i$ <0.005 จะให้ $\hat{\omega}_i$ = 0.005
3. ถ้า $\hat{\omega}_{i+1}$ - $\hat{\omega}_i$ < 0.0391 จะให้ $\hat{\omega}_{i+1} = \hat{\omega}_i$ + 0.0391 ,i = 1,...9
4. ถ้า $\hat{\omega}_{10}$ >3.135 จะให้ $\hat{\omega}_{10}$ =3.135

ขั้นตอนการเข้ารหัสพารามิเตอร์ LSF นี้สามารถอธิบายได้ดังนี้ ตัวทำนายสัญญาณ MA แต่ละตัวจากทั้ง 2 ตัวนั้นจะประมาณค่าจากค่าสัมประสิทธิ์ LSF ในปัจจุบัน ค่าที่ประมาณขึ้นมาได้จะเป็นค่า Weighted mean-squared error ต่ำสุด

$$E_{lsf} = \sum_{i=1}^{10} w_i(\omega_i - \hat{\omega}_i)^2 \qquad (3.21)$$

โดยค่าน้ำหนักถ่วง w_i สามารถปรับเปลี่ยนได้ตามฟังก์ชันของสัมประสิทธิ์ LSF ที่ยังไม่ได้ควอนไทซ์ดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$w_1 = \begin{cases} 1.0 & if \quad \omega_2 - 0.04\pi - 1 > 0 \\ 10(\omega_2 - 0.04\pi - 1)^2 + 1 & otherwise \end{cases}$$

$$w_i : 2 \le i \le 9 = \begin{cases} 1.0 & if \quad \omega_{i+1} - \omega_{i-1} - 1 > 0 \\ 10(\omega_{i+1} - \omega_{i-1} - 1)^2 + 1 & otherwise \end{cases} \qquad (3.22)$$

$$w_{10} = \begin{cases} 1.0 & if \quad -\omega_9 + 0.92\pi - 1 > 0 \\ 10(-\omega_9 - 0.92\pi - 1)^2 + 1 & otherwise \end{cases}$$

โดยที่ค่าถ่วงน้ำหนักของ w_5, w_6 ต้องคูณด้วย 1.2 และเวกเตอร์ที่ถูกควอนไทซ์สำหรับแต่ละเฟรม m ได้มาจาก

$$l_i = \left[\omega_i^m - \sum_{k=1}^{4} \hat{p}_{i,k} \hat{l}_i^{(m-k)} \right] / \left[1 - \sum_{k=1}^{4} \hat{p}_{i,k} \right] \qquad i = 1,...10 \qquad (3.23)$$

3.3.2.5 การประมาณค่าในช่วงสัมประสิทธิ์ LSP

สัมประสิทธิ์ตัวทำนายเชิงเส้นทั้งที่ถูกควอนไทซ์และที่ไม่ได้ถูกควอนไทซ์ จะถูกใช้ในเฟรมย่อยที่สอง ส่วนในเฟรมย่อยแรกนั้นสัมประสิทธิ์ตัวทำนายเชิงเส้นจะได้มาจากการประมาณค่าในช่วงจากพารามิเตอร์ของเฟรมข้างเคียง จากค่าสัมประสิทธิ์ LSP ในโดเมนของโคไซน์กำหนดให้ $q_i^{(current)}$ คือค่าสัมประสิทธิ์ LSP ที่คำนวณได้จากเฟรมปัจจุบันขนาด 10 มิลลิวินาที และค่า $q_i^{(previous)}$ คือค่าสัมประสิทธิ์ LSP ที่คำนวณได้จากเฟรมก่อนหน้า สัมประสิทธิ์ LSP ที่ได้จากการประมาณค่าในช่วง(ที่ไม่ได้ถูกควอนไทซ์) ในเฟรมย่อยแต่ละเฟรมจะมีค่าดังนี้

$$\begin{aligned} &\text{เฟรมย่อยที่ } 1 : q_i^{(1)} = 0.5 q_i^{(Previous)} + 0.5 q_i^{(current)} \quad i = 1,...10 \\ &\text{เฟรมย่อยที่ } 2 : q_i^{(2)} = 0.5 q_i^{(current)} \quad i = 1,...10 \end{aligned} \qquad (3.24)$$

ค่าสัมประสิทธิ์ LSP ที่ควอนไทซ์แล้วก็จะมีขั้นตอนการประมาณค่าในช่วงเหมือนกัน เพียงแต่แทนสัญลักษณ์จาก q_i เป็น $\hat{q}_i$

3.3.2.6 การเปลี่ยนสัมประสิทธิ์ LSP เป็น LP

ค่าสัมประสิทธิ์ LSP ถูกควอนไทซ์และประมาณค่าในช่วงนั้นจะถูกเปลี่ยนกลับเป็นค่าสัมประสิทธิ์ตัวทำนายเชิงเส้นคือ a_i โดยการหาค่าสัมประสิทธิ์ของ $F_1(z)$ และ $F_2(z)$ จากสมการ

ที่ (3.13) และ (3.14) เมื่อเราทราบค่าสัมประสิทธิ์ LSP ที่ถูกควอนไทซ์และทำการประมาณค่าในช่วงแล้ว ดังนั้นค่าสัมประสิทธิ์ $f_1(i)$ จะคำนวณได้จาก q_i โดยวิธีการวนซ้ำดังนี้

$$
\begin{aligned}
&for\ i=1\ to\ 5\\
&\quad f_1(i) = -2q_{2i-1}f_1(i-1) + 2f_1(i-2)\\
&\quad for\ j=i-1\ downto\ 1\\
&\qquad f_1^i(j) = f_1^{[i-1]}(j) - 2q_{2i-1}f_1^{i-1}(j-1) + f_1^{[i-1]}(j-2)\\
&\quad end\\
&end
\end{aligned}
$$

โดยค่าเริ่มต้น $f_1(0)=1$ และ $f_1(-1)=0$ และค่า $f_2(i)$ คำนวณได้จากวิธีเดียวกันโดยเปลี่ยนจาก $2q_{2i-1}$ เป็น $2q_{2i}$

เนื่องจากทั้ง $F_1(z)$ และ $F_2(z)$ เมื่อถูกคูณด้วย $1+z^{-1}$ และ $1-z^{-1}$ ตามลำดับแล้วจะได้ค่า $F'_1(z)$ และ $F'_2(z)$ ดังนั้นสัมประสิทธิ์ $f_1(i)$ และ $f_2(i)$ ที่หาได้จะเป็น

$$
\begin{aligned}
f'_1(i) &= f_1(i) + f_1(i-1) \qquad i=1,\ldots5\\
f'_2(i) &= f_2(i) + f_2(i-1) \qquad i=1,\ldots5
\end{aligned}
\tag{3.25}
$$

สุดท้ายค่าสัมประสิทธิ์ตัวทำนายเชิงเส้น จะคำนวณหาได้จาก f_1' และ f'_2 ได้ดังนี้

$$
a_i = \begin{cases} 0.5f_1'(i) + 0.5f_2'(i) & i=1,\ldots5 \\ 0.5f_1'(11-i) - 0.5f_2'(11-i) & i=6,\ldots10 \end{cases}
\tag{3.26}
$$

3.3.3 Perceptual weighting

วงจรกรองสัญญาณ Perceptual weighting นั้นมีโครงสร้างมาจากค่าสัมประสิทธิ์ที่ยังไม่ควอนไทซ์ a_i ของวงจรกรองตัวทำนายเชิงเส้น มีคุณสมบัติดังนี้

$$
W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} = \frac{1+\sum_{i=1}^{10}\gamma_1^i a_i z^{-i}}{1+\sum_{i=1}^{10}\gamma_2^i a_i z^{-i}}
\tag{3.27}
$$

ค่า γ_1 และ γ_2 ใช้ในการกำหนดความถี่ตอบสนองของวงจรกรอง W(z) ส่วนค่าสัมประสิทธิ์ reflection k_i จะถูกเปลี่ยนเป็นสัมประสิทธิ์ LAR (Log Area Ratio) o_i ดังนี้

$$
o_i = \log\frac{(1.0+k_i)}{(1.0-k_i)} \qquad i=1,2
\tag{3.28}
$$

สัมประสิทธิ์ LAR ที่ได้จากเฟรมขนาด 10 มิลลิวินาที ของเฟรมปัจจุบันนั้นจะถูกใช้ในเฟรมย่อยที่สอง ส่วนสัมประสิทธิ์ LAR ของเฟรมย่อยแรกนั้นจะได้มาจากการประมาณค่าในช่วงกับสัมประสิทธิ์ LAR ของเฟรมก่อนหน้า ดังนั้นค่าสัมประสิทธิ์ LAR ในเฟรมย่อยแต่ละเฟรมมีค่าดังนี้

$$\begin{aligned} Subframe\,1: o_i^{(1)} &= 0.5o_i^{(previous)} + 0.5o_i^{(current)} \qquad i = 1,2 \\ Subframe\,2: o_i^{2} &= o_i^{(current)} \qquad i = 1,2 \end{aligned} \tag{3.29}$$

เอนเวโลปของสเปกตรัมนั้นถูกกำหนดให้ flat (flat=1) หรือ tilt (flat=0) เพื่อหลีกเลี่ยงการเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็วจะใช้ค่าของ flat ในเฟรมย่อยก่อนหน้า

$$flat^{(m)} = \begin{cases} 0 & if\ \ o_1^{(m)} < -1.74\ and\ o_2^{(m)} > 0.65\ \ and\ flat^{(m-1)} = 1 \\ 1 & if\ \ (o_1^{m} > -1.52\ or\ \ o_2^{(m)} < 0.43)\ \ and\ \ flat^{(m-1)} = 0 \\ flat^{(m-1)} & otherwise \end{cases} \tag{3.30}$$

ถ้าสเปกตรัมของสัมประสิทธิ์ที่ได้จากการประมาณค่าในช่วงของเฟรมย่อยเป็น flat ($flat^m$=1) จะกำหนดให้ $\gamma_1 = 0.94$ และ $\gamma_2 = 0.6$ ถ้าสเปกตรัมของสัมประสิทธิ์ที่ได้เป็น tilt ($flat^m$= 0) จะได้ $\gamma_1 = 0.98$ และ γ_2 ปรับเปลี่ยนได้ตามขนาดของเรโซแนนซ์ที่เกิดขึ้นในวงจรกรองตัวทำนายเชิงเส้นซึ่งมีค่าอยู่ระหว่าง 0.4 ถึง 0.7 การปรับเปลี่ยนทำได้โดยการวัดค่าระยะต่ำสุดระหว่างสัมประสิทธิ์ LSP 2 ชุดที่ได้จากเฟรมปัจจุบัน ดังนี้

$$d_{\min} = \min[\omega_{i+1} - \omega_i] \qquad i = 1,...9 \tag{3.31}$$

ค่า γ_2 นี้จะคำนวณได้จากความสัมพันธ์ดังนี้

$$\gamma_2 = -6.0\, d_{\min} + 1.0 \qquad 0.4 \le \gamma_2 \le 0.7 \tag{3.32}$$

สัญญาณเสียงที่ถูกถ่วงน้ำหนักในเฟรมย่อยจะได้มาจาก

$$sw(n) = s(n) + \sum_{i=1}^{10} a_i \gamma_1^i s(n-i) - \sum_{i=1}^{10} a_i \gamma_2^i sw(n-i) \qquad n = 0,...39 \tag{3.33}$$

3.3.4 การวิเคราะห์หาพิตช์ใน open-loop

การหาพิตช์วงเปิดนั้นจะประมาณค่าได้โดยใช้สัญญาณเสียงที่ถูกถ่วงน้ำหนักแล้ว $sw(n)$ ตามสมการ (3.33) แล้วทำการหาค่าที่สหสัมพันธ์กันมากที่สุด 3 ค่าดังนี้

$$R(k) = \sum_{n=0}^{79} sw(n)sw(n-k) \tag{3.34}$$

และหาค่าใน 3 ช่วงดังนี้

$$i = 1:80,...,143$$
$$i = 2:40,...,79$$
$$i = 3:20,...,39$$

และค่าที่มากที่สุดของ $R(t_i)$, i =1,...,3 จะถูกทำให้เป็นบรรทัดฐานดังนี้

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_n sw^2(n-t_i)}} \qquad i = 1,...,3 \tag{3.35}$$

ค่าที่มากที่สุดระหว่างสามค่าที่ถูกทำให้เป็นบรรทัดฐานนี้จะถูกเลือกเพื่อใช้ในการหาค่าการประวิงเวลา ซึ่งกระทำโดยการถ่วงน้ำหนักค่าอัตสหสัมพันธ์ที่ถูกทำให้เป็นบรรทัดฐานนี้ตามค่าการประวิงเวลา ทำให้ได้ค่าการประวิงเวลา T_{op} ในวงเปิดดังนี้

$$T_{op} = t_1$$
$$R'(T_{op}) = R'(t_1)$$
$$if\ R'(t_2) \geq 0.85\ R'(T_{op})$$
$$R'(T_{op}) = R'(t_2)$$
$$T_{op} = t_2$$
$$end$$
$$if\ R'(t_3) \geq 0.85\ R'(T_{op})$$
$$R'(T_{op}) = R'(t_3)$$
$$T_{op} = t_3$$
$$end$$

ในขั้นตอนนี้จะแบ่งช่วงการประวิงเวลาออกเป็น 3 ช่วงและใช้ค่าที่น้อยที่สุดเพื่อหลีกเลี่ยงการเลือกพิตช์หลายค่า

3.3.5 การคำนวณหาการตอบสนองอิมพัลส์

การตอบสนองอิมพัลส์ (impulse response) h(n) ของวงจรกรองสัญญาณ weighted synthesis $W(z)/\hat{A}(z)$ นั้นจำเป็นต้องใช้ในการค้นหาใน Adaptive codebook และ fixed-codebook โดยจะทำการคำนวณทุกๆ เฟรมย่อยโดยการกรองสัญญาณที่ประกอบด้วยสัญญาณ

สัมประสิทธิ์ของวงจรกรอง $A(z/\gamma_1)$ ที่ขยายด้วยรากที่เป็นศูนย์ด้วยวงจรกรอง $1/\hat{A}(z)$ และ $A(z/\gamma_2)$

3.3.6 การคำนวณหาสัญญาณเป้าหมาย

สัญญาณเป้าหมาย x(n) ที่ใช้ในการค้นหาใน Adaptive-codebook จะคำนวณมาจากการลบการตอบสนองของวงจรกรอง weighted synthesis โดยที่ $W(z)/\hat{A}(z) = A(z/\gamma_1)/[\hat{A}(z)A(z/\gamma_2)]$ ที่มีขาเข้าเป็นศูนย์ ออกจากสัญญาณเสียงที่ถูกถ่วงน้ำหนัก sw(n) ตามสมการที่ (3.33) ซึ่งจะทำการคำนวณทุกๆเฟรมย่อย

ขั้นตอนที่ใช้ในการคำนวณหาสัญญาณเป้าหมายที่ใช้ในข้อกำหนดนี้คือการกรองตัวทำนายเชิงเส้น residual r(n) โดยการรวมวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์ $1/\hat{A}(z)$ และวงจรกรองสัญญาณถ่วงน้ำหนัก $A(z/\gamma_1)/A(z/\gamma_2)$ หลังจากการหาสัญญาณกระตุ้นของเฟรมย่อยแต่ละเฟรมแล้วค่าเริ่มต้นของ วงจรกรองสัญญาณเหล่านี้จะถูกปรับให้ทันกาลโดยการกรองสัญญาณที่เกิดจากผลต่างระหว่างสัญญาณ residual กับสัญญาณกระตุ้น

สัญญาณ residual ที่จำเป็นต้องใช้ในการหาเวกเตอร์ของเป้า จะถูกใช้ในการค้นหาใน Adaptive-codebook กล่าวคือขึ้นตอนการหาค่าการประวิงเวลาของ adaptive-codebook จะน้อยกว่าขนาดของเฟรมย่อย โดยสัญญาณ residual ของตัวทำนายเชิงเส้นมีค่าดังนี้

$$r(n) = s(n) + \sum_{i=1}^{10} a_i s(n-i) \qquad n = 0,...39 \qquad (3.36)$$

3.3.7 การค้นหา adaptive-codebook

พารามิเตอร์ของ adaptive-codebook คือการประวิงเวลาอัตราขยาย โดยใช้วงจรกรองสัญญาณพิตช์ซึ่งจะทำการค้นหาสัญญาณกระตุ้นที่มีการประวิงเวลาน้อยกว่าขนาดของเฟรมย่อย ขั้นตอนการค้นหานั้น สัญญาณกระตุ้นจะถูกขยายด้วยสัญญาณ residual ของตัวทำนายเชิงเส้นเพื่อทำให้การค้นหาในวงปิดเป็นไปโดยง่าย ดังนั้นการค้นหา adaptive codebook จะกระทำทุกๆเฟรมย่อย ในเฟรมย่อยแรกส่วนของการประวิงเวลาของพิตช์ T_1 ส่วนในเฟรมย่อยที่สองค่าการประวิงเวลา T_2 โดยคำนวณค่าเป็นจำนวนเต็ม

ในเฟรมย่อยแต่ละเฟรมนั้นการหาค่าการประวิงเวลาที่ถูกต้องที่สุดทำได้โดยการวิเคราะห์ในวงปิด เพื่อให้ค่า MSE ของค่าน้ำหนักถ่วงมีค่าต่ำสุด ในเฟรมย่อยแรกหลังจากได้ค่าการประวิงเวลา T_1 โดยการค้นหาจากช่วงสั้นรอบๆ ค่าการประวิงเวลา T_{op} ที่ได้จากวงเปิด ขอบเขตของการค้นหา t_{min} และ t_{max} กำหนดได้ดังนี้

$$t_{\min} = T_{op} - 3$$
$$if\ t_{\min} < 20\ then\ t_{\min} = 20$$
$$t_{\max} = t_{\min} + 6$$
$$if\ t_{\max} > 143\ then$$
$$t_{\max} = 143$$
$$t_{\min} = t_{\max} - 6$$
$$end$$

สำหรับเฟรมย่อยที่สองการวิเคราะห์การประวิงเวลาของพิตช์ในวงปิด จะค้นหารอบๆ ค่าพิตช์ที่ได้จากเฟรมย่อยแรกเพื่อหาค่าการประวิงเวลา T_2 ที่ดีที่สุดโดยขอบเขตในการค้นหาจะอยู่ระหว่าง t_{min} และ t_{max} เมื่อ t_{min} และ t_{max} คือค่าที่ได้จาก T_1 ดังนี้

$$t_{\min} = \text{int}(T_1) - 5$$
$$if\ t_{\min} < 20\ \ then\ \ t_{\min} = 20$$
$$t_{\max} = t_{\min} + 9$$
$$if\ t_{\max} > 143\ then$$
$$t_{\max} = 143$$
$$t_{\min} = t_{\max} - 9$$
$$end$$

ในการค้นหาพิตช์ของวงปิดจะหาค่าที่น้อยที่สุดของ mean-squared weighted error ระหว่างสัญญาณเสียงต้นแบบกับสัญญาณเสียงที่สร้างขึ้นมา โดยการหาค่าที่มากที่สุดจากพจน์ดังนี้

$$R(k) = \frac{\sum_{n=0}^{39} x(n) y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k(n) y_k(n)}} \qquad (3.37)$$

เมื่อ $x(n)$ คือสัญญาณเป้าหมาย และ $y_k(n)$ คือสัญญาณกระตุ้นที่ผ่านวงจรกรองสัญญาณที่มีการประวิงเวลา k (สัญญาณกระตุ้นคอนโวลูชันกับ h(n)) อนึ่งช่วงการค้นหานั้นจะถูกจำกัดอยู่รอบๆ ค่าที่ถูกเลือกไว้ก่อนแล้ว ซึ่งมาจากพิตช์ T_{op} ในการวิเคราะห์ในวงเปิด สำหรับเฟรมย่อยแรกและ T_1 สำหรับเฟรมย่อยที่สอง

การคำนวณสัญญาณ $y_k(n)$ สำหรับแต่ละค่าการประวิงเวลาจาก t_{min} สำหรับค่าการประวิงเวลาอื่นๆ ในช่วงการค้นหาคือ $k = t_{min} + 1, \ldots, t_{max}$ โดยใช้วิธีรีเคอร์ซีฟดังนี้

$$y_k(n) = y_{k-1}(n-1) + u(-k)h(n) \qquad n = 39,\ldots 0 \qquad (3.38)$$

เมื่อ u(n) , n = -143,...39, คือบัฟเฟอร์ของสัญญาณกระตุ้น และ $y_{k-1}(-1) = 0$ จะเห็นได้ว่าในขั้นตอนการค้นหานั้นสัญญาณตัวอย่าง u(n), n = 0,...39 นั้นไม่ทราบค่า จึงทำให้จำเป็นต้องหาค่าการประวิงเวลาของพิตช์ในช่วงที่น้อยกว่า 40 วิธีการอย่างง่ายคือสัญญาณ residual ของตัวทำนายเชิงเส้นจะถูกสำเนาไปเป็น u(n) เพื่อทำให้สมการที่ (3.38) เป็นจริงสำหรับทุกค่าการประวิงเวลา

สำหรับการหาค่า T_2 และ T_1 ถ้าค่าที่น้อยที่สุดของค่าประวิงเวลาของวงรอบปิดที่เป็นจำนวนเต็มน้อยกว่า 85 จะต้องมีการตรวจสอบค่าเศษส่วนรอบๆ ค่าจำนวนเต็มของค่าประวิงเวลา การค้นหาค่าพิตช์แบบเศษส่วน สามารถทำได้โดยใช้การประมาณค่าในช่วงของค่าบรรทัดฐานสหสัมพันธ์ (normalized correlation) ในสมการที่ 3.37 และหาที่มากที่สุด การประมาณค่าในช่วงทำได้โดยใช้วงจรกรองแบบ FIR ใช้หน้าต่างแบบ Hamming ที่มีจุดตัดความถี่ (-3 dB) ที่ 3600 เฮริตซ์ การหาค่า R(k) สำหรับค่าเศษส่วน $-\frac{2}{3}$ $-\frac{1}{3}$ 0 $\frac{1}{3}$ $\frac{2}{3}$ จะใช้สูตรการประมาณค่าในช่วงดังนี้

$$R(k)_t = \sum_{i=0}^{3} R(k-i)\,b_{12}(t+3i) + \sum_{i=0}^{3} R(k+1+i)\,b_{12}(3-t+3i) \qquad t = 0,1,2 \qquad (3.39)$$

เมื่อ t = 0,1,2 สำหรับค่าเศษส่วน 0 1/3 2/3 ตามลำดับ สังเกตว่าจำเป็นจะต้องคำนวณค่าสหสัมพันธ์ในสมการที่ 3.37 ด้วยการใช้ย่าน $t_{min} - 4$, $t_{max} + 4$ ตามคุณสมบัติของการประมาณค่าในช่วง

3.3.7.1 การสร้างเวกเตอร์ adaptive-codebook

เมื่อทำการหาค่าการประวิงเวลาของพิตช์ได้แล้วจะทำการคำนวณหาเวกเตอร์ของ codebook v(n) โดยหาจากการประมาณค่าในช่วงจากสัญญาณกระตุ้นของ u(n) ในอดีตที่มีการประวิงเวลาเป็นจำนวนเต็ม k และเศษส่วน t ดังนี้

$$v(n) = \sum_{i=0}^{9} u(n-k+i)\,b_{30}(t+3i) + \sum_{i=0}^{9} u(n-k+1+i)\,b_{30}(3-t+3i)$$
$$n = 0,...39 \quad t = 0,1,2 \qquad (3.40)$$

การประมาณค่าในช่วงโดยใช้วงจรกรองแบบ FIR ใช้หน้าต่างแบบ Hamming มีจุดตัดความถี่อยู่ที่ (-3 dB) 3600 เฮริตซ์

3.3.7.2 การเข้ารหัสค่าการประวิงเวลาของ adaptive-codebook

การประวิงเวลาของพิตช์ T_1 จะถูกเข้ารหัสขนาด 8 บิตในเฟรมย่อยแรก ส่วนการประวิงเวลา T_2 ของเฟรมย่อยที่สองนั้นจะถูกเข้ารหัสขนาด 5 บิต ค่าพารามิเตอร์ของพิตช์ P1 จะได้จาก

$$P1 = \begin{cases} 3(\text{int}(T_1) - 19) + frac - 1 & if \ T_1 = [19,...,85], frac = [-1,0,1] \\ (\text{int}(T_1) - 85) + 197 & if \ T_1 = [86,...143], frac = 0 \end{cases} \quad (3.41)$$

และค่าการประวิงเวลาของพิตช์ T_2 คือการเข้ารหัสที่มีความสัมพันธ์กับค่า T_1 ซึ่งหาได้ดังนี้

$$P2 = 3(\text{int}(T_2) - t_{\min}) + frac + 2 \quad (3.42)$$

เมื่อ t_{min} ได้มาจาก T_1 ในหัวข้อที่ 3.3.7

การสร้างตัวเข้ารหัสถูกออกแบบให้มีความทนทานต่อการผิดพลาดของบิตข้อมูลนั้นได้โดยใช้พาริตีบิต P0 ซึ่งคำนวณมาจากตัวชี้ P1 ของเฟรมย่อยแรก พาริตีบิตที่ได้นั้นมาจากการทำ XOR จากบิตที่มีนัยสำคัญสูงสุด 6 บิตของ P1 ที่ตัวถอดรหัสพาริตีนี้จะถูกคำนวณเพื่อใช้ในการตรวจสอบการผิดพลาดของบิตข้อมูล ซึ่งถ้าได้ค่าไม่ตรงกันก็จะใช้วิธีการแก้ไขข้อผิดพลาดเข้ามาช่วย

3.3.7.3 การคำนวณหาอัตราขยายของ adaptive-codebook

เมื่อหาค่าการประวิงเวลาของ adaptive-codebook ได้แล้วจะทำการคำนวณหาอัตราขยาย g_p ดังนี้

$$g_p = \frac{\sum_{n=0}^{39} x(n)y(n)}{\sum_{n=0}^{39} y(n)y(n)} \qquad 0 \le g_p \le 1.2 \quad (3.43)$$

เมื่อ $x(n)$ คือ สัญญาณของเป้า ส่วน $y(n)$ คือสัญญาณที่ได้จากการกรองเวกเตอร์ของ adaptive-codebook (การตอบสนองของ W(z)/A(z) ที่มีขาเข้าเป็นศูนย์คือ $v(n)$) โดยได้มาจากการคอนโวลูชันระหว่าง $v(n)$ กับ $h(n)$ ดังนี้

$$y(n) = \sum_{i=0}^{n} v(i)h(n-i) \qquad n = 0,...39 \quad (3.44)$$

3.3.8 โครงสร้างและการค้นหา Fixed codebook

Fixed codebook คือ พื้นฐานโครงสร้างแบบพีชคณิตที่ออกแบบโดยใช้ Interleaved Single-Pulse Permutation (ISPP) ซึ่ง Codevector จะประกอบด้วย 4 พัลส์ (มีค่าไม่เป็นศูนย์) โดยแต่ละพัลส์มีขนาดเป็น +1 และ -1 และสามารถกำหนดตำแหน่งของพัลส์ดังตาราง

ตารางที่ 3.2 โครงสร้างของ Fixed-codebook

Pulse	Sign	Positions
i_0	$s_0 : \pm 1$	$m_0 : 0,5,10,15,20,25,30,35$
i_1	$s_1 : \pm 1$	$m_1 : 1,6,11,16,21,26,31,36$
i_2	$s_2 : \pm 1$	$m_2 : 2,7,12,17,22,27,32,37$
i_3	$s_3 : \pm 1$	$m_3 : 3,8,13,18,23,28,33,38$ $4,9,14,19,24,29,34,39$

Codebook c(n) สร้างจากเวกเตอร์ศูนย์ขนาด 1x40 และหาตำแหน่งของพัลส์ทั้ง 4 ที่มีการคูณด้วยเครื่องหมายในแต่ละตำแหน่งดังนี้

$$c(n) = s_0\delta(n-m_0) + s_1\delta(n-m_1) + s_2\delta(n-m_2) + s_3\delta(n-m_3) \quad n = 0,...,39 \qquad (3.45)$$

เมื่อ $\delta(0)$ คือพัลส์ขนาดหนึ่งหน่วย สิ่งที่พิเศษอย่างหนึ่งของ codebook นี้คือ codebook ที่ถูกเลือกนั้นจะถูกกรองด้วย adaptive pre-filter P(z) ซึ่งทำให้ส่วนที่เป็นฮาร์มอนิกนั้นเด่นขึ้นมาเพื่อตรวจสอบคุณภาพของเสียงที่สร้างขึ้นมาได้ วงจรกรองนี้คือ

$$P(z) = 1/(1-\beta z^{-T}) \qquad (3.46)$$

เมื่อ T คือส่วนที่เป็นจำนวนเต็มของค่าการประวิงเวลาพิตช์ของเฟรมย่อยปัจจุบันและ β คือ อัตราขยายของพิตช์ ซึ่งจะมีค่าปรับเปลี่ยนได้โดยการควอนไทซ์อัตราขยายของ adaptive codebook จากเฟรมย่อยก่อนหน้านี้ดังนี้

$$\beta = \hat{g}_p^{(m-1)} \quad \text{โดยที่} \quad 0.2 \le \beta \le 0.8 \qquad (3.47)$$

ถ้าค่าประวิงเวลาน้อยกว่า 40 codebook c(n) ของสมการที่ 3.45 จะเปลี่ยนเป็น

$$c(n) = \begin{cases} c(n) & n = 0,...T-1 \\ c(n) + \beta c(n-T) & n = T,...39 \end{cases} \quad (3.48)$$

การปรับเปลี่ยนของ fixed-codebook นี้จะต้องทำการปรับเปลี่ยนการตอบสนองอิมพัลส์ h(n) ตามสมการดังนี้

$$h(n) = \begin{cases} h(n) & n = 0,...T-1 \\ h(n) + \beta h(n-T) & n = T,...,39 \end{cases} \quad (3.49)$$

3.3.8.1 ขั้นตอนการค้นหา fixed-codebook

การหา fixed-codebook นั้นทำได้โดยการหาค่าต่ำสุดของ mean-squared error ระหว่างสัญญาณเสียงที่ถูกถ่วงน้ำหนัก $sw(n)$ ตามสมการที่ (3.33) และสัญญาณเสียงที่สร้างกลับขึ้นมาที่ถูกถ่วงน้ำหนักแล้ว สัญญาณของเป้าที่ใช้ในการหาพิตช์ในวงเปิด จะถูกปรับให้ทันกาลโดยการลบด้วยค่าของ adaptive-codebook ดังนี้

$$x'(n) = x(n) - g_p y(n) \qquad n = 0,...39 \quad (3.50)$$

เมื่อ y(n) คือ เวกเตอร์ของ adaptive codebook ที่ได้จากสมการที่ (3.44) และ g_p คือ อัตราขยายของ adaptive codebook จากสมการที่ (3.43)

การกำหนดเมทริกซ์ Φ มีสมบัติเป็นเมทริกซ์แบบสมมาตรประกอบด้วยสมาชิกที่เป็นค่าอัตสหสัมพันธ์ของ h(n) ดังนี้

$$\phi(i,j) = \sum_{n=j}^{39} h(n-i)h(n-j) \qquad i = 0,...39 \qquad j = i,...,39 \quad (3.51)$$

สัญญาณอัตสหสัมพันธ์ d(n) ได้จากสัญญาณเป้าหมาย $\mathbf{x'(n)}$ และผลตอบสนองอิมพัลซ์ h(n) ดังนี้

$$d(n) = \sum_{i=n}^{39} x'(i)h(i-n) \qquad n = 0,...39 \quad (3.52)$$

ถ้า c_k คือเวกเตอร์ของ fixed-codebook ที่ k จะสามารถหา codebook ได้โดยทำให้พจน์ต่อไปนี้มีค่ามากที่สุด

$$\frac{C_k^2}{E_k} = \frac{(\sum_{n=0}^{39} d(n)c_k(n))^2}{c_k' \Phi c_k} \quad (3.53)$$

เมื่อ c_k' คือทรานโพสต์ของ c_k

สัญญาณ d(n) และเมทริกซ์ Φ จะถูกคำนวณก่อนที่จะหา codebook จะเห็นได้ว่าจะต้องมีการคำนวณหาค่าต่างๆ ดังนั้นจึงต้องออกแบบขั้นตอนการจัดเก็บข้อมูลอย่างมีประสิทธิภาพเพื่อเพิ่มความเร็วในการค้นหา

โครงสร้างที่เป็นแบบพีชคณิตของ codebook C จะทำให้ขั้นตอนในการค้นหาเป็นไปอย่างรวดเร็ว แต่เวกเตอร์ของ codebook c_k จะมีจำนวนพัลส์ 4 พัลส์เท่านั้นที่ไม่เป็นศูนย์ การทำอัตสหสัมพันธ์ของตัวเศษในสมการที่ 3.53 สำหรับกำหนดเวกเตอร์ c_k ดังนี้

$$C = \sum_{i=0}^{3} s_i d(m_i) \tag{3.54}$$

เมื่อ m_i คือตำแหน่งของพัลส์ที่ i และ s_i คือขนาดของพัลส์นั้น และพลังงานในตัวส่วนของสมการที่ (3.53) คือ

$$E = \sum_{i=0}^{3} \phi(m_i, m_i) + 2\sum_{i=0}^{2} \sum_{j=i+1}^{3} s_i s_j \phi(m_i, m_i) \tag{3.55}$$

เพื่อทำให้ขั้นตอนการทำงานนี้ง่ายขึ้น ขนาดของพัลส์จะถูกกำหนดโดยนำค่ามาจากการควอนไทซ์สัญญาณ d(n) ทำได้โดยการกำหนดให้เครื่องหมายของพัลส์ที่ตำแหน่งใดๆ มีค่าเท่ากับเครื่องหมายของสัญญาณ d(n) ที่ตำแหน่งนั้น ดังนั้นจะมีขั้นตอนก่อนทำการหา codebook คือขั้นแรกสัญญาณ d(n) จะถูกแบ่งเป็นสองส่วนได้แก่ค่าสัมบูรณ์ $|d(n)|$ และเครื่องหมาย sign[d(n)] ขั้นที่สองเมทริกซ์ จะถูกปรับค่าโดยจะรวมเครื่องหมายของสัญญาณดังกล่าวเข้าไปด้วยดังนี้

$$\phi'(i,j) = sign[d(i)]sign[d(j)]\phi(i,j) \qquad i = 0,...39 \qquad j = i+1,...39 \tag{3.56}$$

ค่าในแนวทแยงมุมของ Φ จะถูกปรับขนาดโดยการหารสมการที่ (3.55)

$$\phi'(i,i) = 0.5\phi'(i,i) \qquad i = 0,...39 \tag{3.57}$$

การทำสหสัมพันธ์ในสมการ (3.54) หาได้จาก

$$C = |d(m_0)| + |d(m_1)| + |d(m_2)| + |d(m_3)| \tag{3.58}$$

และพลังงานในสมการที่ (3.55) กำหนดดังนี้

$$E/2 = \phi'(m_0,m_0) + \phi'(m_1,m_1) + \phi'(m_0,m_1) + \phi'(m_2,m_2) + \phi'(m_0,m_2) + \phi'(m_1,m_2) + \phi'(m_3,m_3) + \phi'(m_0,m_3) + \phi'(m_1,m_3) + \phi'(m_2,m_3) \quad (3.59)$$

ขอบเขตของการค้นหานั้น จะใช้วิธีการคำนวณหาค่า threshold ก่อนที่จะคำนวณในลูป และถ้าค่า threshold มีค่าเกินที่กำหนดก็จะทำการคำนวณ ค่าตัวเลขสูงสุดของลูปเวลาสามารถกำหนดให้มีเปอร์เซ็นต์ที่ต่ำในการค้นหาชุดรหัส ค่า threshold จะถูกคำนวณบนพื้นฐานของค่าสหสัมพันธ์ C และค่าสัมบูรณ์สูงสุดของค่าสหสัมพันธ์และค่าเฉลี่ยของสหสัมพันธ์เป็นตัวช่วยสำหรับสามพัลส์แรก โดยที่ค่า max_3 และ av_3 จะต้องหาก่อนที่จะทำการค้นหาชุดรหัส ค่า threshold สามารถหาได้ดังนี้

$$thr_3 = av_3 + K_3(\max_3 - av_3) \quad (3.60)$$

ลูปการคำนวณที่ 4 จะถูกคำนวณถ้าค่าสัมบูรณ์ของสหสัมพันธ์มีค่าเกินค่า thr_3 เมื่อ $0 \le k_3 < 1$ โดยที่ค่าของ k_3 จะเป็นตัวควบคุมเปอร์เซ็นต์ในการค้นหาและจะกำหนดให้ $k_3 = 0.4$ จะเห็นได้ว่าเวลาที่ใช้ในการค้นหาผลลัพธ์จะเปลี่ยนแปลงได้

3.3.8.2 การคำนวณการเข้ารหัสของ fixed-codebook

ตำแหน่งของพัลส์ i_0,i_1,i_2 จะเข้ารหัสแต่ละชุดขนาด 3 บิต และตำแหน่งของ i_3 จะเข้ารหัส 4 บิตและแต่ละพัลส์จะมีการเข้ารหัสของขนาดของพัลส์อีกพัลส์ละ 1 บิต ซึ่งรวมแล้วทั้งหมด 17 บิตสำหรับพัลส์ทั้งหมด 4 พัลส์ และกำหนดให้ s = 1 ถ้าเครื่องหมายเป็นบวกและ s = 0 ถ้าเครื่องหมายเป็นลบ และเครื่องหมายที่เข้ารหัสได้มาจาก

$$S = s_0 + 2s_1 + 4s_2 + 8s_3 \quad (3.61)$$

และการเข้ารหัสของ fixed-codebook ได้มาจาก

$$C = (m_0/5) + 8(m_1/5) + 64(m_2/5) + 512(2m_3/5) + jx) \quad (3.62)$$

และ jx = 0 ถ้า m_3 = 3,8,...38, และ jx = 1 ถ้า m_3 = 4,9,...,39

3.3.9 การควอนไทซ์อัตราขยาย

อัตราการขยายของ adaptive-codebook (อัตราขยายพิตช์) และอัตราขยายของ fixed codebook จะถูกควอนไทซ์แบบเวกเตอร์ที่มีขนาด 7 บิต การหาอัตราขยายของ codebook ทำได้โดยหาค่าต่ำสุดของ mean-squared weighted error ระหว่างสัญญาณเสียงต้นฉบับกับสัญญาณเสียงที่สร้างขึ้นมาดังนี้

$$E = x'x + g_p^2 y'y + g_c^2 z'z - 2g_p x'y - 2g_c x'z + 2g_p g_c y'z \tag{3.63}$$

เมื่อ x คือเวกเตอร์เป้าหมาย y คือเวกเตอร์ adaptive-codebook ที่ผ่านการกรองสัญญาณตามสมการที่ (3.44) และ z คือเวกเตอร์ fixed-codebook ที่คอนโวลูชันกับ h(n)

$$z(n) = \sum_{i=0}^{n} c(i)h(n-i) \qquad n = 0,...39 \tag{3.64}$$

3.3.9.1 **การทำนายอัตราขยาย**

อัตราขยายของ fixed-codebook g_c สามารถหาได้จาก

$$g_c = \gamma g'_c \tag{3.65}$$

เมื่อ g'_c คืออัตราขยายที่ทำนายได้โดยใช้พลังงานของ fixed-codebook ก่อนหน้า ส่วน γ คือตัวประกอบที่ใช้กำหนดค่าความถูกต้อง โดยพลังงานเฉลี่ยของ fixed-codebook มีค่าดังนี้

$$E = 10\log\left(\frac{1}{40}\sum_{n=0}^{39} c(n)^2\right) \tag{3.66}$$

หลังจากทำการปรับขนาดเวกเตอร์ c(n) ด้วยอัตราขยายของ fixed-codebook g_c แล้ว พลังงานของ fixed-codebook ที่ถูกปรับขนาดจะมีค่าเป็น 20 log g_c + E ให้ $E^{(m)}$ คือพลังงานเฉลี่ย(เป็น dB) ของ fixed-codebook ที่ถูกปรับขนาดในเฟรมย่อยแต่ละเฟรม m มีค่าดังนี้

$$E^{(m)} = 20\log g_c + E - \overline{E} \tag{3.67}$$

เมื่อ $\overline{E}$ = 30 dB คือพลังงานเฉลี่ยของสัญญาณกระตุ้นของ fixed-codebook ซึ่งอัตราขยายสามารถเขียนได้ในพจน์ของ $E^{(m)}$, E และ $\overline{E}$ ดังนี้

$$g_c = 10^{(E^{(m)} + \overline{E} - E)/20} \tag{3.68}$$

อัตราขยายที่ทำนายได้ g'_c จะหาได้จากการทำนายค่า log-energy ของ fixed-codebook เฟรมปัจจุบันเพิ่มเติมจาก log-energy ของ fixed-codebook ก่อนหน้า โดยใช้ตัวทำนายสัญญาณ MA ที่มีอันดับ 4 พลังงานที่ทำนายได้นั้นมีค่า

$$\tilde{E}^{(m)} = \sum_{i=1}^{4} b_i \hat{U}^{(m-i)} \tag{3.69}$$

เมื่อ $[b_1\ b_2\ b_3\ b_4] = [0.68\ 0.58\ 0.34\ 0.19]$ คือสัมประสิทธ์ของตัวทำนายสัญญาณ MA และ $\hat{U}^{(m-i)}$ คือค่าควอนไทซ์ของค่าผิดพลาดจากการทำนาย $U^{(m)}$ ที่เฟรมย่อย m มีค่าดังนี้

$$U^m = E^m - \tilde{E}^m \tag{3.70}$$

อัตราขยายที่ทำนายได้ g_c' จะหาได้โดยการแทนค่า E^m ด้วยค่าของตัวเองที่ทำนายได้จากสมการที่ (3.68) ดังนี้

$$g_c' = 10^{(\tilde{E}^{(m)}+\bar{E}-E)/20} \tag{3.71}$$

ค่าตัวประกอบที่กำหนดค่าความถูกต้อง γ จะสัมพันธ์กับค่าผิดพลาดที่ได้จากการทำนายอัตราขยายดังนี้

$$U^{(m)} = E^{(m)} - \tilde{E}^{(m)} = 20\log(\gamma) \tag{3.72}$$

3.3.9.2 การค้นหา codebook สำหรับการควอนไทซ์อัตราขยาย

อัตราขยายของ adaptive codebook g_p และตัวประกอบ γ คือ เวกเตอร์ที่ควอนไทซ์โดยใช้ codebook ที่มีโครงสร้างแบบสังยุคที่มี 2-stage ในสเตจแรกจะประกอบด้วย codebook gA ขนาด 3 บิต และในสเตจที่สองประกอบด้วย codebook gB ขนาด 4 บิต ค่าแรกใน codebook นั้นจะใช้แทนค่าอัตราขยายของ Adaptive codebook g_p ที่ควอนไทซ์แล้ว และค่าที่สองจะแทนอัตราขยายของ fixed-codebook g_c ที่ถูกควอนไทซ์ด้วยตัวประกอบที่ควบคุมความถูกต้อง γ แล้ว กำหนดให้ตัวชี้ codebook GA และ GB สำหรับ gA และ gB ตามลำดับ และอัตราขยายของ adaptive-codebook ที่ควอนไทซ์แล้วกำหนดได้ดังนี้

$$\hat{g}_p = gA_1(GA) + gB_1(GB) \tag{3.73}$$

และอัตราขยายของ fixed-codebook ที่ควอนไทซ์แล้วกำหนดดังนี้

$$\hat{g}_c = g_c'\hat{\gamma} = g_c'(gA_2(GA) + gB_2(GB)) \tag{3.74}$$

โครงสร้างที่เป็นคอนจูเกตนี้ช่วยทำให้การค้นหาเป็นไปโดยง่ายโดยการใช้วิธีการคัดเลือกล่วงหน้าในส่วน codebook gA จะมีทั้งหมด 8 เวกเตอร์ ในขั้นตอนการคัดเลือกล่วงหน้านั้นกลุ่มของเวกเตอร์ที่มีค่าเข้าใกล้ g_c จะถูกเลือกขึ้นมา 4 เวกเตอร์ ในทำนองเดียวกัน codebook gB ที่มีทั้งหมด 16 เวกเตอร์ จะได้กลุ่มของเวกเตอร์ 8 เวกเตอร์ที่มีค่าเข้าใกล้ g_p ถูกเลือกขึ้นมาเป็นผลให้ได้เวกเตอร์ที่ดีที่สุดมาครึ่งหนึ่ง จากนั้นจึงทำการหาเวกเตอร์ที่ดีที่สุดจาก 32 แบบที่

เป็นไปได้ทั้งหมด โดยมีเงื่อนไขของการเลือกเพื่อทำให้ weight mean squared error มีค่าต่ำสุด ตามสมการที่ (3.63)

3.3.10 การปรับให้ทันกาลในหน่วยความจำ

การปรับให้ทันกาลของค่าพารามิเตอร์ต่างๆ ของตัววงจรกรองสัญญาณที่ใช้สังเคราะห์เสียงนั้นจะทำการคำสวณหาสัญญาณของเป้า ในเฟรมย่อยถัดไป หลังจากทำการควอนไทซ์อัตราขยายทั้งสองแล้วจะได้สัญญาณกระตุ้น u(n) ของเฟรมย่อยปัจจุบันมีค่าดังนี้

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \qquad n = 0,...39 \tag{3.75}$$

เมื่อ g_p และ g_c คืออัตราขยายของ adaptive-codebook และ fixed-codebook ที่ควอนไทซ์แล้วตามลำดับ v(n) คือเวกเตอร์ของ adaptive-codebook และ c(n) คือเวกเตอร์ของ fixed-codebook ที่รวมส่วนขยายสัญญาณฮาร์มอนิกแล้ว วงจรกรองสัญญาณที่ใช้สังเคราะห์เสียงนั้นจะถูกปรับให้ทันกาลโดยการกรองสัญญาณ r(n) – u(n) (ผลต่างระหว่าง residual กับสัญญาณกระตุ้น) ผ่านวงจรกรองสัญญาณ 1/A(z) และวงจรกรองสัญญาณ $A(z/\gamma_1)/A(z/\gamma_2)$ ทุกๆ เฟรมย่อย (40 ตัวอย่างสัญญาณ) แล้วบันทึกค่าให้กับวงจรกรองสัญญาณที่ใช้สังเคราะห์เสียงต่อไป จะเห็นว่าจะต้องใช้วงจรกรองสัญญาณ 3 ชุด ในการทำงานนี้ วิธีที่ง่ายกว่านี้จะใช้วงจรกรองสัญญาณเพียงชุดเดียวสามารถทำได้โดยสร้างสัญญาณเสียงขึ้นมาใหม่ $\hat{s}(n)$ โดยคำนวณจากการกรองสัญญาณกระตุ้นด้วย 1/(A(z) ขาออกของวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์เสียงที่มีขาเข้าเป็น r(n)-u(n) มีค่าเทียบเท่ากับ $e(n) = s(n) - \hat{s}(n)$ การปรับให้ทันกาลพารามิเตอร์วงจรกรองสัญญาณ $A(z/\gamma_1)/A(z/\gamma_2)$ นั้นจะนำสัญญาณผิดพลาด e(n) มากรองผ่านวงจรกรองสัญญาณเพื่อหาค่า perceptually weighted error ew(n) อย่างไรก็ตามสัญญาณ ew(n) จะหาได้จากสมการดังนี้

$$ew(n) = x(n) - \hat{g}_p y(n) - \hat{g}_c z(n) \tag{3.76}$$

เมื่อหาค่าสัญญาณ x(n) , y(n) และ z(n) ได้ก็จะทำการปรับให้ทันกาลโดยการคำนวณหา ew(n) ตามสมการที่ (3.76) สำหรับ n = 30,...,39 ซึ่งจะลดการใช้วงจรกรองสัญญาณลงมา 2 ชุด

3.4 รายละเอียดการทำงานของตัวถอดรหัส CS-ACELP

บล็อกไดอะแกรมการทำงานของตัวถอดรหัสแบบ CS-ACELP แสดงดังรูปที่ 3.5 สามารถอธิบายการทำงานได้ดังนี้

ขั้นแรกพารามิเตอร์ที่ถูกส่งมาจะถูกถอดรหัสออกมาก่อน ซึ่งได้แก่ สัมประสิทธิ์ของตัวทำนายเชิงเส้น เวกเตอร์ adaptive-codebook เวกเตอร์ fixed-codebook และอัตราขยาย พารามิเตอร์ที่ส่งผ่านมานั้นแสดงดังในตารางที่ 3.2 พารามิเตอร์ที่ใช้ในตัวถอดรหัสนี้จะถูกคำนวณกลับขึ้นมาเพื่อนำมาใช้ในการสร้างสัญญาณเสียงขึ้นมา สัญญาณเสียงที่สร้างกลับขึ้นมานั้นจะถูกขยายโดยการทำ post processing ซึ่งประกอบด้วย postfilter วงจรกรองสัญญาณความถี่สูงผ่าน ตัวปรับขยายสัญญาณ

3.4.1 ขั้นตอนการถอดรหัสพารามิเตอร์

3.4.1.1 การถอดรหัสพารามิเตอร์ของวงจรกรองสัญญาณ LP

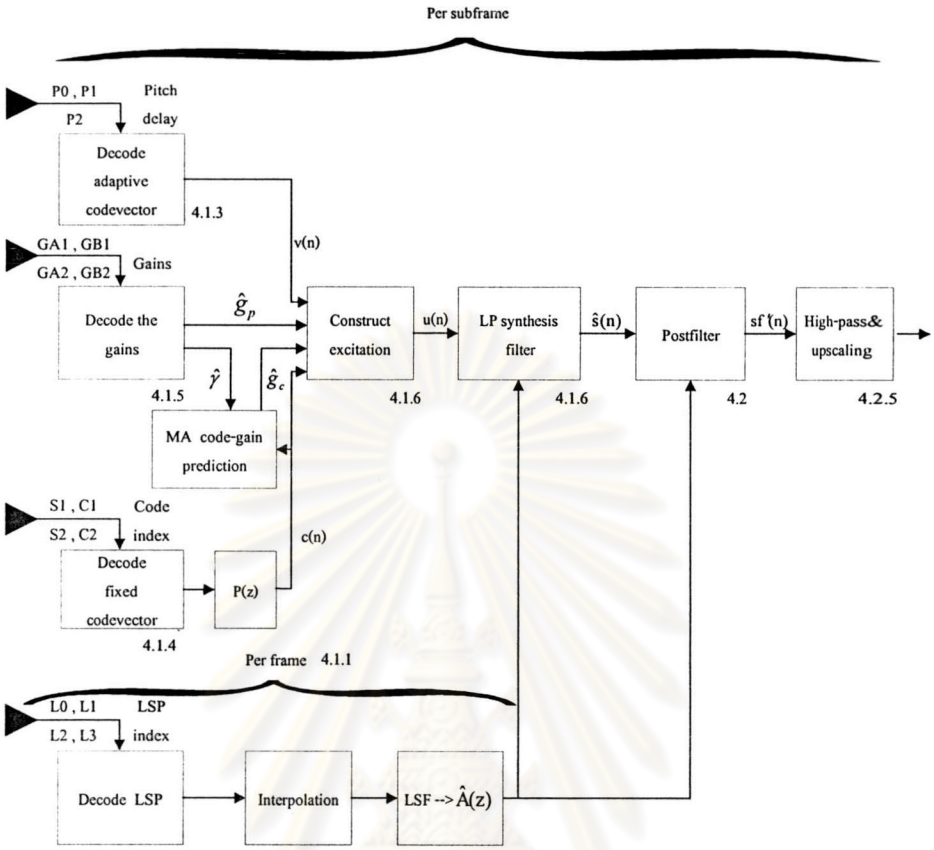
พารามิเตอร์ L0 L1 L2 L3 ที่รับได้นั้นจะใช้ในการสร้างสัมประสิทธิ์ LSP ที่ถูกควอนไทซ์และทำการประมาณค่าในช่วง 2 ชุด (ของเฟรมย่อยแต่ละเฟรม) ในแต่ละเฟรมย่อยแต่ละเฟรมนั้นสัมประสิทธิ์ LSP ที่มาจากการประมาณค่าในช่วงจะถูกเปลี่ยนเป็นสัมประสิทธิ์ตัวทำนายเชิงเส้น a_i ซึ่งถูกใช้การสังเคราะห์สัญญาณเสียงในเฟรมย่อยแต่ละเฟรม

ขั้นตอนที่เกิดขึ้นในแต่ละเฟรม

1. ถอดรหัสเวกเตอร์ของ adaptive-codebook
2. ถอดรหัสเวกเตอร์ของ fixed-codebook
3. ถอดรหัสอัตราขยายของ adaptive-codebook และ fixed-codebook
4. ทำการคำนวณเพื่อสร้างสัญญาณเสียงกลับคืน

3.4.1.2 การคำนวณหาพาริตี้

จะทำการสร้างสัญญาณกระตุ้นขึ้นมาก่อน และคำนวณหาพาริตีบิตจากพารามิเตอร์ P1 ของค่าการประวิงเวลาใน adaptive-codebook ซึ่งถ้าพาริตีบิตที่คำนวณขึ้นมาได้ไม่ตรงกับพาริตีบิต P0 ที่ส่งมาหมายความว่าเกิดการผิดพลาดในการส่งข้อมูล



รูปที่ 3.5 บล็อกไดอะแกรมการทำงานของตัวถอดรหัสแบบ CS-ACELP

3.4.1.3 การถอดรหัสเวกเตอร์ adaptive-codebook

ถ้าไม่เกิดข้อผิดพลาดจากการรับตัวชี้ adaptive-codebook P1 จะใช้ P1 ในการหาส่วนของจำนวนเต็มและส่วนของเศษของค่าประวิงเวลาพิตช์ T_1 ได้ดังนี้

$$
\begin{aligned}
&\textit{if } P1 < 197 \\
&\qquad \text{int}(T_1) = (P1+2)/3 + 19 \\
&\qquad \textit{frac} \quad = P1 - 3\,\text{int}(T_1) + 58 \\
&\textit{else} \\
&\qquad \text{int}(T_1) = P1 - 112 \\
&\qquad \textit{frac} \quad = 0 \\
&\textit{end}
\end{aligned}
$$

I21227263

และส่วนของจำนวนเต็มและส่วนของเศษ ของ T_2 จะได้มาจาก P2 และ t_{min} เมื่อ t_{min} ได้มาจาก T_1 ดังนี้

$$t_{\min} = \text{int}(T_1) - 5$$
$$if\ t_{\min} < 20\ then\ t_{\min} = 20$$
$$t_{\max} = t_{\min} + 9$$
$$if\ t_{\max} > 143\ then$$
$$t_{\max} = 143$$
$$t_{\min} = t_{\max} - 9$$
$$end$$

และ T_2 จะได้จาก

$$\text{int}(T_2) = (P2 + 2)/3 - 1 + t_{\min}$$
$$frac = P2 - 2 - 3((P2 + 2)/3 - 1)$$

เวกเตอร์ Adaptive-codebook v(n) จะหาได้จากการประมาณค่าในช่วงในส่วนของสัญญาณกระตุ้น u(n) ในอดีตที่เก็บไว้ในบัฟเฟอร์ โดยใช้สมการที่ (3.40)

3.4.1.4 การถอดรหัสเวกเตอร์ fixed-codebook

พารามิเตอร์ C ของ fixed-codebook ที่รับได้จะถูกใช้ในการหาตำแหน่งของพัลส์ของสัญญาณกระตุ้น ส่วนเครื่องหมายของพัลส์จะได้มาจากพารามิเตอร์ S ตำแหน่งและเครื่องหมายของพัลส์จะถูกถอดรหัสมาจากเวกเตอร์ fixed-codebook c(n) โดยใช้สมการที่ (3.45) ถ้าส่วนที่เป็นจำนวนเต็มของค่าการประวิงเวลา T มีค่าน้อยกว่าขนาดเฟรมย่อย 40 เวกเตอร์ c(n) จะถูกปรับเปลี่ยนตามสมการที่ (3.73)

3.4.1.5 การถอดรหัสอัตราขยายของ adaptive-codebook และ fixed-codebook

พารามิเตอร์ที่รับได้จะมีอัตราขยายของ adaptive-codebook g_p และตัวประกอบที่ใช้กำหนดความถูกต้องของอัตราขยายของ fixed-codebook γ การประมาณค่าอัตราขยาย fixed-codebook g'_c โดยใช้สมการที่ (3.71) เวกเตอร์ fixed-codebook จะได้มาจากผลคูณค่าตัวประกอบที่ใช้ควบคุมความถูกต้องด้วยอัตราขยายที่ทำนายได้ในสมการที่ (3.74) อัตราขยาย adaptive-codebook จะถูกสร้างกลับขึ้นมาใหม่โดยใช้สมการที่ (3.73)

ตารางที่ 3.3 รายละเอียดของพารามิเตอร์ที่ได้จากการถอดรหัส (บิต MSB ถูกส่งมาก่อน)

สัญลักษณ์	รายละเอียด	จำนวนบิต
L0	ตัวทำนายสัญญาณ MA ตัวชี้ควอนไทเซอร์ LSP	1
L1	เวกเตอร์ของควอนไทเซอร์ในสเตจแรก	7
L2	เวกเตอร์ส่วนล่างของควอนไทเซอร์ LSP ในสเตจที่สอง	5
L3	เวกเตอร์ส่วนบนของควอนไทเซอร์ LSP ในสเตจที่สอง	5
P1	การประวิงเวลาของพิตช์ในเฟรมย่อยแรก	8
P0	พาริตีบิตของค่าการประวิงเวลาของพิตช์	1
C1	fixed-codebook ในเฟรมย่อยแรก	13
S1	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยแรก	4
GA1	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยแรก	3
GB1	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยแรก	4
P2	การประวิงเวลาของพิตช์ เฟรมย่อยที่สอง	5
C2	Fixed-codebook ในเฟรมย่อยที่สอง	13
S2	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยที่สอง	4
GA2	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยที่สอง	3
GB2	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยที่สอง	4
จำนวนบิตรวม		80

3.4.1.6 การสังเคราะห์สัญญาณเสียง

สัญญาณกระตุ้น u(n) (สมการที่ 3.75) เป็นสัญญาณอินพุตที่ป้อนให้กับวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์หรือตัวทำนายเชิงเส้น จะทำให้สามารถสร้างสัญญาณเสียงกลับขึ้นมาในเฟรมย่อยแต่ละเฟรมดังนี้

$$\hat{s}(n) = u(n) - \sum_{i=1}^{10} \hat{a}_i \hat{s}(n-i) \qquad n = 0,...39 \tag{3.77}$$

เมื่อ a_i คือ สัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้นของเฟรมย่อยปัจจุบัน สัญญาณเสียงที่สร้างกลับขึ้นมา $\hat{s}(n)$ จะถูกป้อนให้กับขั้นตอนสุดท้ายต่อไป

3.4.2 Post-precessing

ส่วนของการประมวลผลขึ้นสุดท้ายนั้นประกอบด้วยฟังก์ชัน 3 ฟังก์ชัน คือ adaptive postfiltering วงจรกรองสัญญาณความถี่สูงผ่านและตัวปรับขนาดสัญญาณขึ้น ส่วนของ adaptive ส่วนของ adaptive post filtering มาจากการต่อคาสเคดของวงจรกรองสัญญาณ 3 วงจรคือ long-term postfilter $H_p(z)$ short-term postfilter $H_f(z)$ และ tilt compensation filter $H_t(z)$ และมีส่วนการควบคุมอัตราการขยายปรับค่าได้เป็นส่วนสุดท้าย สัมประสิทธิ์ของ postfilter จะถูกปรับให้ทันกาลทุกๆ 5 มิลลิวินาที (เฟรมย่อย) ขั้นตอนในการทำ postfiltering นั้นจะทำโดยนำสัญญาณเสียง $\hat{s}(n)$ มาหาส่วนกลับของการกรองผ่าน $A(z/\gamma_n)$ เพื่อสร้างสัญญาณ residual r(n) สัญญาณนี้จะถูกใช้ในการคำนวณหาการประวิงเวลา T และ อัตราการขยาย g_t ของ long-term postfilter $H_p(z)$ และสัญญาณ r(n) จะถูกกรองผ่าน long-term postfilter $H_p(z)$ และวงจรกรองสัญญาณสังเคราะห์ $1/[g_f\hat{A}(z/\gamma_d)]$ สุดท้ายสัญญาณออกของวงจรกรองสัญญาณสังเคราะห์จะถูกป้อนผ่าน tilt compensation filter $H_t(z)$ เพื่อสร้างสัญญาณเสียง sf(n) เพื่อให้พลังงานมีค่าเท่ากับ s(n) จะได้สัญญาณ $sf'(n)$ ป้อนผ่านวงจรกรองสัญญาณความถี่สูงผ่านและวงจรปรับขยายสัญญาณและได้สัญญาณเสียงที่ต้องการออกมา

3.4.2.1 Long-term postfilter

Long-term postfilter กำหนดให้

$$H_p(z) = \frac{1}{1+\gamma_p gl}(1+\gamma_p gl z^{-T}) \tag{3.78}$$

เมื่อ T คือ ค่าการประวิงเวลาของพิตช์และ gl คือสัมประสิทธิ์ของอัตราขยายซึ่งมีค่าไม่เกิน 1 และจะมีค่าเป็นศูนย์ถ้าอัตราขยายของการทำนายสัญญาณของ long-term นี้มีค่าน้อยกว่า 3 dB ตัวประกอบ γ_p จะควบคุมปริมาณการทำ long-term postfiltering และมีค่า γ_p = 0.5 การประวิงเวลาและอัตราขยายของ long-term จะถูกคำนวณมาจากสัญญาณ residual r(n) โดยการกรองสัญญาณเสียง $\hat{s}(n)$ ผ่าน $A(z/\gamma_n)$ ซึ่งเป็น short-term postfilter

$$\hat{r}(n) = \hat{s}(n) + \sum_{i=1}^{10} \gamma_n^i \hat{a}_i \hat{s}(n-i) \tag{3.79}$$

การประวิงเวลาของ long-term คำนวณได้จากสองขั้นตอน ในขั้นแรกจะทำการเลือกจำนวนเต็ม T_0 ที่ดีที่สุดในช่วง [int(T_1)-1 , int(T_1)+1] เมื่อ int(T_1) คือส่วนของจำนวนเต็มจากค่าการประวิงเวลาของพิตช์ T_1 ของเฟรมย่อยแรกที่ส่งมา ซึ่งค่านี้จะทำให้สหสัมพันธ์ที่ได้มีค่าสูงสุดดังนี้

$$R(k) = \sum_{n=0}^{39} \hat{r}(n)\hat{r}(n-k) \tag{3.80}$$

ในขั้นที่สองจะเลือกเศษส่วนของค่าการประวิงเวลา T ที่ดีที่สุดที่มีความละเอียด 1/8 รอบๆ ค่า T_0 การเลือกค่านี้ทำได้โดยการทำให้เป็นบรรทัดฐานด้วยสหสัมพันธ์สูงสุดดังนี้

$$R'(k) = \frac{\sum_{n=0}^{39} \hat{r}(n)\hat{r}_k(n)}{\sqrt{\sum_{n=0}^{39} \hat{r}_k(n)\hat{r}_k(n)}} \tag{3.81}$$

เมื่อ $\hat{r}_k(n)$ คือ สัญญาณ residual ที่ประวิงเวลา k และสามารถหาค่าประวิงเวลาที่ดีที่สุด T ได้จากนั้น สหสัมพันธ์ $R'(k)$ จะถูกทำให้เป็นบรรทัดฐานด้วยรากที่สองของพลังงานของ $\hat{r}(n)$ ค่ายกกำลังสองของสหสัมพันธ์จะถูกใช้ในการหา ถ้า long-term filter ไม่สามารถทำได้ โดยการกำหนดให้ gl = 0 ถ้า

$$\frac{R'(T)^2}{\sum_{n=0}^{39} \hat{r}(n)\hat{r}(n)} < 0.5 \tag{3.82}$$

นอกจากนั้นค่าของ gl คำนวณได้จาก

$$gl = \frac{\sum_{n=0}^{39} \hat{r}(n)\hat{r}_k(n)}{\sum_{n=0}^{39} \hat{r}_k(n)\hat{r}_k(n)} \qquad 0 \le gl \le 1.0 \tag{3.83}$$

สัญญาณ $\hat{r}_k(n)$ จะถูกคำนวณหาก่อนโดยใช้วงจรกรองสัญญาณประมาณค่าในช่วงที่มีความยาว 33 หลังจากเลือกค่า T แล้ว $\hat{r}_k(n)$ จะถูกคำนวณอีกครั้งด้วยวงจรกรองสัญญาณการประมาณค่าในช่วงที่มีความยาวมากขึ้นเป็น 129 สัญญาณที่คำนวณได้ใหม่นี้จะไปแทนที่สัญญาณเก่า ถ้า $R'(T)$ มีค่าเพิ่มขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.4.2.2 Short-term postfilter

short-term postfilter มีฟังก์ชันดังนี้

$$H_f(z) = \frac{1}{gf}\frac{\hat{A}(z/\gamma_n)}{\hat{A}(z/\gamma_d)} = \frac{1}{gf}\frac{1+\sum_{i=1}^{10}\gamma_n^i \hat{a}_i z^{-i}}{1+\sum_{i=1}^{10}\gamma_d^i \hat{a}_i z^{-i}} \tag{3.84}$$

เมื่อ $\hat{A}(z)$ คือส่วนกลับของวงจรกรองตัวทำนายเชิงเส้น ที่มีค่าสัมประสิทธิ์ถูกควอนไทซ์ที่คำนวณได้จากพารามิเตอร์ที่รับได้ (การวิเคราะห์สัมประสิทธิ์ตัวทำนายเชิงเส้น ไม่ได้ทำที่ตัวถอดรหัส) และตัวประกอบ γ_n และ γ_d จะควบคุมขนาดของการทำ short-term postfiltering และมีค่า γ_n = 0.55 และ γ_d = 0.7 อัตราขยาย g_f จะคำนวณโดยใช้การตอบสนองอิมพัลส์ $h_f(n)$ ของวงจรกรองสัญญาณ $A(z/\gamma_n)/A(z/\gamma_d)$ ดังนี้

$$g_f = \sum_{n=0}^{19}\left|h_f(n)\right| \tag{3.85}$$

3.4.2.3 Tilt compensation

วงจรกรองสัญญาณที่ใช้ชดเชย tilt คือ $H_t(z)$ ซึ่งเกิดจาก short-term postfilter จะมีค่าดังนี้

$$H_t(z) = \frac{1}{g_t}(1+\gamma_t k_1' z^{-1}) \tag{3.86}$$

เมื่อ $\gamma_t k_t'$ คือตัวประกอบ tilt ซึ่ง k_1' จะเป็นสัมประสิทธิ์ reflection ตัวแรกและคำนวณได้จาก $h_f(n)$

$$k_1' = -\frac{r_h(1)}{r_h(0)} \qquad r_h(i) = \sum_{j=0}^{19-i} h_f(j)h_f(j+i) \tag{3.87}$$

พจน์อัตราขยาย $g_t = 1-\left|\gamma_t k_1'\right|$ จะชดเชยการลดลงเนื่องจากผลกระทบของ g_f ใน $H_t(z)$ ค่า γ_t นี้จะมีค่าขึ้นอยู่กับเครื่องหมายของ k_1' กล่าวคือถ้า k_1' มีค่าเป็นลบจะได้ $\gamma_t = 0.9$ และถ้า k_1' มีค่าเป็นบวกจะได้ γ_t = 0.2

3.4.2.4 การควบคุมอัตราขยายแบบปรับค่าได้

การควบคุมอัตราขยายแบบปรับค่าได้ (adaptive gain control) ถูกใช้ในการชดเชยความแตกต่างระหว่างขนาดสัญญาณเสียงที่สร้างกลับขึ้นมา $\hat{s}(n)$ กับสัญญาณที่ได้จาก

postfilter $sf(n)$ ค่าตัวประกอบ G ที่ใช้ควบคุมปรับอัตราขยายในเฟรมย่อยปัจจุบันนั้นคำนวณได้จาก

$$G = \frac{\sum_{n=0}^{39} |\hat{s}(n)|}{\sum_{n=0}^{39} |sf(n)|} \tag{3.88}$$

สัญญาณที่ผ่านการควบคุม $sf'(n)$ แล้วจะมีค่า

$$sf'(n) = g^{(n)} sf(n) \qquad n = 0,...39 \tag{3.89}$$

เมื่อ $g^{(n)}$ จะถูกปรับให้ทันกาลตัวอย่างต่อตัวอย่างดังนี้

$$g^{(n)} = 0.85 g^{(n-1)} + 0.15G \qquad n = 0,...39 \tag{3.90}$$

โดยค่าเริ่มต้น $g^{(-1)} = 1.0$ โดยในเฟรมย่อยถัดไปจะกำหนดให้ $g^{(-1)} = g^{(39)}$ ของเฟรมก่อนหน้า

3.4.2.5 การกรองสัญญาณความถี่สูงผ่านและปรับขยายขนาด

วงจรกรองสัญญาณความถี่สูงผ่านที่มีความถี่ตัด 100 Hz จะถูกใช้กับสัญญาณเสียงที่สร้างกลับขึ้นมาใหม่ที่ได้จากการทำ postfilter $sf'(n)$ ซึ่งมีคุณสมบัติดังนี้

$$H_{h2}(z) = \frac{0.93980581 - 1.8795834z^{-1} + 0.93980581z^{-2}}{1 - 1.9330735z^{-1} + 0.93589199z^{-2}} \tag{3.91}$$

สัญญาณที่ได้จากวงจรกรองสัญญาณนี้จะมีอัตราขยาย 2 เท่าเพื่อชดเชยที่เกิดจากตัวเข้ารหัส

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย