การบีบอัดเสียงพูดภาษาไทยโดยใช้การเข้ารหัส MP-CELP
ตามข้อกำหนดของ MPEG-4

นายสุภัทรชัย ชมพันธุ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2543

ISBN 974-13-0048-4

THAI SPEECH COMPRESSION BASED ON MP-CELP ACCORDING TO MPEG-4 REQUIREMENTS

MR. SUPATTARACHAI CHOMPUN

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

Academic Year 2000
ISBN 974-13-0048-4

หัวข้อวิทยานิพนธ์	การบีบอัดเสียงพูดภาษาไทยโดยใช้การเข้ารหัส MP-CELP ตามข้อกำหนดของ MPEG-4
โดย	นาย สุภัทรชัย ชมพันธุ์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์

…… ประธานกรรมการ
(ศาสตราจารย์ ดร.ประสิทธิ์ ประพิณมงคลการ)

……... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)

……กรรมการ
(ดร.เสถียร เตรียมล้ำเลิศ)

สุภัทรชัย ชมพันธุ์ : การบีบอัดเสียงพูดภาษาไทยโดยใช้การเข้ารหัส MP-CELP ตามข้อกำหนดของ MPEG-4. (THAI SPEECH COMPRESSION BASED ON MP-CELP ACCORDING TO MPEG-4 REQUIREMENTS) อ.ที่ปรึกษา: รศ. ดร.สมชาย จิตะพันธ์กุล 126 หน้า. ISBN 974-13-0048-4.

วิทยานิพนธ์นี้นำเสนอการเข้ารหัสเสียงพูดภาษาไทย ตามข้อกำหนดของมาตรฐานการเข้ารหัสเสียงพูดธรรมชาติ MPEG-4 (Moving Picture Expert Group - 4) หลักการเข้ารหัสนี้อยู่บนพื้นฐานของการเข้ารหัสด้วยวิธี MP-CELP (Multi Pulse-based Code-Excited Linear Prediction) โดยสมบัติของ MP-CELP เอง และการเพิ่มส่วนขยายเข้าไป ทำให้สามารถรองรับการทำงานที่หลายอัตราการเข้ารหัส (Multiple Bitrate) และสามารถปรับระดับอัตราการเข้ารหัสได้ (Bitrate Scalability) ตามลำดับ

เทคนิคการวิเคราะห์พิตช์ด้วยความละเอียดสูง (High Pitch Delay Resolution technique) ที่ระดับความละเอียด 1/2 1/3 และ 1/4 ถูกนำเสนอและประยุกต์ใช้เพื่อปรับปรุงการเข้ารหัสเสียงพูดภาษาไทยด้วยวิธี MP-CELP สำหรับการวิเคราะห์พิตช์ดีเลย์

การเข้ารหัสเสียงพูดที่จำลองขึ้น สามารถปรับปรุงคุณภาพเสียงพูดภาษาไทยให้อยู่ในระดับที่เท่าเทียมกับเสียงพูดภาษาอังกฤษ ด้วยอัตราการเข้ารหัสที่เพิ่มขึ้น 200-400 bps คือจากเดิม 5,600-14,600 bps เป็น 5,800-15,000 bps สำหรับส่งข้อมูลเศษส่วนพิตช์ เทียบได้กับอัตราบีบอัด 4.27-11.03 เท่า ส่วนเวลาประวิงจะเท่ากับมาตรฐานการเข้ารหัส ITU G.729 คือ 15 มิลลิวินาที

ภาควิชา วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต.....
สาขาวิชา วิศวกรรมไฟฟ้า	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2543	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4170597021 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: MP-CELP / MPEG-4 / SPEECH CODING / THAI SPEECH / HPDR

SUPATTARACHAI CHOMPUN : THAI SPEECH COMPRESSION BASED ON MP-CELP ACCORDING TO MPEG-4 REQUIREMENTS. THESIS ADVISOR : ASSOC. PROF. DR. SOMCHAI JITAPUNKUL, Dr. Ing, 126 pp. ISBN 974-13-0048-4.

This thesis proposes Thai speech coding according to the natural speech coding of MPEG-4 standards. The operation principle of this codec is based on the MP-CELP coding. By the MP-CELP's attributes and embedding enhancement layers, it can support the special functionalities of multiple bitrates and bitrate scalabilities.

In the pitch delay analysis, high pitch delay resolution technique of 1/2, 1/3 and 1/4 pitch fractions is proposed and adopted to improve Thai speech MP-CELP coding quality.

By simulating the proposed codec, the results show improvement of Thai speech quality, nearly equivalent to that of English. The operating bitrates are increased by 200-400 bps for the additional pitch fraction information from 5,600-14,600 bps to 5,800-15,000 bps corresponding to the compression ratio of 4.27-11.03, while the coding delay of 15 ms is equal to that of the ITU G.729 stardard.

Department Electrical Engineering Student's signature .

Field of Study Electrical Engineering Advisor's signature .

Academic year 2000 Co-advisor's signature .

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งได้ให้คำแนะนำ ข้อคิดเห็น และสนับสนุนอุปกรณ์เครื่องมือต่างๆ ในการทำวิจัยมาด้วยดีตลอด ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

ขอขอบคุณเพื่อนพี่น้องนิสิตที่อยู่ภายในห้องปฏิบัติการวิจัยกรรมวิธีสัญญาณดิจิทัล (Digital Signal Processing Research Laboratory) ที่ได้ช่วยเหลือเกี่ยวกับข้อมูลเสียงพูด การประเมินผล รวมถึงคำแนะนำ ตลอดระยะเวลาการทำวิจัยอย่างดียิ่ง

นอกจากนี้ขอขอบคุณเพื่อนพี่น้องนิสิต ที่อยู่ภายในห้องปฏิบัติการวิจัยระบบโทรคมนาคม ที่ได้ช่วยเหลือในส่วนของประเมินผลการทดลอง และเป็นกำลังใจที่ดียิ่งต่อผู้วิจัย

และผู้วิจัยต้องขอขอบคุณสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่ได้สนับสนุนทุนการศึกษาให้ผู้วิจัยตามโครงการ Telecommunication Consortium

ท้ายนี้ผู้วิจัยขอกราบขอบพระคุณบิดามารดา ที่ให้การสนับสนุนแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สารบัญ

สารบัญ (ต่อ)

บทที่ หน้า

สารบัญ (ต่อ)

สารบัญ (ต่อ)

สารบัญ (ต่อ)

สารบัญตาราง

หน้า

สารบัญตาราง (ต่อ)

หน้า

สารบัญภาพ

หน้า

สารบัญภาพ (ต่อ)

หน้า

สารบัญภาพ (ต่อ)

หน้า

สารบัญภาพ (ต่อ)

หน้า

สารบัญภาพ (ต่อ)

หน้า

บัญชีสัญลักษณ์

$1/A(z)$	วงจรกรองสัญญาณสังเคราะห์เสียง LP
$H_{h1}(z)$	วงจรกรองสัญญาณแบบความถี่สูงผ่านด้านเข้า
$H_p(z)$	postfilter ส่วน long-term
$H_f(z)$	postfilter ส่วน short-term
$H_t(z)$	วงจรกรอง tilt-compensation
$H_{h2}(z)$	วงจรกรองสัญญาณแบบความถี่สูงผ่านด้านออก
$P(z)$	วงจรกรองสำหรับ fixed-codebook
$W(z)$	วงจรกรอง weighting
c(n)	fixed-codebook
d(n)	คอร์รีเลชันระหว่างสัญญาณของเป้า กับ h(n)
ew(n)	สัญญาณความผิดพลาด
h(n)	การตอบสนองอิมพัลส์ของวงจรกรอง weighting and synthesis
r(n)	สัญญาณ residual
s(n)	สัญญาณเสียงก่อนวงจร pre-processing
$\hat{s}(n)$	สัญญาณเสียงที่สังเคราะห์กลับขึ้นมา
$s'(n)$	สัญญาณเสียงในช่วงที่กำหนด
sf(n)	ขาออกที่ได้จาก postfilter
$sf'(n)$	ขาออกที่ได้จาก postfilter ที่ปรับขนาดสัญญาณแล้ว
sw(n)	สัญญาณเสียงที่ weighted แล้ว
x(n)	สัญญาณของเป้า
$x'(n)$	สัญญาณของเป้าที่สอง
u(n)	เอ็กไซเทชันของการสังเคราะห์เสียงด้วย LP
v(n)	adaptive-codebook
y(n)	คอนโวลูชัน v(n)*h(n)
z(n)	คอนโวลูชัน c(n)*h(n)
g_p	อัตราขยาย adaptive-codebook
g_c	อัตราขยาย fixed-codebook
g_l	เทอมอัตราขยายของ long-term posfilter
g_f	เทอมอัตราขยายของ short-term posfilter
g_t	เทอมอัตราขยายของ tilt postfilter

บัญชีสัญลักษณ์ (ต่อ)

G	อัตราขยายที่ถูกนอมอลไลซ์
T_{op}	open-loop pitch delay
a_i	สัมประสิทธิ์ของ LP ($a_0 = 1.0$)
k_i	สัมประสิทธิ์ reflection coefficient
k'_1	reflection coefficient ของ tilt postfilter
o_i	สัมประสิทธิ์ LAR
ω_i	ความถี่นอมอลไลซ์ LSF
$p_{i,j}$	ตัวทำนาย MA สำหรับการควอนไทซ์ LSF
q_i	สัมประสิทธิ์ LSP
r(k)	สัมประสิทธิ์ออโต้คอร์รีเลชัน
r'(k)	สัมประสิทธิ์ออโต้คอร์รีเลชันที่โมดิฟายด์
w_i	สัมประสิทธิ์ตัวถ่วง LSP
l_i	ขาออกจากตัวควอนไทซ์ LSP
f_s	อัตราสุ่ม
BW	แบนด์วิดธ์
γ_1	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง perceptual weighting
γ_2	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง perceptual weighting
γ_n	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง post filter
γ_d	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง post filter
γ_p	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง pitch post filter
γ_t	สัมประสิทธิ์ตัวถ่วงของวงจรกรอง tilt post filter
C	fixed-codebook (algebraic)
$L0$	codebook ของตัวทำนาย MA
$L1$	codebook ของ LSP ในสเตจที่ 1
$L2$	codebook ของ LSP ในสเตจที่ 2 (low part)
$L3$	codebook ของ LSP ในสเตจที่ 2 (high part)
gA	gain codebook (สเตจที่ 1)
gB	gain codebook (สเตจที่ 2)
w_{lag}	correlation lag window
w_{lp}	LP analysis window

บัญชีสัญลักษณ์ (ต่อ)

a(n)	ค่าถ่วงน้ำหนัก ในการคำนวณสัญญาณกระตุ้นแบบปรับตัว
b(n)	ค่าถ่วงน้ำหนัก ในการคำนวณสัญญาณกระตุ้นแบบปรับตัว ในเทคนิค HPDR
w(n)	Hamming window function

บทที่ 1

บทนำ

การสื่อสารโทรคมนาคมในปัจจุบันและอนาคตอันใกล้ ได้ให้ความสำคัญกับการส่งข่าวสารผ่านช่องสัญญาณที่มีขนาดจำกัด อย่างมีประสิทธิภาพ ข่าวสารเหล่านี้อยู่ในรูปของสัญญาณเสียง ภาพนิ่ง ภาพเคลื่อนไหว หรือข้อมูล เป็นต้น ปริมาณความต้องการการเข้าใช้ของผู้บริโภคมีแนวโน้มสูงขึ้น ส่งผลให้ขนาดของข่าวสารเพิ่มขึ้นอย่างมาก เพื่อตอบสนองความต้องการดังกล่าวจึงจำเป็นต้องมีการเพิ่มความจุของช่องสัญญาณ และวิธีการหนึ่งในการเพิ่มความจุของช่องสัญญาณคือการบีบอัดสัญญาณ เพื่อให้สามารถใช้งานช่องสัญญาณที่มีแบนด์วิดท์จำกัดได้อย่างมีประสิทธิภาพ

การเข้ารหัส หรือการบีบอัดสัญญาณเสียงมีบทบาทสำคัญต่อการออกแบบระบบโทรศัพท์เคลื่อนที่ ทั้งนี้เพราะแบนด์วิดท์ที่จัดสรรให้กับการใช้งานของโทรศัพท์เคลื่อนที่มีขนาดจำกัด จึงจำเป็นต้องใช้งานช่องสัญญาณที่มีอยู่ให้เกิดประโยชน์สูงสุด การเข้ารหัสจึงต้องคำนึงถึงอัตราบิตต่ำและยังคงรักษาคุณภาพเสียงในระดับที่ดีพอสำหรับการใช้งาน การที่จะตอบสนองความต้องการทั้งสองจึงเป็นปัญหาสำคัญมากในการพัฒนาตัวเข้ารหัสเสียงพูด [1]

ในปี ค.ศ.1995 กลุ่มทำงานใน ISO ในนามของ Moving Picture Experts Group (MPEG) ได้เริ่มดำเนินงานในการกำหนดมาตรฐานเครื่องมือการเข้ารหัสที่มีอัตราการเข้ารหัสต่ำ เพื่อใช้งานในเครือข่ายอินเตอร์เน็ต และระบบสื่อสารเคลื่อนที่ ช่องสัญญาณส่งผ่านมีแบนด์วิดท์จำกัด โครงการนี้รู้จักกันในนามของ MPEG-4 ซึ่งถูกกำหนดเป็นมาตรฐานสากล ISO14496 ในปี ค.ศ.1999 [2 3 และ 4]

มาตรฐาน MPEG-4 จะพิจารณาสัญญาณเสียงและภาพเป็นแบบ object-based เพื่อรองรับสื่อประสม เช่น รายการภาพยนตร์ รายการวิทยุ และถูกส่งในลักษณะของ media object วิธีการเข้ารหัสสำหรับแต่ละชนิดของวัตถุถูกกำหนดด้วยมาตรฐาน MPEG-4 audio และ MPEG-4 video เมื่อแต่ละวัตถุถูกส่งไปยังปลายทาง ชิ้นส่วนต่างๆ จะถูกนำมาประกอบรวมกันเป็น audio visual scene โดยมีรูปแบบเฉพาะสำหรับการจัดการ [3 5 และ 6]

วิทยานิพนธ์นี้จะเน้นเฉพาะการเข้ารหัสเสียงพูดธรรมชาติแถบแคบ ตามข้อกำหนดของมาตรฐาน MPEG-4 audio และปรับปรุงการเข้ารหัสให้เหมาะสมกับเสียงพูดภาษาไทย ที่เป็นเสียงดนตรี (Tonal Language)

ในบทนี้จะกล่าวถึงวัตถุประสงค์ ขอบเขตของวิทยานิพนธ์ ขั้นตอนการดำเนินงาน ประโยชน์ที่ได้รับจากวิทยานิพนธ์ ภาพรวมของเนื้อหาในแต่ละบทของวิทยานิพนธ์

บทที่ 2 จะกล่าวถึงการพัฒนาปรับปรุงการเข้ารหัสเสียงพูดของมาตรฐาน ITU จากอดีตจนถึงปัจจุบัน การวัดสมรรถนะของการเข้ารหัสเสียงพูด การเปรียบเทียบสมรรถนะการเข้ารหัสเสียงพูดโดยวิธีต่างๆ การเข้ารหัสเสียงพูดโดยวิธี CS-ACELP ตามมาตรฐาน ITU G.729 ซึ่งเป็นการเข้ารหัสเสียงพูดมาตรฐานที่ใช้อ้างอิงในวิทยานิพนธ์นี้ และสุดท้ายคือข้อกำหนดการเข้ารหัสเสียงพูดของมาตรฐาน MPEG-4

บทที่ 3 จะกล่าวถึงหลักการทำงานโดยรวมของตัวเข้ารหัสหลักและตัวถอดรหัส MP-CELP ที่อัตราการเข้ารหัส 3 ค่า คือ 5600 8200 และ 12200 bps การปรับระดับอัตราการเข้ารหัส ทั้งนี้ตามข้อกำหนดของมาตรฐาน MPEG-4 คือสามารถปรับระดับอัตราการเข้ารหัสเพิ่มขึ้น 3 ขั้น ขั้นละ 800 bps ความแตกต่างระหว่างคุณลักษณะของเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ และสุดท้ายคือการปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ให้เหมาะสมกับเสียงพูดภาษาไทย

บทที่ 4 นำเสนอผลการทดลองการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ออกเป็น 2 ขั้นตอน ขั้นตอนแรกคือการทำการทดลองเพื่อเปรียบเทียบคุณภาพเสียงที่ผ่านการเข้ารหัสโดยวิธี MP-CELP ระหว่างภาษาไทยกับภาษาอังกฤษ จากนั้นทำการทดลองเพื่อเพิ่มประสิทธิภาพการเข้ารหัสเสียงพูดภาษาไทยโดยการเพิ่มความละเอียดของค่าพิตช์

บทที่ 5 กล่าวถึงสรุปผลการวิจัย และข้อเสนอแนะสำหรับการวิจัยในอนาคต

วัตถุประสงค์ของวิทยานิพนธ์

1. เพื่อพัฒนากรรมวิธีเข้ารหัสสัญญาณเสียงพูดด้วยอัตราบิบอัดสูงๆ อย่างมีประสิทธิภาพ
2. เพื่อพัฒนากรรมวิธีเข้ารหัสสัญญาณเสียงพูดภาษาไทยสำหรับอัตราบิตต่ำ

ขอบเขตของวิทยานิพนธ์

พัฒนาโปรแกรมบีบอัดเสียงพูด จากระบบการบีบอัดเสียงพูดที่ออกแบบขึ้น ให้สอดคล้องกับข้อกำหนดของ MPEG-4 และสามารถบีบอัดเสียงพูดภาษาไทย ให้มีคุณภาพเทียบเท่าการเข้ารหัสตามมาตรฐาน ITU G.729

วิธีการดำเนินงาน

1. ศึกษากรรมวิธีเข้ารหัสสัญญาณเสียงพูดแบบต่างๆ เพื่อหาจุดอ่อน
2. ศึกษาแนวคิด ทฤษฎี และการประยุกต์ใช้ของการเข้ารหัส MP-CELP

3. จำลองการเข้ารหัส MP-CELP รวมถึงเพิ่มหน้าที่การทำงาน เพื่อให้สอดคล้องกับข้อกำหนดของ MPEG-4

4. พัฒนาโปรแกรมเพื่อให้เหมาะสมกับการบีบอัดเสียงพูดภาษาไทย

5. เปรียบเทียบผลของการบีบอัดเสียงพูดภาษาไทยที่ใช้โปรแกรมที่พัฒนาขึ้น กับการเข้ารหัสมาตรฐานของ ITU G.729

6. ทดสอบและแก้ไขกรรมวิธีที่พัฒนาขึ้น

7. สรุปผล วิจารณ์ผล และเขียนวิทยานิพนธ์

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้เรียนรู้ทฤษฎีการเข้ารหัสแบบ MP-CELP ในการประมวลผลสัญญาณดิจิทัล

2. ได้เรียนรู้วิธีการบีบอัดเสียงพูด และข้อจำกัดต่างๆ ในการบีบอัดเสียงพูด

3. โปรแกรมการบีบอัดเสียงพูดที่พัฒนาขึ้นสามารถในไปใช้ได้จริง หรือนำไปพัฒนาต่อเพื่อให้ใช้งานได้จริง

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้ จะกล่าวถึง การพัฒนาปรับปรุงการเข้ารหัสเสียงพูดของมาตรฐาน ITU จากอดีตจนถึงปัจจุบัน การวัดสมรรถนะของการเข้ารหัสเสียงพูด การเปรียบเทียบสมรรถนะการเข้ารหัสเสียงพูดโดยวิธีต่างๆ การเข้ารหัสเสียงพูดโดยวิธี CS-ACELP (Conjugate Structure Algebraic Code Excited Linear Prediction) ตามมาตรฐาน ITU G.729 ซึ่งเป็นการเข้ารหัสเสียงพูดมาตรฐานที่ใช้อ้างอิงในวิทยานิพนธ์นี้ และสุดท้ายคือข้อกำหนดการเข้ารหัสเสียงพูดของมาตรฐาน MPEG-4

2.1 การพัฒนาปรับปรุงการเข้ารหัสเสียงพูดของมาตรฐาน ITU

ในการเข้ารหัสเสียงพูดที่มีความต้องการคุณภาพของเสียงที่สูงนั้น ในอดีตจำเป็นต้องใช้อัตราการเข้ารหัสสูง เช่น การเข้ารหัสด้วยวิธี PCM (Pulse Code Modulation) ในมาตรฐาน ITU G.711 หรือ ADPCM (Adaptive Differential Pulse Code Modulation) ในมาตรฐาน ITU G.721 ต่อมาได้มีการพัฒนากรรมวิธีการเข้ารหัสเสียงพูดหรือการบีบอัดเสียงพูดเป็นจำนวนมาก เพื่อต้องการให้คุณภาพของเสียงที่สูงในขณะที่ใช้อัตราการเข้ารหัสต่ำลง เช่น การเข้ารหัสด้วยวิธี CELP (Code Excited Linear Prediction) MPLPC (Multiple Pulse Linear Prediction Coding) ATC (Adaptive Transformation Coding) เป็นต้น แต่เทคนิคการเข้ารหัสเหล่านี้มีเวลาประวิงค่อนข้างสูงคือประมาณ 40-60 มิลลิวินาที เนื่องมาจากมีการเก็บตัวอย่างเสียงในแต่ละเฟรมจำนวนมาก จึงได้มีการพัฒนากรรมวิธีการเข้ารหัสที่เรียกว่า Low-Delay CELP (LD-CELP) ที่สามารถลดช่วงเวลาประวิงให้ต่ำกว่า 2 มิลลิวินาที และได้ถูกกำหนดเป็นมาตรฐาน ITU G.728 ที่อัตราการเข้ารหัส 16 kb/s และยังคงคุณภาพของเสียงได้ทัดเทียมการเข้ารหัสด้วยวิธี ADPCM [7 และ 8]

การเข้ารหัสด้วยวิธี Low-Delay CELP มีปริมาณการคำนวณที่ค่อนข้างสูงมาก ตัวเข้ารหัสและถอดรหัสก็มีความซับซ้อนสูง [8 9 และ 10] ทำให้ในทางปฏิบัติจำเป็นต้องใช้ตัวประมวลผลสัญญาณที่มีประสิทธิภาพสูง จึงเป็นข้อด้อยของการเข้ารหัสชนิดนี้ ในปี ค.ศ. 1990 ทางกลุ่ม study group (SG) 15 ของ ITU ได้กำหนดความต้องการของตัวเข้ารหัสเสียงที่มีคุณภาพสำหรับการสื่อสารผ่านเครือข่ายที่มีการรบกวนสูง (toll quality) โดยมีอัตราบิตข้อมูลที่ 8 kbps และทนต่อความผิดพลาดของบิตข้อมูลที่ 10^{-3} ได้ ซึ่งเทียบเท่าหรือดีกว่ามาตรฐาน G.726 ADPCM และมีการประวิงเวลาที่เกิดจากการเข้ารหัสไม่เกิน 5 มิลลิวินาที สำหรับใช้ในระบบสื่อสารไร้สายเคลื่อนที่ส่วนบุคคล โดยกำหนดระยะเวลาสิ้นสุดโครงการนี้ให้พอดีกับตารางการกำหนดมาตรฐานของกลุ่มที่ศึกษาระบบสื่อสาร ไร้สายเคลื่อนที่ส่วนบุคคลสำหรับใช้ในยุคที่ 3 ของ ITU เอง [11] ซึ่งเคยรู้จัก

กันในชื่อของ FPLMTS (Future Public Land mobile Telephone System) และต่อมาเปลี่ยนเป็น IMT-2000 (IMT: International Mobile Telephone) ที่ต้องการการเข้ารหัสสัญญาณเสียงที่อัตราข้อมูลที่ 8 kbps การประวิงเวลารวมในการสื่อสารเมื่อคิดทางเดียวมีค่าไม่เกิน 25 มิลลิวินาที (การประวิงเวลาเนื่องจากการเข้ารหัสน้อยกว่า 16 มิลลิวินาที แต่ต้องการให้ไม่เกิน 5 มิลลิวินาที เพื่อใช้กับการสื่อสารในระบบโทรศัพท์เคลื่อนที่ผ่านดาวเทียมได้ด้วย) มีความซับซ้อนไม่มากเกินไป และสามารถทนต่อสัญญาณรบกวนในช่องการสื่อสารได้ นอกจากนี้ยังได้แบ่งระดับความต้องการออกเป็น 3 ระดับ คือสำหรับใช้ภายในอาคาร (class A) ภายนอกอาคาร (class B) และสื่อสารผ่านดาวเทียม (class C) สำหรับผลการรบกวนที่เกิดขึ้นในช่องการสื่อสารนั้น ต้องการให้รหัสที่ได้นั้นมีความสามารถทนทานต่อสัญญาณรบกวนในช่องสื่อสารได้โดยคุณภาพของเสียงที่ได้ไม่ลดลงมากนัก [12 และ 13]

อย่างไรก็ตามไม่มีวิธีการใดที่ผ่านการทดสอบตามข้อกำหนด ภายในกำหนดเวลาไม่เกินเดือนกรกฎาคม ปี ค.ศ. 1991 ดังนั้นทาง SG15 จึงได้แก้ไขข้อกำหนดใหม่โดยยอมให้มีการประวิงเวลาที่เกิดจากการเข้ารหัสและถอดรหัสรวมกันไม่เกิน 32 มิลลิวินาที (การประวิงเวลาที่เกิดจากการเข้ารหัสไม่เกิน 16 มิลลิวินาที) แต่ยังคงต้องการให้มีค่าไม่เกิน 5 มิลลิวินาที ถ้าเป็นไปได้ ทาง SG15 จึงได้กำหนดความต้องการของมาตรฐานการเข้ารหัสใหม่ [14] ในเดือนพฤศจิกายน ปี ค.ศ. 1992 ได้มีผู้เสนอวิธีการเข้ารหัสตามข้อกำหนด 2 แบบ คือจาก NTT (ประเทศญี่ปุ่น) ที่มีการแบ่งสัญญาณเสียงออกเป็นเฟรมๆ ละ 13 มิลลิวินาที และเรียกวิธีนี้ว่า conjugate structure CELP (CS-CELP) และอีกวิธีจาก France Telecom ร่วมมือกับทีมนักวิจัยจากมหาวิทยาลัย Sherbooke ของแคนาดา ซึ่งข้อมูลถูกแบ่งเป็นเฟรมๆ ละ 12 มิลลิวินาที เรียกว่า algebraic CELP (A-CELP) ทั้งสองวิธีต่างก็ได้รับการยอมรับจาก SG15 อย่างไรก็ตามทาง SG15 ต้องการให้มีการปรับปรุงให้แต่ละเฟรมข้อมูลมีค่าประมาณ 10 มิลลิวินาที เพื่อลดค่าการประวิงเวลาที่เกิดขึ้นให้เหลือน้อยที่สุด และได้มีการทดสอบในเดือนกันยายน ปี ค.ศ. 1993 ที่ประเทศสหพันธรัฐเยอรมัน โดยกำหนดให้สร้างขึ้นด้วยการประมวลผลแบบ fixed point และมีทดสอบกับผู้พูดจำนวน 6 ภาษา ได้แก่ ภาษาญี่ปุ่น อิตาลี เยอรมัน ฝรั่งเศส นอร์เวย์ และอังกฤษ ภายใต้สภาพแวดล้อมที่แตกต่างกัน [8] อย่างไรก็ตามหลังจากการทดสอบได้มีการปรับปรุงและพัฒนาวิธีการให้ดีขึ้นโดยความร่วมมือจาก AT&T และในที่สุดทาง ITU ก็เลือกวิธีการเข้ารหัสแบบ Conjugate Structure and Algebraic CELP (CS-ACELP) โดยเป็นการรวมทั้งสองวิธีเข้าด้วยกันได้เป็นมาตรฐาน G.729 ในเดือนกุมภาพันธ์ ปี ค.ศ. 1995 นับเป็นมาตรฐานการเข้ารหัสเสียงพูดล่าสุดของ ITU [15 16 และ 17]

2.2 การวัดสมรรถนะของการเข้ารหัสเสียงพูด

สมรรถนะในการเข้ารหัสเสียงพูดจะพิจารณาจากคุณสมบัติต่างๆ เช่น อัตราการเข้ารหัส (bit rate) คุณภาพเสียงที่ผ่านการเข้ารหัส (speech quality) ความซับซ้อนของการเข้ารหัส (complexity มีหน่วยเป็น MIPS: Million Instructions Per Second) ค่าประวิงเวลา (delay time) ความทนทานต่อความผิดพลาดที่เกิดภายในช่องสัญญาณ (robustness) หรือการแทรกสอดที่เกิดจากเสียงอื่นๆ (acoustic interferences) เช่น เสียงรบกวน สัญญาณ DTMF ในระบบโทรศัพท์ สัญญาณของโมเด็ม เป็นต้น [18 19 และ 20]

ในการสื่อสารข้อมูลแบบดิจิทัล คุณภาพของเสียงพูดถูกแบ่งออกเป็น 4 ระดับได้แก่

1. ระดับกระจายเสียง (broadcast) เสียงพูดในระดับนี้จะอ้างถึงเสียงพูดบรรยายที่มีคุณภาพสูง โดยปกติจะมีอัตราการเข้ารหัสตั้งแต่ 64 kbps ขึ้นไป

2. ระดับเครือข่าย (toll หรือ network) คุณภาพเสียงจะสามารถเทียบได้กับเสียงพูดในระบบอนาลอก ในช่วงความถี่ 200-3200 เฮิรตซ์ โดยปกติจะมีอัตราการเข้ารหัสตั้งแต่ 16 kbps ขึ้นไป

3. ระดับสื่อสาร (communication) ยอมให้คุณภาพเสียงลดลงได้บ้างแต่ยังคงความเป็นธรรมชาติของเสียงอยู่ มีคุณภาพเพียงพอที่จะใช้ในการสื่อสาร สามารถสร้างได้โดยอัตราการเข้ารหัสตั้งแต่ 4.8 kbps ขึ้นไป แต่ปัจจุบันมีเป้าหมายให้ลดลงมาที่ 4.0 kbps

4. ระดับสังเคราะห์ (synthetic) สามารถรับฟังได้เข้าใจ แต่ไม่เป็นธรรมชาติ และสูญเสียคุณสมบัติในการรู้จำเจ้าของเสียงพูด มีอัตราการเข้ารหัสต่ำกว่า 4.0 kbps

การวัดคุณภาพของการเข้ารหัสนั้นเป็นงานที่สำคัญและมีความยุ่งยากอยู่มาก วิธีการหนึ่งที่นิยมใช้กันอย่างแพร่หลาย คือการใช้ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวน (Power Signal-to-Noise Ratio : PSNR) เป็นการวัดค่าเชิงวัตถุ (objective measurement) ใช้สำหรับวัดคุณสมบัติของอัลกอริทึมในการบีบอัดข้อมูล โดยมีการคำนวณตามสมการ

$$\mathrm{PSNR} = 10\log_{10}\left\{\frac{\sum_{n=0}^{M-1} s^2(n)}{\sum_{n=0}^{M-1}(s(n)-s'(n))^2}\right\} \qquad (2\text{-}1)$$

โดย s(n) คือสัญญาณเสียงดั้งเดิม ส่วน $s'(n)$ คือสัญญาณเสียงที่ผ่านการเข้ารหัสแล้ว ค่า PSNR นั้นถือได้ว่าเป็นการวัดแบบช่วงยาว (long-term) สำหรับวัดหาความถูกต้องของการสร้างสัญญาณเสียงกลับขึ้นมาใหม่

การเปลี่ยนแปลงอย่างฉับพลันสามารถตรวจจับและประเมินได้โดยการใช้ PSNR ในช่วงสั้น (short-term) คือการคำนวณ PSNR สำหรับแต่ละส่วนของเสียงพูดที่มีอยู่ N จุด เรียกการวัดเช่น

นี้ว่า การหาอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน (SegPSNR: Segmental Power Signal-to-Noise Ratio) โดยมีการคำนวณตามสมการ

$$\text{SegPSNR} = \frac{10}{L}\sum_{i=0}^{L-1}\log_{10}\left\{\frac{\sum_{n=0}^{N-1} s^2(iN+n)}{\sum_{n=0}^{N-1}(s(iN+n)-s'(iN+n))^2}\right\} \qquad (2\text{-}2)$$

เนื่องจากการหาค่าเฉลี่ยของสมการ (2-2) เกิดหลังการคำนวณค่าลอการิทึม SegPSNR จะแสดงข้อผิดพลาดของการเข้ารหัสที่การทำงานมีการเปลี่ยนแปลงไปเรื่อยๆ ได้มากกว่าค่า PSNR ธรรมดา นอกจากนี้ยังมีวิธีการอื่นๆ อีกเช่น articulation index log special distance และ Euclidian distance ซึ่งวิธีเหล่านี้เป็นการวัดเชิงวัตถุทั้งสิ้น มิได้พิจารณาถึงคุณภาพในการรับฟังเสียงของหูมนุษย์ แต่ในการออกแบบอัลกอริทึมที่ใช้อัตราการเข้ารหัสต่ำเกือบทั้งหมดจะมีพื้นฐานมาจากบรรทัดฐานของการรับฟังเสียงของมนุษย์

วิธีการวัดคุณภาพเสียงอีกประเภทหนึ่งคือใช้การรับรู้และความรู้สึกของมนุษย์เป็นเกณฑ์ในการตัดสิน คือการวัดในเชิงของผู้ฟัง (subjective measurement) มีหลายวิธีได้แก่ Diagnostic Rhyme Test (DRT) Diagnostic Acceptability Measure (DAM) และ Mean Opinion Score (MOS)

วิธี MOS เป็นวิธีที่ใช้กันอย่างแพร่หลาย จะใช้ผู้ฟังประมาณ 12-24 คน (อาจเปลี่ยนไปแล้วแต่มาตรฐานที่จะเป็นตัวกำหนด) ซึ่งถ้าทำการทดสอบตามมาตรฐาน ITU นั้นจะมีรายละเอียดขั้นตอนในการทดสอบดังนี้ [21]

1. ใช้คนกลุ่มเล็กๆ ทำการอ่านประโยคทดสอบและทำการอัดสัญญาณเสียงเหล่านี้ไว้
2. ทำการเข้ารหัสสัญญาณเสียงเหล่านี้ด้วยวิธีที่ต้องการจะทดสอบ
3. ทดสอบคุณภาพของเสียงกับกลุ่มคนประมาณ 12-24 คน โดยที่แต่ละคนจะให้คะแนนที่มีค่าอยู่ระหว่าง 1-5 ตามคุณภาพของสัญญาณที่ตัวเองรู้สึก รายละเอียดของคะแนนแต่ละขั้นแสดงไว้ในตารางที่ 2.1
4. นำค่าเฉลี่ยที่ได้ไปใช้ ซึ่งมีชื่อเรียกว่า Mean Opinion Score (MOS)

ระบบการใช้งานแต่ละระบบมีความต้องการคุณภาพเสียงหรือค่า MOS ที่แตกต่างกันออกไป สามารถสรุปค่า MOS ที่เหมาะสมกับการใช้งานต่างๆ ได้ดังตารางที่ 2.2

ตารางที่ 2.1 รายละเอียดวิธีการให้คะแนนในการวัด MOS

คะแนน	คุณภาพของเสียง
5	ดีมาก (คุณภาพเสียงชัดเจนและเข้าใจง่าย)
4	ดี (คุณภาพเสียงดีและเข้าใจง่าย แต่อาจได้ยินเสียงรบกวนบ้าง)
3	พอใช้ (คุณภาพเสียงเข้าใจได้แต่อาจต้องอาศัยความตั้งใจ หรือบางทีต้องขอให้พูดซ้ำ)
2	เลว (คุณภาพเสียงจะเข้าใจได้ก็ต่อเมื่อมีความตั้งใจมากๆ และบ่อยครั้งที่ต้องขอให้พูดซ้ำ)
1	เลวมาก (ฟังไม่รู้เรื่องเลย)

ตารางที่ 2.2 ค่า MOS ที่เหมาะสมกับการใช้งานในระบบต่างๆ

MOS	การใช้งาน
4.5-5.0	Broadcast Quality
4.0-4.5	Network or “Toll” Quality
3.5-4.0	Communication Quality
2.5-3.5	Synthetic Quality

2.3 การเปรียบเทียบสมรรถนะการเข้ารหัสเสียงพูดโดยวิธีต่างๆ

สมรรถนะการเข้ารหัสเสียงพูดโดยวิธีต่างๆ ที่รู้จักกันอย่างแพร่หลาย แสดงในตารางที่ 2.3 มีทั้งที่ได้ถูกกำหนดเป็นมาตรฐานใน ITU และที่ถูกกำหนดใช้ในระบบสื่อสารในองค์กรต่างๆ โดยได้นำเสนอในเชิงเปรียบเทียบคุณสมบัติด้านต่างๆ คือ อัตราการเข้ารหัส เวลาประวิงในการเข้ารหัส ความซับซ้อนในการเข้ารหัส คุณภาพเสียงพูดที่ผ่านการเข้ารหัส และช่องสุดท้ายแสดงแหล่งที่ประยุกต์ใช้งานของการเข้ารหัสโดยวิธีนั้นๆ [11 22 23 24 25 26 และ 27].

ตารางที่ 2.3 แสดงการเปรียบเทียบคุณสมบัติการเข้ารหัสเสียงพูดแบบต่างๆ

Scheme	Bit rate (kb/s)	Delay (ms)	Complexity (MIPS)	Quality (MOS)	Application
G.711 PCM	64	0.125	0.01	4.1 Toll	
G.721 ADPCM	32	0.125	2	4.3 Toll	
G.726 ADPCM	16 24 32 40	0.125	2	Toll	Wideband
G.722 wideband	48 56 64	0.125		Toll	Wideband
G.728 LD-CELP	16	2	30	4.2	
G.729 CS-ACELP	8	15	20	3.7	Nokia IS-641 North America
G.729 CS-ACELP Annex A	8	15	11	3.7	
G.723.1 MPC-MLO	5.3 6.4	67.5	16		Internet phone
RPE-LTP (GSM)	13	20	6	3.47+	European Global System
MPE-LPC	9.6		11	3.4	Sky phone
IS-54 VSELP (TIA)	7.95	20	13.5	3.45+	D-AMPS
PDC VSELP (RCR Japan)	6.7	20	13.5	3.5	Japan (JDS)
GSM-HR	6.5	20	(1.5 wrt GSM)	4	ETSI
Type A (ADPCM+VQ)	8.1	7.5	(8 wrt GSM)	3.5	UMTS
Type B (GSM FR)	7.9	20	(2 wrt GSM)	3.8	UMTS
IS-96 Q-CELP (TIA)	1.2-8 (3.5avg)	20	(1 wrt GSM)	4-	CDMA in USA

ตารางที่ 2.3 แสดงการเปรียบเทียบคุณสมบัติการเข้ารหัสเสียงพูดแบบต่างๆ (ต่อ)

Scheme	Bit rate (kb/s)	Delay (ms)	Complexity (MIPS)	Quality (MOS)	Application
TGMS codec (Third Generation Mobile Systems)	<8	<10	(0.5 wrt GSM)	4.5+	
PDC PSI-CELP (RCR-Japan)	3.45	40			Japan (JDS)
FS-1015 LPC 10E	2.4	22.5	7	2.3	US-DoD
FS-1016 CELP	4.8	30	16	3.2	US-DoD
MELP	2.4	22.5			US-DoD
STC-1	4.8		13	3.52	
STC-2	2.4		13	2.9	
IMBE	4.1		13	3.4	

2.4 การเข้ารหัสเสียงพูดโดยวิธี CS-ACELP ตามมาตรฐาน ITU G.729

การเข้ารหัสเสียงพูดโดยวิธี CS-ACELP เป็นมาตรฐานล่าสุดของ ITU ถูกออกแบบมาเพื่อประมวลผลสัญญาณดิจิทัลที่ได้จากการกรองสัญญาณ ในช่วงแบนด์วิดท์ของระบบโทรศัพท์มีขาเข้าเป็นสัญญาณอนาลอก (ITU G.712) ที่อัตราสุ่ม 8000 Hz แล้วเปลี่ยนเป็นรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิตแล้วจึงป้อนให้กับขาเข้าของตัวเข้ารหัสนี้ ขาออกของตัวถอดรหัสจะถูกเปลี่ยนกลับมาเป็นสัญญาณในลักษณะตรงข้ามกัน คุณสมบัติอื่นๆ ของขาเข้าและขาออกนั้นจะเป็นไปตามข้อกำหนดในมาตรฐาน ITU G.711 ของรหัส PCM ที่มีอัตรา 64 kbps ซึ่งจะถูกเปลี่ยนเป็นรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิตก่อนการเข้ารหัส และจากรหัส PCM แบบเชิงเส้นที่มีความละเอียด 16 บิตก็จะถูกเปลี่ยนเป็นรหัสที่เหมาะสมต่อไปหลังจากการถอดรหัสแล้ว ซึ่งกระแสบิตที่ได้จากการเข้ารหัสที่จะส่งไปถอดรหัสนั้นจะถูกกำหนดให้เป็นไปตามมาตรฐานอัตราการเข้ารหัสอยู่ที่ 8 kpbs [8 และ 15]

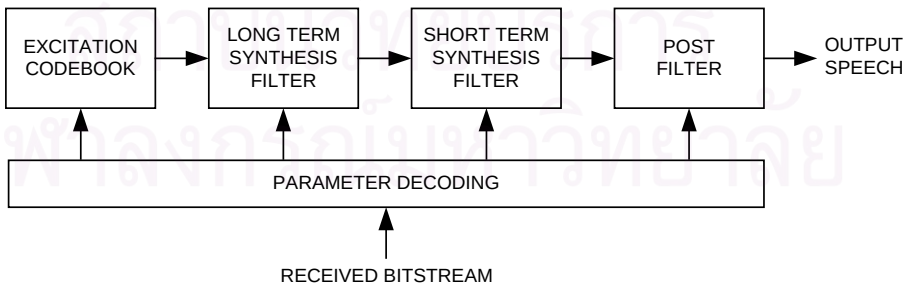
ตัวเข้ารหัสสัญญาณเสียงแบบ CS-ACELP ใช้หลักการเข้ารหัสสัญญาณเสียงแบบ code-excited linear-predictive (CELP) โดยตัวเข้ารหัสจะแบ่งสัญญาณเสียงออกเป็นเฟรมๆ ละ 10 มิลลิวินาที ในแต่ละเฟรมจะมีตัวอย่างที่ถูกสุ่ม 80 ตัวอย่าง โดยอัตราสุ่มจะเท่ากับ 8000 ตัวอย่างต่อวินาที สัญญาณเสียงในทุกๆ เฟรมจะถูกวิเคราะห์เพื่อหาพารามิเตอร์ต่างๆ ตามหลักการของ CELP (ได้แก่ สัมประสิทธิ์ของวงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียง adaptive-codebook

fixed-codebook และอัตราขยาย) พารามิเตอร์เหล่านี้จะถูกนำมาเข้ารหัสและส่งไปยังปลายทาง รายละเอียดของบิตต่างๆ ที่ได้จากการเข้ารหัสพารามิเตอร์เหล่านี้แสดงในตารางที่ 2.4 ที่ตัวถอดรหัสพารามิเตอร์เหล่านี้ถูกใช้ในการสร้างเอ็กไซเทชัน และสัมประสิทธิ์ของวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์สัญญาณเสียง สัญญาณเสียงจะถูกสังเคราะห์โดยการกรองสัญญาณเอ็กไซเทชันผ่านวงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียงแบบ short-term ตามรายละเอียดในรูปที่ 2.1

ตารางที่ 2.4 การจัดสรรบิตของการเข้ารหัสเสียงพูดโดยวิธี CS-ACELP

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

วงจรกรองสัญญาณที่ใช้สังเคราะห์สัญญาณเสียงแบบ short-term นั้นจะเป็นวงจรกรองสัญญาณที่ใช้ทำนายสัญญาณแบบเชิงเส้น (linear prediction) LP ที่มีอันดับเท่ากับ 10 สำหรับวงจรกรองสัญญาณแบบ long-term ที่ใช้ในการสังเคราะห์พิตช์ (pitch) จะถูกสร้างโดยใช้วิธี adaptive-codebook สัญญาณเสียงที่ได้จากการสังเคราะห์จะถูกปรับแต่งโดย postfilter



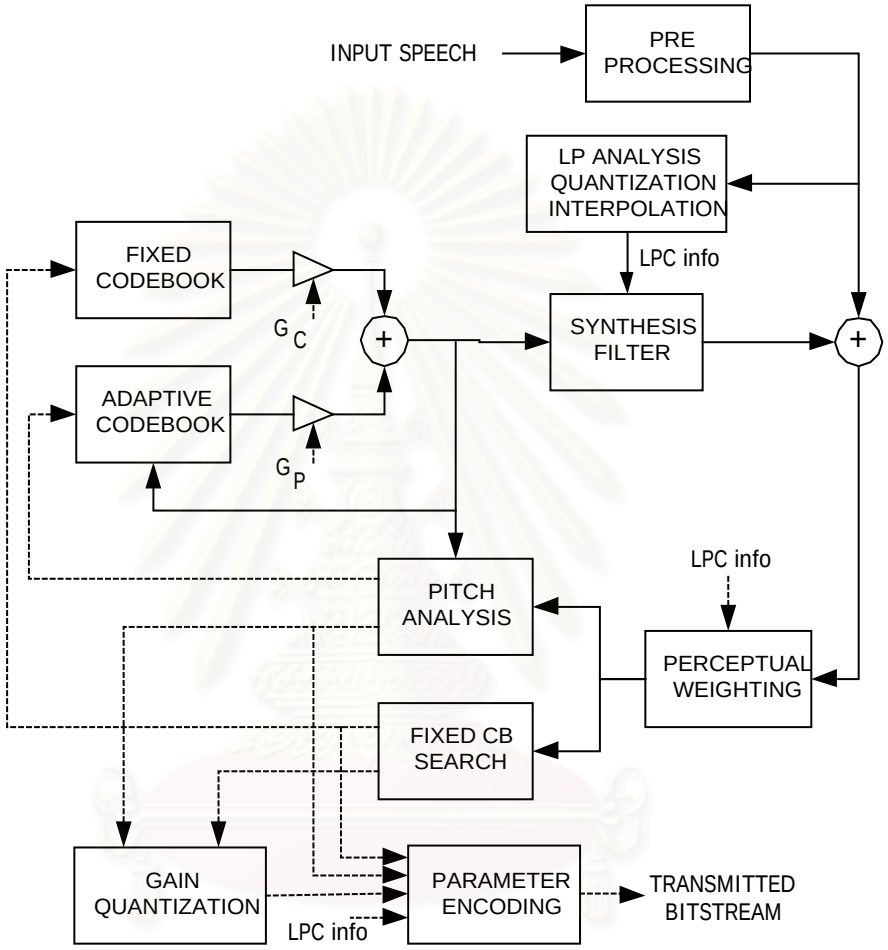
รูปที่ 2.1 ไดอะแกรมการทำงานของการสังเคราะห์สัญญาณเสียงด้วยวิธี CELP

2.4.1 ตัวเข้ารหัส

หลักการของการเข้ารหัสแสดงในรูปที่ 2.2 สัญญาณเข้าจะถูกกรองแบบความถี่สูงผ่านและถูกปรับลดขนาดโดย pre-processing สัญญาณที่ได้จาก pre-processing นี้จะเป็นสัญญาณขาเข้าให้กับทุกๆ วงจรที่เหลือเพื่อใช้ในการวิเคราะห์ การวิเคราะห์ LP นั้นจะแบ่งสัญญาณเสียงเป็นเฟรมๆ ละ 10 มิลลิวินาที เพื่อใช้ในการคำนวณหาสัมประสิทธิ์ของวงจรกรองสัญญาณ LP สัมประสิทธิ์เหล่านี้จะถูกแปลงเป็น line spectrum pairs (LSP) และทำการควอนไทซ์โดยใช้วิธีการเวกเตอร์ควอนไทซ์แบบ 2 stage ที่มีความละเอียด 18 บิต สัญญาณเอ็กไซเทชันจะถูกคำนวณหาจากการใช้วิธีการ analysis-by-synthesis ในการค้นหาเพื่อให้ค่าความผิดพลาดระหว่างสัญญาณเสียงดั้งเดิมกับสัญญาณเสียงที่สังเคราะห์ขึ้นมาได้มีค่าน้อยที่สุดโดยใช้วิธีการวัดด้วย perceptual weighted distortion โดยใช้การกรองสัญญาณความผิดพลาดด้วยวงจรกรองสัญญาณแบบ perceptual weighting ที่มีสัมประสิทธิ์ที่ได้มาจากสัมประสิทธิ์ของวงจรกรองสัญญาณ LP ที่ถูก unquantized แล้ว จำนวนของการทำ perceptual weighting นั้นสามารถปรับได้เพื่อให้สมรรถนะการทำงานที่ได้มีผลตอบสนองต่อสัญญาณเข้าเท่ากันทุกๆ ความถี่ (flat frequency-response)

พารามิเตอร์ของเอ็กไซเทชัน (พารามิเตอร์ของ fixed-codebook และ adaptive-codebook) จะถูกคำนวณหาทุกๆ เฟรมย่อย โดยแต่ละเฟรมย่อยนี้จะมีขนาด 5 มิลลิวินาที (40 ตัวอย่างสัญญาณที่อัตราสุ่ม 8000 เฮิรตซ์) สัมประสิทธิ์ของวงจร กรองสัญญาณ LP ทั้งที่ควอนไทซ์และ unquantized จะถูกใช้ในเฟรมย่อยที่ 2 ส่วนในเฟรมย่อยที่ 1 นั้นจะใช้วิธีการหาสัมประสิทธิ์ของวงจรกรองสัญญาณ LP โดยการประมาณค่าในช่วง (ทั้ง quantized และ unquantized) สำหรับ open-loop pitch delay จะถูกประมาณหนึ่งครั้งในแต่ละเฟรมจากสัญญาณที่ได้จากวงจรกรองสัญญาณ perceptual weighting สำหรับกระบวนการที่จะกล่าวต่อไปนี้จะกระทำทุกๆ เฟรมย่อย สัญญาณของเป้า x(n) จะถูกคำนวณโดยการกรองสัญญาณ residual ของ LP ผ่านวงจรกรองสัญญาณ weighted synthesis W(z)/A(z) โดยเมื่อเริ่มการทำงาน วงจรกรองสัญญาณเหล่านี้จะถูกปรับให้ทันกาลจากการกรองค่าความผิดพลาดระหว่าง LP residual กับเอ็กไซเทชัน นั่นหมายถึงการหักล้างด้วยวงจรกรองสัญญาณแบบ weighted synthesis ที่ zero-input response ออกจากสัมประสิทธิ์ของสัญญาณเสียง การตอบสนองอิมพัลส์ h(n) ของวงจรกรองสัญญาณ weighted synthesis จะถูกคำนวณหา ส่วนการวิเคราะห์หา closed-loop pitch (เพื่อหาค่าประวิงเวลา และอัตราขยายของ adaptive-codebook) จะใช้สัญญาณของเป้า x(n) และการตอบสนองอิมพัลส์ h(n) โดยการค้นหารอบค่าของ open-loop pitch delay ค่า pitch delay นี้จะถูกเข้ารหัสด้วยความละเอียด 8 บิตในเฟรมย่อยแรก ส่วนที่เหลือจะเข้ารหัสด้วยความละเอียด 5 บิตในเฟรมย่อยที่สอง สัญญาณของเป้า x(n) จะถูกปรับค่าโดยการนำค่าจาก adaptive-codebook มาหักล้าง และค่าสัญญาณของเป้าที่ได้ใหม่ $x'(n)$ นี้จะถูกใช้ในการค้นหา fixed-codebook เพื่อให้ได้เอ็กไซเทชันที่ถูกต้องที่สุดต่อไป

โดยโครงสร้างของ fixed-codebook นี้จะเป็นแบบ algebraic ที่มีความละเอียด 17 บิต อัตราขยายของ adaptive-codebook และ fixed-codebook จะถูกควอนไทซ์ที่มีความละเอียด 7 บิต (โดยจะใช้ตัวทำนายสัญญาณแบบ MA-Moving Average กับอัตราขยายของ fixed-codebook) สุดท้ายค่าต่างๆ ของวงจรกรองสัญญาณจะถูกปรับค่าตามเอ็กไซเทชันที่หาได้

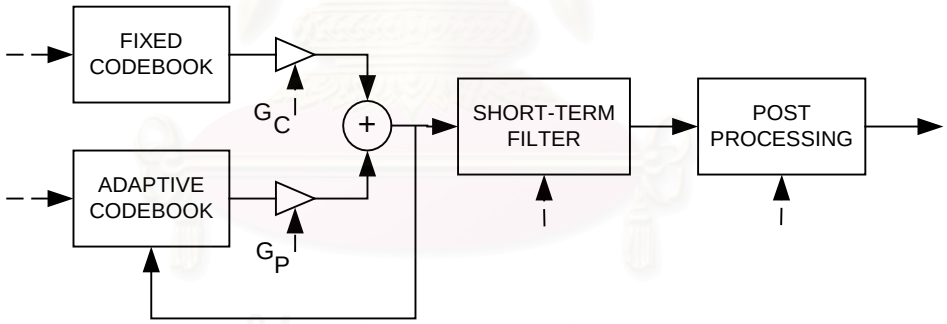


รูปที่ 2.2 หลักการเข้ารหัสของตัวเข้ารหัสแบบ CS-ACELP

2.4.2 ตัวถอดรหัส

หลักการทำงานของตัวถอดรหัสแสดงตามรูปที่ 2.3 อันดับแรกพารามิเตอร์ที่รับมาได้ในรูปของกระแสของบิตข้อมูลจะถูกแยกแยะออกตามเฟรมๆ ละ 10 มิลลิวินาที และถอดรหัสออกมาเป็นค่าของพารามิเตอร์ต่างๆ ได้แก่สัมประสิทธิ์ 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 นั้นจะมีขั้นตอนการทำงานดังนี้

- เอ็กไซเทชันจะถูกสร้างขึ้นโดยการรวม adaptive-codebook และ fixed-codebook เข้าด้วยกันตามอัตราขยายของแต่ละตัว
- สัญญาณเสียงจะถูกสังเคราะห์ขึ้น โดยนำสัญญาณเอ็กไซเทชันที่ได้มากรองด้วยวงจรกรองสัญญาณสังเคราะห์เสียง LP
- สัญญาณเสียงที่สังเคราะห์ได้จะนำไปผ่านวงจร post-processing ซึ่งประกอบด้วย adaptive postfilter ที่สร้างจากวงจรกรองสัญญาณแบบ long-term และ short-term วงจรกรองสัญญาณความถี่สูงผ่าน และวงจรปรับขนาดสัญญาณ



รูปที่ 2.3 หลักการทำงานของตัวถอดรหัสแบบ CS-ACELP

2.4.3 การประวิงเวลา

การเข้ารหัสของตัวเข้ารหัสสัญญาณเสียงพูดและสัญญาณเสียงอื่นๆ นี้จะทำงานกับสัญญาณเสียงที่แบ่งออกเป็นเฟรมๆ ละ 10 มิลลิวินาที นอกจากนี้จะต้องมีส่วนวิเคราะห์สัญญาณล่วงหน้าอีก 5 มิลลิวินาที ทำให้การประวิงเวลาที่เกิดขึ้นเนื่องจากการทำงานตามอัลกอริทึมนี้มีค่า

15 มิลลิวินาที สำหรับการประวิงเวลาทั้งหมดอันเนื่องมาจากสาเหตุอื่นๆด้วยนั้น ขึ้นกับสภาพแวดล้อมในการใช้งานจริง ซึ่งอาจมีสาเหตุมาจาก

- เวลาที่ใช้ในการประมวลผลสำหรับการเข้าและถอดรหัส
- เวลาที่ใช้ในการส่งผ่านข้อมูลในช่องสัญญาณ
- เวลาที่เกิดจากการทำมัลติเพลกซ์เพื่อใช้ในการรวมข้อมูลของสัญญาณเสียงกับข้อมูลอื่นๆ

จากบทความ “Thai Speech Compression Using CS-ACELP according to ITU G.729 Standard” [28] ได้จำลองการเข้ารหัสเสียงพูดโดยวิธี CS-ACELP และทดสอบสมรรถนะการเข้ารหัส โดยเปรียบเทียบระหว่างเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ ได้ผลว่า คุณภาพเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสแล้ว ด้อยกว่าคุณภาพเสียงพูดภาษาอังกฤษที่ผ่านการเข้ารหัสในสภาพแวดล้อมเดียวกัน ประมาณ 0.25-0.28 dB สำหรับเสียงพูดของชาย และประมาณ 0.28-0.34 dB สำหรับเสียงพูดของหญิง แสดงให้เห็นว่า การเข้ารหัสเสียงพูดโดยวิธีนี้ จะให้สมรรถนะตกลงเมื่อนำมาใช้กับเสียงพูดที่เป็นภาษาไทย

ในวิทยานิพนธ์นี้ จะใช้การเข้ารหัสเสียงพูดโดยวิธีนี้ เป็นตัวอ้างอิงกับการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ที่นำเสนอตามข้อกำหนดการเข้ารหัสเสียงพูดมาตรฐาน MPEG-4

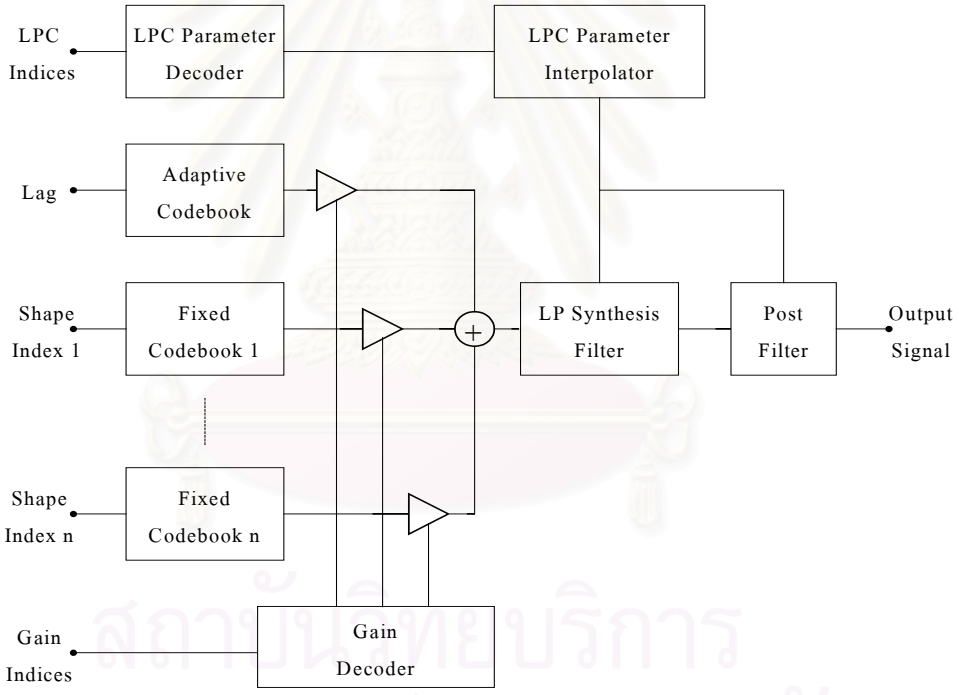
2.5 ข้อกำหนดการเข้ารหัสเสียงพูดมาตรฐาน MPEG-4

ในปี ค.ศ.1995 Moving Pictures Experts Group (MPEG) กลุ่มทำงานใน ISO ได้เริ่มดำเนินงานในการกำหนดมาตรฐานเครื่องมือการเข้ารหัสที่อัตราการเข้ารหัสต่ำ เพื่อใช้งานใน internet และช่องสัญญาณส่งผ่านที่มี bandwidth จำกัด โครงการนี้รู้จักในนาม MPEG-4 ซึ่งจะถูกกำหนดเป็นมาตรฐานสากล ISO 14496 ในปี ค.ศ.1999 นี้ [2 และ 3]

มาตรฐาน MPEG-4 จะพิจารณาสัญญาณเสียงหรือภาพเป็นแบบ object-based เพื่อรองรับ multimedia เช่น รายการภาพยนตร์ รายการวิทยุ ซึ่งจะถูกส่งในลักษณะ media object เช่น streaming video segments streaming video still images streaming audio tracks synthetic visual graphics วิธีการเข้ารหัสสำหรับแต่ละชนิดของ media object ถูกกำหนดในมาตรฐาน MPEG-4 audio และ MPEG-4 video เมื่อแต่ละ object ถูกส่งไปยังปลายทาง ชิ้นส่วนต่างๆ จะถูกนำมาประกอบรวมกันเป็น audio visual scene โดยมีรูปแบบเฉพาะสำหรับการจัดการ [5 และ 6]

2.5.1 MPEG-4 CELP coder

ในเบื้องต้น ตัวถอดรหัส CELP ประกอบไปด้วย แหล่งกำเนิดสัญญาณกระตุ้น และฟิลเตอร์สังเคราะห์อย่างละชุด นอกจากนี้จะมีส่วนเพิ่มเติมคือ postfilter แหล่งกำเนิดสัญญาณกระตุ้นมีทั้งส่วนที่เป็นรายคาบซึ่งมี adaptive codebook และส่วนที่เป็น random ซึ่งมีหนึ่ง fixed codebook หรือมากกว่า ที่ตัวถอดรหัสนี้ สัญญาณกระตุ้นถูกสร้างขึ้นโดยใช้ดัชนีของ codebook ซึ่งประกอบไปด้วย pitch lag สำหรับ adaptive codebook และ shape index สำหรับ fixed codebook และดัชนีอัตราขยาย ซึ่งประกอบไปด้วย adaptive และ fixed codebook gains สัญญาณกระตุ้นนี้จะถูกกรองด้วยฟิลเตอร์ LP สังเคราะห์ ที่ได้สัมประสิทธิ์ LPC จากการประมาณจากดัชนี LPC ในแต่ละเฟรม สุดท้ายสัญญาณเสียงที่สังเคราะห์ได้จะถูกเพิ่มคุณภาพด้วย postfilter รูปที่ 2.4 แสดงบล็อกไดอะแกรมของ ตัวถอดรหัส CELP ตามข้อกำหนดของมาตรฐาน MPEG-4 [3]



รูปที่ 2.4 บล็อกไดอะแกรมของตัวถอดรหัส CELP ตามข้อกำหนด MPEG4

สิ่งที่ทำให้ MPEG-4 CELP แตกต่างไปจาก CELP ดั้งเดิมคือความยืดหยุ่นที่มีมากขึ้น คือ CELP ดั้งเดิมจะรองรับการประยุกต์ใช้งานที่อัตราการเข้ารหัสคงที่ค่าหนึ่งๆ สำหรับ MPEG-4 แล้วนอกจากจะรองรับอัตราการบีบอัดที่สูงแล้วยังสามารถทำงานที่อัตราการเข้ารหัสต่างๆ กันด้วยความสามารถปรับขนาดอัตราการเข้ารหัส หรือคุณลักษณะอื่น นั่นคือจะสามารถรองรับหน้าที่การทำงานเพิ่มเติมด้านอื่นต่อไปนี้ด้วย

2.5.2 หน้าที่การทำงานของ MPEG-4 CELP

หน้าที่การทำงานหลักของตัวเข้ารหัสหลัก MPEG-4 CELP ที่เพิ่มเติมจากการเข้ารหัสแบบ CELP ดั้งเดิม [3] มีดังต่อไปนี้

2.5.2.1 Multiple bit rates

การเข้ารหัสเสียงธรรมชาติจะรองรับอัตราการเข้ารหัสหลายอัตรา เพื่อให้สามารถปรับเปลี่ยนตามความต้องการของระบบได้

2.5.2.2 Bit-rate Scalability

สามารถปรับระดับอัตราการเข้ารหัสได้โดยเพิ่มส่วนขยาย enhancement layer จะทำให้อัตราการเข้ารหัสเพิ่มขึ้นเป็นขั้นๆ และจำนวนขั้นสูงสุดสำหรับปรับระดับอัตราการเข้ารหัสจะอยู่ที่ 3 ขั้น

วิทยานิพนธ์นี้จึงได้นำเสนอการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ที่สนองตอบข้อกำหนดของมาตรฐาน MPEG-4 ในข้างต้น เนื่องจาก CS-ACELP เป็นการเข้ารหัสที่อัตราคงที่ จึงได้มีการพัฒนาปรับปรุง MP-CELP เพื่อให้สามารถปรับระดับอัตราการเข้ารหัส โดยมีหลักการทำงานของตัวเข้ารหัสและตัวถอดรหัสอยู่ในบทต่อไป

บทที่ 3

การเข้ารหัสเสียงพูดโดยวิธี MP-CELP

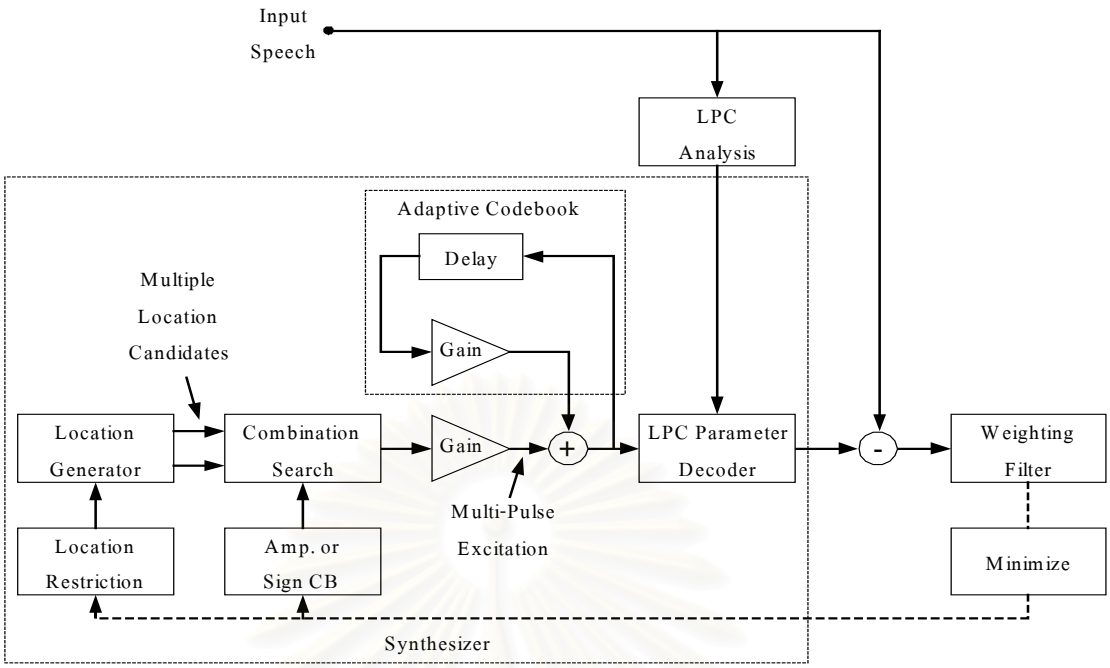
เนื้อหาในบทนี้จะกล่าวถึง หลักการทำงานโดยรวมของตัวเข้ารหัสหลักและตัวถอดรหัส MP-CELP ที่อัตราการเข้ารหัส 3 ค่า คือ 5600 8200 และ 12200 bps การปรับระดับอัตราการเข้ารหัส ทั้งนี้ตามข้อกำหนดของมาตรฐาน MPEG-4 คือสามารถปรับระดับอัตราการเข้ารหัสได้ ความแตกต่างระหว่างคุณลักษณะของเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ และสุดท้ายคือการปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ให้เหมาะสมกับเสียงพูดภาษาไทย

วิทยานิพนธ์นี้นำเสนอการเข้ารหัสเสียงพูดที่อัตราสุ่ม 8 kHz (เสียงแถบแคบ) ใช้การเข้ารหัสเสียงพูด MP-CELP ที่เสนอโดยนักวิจัยชาวญี่ปุ่น Ozawa [29] เนื่องจากสามารถตอบสนองข้อกำหนดของ MPEG-4 ที่กล่าวในบทที่ 2 ได้ในระดับหนึ่ง หลักการทำงานของตัวเข้ารหัสหลัก และส่วนขยายอธิบายได้ดังต่อไปนี้

3.1 หลักการทำงานโดยรวมของตัวเข้ารหัสหลัก MP-CELP

3.1.1 ตัวเข้ารหัสหลัก MP-CELP

MP-CELP เป็นการเข้ารหัสเสียงพูดที่พัฒนามาจาก CS-ACELP ตามมาตรฐาน ITU G.729 มีส่วนประกอบของการเข้ารหัสหลักเหมือนกับ CS-ACELP ดังแสดงด้วยบล็อกไดอะแกรมในรูปที่ 3.1 จะแตกต่างกันตรงแหล่งกำเนิดสัญญาณกระตุ้นในส่วนของ fixed codebook คือ CS-CELP จะสร้างสัญญาณกระตุ้นที่ประกอบด้วยพัลส์จำนวน 4 พัลส์ที่ตำแหน่งต่างๆใน 40 ตำแหน่งและถูกกำหนดโดย sign codebook บนพื้นฐานโครงสร้างพีชคณิต (algebraic) ส่วน MP-CELP จะสร้างสัญญาณกระตุ้นประกอบด้วยพัลส์ที่ปรับเปลี่ยนจำนวนได้ โดยถูกสร้างจาก 4 ส่วนหลักด้วยกัน ส่วนที่ 1 คือ location restriction จะกำหนดตำแหน่งของพัลส์ที่เป็นไปได้ บนพื้นฐานโครงสร้างพีชคณิต ส่วนที่ 2 คือ location generator จะสร้างชุดตำแหน่งของพัลส์และตำแหน่งของแต่ละพัลส์ในชุดนี้จะถูกจำกัดโดยส่วนแรก ส่วนที่ 3 คือ amplitude/sign codebook จะสร้างชุดขนาดหรือเครื่องหมายของพัลส์ ในวิทยานิพนธ์นี้เลือกใช้ sign codebookตาม [4] เพราะใช้จำนวนบิตในการเข้ารหัสต่ำกว่า ส่วนที่ 4 คือ combination search จะเป็นส่วนที่เลือกชุดตำแหน่งและชุดเครื่องหมายของพัลส์ที่เหมาะสมที่สุด คือทำให้ผลต่างของสัญญาณเสียงขาเข้ากับสัญญาณเสียงสังเคราะห์มีค่าต่ำที่สุด



รูปที่ 3.1 หลักการทำงานของตัวเข้ารหัสหลักแบบ MP-CELP

จุดเด่นของการเข้ารหัสด้วยวิธีนี้ คือ สามารถออกแบบสัญญาณกระตุ้นในส่วนของ fixed codebook ให้สามารถปรับเปลี่ยนจำนวนพัลส์ได้ มีผลให้สามารถประยุกต์เป็นการเข้ารหัสเสียงพูดตามข้อกำหนดของมาตรฐาน MPEG-4 ได้ เพราะการรองรับการทำงานที่หลายอัตราการเข้ารหัส และ การปรับระดับอัตราการเข้ารหัส จะกระทำได้โดยการปรับเปลี่ยนจำนวนพัลส์ของสัญญาณกระตุ้นในส่วนของ fixed codebook นี้ [30 31 และ 32] จะกล่าวถึงรายละเอียดในหัวข้อถัดไป

3.1.2 รายละเอียดการทำงานของตัวเข้ารหัสหลัก MP-CELP

ในหัวข้อนี้ จะได้อธิบายการทำงานของแต่ละฟังก์ชันของตัวเข้ารหัสหลักตามรูปที่ 3.2 โดยความหมายของสัญญาณต่างๆ เป็นไปตามบัญชีสัญลักษณ์ การใช้งานฟังก์ชันต่างๆ ที่นำเสนอในวิทยานิพนธ์นี้จะยึดตามงานวิจัยของ Ozawa เป็นหลัก นั่นก็คือการตั้งค่าพารามิเตอร์ของการเข้ารหัสจะเป็นไปตามบทความของกลุ่มนักวิจัยนำโดย Ozawa [4 29 และ 30]

3.1.2.1 Pre-processing

สัญญาณเข้านั้นกำหนดให้เป็นสัญญาณ PCM ที่มีความละเอียด 16 บิต โดยฟังก์ชันทั้งสองส่วนใน pre-processing นั้นจะทำหน้าที่ 2 ประการ คือการปรับขนาดสัญญาณ และการกรองความถี่สูงผ่าน

การปรับขนาดสัญญาณนั้น จะทำการทอนสัญญาณเสียงพูดที่เป็นสัญญาณเข้าลงครึ่งหนึ่งเพื่อลดโอกาสเกิดการล้นในการคำนวณด้วยตัวประมวลผลแบบ fixed-point ส่วนวงจรกรอง

สัญญาณความถี่สูงผ่านมีไว้เพื่อใช้กำจัดสัญญาณความถี่ต่ำในช่วงที่ไม่ต้องการ โดยใช้วงจรกรองสัญญาณที่มีอันดับเท่ากับ 2 มีทั้งส่วนของ pole และ zero ที่มีความถี่ตัดที่ 140 Hz ทั้งวงจรปรับขนาดสัญญาณและวงจรกรองสัญญาณความถี่สูงผ่านนั้นประกอบรวมกันเป็นส่วนเดียวกันโดยการกำหนดให้สัมประสิทธิ์ของการขยายของวงจรกรองสัญญาณให้มีค่า 0.5 ซึ่งเป็นไปตามสมการ

$$H_{h1}(z) = \frac{0.46363718 - 0.92724705z^{-1} + 0.46363718z^{-2}}{1 - 1.9059465z^{-1} + 0.9114024z^{-2}} \qquad (3\text{-}1)$$

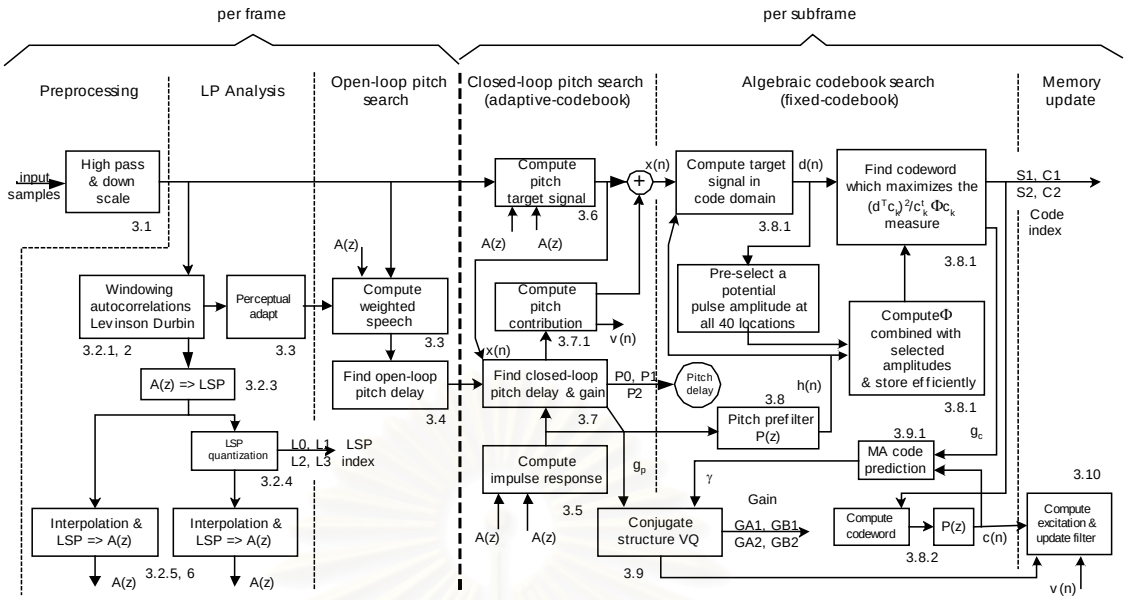
สัญญาณเข้า s(n) นั้นจะถูกกรองด้วย $H_{h1}(z)$ และสัญญาณที่ได้จะถูกป้อนให้กับวงจรส่วนต่างๆ ในการประมวลผลต่อไป

3.1.2.2 Linear prediction analysis and quantization

วงจรกรองสัญญาณที่ใช้ในการสังเคราะห์นั้นเป็นแบบระยะสั้น [33 และ 34] ที่มีโครงสร้างเป็นวงจรกรองทำนายเชิงเส้นที่มีอันดับเท่ากับ 10 ซึ่งฟังก์ชันส่งผ่านของวงจรเป็นไปตามสมการ

$$\frac{1}{A(z)} = \frac{1}{1 + \sum_{i=1}^{10} a_i z^{-i}} \quad , \ i=1,\ldots,10 \qquad (3\text{-}2)$$

เมื่อ a_i คือสัมประสิทธิ์ (ที่ถูกควอนไทซ์แล้ว) ของตัวทำนายเชิงเส้น การวิเคราะห์สัมประสิทธิ์นี้จะกระทำหนึ่งครั้งต่อเฟรม โดยใช้อัลกอริทึมของ Levinson บนพื้นฐานของกรรมวิธีอัตสหสัมพันธ์ (autocorrelation method) จากข้อมูลแบบไม่สมมาตรขนาด 30 มิลลิวินาที ของตัวอย่างของเสียงขาเข้าจำนวน 80 ตัวอย่าง (10 มิลลิวินาที) หลังจากได้สัมประสิทธิ์แล้วจะถูกเปลี่ยนเป็นค่า LSP เพื่อนำไปควอนไทซ์และใช้ในการประมาณค่าในช่วง ค่าที่ได้นี้จะถูกเปลี่ยนกลับเป็นสัมประสิทธิ์เพื่อใช้ในการสังเคราะห์และส่งให้วงจรกรองถ่วงน้ำหนัก (weighting filter) ใช้ในเฟรมย่อยแต่ละเฟรมต่อไป



รูปที่ 3.2 โครงสร้างการทำงานของตัวเข้ารหัส MP-CELP

3.1.2.2.1 ชุดสัญญาณสุ่มและวิธีการคำนวณอัตสหสัมพันธ์

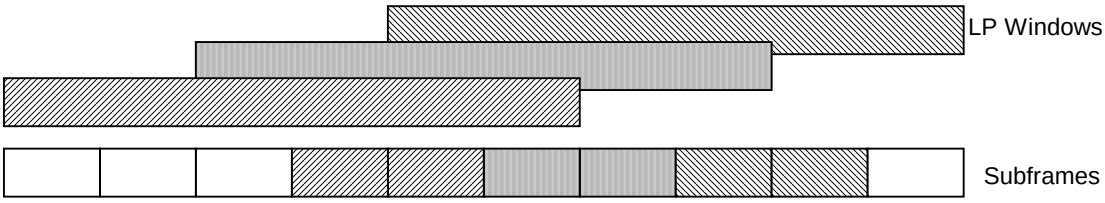
ชุดสัญญาณสุ่มที่ใช้ในการวิเคราะห์ตัวทำนายเชิงเส้นนั้นประกอบด้วย 2 ส่วน โดยส่วนแรกคือหน้าต่าง Hamming และในส่วนที่สองนั้นจะเป็นฟังก์ชันของโคไซน์ ดังนี้

$$\omega_p(n) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi n}{399}) & n = 0\text{K } 199 \\ \cos(\frac{2\pi(n-200)}{159}) & n = 200\text{K } 239 \end{cases} \qquad (3\text{-}3)$$

การวิเคราะห์ตัวทำนายเชิงเส้นนั้นจะใช้สัญญาณสุ่มล่วงหน้ามา 5 มิลลิวินาที หรือ 40 ตัวอย่างจากเฟรมถัดจากเฟรมปัจจุบัน หมายความว่าอัลกอริทึมนี้จะมีการประวิงเวลาที่ตัวเข้ารหัสเพิ่มขึ้นอีก 5 มิลลิวินาที และใช้อีก 120 ตัวอย่างจากเฟรมก่อนหน้าเฟรมปัจจุบัน รวมแล้วจะใช้จำนวนตัวอย่างทั้งสิ้น 240 ตัวอย่าง ซึ่งอธิบายได้ตามรูปที่ 3.3

ดังนั้นชุดของสัญญาณสุ่มขาเข้าที่ได้คือ

$$s'(n) = \omega_p(n)s(n), \quad n = 0,\text{K },239 \qquad (3\text{-}4)$$



รูปที่ 3.3 ชุดของสัญญาณสุ่มที่ใช้ในการวิเคราะห์ตัวทำนายเชิงเส้น โดยสีที่แรเงาแตกต่างกันหมายถึงเฟรมที่ต่างกัน

ดังนั้นสัมประสิทธิ์ของอัตสหสัมพันธ์ จะมีค่า

$$r(k) = \sum_{n=k}^{239} s'(n)s'(n-k), \quad k = 0,\text{K},10 \tag{3-5}$$

เพื่อหลีกเลี่ยงปัญหาการคำนวณเนื่องจากสัญญาณเข้าที่มีขนาดต่ำจะกำหนดให้ $r(0)$ มีค่าต่ำสุดคือเป็น 1.0 และกำหนดให้มีแบนด์วิดท์ 60 Hz โดยการคูณสัมประสิทธิ์ของอัตสหสัมพันธ์ ที่ได้ด้วย

$$\omega_{lag}(k) = \exp\left[-\frac{1}{2}\left(\frac{2\pi \quad k}{f_s}\right)^2\right] \quad ,k = 1,\text{K},10 \tag{3-6}$$

เมื่อ BW = 60 Hz คือแบนด์วิดท์ที่กำหนด และ f_s คือ ความถี่ที่ใช้ในการสุ่ม มีค่า 8000 Hz นอกจากนี้ $r(0)$ จะถูกคูณด้วยค่า 1.0001 [15] ซึ่งเป็น white-noise factor ดังนั้นสัมประสิทธิ์ของอัตสหสัมพันธ์ ใหม่จะมีค่าเป็น

$$\begin{aligned} r'(0) &= 1.0001r(0) \\ r'(k) &= \omega_{lag}(k)\, r(k) \qquad k = 1\ldots, 10 \end{aligned} \tag{3-7}$$

3.1.2.2.2 อัลกอริทึม Levinson-Durbin

ค่าสัมประสิทธิ์ของอัตสหสัมพันธ์ ใหม่ $r'(k)$ จะถูกนำมาใช้หาสัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้น คือ a_i เมื่อ $i = 1,\ldots, 10$ โดยการหาจากสมการนี้

$$\sum_{i=1}^{10} a_i r'(|\, i-k\,|) = -r'(k) \quad k = 1,...,10 \tag{3-8}$$

เราสามารถหาคำตอบของสมการที่ (3-8) ได้โดยใช้อัลกอริทึม Levinson-Durbin เป็นแบบวนซ้ำดังนี้

$E^{[0]} = r'(0)$

For i = 1 to 10

$a_0^{[i-1]} = 1$

$k_i = -[\sum_{j=0}^{i-1} a_j^{[i-1]} r'(i-j)] / E^{[i-1]}$

$a_j^{[i]} = k_i$

for j = 1 to i – 1

$a_j^{[i]} = a_j^{[i-1]} + k_i a_{i-j}^{[i-1]}$

end

$E^{[i]} = (1 - k_i^2) E^{[i-1]}$

end

โดยคำตอบสุดท้ายคือ $a_j = a_j^{[10]}$ เมื่อ j = 1, …, 10 และ a_0 = 1.0

3.1.2.2.3 การเปลี่ยนสัมประสิทธิ์ตัวทำนายเชิงเส้นเป็น LSP

สัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้น คือ a_i จะถูกเปลี่ยนเป็นสัมประสิทธิ์ของ LSP เพื่อนำไปควอนไทซ์และทำการประมาณค่าในช่วงต่อไป [4 และ 15] โดยสัมประสิทธิ์ LSP ของวงจรกรองสัญญาณที่มีอันดับเท่ากับ 10 สามารถกำหนดได้ตามสมการผลรวมและผลลบของโพลิ-โนเมียลตามลำดับดังนี้

$$F_1'(z) = A(z) + z^{-11}A(z^{-1}) \quad (3\text{-}9)$$

และ

$$F_2'(z) = A(z) - z^{11}A(z^{-1}) \quad (3\text{-}10)$$

โพลิโนเมียล $F_1'(z)$ จะสมมาตร ส่วน $F_2'(z)$ จะไม่สมมาตร ซึ่งสามารถพิสูจน์ได้จากรากของสมการโพลิโนเมียลนั้นจะอยู่บนวงกลมหนึ่งหน่วยและแตกต่างออกไปสำหรับอีกสมการ โดย $F_1'(z)$ จะมีรากที่ $z = -1$ ($\omega = \pi$) และ $F_2'(z)$ จะมีรากที่ $z = 1$ ($\omega = 0$) รากทั้งสองนี้ทำให้ได้โพลิโนเมียลชุดใหม่คือ

$$F_1(z) = F_1'(z)/(1 + z^{-1}) \quad (3\text{-}11)$$

และ

$$F_2(z) = F_2'(z)/(1 - z^{-1}) \quad (3\text{-}12)$$

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

$$F_1(z) = \prod_{i=1,3,...,9} (1 - 2q_i z^{-1} + z^{-2}) \quad (3\text{-}13)$$

และ

$$F_2(z) = \prod_{i=2,4,...,10} (1 - 2q_i z^{-1} + z^{-2}) \quad (3\text{-}14)$$

เมื่อ $q_i = \cos(\omega_i)$ ซึ่งสัมประสิทธิ์ ω_i นี้เรียกว่า line spectral frequency (LSF) และจะมีคุณสมบัติ $0 < \omega_1 < \omega_2 < \ldots < \omega_{10} < \pi$ ส่วนสัมประสิทธิ์ q_i นั้นคือสัมประสิทธิ์ LSP ที่อยู่ในโดเมนของโคไซน์

ทั้งโพลิโนเมียล $F_1(z)$ และ $F_2(z)$ นั้นจะสมมาตรเฉพาะสัมประสิทธิ์ 5 ตัวแรกเท่านั้น ซึ่งจะต้องทำการคำนวณหา สัมประสิทธิ์ของโพลิโนเมียลเหล่านี้สามารถหาได้จากวิธีการรีเคอร์ซีฟดังนี้

$$\begin{aligned} f_1(i+1) &= a_{i+1} + a_{10-i} - f_1(i), \quad i = 0, \ldots, 4 \\ f_2(i+1) &= a_{i+1} + a_{10-i} - f_2(i), \quad i = 0, \ldots, 4 \end{aligned} \quad (3\text{-}15)$$

เมื่อ $f_1(0) = f_2(0) = 1.0$ สัมประสิทธิ์ LSP นี้จะถูกคำนวณโดยการแทนค่าโพลิโนเมียล $F_1(z)$ และ $F_2(z)$ ที่ z ตำแหน่ง 60 จุด ระหว่าง 0 ถึง π และตรวจสอบการเปลี่ยนของเครื่องหมาย การเปลี่ยนเครื่องหมายนั้นจะมีความสำคัญต่อราก และช่วงที่มีการเปลี่ยนเครื่องหมายนี้จะถูกหารลง 4 เท่าเพื่อทำให้สามารถหาค่ารากแม่นยำมากยิ่งขึ้น โพลิโนเมียลแบบ Chebyshev จะถูกใช้หาค่าของ $F_1(z)$ และ $F_2(z)$ วิธีนี้จะให้รากที่อยู่ในโดเมนของโคไซน์โดยตรง โพลิโนเมียล $F_1(z)$ และ $F_2(z)$ นี้จะแทน $z = e^{j\omega}$ ได้ดังนี้

$$F(\omega) = 2\, e^{-j5\omega} C(x) \quad (3\text{-}16)$$

เมื่อ

$$C(x) = T_5(x) + f(1)T_4(x) + f(2)T_3(x) + f(3)T_2(x) + f(4)T_1(x) + f(5)/2 \quad (3\text{-}17)$$

เมื่อ $T_m(x) = \cos(m\omega)$ คือ โพลิโนเมียลแบบ Chebyshev อันดับที่ m และ $f(i)$, $i = 1\ldots.5$ คือสัมประสิทธิ์ของทั้ง $F_1(z)$ หรือ $F_2(z)$ ตามสมการที่ (17) โพลิโนเมียล $C(x)$ จะหาได้จาก $x = \cos(\omega)$ โดยใช้วิธีการแบบรีเคอร์ซีฟดังนี้

for $k = 4$ to 1

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

end

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

เมื่อค่าเริ่มต้น $b_5 = 1$ และ $b_6 = 0$

3.1.2.2.4 การควอนไทซ์สัมประสิทธิ์ LSP

สัมประสิทธิ์ LSP จะถูกควอนไทซ์ด้วยค่า LSF แทนด้วย ω_i โดยจะทำการทำให้เป็นบรรทัดฐานความถี่ในช่วง [0, π] นั่นคือ

$$\omega_i = \arccos(q_i), \qquad i = 1,\ldots, 10 \qquad (3\text{-}18)$$

ตัวทำนายสัญญาณแบบ MA ที่มีอันดับ 4 ถูกใช้ในการทำนายค่าสัมประสิทธิ์ LSF ของเฟรมปัจจุบัน ค่าความแตกต่างระหว่างค่าสัมประสิทธิ์ที่คำนวณได้กับค่าที่ทำนายได้จะถูกควอนไทซ์ด้วยตัวควอนไทซ์แบบเวกเตอร์(VQ) แบบ 2 ตอน โดยตอนแรกจะเป็น VQ ที่มีขนาด 1x10 ใช้ codebook $L1$ สำหรับข้อมูลจำนวน 128 รูปแบบ (7 บิต) ส่วนในสเตจที่สองเป็น VQ แบบ 10 บิตที่ถูกสร้างขึ้นจากชุด VQ ที่มีขนาด 1x5 จำนวน 2 ตัวโดยมี codebook เป็น $L2$ และ $L3$ สำหรับข้อมูล 32 รูปแบบ (5 บิต) ในแต่ละ VQ [35 และ 36]

ก่อนอื่นจะอธิบายขั้นตอนในการถอดรหัสก่อนซึ่งจะทำให้เข้าใจขั้นตอนในการควอนไทซ์ได้ง่ายขึ้น โดยแต่ละสัมประสิทธิ์ที่ได้จากการรวมกันของ codebook ทั้งสองคือ

$$l_i = \begin{cases} L1_i(J1) + L2_i(J2) & i = 1,....,5 \\ L1_i(J1) + L3_{i-5}(J3) & i = 6,...,10 \end{cases} \qquad (3\text{-}19)$$

เมื่อ $J1$, $J2$ และ $J3$ คือตัวชี้ codebook เพื่อหลีกเลี่ยงการเกิด sharp resonance ในวงจรกรองตัวทำนายเชิงเส้น ที่ควอนไทซ์แล้ว สัมประสิทธิ์ l_i จะต้องจัดเรียงให้ค่าสัมประสิทธิ์ของตัวข้างเคียงมีค่าระยะทาง J ต่ำสุด โดยการเรียงนี้สามารถกระทำได้ดังนี้

for $i = 2,\ldots,10$

if $(l_{i-1} > l_i - \mathrm{J})$

$$l_{i-1} = (l_i + l_{i-1} - \mathrm{J})/2$$

$$l_i = (l_i + l_{i-1} + \mathrm{J})/2$$

end

end

หลังจากเสร็จขั้นตอนการเรียงลำดับนี้แล้วเราจะได้สัมประสิทธิ์ LSF ที่ผ่านการ ควอนไทซ์แล้ว $\omega_i^{(m)}$ สำหรับเฟรมที่ m จากการรวมกันของขาออกของตัวควอนไทซ์ก่อนหน้า $l_i^{(m-k)}$ และขาออกของตัวควอนไทซ์ในปัจจุบัน $l_i^{(m)}$ ดังนี้

$$\omega_i^{(m)} = (1-\sum_{k=1}^{4} p_{i,k}) l_i^{(m)} + \sum_{k=1}^{4} p_{i,k} l_i^{(m-k)}, \quad i = 1,...,10 \tag{3-20}$$

เมื่อ $p_{i,k}$ คือสัมประสิทธิ์ของตัวทำนายสัญญาณ MA ซึ่งตัวทำนายสัญญาณ MA ที่ใช้นี้จะถูกกำหนดโดยบิต $C0$ เมื่อเริ่มต้นการทำงานค่า $l_i^{(k)}$ จะกำหนดโดย $l_i = i\pi/11$ สำหรับทุกค่า $k < 0$

หลังจากทำการคำนวณหา ω_i แล้วต่อไปวงจรกรองจะตรวจสอบความมีเสถียรภาพดังนี้

1. ทำการเรียงลำดับสัมประสิทธิ์ ω_i จากน้อยไปหามาก
2. ถ้า $\omega_1 < 0.005$ จะให้ $\omega_1 = 0.005$
3. ถ้า $\omega_{i+1} - \omega_i < 0.0391$ จะให้ $\omega_{i+1} = \omega_i + 0.0391, \quad i = 1,\ldots, 9$
4. ถ้า $\omega_{10} > 3.135$ จะให้ $\omega_{10} = 3.135$

ขั้นตอนในการเข้ารหัสพารามิเตอร์ LSF นี้สามารถอธิบายได้ดังนี้ ตัวทำนายสัญญาณ MA แต่ละตัวจากทั้ง 2 ตัวนั้นจะประมาณค่าจากค่าสัมประสิทธิ์ LSF ปัจจุบัน ค่าที่ประมาณขึ้นมาได้จะกำหนดให้เป็นค่า weighted mean-squared error ต่ำสุด

$$E_{lsf} = \sum_{i=1}^{10} w_i (\omega_i - \omega_i)^2 \tag{3-21}$$

โดยค่าน้ำหนักถ่วง w_i สามารถปรับเปลี่ยนได้ตามฟังก์ชันของสัมประสิทธิ์ LSF ที่ยังไม่ได้ควอนไทซ์ดังนี้

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

$$w_i \text{ สำหรับ } 2 \le i \le 9 = \begin{cases} 1.0 & if\ \omega_{i+1} - \omega_{i-1} - 1 > 0 \\ 10(\omega_{i+1} - \omega_{i-1} - 1)^2 + 1 & otherwise \end{cases} \tag{3-22}$$

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

เวกเตอร์ที่ถูกควอนไทซ์ในแต่ละเฟรม m ได้มาจาก

$$l_i = [\omega_i^{(m)} - \sum_{k=1}^{4} p_{i,k} l_i^{(m-k)}] / (1 - \sum_{k=1}^{4} p_{i,k}), \quad i = 1,...,10 \tag{3-23}$$

โดย codebook แรก ($L1$) จะถูกหาก่อนและทำให้ J1 ที่มีค่า mean-squared error (ที่ยังไม่ได้ให้น้ำหนัก) ต่ำสุดถูกเลือกขึ้นมา จากนั้นจึงทำการค้นหา codebook ที่สอง ($L2$) ซึ่งใช้กำหนดส่วน

ล่างของสเตจที่สอง ค่าน้ำหนักถ่วง MSE ในสมการที่ (3-21) จะถูกคำนวณหาและจะได้เวกเตอร์ *J*2 เมื่อค่าผิดพลาดในส่วนล่างนี้มีค่าต่ำสุด เมื่อทำการค้นหา codebook *J*1 และที่ได้จากส่วนล่างในสเตจที่สองคือ *J*2 แล้วจึงทำการค้นหา codebook *J*3 ในส่วนบนจากสเตจที่สองต่อไป โดยเวกเตอร์ *J*3 ที่ได้จะทำให้ค่า MSE ของน้ำหนักถ่วงนี้มีค่าต่ำสุด เวกเตอร์ l_i จะถูกนำมาเรียงลำดับใหม่สองครั้งโดยใช้ขั้นตอนตามที่กล่าวไปแล้ว กระบวนการนี้จะทำแยกกันสำหรับตัวทำนาย MA ทั้งสองชุด ตามที่ได้กล่าวไปแล้วนั้นเวกเตอร์ l_i ที่ได้จากกระบวนการนี้จะถูกนำมาเรียงลำดับใหม่สองครั้งแล้วจึงตรวจสอบเสถียรภาพ และจะได้ค่าสัมประสิทธิ์ของ LSF ω_1 ที่ถูก ควอนไทซ์แล้ว

3.1.2.2.5 การประมาณค่าในช่วงสัมประสิทธิ์ LSP

สัมประสิทธิ์ตัวทำนายเชิงเส้นทั้งที่ถูกควอนไทซ์และที่ไม่ได้ถูกควอนไทซ์ จะถูกใช้ในเฟรมย่อยที่สอง ส่วนในเฟรมย่อยแรกนั้นสัมประสิทธิ์ตัวทำนายเชิงเส้นจะได้มาจากการประมาณค่าในช่วงจากพารามิเตอร์ของเฟรมข้างเคียง จากค่าสัมประสิทธิ์ LSP ในโดเมนของโคไซน์ กำหนดให้ $q_i^{(\text{current})}$ คือค่าสัมประสิทธิ์ LSP ที่คำนวณได้จากเฟรมปัจจุบันขนาด 10 มิลลิวินาที และ $q_i^{(\text{previous})}$ คือค่าสัมประสิทธิ์ LSP ที่คำนวณได้จากเฟรมก่อนหน้าขนาด 10 มิลลิวินาที สัมประสิทธิ์ LSP ที่ได้จากการประมาณค่าในช่วง (ที่ไม่ได้ถูกควอนไทซ์) ในเฟรมย่อยแต่ละเฟรมจะมีค่าดังนี้

$$\begin{aligned} &\text{เฟรมย่อยที่ } 1: q_i^{(1)} = 0.5\, q_i^{(\text{previous})} + 0.5 q_i^{(\text{current})}, && i = 1,\ldots, 10, \\ &\text{เฟรมย่อยที่ } 2: q_i^{(2)} = q_i^{(\text{current})}, && i = 1,\ldots, 10, \end{aligned} \qquad (3\text{-}24)$$

ค่าสัมประสิทธิ์ LSP ที่ควอนไทซ์แล้วก็จะมีขั้นตอนการประมาณค่าในช่วงเหมือนกันเพียงแต่แทนสัญลักษณ์ในสมการที่ (24) จาก q_i ด้วย $\hat{q}_i$

3.1.2.2.6 การเปลี่ยนสัมประสิทธิ์ LSP เป็นสัมประสิทธิ์ตัวทำนายเชิงเส้น

ค่าสัมประสิทธิ์ LSP ที่ถูกควอนไทซ์และประมาณค่าในช่วงนั้นจะถูกเปลี่ยนกลับเป็นค่าสัมประสิทธิ์ตัวทำนายเชิงเส้น คือ a_i โดยการหาค่าสัมประสิทธิ์ของ $F_1(z)$ และ $F_2(z)$ จากสมการที่ (3-13) และ (3-14) เมื่อเราทราบค่าสัมประสิทธิ์ LSP ที่ถูกควอนไทซ์และทำการประมาณค่าในช่วงแล้ว ดังนั้นค่าสัมประสิทธิ์ $f_1(i)$ จะคำนวณได้จาก q_i โดยใช้วิธีการแบบวนซ้ำดังนี้ [37 และ 38]

for i = 1 to 5

$f_1(i) = -2q_{2i-1} f_1(i - 1) + 2 f_1(i - 2)$

for $j = i - 1$ down to 1

$f_1^{[i]}(j) = f_1^{[i-1]}(j) - 2q_{2i-1} f_1^{[i-1]}(j - 1) + f_1^{[i-1]}(j - 2)$

end

end

โดยมีค่าเริ่มต้น $f_1(0) = 1$ และ $f_1(-1) = 0$ ค่าสัมประสิทธิ์ $f_2(i)$ จะคำนวณหาได้จากวิธีเดียวกันโดยเปลี่ยนจาก $2q_{2i-1}$ 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

$$\begin{aligned} f_1'(i) &= f_1(i) + f_1(i - 1), \quad i = 1, \ldots, 5, \\ f_2'(i) &= f_2(i) + f_2(i - 1), \quad i = 1, \ldots, 5, \end{aligned} \tag{3-25}$$

สุดท้ายค่าสัมประสิทธิ์ตัวทำนายเชิงเส้น จะคำนวณหาได้จาก $f_1'(i)$ และ $f_2'(i)$ ดังนี้

$$a_i = \begin{cases} 0.5 f_1'(i) + 0.5 f_2'(i), & i = 1,...,5 \\ 0.5 f_1'(11 - i) - 0.5 f_2'(11 - i), & i = 6,...,10 \end{cases} \tag{3-26}$$

3.1.2.3 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 ใช้ในการกำหนดความถี่ตอบสนองของวงจรกรองสัญญาณ $W(z)$ โดยการปรับค่าตัวแปรเหล่านี้จะทำให้การถ่วงน้ำหนักที่เกิดขึ้นเป็นไปอย่างมีประสิทธิภาพ ซึ่งทำได้โดยให้ γ_1 และ γ_2 เป็นฟังก์ชันเชิงสเปกตรัมของสัญญาณเข้า การปรับแต่งค่านี้จะทำทุกๆ เฟรม แต่ขั้นตอนในการทำประมาณค่าในช่วงนั้น จะทำที่เฟรมย่อยแรกเพื่อทำให้การปรับแต่งที่ได้เป็นไปอย่างต่อเนื่อง รูปร่างของสเปกตรัมนี้ได้มาจากวงจรกรองทำนายสัญญาณที่มีอันดับ 2 โดยใช้วิธี Levinson-

Durbin (หัวข้อที่ 3.1.2.2) ค่าสัมประสิทธิ์ reflection k_i จะถูกเปลี่ยนเป็นสัมประสิทธิ์ LAR (Log Area Ratio) o_i ดังนี้

$$o_i = \log\frac{(1.0 + k_i)}{(1.0 - k_i} \quad i = 1,2 \tag{3-28}$$

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

$$\begin{aligned} &\text{เฟรมย่อยที่ 1 :} \quad o_i^{(1)} = 0.5\, o_i^{(\text{previous})} + 0.5\, o_i^{(\text{current})}, \quad i = 1,2 \\ &\text{เฟรมย่อยที่ 2 :} \quad o_i^{(2)} = o_i^{(\text{current})}, \quad i = 1,2 \end{aligned} \tag{3-29}$$

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

$$flat^{(m)} = \begin{cases} 0 & \text{if } o_1^{(m)} < -1.74 \text{ and } o_2^{(m)} > 0.65 \text{ and } flat^{(m-1)} = 1, \\ 1 & \text{if } (o_1^{(m)} > -1.52 \text{ or } o_2^{(m)} < 0.43) \text{ and } flat^{(m-1)} = 0, \\ flat^{(m-1)} & \text{otherwise} \end{cases} \tag{3-30}$$

ถ้าสเปกตรัมของสัมประสิทธิ์ที่ได้จากการประมาณค่าในช่วงของเฟรมย่อยเป็น flat ($flat^{(m)}$ = 1) จะกำหนดให้ $\gamma_1 = 0.94$ และ $\gamma_2 = 0.6$ ถ้าสเปกตรัมของสัมประสิทธิ์ที่ได้เป็น tilt ($flat^{(m)} = 0$) จะกำหนดให้ $\gamma_1 = 0.98$ และ γ_2 ปรับเปลี่ยนได้ตามขนาดของเรโซแนนซ์ที่เกิดขึ้นในวงจรกรองตัวทำนายเชิงเส้น ซึ่งมีค่าระหว่าง 0.4 ถึง 0.7 และถ้าเรโซแนนซ์ที่เกิดมีค่ามากก็จะกำหนดให้ γ_2 มีค่าเท่ากับ 0.7 การปรับเปลี่ยนทำได้โดยการวัดค่าระยะต่ำสุดระหว่างสัมประสิทธิ์ LSP 2 ชุดที่ได้จากเฟรมปัจจุบัน ดังนี้

$$d_{min} = min[\omega_{i+1} - \omega_i], \quad i = 1, \ldots, 9 \tag{3-31}$$

ค่า γ_2 นี้จะคำนวณได้โดยใช้ความสัมพันธ์ดังนี้

$$\gamma_2 = -0.6d_{min} + 1.0, \quad \text{โดย } 0.4 \le \gamma_2 \le 0.7 \tag{3-32}$$

สัญญาณเสียงที่ถูกถ่วงน้ำหนักในเฟรมย่อยนั้นจะได้มาจาก

$$sw(n) = s(n) + \sum_{i=1}^{10} a_i \gamma_1^i s(n-i) - \sum a_i \gamma_2^i sw(n-i), \quad i = 0,...,39 \tag{3-33}$$

สัญญาณเสียงที่ถูกถ่วงน้ำหนักนี้จะถูกใช้ในการประมาณค่าการประวิงเวลาของพิตช์ในแต่ละเฟรม [39]

3.1.2.4 การวิเคราะห์หาพิตช์ในวงเปิด

เพื่อลดความซับซ้อนในการค้นหาค่าการประวิงเวลาใน adaptive-codebook ช่วงในการค้นหาควรจะถูกจำกัดอยู่ระหว่างค่าการประวิงเวลา T_{op} ซึ่งหามาจากการวิเคราะห์พิตช์วงเปิดได้ [40 และ 41] การวิเคราะห์นี้จะทำทุกๆ เฟรม (10 มิลลิวินาที) โดยพิตช์วงเปิดนั้นจะประมาณค่าได้โดยใช้สัญญาณเสียงที่ถูกถ่วงน้ำหนักแล้ว $sw(n)$ ตามสมการที่ (3-33) แล้วทำการหาค่าที่อัตสหสัมพันธ์กันมากที่สุด 3 ค่าดังนี้

$$R(k) = \sum_{n=0}^{79} sw(n)sw(n-k) \tag{3-34}$$

ซึ่งจะหาค่าใน 3 ช่วงดังนี้

$i = 1$: 80,…,143,

$i = 2$: 40,…,79,

$i = 3$: 20,…,39

ค่าที่มากที่สุดของ $R(t_i)$, $i = 1,2,3$ จะถูกทำให้เป็นบรรทัดฐานดังนี้

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_n sw^2(n-t_i)}}, \quad i = 1,2,3 \tag{3-35}$$

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

ขั้นตอนนี้จะแบ่งช่วงการประวิงเวลาออกเป็น 3 ช่วงและใช้ค่าที่น้อยที่สุดเพื่อเลี่ยงการเลือกพิตช์หลายค่า

3.1.2.5 การคำนวณหาการตอบสนองอิมพัลส์

การตอบสนองอิมพัลส์ (impulse response) $h(n)$ ของวงจรกรองสัญญาณ weighted synthesis $W(z)/\hat{A}(z) = A(z/\gamma_1)/[\hat{A}(z)A(z/\gamma_2)]$ นั้นจำเป็นต้องใช้ในการค้นหาใน adaptive-codebook และ fixed-codebook โดยจะทำการคำนวณทุกๆ เฟรมย่อยโดยการกรองสัญญาณที่ประกอบด้วยสัมประสิทธิ์ของวงจรกรอง $A(z/\gamma_1)$ ที่ขยายด้วยรากที่เป็นศูนย์ด้วยวงจรกรอง $1/\hat{A}(z)$ และ $A(z/\gamma_2)$

3.1.2.6 การคำนวณหาสัญญาณของเป้า

สัญญาณของเป้า $x(n)$ ที่ใช้ในการค้นหาใน adaptive-codebook จะคำนวณมาจากการลบการตอบสนองของวงจรกรอง weighted synthesis $W(z)/\hat{A}(z) = A(z/\gamma_1)/[\hat{A}(z)A(z/\gamma_2)]$ ที่มีขาเข้าเป็นศูนย์ ออกจากสัญญาณเสียงที่ถูกถ่วงน้ำหนัก $sw(n)$ ตามสมการที่ (3-35) ซึ่งจะทำการคำนวณทุกๆ เฟรมย่อย

ขั้นตอนที่ใช้ในการคำนวณหาสัญญาณของเป้า ที่ใช้ในข้อกำหนดนี้คือการกรองตัวทำนายเชิงเส้น residual $r(n)$ โดยการรวมวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์ $1/A(z)$ และ วงจรกรองสัญญาณถ่วงน้ำหนัก $A(z/\gamma_1)/A(z/\gamma_2)$ หลังจากการหาสัญญาณกระตุ้นของเฟรมย่อยแต่ละเฟรมแล้วค่าเริ่มต้นของ วงจรกรองสัญญาณเหล่านี้จะถูกปรับให้ทันกาลโดยการกรองสัญญาณที่เกิดจาก

ผลต่างระหว่างสัญญาณ residual กับสัญญาณกระตุ้น การปรับให้ทันกาลในหน่วยความจำของวงจรกรองสัญญาณเหล่านี้ได้อธิบายไว้ในหัวข้อที่ 3.1.2.10

สัญญาณ residual ที่จำเป็นต้องใช้ในการหาเวกเตอร์ของเป้า จะถูกใช้ในการค้นหาใน adaptive-codebook กล่าวคือขั้นตอนการหาค่าการประวิงเวลาของ adaptive-codebook จะน้อยกว่าขนาดของเฟรมย่อย จะได้อธิบายในหัวข้อถัดไป โดยสัญญาณ residual ของตัวทำนายเชิงเส้น มีค่าดังนี้

$$r(n) = s(n) + \sum_{i=1}^{10} a_i s(n-i), \quad n = 0,...,39 \tag{3-36}$$

3.1.2.7 การค้นหาของ adaptive-codebook

ตัวพารามิเตอร์ของ adaptive-codebook คือการประวิงเวลาและอัตราขยาย โดยใช้วงจรกรองสัญญาณพิตช์ซึ่งจะทำการค้นหาสัญญาณกระตุ้นที่มีการประวิงเวลาน้อยกว่าความยาวน้อยกว่าขนาดของเฟรมย่อย ขั้นตอนการค้นหานั้น สัญญาณกระตุ้นจะถูกขยายด้วยสัญญาณ residual ของตัวทำนายเชิงเส้น เพื่อทำให้การค้นหาในวงปิดเป็นไปโดยง่าย ดังนั้นการค้นหา adaptive-codebook นั้นจะกระทำทุกๆ เฟรมย่อย ในเฟรมย่อยแรกส่วนของการประวิงเวลาของพิตช์ T_1 ส่วนในเฟรมย่อยที่สองค่าการประวิงเวลา T_2 โดยคำนวณค่าเป็นจำนวนเต็ม [42 43 และ 44]

ในเฟรมย่อยแต่ละเฟรมนั้นการหาค่าการประวิงเวลาที่ถูกต้องที่สุดทำได้โดยการวิเคราะห์ในวงปิด เพื่อให้ค่า MSE ของค่าน้ำหนักถ่วงมีค่าต่ำสุด ในเฟรมย่อยแรกหลังจากได้ค่าการประวิงเวลา T_1 โดยการค้นหาจากช่วงสั้นรอบๆ ค่าการประวิงเวลา T_{op} ที่ได้จากวงเปิด (ดูหัวข้อ 3.1.2.4) ซึ่งขอบเขตการค้นหา 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 ที่ดีที่สุด โดยขอบเขตในการค้นหาจะอยู่ระหว่าง 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

ในการค้นหาพิตช์ของวงปิด นั้นจะหาค่าที่น้อยที่สุดของ mean-squared weighted error ระหว่างสัญญาณเสียงดังเดิมกับสัญญาณเสียงที่สร้างขึ้นมา โดยการหาค่าที่มากที่สุดจากพจน์นี้

$$R(k) = \frac{\sum_{n=0}^{39} x(n) y_k(n)}{\sqrt{\sum_{n=0}^{39} y_k(n) y_k(n)}} \qquad (3\text{-}37)$$

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

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

$$y_k(n) = y_{k-1}(n - 1) + u(-k)h(n) \qquad n = 39, \ldots, 0 \qquad (3\text{-}38)$$

เมื่อ $u(n)$, $n = -143, \ldots, 39$ คือบัฟเฟอร์ของสัญญาณกระตุ้น และ $y_{k-1}(-1) = 0$ จะเห็นได้ว่าในขั้นตอนการค้นหานั้นสัญญาณตัวอย่าง $u(n)$, $n = 0, \ldots, 39$ นั้นไม่ทราบค่า จึงทำให้จำเป็นต้องหาค่าการประวิงเวลาของพิตช์ในช่วงที่น้อยกว่า 40 วิธีการอย่างง่ายคือสัญญาณ residual ของ ตัวทำนายเชิงเส้น จะถูกสำเนาไปเป็น $u(n)$ เพื่อทำให้สมการที่ (3-36) เป็นจริงสำหรับทุกค่าการประวิงเวลา

3.1.2.7.1 การสร้างเวกเตอร์ adaptive-codebook

เมื่อทำการหาค่าการประวิงเวลาของพิตช์ได้แล้วจะทำการคำนวณหาเวกเตอร์ของ codebook $v(n)$ โดยหาจากการประมาณค่าในช่วงจากสัญญาณกระตุ้น $u(n)$ ในอดีตที่มีการประวิงเวลาเป็นจำนวนเต็ม k ดังนี้

$$v(n) = \sum_{i=0}^{9} u(n-k-i)a(i) + \sum_{i=0}^{9} u(n-k+1+i)a(i), \quad n = 0,...,39 \qquad (3\text{-}39)$$

a(i) เป็นค่าถ่วงน้ำหนักเชิงเส้น

3.1.2.7.2 การเข้ารหัสค่าการประวิงเวลาของ adaptive-codebook

การประวิงเวลาของพิตช์ T_1 จะถูกเข้ารหัสขนาด 6 บิตในเฟรมย่อยแรก ส่วนการประวิงเวลา T_2 ของเฟรมย่อยที่สองนั้นจะถูกเข้ารหัสขนาด 3 บิต เพราะฉะนั้นพารามิเตอร์พิตช์ $P1$ ของเฟรมย่อยแรก และ $P2$ ของเฟรมย่อยหลัง

$$P1 = T_1 \qquad \text{T1} = 20,\ldots,143 \qquad (3\text{-}40)$$

$$P2 = T_2 \qquad \text{T1} = 20,\ldots,143 \qquad (3\text{-}41)$$

การสร้างตัวเข้ารหัสถูกออกแบบให้มีความทนทานต่อการผิดพลาดของบิตข้อมูลนั้นโดยใช้พาริตีบิต $P0$ ซึ่งคำนวณมาจากตัวชี้ $P1$ ของเฟรมย่อยแรก พาริตีบิตที่ได้นั้นมาจากการทำ XOR จากบิตที่มีนัยสำคัญสูงสุด 6 บิตของ $P1$ ที่ตัวถอดรหัสพาริตีบิตนี้จะถูกคำนวณเพื่อใช้ในการตรวจสอบการผิดพลาดของบิตข้อมูล ซึ่งถ้าได้ค่าไม่ตรงกันก็จะใช้วิธีการแก้ไขข้อผิดพลาดเข้ามาช่วย

3.1.2.7.3 การคำนวณหาอัตราขยายของ adaptive-codebook

เมื่อหาค่าการประวิงเวลาของ adaptive-codebook ได้แล้วจะทำการคำนวณหาอัตราขยาย g_p ดังนี้

$$g_p = \frac{\sum_{n=0}^{39} x(n)y(n)}{\sum_{n=0}^{39} y(n)y(n)} \qquad \text{โดย } 0 \le g_p \le 1.2 \qquad (3\text{-}42)$$

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

$$y(n) = \sum_{i=0}^{n} v(i)h(n-i), \quad n = 0,...,39 \qquad (3\text{-}43)$$

3.1.2.8 โครงสร้างและการหา fixed-codebook

ในวิทยานิพนธ์ กำหนดให้ codevector ที่เป็นมัลติพัลส์ ประกอบด้วยพัลส์ 1 พัลส์, 5 พัลส์ และ 10 พัลส์ (ที่มีค่าไม่เป็นศูนย์) การเลือกจำนวนพัลส์นี้เพราะเมื่อมีการเพิ่มส่วนขยายเพื่อปรับอัตราการเข้ารหัสในหัวข้อ 3.3 จะไม่ทำให้จำนวนพัลส์ซ้ำซ้อนกัน และอัตราการเข้ารหัสจะครอบ

คลุมเต็มช่วงกว้างที่เป็นไปได้ แต่ละพัลส์มีขนาดเป็น +1 และ –1 เริ่มจากส่วน sign codebook จะกำหนดเครื่องหมายสำหรับพัลส์แต่ละพัลส์ล่วงหน้า โดยใช้ d(n) ที่เป็นสหสัมพันธ์ข้ามระหว่าง x' (n) และ h(n) เป็นหลักในการกำหนด ส่วน location generator จะกำหนดตำแหน่งของพัลส์เพื่อส่งผ่านไปยังกระบวนการค้นหาอีกที และ location restriction มีโครงสร้างเป็น codebook แบบพีชคณิต (algebraic) ตามตารางที่ 3.1 3.2 และ 3.3 ออกแบบโดยใช้ interleaved single-pulse permutation (ISPP) [4 และ 30]

ตารางที่ 3.1 โครงสร้างของ fixed-codebook *C* กรณี 1 พัลส์

พัลส์	ขนาด	ตำแหน่ง
i_0	s_0 : ±1	m_0 : 0, 5, 10, 15, 20, 25, 30, 35 1, 6, 11, 16, 21, 26, 31, 36 2, 7, 12, 17, 22, 27, 32, 37 3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39

ตารางที่ 3.2 โครงสร้างของ fixed-codebook *C* กรณี 5 พัลส์

พัลส์	ขนาด	ตำแหน่ง
i_0	s_0 : ±1	m_0 : 0, 5, 10, 15, 20, 25, 30, 35
i_1	s_1 : ±1	m_1 : 1, 6, 11, 16, 21, 26, 31, 36
i_2	s_2 : ±1	m_2 : 2, 7, 12, 17, 22, 27, 32, 37
i_3	s_3 : ±1	m_3 : 3, 8, 13, 18, 23, 28, 33, 38
i_4	s_4 : ±1	m_4 : 4, 9, 14, 19, 24, 29, 34, 39

ตารางที่ 3.3 โครงสร้างของ fixed-codebook *C* กรณี 10 พัลส์

พัลส์	ขนาด	ตำแหน่ง
i_0 , i_5	s_0 , s_5: ±1	m_0 , m_5 : 0, 5, 10, 15, 20, 25, 30, 35
i_1 , i_6	s_1 , s_6: ±1	m_1 , m_6 : 1, 6, 11, 16, 21, 26, 31, 36
i_2 , i_7	s_2 , s_7: ±1	m_2 , m_7 : 2, 7, 12, 17, 22, 27, 32, 37
i_3 , i_8	s_3 , s_8: ±1	m_3 , m_8 : 3, 8, 13, 18, 23, 28, 33, 38
i_4 , i_9	s_4 , s_9: ±1	m_4 , m_9 : 4, 9, 14, 19, 24, 29, 34, 39

เวกเตอร์ของ codebook $c(n)$ จะถูกสร้างขึ้นโดยใช้เวกเตอร์ศูนย์ที่มีขนาดเป็น 1x40 และหาตำแหน่งของพัลส์ทั้ง 4 ที่มีการคูณด้วยเครื่องหมายในแต่ละตำแหน่งดังนี้

กรณี 1 พัลส์

$$c(n) = s_0\delta(n - m_0), \qquad n = 0, \ldots, 39 \quad (3\text{-}44a)$$

กรณี 5 พัลส์

$$c(n) = s_0\delta(n - m_0) + s_1\delta(n - m_1) + s_2\delta(n - m_2) + s_3\delta(n - m_3) + s_4\delta(n - m_4), \qquad n = 0, \ldots, 39 \quad (3\text{-}44b)$$

กรณี 10 พัลส์

$$c(n) = s_0\delta(n - m_0) + s_1\delta(n - m_1) + s_2\delta(n - m_2) + s_3\delta(n - m_3) + s_4\delta(n - m_4) + s_5\delta(n - m_5) + s_6\delta(n - m_6) + s_7\delta(n - m_7) + s_8\delta(n - m_8) + s_9\delta(n - m_9), \qquad n = 0, \ldots, 39 \quad (3\text{-}44c)$$

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

$$P(z) = 1/(1 - \beta z^{-T}) \qquad (3\text{-}45)$$

เมื่อ T คือส่วนที่เป็นจำนวนเต็มของค่าการประวิงเวลาพิตช์ของเฟรมย่อยปัจจุบัน และ β คืออัตราขยายของพิตช์ ซึ่งจะมีค่าปรับเปลี่ยนได้โดยการควอนไทซ์อัตราขยายของ adaptive-codebook จากเฟรมย่อยก่อนหน้าดังนี้

$$\beta = g_p^{(m-1)} \quad \text{โดย } 0.2 \le \beta \le 0.8 \qquad (3\text{-}46)$$

ถ้าการประวิงเวลามีค่าน้อยกว่า 40 ตัวอย่าง codebook $c(n)$ ในสมการที่ (3-44a b และ c) จะถูกปรับเปลี่ยนตามนี้

$$c(n) = \begin{cases} c(n), & n = 0,\ldots,T-1, \\ c(n) + \beta c(n-T), & n = T,\ldots,39 \end{cases} \qquad (3\text{-}47)$$

การปรับเปลี่ยนของ fixed-codebook นี้จะต้องทำการปรับเปลี่ยนการตอบสนองอิมพัลส์ $h(n)$ ตามสมการดังนี้

$$h(n) = \begin{cases} h(n), & n = 0,...,T-1, \\ h(n) + \beta h(n-T), & n = T,...,39 \end{cases} \quad (3\text{-}48)$$

3.1.2.8.1 ขั้นตอนการหา fixed-codebook

การหา fixed-codebook นั้นทำได้โดยการหาค่าต่ำสุดของ mean-squared error ระหว่างสัญญาณเสียงที่ถูกถ่วงน้ำหนัก $sw(n)$ ตามสมการที่ (3-33) และสัญญาณเสียงที่สร้างกลับขึ้นมาที่ถูกถ่วงน้ำหนักแล้ว สัญญาณของเป้า ที่ใช้ในการหาพิตช์ในวงปิด จะถูกปรับให้ทันกาลโดยการลบด้วยค่าของ adaptive-codebook ดังนี้

$$x'(n) = x(n) - g_p y(n) \qquad n = 0, \ldots, 39 \quad (3\text{-}49)$$

เมื่อ $y(n)$ คือเวกเตอร์ของ adaptive-codebook ที่ผ่านการกรองสัญญาณตามสมการที่ (3-43) แล้ว และ g_p คืออัตราขยายของ adaptive-codebook ตามสมการที่ (3-42)

กำหนดเมทริกซ์ Φ มีสมบัติเป็นเมทริกซ์แบบสมมาตรประกอบด้วยสมาชิกที่เป็นค่าอัตสหสัมพันธ์ของ $h(n)$ ดังนี้

$$\phi(i,j) = \sum_{n=j}^{39} h(n-i)h(n-j), \quad i = 0,..,39, \quad j = i,...,39 \quad (3\text{-}50)$$

สัญญาณสหสัมพันธ์ข้าม $d(n)$ ได้มาจากสัญญาณของเป้า $x'(n)$ และการตอบสนองอิมพัลส์ $h(n)$ ดังนี้

$$d(n) = \sum_{i=n}^{39} x'(i)h(i-n), \quad n = 0,...,39 \quad (3\text{-}51)$$

ถ้า c_k คือเวกเตอร์ของ fixed-codebook ที่ k จะสามารถหา codebook ได้โดยทำให้พจน์ต่อไปนี้มีค่ามากที่สุด

$$\frac{C_k^2}{E_k} = \frac{\left(\sum_{n=0}^{39} d(n)c_k(n)\right)^2}{c_k^t \Phi c_k} \quad (3\text{-}52)$$

เมื่อ c_k^t คือค่าทรานสโพสต์ของ c_k

สัญญาณ $d(n)$ และเมทริกซ์ Φ จะถูกคำนวณก่อนที่จะหา codebook จะเห็นได้ว่าจะต้องมีการคำนวณหาค่าต่างๆ ดังนั้นจึงต้องออกแบบขั้นตอนการจัดเก็บข้อมูลอย่างมีประสิทธิภาพเพื่อเพิ่มความเร็วในการค้นหา

โครงสร้างที่เป็นแบบ algebraic ของ codebook C จะทำให้ขั้นตอนในการค้นหาเป็นไปอย่างรวดเร็ว แต่เวกเตอร์ของ codebook c_k จะมีจำนวนพัลส์เพียง 1 5 หรือ 10 พัลส์เท่านั้นที่มีค่าไม่เป็นศูนย์ การทำสหสัมพันธ์ข้ามตามสมการที่ (3-52) สำหรับเวกเตอร์ c_k จึงสามารถแสดงได้ดังนี้

กรณี 1 พัลส์

$$C = \sum_{i=0}^{0} s_i d(m_i) \tag{3-53a}$$

กรณี 5 พัลส์

$$C = \sum_{i=0}^{4} s_i d(m_i) \tag{3-53b}$$

กรณี 10 พัลส์

$$C = \sum_{i=0}^{9} s_i d(m_i) \tag{3-53c}$$

เมื่อ m_i คือตำแหน่งของพัลส์ที่ i และ s_i คือขนาดของพัลส์นั้น พลังงานของสัญญาณตามสมการที่ (3-52) นั้นมีค่าดังนี้

กรณี 1 พัลส์

$$E = \sum_{i=0}^{0} \varphi(m_i, m_i) \tag{3-54a}$$

กรณี 5 พัลส์

$$E = \sum_{i=0}^{4} \varphi(m_i, m_i) + 2\sum_{i=0}^{3}\sum_{j=i+1}^{4} s_i s_j \varphi(m_i, m_j) \tag{3-54b}$$

กรณี 10 พัลส์

$$E = \sum_{i=0}^{9} \varphi(m_i, m_i) + 2\sum_{i=0}^{8}\sum_{j=i+1}^{9} s_i s_j \varphi(m_i, m_j) \tag{3-54c}$$

เพื่อทำให้ขั้นตอนการทำงานนี้ง่ายขึ้น ขนาดของพัลส์จะถูกกำหนดโดยนำค่ามาจากการควอนไทซ์สัญญาณ $d(n)$ ทำได้โดยการกำหนดให้เครื่องหมายของพัลส์ที่ตำแหน่งใดๆ มีค่าเท่ากับเครื่องหมายของสัญญาณ $d(n)$ ที่ตำแหน่งนั้น ดังนั้นจะมีขั้นตอนก่อนทำการหา codebook คือ ขั้นแรกสัญญาณ $d(n)$ จะถูกแบ่งเป็นสองส่วนได้แก่ค่าสัมบูรณ์ $|d(n)|$ และเครื่องหมาย $\text{sign}[d(n)]$ ขั้นที่สองเมทริกซ์ Φ จะถูกปรับค่าโดยจะรวมเครื่องหมายของสัญญาณดังกล่าวเข้าไปด้วยดังนี้

$$\phi'(i,j) = \text{sign}[d(i)]\ \text{sign}[d(j)]\ \phi(i,j), \qquad i = 0,\ldots, 39,\ \ j = i+1, \ldots, 39 \tag{3-55}$$

ค่าในแนวทแยงมุมของ Φ จะถูกปรับขนาดโดยการหารสมการที่ (3-54) ด้วย 2

$$\phi'(i,j) = 0.5\phi(i,j), \qquad i = 0,\ldots, 39 \tag{3-56}$$

การทำสหสัมพันธ์ข้ามในสมการที่ (3-53) จะหาได้จาก

กรณี 1 พัลส์

$$C = |d(m_0)| \tag{3-57a}$$

กรณี 5 พัลส์

$$C = |d(m_0)| + |d(m_1)| + |d(m_2)| + |d(m_3)| + |d(m_4)| \tag{3-57b}$$

กรณี 10 พัลส์

$$C = |d(m_0)| + |d(m_1)| + |d(m_2)| + |d(m_3)| + |d(m_4)| + |d(m_5)| + |d(m_6)| + |d(m_7)| + |d(m_8)| + |d(m_9)| \tag{3-57c}$$

และพลังงานในสมการที่ (3-54) จะมีค่า

กรณี 1 พัลส์

$$E/2 = \phi'(m_0, m_0) \tag{3-55a}$$

กรณี 5 พัลส์

$$\begin{aligned} E/2 = {} & \phi'(m_0, m_0) \\ & + \phi'(m_1, m_1) + \phi'(m_0, m_1) \\ & + \phi'(m_2, m_2) + \phi'(m_0, m_2) + \phi'(m_1, m_2) \\ & + \phi'(m_3, m_3) + \phi'(m_0, m_3) + \phi'(m_1, m_3) + \phi'(m_2, m_3) \\ & + \phi'(m_4, m_4) + \phi'(m_0, m_4) + \phi'(m_1, m_4) + \phi'(m_2, m_4) + \phi'(m_3, m_4) \end{aligned} \tag{3-55b}$$

กรณี 10 พัลส์

ส่วนของ 5 พัลส์แรก

$$\begin{aligned} E/2 = {} & \phi'(m_0, m_0) \\ & + \phi'(m_1, m_1) + \phi'(m_0, m_1) \\ & + \phi'(m_2, m_2) + \phi'(m_0, m_2) + \phi'(m_1, m_2) \\ & + \phi'(m_3, m_3) + \phi'(m_0, m_3) + \phi'(m_1, m_3) + \phi'(m_2, m_3) \\ & + \phi'(m_4, m_4) + \phi'(m_0, m_4) + \phi'(m_1, m_4) + \phi'(m_2, m_4) + \phi'(m_3, m_4) \end{aligned} \tag{3-55ca}$$

ส่วนของ 5 พัลส์หลัง

$$
\begin{aligned}
E/2 = \quad & \phi'(m_5, m_5) \\
& + \phi'(m_6, m_6) + \phi'(m_5, m_6) \\
& + \phi'(m_7, m_7) + \phi'(m_5, m_7) + \phi'(m_6, m_7) \\
& + \phi'(m_8, m_8) + \phi'(m_5, m_8) + \phi'(m_6, m_8) + \phi'(m_7, m_8) \\
& + \phi'(m_9, m_9) + \phi'(m_5, m_9) + \phi'(m_6, m_9) + \phi'(m_7, m_9) + \phi'(m_8, m_9)
\end{aligned}
\tag{3-55cb}
$$

สมการ (3.55ca) ใช้สำหรับการค้นหา 5 พัลส์แรก ส่วนสมการ (3.55cb) ใช้สำหรับการค้นหา 5 พัลส์หลัง นั่นคือตำแหน่งหลังจะเป็นตำแหน่งที่ให้ความสำคัญรองจาก 5 พัลส์แรก

3.1.2.8.2 การเข้ารหัส fixed-codebook

กรณี 1 พัลส์

ตำแหน่งของแต่ละพัลส์ i_0 จะถูกเข้ารหัสที่มีความละเอียด 6 บิต คือเป็นตำแหน่ง track 3 บิต และตำแหน่งของ offset อีก 3 บิต โดยขนาดของพัลส์แต่ละพัลส์นั้นจะถูกเข้ารหัสขนาด 1 บิต ไว้ ดังนั้นจะมีทั้งหมด 7 บิต กำหนดให้ $s = 1$ ถ้าเครื่องหมายมีค่าเป็นบวก และ $s = 0$ ถ้าค่าที่ได้เป็นลบ ดังนั้นคำรหัสของเครื่องหมายที่ได้ จะเป็น

$$S = s_0 \tag{3-56a}$$

และคำรหัสของตำแหน่งพัลส์ ของ fixed-codebook จะมีค่าดังนี้

$$C = 8(m_0/5) + \mathrm{x} \tag{3-57a}$$

เมื่อ $x = 0$ ถ้า $m_0 = 0, 5, \ldots, 35$

$x = 1$ ถ้า $m_0 = 1, 6, \ldots, 36$

$x = 2$ ถ้า $m_0 = 2, 7, \ldots, 37$

$x = 3$ ถ้า $m_0 = 3, 8, \ldots, 38$

และ $x = 4$ ถ้า $m_0 = 4, 9, \ldots, 39$

กรณี 5 พัลส์

ตำแหน่งของแต่ละพัลส์ i_0 i_1 i_2 i_3 และ i_4 จะถูกเข้ารหัสที่มีความละเอียด 3 บิต โดยขนาดของแต่ละพัลส์นั้นจะถูกเข้ารหัสขนาด 1 บิตไว้ ดังนั้นจะมีทั้งหมด 20 บิตจากพัลส์ทั้งห้า โดย

กำหนดให้ $s = 1$ ถ้าเครื่องหมายมีค่าเป็นบวก และ $s = 0$ ถ้าค่าที่ได้เป็นลบ ดังนั้นคำรหัสของเครื่องหมายที่ได้ จะเป็น

$$S = s_0 + 2s_1 + 4s_2 + 8s_3 + 16 s_4 \quad (3\text{-}56b)$$

และคำรหัสของตำแหน่งพัลส์ ของ fixed-codebook จะมีค่าดังนี้

$$C = (m_0/5) + 8(m_1/5) + 64(m_2/5) + 512(m_3/5) + 4096(m_4/5) \quad (3\text{-}57b)$$

กรณี 10 พัลส์

ตำแหน่งของแต่ละพัลส์ $i_0\ i_1\ i_2\ i_3\ i_4\ i_5\ i_6\ i_7\ i_8$ และ i_9 จะถูกเข้ารหัสที่มีความละเอียด 3 บิต โดยขนาดของแต่ละพัลส์นั้นจะถูกเข้ารหัสขนาด 1 บิตไว้ ดังนั้นจะมีทั้งหมด 40 บิตจากพัลส์ทั้งห้า โดยกำหนดให้ $s = 1$ ถ้าเครื่องหมายมีค่าเป็นบวก และ $s = 0$ ถ้าค่าที่ได้เป็นลบ ดังนั้นคำรหัสของเครื่องหมายที่ได้ จะเป็น

$$S = s_0 + 2s_1 + 4s_2 + 8s_3 + 16 s_4 + 32s_5 + 64s_6 + 128s_7 + 256s_8 + 512 s_9 \quad (3\text{-}56c)$$

และคำรหัสของตำแหน่งพัลส์ ของ fixed-codebook จะมีค่าดังนี้

$$C1 = (m_0/5) + 8(m_1/5) + 64(m_2/5) + 512(m_3/5) + 4096(m_4/5) \quad (3\text{-}57ca)$$

และ $$C2 = (m_5/5) + 8(m_6/5) + 64(m_7/5) + 512(m_8/5) + 4096(m_9/5) \quad (3\text{-}57cb)$$

3.1.2.9 การควอนไทซ์อัตราขยาย

อัตราขยายของ adaptive-codebook (อัตราขยายพิตช์) และอัตราขยายของ fixed-codebook จะถูกควอนไทซ์แบบเวกเตอร์ที่มีขนาด 7 บิต การหาอัตราขยายของ codebook ทำโดยหาค่าต่ำสุดของ mean squared weighted error ระหว่างสัญญาณเสียงเดิมกับสัญญาณเสียงที่สร้างขึ้นมาดังนี้

$$E = \mathrm{x}^t\mathrm{x} + g_p^2\mathrm{y}^t\mathrm{y} + g_c^2\mathrm{z}^t\mathrm{z} - 2g_p\mathrm{x}^t\mathrm{y} - 2g_c\mathrm{x}^t\mathrm{z} + 2g_pg_c\mathrm{y}^t\mathrm{z} \quad (3\text{-}58)$$

เมื่อ x คือเวกเตอร์ของเป้า (ดูหัวข้อที่ 3.1.2.6) , y คือเวกเตอร์ของ adaptive-codebook ที่ถูกกรองสัญญาณแล้วตามสมการที่ (3-43) และ z คือเวกเตอร์ของ fixed-codebook ที่คอนโวลูชันกับ $h(n)$

$$z(n) = \sum_{i=0}^{n} c(i)h(n-i) \quad n = 0,...,39 \quad (3\text{-}59)$$

3.1.2.9.1 การทำนายอัตราขยาย

อัตราขยายของ fixed-codebook g_c เขียนได้ดังนี้

$$g_c = \gamma g_c' \tag{3-60}$$

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

$$E = 10\log\left(\frac{1}{40}\sum_{n=0}^{39} c(n)^2\right) \tag{3-61}$$

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

$$E^{(\mathrm{m})} = 20\log g_c + E - \overline{E} \tag{3-62}$$

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

$$g_c = 10^{(E^{(m)}+\overline{E}-E)/20} \tag{3-63}$$

อัตราขยายที่ทำนายได้ g_c' จะหาได้จากการทำนายค่า log-energy ของ fixed-codebook ปัจจุบันเพิ่มเติมจาก log-energy ของ fixed-codebook ก่อนหน้า โดยใช้ตัวทำนายสัญญาณ MA ที่มีอันดับ 4 พลังงานที่ทำนายได้นั้นมีค่า

$$\tilde{E}^{(m)} = \sum_{i=1}^{4} b_i \hat{U}^{(m-i)} \tag{3-64}$$

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

$$U^{(m)} = E^{(m)} - \tilde{E}^{(m)} \tag{3-65}$$

อัตราขยายที่ทำนายได้ g_c' จะหาได้โดยการแทน $E^{(\mathrm{m})}$ ด้วยค่าของตัวเองที่ทำนายได้จากสมการที่ (3-63) ดังนี้

$$g_c' = 10^{(\tilde{E}^{(m)}+\overline{E}-E)/20} \tag{3-66}$$

ค่าตัวประกอบที่กำหนดค่าความถูกต้อง γ จะสัมพันธ์กับค่าผิดพลาดที่ได้จากการทำนายอัตราขยายดังนี้

$$U^{(m)} = E^{(m)} - \tilde{E}^{(m)} = 20\log(\gamma) \tag{3.67}$$

3.1.2.9.2 การหา codebook สำหรับการควอนไทซ์อัตราขยาย

อัตราขยายของ adaptive-codebook g_p และตัวประกอบ คือเวกเตอร์ที่ควอนไทซ์โดยใช้ 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 อัตราขยายของ adaptive-codebook ที่ควอนไทซ์จะมีค่า

$$\hat{g}_p = gA_1(GA) + gB_1(GB) \tag{3-68}$$

และอัตราขยายของ fixed-codebook ที่ควอนไทซ์แล้วมีค่า

$$\hat{g}_c = g_c'\gamma = g_c'(gA_2(GA) + gB_2(GB)) \tag{3-69}$$

โครงสร้างที่เป็นคอนจูเกตนี้ช่วยทำให้การค้นหาเป็นไปโดยง่ายโดยการใช้วิธีการคัดเลือกล่วงหน้า ในส่วน codebook *gA* จะมีทั้งหมด 8 เวกเตอร์ ในขั้นตอนการคัดเลือกล่วงหน้านั้น กลุ่มของเวกเตอร์ที่มีค่าเข้าใกล้ g_c จะถูกเลือกขึ้นมา 4 เวกเตอร์ ในทำนองเดียวกัน codebook *gB* ที่มีทั้งหมด 16 เวกเตอร์ จะได้กลุ่มของเวกเตอร์ 8 เวกเตอร์ที่มีค่าเข้าใกล้ g_p ถูกเลือกขึ้นมา เป็นผลให้ได้เวกเตอร์ที่ดีที่สุดมาครึ่งหนึ่ง จากนั้นจึงทำการหาเวกเตอร์ที่ดีที่สุดจาก 32 แบบที่เป็นไปได้ทั้งหมด โดยมีเงื่อนไขของการเลือก เพื่อทำให้ weighted mean squared error มีค่าต่ำสุดตามสมการที่ (3-58)

3.1.2.10 การปรับให้ทันกาลในหน่วยความจำ

การปรับให้ทันกาลค่าพารามิเตอร์ต่างๆ ของตัววงจรกรองสัญญาณที่ใช้สังเคราะห์เสียงนั้น จะทำการคำนวณหาสัญญาณของเป้า ในเฟรมย่อยถัดไป หลังจากทำการควอนไทซ์อัตราขยายทั้งสองแล้ว จะได้สัญญาณกระตุ้น $u(n)$ ของเฟรมย่อยปัจจุบันมีค่าดังนี้

$$u(n) = g_p v(n) + g_c c(n),\ n = 0, \ldots, 39 \qquad (3\text{-}70)$$

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 ชุดในการทำงานนี้ วิธีการที่ง่ายกว่านั้นซึ่งใช้วงจรกรองสัญญาณเพียงชุดเดียวสามารถทำได้โดยสร้างสัญญาณเสียงขึ้นมาใหม่ $\hat{s}(n)$ 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

$$ew(n) = x(n) - \hat{g}_p y(n) - \hat{g}_c z(n) \qquad (3\text{-}71)$$

เมื่อหาค่าสัญญาณ $x(n)$, $y(n)$ และ $z(n)$ ได้ก็จะทำการปรับให้ทันกาลโดยการคำนวณหา $ew(n)$ ตามสมการที่ (3-71) สำหรับ $n = 30, \ldots, 39$ ซึ่งจะลดการใช้วงจรกรองสัญญาณลง 2 ชุด

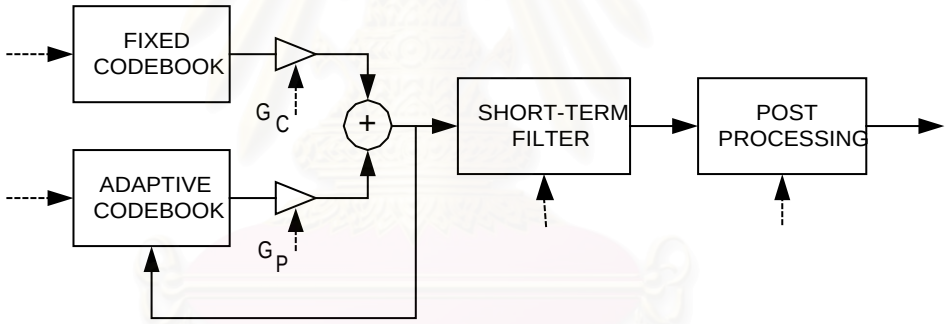
3.2 หลักการทำงานโดยรวมของตัวถอดรหัส MP-CELP

3.2.1 ตัวถอดรหัส MP-CELP

หลักการทำงานของตัวถอดรหัสแสดงตามรูปที่ 3.4 มีองค์ประกอบต่างๆ เหมือนตัวถอดรหัสของ CS-ACELP อันดับแรกพารามิเตอร์ที่รับมาในรูปของกระแสของบิตข้อมูลจะถูกแยกแยะ

ออกตามเฟรมๆ ละ 10 มิลลิวินาที และถอดรหัสออกมาเป็นค่าของพารามิเตอร์ต่างๆ ได้แก่ สัมประสิทธิ์LSP fractional pitch delay 2 ค่า fixed-codebook 2 ชุด และอัตราขยายของ fixed-codebook กับ adaptive-codebook อย่างละ 2 ชุด สัมประสิทธิ์ LSP นั้นจะถูกนำมาทำการประมาณค่าในช่วงแล้วเปลี่ยนเป็นสัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้น ของเฟรมย่อยแต่ละเฟรม โดยเฟรมย่อยแต่ละเฟรมที่มีขนาด 5 มิลลิวินาทีนั้น [4 15 35 และ 39]
จะมีขั้นตอนการทำงานดังนี้

- สัญญาณกระตุ้นจะถูกสร้างขึ้นโดยการรวม adaptive-codebook และ fixed-codebook เข้าด้วยกันตามอัตราขยายของแต่ละตัว
- สัญญาณเสียงจะถูกสังเคราะห์ขึ้น โดยนำสัญญาณกระตุ้นที่ได้มากรองด้วยวงจรกรองตัวทำนายเชิงเส้นสังเคราะห์เสียง
- สัญญาณเสียงที่สังเคราะห์ได้จะนำไปผ่านวงจร post-processing ซึ่งประกอบด้วย adaptive postfilter ที่สร้างจากวงจรกรองสัญญาณแบบ long-term และ short-term, วงจรกรองสัญญาณความถี่สูงผ่าน และวงจรปรับขนาดสัญญาณ



รูปที่ 3.4 หลักการทำงานของตัวถอดรหัสแบบ MP-CELP

3.2.2 รายละเอียดการทำงานของตัวถอดรหัส MP-CELP

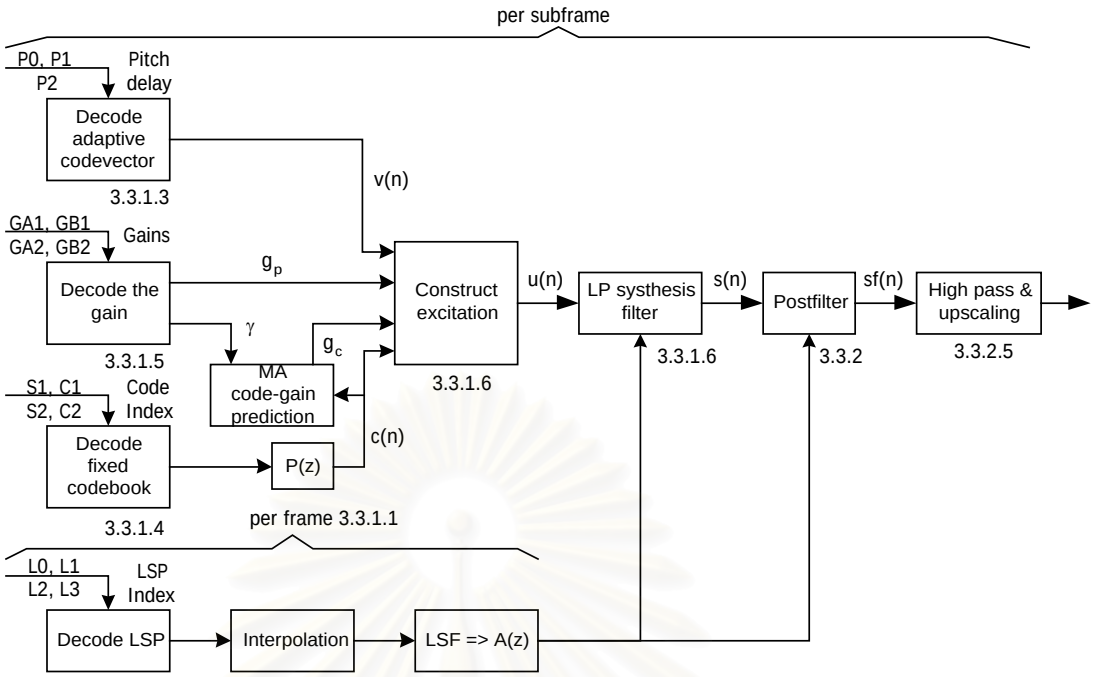
รายละเอียดการทำงานของตัวถอดรหัส MP-CELP อธิบายได้ตามฟังก์ชันต่างๆ ในรูปที่ 3.5 ขั้นแรกนั้นพารามิเตอร์จะถูกถอดรหัสออกมาก่อน (สัมประสิทธิ์ของตัวทำนายเชิงเส้น, เวกเตอร์ adaptive-codebook, เวกเตอร์ fixed-codebook และ อัตราขยาย) พารามิเตอร์ที่ส่งผ่านมานั้นมีรายละเอียดตามตารางที่ 3.4 พารามิเตอร์ที่ใช้ในตัวถอดรหัสนี้จะถูกคำนวณกลับขึ้นมาเพื่อนำมาใช้สร้างสัญญาณเสียงขึ้น สัญญาณเสียงที่สร้างกลับขึ้นมานั้นจะถูกขยายโดยการทำ post-processing ซึ่งประกอบด้วย postfilter, วงจรกรองสัญญาณความถี่สูงผ่าน และตัวปรับขยายสัญญาณ สุดท้าย

จะอธิบายถึงการแก้ไขความผิดพลาดเมื่อพาริตีบิตมีค่าไม่ตรงกับที่ส่งมา หรือแฟลกที่ให้ลบเฟรม (frame erasure flag) ถูกส่งมา รายละเอียดของสัญญาณของตัวถอดรหัสแสดงในรูปที่ 3.5

ตารางที่ 3.4 รายละเอียดของพารามิเตอร์ต่างๆ และลำดับการเรียงบิตข้อมูล
(บิตที่มีนัยสำคัญสูงสุด MSB จะถูกส่งมาก่อน)

สัญลักษณ์	รายละเอียด	จำนวนบิต
*L*0	ตัวทำนายสัญญาณ MA ตัวชี้ควอนไทเซอร์ LSP	1
*L*1	เวกเตอร์ของควอนไทเซอร์ LSP ในสเตจแรก	7
*L*2	เวกเตอร์ส่วนล่างของควอนไทเซอร์ LSP ในสเตจที่สอง	5
*L*3	เวกเตอร์ของควอนไทเซอร์ LSP ในสเตจที่สอง	5
*P*1	การประวิงเวลาของพิตช์ในเฟรมย่อยแรก	6
*P*0	พาริตีบิตของค่าการประวิงเวลาของพิตช์	1
*C*1	fixed-codebook ในเฟรมย่อยแรก	6/15/30
*S*1	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยแรก	1/5/10
*GA*1	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยแรก	3
*GB*1	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยแรก	4
*P*2	การประวิงเวลาของพิตช์ในเฟรมย่อยที่สอง	3
*C*2	fixed-codebook ในเฟรมย่อยที่สอง	6/15/30
*S*2	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยที่สอง	1/5/10
*GA*2	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยที่สอง	3
*GB*2	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยที่สอง	4

กรณีตัวเข้ารหัสหลักที่สัญญาณกระตุ้นในส่วนของ fixed codebook มี 1 พัลส์ 5 พัลส์ และ 10 พัลส์ จะทำงานด้วยอัตราการเข้ารหัส 5,600 8,200 และ 12,200 bps ตามลำดับ



รูปที่ 3.5 สัญญาณต่างๆ ในตัวถอดรหัส MP-CELP

3.2.2.1 ขั้นตอนการถอดรหัสพารามิเตอร์

การถอดรหัสพารามิเตอร์มีขั้นตอนดังนี้

3.2.2.1.1 การถอดรหัสพารามิเตอร์ของวงจรกรองตัวทำนายเชิงเส้น

พารามิเตอร์ *L0 L1 L2* และ *L3* ที่รับได้นั้นถูกใช้ในการสร้างสัมประสิทธิ์ LSP ที่ถูกควอนไทซ์ และทำการประมาณค่าในช่วง 2 ชุด (ของเฟรมย่อยแต่ละเฟรม) ในเฟรมย่อยแต่ละเฟรมนั้น สัมประสิทธิ์ LSP ที่มาจากการประมาณค่าในช่วงจะถูกเปลี่ยนเป็นสัมประสิทธิ์ตัวทำนายเชิงเส้น a_i ซึ่งถูกใช้ในการสังเคราะห์สัญญาณเสียงในเฟรมย่อยแต่ละเฟรม

ขั้นตอนที่เกิดขึ้นในเฟรมย่อยแต่ละเฟรมมีดังนี้

1. ถอดรหัสเวกเตอร์ adaptive-codebook
2. ถอดรหัสเวกเตอร์ fixed-codebook
3. ถอดรหัสอัตราขยายของ adaptive-codebook และ fixed-codebook
4. ทำการคำนวณเพื่อสร้างสัญญาณเสียงกลับขึ้นมา

3.2.2.1.2 การคำนวณหาพารีตีบิต

จะทำการสร้างสัญญาณกระตุ้นขึ้นมาก่อน และคำนวณหาพาริตีบิตจากพารามิเตอร์ $P1$ ของค่าการประวิงเวลาใน adaptive-codebook ซึ่งถ้าพาริตีบิตที่คำนวณขึ้นมาได้ไม่ตรงกับพาริตีบิต $P0$ ที่ส่งมา หมายความว่าเกิดบิตผิดพลาดขึ้นในการส่งข้อมูล

ถ้าตรวจสอบพาริตีแล้วพบว่ามีความผิดพลาดเกิดกับ $P1$ ค่าการประวิงเวลา $T1$ จะถูกปรับให้มีค่าเป็นจำนวนเต็มของ $T2$ ของเฟรมก่อนหน้า

3.2.2.1.3 การถอดรหัสเวกเตอร์ adaptive-codebook

จากการตรวจสอบจาก $P0$ ถ้าไม่มีการผิดพลาดเกิดขึ้นกับ adaptive-codebook $P1$ ที่รับได้ $T1$ จะได้มาจาก $P1$ ส่วน $T2$ จะได้มาจาก $P2$ แต่หาก $P0$ บ่งชี้ความผิดพลาดของ $P1$ จะมีการนำค่าพิตช์ดีเลย์ $P2$ ของเฟรมก่อนหน้า มาใช้แทน $P1$ ของเฟรมปัจจุบัน

เวกเตอร์ของ adaptive-codebook $v(n)$ จะได้มาจากการประมาณค่าในช่วงของสัญญาณกระตุ้น $u(n)$ ในอดีตที่เก็บไว้ในบัฟเฟอร์ โดยย้อนกลับไปด้วยค่าการประวิงเวลา $T1$ หรือ $T2$ โดยใช้สมการที่ (3-39)

3.2.2.1.4 การถอดรหัสเวกเตอร์ fixed-codebook

พารามิเตอร์ C ของ fixed-codebook ที่รับได้จะถูกใช้ในการหาตำแหน่งของพัลส์ของสัญญาณกระตุ้น ส่วนเครื่องหมายของพัลส์จะได้มาจากพารามิเตอร์ S ตำแหน่งและเครื่องหมายของพัลส์จะถูกถอดรหัสมาจากเวกเตอร์ fixed-codebook $c(n)$ โดยใช้สมการที่ (3-44) ถ้าส่วนที่เป็นจำนวนเต็มของค่าการประวิงเวลา T มีค่าน้อยกว่าขนาดเฟรมย่อย 40 เวกเตอร์ $c(n)$ จะถูกปรับเปลี่ยนตามสมการที่ (3-47)

3.2.2.1.5 การถอดรหัสอัตราขยายของ adaptive-codebook และ fixed-codebook

พารามิเตอร์ที่รับได้จะมีอัตราขยายของ adaptive-codebook g_p และตัวประกอบที่ใช้กำหนดความถูกต้องของอัตราขยายของ fixed-codebook γ การประมาณค่าอัตราขยาย fixed-codebook g_c' ทำได้โดยใช้สมการที่ (3-66) เวกเตอร์ fixed-codebook จะได้มาจากผลคูณค่าตัวประกอบที่ใช้ควบคุมความถูกต้องด้วยอัตราขยายที่ทำนายได้ (สมการที่ (3.69)) อัตราขยาย adaptive-codebook จะถูกสร้างกลับขึ้นมาใหม่โดยใช้สมการที่ (3-68)

3.2.2.1.6 การสังเคราะห์สัญญาณเสียง

สัญญาณกระตุ้น $u(n)$ (ดูสมการที่ (3-70)) ซึ่งเป็นขาเข้าที่ถูกป้อนให้กับวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์หรือตัวทำนายเชิงเส้น จะทำให้สามารถสร้างสัญญาณเสียงกลับขึ้นมาในเฟรมย่อยแต่ละเฟรมดังนี้

$$\hat{s}(n) = u(n) - \sum_{i=1}^{10} a_i \hat{s}(n-i), \quad n = 0,...,39 \tag{3-72}$$

เมื่อ a_i คือสัมประสิทธิ์ของวงจรกรองตัวทำนายเชิงเส้น ของเฟรมย่อยปัจจุบัน สัญญาณเสียงที่สร้างกลับขึ้นมา $\hat{s}(n)$ จะถูกป้อนให้กระบวนการในส่วนสุดท้าย จะได้กล่าวถึงในหัวข้อถัดไป

3.2.2.2 Post-processing

ส่วนของ post-processing นั้นประกอบด้วยฟังก์ชัน 3 ฟังก์ชันคือ adaptive postfiltering วงจรกรองสัญญาณความถี่สูงผ่าน และตัวปรับขนาดสัญญาณขึ้น ส่วนของ adaptive post filtering มาจากการต่อคาสเคดวงจรกรองสัญญาณ 3 วงจรคือ long-term postfilter $H_p(z)$ short-term postfilter $H_f(z)$ และ tilt compensation filter $H_t(z)$ และมีส่วนการควบคุมอัตราขยายแบบปรับค่าได้เป็นส่วนสุดท้าย สัมประสิทธิ์ของ postfilter จะถูกปรับให้ทันกาลทุกๆ 5 มิลลิวินาที (เฟรมย่อย) ขั้นตอนในการทำ postfiltering นั้นจะทำโดยนำสัญญาณเสียง $\hat{s}(n)$ มาหาส่วนกลับของการกรองผ่าน $A(z/\gamma_n)$ เพื่อสร้างสัญญาณ residual $r(n)$ สัญญาณนี้จะถูกใช้ในการคำนวณหาการประวิงเวลา T และอัตราขยาย g_l ของ long-term postfilter $H_p(z)$ และสัญญาณ $r(n)$ จะถูกกรองผ่าน long-term postfilter $H_p(z)$ และวงจรกรองสัญญาณสังเคราะห์ $1/[g_l A(z/\gamma_n)]$ สุดท้ายสัญญาณออกของวงจรกรองสัญญาณสังเคราะห์จะถูกป้อนผ่าน tilt compensation filter $H_t(z)$ เพื่อสร้างสัญญาณเสียง $sf(n)$ ส่วนของการควบคุมอัตราขยายแบบปรับค่าได้จะนำมาใช้กับสัญญาณ $sf(n)$ เพื่อให้พลังงานมีค่าเท่ากับ $s(n)$ จะได้สัญญาณ $sf'(n)$ ป้อนผ่านวงจรกรองสัญญาณความถี่สูงผ่านและวงจรปรับขยายสัญญาณและได้สัญญาณเสียงที่ต้องการออกมา

3.2.2.2.1 Long-term postfilter

วงจร long-term postfilter มีฟังก์ชันส่งผ่านเป็นดังนี้

$$H_p(z) = \frac{1}{1+\gamma_p g_l}(1+\gamma_p g_l z^{-T}) \tag{3-73}$$

เมื่อ T คือค่าการประวิงเวลาของพิตช์และ g_l คือสัมประสิทธิ์ของอัตราขยายซึ่งมีค่าไม่เกิน 1 และจะมีค่าเป็นศูนย์ถ้าอัตราขยายของการทำนายสัญญาณของ long-term นี้มีค่าน้อยกว่า 3 dB ตัวประกอบ γ_p จะควบคุมปริมาณการทำ long-term postfiltering และมีค่า $\gamma_p = 0.5$ การประวิงเวลา

และอัตราขยายของ long-term จะถูกคำนวณมาจากสัญญาณ residual $r(n)$ โดยการกรองสัญญาณเสียง $\hat{s}(n)$ ผ่าน $A(z/\gamma_n)$ ซึ่งเป็น short-term postfilter

$$r(n) = s(n) + \sum_{i=1}^{10} \gamma_n^i a_i s(n-i) \tag{3-74}$$

การประวิงเวลาของ long-term คำนวณได้จากสองขั้นตอน ในขั้นแรกจะทำการเลือกจำนวนเต็ม T_0 ที่ดีที่สุดในช่วง [int(T_1) – 1, int(T_1) + 1] เมื่อ int(T_1) คือส่วนของจำนวนเต็มจากค่าการประวิงเวลาพิตช์ T_1 ของเฟรมย่อยแรกที่ส่งมา ซึ่งค่านี้จะทำให้อัตสหสัมพันธ์ที่ได้มีค่าสูงสุดดังนี้

$$R(k) = \sum_{n=0}^{39} r(n) r(n-k) \tag{3-75}$$

ในขั้นที่สองจะเลือกเศษส่วนของค่าการประวิงเวลา T ที่ดีที่สุดที่มีความละเอียด 1/8 รอบๆ ค่า T_0 การเลือกค่านี้ทำได้โดยการทำให้เป็นบรรทัดฐานด้วยอัตสหสัมพันธ์สูงสุดดังนี้

$$R'(k) = \frac{\sum_{n=0}^{39} r(n)\hat{r}_k(n)}{\sqrt{\sum_{n=0}^{39} \hat{r}_k(n)\hat{r}_k(n)}} \tag{3-76}$$

เมื่อ $\hat{r}_k(n)$ คือสัญญาณ residual ที่เวลาประวิง k และสามารถหาค่าประวิงเวลาที่ดีที่สุด T ได้ จากนั้นอัตสหสัมพันธ์ $R'(T)$ จะถูกทำให้เป็นบรรทัดฐานด้วยรากที่สองของพลังงานของ $r(n)$

ค่าของ g_l คำนวณได้จาก

$$g_l = \frac{\sum_{n=0}^{39} r(n)\hat{r}_k(n)}{\sum_{n=0}^{39} \hat{r}_k(n)\hat{r}_k(n)}, \quad 0 \le g_l \le 1.0 \tag{3-77}$$

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

3.2.2.2.2 Short-term postfilter

วงจร short-term postfilter มีฟังก์ชันส่งผ่านเป็นดังนี้

$$H_f(z) = \frac{1}{g_f}\frac{A(z/\gamma_n)}{A(z/\gamma_d)} = \frac{1}{g_f}\frac{1+\sum_{i=1}^{10}\gamma_n^i a_i z^{-i}}{1+\sum_{i=1}^{10}\gamma_d^i a_i z^{-i}} \quad (3\text{-}78)$$

เมื่อ $A(z)$ คือส่วนกลับของวงจรกรองตัวทำนายเชิงเส้น ที่ค่าสัมประสิทธิ์ถูกควอนไตซ์ที่คำนวณได้จากพารามิเตอร์ที่รับได้ (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

$$g_f = \sum_{n=0}^{19} |h_f(n)| \quad (3\text{-}79)$$

3.3.2.3 Tilt compensation

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

$$H_t(z) = \frac{1}{g_t}(1+\gamma_t k_1' z^{-1}) \quad (3\text{-}80)$$

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

$$k_1' = -\frac{r_h(1)}{r_h(0)}; \quad r_h(i) = \sum_{j=0}^{19-i} h_f(j)h_f(j+i) \quad (3\text{-}81)$$

พจน์อัตราขยาย g_t = 1 - $|\gamma_t k_1'|$ จะชดเชยการลดลงเนื่องจากผลกระทบของ g_f ใน $H_f(z)$ ค่า γ_t นี้จะมีค่าขึ้นอยู่กับเครื่องหมายของ k_1' กล่าวคือถ้า k_1' มีค่าเป็นลบ จะได้ γ_t = 0.9 และ ถ้า k_1' มีค่าเป็นบวกจะได้ γ_t = 0.2

3.2.2.2.4 การควบคุมอัตราขยายแบบปรับค่าได้

การควบคุมอัตราขยายแบบปรับค่าได้ (adaptive gain control) ถูกใช้ในการชดเชยความแตกต่างระหว่างขนาดสัญญาณเสียงที่สร้างกลับขึ้นมา $\hat{s}(n)$ กับสัญญาณที่ได้จาก postfilter $sf(n)$ ค่าตัวประกอบ G ที่ใช้ควบคุมปรับอัตราขยายในเฟรมย่อยปัจจุบันนั้นคำนวณได้จาก

$$G = \frac{\sum_{n=0}^{39} |\hat{s}(n)|}{\sum_{n=0}^{39} |sf(n)|} \tag{3-82}$$

สัญญาณที่ผ่านการควบคุม $sf'(n)$ แล้วจะมีค่า

$$sf'(n) = g^{(n)} sf(n), \qquad n = 0,\ldots, 39 \tag{3-83}$$

เมื่อ $g^{(n)}$ จะถูกปรับให้ทันกาลตัวอย่างต่อตัวอย่างดังนี้

$$g^{(n)} = 0.9875\, g^{(n-1)} + 0.125G, \qquad n = 0, \ldots, 39 \tag{3-84}$$

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

3.2.2.2.5 การกรองสัญญาณความถี่สูงผ่านและปรับขยายขนาด

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

$$H_{h2}(z) = \frac{0.93980581 - 1.8795834\, z^{-1} + 0.93980581\, z^{-2}}{1 - 1.9330735\, z^{-1} + 0.93589199\, z^{-2}} \tag{3-85}$$

สัญญาณที่ได้จากวงจรกรองสัญญาณนี้จะมีอัตราขยาย 2 เท่าเพื่อชดเชยที่เกิดจากตัวเข้ารหัส

3.2.2.3 การกำหนดค่าเริ่มต้นให้กับตัวเข้ารหัสและถอดรหัส

พารามิเตอร์ต่างๆ ของตัวเข้ารหัสและถอดรหัสจะมีค่าเริ่มต้นเป็น 0 ยกเว้นพารามิเตอร์ต่างๆ ที่แสดงอยู่ในตารางที่ 3.5

ตารางที่ 3.5 รายละเอียดของพารามิเตอร์ที่มีค่าเริ่มต้นไม่เป็นศูนย์

ตัวแปร	หัวข้อ	ค่าเริ่มต้น
β	3.1.2.8	0.8
$g^{(-1)}$	3.2.2.2.4	1.0
l_i	3.1.2.2.4	$i\pi/11$
q_i	3.1.2.2.4	$\arccos(i\pi/11)$
$\hat{U}^{(k)}$	3.1.2.9.1	-14

3.2.2.4 การแก้ไขข้อผิดพลาด

ขั้นตอนในการแก้ไขข้อผิดพลาดถูกกำหนดให้มีขึ้นในตัวถอดรหัส เพื่อทำให้คุณภาพของเสียงที่สร้างกลับขึ้นมานั้นไม่ลดลงเนื่องจากไม่สามารถสร้างเฟรมข้อมูลที่รับมาได้ ขั้นตอนการแก้ไขข้อผิดพลาดนี้จะทำงานเมื่อพารามิเตอร์ของเฟรมเกิดการสูญหาย

วัตถุประสงค์ในการแก้ไขข้อผิดพลาดนั้น มีไว้เพื่อใช้สร้างเฟรมปัจจุบันจากเฟรมก่อนหน้าที่รับมาได้เนื่องจากมีการลบเฟรมที่มีข้อผิดพลาดเกิดขึ้น ขั้นตอนในการแทนที่สัญญาณกระตุ้นที่ไม่ถูกต้องด้วยสัญญาณกระตุ้นที่มีคุณสมบัติใกล้เคียงกันโดยใช้การทำนายค่าอัตราขยายใน long-term ซึ่งคำนวณมาจากการวิเคราะห์ long-term postfilter โดยจะหาตัวทำนายแบบ long-term เพื่อใช้ทำนายค่าอัตราขยายซึ่งมีค่ามากกว่า 3 dB กระบวนการในการแก้ไขข้อผิดพลาดนั้นเฟรมจะกำหนดให้เป็นรายคาบ ถ้ามีเฟรมย่อยอย่างน้อยหนึ่งเฟรมย่อยมีอัตราขยายที่ได้จากการทำนายแบบ long-term มากกว่า 3 dB นอกนั้นกำหนดให้ไม่เป็นรายคาบ ขั้นตอนในการสร้างเฟรมที่ถูกลบนั้นมีดังนี้

1. หาพารามิเตอร์ของวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์เสียง
2. ลดอัตราขยายของ adaptive-codebook และ fixed-codebook
3. ลดค่าพารามิเตอร์ของตัวทำนายอัตราขยาย
4. สร้างสัญญาณกระตุ้นที่นำมาใช้แทน

3.2.2.4.1 พารามิเตอร์ของวงจรกรองสัญญาณที่ใช้ในการสังเคราะห์เสียง

พารามิเตอร์วงจรกรองสัญญาณที่ใช้ในการสังเคราะห์เสียงของเฟรมที่ถูกลบนั้น จะใช้พารามิเตอร์ตัวทำนายเชิงเส้น ของเฟรมก่อนหน้าที่สมบูรณ์แทน ซึ่งหน่วยความจำของตัวทำนาย MA LSF จะประกอบด้วยค่าพารามิเตอร์ l_i ที่รับมาได้ แต่พารามิเตอร์นี้ไม่สามารถนำมาใช้กับเฟรมปัจจุบัน m ได้ แต่จะต้องทำการคำนวณใหม่จากพารามิเตอร์ LSF ω_i และพารามิเตอร์จากหน่วยความจำของตัวทำนายโดยใช้

$$l_i = [\omega_i^{(m)} - \sum_{k=1}^{4} p_{i,k} l_i^{(m-1)}]/(1 - \sum_{k=1}^{4} p_{i,k}), \quad i = 1,...,10 \tag{3-86}$$

เมื่อสัมประสิทธิ์ของตัวทำนาย MA $p_{i,k}$ มาจากค่าในเฟรมสมบูรณ์ก่อนหน้า

3.2.2.4.2 การปรับลดอัตราขยายของ adaptive-codebook และ fixed-codebook

อัตราขยายของ fixed-codebook จะถูกปรับลดขนาดลงเมื่อเทียบกับค่าในเฟรมก่อนหน้าและมีค่าดังนี้

$$g_c^{(m)} = 0.98 g_c^{(m-1)} \quad (3\text{-}87)$$

เมื่อ m คือตัวบ่งบอกลำดับของเฟรมย่อย ส่วนอัตราขยายของ adaptive-codebook นั้นจะถูกปรับลดลงเมื่อเทียบกับค่าในเฟรมก่อนหน้าและมีค่าดังนี้

$$g_p^{(m)} = 0.9 g_p^{(m-1)} \qquad \text{โดย } g_p^{(m)} < 0.9 \quad (3\text{-}88)$$

3.2.2.4.3 การลดค่าพารามิเตอร์ของตัวทำนายอัตราขยาย

ตัวทำนายอัตราขยายจะใช้พลังงานของเฟรมก่อนหน้าเป็นเงื่อนไขในการเลือกเวกเตอร์ของ fixed-codebook เพื่อหลีกเลี่ยงปัญหาที่เกิดจากข้อผิดพลาดในการส่งข้อมูลเมื่อเฟรมที่รับได้เฟรมหนึ่งครบถ้วนสมบูรณ์ ก็จะทำให้มีการปรับให้ทันกาลค่าในหน่วยความจำของตัวทำนายอัตราขยายด้วยพลังงานของ codebook ที่ถูกปรับลดอัตราขยายลง ค่าของ $U^{(m)}$ สำหรับเฟรมย่อยปัจจุบัน m จะถูกกำหนดให้มีค่าน้อยลง 4 dB ดังนี้

$$U^{(m)} = (0.25\sum_{i=1}^{4} U^{(m-i)}) - 4.0 \qquad \text{โดย } U^{(m)} \geq -14 \quad (3\text{-}89)$$

3.2.2.4.4 การสร้างสัญญาณกระตุ้นที่นำมาใช้แทน

สัญญาณกระตุ้นที่จะสร้างขึ้นแทนนี้จะขึ้นอยู่กับเงื่อนไขของความเป็นรายคาบ ตามที่ได้กล่าวไปแล้ว ถ้าเฟรมที่สร้างกลับขึ้นมาเฟรมสุดท้ายถูกกำหนดให้เป็นรายคาบ ดังนั้นเฟรมปัจจุบันก็ควรจะเป็นรายคาบด้วย ในกรณีนี้จะใช้ adaptive-codebook เท่านั้น ส่วน fixed-codebook จะกำหนดให้มีค่าเป็นศูนย์ ค่าการประวิงเวลาของพิตช์จะได้มาจากส่วนที่เป็นจำนวนเต็มของค่าการประวิงเวลาของพิตช์ของเฟรมก่อนหน้า เพื่อหลีกเลี่ยงปัญหาการเพิ่มขึ้นของค่าการประวิงเวลามากเกินไปสำหรับเฟรมย่อยถัดไปจึงกำหนดให้ค่าการประวิงเวลานี้มีค่าไม่เกิน 143 ส่วนอัตราขยายของ adaptive-codebook นั้นจะใช้ค่าที่ถูกปรับลดอัตราขยายลงและมีค่าเป็นไปตามสมการที่ (3-88)

แต่ถ้าเฟรมสุดท้ายถูกกำหนดให้ไม่เป็นรายคาบ ดังนั้นในเฟรมถัดมาก็ควรไม่เป็นรายคาบด้วย โดยจะกำหนดให้เวกเตอร์ของ adaptive-codebook มีค่าเป็นศูนย์ ส่วน fixed-codebook จะได้มาจากการสุ่มเลือกจากตัวชี้ codebook และ sign ตัวกำเนิดค่าสุ่มนั้นจะมีฟังก์ชันดังนี้

$$seed = seed*31821 + 13849 \quad (3\text{-}90)$$

โดยค่า *seed* เริ่มต้นมีค่าเป็น 21845 ตัวชี้ codebook จะมาจากเลข 13 ตำแหน่งสุดท้ายที่มีนัยสำคัญต่ำสุดของค่าที่สุ่มมาได้ ส่วนเครื่องหมายของ fixed-codebook นั้นจะมาจากเลข 4 ตำแหน่งสุดท้ายที่มีนัยสำคัญต่ำสุดของค่าที่สุ่มได้ในครั้งต่อไป โดยอัตราขยายของ fixed-codebook นี้จะมีค่าเป็นไปตามสมการที่ (3-87)

3.3 การปรับระดับอัตราการเข้ารหัส

ตัวเข้ารหัสหลัก MP-CELP ที่นำเสนอในหัวข้อ 3.1 มี 3 ตัวด้วยกัน คือ ตัวเข้ารหัสหลักที่สัญญาณกระตุ้นในส่วนของ fixed codebook มี 1 พัลส์ 5 พัลส์ และ 10 พัลส์ ในการปรับระดับอัตราการเข้ารหัส จะกระทำโดยต่อส่วนขยาย (enhancement layer) เข้ากับตัวเข้ารหัสหลัก เพื่อเพิ่มจำนวนพัลส์ในส่วนของ fixed codebook ครั้งละ 1 พัลส์ [4 30 และ 31] บิตข้อมูลในแต่ละเฟรมย่อยที่ต้องส่งเพิ่มคือ ตำแหน่งของพัลส์ที่เพิ่มอีก 3 บิต และเครื่องหมายของพัลส์ที่เพิ่มอีก 1 บิต เพราะฉะนั้นในแต่ละเฟรมจะต้องส่งบิตข้อมูลเพิ่มขึ้น 8 บิต ทำให้การปรับระดับอัตราการเข้ารหัสแต่ละครั้งจะเพิ่มอัตราบิตขึ้น 800 bps

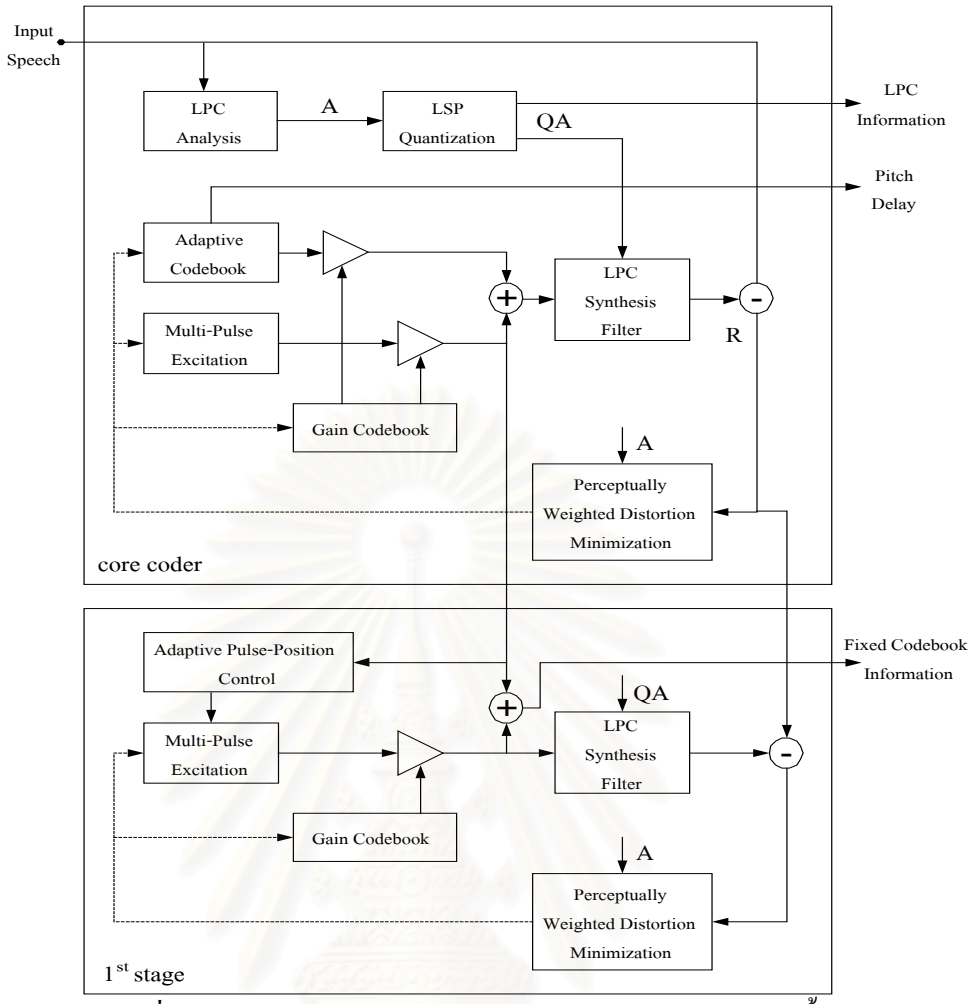
ตามข้อกำหนดของมาตรฐาน MPEG-4 กำหนดให้มีการปรับระดับอัตราการเข้ารหัสได้มากที่สุด 3 ระดับ [3] วิทยานิพนธ์นี้ จึงนำเสนอการปรับระดับอัตราการเข้ารหัส 3 ขั้น และนำเสนอรายละเอียดของแต่ละขั้นดังต่อไปนี้

3.3.1 การปรับระดับอัตราการเข้ารหัส 1 ขั้น (1 enhancement layer)

การทำงานของตัวเข้ารหัส MP-CELP ที่มีการปรับระดับอัตราการเข้ารหัส 1 ขั้น แสดงด้วยบล็อกไดอะแกรมในรูปที่ 3.6

ส่วนขยายปรับระดับอัตราเข้ารหัสจะทำการเข้ารหัสสัญญาณเศษเหลือ (residual signal) ที่สร้างขึ้นจากในส่วนของตัวเข้ารหัสหลัก โดยใช้การควอนไทซ์สัญญาณมัลติพัลส์ เช่นเดียวกับในส่วนของตัวเข้ารหัสหลัก จุดแตกต่างที่สำคัญคือมีการเพิ่มตัวควบคุมตำแหน่งพัลส์แบบปรับแต่งได้ (adaptive pulse position control) เพื่อทำหน้าที่ควบคุมตำแหน่งของพัลส์ที่เกิดขึ้นใหม่ในส่วนขยายปรับระดับอัตราเข้ารหัสไม่ให้ไปซ้ำกับตำแหน่งพัลส์เดิมในตัวเข้ารหัสหลัก กระบวนการนี้จะทำโดยมีการปรับโครงสร้าง codebook แบบพีชคณิตก่อนล่วงหน้า เพื่อกำหนดตำแหน่งพัลส์ที่เป็นไปได้ทั้งหมดของพัลส์ที่จะเกิดขึ้นใหม่ นั่นคือตัวควบคุมตำแหน่งพัลส์แบบปรับแต่งได้ จะรับข้อมูลตำแหน่งพัลส์เดิมจากตัวเข้ารหัส และจะควบคุมการควอนไทซ์สัญญาณมัลติพัลส์ดังไดอะแกรมในรูปที่ 3.6

การค้นหาตำแหน่งของพัลส์ใหม่ที่จะเกิดขึ้นในส่วนของการควอนไทซ์สัญญาณมัลติพัลส์จะใช้หลักการเดียวกับตัวเข้ารหัสหลัก คือหาตำแหน่งพัลส์ที่ทำให้ mean squared error ระหว่างสัญญาณที่ถูกถ่วงน้ำหนักกับสัญญาณเศษเหลือจากตัวเข้ารหัสหลัก มีค่าต่ำที่สุด ตามหัวข้อ 3.1.2.8.1

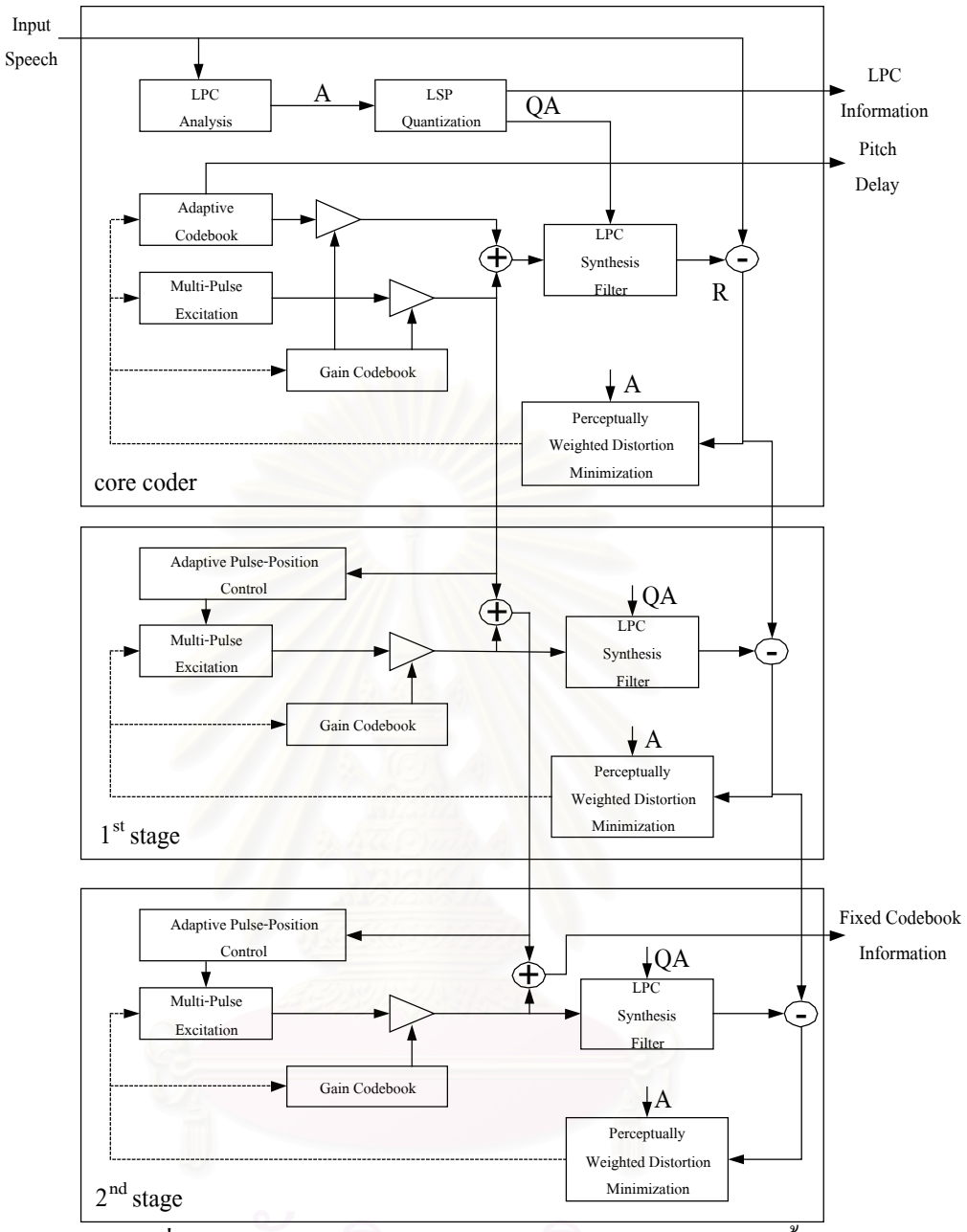


รูปที่ 3.6 บล็อกไดอะแกรมการปรับระดับอัตราการเข้ารหัส 1 ขั้น

3.3.2 การปรับระดับอัตราการเข้ารหัส 2 ขั้น (2 enhancement layers)

การทำงานของตัวเข้ารหัส MP-CELP ที่มีการปรับระดับอัตราการเข้ารหัส 2 ขั้น แสดงด้วยบล็อกไดอะแกรมในรูปที่ 3.7

เป็นการเข้ารหัสสัญญาณเศษเหลือของส่วนขยายปรับระดับอัตราการเข้ารหัสขั้นที่ 1 โดยส่วนขยายปรับระดับอัตราการเข้ารหัสขั้นที่ 2 จะมีองค์ประกอบและหลักการทำงานเหมือนกับส่วนขยายในขั้นที่ 1 ทุกประการ

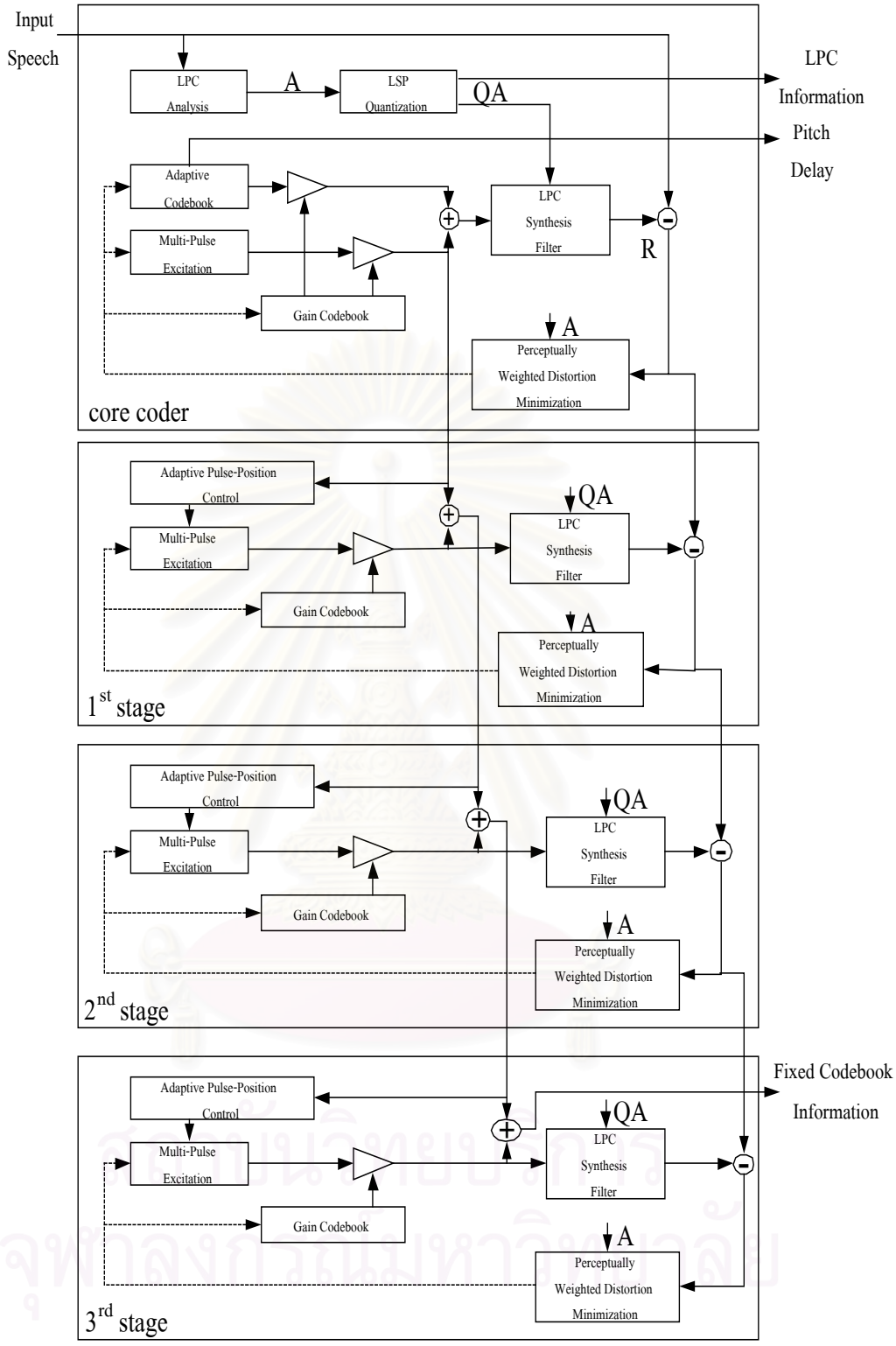


รูปที่ 3.7 บล็อกไดอะแกรมการปรับระดับอัตราการเข้ารหัส 2 ขั้น

3.3.3 การปรับระดับอัตราการเข้ารหัส 3 ขั้น (3 enhancement layers)

การทำงานของตัวเข้ารหัส MP-CELP ที่มีการปรับระดับอัตราการเข้ารหัส 3 ขั้น แสดงด้วยบล็อกไดอะแกรมในรูปที่ 3.8

เป็นการเข้ารหัสสัญญาณเศษเหลือของส่วนขยายปรับระดับอัตราการเข้ารหัสขั้นที่ 2 โดยส่วนขยายปรับระดับอัตราการเข้ารหัสขั้นที่ 3 จะมีองค์ประกอบและหลักการทำงานเหมือนกับส่วนขยายในขั้นที่ 1 ทุกประการ



รูปที่ 3.8 บล็อกไดอะแกรมการปรับระดับอัตราการเข้ารหัส 3 ขั้น

3.3.4 การจัดสรรบิตสำหรับการปรับระดับอัตราการเข้ารหัส

จำนวนบิตที่ถูกจัดสรรให้กับพารามิเตอร์แต่ละตัว แสดงดังในตารางที่ 3.6 พารามิเตอร์ *C*1 *S*1 *C*2 และ *S*2 จะถูกกำหนดจำนวนบิตให้ขึ้นกับจำนวนขั้นของการปรับระดับอัตราการเข้ารหัส กรณีตัวเข้ารหัสหลักที่ไม่มีการปรับระดับอัตราการเข้ารหัส จะทำงานด้วยอัตราการเข้ารหัส 5,600 8,200 และ 12,200 bps สำหรับสัญญาณกระตุ้นในส่วนของ fixed codebook มี 1 พัลส์ 5 พัลส์ และ 10 พัลส์ ตามลำดับ

ตารางที่ 3.6 รายละเอียดของพารามิเตอร์ต่างๆ และลำดับการเรียงบิตข้อมูล
(บิตที่มีนัยสำคัญสูงสุด MSB จะถูกส่งมาก่อน)

สัญญลักษณ์	รายละเอียด	จำนวนบิต
*L*0	ตัวทำนายสัญญาณ MA ตัวชี้ควอนไทเซอร์ LSP	1
*L*1	เวกเตอร์ของควอนไทเซอร์ LSP ในสเตจแรก	7
*L*2	เวกเตอร์ส่วนล่างของควอนไทเซอร์ LSP ในสเตจที่สอง	5
*L*3	เวกเตอร์ของควอนไทเซอร์ LSP ในสเตจที่สอง	5
*P*1	การประวิงเวลาของพิตซ์ในเฟรมย่อยแรก	6
*P*0	พาริตีบิตของค่าการประวิงเวลาของพิตซ์	1
*C*1	fixed-codebook ในเฟรมย่อยแรก	6+3e/15+3e/30+3e
*S*1	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยแรก	1+e/5+e/10+e
*GA*1	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยแรก	3
*GB*1	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยแรก	4
*P*2	การประวิงเวลาของพิตซ์ในเฟรมย่อยที่สอง	3
*C*2	fixed-codebook ในเฟรมย่อยที่สอง	6+3e/15+3e/30+3e
*S*2	เครื่องหมายของพัลส์ของ fixed-codebook เฟรมย่อยที่สอง	1+e/5+e/10+e
*GA*2	อัตราขยายของ codebook (สเตจที่ 1) เฟรมย่อยที่สอง	3
*GB*2	อัตราขยายของ codebook (สเตจที่ 2) เฟรมย่อยที่สอง	4

หมายเหตุ e คือจำนวนขั้นของการปรับระดับอัตราการเข้ารหัส

กรณีตัวเข้ารหัสหลักที่สัญญาณกระตุ้นในส่วนของ fixed codebook มี 1 พัลส์ มีการปรับระดับอัตราการเข้ารหัส จะทำงานด้วยอัตราการเข้ารหัส 6,400 7,200 และ 8,000 bps สำหรับการปรับระดับอัตราการเข้ารหัส 1 2 และ 3 ขั้น ตามลำดับ

กรณีตัวเข้ารหัสหลักที่สัญญาณกระตุ้นในส่วนของ fixed codebook มี 5 พัลส์ มีการปรับระดับอัตราการเข้ารหัส จะทำงานด้วยอัตราการเข้ารหัส 9,000 9,800 และ 10,600 bps สำหรับการปรับระดับอัตราการเข้ารหัส 1 2 และ 3 ขั้น ตามลำดับ

และกรณีตัวเข้ารหัสหลักที่สัญญาณกระตุ้นในส่วนของ fixed codebook มี 10 พัลส์ มีการปรับระดับอัตราการเข้ารหัส จะทำงานด้วยอัตราการเข้ารหัส 13,000 13,800 และ 14,600 bps สำหรับการปรับระดับอัตราการเข้ารหัส 1 2 และ 3 ขั้น ตามลำดับ

3.4 ความแตกต่างระหว่างคุณลักษณะของเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ

3.4.1 คุณลักษณะของเสียงพูดภาษาอังกฤษ

เสียงพูดในภาษาอังกฤษ ประกอบไปด้วยประโยคลักษณะต่างๆ [45] และในแต่ละประโยคจะประกอบด้วยส่วนย่อยเป็นคำ และในทางเวลา ส่วนย่อยของคำจะมีส่วนย่อยเป็นพยางค์ (syllable) แสดงเป็นสัญลักษณ์ได้ดังนี้ $C_i(C_i)(C_i)\ V(V)\ C_f(C_f)(C_f)$

$C_i(C_i)(C_i)$ คือสัญลักษณ์แสดงพยัญชนะต้น และ $C_f(C_f)(C_f)$ คือสัญลักษณ์แสดงพยัญชนะสะกด อาจประกอบไปด้วยพยัญชนะเดี่ยว พยัญชนะควบ 2 ตัว หรือ 3 ตัว ก็ได้

หน่วยพยัญชนะสำหรับพยัญชนะต้นและพยัญชนะสะกดมีทั้งสิ้น 24 ตัวด้วยกัน แสดงไว้ในตารางที่ 3.7 โดยแบ่งตามแหล่งการกำเนิดเสียง

ตารางที่ 3.7 หน่วยพยัญชนะของภาษาอังกฤษ

ตำแหน่งการกำเนิดเสียง	bilabial	Labio-dental	dental	alveolar	Post-alveolar	palatal	velar	glottal
plosive	p b p b			t d t d			k g k g	
Affricate					ʧ ʤ ch j			
fricative		f v f v	ɵ ð think this	s z s z	ʃ ʒ sh measure			h h
nasal	m m			n n			ŋ ng	
lateral				l l				
approximant	w w				r r	j y		

V(V) คือสัญลักษณ์แสดงสระ มีทั้งที่เป็นสระเดี่ยวเสียงสั้น สระเดี่ยวเสียงยาว และสระผสม แสดงอยู่ในตารางที่ 3.8

ตารางที่ 3.8 หน่วยสระของภาษาอังกฤษ

สระเสียงสั้น	ɪ pit	e pet	æ pat	ʌ putt	ɒ pot	ʊ put	ə another	
สระเสียงยาว	i: bean	ɑ: barn	ɔ: born	u: boon	ɜ: burn			
สระผสม	eɪ bay	ɑɪ buy	ɔɪ boy	əʊ no	ɑʊ now	iə peer	eə pair	ʊə poor

3.5.2 คุณลักษณะของเสียงพูดภาษาไทย

เสียงพูดในภาษาไทยจะมีหน่วยย่อยพยางค์โดยสามารถแสดงเป็นลักษณ์ $C_i(C_i)\ V(V)^{T}\ C_f$ $C_i(C_i)$ คือสัญลักษณ์แสดงพยัญชนะต้น C_f คือสัญลักษณ์แสดงพยัญชนะสะกด พยัญชนะต้นอาจประกอบด้วยพยัญชนะเดี่ยวหรือพยัญชนะควบ ส่วนตัวสะกดจะเป็นได้เพียงพยัญชนะเดี่ยว หน่วยพยัญชนะสำหรับพยัญชนะต้นมีทั้งสิ้น 21 ตัวด้วยกัน [46 และ 47] ดังแสดงในตารางที่ 3.9

ตารางที่ 3.9 หน่วยพยัญชนะต้นของภาษาไทย

ตำแหน่งการกำเนิดเสียง		Labial ริมฝีปาก	Alveolar ปุ่มเหงือก	Palatal ฐานเพดาน	Velar เพดานอ่อน	Glottal เส้นเสียง
เสียงกัก	เสียงไม่ก้อง ไม่พ่นลม	p ป	t ต,ฏ	c จ	k ก	? อ
	เสียงไม่ก้อง พ่นลม	ph พ,ภ,ผ	th ท,ธ,ฒ,ฑ,ถ,ฐ	ch ช,ฌ,ฉ	kh ค,ฆ,ข	
	เสียงก้อง	b บ	d ด,ฎ,ฑ			
เสียงไม่กัก	นาสิก	m ม	n น,ณ		ng ง	
	เสียดแทรก	f ฟ,ฝ	s ซ,ศ,ษ,ส			h ฮ,ห
	ลิ้นรัว		r ร,ฤ			
	ข้างลิ้น		l ล,ฬ			
	เสียงเปิด	w ว		j ย,ญ		

ส่วนพยัญชนะสะกดจะมีเพียง 8 ตัว ที่สามารถเกิดขึ้นได้ดังตารางที่ 3.10

ตารางที่ 3.10 หน่วยพยัญชนะสะกดของภาษาไทย

ตำแหน่งการกำเนิดเสียง		Labial ริมฝีปาก	Alveolar ปุ่มเหงือก	Palatal ฐานเพดาน	Velar เพดานอ่อน
เสียงกัก		p บ,ป,พ,ภ ฟ	t ด,ฎ,ต,ฏ ท,ธ,ฒ,ฑ,ถ,ฐ จ,ช ซ,ศ,ษ,ส		k ก,ต,ฏ,ข
เสียงไม่กัก	นาสิก	m ม	n น,ณ,ร,ล,ฬ,ญ		ng ง
	ข้างลิ้น	w ว		j ย	

V(V) คือสัญลักษณ์แสดงสระ มีจำนวนทั้งสิ้น 18 ตัว ประกอบด้วยสระเสียงสั้น 9 ตัว และสระเสียงยาว 9 ตัว คู่เสียงสระสั้นและยาว ถูกแบ่งแยกตามตำแหน่งของลิ้น แสดงอยู่ในตารางที่ 3.11

ตารางที่ 3.11 หน่วยสระของภาษาไทย

ตำแหน่งของลิ้น	หน้า	กลาง	หลัง
สูง	i , ii ิ , ี	v, vv ึ , ื	u, uu ุ , ู
กลาง	e, ee เ_ะ,เ_	g, gg เ_อะ&เ◌็,เ_อ&เ◌็_	o, oo โ_ะ&CC,โ_
ต่ำ	x, xx แ_ะ,แ_	a, aa _ะ&◌ั,_า	@, @@ เ_าะ,_อ
ผสม	ia, iia เ◌ียะ,เ◌ีย	va, vva เ◌ือะ,เ◌ือ	ua, uua วะ,◌ัว

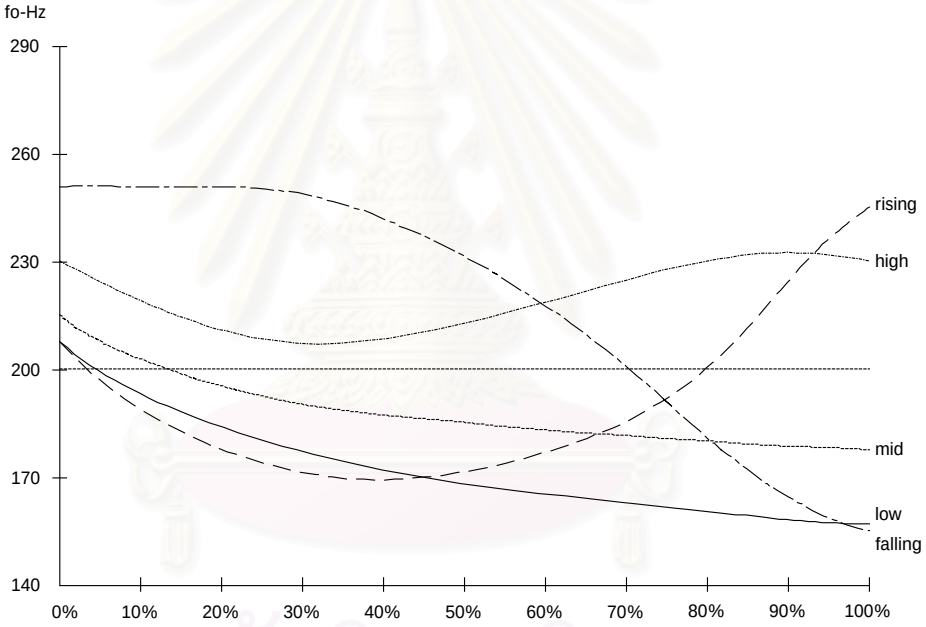
T คือสัญลักษณ์แสดงวรรณยุกต์ (เสียงดนตรี) ของคำหรือพยางค์นั้นๆ มี 4 รูป แต่มี 5 เสียง โดยตำแหน่งในการเขียนจะอยู่ที่พยัญชนะต้นเดี่ยว หรือพยัญชนะต้นตัวที่ 2 สำหรับพยัญชนะควบ แต่ในการวิเคราะห์ค่าคุณลักษณะของเสียง ความแตกต่างที่จะบ่งบอกถึงวรรณยุกต์จะเกิดอยู่ที่เสียงสระ สัญลักษณ์ T จึงปรากฏอยู่ที่เหนือสัญลักษณ์ V วรรณยุกต์ที่เติมในแต่ละคำ จะมีผลให้ความ

หมายของคำเปลี่ยนแปลงไป โดยอาจไม่เกี่ยวข้องกันเลย ดังรูป เสียงวรรณยุกต์ และตัวอย่าง แสดงอยู่ในตารางที่ 3.12

ตารางที่ 3.12 หน่วยวรรณยุกต์ของภาษาไทย

วรรณยุกต์	สามัญ	เอก	โท	ตรี	จัตวา
รูป	-(ไม่มีรูป)				
ตัวอย่าง	ตอง	ต่อง	ต้อง	ต๊อง	ต๋อง

ค่าความถี่มูลฐาน (fundamental frequency) เป็นค่าคุณลักษณะสำคัญ ที่บ่งบอกความเป็นรายคาบของเสียง (voice periodicity) มีการเปลี่ยนแปลงที่เป็นลักษณะเฉพาะของวรรณยุกต์แต่ละเสียง [46 47 และ 48]



รูปที่ 3.9 ลักษณะสมบัติของวรรณยุกต์ต่างๆ ของเสียงพูดภาษาไทย

จากรูปที่ 3.9 แสดงถึงการเปลี่ยนแปลงความถี่มูลฐานในแต่ละวรรณยุกต์ มีความแตกต่างกันและมีลักษณะเฉพาะตัว สามารถแบ่งเป็นกลุ่มได้ 2 กลุ่ม คือ กลุ่มสถิต ประกอบด้วย 3 วรรณยุกต์ได้แก่ เสียงสูง (high) เสียงกลาง (mid) เสียงต่ำ (low) และกลุ่มพลวัต ประกอบด้วย 2 วรรณยุกต์ด้วยกัน คือ เสียงสูงขึ้น (rising) เสียงตก (falling)

สรุปได้ว่า เสียงวรรณยุกต์ทั้งห้า สามารถแยกแยะออกจากกันโดยอาศัย 2 คุณลักษณะหลัก คือ ความสูงของพิตช์ และทิศทางของพิตช์ คุณลักษณะทั้งสองนี้บ่งบอกถึงการเปลี่ยนแปลงของพิตช์ผ่านช่วงเวลาในแต่ละคำหรือพยางค์ในภาษาไทย

3.5 การปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP กับเสียงพูดภาษาไทย

จุดแตกต่างที่สำคัญระหว่างเสียงพูดภาษาดนตรี (tonal language) และเสียงพูดที่ไม่ใช่ภาษาดนตรี (toneless language) คือวรรณยุกต์ (tone) เช่นเสียงพูดภาษาไทยมี 5 เสียงวรรณยุกต์ แต่ละเสียงวรรณยุกต์จะมีลักษณะสมบัติของค่าความถี่มูลฐานเฉพาะตัว ในอีกแง่หนึ่ง ค่าความถี่มูลฐานก็คือส่วนกลับของค่าพิตช์ดีเลย์ที่เป็นตัวชี้บ่งความเป็นรายคาบของเสียง และเป็นพารามิเตอร์ตัวหนึ่งของตัวเข้ารหัสที่อยู่บนพื้นฐานของ CELP เช่น CS-ACELP และ MP-CELP เป็นต้น

จากการศึกษาเบื้องต้นเกี่ยวกับ การเข้ารหัสเสียงพูดภาษาไทยโดยใช้วิธี CS-ACELP ตามมาตรฐาน ITU G.729 [15 และ 28] พบว่าคุณภาพการเข้ารหัสของเสียงพูดภาษาไทย จะด้อยกว่าของเสียงพูดภาษาอังกฤษ ทั้งนี้เนื่องมาจากความแม่นยำในการวิเคราะห์ค่าพิตช์ดีเลย์ที่แสดงถึงข้อมูลวรรณยุกต์ของภาษาไทยยังไม่เพียงพอ และนอกจากนี้ยังพบว่าคุณภาพการเข้ารหัสของเสียงพูดของเพศหญิงต่ำกว่าของเพศชาย โดยร่วมกับการศึกษาเบื้องต้นเกี่ยวกับการแยกแยะพยัญชนะไทยโดยใช้เสียงวรรณยุกต์ [46] พบว่า เสียงพูดของเพศหญิงจะมีอัตราการเปลี่ยนแปลงของลักษณะสมบัติของค่าความถี่มูลฐานสูงกว่าของเพศชาย แสดงให้เห็นว่า การตรวจจับการเปลี่ยนแปลงของพิตช์ของตัวเข้ารหัสโดยวิธี CS-ACELP นี้ ยังไม่เพียงพอสำหรับเสียงพูดของเพศหญิง เพื่อให้เทียบเท่ากับของเพศชาย

เพราะคำนึงถึงการเปลี่ยนแปลงของพิตช์ที่จะบ่งบอกถึงเสียงวรรณยุกต์ต่างๆ ฉะนั้นสำหรับการเข้ารหัสที่อยู่บนพื้นฐานของ CELP ที่มีพารามิเตอร์ค่าประวิงเวลาของพิตช์ร่วมด้วย การให้ความสำคัญกับพิตช์โดยส่งค่าพิตช์ดีเลย์ ด้วยความละเอียดสูงๆ จะทำให้การเข้ารหัสสามารถตรวจจับการเปลี่ยนแปลงของพิตช์ได้ดี และแม่นยำมากยิ่งขึ้น โดยเฉพาะกับเสียงพูดภาษาดนตรี เช่น เสียงพูดภาษาไทย เป็นต้น

วิทยานิพนธ์นี้จึงนำเสนอการปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ด้วยเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง (High Pitch Delay Resolutions: HPDR) จากเดิมที่ค่าพิตช์ดีเลย์ถูกวิเคราะห์และส่งเป็นจำนวนเต็ม (จำนวนตัวอย่าง 1 ตัวอย่าง เท่ากับ 1/8000 วินาที) เปลี่ยนเป็นการวิเคราะห์และส่งเศษส่วนที่ความละเอียด 1/2 1/3 และ 1/4

หลักการของเทคนิคนี้คือ จะมีการเพิ่มส่วนการวิเคราะห์เศษส่วนของพิตช์เข้าไปในการวิเคราะห์พิตช์เดิม

จากหัวข้อ 3.1.2.7 ในเฟรมย่อยแรกจะมีการหาเศษส่วนที่เหมาะสมรอบๆ ค่าจำนวนเต็มพิตช์ T_{op} สำหรับ HPDR ที่เศษส่วนพิตช์ 1/2 จะหาเศษส่วนที่เหมาะสมจาก 0 และ 1/2, สำหรับ HPDR ที่เศษส่วนพิตช์ 1/3 จะหาเศษส่วนที่เหมาะสมจาก –1/3, 0 และ 1/3 และสำหรับ HPDR ที่เศษส่วนพิตช์ 1/4 จะหาเศษส่วนที่เหมาะสมจาก –1/4, 0, 1/4 และ 2/4

การส่งค่าพิตช์ดีเลย์ที่ความละเอียด 1/2 จะมีผลให้ต้องส่งบิตข้อมูลสำหรับเศษส่วนนี้อีก 1 บิตต่อ 1 เฟรมย่อย หรือคือ 2 บิตต่อ 1 เฟรม ทำให้อัตราการเข้ารหัสเพิ่มขึ้น 200 bps ส่วนการส่งค่า

พิตช์ดีเลย์ที่ความละเอียด 1/3 และ 1/4 จะมีผลให้ต้องส่งบิตข้อมูลสำหรับเศษส่วนนี้อีก 2 บิตต่อ 1 เฟรมย่อย หรือคือ 4 บิตต่อ 1 เฟรม ทำให้อัตราการเข้ารหัสเพิ่มขึ้น 400 pbs โดยประยุกต์ใช้กับการเข้ารหัส ที่อัตราการเข้ารหัสหลัก และการปรับระดับทุกๆ อัตรา

3.5.1 การเข้ารหัสและการถอดรหัสค่าการประวิงเวลาของ adaptive-codebook เมื่อใช้เทคนิค HPDR ที่เศษส่วนพิตช์ 1/2

3.5.1.1 การวิเคราะห์หาเศษส่วนพิตช์

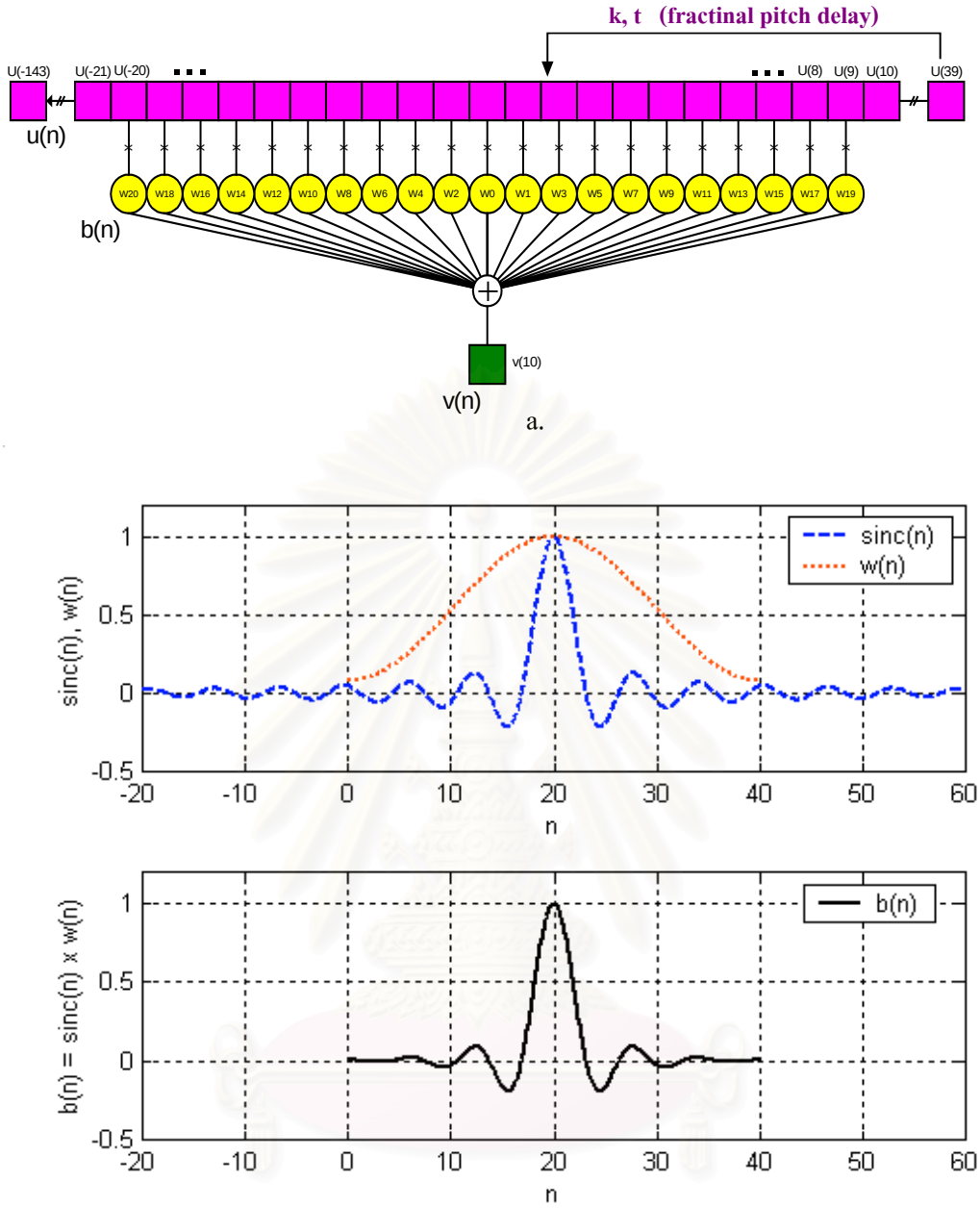
การวิเคราะห์พิตช์ในระดับจำนวนเต็ม ยังคงวิเคราะห์เหมือนเดิมตามหัวข้อ 3.1.2.7 ส่วนการวิเคราะห์หาเศษส่วนพิตช์ จะกระทำโดยการพิจารณาค่าสหสัมพันธ์ข้ามระหว่าง สัญญาณของเป้า x(n) และสัญญาณกระตุ้นในอดีตที่เก็บไว้ในบัฟเฟอร์ ค่าเศษส่วนพิตช์ t ที่ดีที่สุดที่ทำให้ค่าสหสัมพันธ์ข้ามนี้มีค่าสูงสุด จะถูกเลือกเป็นตัวแทนของเฟรมย่อยนั้นๆ ค่าสหสัมพันธ์ข้ามหาได้จากสมการต่อไปนี้ [15 42 และ 43]

$$R(k)_t = \sum_{i=0}^{2} R(k-i)b(t+i.2) + \sum_{i=0}^{2} R(k+1+i)b(2-t+i.2), \quad t=0,1 \qquad (3\text{-}91)$$

นิยาม R(k) จากสมการ (3-37) ค่าถ่วงน้ำหนัก b(n) คือ สัญญาณแบบ FIR b ที่มีโครงสร้างเป็นช่วงสัญญาณของฟังก์ชัน $\sin(x)/x$ ที่ผ่านการจำกัดช่วงด้วย Hammimg Window ตามสมการที่ (3-92) เพื่อมิให้การประมาณค่าในช่วงมีความซับซ้อนมากเกินไป ดังในรูปที่ 3.10b [43] ความกว้างของ Hammimg Window เป็นพารามิเตอร์หนึ่งที่จะปรับให้เหมาะสมกับแต่ละสิ่งแวดล้อม ค่า t เป็น 0 และ 1 สอดคล้องกับเศษส่วน 0 และ 1/2 ตามลำดับ

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n/m), & 0 \le n < m+1 \\ 0, & \text{otherwise} \end{cases} \qquad (3\text{-}92)$$

กระบวนการหาค่าสูงสุดของค่าสหสัมพันธ์ข้างต้น คือการหาค่าพิตช์ที่เหมาะสม ที่จะย้อนกลับไปในอดีตเพื่อนำสัญญาณกระตุ้น u(n) ในบัฟเฟอร์ มาเป็นตัวแทนสัญญาณกระตุ้นในส่วนปรับตัว (adaptive-codebook) ในเฟรมย่อยปัจจุบัน สามารถอธิบายจากรูปที่ 3.10a คือ สัญญาณกระตุ้นในส่วนปรับตัวของเฟรมย่อยปัจจุบัน v(n) ได้จากการประมาณค่าในช่วง คือเป็นผลรวมของผลคูณระหว่าง u(n) กับค่าถ่วงน้ำหนัก b(n)



รูปที่ 3.10 a. ตัวอย่างการสร้างสัญญาณกระตุ้นแบบปรับตัว (adaptive-codebook) เมื่อใช้ค่าถ่วงน้ำหนักจำนวน 21 ตัวอย่าง

b. ค่าถ่วงน้ำหนัก b(n) - กราฟรูปล่าง บนพื้นฐานของ sinc(n) ที่จำกัดช่วงด้วย Hamming window w(n) - กราฟรูปบน

3.5.1.2 การเข้ารหัส

การประวิงเวลาของพิตช์ T_1 และเศษส่วนที่ระดับความละเอียด 1/2 ถูกกำหนดด้วยค่าของ $frac = 0$ และ 1 เพราะฉะนั้นตัวชี้พิตช์ $P1$ จะถูกเข้ารหัสดังนี้

$$P1 = 2(\text{int}(T_1) - 19) + \text{frac}, \quad T_1 = [19,...,143],\ \text{frac} = [0,1] \tag{3-93}$$

ทำนองเดียวกัน P2 ก็ถูกเข้ารหัสด้วยหลักการเดียวกับ P1

การสร้างตัวเข้ารหัสถูกออกแบบให้มีความทนทานต่อการผิดพลาดของบิตข้อมูลนั้นโดยใช้พาริตีบิต $P0$ ซึ่งคำนวณมาจากตัวชี้ $P1$ ของเฟรมย่อยแรก พาริตีบิตที่ได้นั้นมาจากการทำ XOR จากบิตที่มีนัยสำคัญสูงสุด 6 บิตของ $P1$ ที่ตัวถอดรหัสพาริตีบิตนี้จะถูกคำนวณเพื่อใช้ในการตรวจสอบการผิดพลาดของบิตข้อมูล ซึ่งถ้าได้ค่าไม่ตรงกันก็จะใช้วิธีการแก้ไขข้อผิดพลาดเข้ามาช่วย

3.5.1.3 การถอดรหัส

ถ้าไม่มีข้อผิดพลาดเกิดขึ้นแก่ adaptive-codebook $P1$ ที่รับได้ โดยตรวจสอบจาก $P0$ $T1$ และ frac จะได้มาจาก $P1$ ตามสมการที่ (3-93) ส่วน $T2$ และ frac ของเฟรมย่อยที่สอง ก็ได้มาจาก $P2$ ด้วยหลักการเดียวกัน แต่หาก $P0$ บ่งชี้ความผิดพลาดของ $P1$ จะมีการนำค่าพิตช์ดีเลย์ $P2$ ของเฟรมก่อนหน้า มาใช้แทน $P1$ ของเฟรมปัจจุบัน

$$\text{int}(T_1) = P1/2 + 19 \tag{3-94}$$

$$frac = P1 - 2 \text{ x } \text{int}(T_1) + 38 \tag{3-95}$$

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

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.2) + \sum_{i=0}^{W} u(n-k+1+i)b(2-t+i.2) \tag{3-96}$$

โดยที่ n = 0, …, 39 t = 0, 1 และ พารามิเตอร์ W คือครึ่งหนึ่งของจำนวนค่าถ่วงน้ำหนักจะขึ้นกับความกว้างของ Hamming window (m)

3.5.2 การเข้ารหัสและการถอดรหัสค่าการประวิงเวลาของ adaptive-codebook เมื่อใช้เทคนิค HPDR ที่เศษส่วนพิตช์ 1/3

3.5.2.1 การวิเคราะห์หาเศษส่วนพิตช์

การวิเคราะห์พิตช์ในระดับจำนวนเต็ม ยังคงวิเคราะห์เหมือนเดิมตามหัวข้อ 3.1.2.7 ส่วนการวิเคราะห์หาเศษส่วนพิตช์ จะกระทำโดยการพิจารณาค่าสหสัมพันธ์ข้ามระหว่าง สัญญาณของเป้า x(n) และสัญญาณกระตุ้นในอดีตที่เก็บไว้ในบัฟเฟอร์ ค่าเศษส่วนพิตช์ t ที่ดีที่สุดที่ทำให้ค่าสหสัมพันธ์ข้ามนี้มีค่าสูงสุด จะถูกเลือกเป็นตัวแทนของเฟรมย่อยนั้นๆ ค่าสหสัมพันธ์ข้ามหาได้จากสมการต่อไปนี้

$$R(k)_t = \sum_{i=0}^{3} R(k-i)b(t+i.3) + \sum_{i=0}^{3} R(k+1+i)b(3-t+i.3), \quad t = 0,1,2 \qquad (3\text{-}97)$$

นิยาม R(k) จากสมการ (3-37) ค่าถ่วงน้ำหนัก b(n) เป็นค่าเดียวกันกับในหัวข้อ 3.5.1.1 ค่า t เป็น 0 1 และ 2 สอดคล้องกับเศษส่วน 0 1/3 และ 2/3 ตามลำดับ

3.5.2.2 การเข้ารหัส

การประวิงเวลาของพิตช์ T_1 และเศษส่วนที่ระดับความละเอียด 1/3 ถูกกำหนดด้วยค่าของ *frac* = -1 0 และ 1 เพราะฉะนั้นตัวชี้พิตช์ *P*1 จะถูกเข้ารหัสดังนี้

$$P1 = 3 \times (\text{int}(T_1) - 19) + \text{frac} - 1, \quad T_1 = [19,...,143], \text{frac} = [-1,0,1] \qquad (3\text{-}98)$$

ทำนองเดียวกัน P2 ก็ถูกเข้ารหัสด้วยหลักการเดียวกับ P1

การสร้างตัวเข้ารหัสถูกออกแบบให้มีความทนทานต่อการผิดพลาดของบิตข้อมูลนั้นโดยใช้พาริตีบิต *P*0 ซึ่งคำนวณมาจากตัวชี้ *P*1 ของเฟรมย่อยแรก พาริตีบิตที่ได้นั้นมาจากการทำ XOR จากบิตที่มีนัยสำคัญสูงสุด 6 บิตของ *P*1 ที่ตัวถอดรหัสพาริตีบิตนี้จะถูกคำนวณเพื่อใช้ในการตรวจสอบการผิดพลาดของบิตข้อมูบ ซึ่งถ้าได้ค่าไม่ตรงกันก็จะใช้วิธีการแก้ไขข้อผิดพลาดเข้ามาช่วย

3.5.2.3 การถอดรหัส

ถ้าไม่มีการผิดพลาดเกิดขึ้นกับ adaptive-codebook *P*1 ที่รับได้ โดยตรวจสอบจาก *P*0 *T*1 และ frac จะได้มาจาก *P*1 ตามสมการที่ (3-98) ส่วน *T*2 และ frac ของเฟรมย่อยที่สอง ก็ได้มาจาก *P*2 ด้วยหลักการเดียวกัน แต่หาก *P*0 บ่งชี้ความผิดพลาดของ *P*1 จะมีการนำค่าพิตช์ดีเลย์ *P*2 ของเฟรมก่อนหน้า มาใช้แทน *P*1 ของเฟรมปัจจุบัน

$$\text{int}(T_1) = (P1 + 2)/3 + 19 \qquad (3\text{-}99)$$

$$frac = P1 - 3 \text{ x int}(T_1) + 58 \qquad (3\text{-}100)$$

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

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.3) + \sum_{i=0}^{W} u(n-k+1+i)b(3-t+i.3) \qquad (3\text{-}101)$$

โดยที่ n = 0, …, 39 และ t = 0, 1, 2 และ พารามิเตอร์ W คือครึ่งหนึ่งของจำนวนค่าถ่วงน้ำหนักจะขึ้นกับความกว้างของ Hamming window (m)

3.5.3 การเข้ารหัสและการถอดรหัสค่าการประวิงเวลาของ adaptive-codebook เมื่อใช้เทคนิค HPDR ที่เศษส่วนพิตช์ 1/4

3.5.3.1 การวิเคราะห์หาเศษส่วนพิตช์

การวิเคราะห์พิตช์ในระดับจำนวนเต็ม ยังคงวิเคราะห์เหมือนเดิมตามหัวข้อ 3.1.2.7 ส่วนการวิเคราะห์หาเศษส่วนพิตช์ จะกระทำโดยการพิจารณาค่าสหสัมพันธ์ข้ามระหว่าง สัญญาณของเป้า x(n) และสัญญาณกระตุ้นในอดีตที่เก็บไว้ในบัฟเฟอร์ ค่าเศษส่วนพิตช์ t ที่ดีที่สุดที่ทำให้ค่าสหสัมพันธ์ข้ามนี้มีค่าสูงสุด จะถูกเลือกเป็นตัวแทนของเฟรมย่อยนั้นๆ ค่าสหสัมพันธ์ข้ามหาได้จากสมการต่อไปนี้

$$R(k)_t = \sum_{i=0}^{4} R(k-i)b(t+i.4) + \sum_{i=0}^{4} R(k+1+i)b(4-t+i.4), \quad t=0,1,2,3 \quad (3\text{-}102)$$

นิยาม R(k) จากสมการ 3-37 ค่าถ่วงน้ำหนัก b(n) เป็นค่าเดียวกันกับในหัวข้อ 3.5.1.1 ค่า t เป็น 0 1 2 และ 3 สอดคล้องกับเศษส่วน 0 1/4 2/4 และ –1/4 ตามลำดับ

3.5.3.2 การเข้ารหัส

การประวิงเวลาของพิตช์ T_1 และเศษส่วนที่ระดับความละเอียด 1/4 ถูกกำหนดด้วยค่าของ *frac* = -1 0 1 และ 2 เพราะฉะนั้นตัวชี้พิตช์ *P*1 จะถูกเข้ารหัสดังนี้

$$P1 = 4 \times (\text{int}(T_1) - 19) + \text{frac} - 1, \quad T_1 = [19,...,143], \text{frac} = [-1,0,1,2] \quad (3\text{-}103)$$

ทำนองเดียวกัน P2 ก็ถูกเข้ารหัสด้วยหลักการเดียวกับ P1

การสร้างตัวเข้ารหัสถูกออกแบบให้มีความทนทานต่อการผิดพลาดของบิตข้อมูลนั้นโดยใช้พาริตีบิต *P*0 ซึ่งคำนวณมาจากตัวชี้ *P*1 ของเฟรมย่อยแรก พาริตีบิตที่ได้นั้นมาจากการทำ XOR จากบิตที่มีนัยสำคัญสูงสุด 6 บิตของ *P*1 ที่ตัวถอดรหัสพาริตีบิตนี้จะถูกคำนวณเพื่อใช้ในการตรวจสอบการผิดพลาดของบิตข้อมูบ ซึ่งถ้าได้ค่าไม่ตรงกันก็จะใช้วิธีการแก้ไขข้อผิดพลาดเข้ามาช่วย

3.5.3.3 การถอดรหัส

ถ้าไม่มีการผิดพลาดเกิดขึ้นกับ adaptive-codebook *P*1 ที่รับได้ โดยตรวจสอบจาก *P*0 *T*1 และ frac จะได้มาจาก *P*1 ตามสมการที่ (3-103) ส่วน *T*2 และ frac ของเฟรมย่อยที่สอง ก็ได้มาจาก *P*2 ด้วยหลักการเดียวกัน แต่หาก *P*0 บ่งชี้ความผิดพลาดของ *P*1 จะมีการนำค่าพิตช์ดีเลย์ *P*2 ของเฟรมก่อนหน้า มาใช้แทน *P*1 ของเฟรมปัจจุบัน

$$\text{int}(T_1) = (P1 + 2)/4 + 19 \quad (3\text{-}104)$$

$$frac = P1 - 4 \text{ x } \text{int}(T_1) + 77 \quad (3\text{-}105)$$

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

$$v(n) = \sum_{i=0}^{W} u(n-k-i)b(t+i.4) + \sum_{i=0}^{W} u(n-k+1+i)b(4-t+i.4) \quad (3\text{-}106)$$

โดยที่ n = 0, …, 39 และ t = 0, 1, 2, 3 และ พารามิเตอร์ W คือครึ่งหนึ่งของจำนวนค่าถ่วงน้ำหนักจะขึ้นกับความกว้างของ Hamming window (m)

บทที่ 4
การทดลองการเข้ารหัสเสียงพูดโดยวิธี MP-CELP

วิทยานิพนธ์นี้แบ่งการทดลองการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ออกเป็น 2 ขั้นตอน ขั้นตอนแรกคือการทำการทดลองเพื่อเปรียบเทียบคุณภาพเสียงที่ผ่านการเข้ารหัสโดยวิธี MP-CELP ระหว่างภาษาไทยกับภาษาอังกฤษ จากนั้นทำการทดลองเพื่อเพิ่มประสิทธิภาพการเข้ารหัสเสียงพูดภาษาไทยโดยการเพิ่มความละเอียดของค่าพิตช์ คือนำเสนอเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง

อุปกรณ์และวิธีการเก็บข้อมูล

ในการทดลองบีบอัดเสียงพูดนี้ ใช้โปรแกรมที่พัฒนาบนภาษา Microsoft Visual C++ เวอร์ชัน 6.0 ทำงานบนเครื่องคอมพิวเตอร์ที่ใช้ไมโครโปรเซสเซอร์ Pentium III – 733 MHz ในส่วนการบันทึกเสียงจะใช้ A/D convertor ที่อยู่บน Sound Blaster Card ด้วยอัตราสุ่ม 8000 Hz ขนาดข้อมูลตัวอย่างละ 8 บิต โดยเก็บเป็นแฟ้มข้อมูลชนิด .raw และในทางกลับกัน การแสดงผลจะผ่าน D/A converter ที่มีอยู่ใน Sound Blaster Card เช่นกัน

4.1 การเปรียบเทียบคุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยวิธี MP-CELP ระหว่างเสียงพูดภาษาไทย (tonal language) กับเสียงพูดภาษาอังกฤษ (toneless language)

การทดลองแรกนี้ เพื่อวัดความแตกต่างระหว่างคุณภาพเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ ที่ผ่านการเข้ารหัสโดยวิธี MP-CELP เพราะเนื่องมาจากเสียงพูดภาษาไทยเป็นเสียงดนตรี คือส่วนย่อยที่เป็นคำ หรือพยางค์ ที่มีเสียงวรรณยุกต์ต่างกัน (การเปลี่ยนแปลงของพิตช์ผ่านช่วงเวลาในแต่ละคำหรือพยางค์) จะทำให้ความหมายแตกต่างกันออกไปดังที่ได้อธิบายไว้ในหัวข้อ 3.4 มีผลทำให้คุณภาพเสียงที่ผ่านการเข้ารหัสไม่ได้เท่ากับมาตรฐานในภาษาต้นแบบการทดลอง

4.1.1 ประโยคที่ใช้ทดสอบ

ในการทดลอง 4.1 นี้ ได้ใช้ผู้พูดประโยคทดสอบในแต่ละภาษาที่เป็นเจ้าของภาษา มีรายละเอียดดังนี้

ประโยคทดสอบภาษาอังกฤษ ใช้ผู้พูดเจ้าของภาษา เป็นเพศชาย 3 คนและ เพศหญิง 3 คน มีประโยคทดสอบ 36 ประโยค บันทึกจากยูบีซีเคเบิลทีวี ดังแสดงในตารางที่ 4.1 คอลัมน์แรก

ประโยคทดสอบภาษาไทย ใช้ผู้พูดเจ้าของภาษา เป็นเพศชาย 3 คนและ เพศหญิง 3 คน มีประโยคทดสอบ 36 ประโยค ดังแสดงในตารางที่ 4.1 คอลัมน์ที่สอง

ตารางที่ 4.1 ประโยคทดสอบการเข้ารหัส

ลำดับ	ประโยคทดสอบภาษาอังกฤษ	ประโยคทดสอบภาษาไทย
1	After taking off from Cuba this morning.	-คน-ทำ-บาป-อวด-ตัว-ว่า-เก่ง- /khon0/thm0/baap0/?uuat1/tuua0/waa2/keng1/
2	American companies.	-เธอ-ทำ-ฉัน-บ่า-ปวด-ไป-หมด- /thqq0/tham0/chan4/baa1/puuat1/paj0/mot1/
3	An operating System.	-เขา-เป็น-ญาติ-อำ-ภา- /kaw4/pen0/jaat2/?am0/phaa0/
4	Before the end of the year.	-ใน-ปาก-อิน-ทรีย์-มี-ปลา-สอง-ตัว- /naj0/paak1/?in0/sii0/mii0/plaa0/s@@ng4/tuua0/
5	Billion dollars.	-มา-ลี-โดน-หญ้า-ตำ-ที่-ขา- /maa0/lii0/doon0/jaa2/tam0/thii2/khaa4/
6	Cannot confirm.	-คน-กิน-ข้าว-แต่-ป่า-กิน-น้ำ- /khon0/kin0/khaaw2/txx1/paa1/kin0/naam3/
7	Financial Sector.	-แม่-ไป-ตาม-อา-ที่-บ้าน- /mxx2/paj0/taam0/?aa0/thii2/baan2/
8	For the next three months.	-แม่-บอก-ว่า-ตา-มา-อยู่-ที่-บ้าน- /mxx2/b@@k1/waa2/taa0/maa0/juu1/thii2/baan2/
9	High oil prize.	-ฉัน-เห็น-ว่าน-ออก-ดอก-ตั้ง-หลาย-ต้น- /chan0/hen4/waan2/?@@k1/d@@k1/tang2/laaj4/ton2/
10	HKEX is launching an internet IQ website.	-คำ-ว่า-หนอก-หมาย-ถึง-ต้น-คอ-วัว- /kham0/waa2/n@@k1/maaj4/thvng4/ton2/kh@@0/wuua0/
11	I believe that.	-ฉัน-จะ-ลอง-อม-เงิน-คุณ-แม่-ดู- /chan0/ca0/l@@ng0/?om0/ngqn0/kuun0/mxx2/duu0/
12	I talk to you this morning.	-พี่-ลอ-งม-แหวน-ขึ้น-มา-ให้- /phii2/l@@0/ngom0/wxxn0/khvn2/maa0/haj2/
13	I think we have to wait one or two months.	-น้อง-จะ-เอา-ว่าว-อัน-นั้น- /n@@ng3/ca1/?aw0/waaw2/?an0/naan3/
14	Impact from the oil crisis.	-อา-จารย์-บอก-ว่า-วัน-นี้-เป็น-วัน-ดี- /?aa0/caan0/b@@k1/waa2/wan0/nii3/pen0/wan0/dii0/
15	In credit markets.	-ตา-กล้อง-บอก-ให้-เอน-ตัว-ไป-ทาง-ซ้าย- /taa0/kl@ng2/b@@k1/haaj2/?iiang0/tuua0/paj0/taang0/saaj3/
16	In single family.	-ปาก-กา-รา-คา-ห้า-เยน-เท่า-นั้น- /paak0/kaa0/raa0/khaa0/haa2/jen0/thaw2/nan3/
17	In the first half of the year.	-พวก-นั้น-โดน-ปรับ-ราย-ตัว- /phuuak2/nan3/don0/prab1/raaj0/tuua0/
18	In Tokyo, the Nikei is closing up for 10 percent.	-ฝน-ตก-ประ-ปราย-เป็น-ประ-จำ- /fon4/tok1/pra0/praaj0/pen0/pra0/cam0/
19	It is not just the asian crisis.	-เธอ-ปัด-รา-ที่-ข-นม-ออก- /thqq0/pad1/raa0/thii2/kha1/nom4/?@@k1/

ตารางที่ 4.1 ประโยคทดสอบการเข้ารหัส (ต่อ)

ลำดับ	ประโยคทดสอบภาษาอังกฤษ	ประโยคทดสอบภาษาไทย
20	It's not really new.	-มา-นะ-ปะ-ตรา-ที่-หน้า-รถ-ยนต์- /maa0/na2/pa1/traa0/thii2/naa2/rot2/jon0/
21	June and July.	-ส่วน-ภู-มิ-ภาค-รอง-ลง-มา-จาก-ส่วน-กลาง- /suuan1/phuu0/mi2/phaak2/r@@ng0/long0/maa0/caak0/suuan1/klaang0/
22	Radio Networking.	-ตำ-รวจ-ใช้-ผ้า-กรอง-ฝุ่น-ปิด-ปาก- /tam0/ruuat10/chaaj3/phaa2/kr@@ng0/fun1/pid1/paak1/
23	Start first month in August.	-ช่วย-กัน-จับ-ปลีก-ไป-ไกล-ไกล- /chuuaj2/kan0/cap1/liik1/paj0/klaj0/klaj0/
24	Take over target.	-ผม-จะ-ปลีก-ตัว-มา-ทัน-ที- /phom4/ca0/pliik1/tuua0/maa0/than0/thii0/
25	The market capitalization.	-เขา-อยาก-สัก-ลาย-เสือ-ที่-แขน- /khaaw4/jaak1/sak1/laaj0/svva4/thii2/khxxn4/
26	The water temperature is reported warm.	-น้ำ-ใน-สระ-กลาย-เป็น-สี-ดำ- /naam3/naj0/sa1/klaaj0/pen0/sii4/dam0/
27	The Neeke had been down 2 percent.	-เขา-แห่-นาค-เวียน-รอบ-โบสถ์- /khaw4/hxx1/naak2/wiian0/r@@p2/boot1/
28	The government.	-ข้าว-เปลือก-อยู่-หน้า-เกวียน-สอง-กอง- /khaaw2/plqqk1/juu1/naa2/kwiian0/s@@ng4/k@@ng0/
29	There were nearly 5 percent last month.	-คำ-ว่า-เตียบ-แปล-ว่า-ตะ-ลุ่ม- /kham0/waa2/tiip1/plxx0/waa2/ta1/lum2/
30	There will be a lot of people.	-ชาติ-ชาย-โดน-ครู-ติ-อับ-อาย-ขาย-หน้า- /chaat2/chaaj0/doon0/khruu0/ti1/?ap1/?aaj0/khaaj4/naa2/
31	Three million Hongkong dollars.	-เธอ-จบ-จาก-โรง-เรียน-เตรียม-อุ-ดม- /thqq0/cop1/caak1/roong0/riian0/triiam0/?u1/dom0/
32	Turn into a round.	-ตอน-เรียน-ปริญ-ญา-ตรี-อัม-พร-เคย-ได้-รับ-ทุน- /t@@n0/riian0/prin0/ja0/trii0/?am0/ph@@n0/khqqj0/daaj2/rap3/thun0/
33	Two billion dollors.	-เขา-เคือง-ฉัน-ด้วย-เรื่อง-เล็ก-น้อย- /kh@w4/kvvang0/chan4/duuaj2/rvvang2/lek0/n@@j3/
34	Two years ago.	-เธอ-คือ-อัง-ค-ณา-ทิม-ดี- /thqq0/khvv0/@@ang0/kha0/naa0/thim0/dii0/
35	Very Difficult.	-ลำ-ดวน-เป็น-คน-ใจ-คอ-รวน-เร- /lam0/duuan0/pen0/khon0/caj0/kh@@0/ruuan0/ree0/
36	Well, that stratagy looks like it works.	-ปิ-ติ-ปัก-ธง-ที่-รู-อัน-สุด-ท้าย- /pi0/ti0/pak0/thong0/thii2/ruu0/?an0/sud0/thaaj3/

4.1.2 คุณภาพการเข้ารหัสโดยใช้การวัดค่าเชิงวัตถุ (ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน)

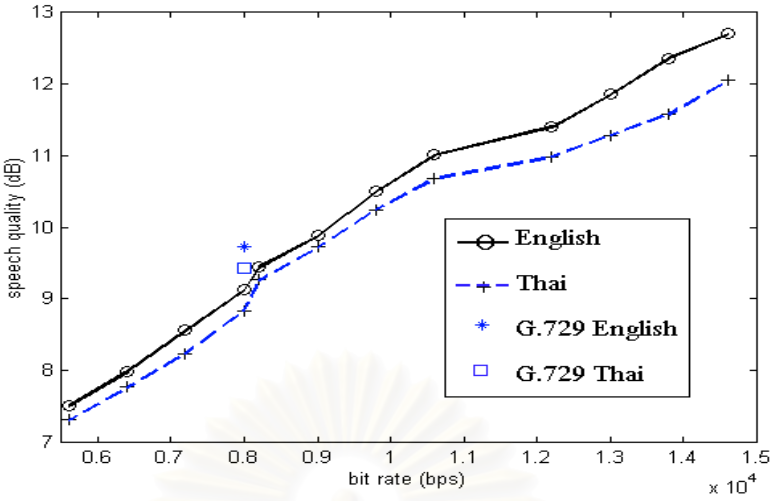
ผลการทดลองจากการเข้ารหัสและถอดรหัส ใช้ประโยคทดสอบในหัวข้อ 4.1.1 และวัดค่าคุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยใช้สมการที่ 2-2 โดยการเข้ารหัสจะใช้ตัวเข้ารหัสหลักทั้งสามอัตราที่นำเสนอไว้ในหัวข้อ 3.1 และใช้การปรับระดับอัตราการเข้ารหัสจากอัตราหลักเป็นอีก 3 อัตรา ที่นำเสนอไว้ในหัวข้อ 3.3 แสดงเป็นค่าคุณภาพเสียงพูดเฉลี่ยในตารางที่ 4.2 และมีค่าคุณภาพเสียงอ้างอิงเพื่อใช้เปรียบเทียบ จากการเข้ารหัสโดยวิธี CS-ACELP (G.729) ในตารางที่ 4.3 และนำเสนอในรูปของกราฟเส้นเพื่อเปรียบเทียบผลในแง่ต่างๆ ดังรูปที่ 4.1

ตารางที่ 4.2 คุณภาพเสียงพูดที่ผ่านการเข้ารหัส MP-CELP โดยใช้ค่า SegPSNR

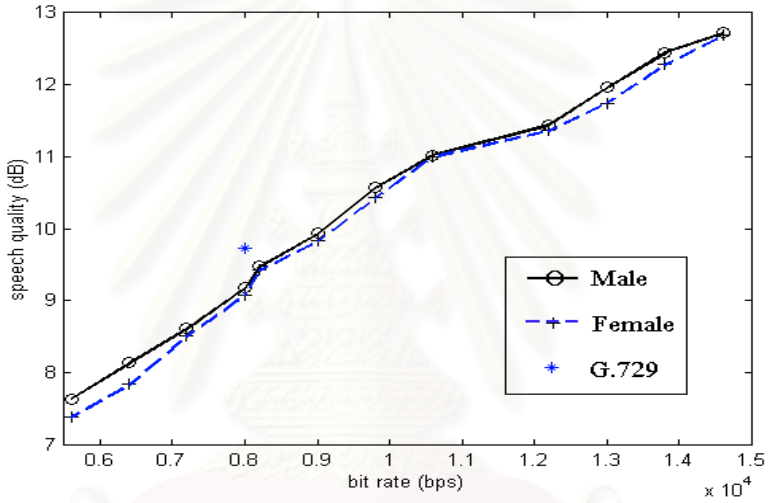
อัตราเข้ารหัสหลัก (bps)	ขั้นการปรับระดับ	อัตราเข้ารหัสรวม (bps)	คุณภาพเสียงเพศชายโดยเฉลี่ย (dB)		คุณภาพเสียงเพศหญิงโดยเฉลี่ย (dB)	
			อังกฤษ	ไทย	อังกฤษ	ไทย
5600	-	5600	7.64	7.41	7.39	7.22
	1	6400	8.14	7.84	7.84	7.70
	2	7200	8.62	8.32	8.51	8.16
	3	8000	9.18	9.01	9.08	8.66
8200	-	8200	9.48	9.35	9.42	9.18
	1	9000	9.93	9.88	9.83	9.56
	2	9800	10.57	10.50	10.43	9.99
	3	10600	11.02	10.83	11.00	10.53
12200	-	12200	11.44	11.09	11.36	10.87
	1	13000	11.96	11.33	11.74	11.23
	2	13800	12.44	11.84	12.27	11.32
	3	14600	12.71	12.19	12.68	11.90

ตารางที่ 4.3 คุณภาพเสียงพูดที่ผ่านการเข้ารหัส CS-CELP (ค่าอ้างอิง) โดยใช้ค่า SegPSNR จากบทความ [28]

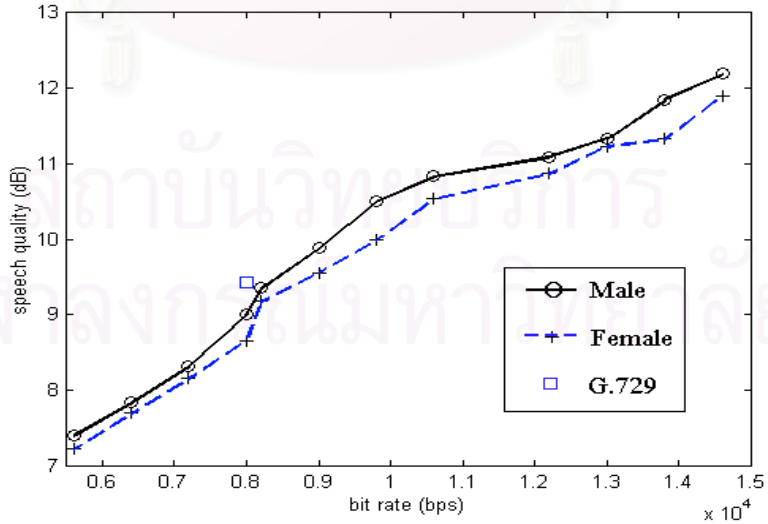
อัตราเข้ารหัส (bps)	คุณภาพเสียงเพศชายโดยเฉลี่ย (dB)		คุณภาพเสียงเพศหญิงโดยเฉลี่ย (dB)	
	อังกฤษ	ไทย	อังกฤษ	ไทย
8000	9.74	9.46	9.72	9.39



a.



b.



c.

รูปที่ 4.1 a. เปรียบเทียบค่า SegPSNR ระหว่างเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ

b. เปรียบเทียบค่า SegPSNR ระหว่างเสียงพูดเพศชายกับเพศหญิง ในภาษาอังกฤษ

c. เปรียบเทียบค่า SegPSNR ระหว่างเสียงพูดเพศชายกับเพศหญิง ในภาษาไทย

4.1.3 คุณภาพการเข้ารหัสโดยใช้การวัดค่าเชิงผู้ฟัง (ค่า MOS)

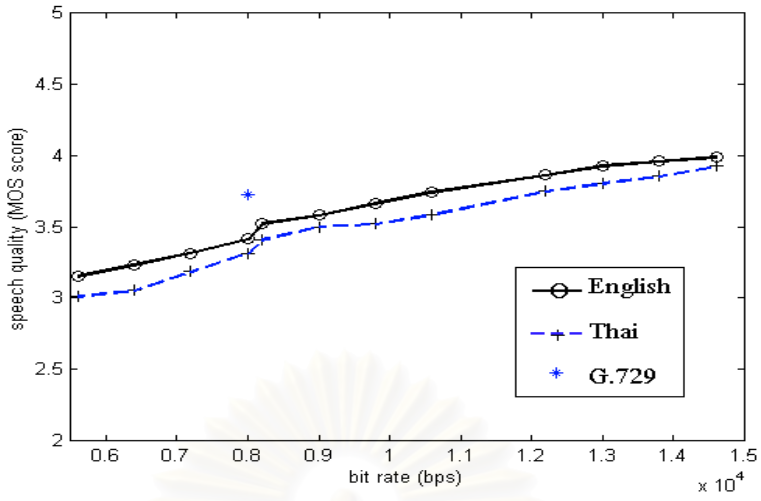
ผลการทดลองจากการเข้ารหัสและถอดรหัส โดยวัดค่าโดยใช้ผู้ฟัง 12 คนเป็นผู้ให้คะแนน โดยมีค่าเป็นคะแนนเต็มจาก 1 ถึง 5 แสดงเป็นค่าคุณภาพเสียงพูดเฉลี่ยในตารางที่ 4.4 และมีค่าคุณภาพเสียงอ้างอิงจากการเข้ารหัสโดยวิธี CS-ACELP (G.729) เพื่อใช้เปรียบเทียบ และปรับเทียบสำหรับผู้ฟังที่ให้คะแนน ในตารางที่ 4.5 พร้อมกับนำเสนอในรูปของกราฟเส้นเพื่อเปรียบเทียบผลในแง่ต่างๆ ดังรูปที่ 4.2

ตารางที่ 4.4 คุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยใช้ค่า MOS

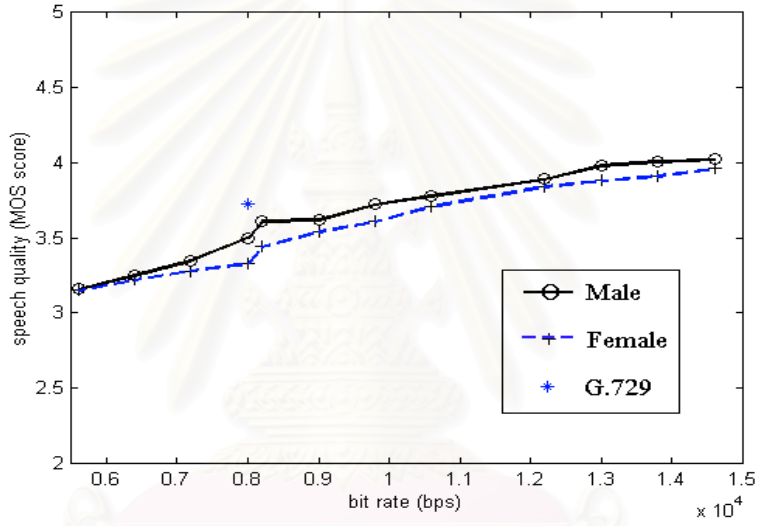
อัตราเข้ารหัสหลัก (bps)	ขั้นการปรับระดับ	อัตราเข้ารหัสรวม (bps)	คุณภาพเสียงเพศชายโดยเฉลี่ย (จุด)		คุณภาพเสียงเพศหญิงโดยเฉลี่ย (จุด)	
			อังกฤษ	ไทย	อังกฤษ	ไทย
5600	-	5600	3.16	3.02	3.15	3.01
	1	6400	3.25	3.08	3.22	3.03
	2	7200	3.35	3.21	3.28	3.16
	3	8000	3.50	3.35	3.33	3.28
8200	-	8200	3.61	3.43	3.44	3.39
	1	9000	3.62	3.50	3.54	3.50
	2	9800	3.72	3.56	3.61	3.48
	3	10600	3.78	3.61	3.71	3.56
12200	-	12200	3.89	3.78	3.84	3.72
	1	13000	3.98	3.81	3.88	3.80
	2	13800	4.01	3.88	3.91	3.83
	3	14600	4.02	3.93	3.96	3.92

ตารางที่ 4.5 คุณภาพเสียงพูดที่ผ่านการเข้ารหัส CS-CELP (ค่าอ้างอิง) โดยใช้ค่า MOS จากบทความ [7 และ 15]

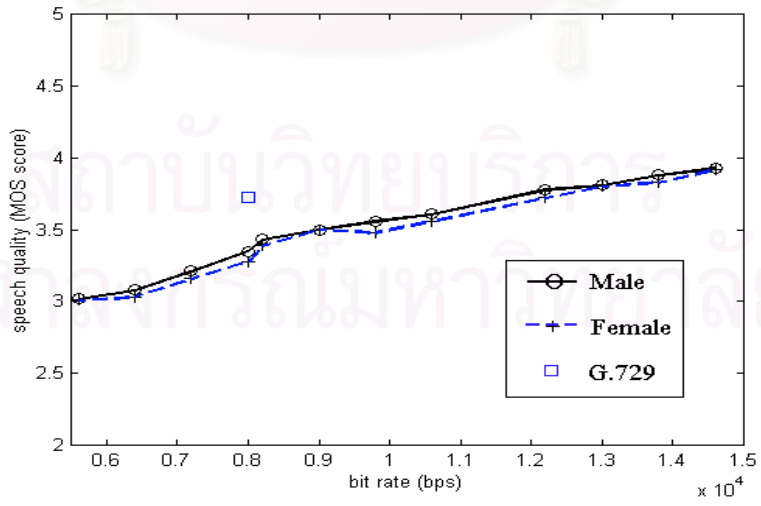
อัตราเข้ารหัส (bps)	คุณภาพเสียงโดยเฉลี่ย (จุด)
8000	3.72



a.



b.



c.

รูปที่ 4.2 a. เปรียบเทียบค่าเชิงผู้ฟังระหว่างเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ

b. เปรียบเทียบค่าเชิงผู้ฟังระหว่างเสียงพูดเพศชายกับเพศหญิง ในภาษาอังกฤษ

c. เปรียบเทียบค่าเชิงผู้ฟังระหว่างเสียงพูดเพศชายกับเพศหญิง ในภาษาไทย

จากกราฟในรูปที่ 4.1 และ 4.2 แสดงค่าคุณภาพเสียง อัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน (SegPSNR) และค่าเชิงผู้ฟัง (ค่า MOS) ตามลำดับ มีแนวโน้มของกราฟที่สอดคล้องกันไม่ว่าจะเป็นเสียงพูดในภาษาไทย ภาษาอังกฤษ ของทั้งเพศชายและเพศหญิง แต่ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วนจะแสดงความแตกต่างของกราฟแต่ละเส้นได้ดีกว่า ในกรณีการเปรียบเทียบระหว่างเสียงพูดภาษาไทยกับเสียงพูดภาษาอังกฤษ และการเปรียบเทียบระหว่างเสียงพูดเพศชายกับเพศหญิงในภาษาไทย ส่วนในกรณีการเปรียบเทียบระหว่างเสียงพูดเพศชายกับเพศหญิงในภาษาอังกฤษ ค่าเชิงผู้ฟังจะแสดงความแตกต่างของกราฟแต่ละเส้นได้ดีกว่า

พิจารณาแยกแยะระหว่างคุณภาพเสียงของเพศชาย และเพศหญิง นำเสนอเป็นกราฟในรูป 4.1b c และรูป 4.2b c ผลการทดลองที่ได้สอดคล้องกัน คือ มีคุณภาพเสียงที่ดีขึ้นเมื่อปรับระดับอัตราการเข้ารหัสขึ้น แต่คุณภาพเสียงของเพศหญิงจะด้อยกว่าคุณภาพเสียงของเพศชายเล็กน้อย คือ ประมาณ 0.13 dB สำหรับเสียงพูดภาษาอังกฤษ และประมาณ 0.27 dB สำหรับเสียงพูดภาษาไทย โดยเฉลี่ยตลอดทุกช่วงอัตราการเข้ารหัส หรือประมาณ 0.09 MOS score สำหรับเสียงพูดภาษาอังกฤษ และประมาณ 0.04 MOS score สำหรับเสียงพูดภาษาไทย โดยเฉลี่ยตลอดทุกช่วงอัตราการเข้ารหัส ผลการทดลองนี้สอดคล้องกับบทความ [45] ที่ใช้การเข้ารหัสบนพื้นฐานของ CELP เหมือนกัน สาเหตุที่เป็นเช่นนี้คือ โดยธรรมชาติแล้ว เสียงพูดของเพศหญิงจะมีความถี่มูลฐานสูงกว่าของเสียงพูดของเพศชาย การเข้ารหัสที่อยู่บนพื้นฐานของ CELP จะมีการวิเคราะห์และส่งค่าพิตช์ที่ความละเอียดในระดับหนึ่ง ทำให้ความถูกต้องของความถี่มูลฐานของเสียงพูดเพศชายจะสูงกว่าของเพศหญิง

พิจารณาเปรียบเทียบระหว่างคุณภาพเสียงพูดภาษาอังกฤษ และภาษาไทย พบว่า คุณภาพเสียงพูดภาษาไทยจะต่ำกว่า คุณภาพเสียงพูดในภาษาอังกฤษ คือประมาณ 0.30 dB สำหรับเสียงพูดของเพศชาย และประมาณ 0.44 dB สำหรับเสียงพูดของเพศหญิง โดยเฉลี่ยตลอดทุกช่วงอัตราการเข้ารหัส หรือประมาณ 0.14 MOS score สำหรับเสียงพูดของเพศชาย และประมาณ 0.10 MOS score สำหรับเสียงพูดของเพศหญิง โดยเฉลี่ยตลอดทุกช่วงอัตราการเข้ารหัส ผลการทดลองนี้สอดคล้องกับบทความ [45] ที่ใช้การเข้ารหัสบนพื้นฐานของ CELP เหมือนกัน สาเหตุที่เป็นเช่นนี้ เพราะเสียงพูดภาษาไทยเป็นเสียงดนตรี ที่ให้ความสำคัญกับการเปลี่ยนแปลงของความถี่มูลฐานหรือพิตช์ ดังที่ได้กล่าวเบื้องต้นในหัวข้อ 3.4 นั่นคือการเข้ารหัส MP-CELP ที่มีการวิเคราะห์และส่งค่าพิตช์ด้วยความละเอียดระดับจำนวนเต็มหนึ่ง ไม่เพียงพอสำหรับการเข้ารหัสเสียงพูดภาษาไทยที่ต้องการคุณภาพเสียงที่ทัดเทียมกับเสียงพูดภาษาอังกฤษ

พิจารณาเปรียบเทียบการเข้ารหัส MP-CELP กับการเข้ารหัสมาตรฐาน ITU G.729 ที่อัตราคงที่ 8000 bps [9] พบว่าคุณภาพเสียงที่ได้ใกล้เคียงกัน โดยการเข้ารหัส MP-CELP ที่อัตรา 8000 bps จะให้ คุณภาพเสียงที่ด้อยกว่าเล็กน้อย การเข้ารหัส MP-CELP ที่อัตรา 8000 bps จะเป็นการ

ปรับระดับอัตราการเข้ารหัส 3 ขั้น จากอัตราการเข้ารหัสหลัก 5600 bps การเพิ่มจำนวนพัลส์ในสัญญาณกระตุ้นแบบคงที่ (fixed-codebook) โดยการหาขนาดและตำแหน่งที่เหมาะสมรองลงไป ซึ่งอาจไม่ใช่ชุดตำแหน่งพัลส์ที่ดีที่สุดที่ทำให้พจน์ในสมการ 3-52 มีค่ามากที่สุด นั่นคือจะไม่ดีเท่ากับการหาขนาดและตำแหน่งที่เหมาะสมที่สุดโดยตรง เหมือนของการเข้ารหัส CS-ACELP

พิจารณาการปรับระดับอัตราการเข้ารหัสในแต่ละขั้น จาก 1 ขั้น เป็น 2 ขั้น และ 3 ขั้น ตามลำดับ มีผลทำให้คุณภาพเสียงที่ผ่านการเข้ารหัสแล้วดีขึ้น คือประมาณขั้นละ 0.42 dB หรือ 0.07 MOS score สำหรับเสียงพูดของเพศชายภาษาไทย ประมาณ 0.40 dB หรือ 0.08 MOS score สำหรับเสียงพูดของเพศหญิงภาษาไทย ประมาณ 0.44 dB หรือ 0.07 MOS score สำหรับเสียงพูดของเพศชายภาษาอังกฤษ ประมาณ 0.39 dB หรือ 0.08 MOS score สำหรับเสียงพูดของเพศหญิงภาษาอังกฤษ แสดงให้เห็นว่าการปรับระดับอัตราการเข้ารหัสโดยการต่อเพิ่มส่วนขยาย (enhancement layer) สามารถเพิ่มคุณภาพเสียงได้ในระดับหนึ่ง

4.2 การเพิ่มประสิทธิภาพการเข้ารหัสเสียงพูดภาษาไทยโดยการเพิ่มความละเอียดของค่าพิตช์

การทดลองที่สอง เพื่อวัดความแตกต่างระหว่างคุณภาพเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสโดยวิธี MP-CELP กรณีที่ยังไม่เพิ่มความละเอียดของค่าพิตช์ดีเลย์ (ดั้งเดิม) กับ คุณภาพเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสโดยวิธี MP-CELP กรณีที่เพิ่มความละเอียดของค่าพิตช์ดีเลย์ในระดับต่างๆ ด้วยเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง (HPDR.1/*) อีกทั้งเพื่อทดสอบว่า การวิเคราะห์และส่งค่าพิตช์ที่ความละเอียดระดับใดจึงจะให้คุณภาพการเข้ารหัสดีที่สุด

4.2.1 ประโยคที่ใช้ทดสอบ

ในการทดลอง 4.2 นี้ ได้ใช้อาสาสมัครผู้พูดภาษาไทย เป็นเพศชาย 16 คนและ เพศหญิง 16 คน และใช้ประโยคทดสอบเดียวกันกับการทดลองในหัวข้อ 4.1 ใช้ทั้ง 36 ประโยค

4.2.2 การหาความกว้างของ Hamming Window ที่เหมาะสมสำหรับการประมาณค่าในช่วงของสัญญาณกระตุ้นแบบปรับตัว

ทดลองปรับความกว้างของ Hamming Window โดยแปรค่าเพื่อให้จำนวนค่าถ่วงน้ำหนักเป็น 5 9 13 17 21 25 29 33 และ 37 ค่า แล้วทดสอบกับการเข้ารหัสเสียงพูดภาษาไทย โดยใช้การเข้ารหัสหลักทั้งสามตัว รูปที่ 4.3 และ 4.4 แสดงลักษณะของค่าถ่วงน้ำหนักที่ได้เมื่อแปรความกว้างของ Hamming Window เป็นค่าต่างๆ ผลการทดลองอยู่ในตารางที่ 4.6 และ 4.7 โดยแสดงเป็นค่า

เฉลี่ยของคุณภาพเสียง ระหว่างเสียงพูดของเพศชายและของเพศหญิง จากประโยคทดสอบในหัวข้อ 4.2.1 พร้อมกับนำเสนอในรูปของกราฟเส้นเพื่อเปรียบเทียบผลดังรูปที่ 4.3 และ 4.4

ตารางที่ 4.6 คุณภาพเสียงพูด (SegPSNR) ที่ความกว้างของ Hamming Window ที่จำนวนค่าถ่วงน้ำหนักต่างๆ โดยใช้เทคนิค HPDR1/2 เทียบกับการไม่ใช้

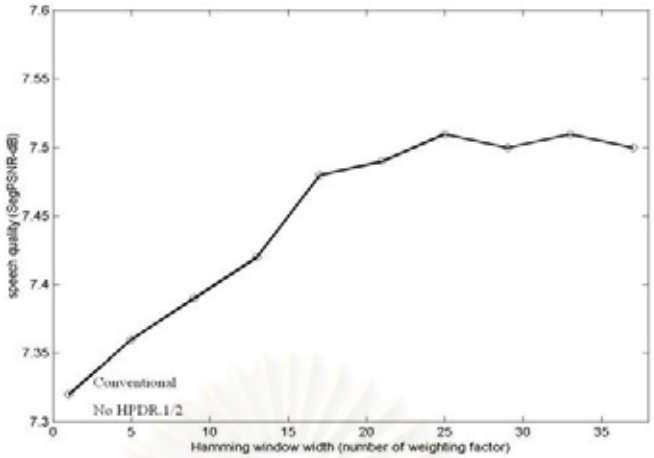
ความกว้าง Hamming Window		คุณภาพเสียง		
		อัตราเข้ารหัสหลัก 5600 bps	อัตราเข้ารหัสหลัก 8200 bps	อัตราเข้ารหัสหลัก 12200 bps
ค่า SegPSNR	-	7.32	9.27	10.98
	5	7.36	9.36	11.02
	9	7.39	9.41	11.09
	13	7.42	9.51	11.13
	17	7.48	9.56	11.15
	21	7.49	9.57	11.18
	25	7.51	9.58	11.17
	29	7.50	9.57	11.19
	33	7.51	9.59	11.21
	37	7.50	9.61	11.19

ตารางที่ 4.7 คุณภาพเสียงพูด (MOS score) ที่ความกว้างของ Hamming Window ที่จำนวนค่าถ่วงน้ำหนักต่างๆ โดยใช้เทคนิค HPDR1/2 เทียบกับการไม่ใช้

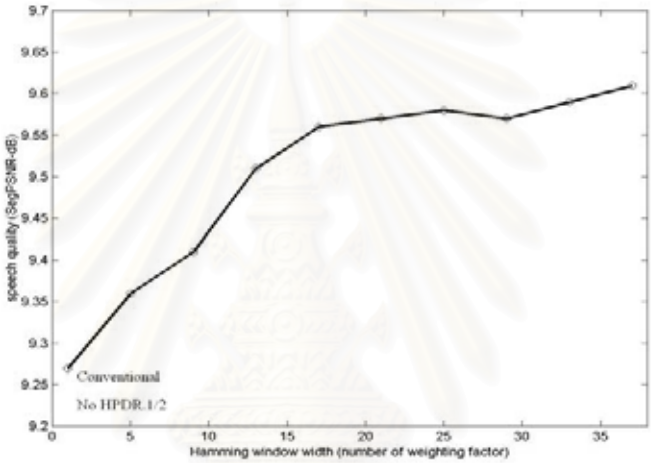
ความกว้าง Hamming Window		คุณภาพเสียง		
		อัตราเข้ารหัสหลัก 5600 bps	อัตราเข้ารหัสหลัก 8200 bps	อัตราเข้ารหัสหลัก 12200 bps
ค่า MOS score	-	3.02	3.41	3.75
	5	3.06	3.44	3.73
	9	3.11	3.48	3.76
	13	3.14	3.51	3.76
	17	3.16	3.52	3.78
	21	3.18	3.55	3.79
	25	3.19	3.55	3.80
	29	3.18	3.56	3.78
	33	3.17	3.53	3.79
	37	3.19	3.57	3.78

เนื่องมาจาก ฟังก์ชันถ่วงน้ำหนัก b(n) ที่ถูกจำกัดช่วงด้วยหน้าต่างนี้ เป็น sin(n)/n จะมีค่าลดลงเมื่อ n มีค่ามากขึ้น ทำให้สามารถตัดทิ้งไปได้ โดยใช้หน้าต่าง Hamming นี้ จากกราฟในรูปที่ 4.3 และ 4.4 จะเห็นได้ว่า เมื่อเพิ่มความกว้างของหน้าต่าง Hamming ไปจนถึงค่าหนึ่ง ค่าคุณภาพภาพเสียงจะเพิ่มขึ้นไม่มาก แสดงว่า ค่าที่ทำให้คุณภาพเสียงเริ่มอิ่มตัว จะสามารถประมาณได้ว่าเป็นค่าที่พอเพียงต่อการประมาณค่าในช่วงสำหรับหาสัญญาณกระตุ้นแบบปรับตัวได้แล้ว ค่าที่วิทยานิพนธ์นี้เลือกใช้ก็คือ ความกว้างของหน้าต่างที่สอดคล้องกับจำนวนค่าถ่วงน้ำหนักที่ 21 ค่า

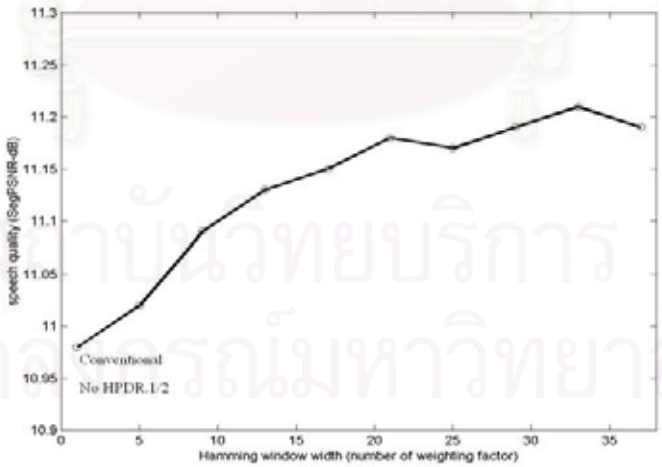
และค่านี้จะใช้เป็นหลักในเทคนิคการวิเคราะห์พิตช์ดีเลย์ที่ความละเอียดสูงทุกๆ ระดับความละเอียด



a.



b.



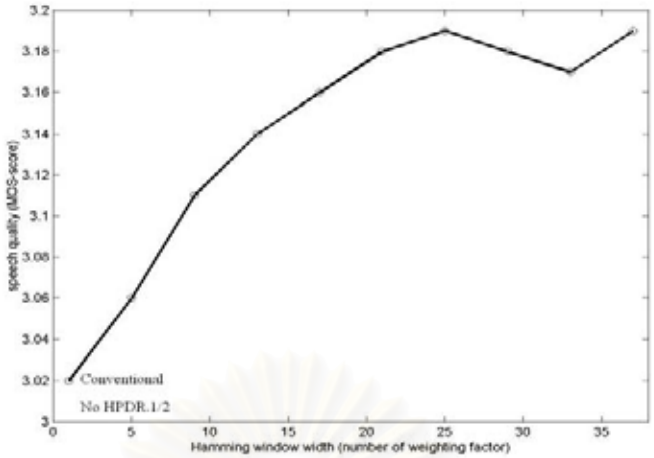
c.

รูปที่ 4.3 คุณภาพเสียงพูด SegPSNR ที่ค่าความกว้างของ Hamming Window ต่าง ๆ

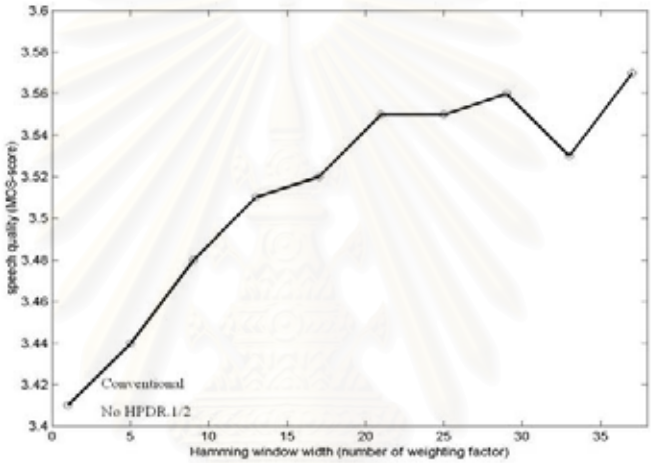
a. โดยใช้อัตราเข้ารหัสหลัก 5600 bps

b. โดยใช้อัตราเข้ารหัสหลัก 8200 bps

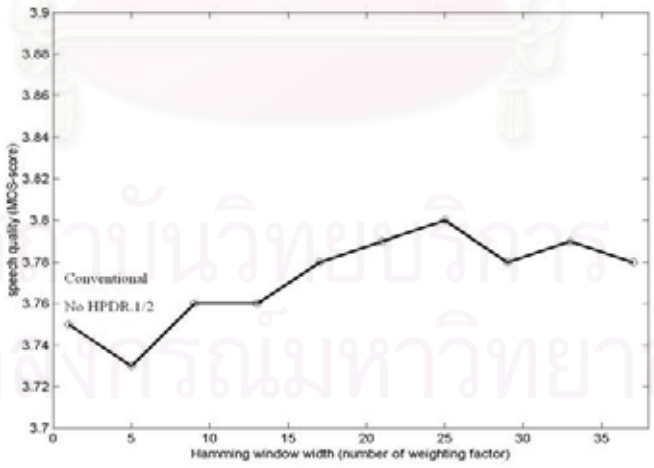
c. โดยใช้อัตราเข้ารหัสหลัก 12200 bps



a.



b.



c.

รูปที่ 4.4 คุณภาพเสียงพูด MOS score ที่ค่าความกว้างของ Hamming Window ต่าง ๆ

a. โดยใช้อัตราเข้ารหัสหลัก 5600 bps

b. โดยใช้อัตราเข้ารหัสหลัก 8200 bps

c. โดยใช้อัตราเข้ารหัสหลัก 12200 bps

4.2.3 คุณภาพการเข้ารหัสโดยใช้การวัดค่าเชิงวัตถุ (ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน)

ผลการทดลองจากการเข้ารหัสและถอดรหัส ใช้ประโยคทดสอบในหัวข้อ 4.1.1 และวัดค่าคุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยใช้สมการที่ 2-2 การเข้ารหัสจะใช้ตัวเข้ารหัสหลักทั้งสามอัตราที่นำเสนอไว้ในหัวข้อ 3.1 และใช้การปรับระดับอัตราการเข้ารหัสจากอัตราหลักเป็นอีก 3 อัตราที่นำเสนอไว้ในหัวข้อ 3.3 โดยเป็นการเปรียบเทียบเป็นกลุ่มอัตราการเข้ารหัส เนื่องจากเทคนิคที่นำเสนอ จะทำให้อัตราการเข้ารหัสเพิ่มขึ้นจากอัตราเข้ารหัสรวม 200, 400 และ 400 pbs สำหรับ HPDR.1/2, HPDR.1/3 และ HPDR.1/4 ตามลำดับ นำเสนอเป็นค่าคุณภาพเสียงพูดเฉลี่ยในตารางที่ 4.8 พร้อมกับนำเสนอในรูปของกราฟเส้นเพื่อเปรียบเทียบผลในแง่ต่างๆ ดังรูปที่ 4.5

ตารางที่ 4.8 คุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยใช้ค่า SegPSNR

อัตราเข้ารหัสหลัก (bps)	ขั้นการปรับระดับ	อัตราเข้ารหัสรวม (bps)	คุณภาพเสียงเพศชาย โดยเฉลี่ย (dB)				คุณภาพเสียงเพศหญิง โดยเฉลี่ย (dB)			
			ดั้งเดิม	HPDR.1/2 +200 bps	HPDR.1/3 +400 bps	HPDR.1/4 +400 bps	ดั้งเดิม	HPDR.1/2 +200 bps	HPDR.1/3 +400 bps	HPDR.1/4 +400 bps
5600	-	5600+	7.41	7.65	7.74	7.80	7.22	7.33	7.66	7.69
	1	6400+	7.84	7.93	8.09	8.11	7.70	7.92	8.03	8.10
	2	7200+	8.32	8.84	8.92	8.98	8.16	8.22	8.89	8.92
	3	8000+	9.01	9.35	9.57	9.57	8.66	9.33	9.44	9.45
8200	-	8200+	9.35	9.65	9.86	9.98	9.18	9.49	9.79	9.83
	1	9000+	9.88	10.17	10.33	10.36	9.56	10.10	10.14	10.23
	2	9800+	10.50	10.65	10.84	10.88	9.99	10.59	10.64	10.68
	3	10600+	10.83	11.05	11.18	11.25	10.53	10.92	11.08	11.18
12200	-	12200+	11.09	11.24	11.39	11.45	10.87	11.11	11.26	11.31
	1	13000+	11.33	11.61	11.86	11.92	11.23	11.43	11.68	11.77
	2	13800+	11.84	12.02	12.22	12.33	11.32	11.89	12.09	12.11
	3	14600+	12.19	12.35	12.40	12.41	11.90	12.16	12.26	12.29

หมายเหตุ + ในช่องอัตราเข้ารหัสรวมหมายถึง อัตราการเข้ารหัสรวมจะเพิ่มขึ้นอีก 200 bps สำหรับช่อง HPDR.1/2 เพิ่มขึ้นอีก 400 bps สำหรับช่อง HPDR.1/3 หรือ 1/4 และไม่เพิ่มขึ้นสำหรับช่องดั้งเดิม

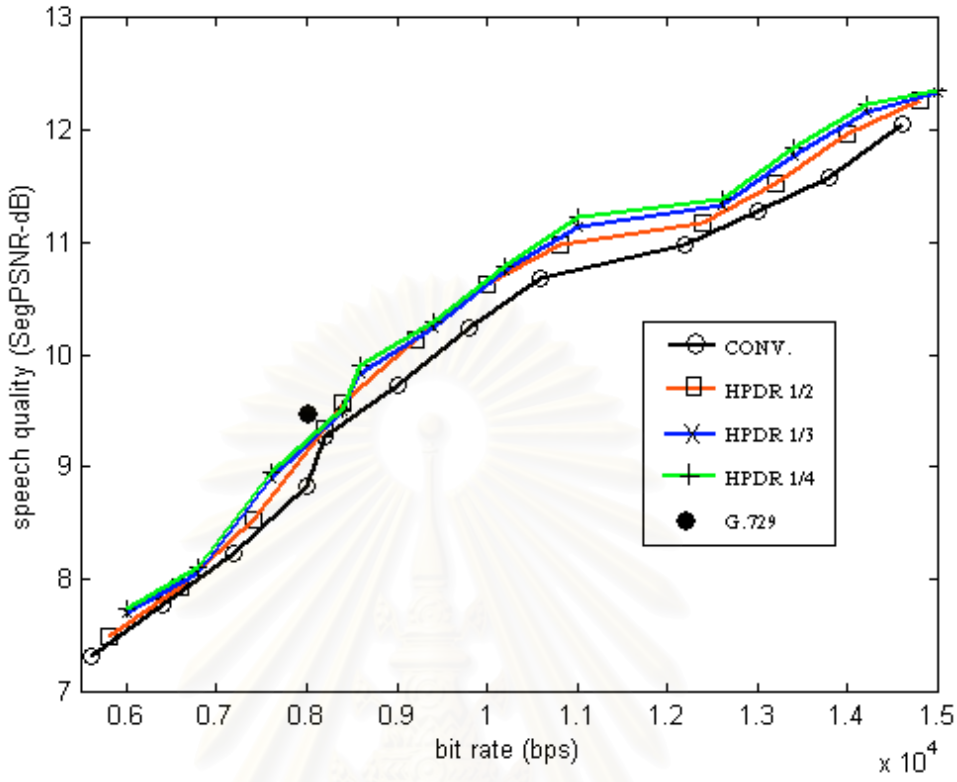
4.2.4 คุณภาพการเข้ารหัสโดยใช้การวัดค่าเชิงผู้ฟัง (ค่า MOS)

ผลการทดลองจากการเข้ารหัสและถอดรหัส โดยวัดค่าโดยใช้ผู้ฟัง 12 คนเป็นผู้ให้คะแนน โดยมีค่าเป็นคะแนนเต็มจาก 1 ถึง 5 นำเสนอเป็นค่าคุณภาพเสียงพูดเฉลี่ยในตารางที่ 4.9 พร้อมกับนำเสนอในรูปของกราฟเส้นเพื่อเปรียบเทียบผลในแง่ต่างๆ ดังรูปที่ 4.6

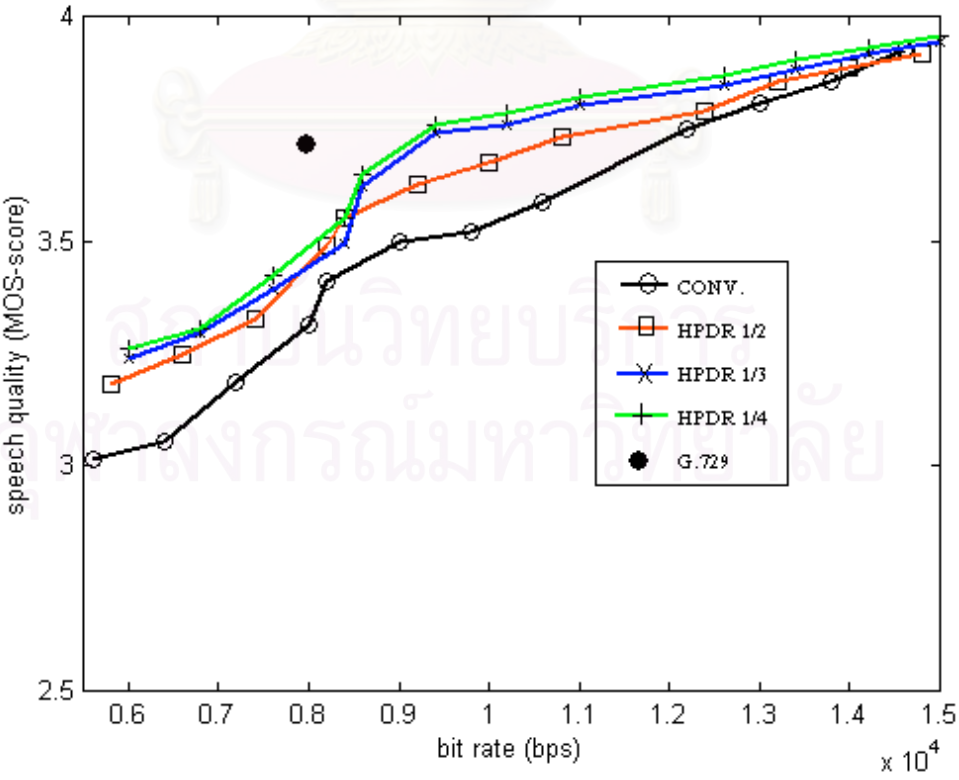
ตารางที่ 4.9 คุณภาพเสียงพูดที่ผ่านการเข้ารหัสโดยใช้ค่า MOS

อัตราเข้ารหัสหลัก (bps)	ขั้นการปรับระดับ	อัตราเข้ารหัสรวม (bps)	คุณภาพเสียงเพศชาย โดยเฉลี่ย (จุด)				คุณภาพเสียงเพศหญิง โดยเฉลี่ย (จุด)			
			ดั้งเดิม	HPDR.1/2 +200 bps	HPDR.1/3 +400 bps	HPDR.1/4 +400 bps	ดั้งเดิม	HPDR.1/2 +200 bps	HPDR.1/3 +400 bps	HPDR.1/4 +400 bps
5600	-	5600+	3.02	3.23	3.29	3.30	3.01	3.13	3.19	3.22
	1	6400+	3.08	3.26	3.32	3.32	3.03	3.24	3.27	3.29
	2	7200+	3.21	3.35	3.42	3.45	3.16	3.30	3.37	3.40
	3	8000+	3.35	3.56	3.51	3.59	3.28	3.42	3.48	3.51
8200	-	8200+	3.43	3.60	3.69	3.70	3.39	3.50	3.55	3.60
	1	9000+	3.50	3.66	3.78	3.81	3.50	3.59	3.70	3.71
	2	9800+	3.56	3.69	3.79	3.82	3.48	3.66	3.73	3.75
	3	10600+	3.61	3.75	3.82	3.84	3.56	3.71	3.78	3.80
12200	-	12200+	3.78	3.80	3.86	3.90	3.72	3.78	3.83	3.84
	1	13000+	3.81	3.87	3.91	3.93	3.80	3.84	3.85	3.88
	2	13800+	3.88	3.89	3.93	3.95	3.83	3.88	3.90	3.91
	3	14600+	3.93	3.94	3.99	3.99	3.92	3.89	3.90	3.92

หมายเหตุ + ในช่องอัตราเข้ารหัสรวมหมายถึง อัตราการเข้ารหัสรวมจะเพิ่มขึ้นอีก 200 bps สำหรับช่อง HPDR.1/2 เพิ่มขึ้นอีก 400 bps สำหรับช่อง HPDR.1/3 หรือ 1/4 และไม่เพิ่มขึ้นสำหรับช่องดั้งเดิม



รูปที่ 4.5 เปรียบเทียบค่า SegPSNRระหว่างHPDR ระดับต่างๆ



รูปที่ 4.6 เปรียบเทียบค่าเชิงผู้ฟังระหว่างHPDR ระดับต่างๆ

ผลการทดลอง กราฟในรูปที่ 4.5 และ 4.6 แสดงให้เห็นว่า การปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP ให้เหมาะสมกับเสียงพูดภาษาไทย โดยใช้เทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง (HPDR) ส่งผลให้คุณภาพเสียงพูดภาษาไทยดีขึ้น และสามารถเรียงลำดับเทคนิคที่ให้ผลการปรับปรุงที่ต่ำสุดไปยังผลที่ดีที่สุด ได้คือ HPDR ที่ความละเอียดระดับ 1/2 HPDR ที่ความละเอียดระดับ 1/3 และ HPDR ที่ความละเอียดระดับ 1/4 ตามลำดับ

เมื่อพิจารณาเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง ที่ความละเอียดระดับ 1/2 จะปรับปรุงคุณภาพเสียงภาษาไทยขึ้นประมาณ 0.24 dB หรือ 0.12 MOS score สำหรับเสียงพูดของเพศชาย และประมาณ 0.35 dB หรือ 0.10 MOS score สำหรับเสียงพูดของเพศหญิง โดยเฉลี่ยตลอดทุกอัตราการเข้ารหัส

เมื่อพิจารณาเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง ที่ความละเอียดระดับ 1/3 จะปรับปรุงคุณภาพเสียงภาษาไทยขึ้นประมาณ 0.40 dB หรือ 0.18 MOS score สำหรับเสียงพูดของเพศชาย และประมาณ 0.55 dB หรือ 0.16 MOS score สำหรับเสียงพูดของเพศหญิง โดยเฉลี่ยตลอดทุกอัตราการเข้ารหัส

เมื่อพิจารณาเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง ที่ความละเอียดระดับ 1/4 จะปรับปรุงคุณภาพเสียงภาษาไทยขึ้นประมาณ 0.45 dB หรือ 0.20 MOS score สำหรับเสียงพูดของเพศชาย และประมาณ 0.60 dB หรือ 0.18 MOS score สำหรับเสียงพูดของเพศหญิง โดยเฉลี่ยตลอดทุกอัตราการเข้ารหัส

จากผลการทดลอง แสดงให้เห็นว่าเทคนิคที่นำเสนอสามารถปรับปรุงคุณภาพเสียงพูดภาษาไทยได้ในระดับหนึ่ง

การที่สามารถวิเคราะห์ และส่งค่าพิตช์ที่ความละเอียดสูงใกล้เคียงกับค่าที่แท้จริงมากยิ่งขึ้น ทำให้การสังเคราะห์เสียงพูดกลับคืนมา มีความถูกต้องแม่นยำมากยิ่งขึ้น และการที่เสียงพูดภาษาไทยเป็นเสียงดนตรี คำแต่ละคำมีเสียงวรรณยุกต์ที่แตกต่างกัน แต่ละวรรณยุกต์จะมีการเปลี่ยนแปลงของความถี่มูลฐานที่ไม่เหมือนกัน [43 44 และ 45] หากวิเคราะห์และส่งค่าพิตช์ไม่ถูกต้องแม่นยำเพียงพอ ย่อมทำให้ไม่สามารถตรวจจับการเปลี่ยนแปลงของความถี่มูลฐานได้ถูกต้องเพียงพอ

เมื่อพิจารณาในแง่ของอัตราการบีบอัด สามารถแสดงเป็นจำนวนเท่าได้ดังตารางที่ 4.10 ช่วงอัตราบีบอัดเริ่มต้นที่ 4.27 เท่า ไปจนถึง 11.43 เท่า

ตารางที่ 4.10 อัตราบีบอัดของการเข้ารหัสที่นำเสนอ

อัตราเข้ารหัสหลัก (bps)	ขั้นการปรับระดับ	อัตราเข้ารหัสรวม (bps)	อัตราบีบอัด (เท่า)			
			ดั้งเดิม	HPDR.1/2 +200 bps	HPDR.1/3 +400 bps	HPDR.1/4 +400 bps
5600	-	5600+	11.43	11.03	10.67	10.67
	1	6400+	10.00	9.70	9.41	9.41
	2	7200+	8.88	8.65	8.42	8.42
	3	8000+	8.00	7.80	7.62	7.62
8200	-	8200+	7.80	7.62	7.44	7.44
	1	9000+	7.11	6.96	6.81	6.81
	2	9800+	6.53	6.40	6.27	6.27
	3	10600+	6.04	5.93	5.82	5.82
12200	-	12200+	5.25	5.16	5.08	5.08
	1	13000+	4.92	4.85	4.78	4.78
	2	13800+	4.64	4.57	4.51	4.51
	3	14600+	4.38	4.32	4.27	4.27

หมายเหตุ + ในช่องอัตราเข้ารหัสรวมหมายถึง อัตราการเข้ารหัสรวมจะเพิ่มขึ้นอีก 200 bps สำหรับช่อง HPDR.1/2 เพิ่มขึ้นอีก 400 bps สำหรับช่อง HPDR.1/3 หรือ 1/4 และไม่เพิ่มขึ้นสำหรับช่องดั้งเดิม

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอการเข้ารหัสเสียงพูดภาษาไทยที่อยู่บนพื้นฐานของการเข้ารหัสเสียงพูดมาตรฐาน MPEG-4 เริ่มจากการจำลองตัวเข้ารหัสและตัวถอดรหัส MP-CELP ที่สอดคล้องตามข้อกำหนดของมาตรฐานการเข้ารหัสเสียงพูดธรรมชาติ MPEG-4 แล้วนำมาทดสอบกับเสียงพูดภาษาอังกฤษ และเสียงพูดภาษาไทย ผลการทดลองแสดงให้เห็นว่า เสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัสด้วยการเข้ารหัสที่จำลองขึ้น มีคุณภาพด้อยกว่าเสียงพูดภาษาอังกฤษ ไม่ว่าจะเป็นค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนเป็นส่วน หรือค่าเชิงผู้ฟัง เพื่อแก้ปัญหาจุดนี้ วิทยานิพนธ์จึงได้นำเสนอการปรับปรุงการเข้ารหัสเสียงพูดโดยวิธี MP-CELP กับเสียงพูดภาษาไทย ด้วยเทคนิคการวิเคราะห์ค่าพิตช์ดีเลย์ด้วยความละเอียดสูง นำมาทดสอบกับเสียงพูดภาษาไทยด้วยฐานข้อมูลเสียงพูดที่บันทึกไว้ในห้องปฏิบัติการวิจัยกรรมวิธีสัญญาณดิจิทัล ที่ประกอบด้วยเสียงของผู้พูดอาสาสมัครเพศชาย และเพศหญิง และมีความหลากหลายในวัยของผู้พูด แล้วนำมาประเมินประสิทธิภาพ พบว่าเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัสโดยปรับปรุงด้วยเทคนิคที่นำเสนอ มีคุณภาพดีขึ้นในระดับที่ใกล้เคียงกับเสียงพูดภาษาอังกฤษ ด้วยอัตราการเข้ารหัสที่สูงขึ้นเล็กน้อยคือ 200-400 bps จากเดิม 5,600-14,600 bps เป็น 5,800-15,000 bps สำหรับส่งข้อมูลเศษส่วนพิตช์ เมื่อเปรียบเทียบเทคนิคการวิเคราะห์พิตช์ที่ความละเอียดระดับต่างๆ พบว่าคุณภาพเสียงพูดที่ผ่านการเข้ารหัสที่ความละเอียดสูงสุดคือ 1/4 จะใกล้เคียงกับที่ความละเอียดสูงสุดคือ 1/3 แต่มีความซับซ้อนมากกว่าถึง 1.5 เท่า สำหรับการวิเคราะห์เศษส่วนพิตช์ที่เพิ่มขึ้น ฉะนั้นเทคนิคการวิเคราะห์พิตช์ที่ความละเอียดระดับ 1/3 จึงมีความเหมาะสมสำหรับการใช้งานกับเสียงพูดภาษาไทยมากที่สุด ในแง่ของเวลาประวิง (delay time) การเข้ารหัสที่นำเสนอจะมีเวลาประวิงที่ใกล้เคียงกับการเข้ารหัสมาตรฐาน ITU G.729 ที่ใช้อัลกอริทึม CS-ACELP คือ 15 มิลลิ-วินาที โดยเป็นเวลาประวิงเชิงอัลกอริทึม ไม่รวมถึงเวลาประวิงในการส่งผ่านข่ายเชื่อมโยงสื่อสารหรือเวลาประวิงการมัลติเพลกซ์ เมื่อมีการรวมข้อมูลเสียงเข้ากับข้อมูลชนิดอื่นเช่น ภาพนิ่ง วิดีโอ หรือฐานข้อมูล เป็นต้น

5.2 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

1. ในขั้นตอนการเก็บข้อมูลเสียงที่นำมาทดสอบกับตัวเข้ารหัสและถอดรหัสที่จำลองขึ้นด้วยอุปกรณ์ในห้องปฏิบัติการวิจัย ถ้าเป็นเสียงพูดที่ดังเกินไป คือเกิดการขริบของส่วนยอดของ

สัญญาณ จะทำให้คุณภาพเสียงที่ได้จากการบีบอัดไม่ดี เนื่องจากสัญญาณช่วงที่ถูกขริบไปนั้นจะส่งผลให้ผลตอบสนองเชิงความถี่ของสัญญาณที่ควรจะเป็นเปลี่ยนแปลงไปมาก ในขณะที่การบีบย่อที่ใช้ในวิทยานิพนธ์นี้ ใช้ LPC ซึ่งเป็นการวิเคราะห์เชิงความถี่ของสัญญาณ ซึ่งจะรับผลกระทบจากข้อบกพร่องส่วนนี้โดยตรง ในการเก็บข้อมูลเสียงที่นำมาทดสอบที่มีขนาดใหญ่เกินขีดจำกัดของอุปกรณ์ จะมีผลให้คุณภาพเสียงที่ผ่านการบีบอัดไม่ดี เพื่อลดปัญหานี้ จึงควรระมัดระวังในการเก็บฐานข้อมูลที่นำมาใช้ทดสอบ

2. การเข้ารหัสที่นำเสนอนี้ ได้นำมาทดสอบกับเสียงพูดภาษาไทยเท่านั้น ซึ่งในบรรดาภาษาที่มีในโลกนี้ ยังมีภาษาอื่นที่เป็นภาษาดนตรีเช่นเดียวกับภาษาไทย การประยุกต์ใช้เทคนิคที่นำเสนอกับเสียงพูดภาษาอื่นๆ จะเป็นแนวทางที่สามารถดำเนินงานวิจัยต่อไปได้

3. การเข้ารหัสที่นำเสนอนี้ ใช้เสียงพูดขาเข้าแถบแคบ คือมีแบนด์วิดท์ 20-3400 Hz หากต้องการให้บริการด้วยคุณภาพเสียงที่สูงขึ้น สามารถปรับปรุงการเข้ารหัสที่นำเสนอให้ใช้เสียงพูดแถบกว้างเป็นสัญญาณขาเข้าแทน โดยปรับเปลี่ยนขนาดของเฟรมในการประมวลผลขั้นต่างๆ ให้เหมาะสม รวมถึงค่าพารามิเตอร์บางอย่าง เพื่อให้สามารถรองรับการทำงานที่ต้องการเสียงที่คุณภาพสูงขึ้นได้อย่างมี ประสิทธิภาพ

รายการอ้างอิง

1. Bristow, G. Electronic speech synthesis New York: McGraw-Hill, 1984.
2. Nomura, T. ISO/IEC JTC 1/SC 29/WG11/M1509 Proposal of Compression Algorithm with Rate Control for MPEG-4/Audio Core Experiments New York: ISO/IEC, 1996.
3. Grill, B., Edler, B., and Funken, R. ISO/IEC JTC 1/SC 29/WG11/N2203 Information Technology – Very low bitrate Audio –Visual coding New York: ISO/IEC, 1996.
4. Nomura, T., Iwadare, M., Serizawa, M., and Ozawa, K. A Bitrate and Bandwidth Scalable CELP Coder. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (May 1998): 341-344.
5. Colomes, C., Jacobson, C., and Scheirer, E. ISO/IEC JTC 1/SC 29/WG11/N2276 Report on the MPEG-4 audio NADIB verification tests New York: ISO/IEC, 1998.
6. Scheirer, E., Kim, S. W., and Dietz, M. ISO/IEC JTC 1/SC 29/WG11/N2425 MPEG-4 Audio verification test results: Audio on Internet New York: ISO/IEC, 1998.
7. Lynch, T. J. Data compression techniques and applications New York: Van Nostrand Reinhold Company, 1985.
8. Schroder, G., and Sherif, M. H. The Road to G.729: ITU 8-kb/s Speech Coding Algorithm with Wireline Quality. IEEE Transactions on Communications Vol 35 (September 1997): 48-54.
9. International Telecommunication Union CCITT recommendation G.728, coding of speech at 16 kbit/s using low-delay code-excited linear prediction Geneva, 1992.
10. Chen, J. A Low-Delay CELP Codec for the CCITT 16 kb/s Speech Coding Standard. IEEE Transactions on Communications Vol 10 (June 1992): 830-849.
11. Evci, C. Speech Codec Aspects for Third Generation Mobile Systems. IEEE Transactions on Vehicular Vol 1 (May 1992): 172-175.
12. Ojanpera, T., and Prasad, R. Wideband CDMA for Third Generation Mobile Communications Norwood: Artech House, 1998.
13. Kondoz, A. M. Digital Speech Coding for Low Bit Rate Communication Systems New York: John Wiley & Sons, 1994.
14. Salami, R. A Toll Quality 8 Kb/s Speech Codec for the Personal Communication System (PCS). IEEE Transactions on Vehicular Vol 43 (August 1994): 808-816.

15. International Telecommunication Union CCITT recommendation G.729, Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP) Geneva: n.p., 1996.

16. Kataoka, A., Moriya, T., and Hayashi, S. An 8-kb/s Conjugate Structure CELP (CS-ACELP) Speech coder. IEEE Transactions on Speech and Audio Processing Vol 4 (November 1996): 401-411.

17. Juan, L., Bigin, L., and Qiuliang F. An 8-kb/s conjugate-structure algebraic CELP (CS-ACELP) speech coding. IEEE Transactions on Signal Processing Vol 2 (October 1998): 1729-1732.

18. Deller, J. R. Jr., Proakis, J. G., and Hansen, J. H. L. Discrete-time processing of speech signals New York: Macmillan, 1993.

19. Donoho, D. L. Unconditional bases are optimal bases for data compression and for statistical estimation Standford University, 1992.

20. Ounnapirak, C. Speech Compression Using Wavelet Transform and LPC Vector Quantization Master Thesis, Electrical Engineering, Engineering, Chulalongkorn University, 1995. (In Thai)

21. Ounnapirak, C., and Jitapunkul, S. Speech Compression Using Wavelets Packets Based on CELP Algorithm. The 18th Conference of Electrical Engineering Vol 1 (November 1995): 918-920. (In Thai)

22. Laflamme, C. 16 kbps Wideband Speech Coding Technique Based on Algebraic CELP. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (April 1991): 13-16.

23. Cox, R. V. Three New Speech Coders From The ITU Cover A Range Of Applications. IEEE Transactions on Communications Vol 35 (September 1997): 40-47.

24. Spanias, A. S. Speech Coding: A Tutorial Review. Proceedings of IEEE Vol 82 (October 1994): 1541-1582.

25. Salami, R. A Toll Quality 8 Kb/s Speech Codec for the Personal Communication System (PCS). IEEE Transactions on Vehicular Vol 43 (August 1994): 808-816.

26. Hellwig, K. Speech Codec for the European Mobile Radio System. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (April 1988): 227-230.

27. Delprat, M., and Evci, C. C. Advance Speech Transmission Techniques for GSM and Beyond. IEEE Transactions on ICT Vol 1 (April 1996): 208-212.

28. Chompun, S., Jitapunkul, S., Tancharoen, D., and Srithanasan, T. Thai Speech Compression Using CS-ACELP Coder Based on ITU G.729 Standard. Proceedings of SNLP Vol 1 (May 2000): 263-268.

29. Ozawa, K. MP-CELP Speech Coding Based on Multi-pulse Vector Quantization and Fast Search. Proceeding of IEICE J79-A (1996): 1655-1663. (In Japanese)

30. Ozawa, K., and Serizawa, M. High Quality Multi-pulse Based CELP Speech Coding at 6.4 kb/s and its Subjective Evaluation. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (May 1998): 529-532.

31. Taumi, S. Low-Delay CELP with Multi-pulse VQ and Fast Search for GSM EFR. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (May 1996): 562-565.

32. Taumi, S. Low-Delay CELP with Multi-pulse VQ and Fast Search for GSM EFR. IEEE Transactions on Acousics, Speech and Signal Processing Vol 1 (May 1996): 562-565.

33. Makhoul, J. Spectral analysis of speech by linear prediction. IEEE Transactions on Audio and Electroacoustics AU-21 (June 1973): 140-148.

34. Kataoka, A. ITU-T Quality 8-kbits/s Standard Speech Codec for Personal Communication Services. Proceeding of International Conference Universal Personal Communication (November 1995): 818-822.

35. Gray, R. M., Vector quantization. IEEE ASSP Magazine Vol 1 (April1984): 4-28.

36. Hussain, Y., and Farvardin, N. Adaptive block transform coding of speech based on LPC vector quantization. IEEE Transactions on Signal Processing Vol 39 (December 1991): 2611-2619.

37. Chen, S. H., and Wang, Y. R. Vector quantization of pitch information in Mandarin speech. IEEE Transactions on Communications Vol 38 (September 1990): 1317-1320.

38. Dedes, I. S., Vaman, D. R., and Chakravarthy, C. V. Variable bit rate adaptive predictive coder. IEEE Transactions on Signal Processing Vol 40 (1992): 511-517.

39. Galand, C. R., Menez, J. E., and Rosso, M. M. Adaptive code excited predictive coding. IEEE Transactions on Signal Processing Vol 40 (1992): 1317-1326.

40. Maksym, J. N. Real-time Pitch extraction by adaptive prediction of the speech waveform. IEEE Transactions on Audio and Electroacoustics AU-21 (June 1973): 149-153.

41. Moorer, J. A. The optimum comb method of pitch period analysis of continuous digitized speech. IEEE Transactions on Acousics, Speech and Signal Processing ASSP-22 (October 1974): 330-338.

42. Haykin, S. Adaptive filter theory Englewood Cliffs, NJ: Prentice-Hall, 1986.

43. Oppenheim, A. V., and Schafer, R. W. Discrete-time signal processing Englewood Cliffs, NJ: Prentice-Hall, 1989.

44. Marcario, R. C. V. Cellular Radio Principles and Design London: Macmillan, 1993.

45. Jones, D. English pronouncing dictionary Cambridge: Cambridge University Press, 1997.

46. Thathong, U., Jitapunkul, S., and Ahkuputra, V. Classification of Thai Consonants Naming Using Thai Tone. Proceedings of ICSLP Vol 1 (October 2000): 105-108.

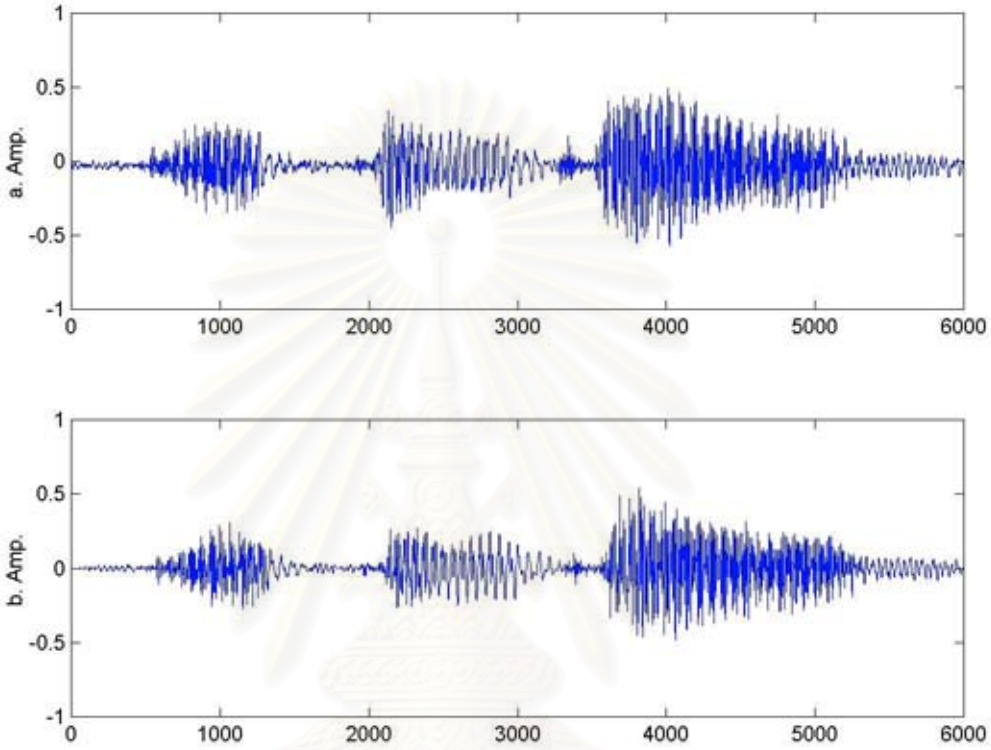
47. Luksaneeyanawin, S. Linguistics Research and Thai Speech Technology. The 5th International Conference on Thai Studies Vol 1 (1993): 51-66.

48. Chompun, S., Jitapunkul, S., and Tancharoen, D. Novel Technique for Tonal Language Speech Compression based on a Bitrate Scalable MP-CELP Coder. IEEE transactions on ITCC Vol 1 (April 2001): 461-464.

ภาคผนวก

ภาคผนวก ก

ตัวอย่างสัญญาณเสียงพูดภาษาอังกฤษที่ผ่านการเข้ารหัสและถอดรหัส ด้วย CS-ACELP (G.729)

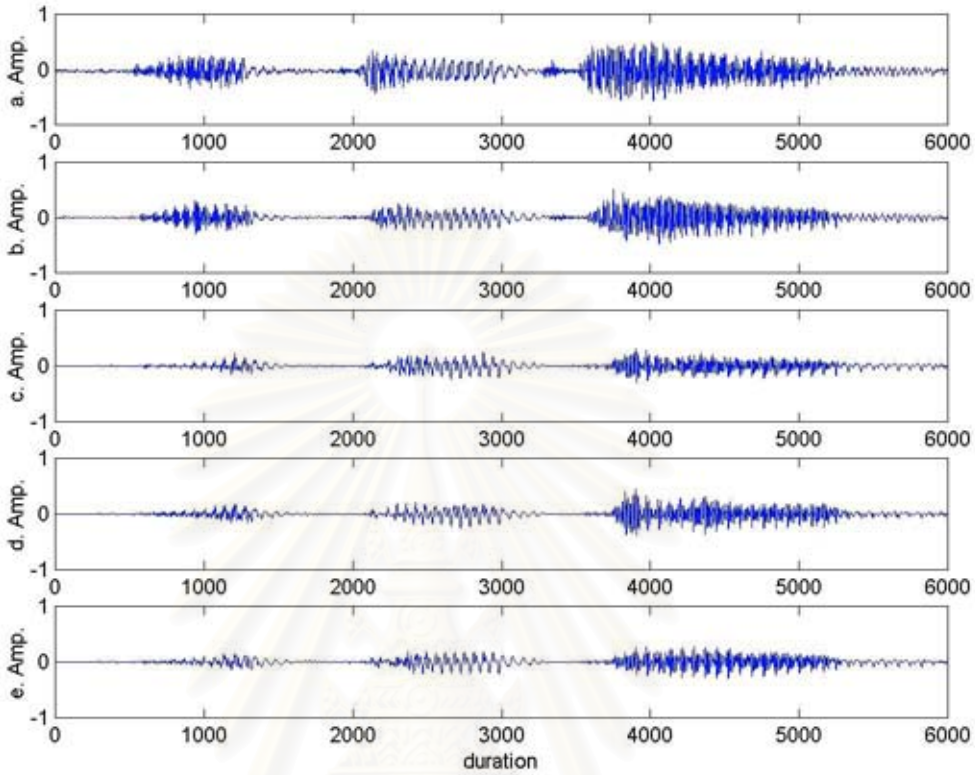


รูปที่ ก.1 สัญญาณเสียงคำว่า I-think-we

a. สัญญาณก่อนเข้ารหัส

b. สัญญาณที่ผ่านการเข้ารหัสด้วย CS-ACELP (G.729)

ตัวอย่างสัญญาณเสียงพูดภาษาอังกฤษที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ



รูปที่ ก.2 สัญญาณเสียงคำว่า I-think-we

c. สัญญาณก่อนเข้ารหัส

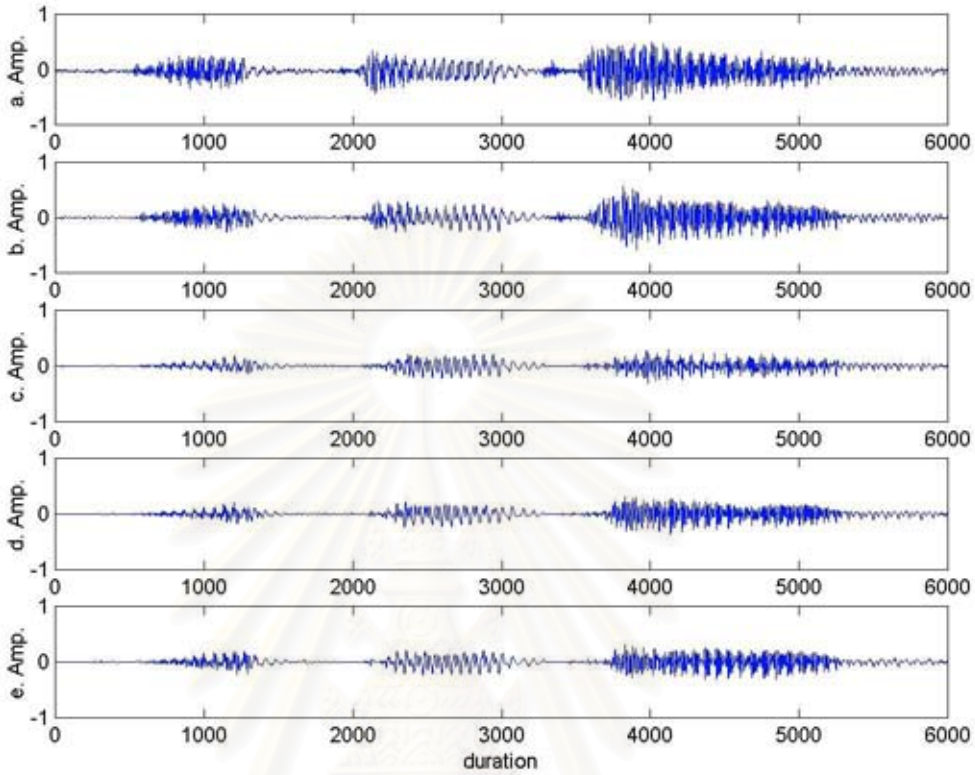
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 3 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 2 ขั้น

f. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 1 ขั้น

g. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ไม่ปรับระดับ

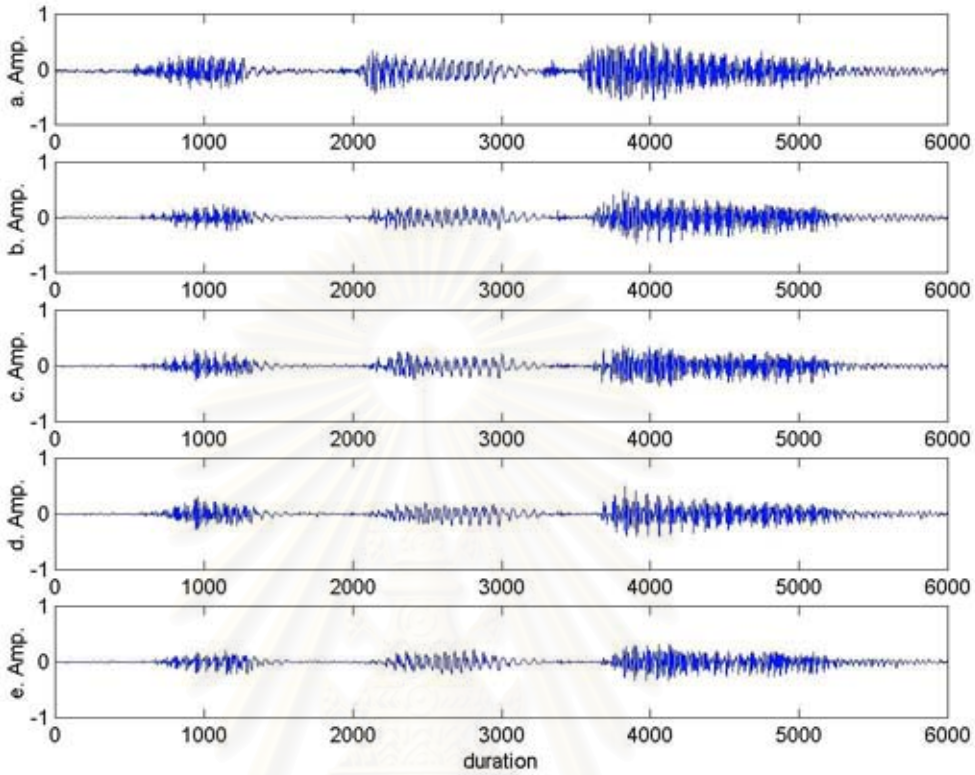
ตัวอย่างสัญญาณเสียงพูดภาษาอังกฤษที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ก.3 สัญญาณเสียงคำว่า I-think-we

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาอังกฤษที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)

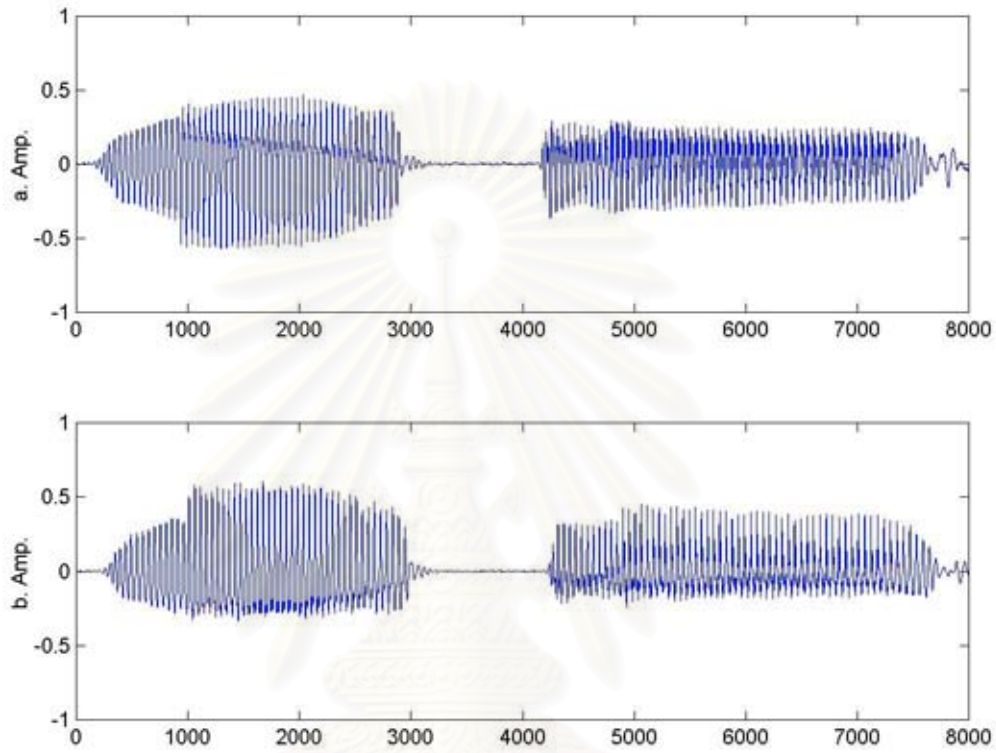


รูปที่ ก.4 สัญญาณเสียงคำว่า I-think-we

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ไม่ปรับระดับ

ภาคผนวก ข

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วย CS-ACELP (G.729)

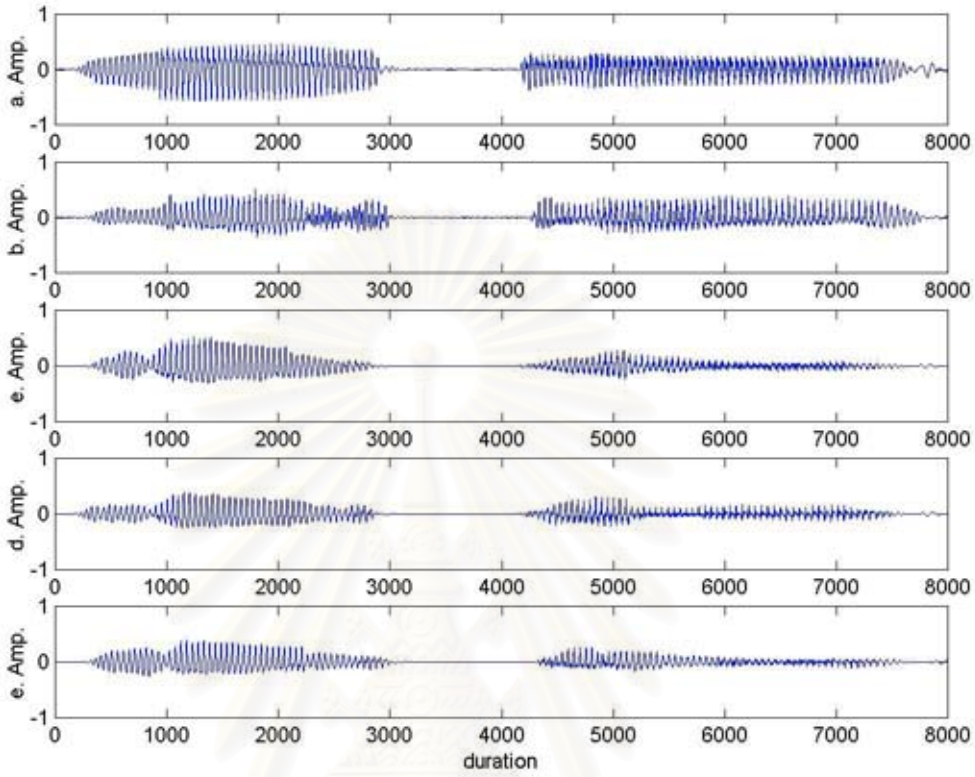


รูปที่ ข.1 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

b. สัญญาณที่ผ่านการเข้ารหัสด้วย CS-ACELP (G.729)

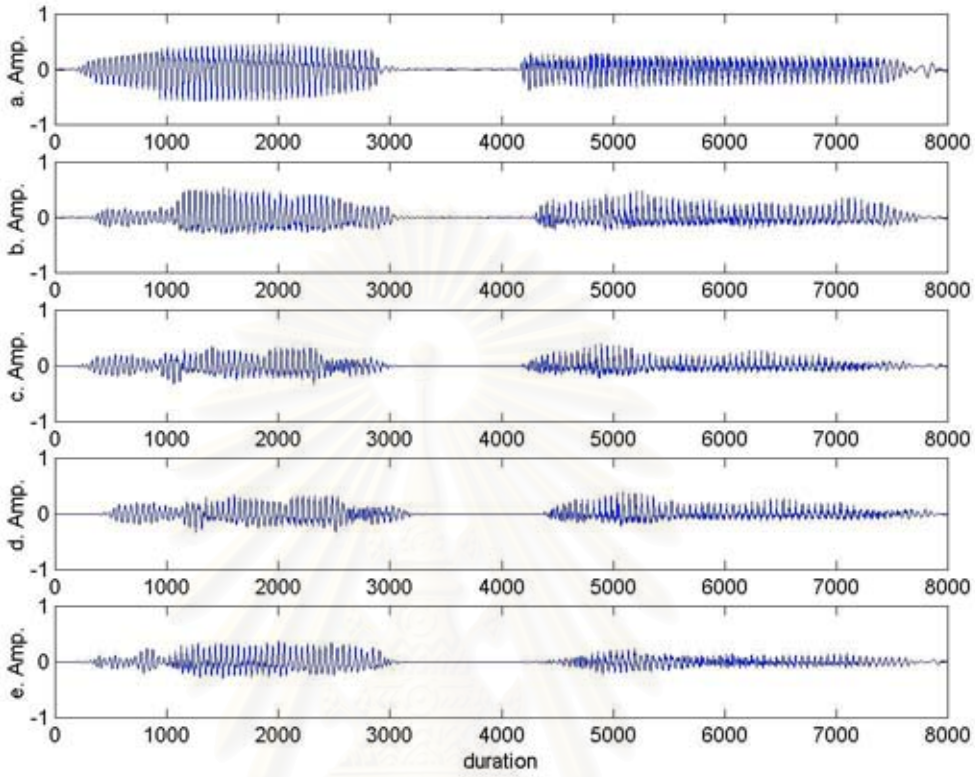
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ



รูปที่ ข.2 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ ไม่ปรับระดับ

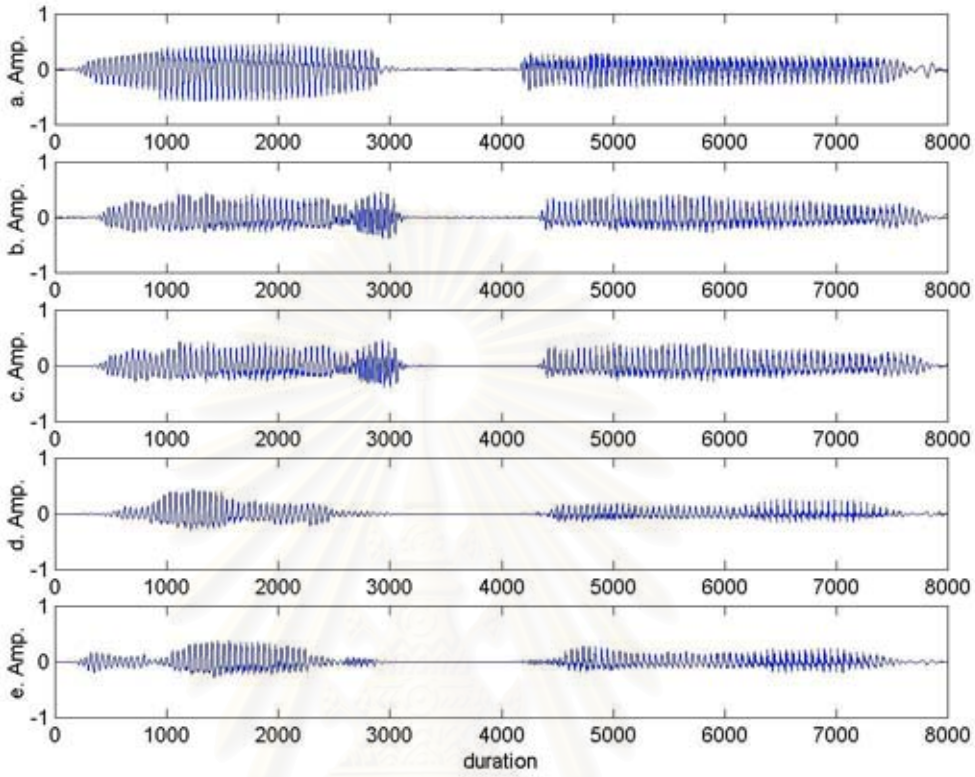
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.3 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ ไม่ปรับระดับ

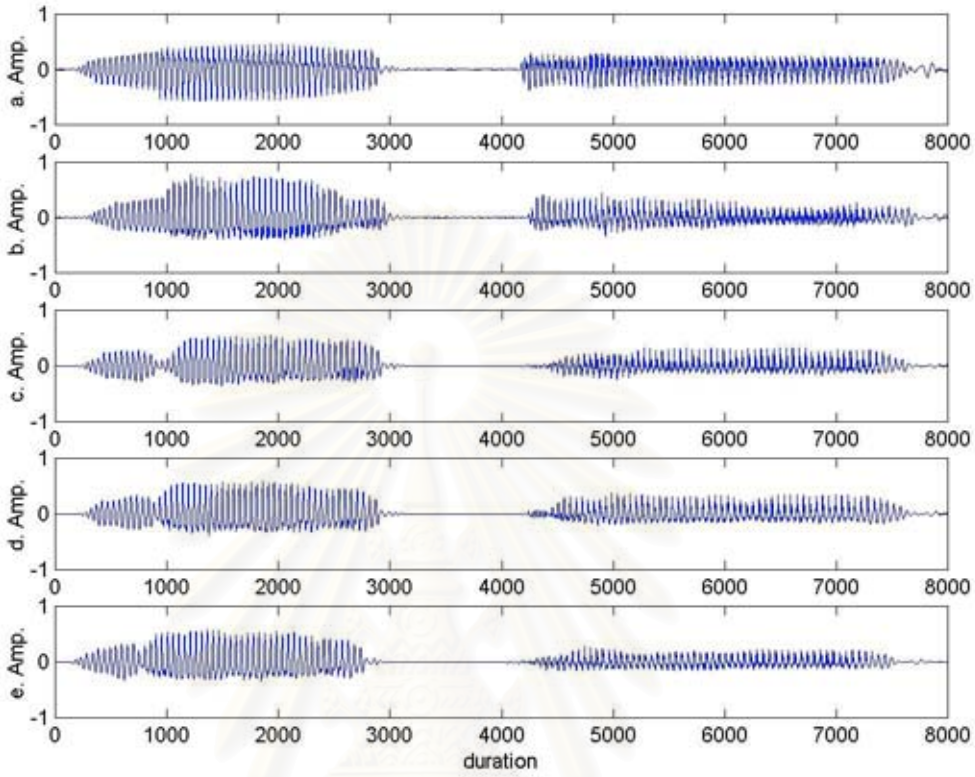
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.4 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ ไม่ปรับระดับ

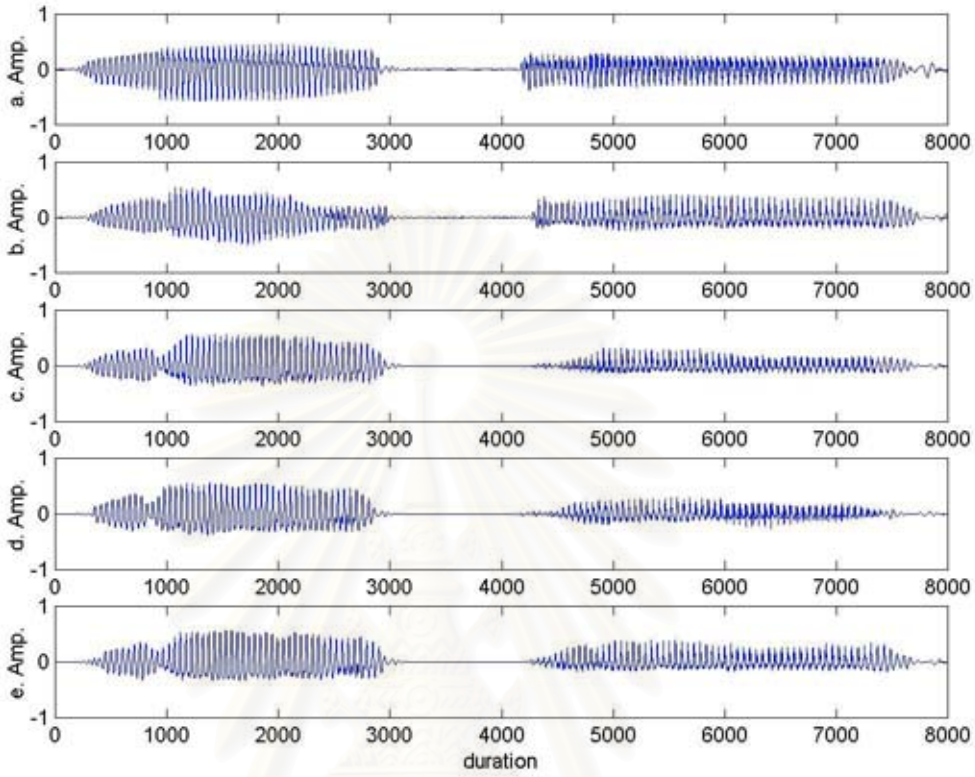
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.5 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/2 ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.6 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

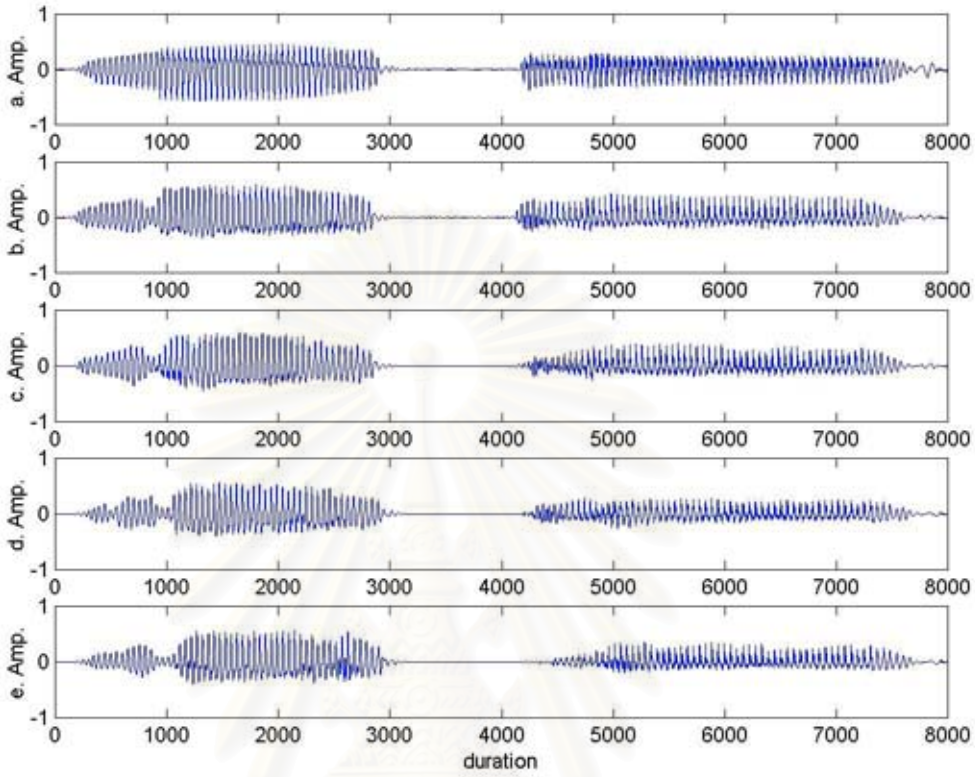
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 3 ขั้น

c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 2 ขั้น

d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 1 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/2 ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.7 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

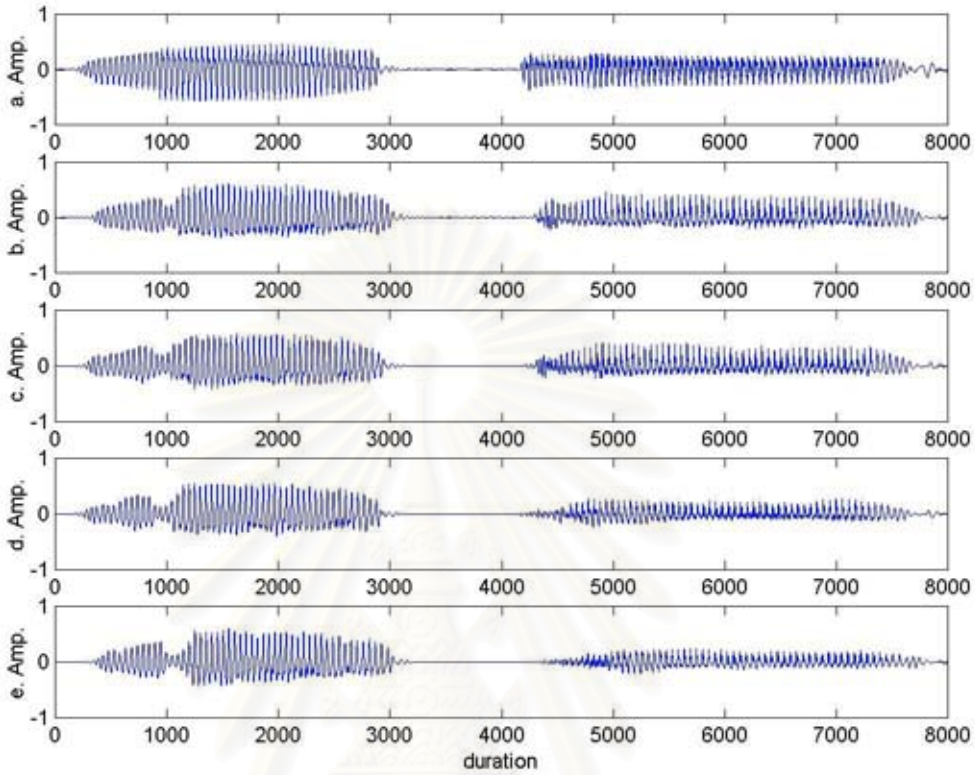
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 3 ขั้น

c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 2 ขั้น

d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/2 ปรับระดับ 1 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/2 ไม่ปรับระดับ

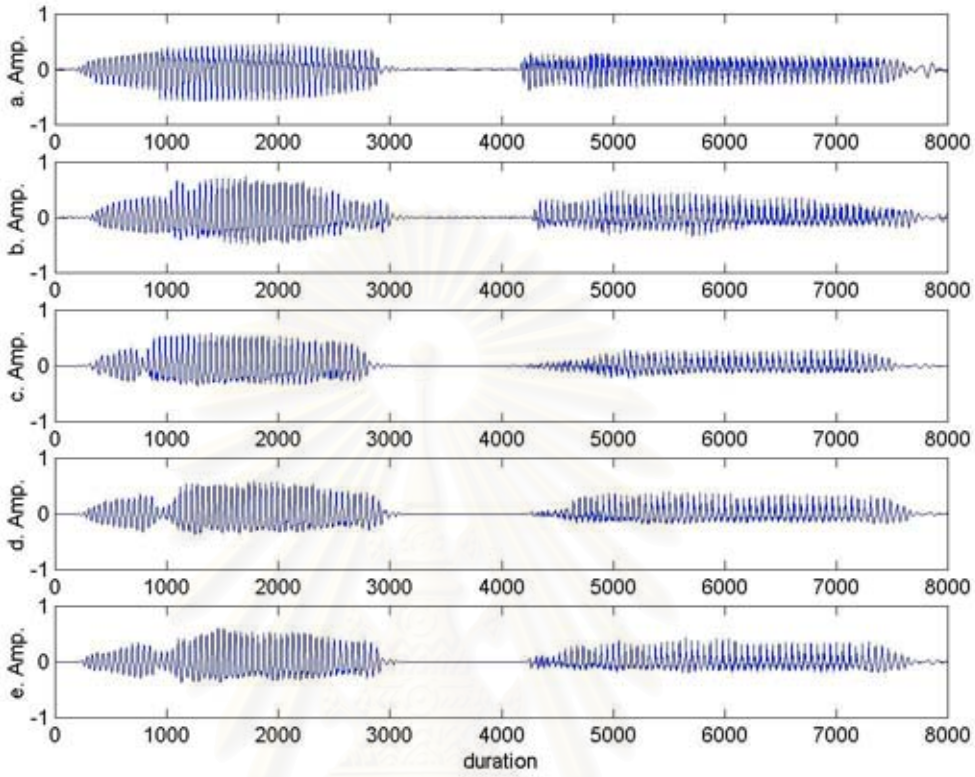
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.8 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/3 ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.9 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

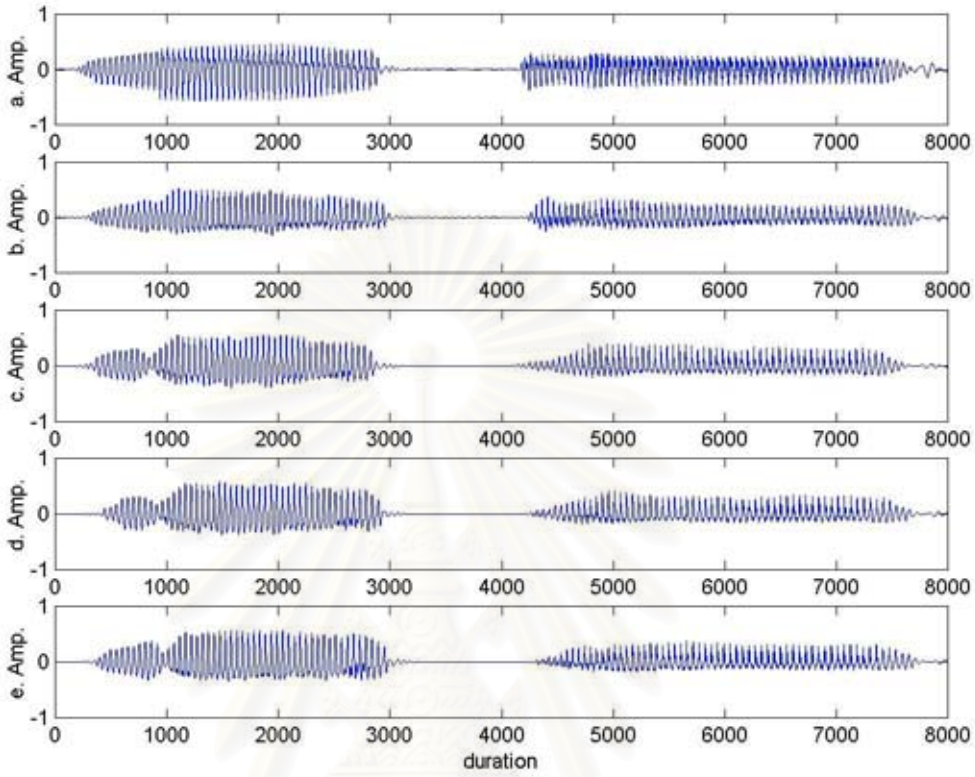
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 3 ขั้น

c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 2 ขั้น

d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 1 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/3 ไม่ปรับระดับ

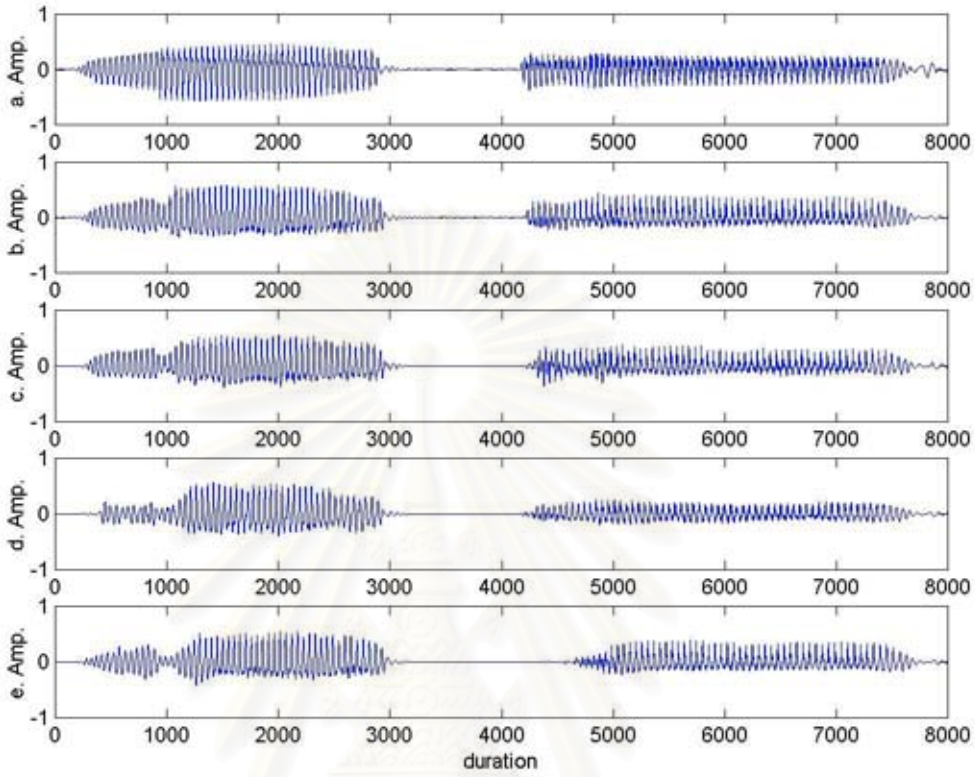
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.10 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/3 ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/3 ไม่ปรับระดับ

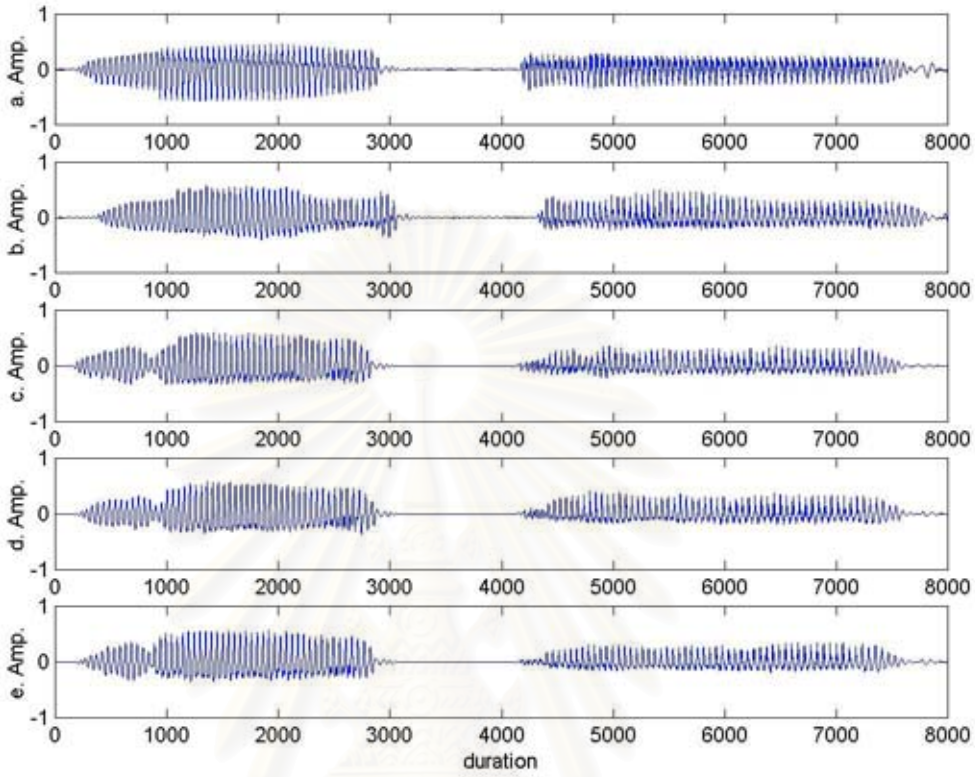
ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.11 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 3 ขั้น
c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 2 ขั้น
d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 1 ขั้น
e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 1 พัลส์ เพิ่ม HPDR 1/4 ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.12 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

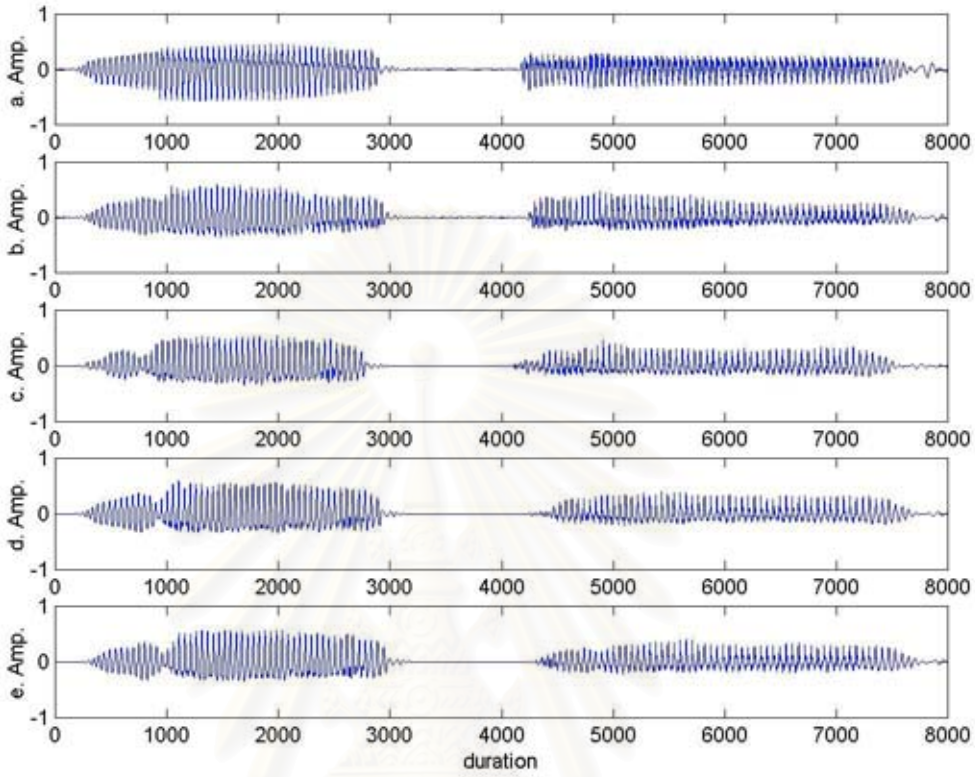
b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 3 ขั้น

c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 2 ขั้น

d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 1 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 5 พัลส์ เพิ่ม HPDR 1/4 ไม่ปรับระดับ

ตัวอย่างสัญญาณเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัส ด้วยการเข้ารหัสที่นำเสนอ (ต่อ)



รูปที่ ข.13 สัญญาณเสียงคำว่า -แม่-บอก- (/mxx2/b@@k1/)

a. สัญญาณก่อนเข้ารหัส

b. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 3 ขั้น

c. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 2 ขั้น

d. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/4 ปรับระดับ 1 ขั้น

e. สัญญาณที่ผ่านการเข้ารหัสด้วย MP-CELP แบบ 10 พัลส์ เพิ่ม HPDR 1/4 ไม่ปรับระดับ

ภาคผนวก ค

บทความทางวิชาการของผู้วิจัยที่ได้รับการตีพิมพ์แล้ว

1 Thai Speech Compression Using CS-ACELP Coder Based on ITU G.729 Standard.

- Proceedings of Symposium on Natural Language Processing.

2 Novel Technique for Tonal Language Speech Compression based on a Bitrate Scalable MP-CELP Coder.

- Proceedings of International Conference on Information Technology: Coding and Computing, IEEE Computer Society.

THAI SPEECH COMPRESSION USING CS-ACELP CODER BASED ON ITU G.729 STANDARD.

S. CHOMPUN, S. JITAPUNKUL, D. TANCHAROEN AND T. SRITHANASAN

Digital Signal Processing Research Laboratory, Department of Electrical Engineering,

Faculty of Engineering, Chulalongkorn University, Bangkok 10330, THAILAND

E-Mail: b0597021@student.chula.ac.th, jsomchai@chula.ac.th

This paper presents the comparison results of speech quality that is encoded and decoded by CS-ACELP coder according to ITU-G.729 standard. The purpose is to test the performance of CS-ACELP coder between Thai speech and English speech. The paper used 2 coding methods; 1) CS-ACELP coder without Voice Activity Detection and 2) CS-ACELP coder with Voice Activity Detection. The objective test was used to measure the speech quality for each case. The results show that both methods give Thai speech quality mostly below than English speech quality, as for methods comparison; both Thai and English, method 2) gives speech quality better than method 1). Eventually, we modified the coder by increasing the order of LP analysis to improve the Thai speech quality.

Key words: Thai speech, speech coding, CS-ACELP, ITU-G.729.

INTRODUCTION

Nowadays, the digital communications are widely developed. The information; audio, images, video or data can be transmitted passthrough wire or wireless network channels. Simultaneously, the number of users to access these networks increases rapidly. Consequently, channel capacity has to be increased, signal compression aims to perform this. As for speech, speech coding was created almost 60 years ago, and improved from then until now (Cox 1997).

The International Telecommunications Union-Telecommunications Sector (ITU-T) has already standardized 64-kb/s μ/A-law PCM, 32-kb/s ADPCM, and 16-kb/s low-delay code-excited linear prediction (LD-CELP). The next step in the progression is an 8-kb/s speech coding algorithm. Since the three existing standards all provide high-quality and short-delay coding, the main requirement for the 8-kb/s algorithm is also initially high-quality and short-delay coding (less than 5 ms)(Kataoka, Moriya, Hayasshi 1996; ISO/JTC 1998).

To achieve high quality and short-delay coding at 8-kb/s, the backward adaptation technique was performed. Many coders used the linear predictive coding (LPC) predictor in a backward-adaptive manner by performing LPC analysis on previously quantized speech. Since the reconstructed signal is available in both the encoder and decoder, this approach does not require that the LPC coefficients be sent to the decoder. However, although this technique is useful for 16-kb/s LD-CELP, an 8-kb/s coder was not found to give high quality when using only backward PLC analysis without pitch prediction. In 1995, CS-ACELP coder was developed and standardized as 8-kb/s G.729.

To verify this circumstance, this paper will study and compare the performance of CS-ACELP coder according to ITU G.729 between Thai and English speeches. The paper will use 2 coding methods: the first is CS-ACELP coder without Voice Activity Detection (VAD) and the second is CS-ACELP coder with Voice Activity Detection. The objective test as segmental signal to noise ratio (segSNR) will be selected to measure the speech quality for each case.

However, since Thai language is a tonal language. So the use of CS-ACELP coder following to ITU G.729 to compress Thai speech will be not guarantee the same quality as English language.

Later, the paper try to modify some parameters in CS-ACELP coder in order to improve its efficient for Thai speech compression. In the 10th linear prediction analysis and filtering, the paper will increase the order to 12 and 14 (Deller, Jr. 1993).

1. CS-ACELP ALGORITHM

The CS-ACELP coder is based on the code-excited linear predictive (CELP) coding model. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples per

second. For every 10 ms frame, the speech signal is analyzed to extracted the parameters of the CELP model (linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains). These parameters are encoded and transmitted. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering this excitation through the short-term synthesis filter based on a 10th order liner prediction filter and the long-term or pitch synthesis filter implemented using adaptive-codebook approach. After computing the reconstructed speech, it is further enhanced by a postfilter (Schroder, Sherif 1997).

The encoding principle is shown in figure 1. The input signal is high-pass filtered and scaled in the pre-processing block. The pre-processing signal serves as the input signal for all subsequent analysis. LP analysis is done once per 10 ms frame to compute the LP coefficients. These coefficients are converted to line spectrum pairs (LSP) and quantized using predictive two-stage vector quantization with 18 bits. The excitation signal is chosen by using an analysis-by-synthesis search procedure in which the error between original and reconstructed speech is minimized according to a perceptually weighted distortion measure. This is done by filtering the error signal with a perceptual weighting filter, whose coefficients are derived from the unquantized LP filter. The amount of perceptual weighting is made adaptive to improve the performance for input signals with a flat frequency-response.

The decoder principle is shown in figure 2. First, the parameters indices are extracted from the received bitstream. These indices are decoded to obtain the coder parameters corresponding to a 10 ms speech frame. These parameters are the LSP coefficients, the 2 fractional pitch delays, the 2 fixed-codebook vectors, and the 2 sets of adaptive and fixed-codebook gains. The LSP coefficients are interpolated and converted to LP coefficients for each subframe. Then, for each 5 ms subframe, the excitation is constructed by adding the adaptive and fixed-codebook vectors scaled by their respective gains, the speech is reconstructed by filtering the excitation through the LP synthesis filter, finally, the reconstructed speech signal is passed through a post-processing stage, which includes an adaptive postfilter based on the long-term and short-term synthesis filter, followed by a high-pass filter and scaling operation.

Voice Activity Detection is in the pre-processing part to decide the input speech frame as voiced or unvoiced speech. Consequently, the unvoiced speech mode neglects the adaptive codebook quantization part because no periodicity is needed while the voiced speech mode still proceeds both fixed and adaptive quantization part.

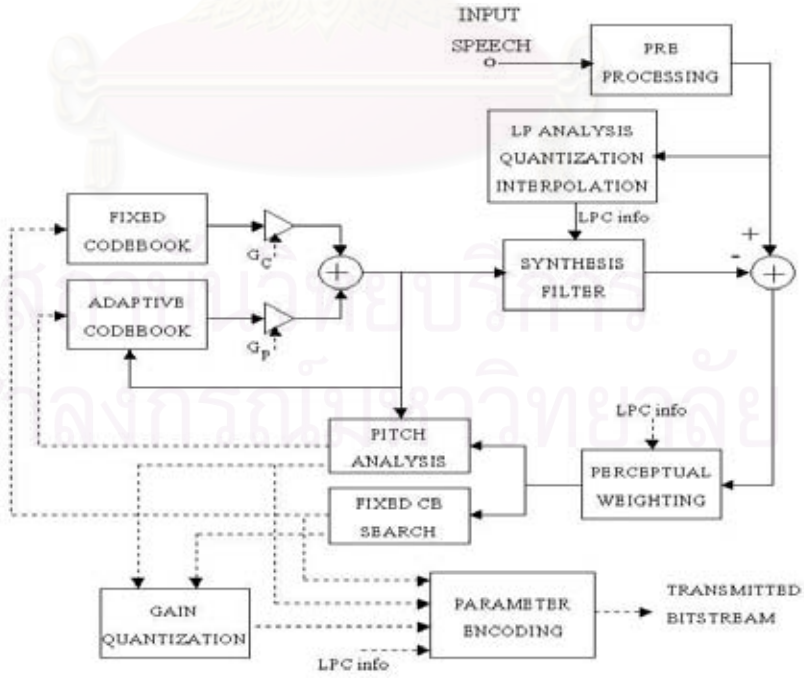


FIGURE 1. Block diagram of G.729 Encoder.

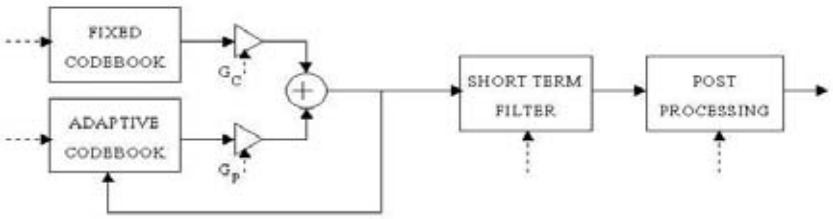


FIGURE 2. Block diagram of G.729 Decoder.

2. METHODS OF CODING

TABLE 1. First set of tested sentences (Ounnapirak, Jitapunkul 1995; Ounnapirak 1995).

Tested sentences (Thai)	Tested sentences (English)
ระบบป้องกันไฟต้องดี	This is a top secret.
ข้าศึกอยู่บนเกาะเยอะแยะ	My name is John K. Smith.
เนื้อหมูเน่าเฟะและมีหนอน	Richard concluded the lecture.
ประโยคทดสอบภาษาไทย	Have a nice day for this trip.
กรุงเทพเป็นมหานคร	Leave me alone in the dark.
ดำเนินการได้ตามสะดวก	Wireless communications.

TABLE 2. Second set of tested sentences (Ounnapirak, Jitapunkul 1995; Ounnapirak 1995).

Tested sentences (Thai)	Tested sentences (English)
อาจารย์เรียนเชิญผู้ปกครอง	Give me a zebra flag please.
ฉันฝากนกฮูกให้กับเธอ	Father played the xylophone.
ระฆังถูกซ่อมแซมใต้โกฏิ	Bob quickly wore his slacks.
จุฬามีกฎลงโทษหนัก	Jam did his paper last night.
เฒ่าเก็บมณฑาบนฐานพระ	Excuse me, may I come in.
เฌอเอมเป็นยารักษาโรค	Good bye, see you tomorrow.

In the experiment, 2 coding methods were used: One was CS-ACELP coder without VAD, another was CS-ACELP coder with VAD. Two sets of sentences were used by those coding methods. Then these two sets of sentences were compared. Each set contained 6 Thai and 6 English sentences. The first set of sentences was in table 1, the second was in table 2. In both sets, speeches of 3 men and 3 women were recorded.

The sentences chosen in these sets covered all of the characters in each language. The second set was performed as same as the first set to compare the results from each one.

3. PERFORMANCE EVALUATION

The quality of speech was evaluated in both 2 sets by using the values of segmental signal to noise ratio defined in eqation (1)(Ounnapirak, Jitapunkul 1995; Ounnapirak 1995). Figure 3 and 4 show the original

signal, the reconstructed signals of both methods of the first sentence in the first set for Thai and English respectively, while table 3 shows segSNR for both methods.

$$\text{SegSNR} = \frac{10}{L}\sum_{i=0}^{L-1}\log_{10}\left\{\frac{\sum_{n=0}^{N-1}s^2(iN+n)}{\sum_{n=0}^{N-1}(s(iN+n)-\hat{s}(iN+n))^2}\right\} \quad (1)$$

Where N is length of segment in bit, L is number of segments, s(n) is the original signal and $\hat{s}(n)$ is the reconstructed signal.

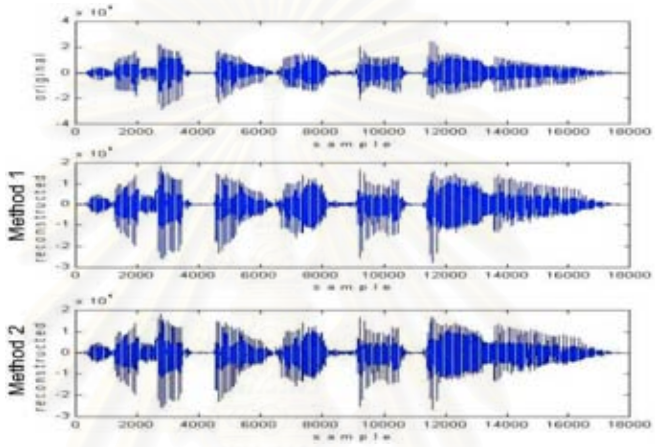


FIGURE 3. Original signal, reconstructed signals of both methods.(Thai in the first set).

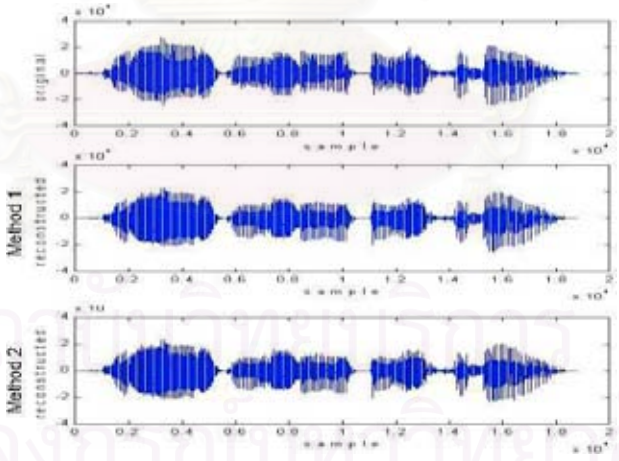


FIGURE 4. Original signal, reconstructed signals of both methods.(English in the first set).

TABLE 3. segSNR of Thai and English speech compression at LP order 10.

Method		SegSNR(dB)			
		Thai		English	
		Male	Female	Male	Female
1st set	1) CS-ACELP	9.45	9.38	9.72	9.72
	2) CS-ACELP with VAD	9.50	9.42	9.75	9.73
2nd set	1) CS-ACELP	9.47	9.40	9.75	9.71
	2) CS-ACELP with VAD	9.49	9.44	9.76	9.72

TABLE 4. segSNR of Thai speech compression at LP order 10, 12 and 14.

Method		SegSNR(dB) - Thai					
		Order 10		Order 12		Order 14	
		Male	Female	Male	Female	Male	Female
1st set	1) CS-ACELP	9.45	9.38	9.53	9.45	9.54	9.48
	2) CS-ACELP with VAD	9.50	9.42	9.62	9.50	9.64	9.52
2nd set	1) CS-ACELP	9.47	9.40	9.54	9.45	9.53	9.48
	2) CS-ACELP with VAD	9.49	9.44	9.60	9.49	9.62	9.51

The results in table 3 show that both methods give Thai speech quality mostly below than English speech quality about 0.25-0.34 dB for the first set, and about 0.27-0.31 dB for the second set. As for methods comparison; both Thai and English, method 2 gives speech quality better than method 1 about 0.01-0.05 dB for the first set, and about 0.01-0.04 dB for the second set. Comparing 2 sets, the results of them were corresponding.

Finally, the order of LP analysis was increased to 12 and 14 to improve the quality of reconstructed speech in case of Thai speech. Table 4 shows the segSNR of both methods. For order 12, it shows the improvement of speech quality about 0.07-0.12 dB for the first set, and about 0.05-0.11 dB for the second set, in comparison to LP order 10. For order 14, it shows the improvement of speech quality about 0.09-0.14 dB for the first set, and about 0.06-0.13 dB for the second set, in comparison to LP order 10. Comparing 2 sets, the results of them were corresponding.

The results were shown that the quality of coding was improved. But the coding rate also increased for allocating the higher order information.

4. CONCLUSION

The ITU G.729 speech coder was applied to Thai Language. By no other modification, the quality of Thai coding is not equivalent to the English Language. After modifying the LP analysis by increasing the LP order from 10 to 12 or 14, the quality of Thai speech coding are truly improved. But the coding rate also increased for allocating the higher order information.

ACKNOWLEDGEMENTS

The authors wish to thank Digital Signal Processing Research Laboratory, Chulalongkorn University for the facility and technical support for this research, and NSTDA for the scholarship passthrough Telecommunication Consortium.

REFERENCES

COX R. V. 1997. Three New Speech Coders From The ITU Cover A Range Of Applications. IEEE Communications Magazine.

KATAOKA A., T. MORIYA, AND S. HAYASHI. 1996. An 8-kb/s Conjugate Structure (CS-ACELP) Speech coder. IEEE Transactions on speech and audio processing, Vol. 4, No. 6.

ISO/JTC 1/SC 29. 1998. N2203CELP. ISO/IEC 14496-3 FCD, ISO/JTC 1/SC 29/WG11.

DELLER AND J.R. JR. 1993. Discrete-Time Processing of Speech Signals. Macmillan, New York.

SCHRODER G. AND M. H. SHERIF. 1997. The Road to G.729: ITU 8-kb/s Speech Coding Algorithm with Wireline Quality. IEEE Communications Magazine.

OUNNAPIRAK C. AND S. JITAPUNKUL. 1995. Speech Compression Using Wavelets Packets Based on CELP Algorithm. 18th conference of Electrical Engineering, Technology Mahanakorn University, Bangkok.

OUNNAPIRAK C. 1995. Speech Compression Using Wavelet Transform and LPC Vector Quantization. Master thesis, Chulalongkorn University, Bangkok.

Novel Technique For Tonal Language Speech Compression Based on A Bitrate Scalable MP-CELP Coder

S. Chompun, S. Jitapunkul and D. Tancharoen
Digital Signal Processing Research Laboratory,
Department of Electrical Engineering,
Faculty of Engineering, Chulalongkorn University,
Bangkok 10330, Thailand
E-Mail: b0597021@student.chula.ac.th, jsomchai@chula.ac.th

Abstract

This paper proposes a modification of flexible Multi-Pulse based Code Excited Linear Predictive (MP-CELP) coder with bitrate scalabilities for tonal language speech in the multimedia applications. The coder consists of a core coder and bitrate scalable tools. The high pitch delay resolutions are applied to the adaptive codebook of core coder for tonal language speech quality improvement. The bitrate scalable tool employs multi-stage excitation coding based on an embedded-coding approach. The multi-pulse excitation codebook at each stage is adaptively produced depending on the selected excitation signal at the previous stage. The experimental results show that the speech quality of the proposed coder is improved above the speech quality of the conventional coder without pitch-resolution adaptation.

1. Introduction

Nowadays the digital communications are widely developed. The audio, images, video or data information can be transmitted passthrough wire or wireless network channels. Simultaneously, the number of users to access these networks increases rapidly. Consequently, channel capacity has to be increased, signal compression aims to perform this [1]. Since the multimedia applications such as videophone and videoconference on ATM and Internet are widely used, the high quality speech coders are highly demanded. These kinds of applications require special considerations for packet loss. To overcome this problem, it is to realize a scalable coder where the synthesized speech signal can be decoded from the received packets, which contain only a part of the whole encoded bitstream. One of standardization activities for such areas is undergoing at the MPEG-4 [2][8].

In 1995, Conjugate-Structure Algebraic Code Excited Linear Predictive (CS-ACELP) coding was developed and standardized as ITU G.729 speech coding at 8 kbps. Later, MP-CELP coder has been proposed to be a scalable coder around this bitrate. This flexible coder employs the multi-pulse excitation which the number of pulses in fixed-entry codebook is selective for bitrate scalability and multiple bitrate functionality according to the MPEG-4 CELP speech coder requirements, see e.g., [2][8].

In MP-CELP, amplitudes or signs for multi-pulse excitation are simultaneously vector quantized. To improve speech quality for background noise conditions, the adaptive pulse location restriction method are applied [3]. This coder operates at various bitrates ranging from 4 to 12 kbps utilizing the flexibility in multi-pulse excitation coding [8].

As for tonal language, such as Thai, a syllable is composed of consonants, vowels and tone [9]. The smallest structure of sounds or syllables in Thai is composed of one vowel unit or one diphthong, one, two or three consonants, and a tone. The structure can be represented as illustrated in figure 1. Ci is initial consonant, Cf is final consonant, V is vowel and T is tone.

Figure 1. Thai syllable structure

The significant difference between tonal and toneless language is tone (T). In tonal language, the words of different tones yield their distinguished meaning. By using the standard speech coder such as CS-ACELP with tonal language, it showed the degraded speech quality when compared to those of toneless language. The reason is that the tone information precision is not enough for tonal language, e.g., [1][9].

This paper proposes a bitrate scalable tonal language speech coder based on a multi-pulse based code excited linear predictive coding [4][5]. The proposed coder provides the bitrate scalabilities which is effective in multimedia communications. Moreover, this coder is improved for the tonal language speech by applying the high pitch delay resolutions to retain the tone information precision. Section 2 describes operation principle for the bitrate scalable MP-CELP coder. The proposed tonal language speech coder is presented in Section 3.

Experimental results are shown in Section 4. Finally, the conclusion of this paper is given in Section 5.

2. Bitrate scalable MP-CELP coder

The operation principle for bitrate scalable MP-CELP coder can be separated into 2 parts, MP-CELP core coder and bitrate scalable tool.

2.1. MP-CELP core coder

The MP-CELP core coder achieves a high coding performance by introducing a multi-pulse vector quantization as depicted in figure 2 [4][5]. The input speech of 10 ms frame is processed through linear prediction (LP) and pitch analysis. The LP coefficients are quantized in the line spectrum pairs (LSP) domain. The pitch delay is encoded by using an adaptive codebook. The residual signal for LP and the pitch analysis is encoded by the multi-pulse excitation scheme. The multi-pulse excitation signal is composed of several non-zero pulses. The pulse positions are restricted in the algebraic-structure codebook and determined by an analysis-by-synthesis approach, e.g., [6][7]. The pulse signs and positions are encoded, while the gains for pitch predictor and the multi-pulse excitation are normalized by the frame energy and encoded.

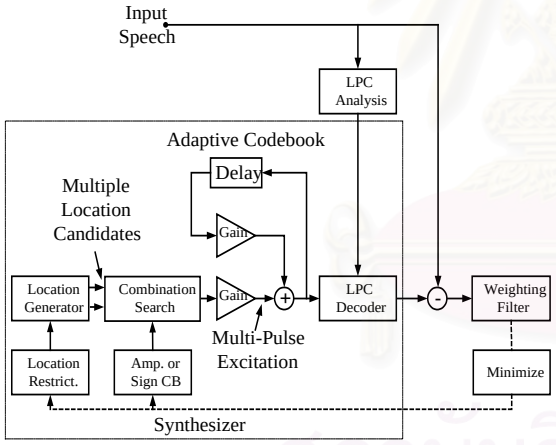


Figure 2. MP-CELP core coder

2.2. Bitrate scalable tool

This paper uses at most 3 stages of the bitrate scalable tools according to the MPEG-4 CELP requirement as illustrated in figure 3 (at the end of paper)[8]. The bitrate scalable tool encodes the residual signal produced at the MP-CELP core coder utilizing the multi-pulse vector quantization. Adaptive pulse position control is employed to change the algebraic-structure codebook at each excitation-coding stage depending on the encoded multi-pulse excitation at the previous stage. The algebraic-structure codebook is adaptively controlled to inhibit the same pulse positions as those of the multi-pulse excitation in the MP-CELP core coder or the previous stage. The pulse positions are determined so that the perceptually weighted distortion between the residual signal and output signal from the scalable tool is minimized. The LP synthesis and perceptually weighted filters are commonly used for both the MP-CELP core coder and the scalable tool.

For this conventional coder, the bit allocation is shown in table 1. The bitrate of core coder is 5600 bps. As for bitrate scalable tool, each stage increases the bitrate of 800 bps. Though, the 1, 2, 3 stages of scalability operate at the total bitrate of 6400, 7200 and 8000 bps respectively.

Table 1. Bit allocation for the conventional coder

Parameter	MP-CELP core coder	Bitrate scalable tool (1 stage)
LSP	18	
Pitch delay	14	
Multi-Pulse	5x2	4x2
Gain	7x2	
Total	56	8
Bitrate (bps)	5600	800

3. Tonal language speech coder

In Thai language, there are 5 different tones, mid(0), low(1), falling(2), high(3) and rising(4), whose characteristics are depicted in figure 4 [9]. Each graph represents the behavior of fundamental frequency (f0) in a period of syllable time where f0 is the inverse of pitch delay time. Though, f0 indicates the periodicity of voice. Investigating the difference between Thai male and Thai female f0 behaviors, Thai female f0 change rate is almost all more than Thai male f0's, see e.g., [10]. This is why the Thai female speech quality encoded by CS-ACELP coder is lower than the Thai male speech quality [1]. Hence, detecting f0 with high precision yields the improvement of the tonal language speech quality.

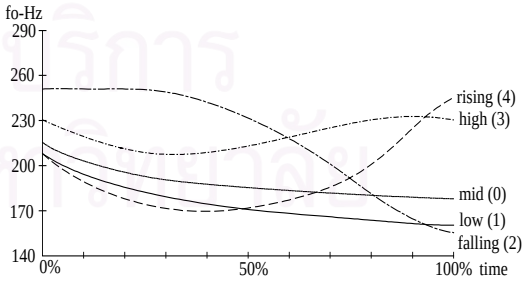


Figure 4. f0 characteristic of 5 tones in Thai

Since pitch delay (or f0) significantly involves in tone of tonal language, this paper proposes an improvement of the bitrate scalable MP-CELP coder by applying the High Pitch Delay Resolutions (HPDR) technique to the pitch analysis of the core coder. The HPDR at pitch fraction of 1/2, 1/3 and 1/4 is adopted to the pitch analysis, consequently, it causes the increments of bitrate as 200, 400 and 400 bps respectively.

The HPDR technique is done by including the pitch fraction analysis within the conventional pitch analysis which finds the optimum fraction around the prior pitch delay integer of the conventional pitch analysis. In order to find the adaptive excitation for the proposed technique, the FIR filter based on a Hamming windowed sin(x)/x function truncated at ±11 and padded with zeros at ±12 is adopted to weight the excitation in the pitch fraction analysis.

4. Experimental results

The coding quality of the proposed coder was evaluated subjectively and objectively by using 36 tested sentences from 10 men and 10 women, some of them were shown in table 2.

Table 2. Thai tested sentences (examples)
0, 1, 2, 3, 4 at each word represent tone of Thai

Order	Tested sentences
1	เขา แห่ นาค เวียน รอบโบสถ์ khaw4 hx:1 na:k2 wi:an0 r@:p2 bo:t1
2	คน ทำ บาป อวด ตัว ว่า เก่ง khon0 tham0 ba:p1 ?u:at1 tu:a0 wa:2 keng1
3	คำ ว่า เตียบ แปร ว่า ตะ ลุ่ม kham0 wa:2 ti:ap1 plx:0 wa:2 ta1 lum2
4	พวก นั้น โดน ปรับ ราย ตัว phu:ak2 nan3 do:n0 prap1 ra:j0 tu:a0
5	เขา เป็น ญาติ อำ ภา khaw4 pen0 ja:t2 ?am0 pha:0
6	น้อง จะ เอา ว่าว อัน นั้น n@:ng3 ca1 ?aw0 waw2 ?an0 nan3
7	เขา อยาก สัก ลาย เสือ ที่ แขน khaw4 ja:k1 sak1 la:j0 sv:a4 thi:2 khx:n4

The effectiveness of the high pitch delay resolutions applied to the conventional coder was evaluated using average segmental SNRs and MOS scores. Comparison tests of each grouped bitrate were conducted and shown in table 3 and 4.

For the objective test (SegSNR), the results showed that both male and female speech quality, every grouped bitrates, the HPDR at pitch fraction of 1/4 gave the maximum value. The order of speech quality from the best to the worst was 1/4's, 1/3's, 1/2's and conventional's respectively. For the subjective test (MOS score), the results were corresponding to those of the objective test.

From the experimental results, the more high resolution be used the more speech quality be obtained. This indicates that the proposed HPDR technique brings the better pitch precision which causes an improvement of the coding quality for tonal language.

Table 3. Objective speech quality (SegSNR)

Stages of scalability		none	1	2	3
Grouped Bitrate (pbs)		5600+	6400+	7200+	8000+
Male	Conv.	7.45	7.81	8.31	9.06
	HPDR.1/2	7.61	7.93	8.79	9.45
	HPDR.1/3	7.70	8.01	8.87	9.59
	HPDR.1/4	7.86	8.04	8.90	9.61
Female	Conv.	7.26	7.66	8.18	8.69
	HPDR.1/2	7.35	7.88	8.21	9.23
	HPDR.1/3	7.61	8.01	8.82	9.38
	HPDR.1/4	7.70	8.06	8.85	9.41

Table 4. Subjective speech quality (MOS Score)

Stages of scalability		none	1	2	3
Grouped Bitrate (pbs)		5600+	6400+	7200+	8000+
Male	Conv.	3.12	3.18	3.28	3.35
	HPDR.1/2	3.18	3.22	3.30	3.36
	HPDR.1/3	3.21	3.30	3.32	3.38
	HPDR.1/4	3.22	3.33	3.34	3.39
Female	Conv.	3.09	3.11	3.20	3.31
	HPDR.1/2	3.13	3.16	3.24	3.33
	HPDR.1/3	3.16	3.23	3.27	3.34
	HPDR.1/4	3.19	3.28	3.30	3.36

5. Summary

A modification of bitrate scalable tonal language speech coder has been proposed. This coder consists of a MP-CELP core coder and the bitrate scalable tools. The high pitch delay resolutions are applied to adaptive codebook of core coder for tonal speech quality improvement. The results show that the coding quality of the proposed coder is better than the conventional coder for Thai language.

6. References

[1] S. Chompun, S. Jitapunkul, D. Tancharoen and T. Srithanasan, "Thai Speech Compression Using CS-ACELP Coder Based on ITU G.729 Standard", *Proc. SNLP*, pp.263-267, Thailand, 2000.

[2] T. Nomura, M. Iwadare, M. Serizawa and K. Ozawa, "A Bitrate and Bandwidth Scalable CELP Coder", Proc. ICASSP, Vol.1, pp.341–344, 1998.

[3] K.Ozawa and M.Serizawa, "High Quality Multi-pulse Based CELP Speech Coding at 6.4 kb/s and its Subjective Evaluation", Proc. ICASSP, pp.529-532, 1998.

[4] S. Taumi, et al., "Low-Delay CELP with Multi-pulse VQ and Fast Search for GSM EFR", Proc. ICASSP, pp.562-565, 1996.

[5] K.Ozawa, et al.,"MP-CELP Speech Coding Based on Multi-pulse Vector Quantization and Fast Search",

Proc. IEICE, Vol.J79-A, No.10, pp1655-1663, 1996 (In Japanese).

[6] C. Laflamme, et al., "16 kbps Wideband Speech Coding Technique Based on Algebraic CELP", Proc. ICASSP, pp.13-16, 1991.

[7] ITU - T Rec. G.729, "Coding of Speech at 8-kbit/s Using Conjugate Structure Algebraic Code-Excited Linear Prediction (CS-ACELP)", COM 15-152-E, 1995.

[8] T. Nomura, et al., "Proposal of Compression Algorithm with Rate Control for MPEG-4/Audio Core Experiments", ISO/IECJTC1 /SC29/WG11/ M1509, 1996.

[9] S. Luksaneeyanawin, "Linguistics Research and Thai Speech Technology", Proc. The 5th International Conference on Thai Studies, School of Oriental and African Studies, University of London, United Kingdom, 1993.

[10] U. Thathong, S. Jitapunkul and V. Ahkuputra, "Classification of Thai Consonants Naming Using Thai Tone", Proc. ICSLP, Beijing China, 2000.

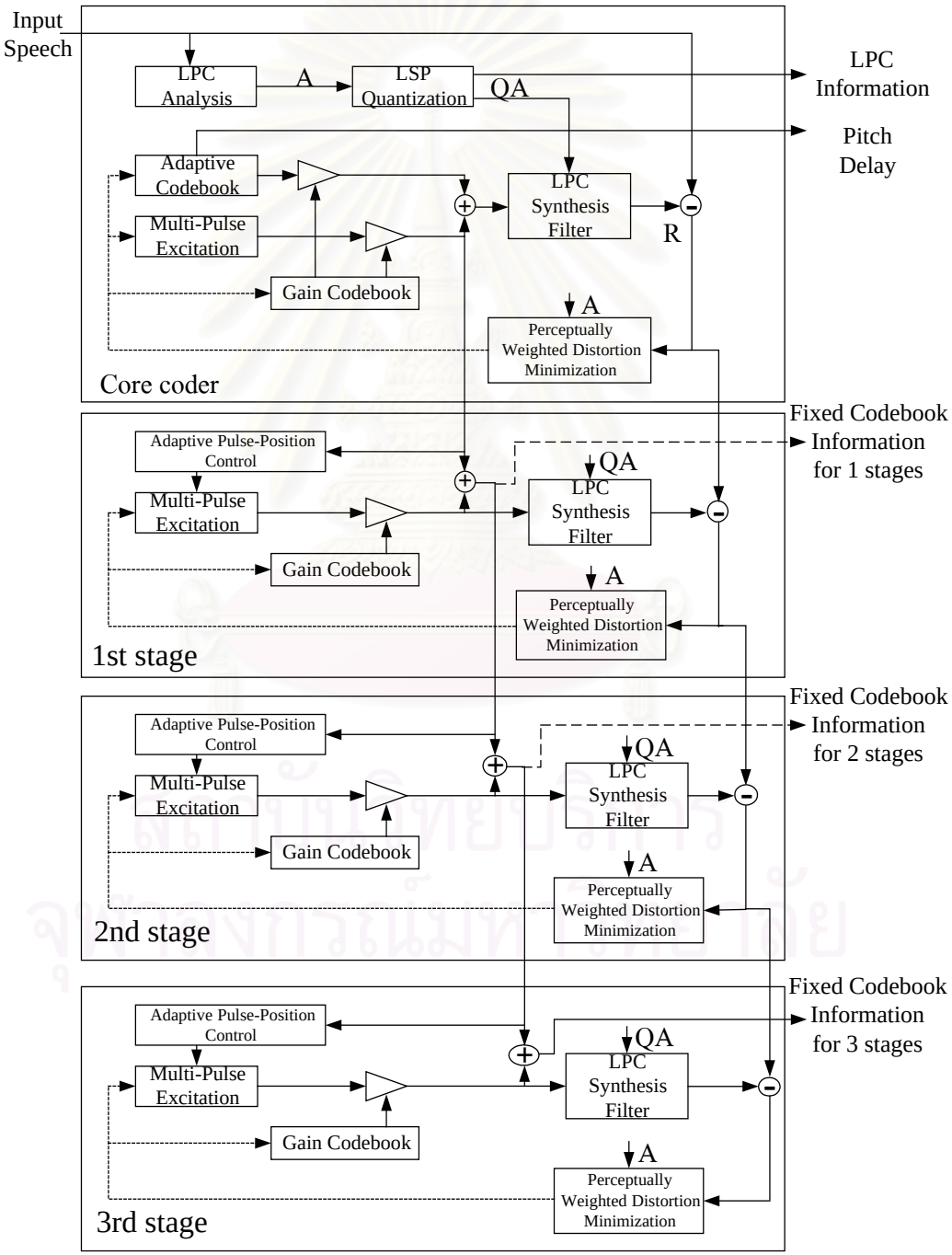


Figure 3. 3-stage bitrate scalable MP-CELP coder

ประวัติผู้เขียนวิทยานิพนธ์

นายสุภัทรชัย ชมพันธุ์ เกิดวันที่ 10 มิถุนายน พ.ศ. 2520 ที่จังหวัดปราจีนบุรี เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2537 สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ที่ห้องปฏิบัติการวิจัยกรรมวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2541