การพัฒนาคุณภาพเสียงของ VoIP CODEC โดยการเรียงต่อกันของโคเดก

ปวิตรา โลกโบว์[1]* และ บวร ปภัสราทร[2]

บทคัดย่อ

ในปัจจุบันมีผู้คิดค้นและพัฒนาโคเดก (CODEC) ขึ้นมามากมายโดยใช้ขั้นตอนวิธีการการบีบอัด (Compression Algorithm) ที่ต่างกันออกไป ทำให้โคเดกแต่ละตัวให้คุณภาพเสียงที่ไม่เท่ากัน ส่งผลให้โอกาสการเกิดการเรียงต่อกันของโคเดก (CODEC Tandem) ระหว่างโคเดกที่มีคุณภาพเสียงที่ต่ำมาก และโคเดกที่มีคุณภาพเสียงที่สูงมากมีมากขึ้น ดังนั้นงานวิจัยฉบับนี้จึงสนใจศึกษาและวิเคราะห์ความเป็นไปได้ในการปรับปรุงคุณภาพเสียงของ VoIP CODEC-G.723.1A โดยใช้เทคนิคการเรียงต่อกันกับโคเดกตัวอื่นที่เป็นที่นิยมใช้งานในช่วงแถบความถี่แคบ (Narrow-band CODEC) รวมทั้ง OPUS ซึ่งเป็นโคเดกที่มีความสามารถหลากหลายและยังอยู่ในการพัฒนาของ Internet Engineering Task Force (IETF) จากการทดลองพบว่า 1) เพื่อให้ได้คุณภาพเสียงที่ดีที่สุดสำหรับ OPUS ค่า Frame Size ควรใช้ที่มีขนาดมากกว่าหรือเท่ากับ 20 มิลลิวินาทีขึ้นไป 2) คุณภาพของเสียงที่ผ่านการเข้ารหัสด้วย G723.1A เพียงครั้งเดียว มีค่า MOS เพิ่มขึ้นเมื่อมีการเข้ารหัสด้วยโคเดกชนิดอื่นบางชนิดรวมถึง OPUS ก่อน ซึ่งในบรรดาโคเดกทั้งหมด OPUS สามารถเพิ่มคุณภาพของเสียงได้มากที่สุด

คำสำคัญ: การเรียงต่อกันของโคเดก G723.1A, OPUS, PESQ

[1] นักศึกษา คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
[2] รองศาสตราจารย์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08-1808-3297 อีเมล: 54440343@st.sit.kmutt.ac.th

รับเมื่อ 7 กุมภาพันธ์ 2556 ตอบรับเมื่อ 25 มีนาคม 2556

Quality Improvement of VoIP CODEC Using Tandem Technique

Pawitra Lokbow[1]* and Borworn Papasratorn[2]

Abstract

Variety of compression algorithms of each VoIP CODEC makes the quality of the output files different. As voices may need to be encoded multiple times while being sent through multiple networks that use different CODEC (so called CODEC tandem), there is a possibility that the CODECs which provide low voice quality may be used together with others that provide much better voice quality. The main purposes of this research are to analyze and seek improvement in the quality of voices encoded using VoIP CODEC-G.723.1A by tandeming with other famous Narrow-band VoIP CODECs. As a comparison result, the quality of voices is improved if voices were encoded by tandem of codecs, compared to those encoded by G.723.1A alone. Furthermore, OPUS CODEC which is developed by Internet Engineering Task Force (IETF) can deliver the most significant improvement among all selected CODECs, especially with the frame size of 20 ms. or above.

Keywords: Codec Tandem, G723.1A, OPUS, PESQ

1 Student, School of Information Technology, King Mongkut's University of Technology Thonburi.

2 Associate Professor, School of Information Technology, King Mongkut's University of Technology Thonburi.

* Corresponding Author, Tel. 08-1808-3297, E-mail: 54440343@st.sit.kmutt.ac.th

Received 7 February 2013; Accepted 25 March 2013

1. บทนำ

ในปัจจุบันเทคโนโลยีด้านการสื่อสารและโทรคมนาคมกำลังก้าวเข้าสู่ยุคที่เรียกว่า โครงข่ายโทรคมนาคมในอนาคต (NGN-Next Generation Network) ซึ่งมีเป้าหมายที่จะหลอมรวมโครงข่ายโทรคมนาคมทั้งโครงข่ายโทรศัพท์บ้านและโครงข่ายโทรศัพท์มือถือเข้ากับโครงข่ายอินเทอร์เน็ต [1] เพื่อบรรลุจุดมุ่งหมายข้างต้น ข้อมูลสนทนาเสียงแบบแอนะล็อก (Analog) จากต้นทางจะถูกแปลงไปเป็นข้อมูลดิจิทัล (Digital) แล้วรวบรวมเป็นกลุ่มข้อมูล (Packet) เพื่อส่งไปยังปลายทาง วิธีการดังกล่าวเป็นที่รู้จักกันดีในชื่อ วีโอไอพี (VoIP-Voice over Internet Protocol) แต่ทั้งนี้เนื่องจากเดิมทีโครงข่ายอินเทอร์เน็ตนั้นมิได้ถูกสร้างมาเพื่อรองรับการสื่อสารแบบเวลาจริง (Real Time Voice Communication) ดังนั้น หัวใจสำคัญอย่างหนึ่งในการพัฒนาระบบวีโอไอพี คือระบบที่พัฒนาขึ้นนั้นจะต้องสามารถรับประกันคุณภาพของการส่งข้อมูลให้เป็นไปตามเงื่อนไขหรือความต้องการ (QoS-Quality-of-Service) [2]

หนึ่งในปัจจัยที่มีผลต่อความแตกต่างของเสียงจากต้นทางและเสียงที่ปลายทาง หรืออีกนัยหนึ่งมีผลต่อคุณภาพของเสียงคือโคเดกที่ใช้ในการแปลงข้อมูลจากแอนะล็อก (Analog) ไปเป็นข้อมูลดิจิทัล (Digital) [3] ในปัจจุบันมีโคเดกที่ถูกสร้างขึ้นมาเพื่อให้เลือกใช้งานมากมาย แต่ละตัวมีลักษณะเฉพาะ ขั้นตอนวิธีในการแปลงข้อมูล (Compression Algorithm) และข้อดีข้อเสียที่แตกต่างกัน [4],[5] โคเดกทุกตัวจะมีลักษณะประจำ (CODEC Attributes) ประกอบด้วย บิตเรต (Bitrate) และแบนด์วิดท์ที่ต้องการใช้ (Require Bandwidth) คุณภาพของโคเดก (Quality) ความซับซ้อนในการเข้ารหัส (Complexity) และค่าหน่วงเวลาที่เกิดจากการเข้ารหัส (Delay) [6],[7] โดยในแต่ละโปรแกรมประยุกต์ (Application) เราสามารถเลือกปรับลักษณะประจำเหล่านี้ได้โดยอาศัยการแลกเปลี่ยน (Trade-off) ของแต่ละตัว เช่น เราสามารถเพิ่ม Quality ของ CODEC ได้โดยการเพิ่มบิตเรตและแบนด์วิดท์ที่ต้องการใช้ แต่ความซับซ้อนในการเข้ารหัสและค่าหน่วงเวลาที่เกิดจากการเข้ารหัสจะเพิ่มขึ้นตามด้วยเช่นกัน [6] ซึ่งในแต่ละโครงข่ายก็จะเลือกใช้โคเดกที่ตนเห็นว่าเหมาะสม ในกรณีที่มีการส่งข้อมูลเสียงผ่านโครงข่ายมากกว่าหนึ่งโครงข่ายก็จะทำให้เกิดการเรียงต่อกันของโคเดก (CODEC Tandem) ซึ่งการเรียงต่อกันนี้เองมีผลให้คุณภาพของเสียงลดลง (Quality Degradation) แต่ทั้งนี้การเข้ารหัสแบบเรียงต่อกันที่มีการวิจัยไปแล้วนั้นล้วนแล้วแต่สนใจที่การนำโคเดกตัวอื่นๆ ที่สนใจมาต่อด้านหลังแล้วทำให้คุณภาพของเสียงลดลง [8],[9]

งานวิจัยนี้นำเสนอการพัฒนาคุณภาพเสียงที่เกิดจากการเข้ารหัสด้วยโคเดก G.723.1A ซึ่งเป็นโคเดกที่นิยมใช้ในการใช้เข้ารหัสเสียงเพื่อรับส่งบนโครงข่ายไอพี [10]-[12] โดยในการทดลองผู้วิจัยเลือกใช้วิธีการเข้ารหัสเสียงด้วยโคเดกที่ต้องการทดสอบก่อนแล้วจึงนำมาเข้ารหัสด้วย G.723.1A ทั้งนี้ในจำนวนโคเดกหลายชนิดที่นำมาทดสอบ งานวิจัยฉบับนี้มุ่งเน้นการทดสอบด้วย OPUS เป็นพิเศษเนื่องจาก OPUS เป็นโคเดกที่มีความสามารถหลากหลายและถูกพยากรณ์ว่าจะมีบทบาทสำคัญในอนาคต อนึ่ง ในปัจจุบันโคเดกชนิดนี้ยังอยู่ในระหว่างการพัฒนาของ Internet Engineering Task Force (IETF) [13]

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การวัดความสามารถในการรับรู้และเข้าใจของผู้ฟัง (Perceptive QoS Evaluation)

การวัดความสามารถในการรับรู้และเข้าใจของผู้ฟัง (Perceptive QoS Evaluation) สามารถแบ่งออกได้เป็น 2 วิธีหลักๆ [2],[14] คือ

1. การวัดคุณภาพของเสียงโดยใช้จิตวิสัยของผู้ฟัง (Subjective Methods) เป็นวิธีการที่ให้ผู้เข้าร่วมการทดสอบพูดและฟังเสียงผ่านระบบโทรศัพท์วีโอไอพี แล้วตัดสินคุณภาพของเสียงที่ถูกส่งผ่านระบบเน็ตเวิร์กด้วยการให้คะแนน หนึ่งในวิธีการวัดคุณภาพของเสียงโดย

ใช้จิตวิสัยของผู้ฟังที่เป็นที่นิยมที่สุด คือ Mean Opinion Score (MOS) โดยวิธีการทดลองจะต้องดำเนินการตาม ITU-T Rec. P.800 [15] ซึ่งค่าเฉลี่ยของคะแนนที่ผู้ฟังต้องให้คะแนนแสดงในตารางที่ 1

ตารางที่ 1 ค่า MOS และคุณภาพของเสียง

คะแนน MOS	คุณภาพของเสียง
5	Excellent - ยอดเยี่ยม
4	Good - ดี
3	Fair - พอใช้
2	Poor - ไม่ดี
1	Bad - แย่

2. การวัดคุณภาพของเสียงโดยใช้เครื่องมือหรือขั้นตอนวิธีของคอมพิวเตอร์ (Objective Methods) การวัดคุณภาพเสียงด้วยวิธีนี้มีข้อดีคือ ใช้เวลาและค่าใช้จ่ายน้อยกว่า เมื่อเทียบกับการวัดคุณภาพของเสียงโดยใช้จิตวิสัยของผู้ฟัง ด้วยเหตุนี้เองจึงทำให้วิธีการวัดคุณภาพเสียงโดยใช้เครื่องมือหรือขั้นตอนวิธีของคอมพิวเตอร์นั้นได้รับความนิยมมากขึ้น [14]

2.2 การวัดคุณภาพเสียงโดยใช้ขั้นตอนวิธีของคอมพิวเตอร์ PESQ (Perceptual Evaluation of Speech Quality)

เป็นการวัดคุณภาพของเสียงโดยการเปรียบเทียบเสียงต้นฉบับ (Original Speech File) กับเสียงที่ผ่านการกระทำใดๆ ก็ตามที่เราสนใจ (Degraded Speech File) [16] ซึ่งผลลัพธ์ที่ได้จากวิธีนี้ถือว่ามีความน่าเชื่อถือมากวิธีหนึ่งและสามารถแปลงค่ากลับออกมาเป็นค่า MOS ด้วยสมการที่แสดงไว้ใน [17]

2.3 CODEC

โคเดกทำหน้าที่ในการแปลงข้อมูลสนทนาเสียงแบบแอนะล็อกไปเป็นข้อมูลดิจิทัลรวบรวมเป็นชุดข้อมูลที่สามารถส่งไปบนเน็ตเวิร์ก โคเดกทุกตัวจะมีลักษณะประจำ (CODEC Attributes) อยู่ 4 ข้อประกอบด้วย [6],[7]

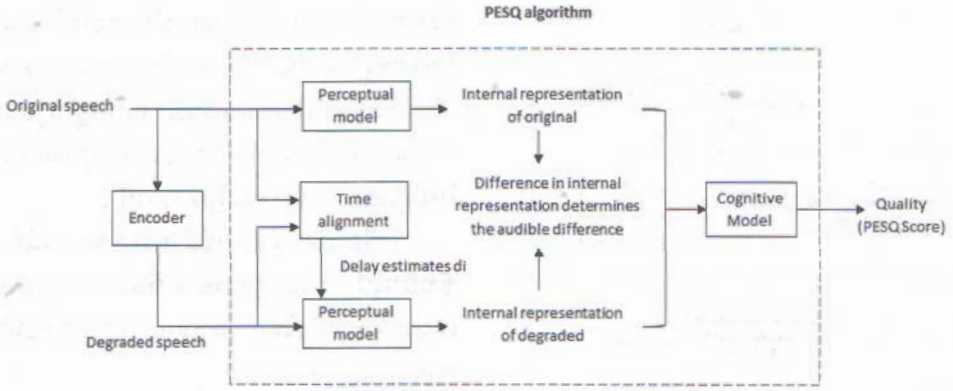
- บิตเรตและแบนด์วิดท์ที่ต้องการใช้
- คุณภาพของโคเดก
- ความซับซ้อนในการเข้ารหัส
- ค่าหน่วงเวลาที่เกิดจากการเข้ารหัส

เนื่องจากบนโครงข่ายเน็ตเวิร์กมีแบนด์วิดท์ที่จำกัดทำให้การแปลงข้อมูลเสียงต้องการการบีบอัด (Compression) เพื่อลดขนาดแบนด์วิดท์ที่ต้องการใช้ในการส่งข้อมูลเสียงนั้น แต่การทำเช่นนี้จะส่งผลกับคุณภาพของเสียง เพราะเหตุนี้เองผู้พัฒนาโคเดกจะต้องมั่นใจว่าเสียงที่ผ่านโคเดกนั้นแม้มีคุณภาพลดลงแต่ผู้ฟังยังสามารถเข้าใจความหมาย (Perception Quality) ของข้อมูลเสียงที่ถูกส่งผ่านมา [18] ทำให้ในปัจจุบันมีโคเดกที่ถูกสร้างขึ้นเพื่อให้เลือกใช้งานมากมาย แต่ละตัวมีลักษณะเฉพาะ ขั้นตอนวิธีในการแปลงข้อมูล และข้อดีข้อเสียที่แตกต่างกัน [4],[5] เราสามารถเลือกปรับลักษณะประจำเหล่านี้ได้โดยอาศัยการแลกเปลี่ยน (Trade-off) ของแต่ละตัว [6] ซึ่งในแต่ละโครงข่ายหรือโปรแกรมประยุกต์ก็จะเลือกใช้โคเดกที่ตนเห็นว่าเหมาะสม

โคเดกยังสามารถแบ่งย่อยตามช่วงความถี่ที่รองรับการใช้งานออกเป็น 2 ช่วง (Range) คือ [4],[16]

- แถบความถี่แคบ (Narrowband-NB) สำหรับเข้ารหัสข้อมูลเสียงโดยใช้ความละเอียดของความถี่เสียง (Sampling Rate) 8 kHz หรือแบนด์วิดท์ต่ำกว่า 4 kHz ตามทฤษฎีของ Nyqust [9]
- แถบความถี่กว้าง (Wideband-WB) สำหรับเข้ารหัสข้อมูลเสียงโดยใช้ความละเอียดของความถี่เสียง 16 kHz หรือแบนด์วิดท์ 7 kHz

นอกจากนี้ในกรณีที่มีการส่งข้อมูลเสียงผ่านโครงข่ายมากกว่าหนึ่งโครงข่ายก็จะทำให้เกิดการเรียงต่อกันของโคเดก (CODEC Tandeming) ทั้งที่เป็นโคเดกตัวเดียวกันและ/หรือต่างกัน ซึ่งอาจมีผลให้คุณภาพของเสียงลดลง (Quality Degradation) [8],[9]



รูปที่ 1 ภาพรวมของหลักการและทฤษฎีที่ถูก อ้างอิงโดย PESQ

2.3.1 OPUS CODEC

OPUS เป็นโคเดกที่อยู่ระหว่างการพัฒนาของ Internet Engineering Task Force (IETF) ด้วยความร่วมมือของ Mozilla Corporation และ Skype Technologies (Microsoft) ในการผสมผสานระหว่างโคเดกสองตัวคือ SILK และ CELP ทำให้ตัวมันสามารถรองรับการใช้งานได้ตั้งแต่แถบความถี่แคบ (Narrowband Speech) ไปจนถึงแถบความถี่ที่กว้างมาก (Fullband Music) ผู้ใช้งานสามารถเลือกจำนวนบิตต่อวินาทีที่แปลงจากข้อมูลแอนะล็อกไปเป็นข้อมูลดิจิทัลได้ตั้งแต่ 6 กิโลบิตต่อวินาที (kbps) ไปจนถึง 510 กิโลบิตต่อวินาที ด้วยความสามารถที่มีอยู่มากนี้เองที่ทำให้ IETF ตั้งใจจะพัฒนาให้ OPUS เป็นโคเดกมาตรฐานในอนาคต เนื่องจากมันสามารถรองรับการใช้งานโปรแกรมประยุกต์สำหรับเสียงพูด (Speech) ไปจนถึงเสียงเพลง (Music) [11],[19]

2.3.2 ITU-T G.711

G.711 หรือที่นิยมเรียกกันว่า PCM (Pulse Code Modulation) ถือว่าเป็นโคเดกพื้นฐานที่สุด มีความซับซ้อนของการทำงานที่น้อยและให้คุณภาพเสียงที่ถือว่าดีที่สุดเมื่อเทียบกับตัวอื่นๆ เนื่องจากแบนด์วิดท์ที่ต้องการของ G.711 คือ 64 kbps โหมดการทำงานของ G.711 มี 2 วิธี คือ μ-law Algorithm (ใช้ในอเมริกาเหนือและญี่ปุ่น) และ A-law Algorithm (ใช้ในยุโรปและนอกเหนือจากที่กล่าวมา) [10]-[12],[20],[21],[28]

2.3.3 ITU-T G.723.1A

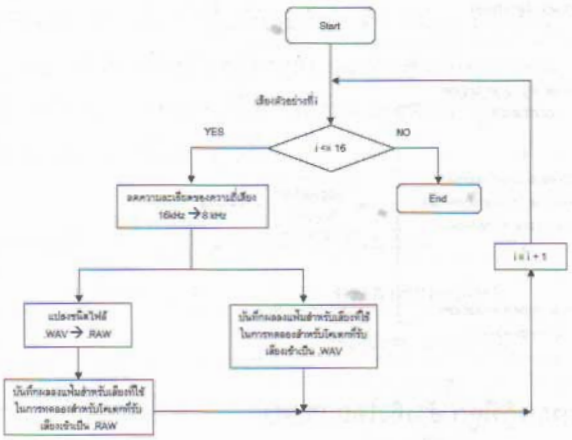
G.723.1A เป็นโคเดกที่มีบิตเรตที่ต่ำมากและสามารถเลือกได้ 2 โหมดซึ่งแต่ละโหมดใช้อัลกอริทึมต่างกัน คือ 5.3 kbps ใช้ Algebraic-Code-Excited Linear-Prediction (ACELP) และอีกตัวคือ 6.3 kbps ใช้ Multipulse Maximum Likelihood Quantization (MP-MLQ) ข้อดีของ G.723.1 คือมีบิตเรตที่ต่ำจึงเหมาะและนิยมใช้ในการรับส่งเสียงผ่านเครือข่ายไอพี แต่เนื่องจากการทำให้บิตเรตต่ำนั้นต้องการการบีบอัดที่มากขึ้นทำให้มีความซับซ้อนของอัลกอริทึมที่สูงส่งผลให้มีค่าหน่วงเวลาจากการเข้ารหัสค่อนข้างมาก [10]-[12],[20],[22]

2.3.4 ITU-T G.726

G.726 หรืออีกชื่อหนึ่งคือ Adaptive Differential Pulse-Codec Modulation-ADPCM ผู้ใช้สามารถเลือกบิตเรตได้ถึง 4 แบบด้วยกันคือ 16 kbps, 24 kbps, 32 kbps และ 40 kbps แต่ที่นิยมใช้งานมากที่สุดคือ 32 kbps เนื่องจากให้คุณภาพเสียงเกือบเทียบเท่า G.711 แต่ใช้แบนด์วิดท์เพียงครึ่งเดียวของ G.711 [10]-[12], [20],[23]

2.3.5 ITU-T G.727

G.727 ถูกสร้างขึ้นมาเพื่อเป็นส่วนขยายของ G.726 ดังนั้นยังคงมีบิตเรตให้เลือก 4 แบบเช่นเดิม [10]-[12],[20],[24]



รูปที่ 2 ขั้นตอนการเตรียมเสียงเพื่อใช้ในการทดลอง

2.3.6 ITU-T G.728

เทคโนโลยีที่ใช้ใน G.728 คือ Low Delay Code Excited Linear Prediction LD-CELP มีบิตเรตเท่ากับ 16 kbps แต่คุณภาพเสียงที่ได้ มีค่าใกล้เคียงกับ G.726 Bitrate 32 kbps [10]-[12],[20],[25]

2.3.7 ETSI SMG11-GSM Full Rate

GSM Full Rate (GSM-FR) หรืออีกชื่อหนึ่งคือ GSM 06.10 ถือได้ว่าเป็นโคเดกที่มีผู้ใช้งานจำนวนมากบนโทรศัพท์มือถือ GSM และมีบิตเรตเท่ากับ 13 kbps [10]-[12],[20]

3. การปรับปรุงคุณภาพเสียงของ VoIP CODEC โดยการเรียงต่อกันของ CODEC

ในเบื้องต้นก่อนที่จะเริ่มการทดลองนำโคเดกมาทำการเรียงต่อกัน จำเป็นต้องมีการจัดเตรียมเสียงเพื่อใช้เป็นต้นฉบับในการทดลอง รวมถึงการเข้ารหัสและถอดรหัสด้วยโคเดกแต่ละตัวเพียงตัวเดียวก่อนเพื่อใช้ในการอ้างอิงและเปรียบเทียบในการทดลองต่อๆ ไปในงานวิจัยฉบับนี้

3.1 การเตรียมเสียงเพื่อใช้ในการทดลอง

ผู้วิจัยเลือกใช้เสียงต้นฉบับที่ถูกจัดทำขึ้นโดย ITU และจัดเก็บไว้ใน Rec. ITU-T P.50 [26] เสียงต้นฉบับดังกล่าวมีคุณสมบัติดังแสดงในตารางที่ 2 แต่เนื่องจากโคเดกที่ผู้วิจัยสนใจถูกจัดอยู่ในช่วงแถบความถี่แคบและเสียงที่รับเข้า (Input Signal) ต้องอยู่ในรูปแบบ .RAW ดังนั้น ผู้วิจัยจึงได้ทำการแปลงสัญญาณเสียงจาก [26] ให้มีคุณสมบัติดังแสดงในตารางที่ 2

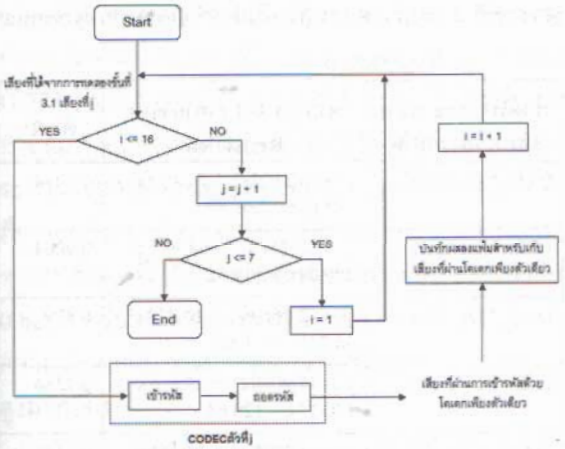
เครื่องมือที่ผู้วิจัยใช้สำหรับขั้นตอนนี้คือ sox [27] ซึ่งเป็นโปรแกรมที่ถูกพัฒนาขึ้นโดย Sourceforge และเหมาะสำหรับใช้ในการดำเนินการต่างๆ กับเสียง (Sound Processing Programs)

ตารางที่ 2 เปรียบเทียบคุณลักษณะของไฟล์เสียงจาก ITU-T Rec. P.50 และเสียงที่ใช้ในการทดลอง

คุณลักษณะ	ITU-T P.50	เสียงที่ใช้ทดลอง
จำนวนเสียงที่ใช้ในการทดลอง	16 signals	16 signals
ภาษา	American-English	American-English
เพศของผู้พูด	ชาย 8 หญิง 8	ชาย 8 หญิง 8
ความละเอียดของความถี่เสียง (Sampling Rate)	16,000 kHz	8,000 Hz
วิธีการเข้ารหัส (Encoding Type)	16-bit linear PCM	16-bit linear PCM
รูปแบบของไฟล์ (File Type)	.WAV	.RAW

3.2 การเข้ารหัสและถอดรหัสเสียงด้วยโคเดกเพียงตัวเดียว

เนื่องจากงานวิจัยฉบับนี้ต้องการค้นหาค่าคุณภาพเสียงที่พัฒนาขึ้น จึงต้องนำเสียงที่ได้จากขั้นตอนที่ 3.1 มาทำการทดลองเพื่อเข้ารหัสและถอดรหัสด้วยโคเดกแต่ละตัวเพียงตัวเดียวก่อนดังรูปที่ 3 เพื่อใช้เป็นตัวเปรียบเทียบในขั้นตอนต่อๆ ไป ซึ่งโค้ดที่ใช้ในการเข้ารหัสและถอดรหัสของ G711, G726, G727, G728 และ GSM



รูปที่ 3 ขั้นตอนการเข้ารหัสและถอดรหัสด้วยโคดเดกแต่ละตัวเพียงตัวเดียว

Full Rate ถูกพัฒนาขึ้นโดย ITU และถูกจัดเก็บไว้ใน Rec. ITU-T G.191 [20] สำหรับ OPUS ถูกจัดเก็บไว้ใน IETF RFC 6716 [19] ทั้งนี้การเข้ารหัสทั้งหมดใช้ค่ามาตรฐาน (Default Value) และไม่ได้ส่งผ่านโครงข่ายจริง

หลังจากผ่านขั้นตอนการเข้ารหัสและถอดรหัสตามขั้นตอนที่ 3.1 และ 3.2 เป็นที่เรียบร้อยแล้ว ขั้นตอนต่อไปคือการทดลองใช้วิธีการเข้ารหัสแบบเรียงต่อกัน ซึ่งขั้นตอนนี้สามารถแบ่งออกได้เป็นสองส่วนคือ การเข้ารหัสและถอดรหัสเสียงและการวัดคุณภาพของเสียงโดยใช้ PESQ ดังที่แสดงไว้ตามรูปที่ 4

3.3 การเข้ารหัสและถอดรหัสเสียงแบบเรียงต่อกัน

ขั้นตอนนี้เป็นการนำผลลัพธ์ที่ได้จากขั้นตอนที่ 3.1 มาทำการเข้ารหัสด้วยโคเดกต่างๆ ที่อยู่ในขอบข่ายของงานวิจัยฉบับนี้ จากนั้นนำมาเข้ารหัสครั้งที่สองด้วยโคเดก G.723.1A โดยที่โคเดกที่ถูกเลือกมาพิจารณามีดังต่อไปนี้ OPUS, G711, G726, G727, G728 และ GSM Full Rate

3.4 การวัดคุณภาพของเสียงโดยใช้ PESQ

การวัดคุณภาพเสียงกระทำโดยการวัดค่า PESQ โดยใช้ขั้นตอนและซอฟท์แวร์ที่ถูกพัฒนาขึ้นโดย ITU และ

รูปที่ 4 ขั้นตอนการปรับปรุงคุณภาพเสียงของ VoIP CODEC โดยการเรียงต่อกันของโคเดก

ถูกจัดเก็บไว้ใน Rec. ITU-T P.862 [16] และเสียงที่ใช้ในการเปรียบเทียบมีดังนี้

- Reference Speech เสียงที่ได้จากการทดลองที่ 3.1
- Degraded Speech เสียงที่ผ่านการเข้ารหัสด้วยโคเดกเพียงตัวเดียวและ เสียงที่ผ่านการเข้ารหัสแบบเรียงต่อกันของโคเดก 2 ตัว

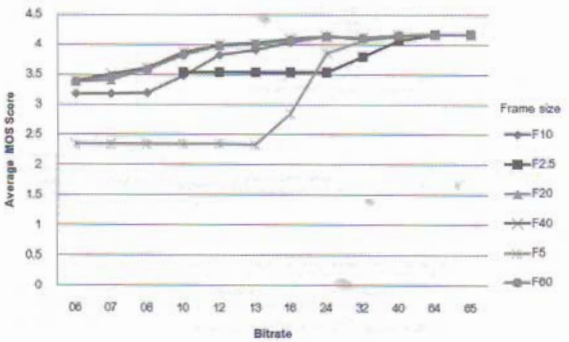
ตัวอย่างคำสั่งการเรียกใช้งาน PESQ เพื่อวัดคุณภาพเสียง pesq +8000 *input_sound** *output_sound**

*ไฟล์เสียงประเภท .WAV PESQ จะข้าม 44-byte แรกที่เป็น Header ไป และจะถือว่าไม่มี Header สำหรับไฟล์ประเภทอื่น

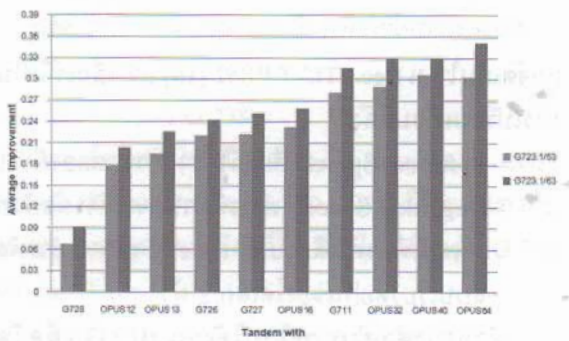
4. ผลการทดลอง

ในงานวิจัยนี้ได้ทำการวัดคุณภาพเสียงตามขนาด Frame Size ของ OPUS ที่มีจำนวนบิตเรตใกล้เคียงกับจำนวนบิตเรตของโคเดกมาตรฐานที่เราใช้ทดลอง นั่นคือ 6, 7, 8, 10, 12, 13, 16, 24, 32, 40, 64 และ 65 kbps ผลที่ได้แสดงในรูปที่ 3

จากรูปที่ 5 แสดงให้เห็นว่าค่าคุณภาพของเสียงเพิ่มขึ้นเมื่อ Frame Size เท่ากับ 20 มิลลิวินาทีหรือสูงกว่า และ Frame Size ที่มีขนาดตั้งแต่ 20 มิลลิวินาทีขึ้นไปให้ค่าคุณภาพเสียงที่ใกล้เคียงกันมาก ดังนั้นการทดลองนี้จึงใช้ Frame Size คงที่เท่ากับ 20 มิลลิวินาที



รูปที่ 5 ค่าคุณภาพของเสียง (MOS) ที่ Frame Size ต่าง ๆ



รูปที่ 6 ค่าเฉลี่ยคุณภาพเสียงที่เพิ่มขึ้นเมื่อเข้ารหัสแบบต่อเนื่องกันด้วยโคเดกชนิดต่าง ๆ

จากการทดลองพบว่า คุณภาพเสียงดีขึ้นเมื่อมีการเรียงต่อกันของโคเดก ซึ่งโคเดกที่ช่วยให้ค่าคุณภาพเสียงเพิ่มขึ้น คือ G711, G726, G727, G728 และ OPUS แต่ในบรรดาโคเดกทั้งหมด OPUS สามารถเพิ่มคุณภาพของเสียงได้มากที่สุด ค่าเฉลี่ยของคุณภาพเสียงที่เพิ่มขึ้นเมื่อเข้ารหัสแบบต่อเนื่องกันด้วยโคเดกชนิดต่าง ๆ แสดงในรูปที่ 6

ผลการทดลองสามารถใช้สร้างสมการทำนายค่าคุณภาพของเสียงที่ผ่านการเข้ารหัสด้วยโคเดกที่เรียงต่อกับ G.723.1A โดยใช้สมการความสัมพันธ์ 3rd Order Polynomial Regression ที่มีสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination-R^2) ไม่น้อยกว่า 0.80 (ยกเว้น OPUS 40) ดังแสดงในตารางที่ 3

ตารางที่ 3 สมการความสัมพันธ์ 3rd Order Polynomial Regression

ลำดับการเข้ารหัสแบบเรียงต่อกัน	สมการ 3rd Polynomial Regression	ค่า R^2
G711 * G723/53	$y = -6.5103x^3 + 62.015x^2 - 195.66x + 207.87$	0.9034
G726 * G723/53	$y = -6.786x^3 + 64.995x^2 - 206.04x + 219.52$	0.9004
G727 * G723/53	$y = -2.0669x^3 + 20.423x^2 - 66.211x + 73.86$	0.819
G728 * G723/53	$y = -3.9048x^3 + 37.698x^2 - 120.15x + 129.65$	0.9254
OPUS/12 * G723/53	$y = -3.2788x^3 + 31.668x^2 - 100.95x + 109.56$	0.8446
OPUS/13 * G723/53	$y = -3.4671x^3 + 33.171x^2 - 104.76x + 112.55$	0.9159
OPUS/16 * G723/53	$y = -4.3888x^3 + 41.77x^2 - 131.42x + 140.05$	0.9015
OPUS/32 * G723/53	$y = -4.0241x^3 + 38.972x^2 - 124.72x + 135.35$	0.8372
OPUS/40 * G723/53	$y = x + (-7.5644x^3 + 72.55x^2 - 232.08x + 247.87)$	0.7697
OPUS/64 * G723/53	$y = -2.6836x^3 + 26.11x^2 - 83.627x + 91.65$	0.8941
G711 * G723/63	$y = -4.2241x^3 + 42.393x^2 - 140.84x + 158.49$	0.8824
G726 * G723/63	$y = -1.8313x^3 + 19.408x^2 - 67.424x + 80.456$	0.8942
G727 * G723/63	$y = -6.181x^3 + 61.672x^2 - 203.93x + 226.98$	0.8912
G728 * G723/63	$y = -1.7147x^3 + 17.924x^2 - 61.431x + 72.505$	0.8428
OPUS/12 * G723/63	$y = -4.2294x^3 + 42.786x^2 - 143.23x + 162.15$	0.9062
OPUS/13 * G723/63	$y = -3.5811x^3 + 36.183x^2 - 120.85x + 136.96$	0.9161
OPUS/16 * G723/63	$y = -4.0564x^3 + 40.94x^2 - 136.67x + 154.48$	0.8334
OPUS/32 * G723/63	$y = -3.8352x^3 + 38.789x^2 - 129.83x + 147.43$	0.8287
OPUS/40 * G723/63	$y = x + (-3.9958x^3 + 41.118x^2 - 141.29x + 162.38)$	0.8422
OPUS/64 * G723/63	$y = -2.4876x^3 + 25.591x^2 - 86.777x + 100.7$	0.8773

ตัวอย่างการพยากรณ์ค่าคุณภาพเสียงโดยใช้สมการความสัมพันธ์ในตารางที่ 3 เช่น ค่าคุณภาพเสียงของ G.723.1A 6.3 kbps คือ 3.5 (ค่า X) เราสามารถทำนายค่าคุณภาพเสียงใหม่ที่ได้หากใช้วิธีเรียงต่อกันของโคเดกระหว่าง OPUS 32 kbps และ G.723.1A 6.3 kbps (ค่า Y) คือ $-3.8352(3.5)^3 + 38.789(3.5)^2 - 129.83(3.5) + 147.43 = 3.75$ ซึ่งค่าที่ได้นี้มีสัมประสิทธิ์ของการตัดสินใจ เท่ากับ 0.8287 หรือ 82.87% (ยิ่งค่า R^2 มีมากเท่าใดยิ่งบอกถึงความแม่นยำในการนำสมการมาใช้เพื่อพยากรณ์ย่อมมีสูงมากขึ้น)

5. อภิปรายผลและสรุป

งานวิจัยฉบับนี้ศึกษาเกี่ยวกับการพัฒนาคุณภาพของเสียงที่ผ่านการเข้ารหัสด้วยโคเดกสองตัวให้มีค่าสูงขึ้นโดยใช้เทคนิคการนำโคเดกตัวอื่นมาต่อไว้ด้านหน้าของโคเดกที่เราต้องการเพิ่มคุณภาพเสียง จากการทดลองพบว่า 1) เพื่อให้ได้คุณภาพเสียงที่ดีที่สุดสำหรับ OPUS ค่า Frame Size ควรใช้ที่มีขนาดมากกว่าหรือเท่ากับ 20 มิลลิวินาทีขึ้นไป และ 2) เสียงที่ผ่านการเข้ารหัสด้วยโคเดกตัวอื่นทุกชนิดที่ใช้ในการทดลอง (ยกเว้น GSM-FR) แล้วจึงตามด้วย G723.1A มีค่าคุณภาพของเสียงที่ดีกว่าเมื่อเปรียบเทียบกับการเข้ารหัสด้วย G723.1A เพียงครั้งเดียว และเมื่อเปรียบเทียบค่าคุณภาพของเสียงที่เพิ่มขึ้นเมื่อมีการเข้ารหัสด้วยโคเดกต่างๆ ข้างต้น พบว่าในบรรดาโคเดกทั้งหมด OPUS สามารถเพิ่มคุณภาพของเสียงได้อย่างมีประสิทธิภาพที่สุด ดังแสดงในรูปที่ 6 ซึ่งผลลัพธ์แสดงให้เห็นว่าคุณภาพของเสียงโดยเฉลี่ยเพิ่มขึ้นสูงที่สุดถึง 10.4 เปอร์เซนต์โดยประมาณ

เมื่อพิจารณาจากผลลัพธ์ของการทดลองจะพบว่าโคเดกทุกตัวที่ช่วยเพิ่มคุณภาพเสียงของ G723.1A นั้นล้วนแล้วแต่มีจำนวนบิตเรตที่มากกว่าทั้งสิ้น ดังนั้นการนำ OPUS เข้ามาใช้งานจึงถือเป็นตัวเลือกที่เหมาะสมเพราะมีความยืดหยุ่นที่สามารถปรับค่าบิตเรตได้อย่างอิสระ และทั้งนี้การเข้ารหัสแบบเรียงต่อกันโดยใช้ OPUS ก่อนนั้นยังแสดงให้เห็นว่าสามารถให้คุณภาพเสียงที่ดีกว่าการเข้ารหัสด้วยวิธีเดียวกันด้วยโคเดกตัวอื่นๆ ที่มีบิตเรตเท่ากันอีกด้วย งานวิจัยฉบับนี้ตั้งข้อสังเกตุว่า เนื่องจาก OPUS เกิดจากการรวมกันของ 2 อัลกอริทึม (Linear Prediction และ Modified Discrete Cosine Transform) หรือ 2 โคเดก (SILK และ CELT) ทำให้เสียงที่ได้จากการถอดรหัสด้วย OPUS Decoder เป็นการนำผลลัพธ์ที่ได้จาก SILK รวมกับ CELT และมีการปรับความละเอียดเสียงให้เหมาะสมในระหว่างการทำงานของตัวถอดรหัส [19] ดังนั้น คุณภาพเสียงที่ได้จึงมีคุณภาพเสียงที่ดีกว่าโคเดกตัวอื่นๆ และเมื่อนำไปต่อเข้ากับ G723.1A ที่ใช้อัลกอริทึม Linear Prediction เหมือนกันจึงเหมาะสมกว่าตัวอื่นๆ

โดยคุณภาพของเสียงที่พัฒนาขึ้นสามารถทำนายได้จากสมการความสัมพันธ์ 3rd Order Polynomial Regression ดังแสดงไว้ในตารางที่ 3 สมการที่ได้นี้สามารถนำมาใช้เพื่อประเมินความเหมาะสมในการนำโคเดกมาต่อกันโดยสามารถทำนายได้ว่า การต่อกันของโคเดกชนิดใดจะให้ผลลัพธ์ที่ดีที่สุด หรือในกรณีใดผลลัพธ์ที่ได้จะไม่คุ้มค่า ทั้งนี้เป็นการช่วยประหยัดเวลาและค่าใช้จ่ายในการทดลองเข้ารหัสเพื่อวัดคุณภาพเสียงด้วยการใช้วิธีต่างๆ เช่น PESQ โดยค่าที่ได้จากการคำนวณนี้มีความถูกต้องมากกว่า 80%

งานวิจัยนี้สามารถนำไปสู่การศึกษาเพิ่มเติมในเรื่องของการนำ OPUS มาช่วยพัฒนาคุณภาพเสียงที่ผ่านการเข้ารหัสด้วยโคเดกตัวอื่นนอกเหนือจาก G.723.1A และเนื่องจาก OPUS สามารถรองรับการใช้งานได้ตั้งแต่ช่วงความถี่แคบ (Narrow-band) ไปจนถึงช่วงเสียงเพลง (Music) ดังนั้นงานวิจัยนี้จึงสามารถนำไปสู่การศึกษาเพิ่มเติมกับโคเดกที่อยู่ในช่วงแถบความถี่กว้าง (Wide-band) หรือแถบความถี่เสียงเพลง

เอกสารอ้างอิง

[1] J.K. Ponder, "Next Generation Networks," in *ICT Trends and Challenges in a Global Era*, 4th July 2005.

[2] S.A. Ahson and M. Ilyas, *VoIP Handbook: Applications, Technologies, Reliability and Security*, CRC Press Taylor & Francis Group, 2008, pp.140.

[3] J.D. Gibson, "Speech coding methods, standards, and applications," *Circuits and Systems Magazine, IEEE*, Vol.5, pp.30-49, 2005.

[4] A. Raake, *Speech Quality of VoIP: Assessment and Prediction*, England: John Wiley & Sons Ltd., 2006, pp. 58-60.

[5] W.C. Hardy, *VoIP Service Quality: Measuring and Evaluating Packet-Switched Voice*, McGraw-Hill Companies Inc., 2003.

[6] R.V. Cox, *Speech Coding*, 2000 CEC Press LLC.

[7] E. Nemer, (2012, December 30). *Handling VoIP Speech Coding Challenges: Part1*, 2002, voip-info.org [Online]. Available: http://www.eetimes.com/design/other/4009262/Handling-VoIP-Speech-Coding-Challenges-Part-1

[8] J.D. Gibson, "Tandem voice communications: digital cellular, VoIP, and voice over Wi-Fi," in *Global Telecommunications Conference, GLOBECOM '04. IEEE*, 2004, vol.2, pp.617-621.

[9] C. Redding, N. DeMinco, and J. Lindner, *Voice Quality Assessment of Vocoders in Tandem Configuration*, Springfield, VA: U.S. Dept. of Commerce, National Telecommunications and Information Administration, 2001.

[10] O. Hersent, *IP Telephony: Deploying VoIP Protocols and IMS Infrastructure*, John Wiley and Sons Ltd, 2011.

[11] N. Unuth, (2012, December 2). "VoiP Codecs," About.com Guide [Online]. Available: http://voip.about.com/od/voipbasics/a/voipcodecs.htm

[12] "VoiP Codec-general overview," (2012, December 2). voip-sip.org [Online]. Available: http://www.voip-sip.org/voip-codecs/

[13] J.M. Valin, K.Vos, and T. Terriberry, "Definition of the opus audio codec," in *IETF draft*, September 2012, pp. 5-8.

[14] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, England: John Wiley & Sons Ltd., 2006, pp. 25-28 and pp. 53-57.

[15] ITU-T, Recommendation P.800, "Methods for subjective determination of transmission quality," August 1996.

[16] ITU-T, Recommendation P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," February 2001.

[17] ITU-T, Recommendation P.862.1 "Mapping function for transforming P.862 raw result scores to MOS-LQO," November 2003.

[18] M. Hasegawa-Johnson and A. Alwan, "Speech Coding: Fundamentals and Applications. Encyclopedia of Telecommunications," 2003.

[19] A. Romo and H. Toukomaa, "Voice Quality Characterization of IETF Opus Codec," Interspeech 2011, Florence, Italy (2011).

[20] ITU-T, Recommendation G.191, "Software tools for speech and audio coding standardization," March 2010.

[21] ITU-T, Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies," November 1988.

[22] ITU-T, Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," May 2006.

[23] ITU-T, Recommendation G.726, "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)," December 1990.

[24] ITU-T, Recommendation G.727, "5-, 4-, 3- and 2-bit/sample embedded adaptive differential pulse code modulation (ADPCM)," December 1990.

[25] ITU-T, Recommendation G.728, "Coding of speech at 16 kbit/s using low-delay code excited linear prediction," September 1992.

[26] ITU-T, Recommendation P.50, "Artificial voices Appendix I: Test signals," February 1998.