การศึกษาและการใช้ประโยชน์การแปลงเวฟเล็ตสำหรับการบีบอัดสัญญาณเสียง
Studying and Exploiting Wavelet Transform for Speech Compression

สุภาธิณี กรสิงห์ [1] จักรี ศรีนนท์ฉัตร[2*]
[1]คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตนครราชสีมา
[2]ห้องปฏิบัติการและวิจัยทางด้านการประมวลผลสัญญาณ
ภาควิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม คณะวิศวกรรมศาสตร์
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี อ.คลองหลวง จ.ปทุมธานี 12110
*E-mail: jakkree.s@en.rmutt.ac.th

บทคัดย่อ

งานวิจัยนี้นำเสนอการศึกษาและการใช้ประโยชน์การแปลงเวฟเล็ตสำหรับการบีบอัดสัญญาณเสียง การทดลองนี้ได้ใช้สัญญาณเสียงทั้งหมด 80 เสียง โดยแบ่งเป็น 4 กลุ่ม ดังนี้ เสียงพูดของผู้หญิงและผู้ชายที่มีความยาว 5 วินาที และ 60 วินาที อย่างละ 20 เสียง จากนั้นนำสัญญาณเสียงไปเข้ากระบวนการคัดเลือกเวฟเล็ตจากเวฟเล็ต 3 ตระกูล คือ ฮาร์เวฟเล็ต ไบออทอลโกนอลแวฟเล็ต และการประมาณค่าไม่ต่อเนื่องของเมเยอร์เวฟเล็ต เพื่อหาเวฟเล็ตที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ ทั้งนี้วิธีการคัดเลือกใช้หลักการหาค่าพลังงานเฉลี่ย การเปรียบเทียบความถี่สเปกโตรแกรม และหลักการของ ไดนามิกไทม์วอร์ปปิง เป็นตัววัดผล หลังจากนั้นนำเวฟเล็ตที่ผ่านการคัดเลือกไปทำการบีบอัดสัญญาณเสียงในระดับที่ 1-3 ซึ่งผลลัพธ์ที่ได้จะถูกนำไปเปรียบเทียบกับเทคนิคการบีบอัดสัญญาณด้วย Federal Standard 1016 Code Excite Linear Prediction (FE 1016 CELP) โดยใช้หลักการของค่าเฉลี่ยผิดพลาดกำลังสองและอัตราส่วนของสัญญาณสูงสุดเป็นตัววัดคุณภาพของการบีบอัดสัญญาณผลการทดสอบพบว่า เวฟเล็ตตระกูลไบออทอลโกนอลให้ประสิทธิภาพในการบีบอัดสูงที่สุด และในการสังเคราะห์เสียงด้วยการคืนกลับเวฟเล็ตแบบไม่ต่อเนื่อง นั้นสัญญาณเสียงของผู้หญิงที่มีความยาว 5 วินาทีให้ประสิทธิภาพเฉลี่ยดีที่สุด ในการเปรียบเทียบค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงพูดนั้นทั้งหมดด้วยการแปลงเวฟเล็ตแบบไม่ต่อเนื่องและ CELP ผลปรากฏว่าการแปลงเวฟเล็ตแบบไม่ต่อเนื่องให้ประสิทธิภาพในการบีบอัดสัญญาณเสียงดีกว่า CELP และยังมีค่าผิดพลาดน้อยกว่าเมื่อนำสัญญาไปเปรียบเทียบกับสัญญาณต้นฉบับ

คำสำคัญ: ฮาร์เวฟเล็ต ไบออทอลโกนอลเวฟเล็ต การประมาณค่าไม่ต่อเนื่องของเมเยอร์เวฟเล็ต การบีบอัดสัญญาณเสียงพูด

Abstract

Recently, speech compression research aims to produce a compact representation of speech sounds such that when reconstructed it is perceived to be close to the original. This thesis presents a studying and comparison of wavelet filter for speech compression.

In the experiments, there are 80 speech signals which are used as input data. These signals can be categorized into 4 groups that consist of male and female speech signal with the length of 5 and 60 seconds respectively. These signals are then pass through to the three types of Wavelet Transform: Haar wavelet, Biorthogonal wavelet and Discrete Approximation of Meyer Wavelet, in order to search the best appropriate for this experiment. To classify wavelet, the energy average, spectrogram and Dynamic Time Warping (DTW) are used. The best appropriate wavelet is then used to compress speech signal in level 1-3. The result of this experiment is then compared with the Federal Standard 1016 Code Excite Linear Prediction (FS 1016 CELP) in the term of speech quality using Means Square Error (MSE) and Peak Signal to Noise Ratio (PSNR).

The results show that Biorthogonal Wavelet provides the best compress efficiency. Also the synthesis speech signal with Invest Discrete Wavelet Transform (IDWT) indicated that the 5 seconds female speech signal provides the best average efficiency. Moreover, the DWT and CELP speech compression is compared in the term of PSNR and MSE. The results show that DWT provides better performance than CELP speech compression and also it gives the errorless when it compares to the original speech signal.

Keywords: Haar wavelet, Biorthogonal wavelet, Discrete approximation of meyer wavelet, Speech signal compression

บทนำ

สัญญาณเสียงพูด (Speech Signal) [1] นั้น เป็นสัญญาณที่มีแถบความถี่ต่ำที่สุด จึงเป็นที่สนใจในการนำมาประมวลผลสัญญาณแบบเวลาจริง (Real Time) เพื่อนำไปพัฒนาต่อเทคโนโลยีต่างๆ และในงานทางการสื่อสารอีกเป็นจำนวนมาก แต่ปัญหาที่สำคัญของสื่อสารด้วยสัญญาณเสียงพูดคือ การที่ระบบสื่อสารนั้นมีแบนด์วิดท์ที่มีจำกัด (Limited Bandwidth) ดังนั้นการพัฒนาให้การประมวลสัญญาณเสียงพูดมีความเร็วแบบเวลาจริงนั้นจึงจำเป็นต้องลดขนาดของสัญญาณลงหรือเรียกว่าการบีบอัดสัญญาณเสียงพูด (Speech Compression) เพื่อให้เหมาะสมกับแบนด์วิดท์ที่มีอย่างจำกัด และเมื่อนำเสียงพูดที่ถูกบีบอัดไปแล้วนั้น มาทำการคืนกลับสัญญาณเสียงหรือการสังเคราะห์สัญญาณเสียงขึ้นมาใหม่ อาจทำให้คุณภาพของสัญญาณเสียงพูดที่ได้มีคุณรูปที่ต่ำลงมาก เนื่องจากการสูญเสียข้อมูลจำนวนมากขณะบีบอัดสัญญาณ ซึ่งในปัจจุบันนี้เทคนิคที่ใช้สำหรับการลดขนาดหรือการบีบอัดสัญญาณเสียงพูดมีหลากหลายเทคนิค ซึ่งแต่ละเทคนิคนั้นจะมีข้อดีและข้อเสียที่แตกต่างกัน และจากการศึกษาพบว่าการบีบอัดสัญญาณด้วยการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform; DWT) เป็นเทคนิคได้รับความสนใจมากในปัจจุบัน

เทคนิค DWT คือ กระบวนการบีบอัดสัญญาณหรือการวิเคราะห์สัญญาณ โดยใช้ตัวกรองความถี่ ซึ่งโดยปกติสัญญาณเสียงมนุษย์จะมีทั้งความถี่ต่ำและความถี่สูง ดังนั้นเมื่อนำสัญญาณเสียงเข้าสู่กระบวนการบีบอัดด้วยเวฟเล็ตสัญญาณจะถูกแบ่งออกเป็น 2 ส่วน คือ การประมาณค่า (Approximation) และรายละเอียด (Detail) ค่าการประมาณจะเป็นส่วนของความถี่ต่ำ และค่ารายละเอียดจะเป็นส่วนของความถี่สูง แต่สำหรับการกรองความถี่ของสัญญาณเสียงพูดนั้น ส่วนที่สำคัญที่สุดก็คือ ส่วนที่เป็นความถี่ต่ำ เนื่องจากเสียงมนุษย์นั้นเมื่อนำส่วนความถี่สูงออกเสียงจะเปลี่ยนแปลงไป แต่ยังคงรู้ว่าผู้พูดต้องการสื่อถึงอะไร แต่ถ้าหากนำส่วนความถี่ต่ำออกไปนั้น จะกลายเป็นเสียงพึมพำเท่านั้น

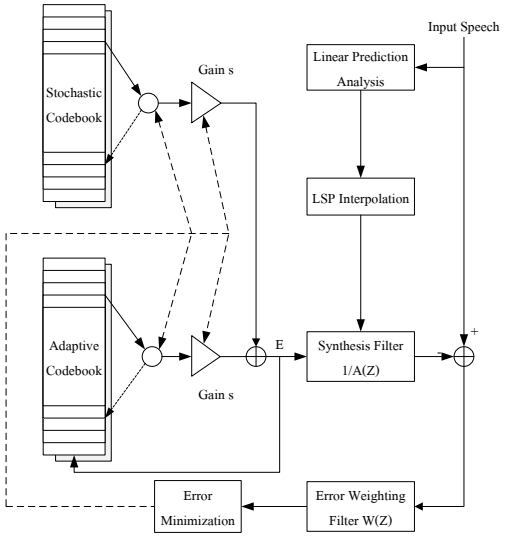
บทความนี้ได้นำเทคนิคการบีบอัดสัญญาณเสียงด้วย DWT ไปประยุกต์ใช้กับเสียงทั้งหมด 80 เสียง และได้นำเทคนิค CELP (Code Excite Linear Prediction) [2-3] มาเปรียบเทียบเพื่อหาคุณภาพของการบีบอัดสัญญาณเสียงพูดหลังจากการคืนกลับสัญญาณหรือการสังเคราะห์สัญญาณนั้นๆ ซึ่ง CELP เป็นเทคนิคที่ได้รับการยอมรับในมาตรฐานตามสากลระหว่างประเทศซึ่งแบ่งออกเป็น 4 มาตรฐาน คือ มาตรฐาน Linear Predict Coefficient-10 (LPC-10) มาตรฐาน Vector Sum Excited Linear Prediction (VSELP) มาตรฐาน Low Delay Code Excited Linear Prediction (LD-CELP) และ มาตรฐาน Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP) แต่ในบทความนี้ได้เลือกใช้การบีบสัญญาณเสียงด้วย CELP ในมาตรฐานของ FS1016 CELP เพื่อนำไปเทียบคุณภาพกับเทคนิค DWT

วิธีการทดลอง

ทฤษฎีการบีบอัดสัญญาณเสียงพูด

การเข้ารหัสสัญญาณเสียงพูดด้วย CELP

CELP ย่อมาจาก Code Excite Linear Prediction [2-3] เป็นเทคโนโลยีที่ใช้สำหรับการบีบอัดสัญญาณเสียงซึ่งได้รับมาตรฐานมากในระดับสากล และได้ถูกนำไปประยุกต์ใช้งานทางด้านงานประมวลผลมากมาย ซึ่งมาตรฐานของ CELP ยังสามารถแยกออกเป็นมาตรฐานย่อยอื่นๆ อีก แต่สำหรับบทความนี้จะศึกษาเฉพาะมาตรฐานของสถาปัตยกรรม Federal Standard 1016 CELP (FS1016 CELP) [4] ซึ่งมีการพัฒนาร่วมกันระหว่างสหรัฐอเมริกาและห้องปฏิบัติการของ AT & Bell สามารถบีบอัดสัญญาณเสียงพูดที่ 4.8 กิโลบิตต่อวินาที และจะเข้ารหัสโดยการแบ่งสัญญาณเสียงพูดเป็นเฟรม (Frame) ซึ่งในหนึ่งเฟรมจะใช้อัตราการสุ่มตัวอย่าง (Sample Rate) 8 kHz และขนาดเฟรม (Frame Size) 30 วินาที (ประมาณ 240 ตัวอย่างต่อ 1 เฟรม)



รูปที่ 1 บล็อกไดอะแกรมการเข้ารหัส FS 1016 CELP [2]

จากรูปที่ 1 เป็นการเข้ารหัสของ FS1016 CELP เริ่มจากการนำสัญญาณเสียงแปลงจากสัญญาณอนาล็อกเป็นดิจิตอล หลังจากนั้นสัญญาณจะถูกนำไปเข้ากระบวนการเข้ารหัส CELP บนพื้นฐานการวิเคราะห์ด้วยการสังเคราะห์หาค่าถ่วงน้ำหนักของเวกเตอร์คอนไตซ์เซชั่น (Vector Quantization; VQ) และจากการทำนายพันธะเชิงเส้น (Linear Prediction) ซึ่งมีการกระตุ้นของสัญญาณด้วย Codebook ที่ใช้งาน 2 ส่วน คือ Adaptive Codebook และ Stochastic Codebook โดยAdaptive Codebook จะถูกนำมา

สร้างสัญญาณกระตุ้นความล่าช้าและ Stochastic Codebook จะถูกนำมาเพื่อแสดงความแตกต่างระหว่างรูปแบบของคลื่นที่เกิดขึ้นจริงและส่วนขยายระยะที่เหมาะสมของการกระตุ้น

การทำนายพันธะเชิงเส้น (Linear Predictive) เป็นเทคนิคที่สำคัญทางด้านการวิเคราะห์เสียงเนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของเสียงพูดเมื่อเทียบกับความเร็วในการประมวลผล หลักการพื้นฐานของการทำนายพันธะเชิงเส้นอาศัยแนวความคิดว่าตัวอย่างสัญญาณเสียงพูดสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณเสียงพูดจากอดีต การวิเคราะห์หาพารามิเตอร์เพื่อใช้ในการทำนายโดยทั่วไปเรียกว่าการเข้ารหัสการทำนายพันธะเชิงเส้น (Linear Predictive Coding; LPC) ในด้านการประมวลผลสัญญาณเสียง การเข้ารหัสการทำนายพันธะเชิงเส้นถูกนำไปใช้ในสองแนวทาง ได้แก่

ก. การเข้ารหัสสัญญาณเสียง เป็นวงจรกรองวิเคราะห์การทำนายพันธะเชิงเส้น (LP Analysis Filter) เพื่อแยกส่วนที่มีการซ้ำซ้อน (Redundancy) ของสัญญาณเสียงออก ส่วนที่เหลือเรียกว่าสัญญาณตกค้าง (Residual Signal)

ข. การสังเคราะห์สัญญาณเสียงพูด เป็นวงจรกรองการทำนายพันธะเชิงเส้นผกผัน (Inverse LP Filter) หรือวงจรกรองสังเคราะห์การทำนายพันธะเชิงเส้น (LP Synthesis Filter) โดยที่ฟังก์ชั่นถ่ายโอนของวงจรกรองดังกล่าวแสดงกรอบสเปกตรัมของสัญญาณเสียงพูด วงจรกรองสังเคราะห์การทำนายพันธะเชิงเส้นแสดงช่องทางเดินเสียงของมนุษย์และใช้หาสัญญาณกระตุ้นที่เหมาะสม แต่เนื่องจากพารามิเตอร์ LPC มีความเสถียรของสัญญาณที่ต่ำ ซึ่งอาจส่งผลต่อการไปคืนกลับสัญญาณเสียงได้ ดังนั้นจึงได้นำค่าพารามิเตอร์จาก LPC ไปพัฒนาต่อด้วย LSP

คู่เส้นสเปกตรัม (Line Spectral Pairs; LSP) หรือความถี่เส้นสเปกตรัม (Line Spectral Frequency; LSF) เป็นพารามิเตอร์รูปแบบหนึ่งที่พัฒนามาจากพารามิเตอร์การทำนายพันธะเชิงเส้น เนื่องจากพารามิเตอร์การทำนายพันธะเชิงเส้นในขั้นตอนการประมาณค่าพารามิเตอร์อาจทำให้เกิดความไม่เสถียรของสัญญาณได้ ซึ่งส่งผลต่อคุณภาพของเสียง ในขณะที่พารามิเตอร์คู่เส้นสเปกตรัมมีคุณสมบัติที่เด่นคือ ค่าพารามิเตอร์อยู่ภายในขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์ และสามารถตรวจสอบเสถียรภาพของวงจรกรองได้ง่าย นอกจากนี้คู่เส้นสเปกตรัมยังแสดงในรูปเชิงความถี่จึงสามารถนำไปใช้ในการหาคุณสมบัติที่แน่นอนในระบบการรับรู้ของคนได้

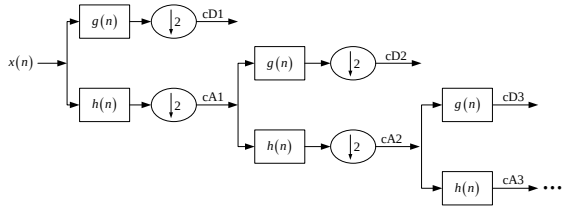
การแปลงเวฟเล็ต

การแปลงเวฟเล็ต (Wavelet Transform)[4-7] เป็นคณิตศาสตร์ที่ใช้ในการวิเคราะห์และสังเคราะห์ลักษณะของสัญญาณซึ่งมีประโยชน์มากในงานทางด้านการประมวลผลสัญญาณ (Signal Processing) ปัจจุบันนี้เวฟเล็ตที่นำมาใช้งานทางด้านการประมวลผลสัญญาณมีหลายตระกูล ที่สามารถเลือกใช้ได้เหมาะสมสำหรับแต่ละงาน โดยสามารถแบ่งเป็น Biorthogonal Wavelet และเวฟเล็ตเชิงตั้งฉากปกติ เช่น Daubechies, Symlet และ Coiflet เป็นต้น แต่ละตระกูลจะเป็นฟังก์ชันพื้นฐานที่มีรูปร่างลักษณะที่แตกต่างกันไป ซึ่งแต่ละตระกูลจะมีค่า Number of Vanishing Moments (NVM) ท้ายชื่อตระกูลเช่น Daubechies 4, 8,..., 20, Symlet 4, 5, 6, ..., 10 และ Coiflet 1, 2, ..., 5 เป็นต้น ถ้าค่า NVM มีค่ามากขึ้นลักษณะของฟังก์ชันพื้นฐานที่เลือกจะมีความราบเรียบ (Smooth) มากขึ้น ประโยชน์ของค่า NVM นี้ก็คือสามารถเลือกชนิดของเวฟเล็ตมาประยุกต์ใช้กับงานที่ต้องการได้อย่างหลากหลายและเหมาะสมมากขึ้น ตระกูลของเวฟเล็ตแม่ที่สำคัญและนิยมใช้กันอย่างแพร่หลายในปัจจุบัน

การแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform) [1] ในทางปฏิบัติแล้วสัญญาณที่นำมาวิเคราะห์โดยคอมพิวเตอร์จะมีลักษณะแบบแบ่งช่วง หรือเป็นสัญญาณที่ถูกแซมปลิ้ง (Sampling) เข้ามา ดังนั้นจึงได้มีพัฒนาการวิเคราะห์การแปลงเวฟเล็ตแบบไม่ต่อเนื่องขึ้น (Discrete Wavelet Transform: DWT) โดยจะใช้ตัวกรองในกระบวนการแปลง ซึ่งถูกพัฒนาขึ้นโดย Mallat ในปี 1988 ระเบียบวิธีการของ Mallat เป็นแบบแผนที่รู้จักกันในกลุ่มผู้ที่ทำการแปลงสัญญาณว่าเป็นการเข้ารหัสแบบ 2 ช่องสัญญาณย่อย (Two-channel Sub band Coder)

สำหรับหลายๆ สัญญาณจะมีส่วนความถี่ต่ำเป็นส่วนสำคัญมากที่สุด ซึ่งจะให้ลักษณะของสัญญาณ ในทางตรงกันข้าม ส่วนความถี่สูงจะบอกความแตกต่างกัน เมื่อพิจารณาเสียงมนุษย์ ถ้าคุณย้ายส่วนความถี่สูงออก เสียงจะต่างออกไป แต่ยังคงรู้ว่าพูดอะไร อย่างไรก็ตามถ้าหากคุณย้ายส่วนความถี่ต่ำออกไปมากพอแล้วจะกลายเป็นเสียงพึมพำเท่านั้น ดังนั้นการวิเคราะห์เวฟเล็ตจึงมักพูดถึงการประมาณค่า (Approximation) และรายละเอียด (Detail) ซึ่งค่าการประมาณจะเป็นส่วนของความถี่ต่ำ และค่ารายละเอียดจะเป็นส่วนของความถี่สูง ถ้าสร้างสัญญาณการแปลงสัญญาณเวฟเล็ตแบบแบ่งช่วงในขั้นตอนเดียว (1 ระดับการแปลง) เมื่อสัญญาณอินพุท 1000 แซมเปิ้ล โดย cA จะเป็นค่าการประมาณ และ cD จะเป็นค่ารายละเอียด

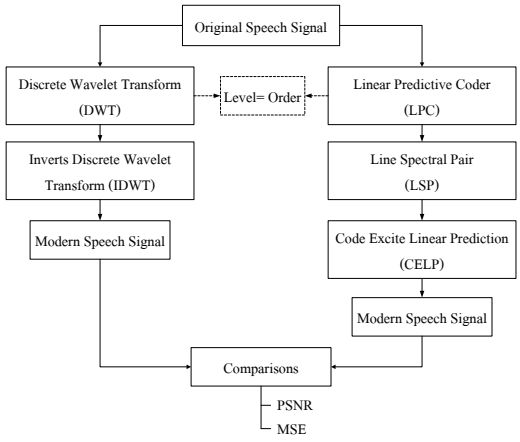
เมื่อพิจารณาการแยกส่วนประกอบหลายๆ ระดับโดยการแยกส่วนประกอบที่สามารถกระทำซ้ำด้วยการประมาณค่าที่ต่อเนื่องตามระดับที่ต้องการ เช่น การแปลงเวฟเล็ตแบบแบ่งช่วงที่ 3 ระดับการแปลงดังรูปที่ 2



รูปที่ 2 การสร้างสัญญาณการแปลงสัญญาณเวฟเล็ตแบบแบ่งช่วงในขั้นตอนเดียว

บทความนี้นำเสนอเทคนิคการบีบอัดสัญญาณเสียงพูด 2 เทคนิค โดยขั้นตอนต่างๆ จะประกอบด้วย การเตรียมสัญญาณเสียงพูดก่อนการบีบอัดสัญญาณ กระบวนการบีบอัดเสียง กระบวนการคืนกลับสัญญาณจากการบีบอัดสัญญาณเสียงพูด และการเปรียบเทียบคุณภาพของสัญญาณเสียงใหม่ที่ใกล้เคียงกับสัญญาณต้นฉบับโดยเปรียบเทียบกันทั้ง 2 เทคนิค ซึ่งขั้นตอนต่างๆ จะแสดงในรูปที่ 3

การเลือกตระกูลเวฟเล็ต เป็นกระบวนการการคัดเลือกเวฟเล็ตเพื่อนำไปใช้ในการบีบอัดสัญญาณเสียง เพื่อเลือกเวฟเล็ตที่เหมาะสมที่สุดสำหรับบทความนี้ โดยจะนำเวฟเล็ตที่ผ่านการคัดเลือกจะถูกบีบอัดสัญญาณเสียงและคืนกลับสัญญาณ เพื่อเปรียบเทียบกับเทคนิค LPC



รูปที่ 3 ขั้นตอนการทำงาน

บทความนี้ใช้การบีบอัดสัญญาณเสียงด้วยวิธี DWT ระดับที่ 2 โดยใช้เวฟเล็ตทั้งหมด 3 ตระกูล คือ Haar wavelet, Biorthogonal wavelet และ Discrete Approximation of Meyer Wavelet ซึ่งในขั้นตอนแรกคือกระบวนการบีบอัดสัญญาณเสียงโดยใช้เวฟเล็ต จะทำการแยกสัญญาณออกเป็นค่าสัมประสิทธิ์ 2 ค่า คือ ค่าสัมประสิทธิ์การประมาณ (Approximate Coefficient) และค่าสัมประสิทธิ์รายละเอียด (Detail Coefficient) ซึ่งการเปรียบเทียบคุณภาพของการบีบอัดสัญญาณเสียงพูดของเวฟเล็ตแต่ละตระกูลสามารถทดสอบและตัดสินใจดังนี้ [8-9]

เตรียมสัญญาณสียงพูด

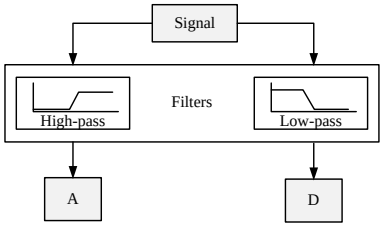
สัญญาณเสียงพูดที่ใช้สำหรับงานวิจัยนี้มีทั้งหมด 80 เสียง ซึ่งเป็นสัญญาณเสียงภาษาอังกฤษและแต่ละสัญญาณเสียงเป็นประโยคคำพูดที่แตกต่างกัน และจะมีอัตราการสุ่มตัวอย่าง (Sampling Rate) อยู่ที่ 8000 Hz โดยข้อมูลมีขนาด 8 บิต ใช้ช่องสัญญาณเดียว (Mono) โดยรูปแบบจะเป็นมาตรฐานของระบบพีซีเอ็ม (Pulse-Code Modulation: PCM) ซึ่งสามารถแบ่งกลุ่มตามเวลาและแยกเป็นสัญญาณเสียงของผู้ชายและผู้หญิงได้ดังนี้

1) เสียงผู้ชายที่เวลา 5 วินาที และ 60 วินาที อย่างละ 20 เสียง

2) เสียงผู้หญิงที่เวลา 5 วินาที และ 60 วินาที อย่างละ 20 เสียง

การบีบอัดสัญญาณเสียงพูดเบื้องต้นเพื่อคัดเลือกเวฟเล็ตที่เหมาะสม

งานวิจัยนี้ใช้การบีบอัดสัญญาณเสียงด้วยวิธี DWT ระดับที่ 2 ซึ่งเวฟเล็ตที่นำมาทำการวิจัยทั้งหมด 3 ตระกูล คือ เวฟเล็ต Haar เวฟเล็ต Biorthogonal และ เวฟเล็ต Discrete Approximation of Meyer ซึ่งในกระบวนการบีบอัดสัญญาณเสียงโดยใช้เวฟเล็ตนั้น จะทำการแยกสัญญาณออกเป็นค่าสัมประสิทธิ์สองค่าคือค่าสัมประสิทธิ์การประมาณ (Approximate Coefficient) และค่าสัมประสิทธิ์รายละเอียด (Detail Coefficient) โดยนำข้อมูลเหล่านี้ไปประมวลผลในกระบวนการอื่นต่อไปดังรูปที่ 4



รูปที่ 4 การแยกองค์ประกอบของความถี่ต่ำและความถี่สูงของการแปลงเวฟเล็ต [8]

การปรับปรุงสัญญาณเสียงพูดใหม่

นำสัญญาณเสียงพูดใหม่ที่ได้มาปรับปรุงอัตราการสุ่มตัวอย่าง (Sample Rate) เพื่อให้ได้คุณภาพของสัญญาณเสียงที่ดีขึ้นดังสมการที่ 1

$$x_{new} = \frac{p}{q} \times sampling\ rate_{new} \times t_{sec} \qquad (1)$$

โดยที่ x_{new} คือ สัญญาณเสียงพูดใหม่

t_{sec} คือ เวลาในหน่วยของวินาที (ในการทดลองนี้คือ 5 วินาที และ 6 วินาที)

$\frac{p}{q}$ คือ อัตราส่วนที่ใช้ในการ Resample

ในงานวิจัยนี้จะใช้อัตราส่วน (p/q) ทั้งหมด 5 อัตราส่วน คือ 3/2, 4/2, 5/2, 3 และ 4 ซึ่งค่า p/q ทั้งหมดนี้ได้ผ่านการทดลองเพื่อหาค่าที่เหมาะสมแล้วโดยอาศัยเวลาเป็นเกณฑ์อ้างอิง ซึ่งค่าที่ได้เมื่อนำไปทดสอบแล้วจะทำให้ได้ค่าสัญญาณเสียงพูดที่มีเวลาเท่าเดิมดังนั้นจากการคำนวณจึงทำให้ได้ค่า Sample Rate และค่าอัตราการบีบอัดสัญญาณเสียงพูดดังตารางที่ 1

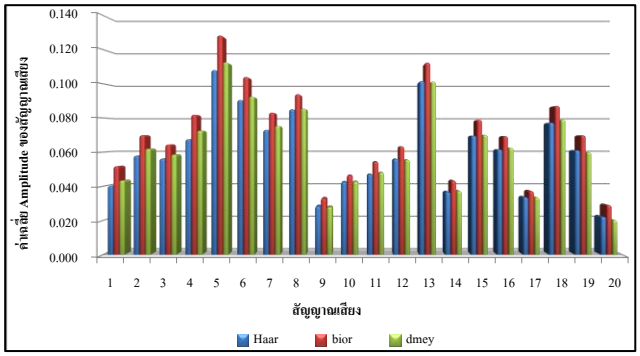
ตารางที่ 1 อัตราการสุ่มตัวอย่างและอัตราการบีบอัดสัญญาณเสียงพูดของอัตราส่วน p/q แต่ละค่า

อัตราส่วนที่	P	q	p/q	Sampling rate	Compression Rate
1	3	2	1.5	3000	2.67
2	4	2	2	4000	2.00
3	5	2	2.5	5000	1.60
4	3	1	3	6000	1.33
5	4	1	4	8000	1.00

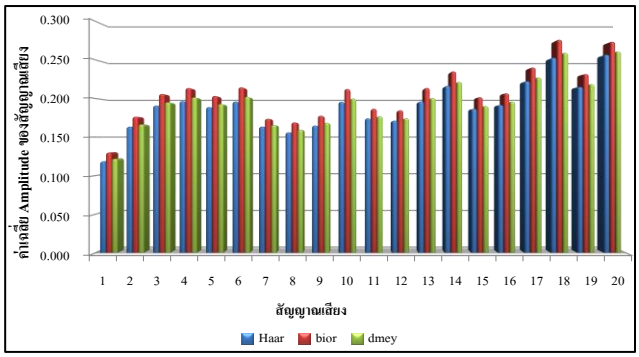
จากตารางที่ 1 จะสังเกตเห็นว่าที่ค่าอัตราส่วน p/q ที่ 3/2 หรือ 1.5 จะสามารถลดค่า Sampling Rate เท่ากับ 3000 Hz จาก 8000 Hz หรือลดลง 2.67 เท่าของสัญญาณเสียงพูดต้นฉบับ *เปรียบเทียบคุณภาพของการบีบอัดสัญญาณเสียงของเวฟเล็ตแต่ละตระกูล* ในการบีบอัดสัญญาณเสียงครั้งนี้ เพื่อหาเวฟเล็ตที่ดีที่สุดจากทั้ง 3 ตระกูล โดยวิธีทดสอบและตัดสินใจดังนี้

ก. การหาค่าพลังงานเฉลี่ยของสัญญาณเสียง โดยนำสัญญาณเสียงที่ผ่านการปรับปรุงสัญญาณเสียงทั้ง 80 เสียง คำนวณหาค่าพลังงานเฉลี่ย เพื่อเปรียบเทียบคุณภาพของสัญญาณเสียงพูดหลังจากการบีบอัดของเวฟเล็ตทั้ง 3 ตระกูล โดยวิธีการหาค่าพลังงานเฉลี่ยตามสมการที่ 2 ซึ่งถ้าเวฟเล็ตตระกูลใดมีค่าพลังงานเฉลี่ยที่มากที่สุดจะเป็นเวฟเล็ตที่ถูกเลือกมาใช้งาน ซึ่งในที่นี้ก็คือ Biorthogonal เวฟเล็ต เนื่องจากให้ค่าความผิดพลาดน้อยที่สุด และแสดงกราฟค่าพลังงานเฉลี่ยในรูปที่ 5-8

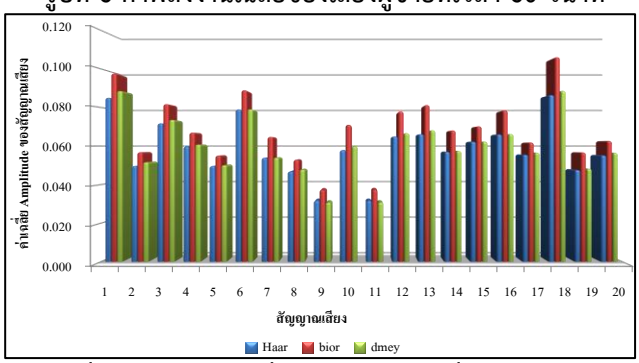
$$x = \frac{1}{n}\sum_{i=1}^{n} |S_i| \qquad (2)$$



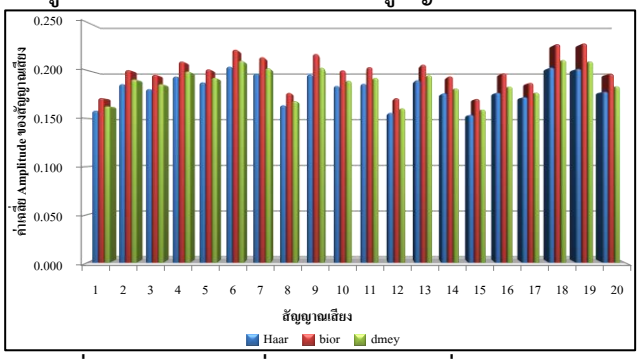
รูปที่ 5 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 5 วินาที



รูปที่ 6 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 60 วินาที

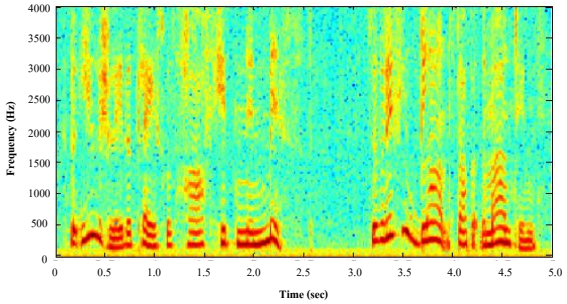


รูปที่ 7 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 5 วินาที

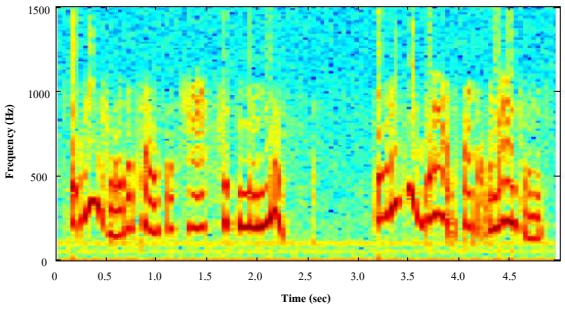


รูปที่ 8 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 60 วินาที

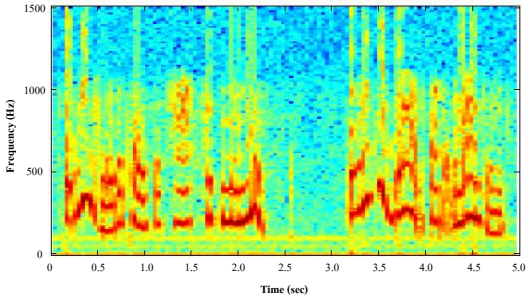
ข. สังเกตจากความถี่สเปกตรัมโดยใช้สเปกโตรแกรม (Spectrogram) ซึ่งโดยทั่วไปจะแบ่งเป็น 2 แบบคือ



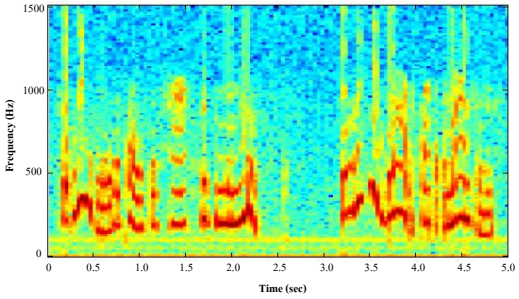
ก. Original



ข. เวฟเล็ต Haar



ค. เวฟเล็ต Biorthogonal



ง. เวฟเล็ต Discrete Approximation of Meyer

รูปที่ 9 ตัวอย่างการเปรียบเทียบความถี่สเปกตรัมของสัญญาณเสียง

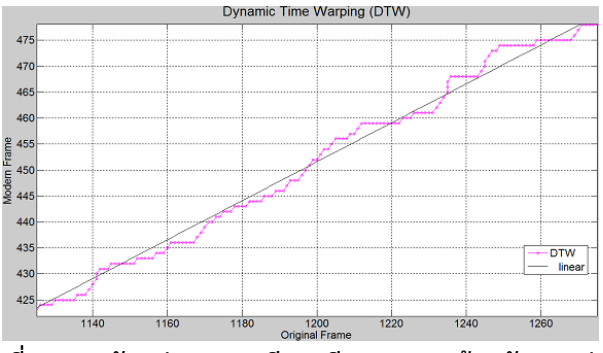
จากรูปที่ 9 เป็นตัวอย่างของการเปรียบเทียบคุณภาพของสัญญาณเสียงพูดหลังจากการบีบอัดด้วยของเวฟเล็ตทั้ง 3 ตระกูล โดยใช้เทคนิคสเปกโตรแกรมมาช่วยในการเปรียบเทียบค่าความถี่ของสัญญาณเสียงหลังจากการปรับปรุงสัญญาณเสียงพูดแล้ว โดยวิเคราะห์จากค่าความเข้มของความถี่ ถ้าเวฟเล็ตตระกูลใดมีลักษณะของกราฟความถี่ที่มีค่าผิดเพี้ยนน้อยที่สุดหรือมีกราฟความถี่ที่คล้ายกับต้นฉบับมากที่สุดจะถือว่าเวฟเล็ตตระกูลนั้นดีที่สุดสำหรับบทความนี้โดยเปรียบเทียบกับสัญญาณเสียงต้นฉบับ

ค. ใช้ Dynamic Time Warping (DTW) เป็นขั้นตอนวิธีสำหรับการเปรียบเทียบความคล้ายกันระหว่างข้อมูล 2 ชุด โดยผลลัพธ์ที่ได้จะให้ค่าระยะทางและวิถีการปรับแนว (Alignment) ที่ดีที่สุดระหว่างข้อมูลทั้งสอง ซึ่งสามารถยืดหรือหดให้รอรับความแปรผันในแกนเวลาได้ ซึ่ง DTW สามารถนำไปประยุกต์ได้กับวิดีโอ เสียง และภาพ รวมไปถึงการประมวลผลสัญญาณต่าง ๆ ที่สามารถแปลงให้อยู่ในรูปของข้อมูลเชิงเส้นได้ ดังนั้นงานวิจัยนี้จึงนำเทคนิคของ DTW มาประยุกต์ใช้ เพื่อจัดการกับคำพูดที่มีความเร็วและการสุ่มข้อมูลที่ไม่เท่ากันแม้จะสื่อความหมายเดียวกัน หรือเพื่อทำให้จังหวะทางเวลาของสัญญาณตัวอย่างมีค่าตรงกับสัญญาณต้นแบบ เพื่อลดข้อผิดพลาดที่จะเกิดขึ้นในการเปรียบเทียบสเปกตรัมของสัญญาณ โดยกำหนดให้ลำดับของ เวกเตอร์ A และเวกเตอร์ B ดังสมการที่ 3 และ การเปรียบเทียบระยะห่างระหว่างเวกเตอร์ 4 A และเวกเตอร์ B แสดงในสมการที่ เมื่อ 5 $d(i,j)$ เป็นระยะห่างระหว่างคู่จุดที่จุด i ของเฟรม A และ j ของเฟรม B โดยที่ a_{in} เป็นค่าลำดับที่ i ของเวกเตอร์ A และ b_{jn} เป็นค่าลำดับที่ j ของเวกเตอร์ B และ K เป็นจำนวนของลักษณะของเวกเตอร์ ซึ่งค่าที่ได้จะเป็นค่าระยะทางที่มีค่าน้อยที่สุด

$$A = a_1, a_2, \ldots, a_I \tag{3}$$

$$B = b_1, b_2, \ldots, b_J \tag{4}$$

$$d(i,j) = \sum_{n=1}^{K} \left(a_{in} - b_{jn}\right)^2 \tag{5}$$



รูปที่ 10 ตัวอย่างการเปรียบเทียบความคล้ายกันระหว่างสัญญาณเสียงต้นฉบับและเสียงใหม่

จากรูปที่ 10 จะเห็นได้ว่าเส้น DTW จะมีลักษณะใกล้เคียงกับเส้นตรงที่เป็นเส้นอ้างอิงมาก ซึ่งเทคนิคถ้าเส้น DTW ใกล้เคียงเส้นตรงมากเท่าไร แสดงให้เห็นว่าสัญญาณเสียงมีความใกล้เคียงกับต้นฉบับมากเท่านั้น ดังนั้นการเปรียบเทียบลักษณะเด่นของสัญญาณเสียงจึงจะใช้ตัวตรวจจับแบบสหสัมพันธ์ (Correlation Detector) โดยการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่างเส้น DTW และ เส้นอ้างอิง ตามสมการที่ 6

$$r = \frac{N\sum XY - \left((\sum X)(\sum Y)\right)}{\sqrt{\left(N\sum X^2 - (\sum X)^2\right)\left(N\sum Y^2 - (\sum Y)^2\right)}} \quad (6)$$

เมื่อ X คือ สัญญาณเสียงต้นฉบับ

Y คือ สัญญาณเสียงที่ถูกคืนกลับ

จากสมการที่ 6 เป็นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ เพื่อเปรียบเทียบประสิทธิภาพของเวฟเล็ตทั้ง 3 ตระกูล ซึ่งผลลัพธ์ของสัญญาณเสียงพูดแต่ละกลุ่มได้แสดงดังตารางที่ 2-5

ตารางที่ 2 การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 5 วินาที

อัตราการสุ่มตัวอย่างของสัญญาณเสียงพูด	ตระกูลเวฟเล็ต		
	Haar	Biorthogonal	Discrete Approximation of Meyer
3000	0.99931	0.99953	0.99902
4000	0.99957	0.99848	0.99876
5000	0.99970	0.99964	0.99965
6000	0.99974	0.99984	0.99988
8000	0.99938	0.99833	0.99787

ตารางที่ 3 การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 5 วินาที

อัตราการสุ่มตัวอย่างของสัญญาณเสียงพูด	ตระกูลเวฟเล็ต		
	Haar	Biorthogonal	Discrete Approximation of Meyer
3000	0.99273	0.97433	0.98839
4000	0.96996	0.96652	0.99299
5000	0.99468	0.99386	0.99208
6000	0.99213	0.97846	0.98747
8000	0.99389	0.99164	0.94263

ตารางที่ 4 การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 60 วินาที

อัตราการสุ่มตัวอย่างของสัญญาณเสียงพูด	ตระกูลเวฟเล็ต		
	Haar	Biorthogonal	Discrete Approximation of Meyer
3000	0.99986	0.99978	0.99986
4000	0.99977	0.99898	0.99941
5000	0.99993	0.99879	0.99870
6000	0.99983	0.99978	0.99980
8000	0.99979	0.99966	0.99986

ตารางที่ 5 การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 60 วินาที

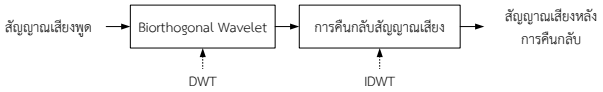
อัตราการสุ่มตัวอย่างของสัญญาณเสียงพูด	ตระกูลเวฟเล็ต		
	Haar	Biorthogonal	Discrete Approximation of Meyer
3000	0.99597	0.98840	0.98443
4000	0.98199	0.97472	0.95501
5000	0.99239	0.99095	0.99226
6000	0.99765	0.99473	0.99533
8000	0.99755	0.99592	0.99883

จากตารางที่ 2-5 เป็นการหาค่าสัมประสิทธิ์สหสัมพันธ์ (r) โดยผลการทดลองแสดงให้เห็นว่าเวฟเล็ตทั้ง 3 ตระกูล มีค่าสัมประสิทธิ์สหสัมพันธ์ที่ใกล้เคียงกันมาก โดยเฉพาะ Haar เวฟเล็ตให้ประสิทธิภาพการบีบอัดสัญญาณเสียงพูดและคืนกลับสัญญาณได้ดีกว่า Biorthogonal เวฟเล็ต และ Discrete Approximation of Meyer เวฟเล็ต เนื่องจากมีค่าสัมประสิทธิ์สหสัมพันธ์ ที่มีค่าเข้าใกล้ 1 มากที่สุด

จากการคัดเลือกตระกูลเวฟเล็ตทั้ง 3 ตระกูล เพื่อหาเวฟเล็ตที่เหมาะสมสำหรับบทความนี้ได้ข้อสรุปเป็นตระกูลเวฟเล็ต Biorthogonal เนื่องจากเวฟเล็ตตระกูลนี้มีฟังก์ชันพื้นฐานที่เหมาะสมกับการวิเคราะห์สัญญาณที่มีลักษณะไม่คงที่ ในรูปที่ 5-8 จะเห็นได้ว่า เวฟเล็ต Biorthogonal ให้ผลที่ดีกว่าเวฟเล็ตตระกูลอื่น ในขั้นตอนต่อไปนี้จะกล่าวถึงการบีบอัดสัญญาณเสียงด้วยเวฟเล็ตเปรียบเทียบกับการบีบอัดสัญญาณเสียงด้วย CELP โดยการบีบอัดนั้นจะใช้ Order สำหรับการบีบอัดที่เท่ากัน คือ Order 10, 20 และ 30

การบีบอัดสัญญาณเสียงพูดโดยใช้เวฟเล็ตตระกูล Biorthogonal

ในขั้นตอนนี้จะนำสัญญาณเสียงที่ได้เตรียมไว้เข้าสู่กระบวนการบีบอัดด้วยเวฟเล็ตดังรูปที่ 11

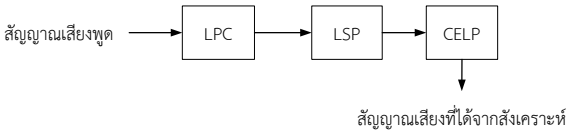


รูปที่ 11 กระบวนการบีบอัดสัญญาณเสียงพูดของเวฟเล็ต

จากรูปที่ 11 สัญญาณเสียงทั้งหมดจะถูกบีบอัดสัญญาณเสียงพูดด้วย Biorthogonal เวฟเล็ตโดยใช้เทคนิคของการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (DWT) และทำการคืนกลับสัญญาณเสียงโดยใช้เทคนิคการแปลงกลับเวฟเล็ตแบบไม่ต่อเนื่อง (Inverse Discrete Wavelet Transform; IDWT)

การสกัดลักษณะของสัญญาณเสียงพูดโดยใช้การเข้ารหัสแบบเชิงเส้น (LPC)

ในขั้นตอนนี้จะนำสัญญาณเสียงชุดเดียวกันกับการบีบอัดสัญญาณเสียงด้วยเวฟเล็ต เพื่อเข้าสู่กระบวนการสกัดลักษณะเด่นของสัญญาณสียงด้วยเทคนิค LPC และใช้หลักการของ CELP เพื่อบีบอัดสัญญาณและสังเคราะห์สัญญาณเสียงขึ้นมาใหม่ดังรูปที่ 12



รูปที่ 12 กระบวนการบีบอัดสัญญาณเสียงพูดของ CELP

จากรูปที่ 12 เสียงทั้งหมดจะถูกนำไปสกัดลักษณะเด่นของสัญญาณเสียงด้วย LPC ที่ Oder 10, 20 และ 40 แต่เนื่องจากเทคนิคของ LPC จะมีประสิทธิภาพไม่ค่อยดี จึงส่งค่าพารามิเตอร์การควอนไตซ์เซชันของ LPC ไปแปลงเป็นค่าพารามิเตอร์ของ LSP เพื่อปรับปรุงการเข้ารหัสสัญญาณเสียงพูดให้มีประสิทธิภาพดีขึ้น หลังจากนั้นนำไปบีบอัดสัญญาณเสียงด้วย CELP และสังเคราะห์สัญญาณเสียงขึ้นมาใหม่โดยใช้ค่าพารามิเตอร์ของ LSP ในการสังเคราะห์ ซึ่ง CELP เป็นเทคโนโลยีในการบีบอัดสัญญาณเสียงที่ได้รับมาตรฐานและเป็นที่รู้จักและน่าจับตามองในปัจจุบัน

เทคนิคการเปรียบเทียบคุณภาพของสัญญาณเสียง

หลังจากการบีบอัดสัญญาณเสียงพูดด้วย DWT และ CELP นั้น ได้ทำการคืนกลับสัญญาณ เสียงด้วยเทคนิคที่แตกต่างกันคือ DWT จะคืนกลับเสียงด้วย IDWT และ CELP สังเคราะห์สัญญาณเสียงขึ้นมาใหม่โดยใช้พารามิเตอร์ของ LSP ซึ่งเสียงที่ได้จะมีความแตกต่างกัน ดังนั้นในบทความนี้จึงใช้เทคนิคค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Mean Square Error; MSE) และอัตราส่วนของสัญญาณต่อสัญญาณรบกวนสูงสุด (Peak Signal to Noise Ratio; PSNR) ในการเปรียบเทียบคุณภาพของสัญญาณเสียง เพื่อหาเทคนิคการบีบอัดสัญญาณเสียงที่เหมาะสมที่สุด

ก. การหาค่าเฉลี่ยความคาดเคลื่อนกำลังสอง (Mean Square Error)

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}(s_i - p_i)^2 \quad (7)$$

โดยที่ s_i คือ สัญญาณเสียงต้นฉบับ

p_i คือ สัญญาณเสียงที่ถูกคืนกลับ

ข. อัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงสุด (Peak Signal to Noise Ratio)

$$PSNR = 10\log_{10}\frac{NX^2}{\|x-r\|^2} \quad (8)$$

โดยที่ N คือ ความยาวของสัญญาณเสียงที่ถูกคืนกลับ

X คือ ค่าสูงสุดของค่าเฉลี่ยกำลังสองของสัญญาณเสียงต้นฉบับ

$\|x-r\|^2$ คือ ความแตกต่างของระดับพลังงานระหว่างสัญญาณ เสียงต้นฉบับกับสัญญาณเสียงคืนกลับ

ผลการทดลอง

จากการบีบอัดสัญญาณเสียงทั้งหมด 80 เสียง โดยแบ่งเป็นเสียงของผู้หญิงและผู้ชายที่มีความยาว 5 วินาที และ 60 วินาที อย่างละ 20 เสียง เพื่อบีบอัดสัญญาณเสียงด้วย DWT ระดับที่ 1- 3 และ CELP Level ที่ 10, 20, 30 โดยเปรียบเทียบผลการทดลองจากค่าความผิดเพี้ยนของ MSE และการหาประสิทธิภาพของ PSNR แสดงให้เห็นว่าคุณภาพของสัญญาณเสียงที่ผ่านการบีบอัดด้วย DWT จะเห็นว่าคุณภาพของสัญญาณเสียงจะดีขึ้นเป็นลำดับโดยเฉพาะอย่างยิ่งการบีบอัดด้วยระดับ 1 เมื่อเปรียบเทียบกับการบีบอัดด้วย CELP ซึ่งสามารถสรุปเป็นค่าเฉลี่ยของเสียงแต่ละประเภทดังตารางที่ 6 และ 7

ตารางที่ 6 ค่าเฉลี่ยของ MSE

ระดับการบีบอัด	เสียงพูด	ค่าเฉลี่ย MSE	
		DWT	CELP
ระดับที่ 3 หรือ Order 10	ผู้หญิง 5 วินาที	0.0039684	0.0394481
	ผู้ชาย 5 วินาที	0.002262	0.0109041
	ผู้หญิง 60 วินาที	0.0081388	0.015538
	ผู้ชาย 60 วินาที	0.0040201	0.0037642
ระดับที่ 2 หรือ Order 20	ผู้หญิง 5 วินาที	0.000699	0.0346344
	ผู้ชาย 5 วินาที	0.001325	0.0098298
	ผู้หญิง 60 วินาที	0.0024015	0.0161028
	ผู้ชาย 60 วินาที	0.0025496	0.0045219
ระดับที่ 1 หรือ Order 30	ผู้หญิง 5 วินาที	0.0002561	0.041775
	ผู้ชาย 5 วินาที	0.0003154	0.0105815
	ผู้หญิง 60 วินาที	0.0005144	0.0168456
	ผู้ชาย 60 วินาที	0.0007185	0.00439

ตารางที่ 7 ค่าเฉลี่ยของ PSNR

ระดับการบีบอัด	เสียง	ค่าเฉลี่ย PSNR	
		DWT	CELP
ระดับที่ 3 หรือ Order 10	ผู้หญิง 5 วินาที	22.964986	13.781483
	ผู้ชาย 5 วินาที	19.55126	19.381605
	ผู้หญิง 60 วินาที	21.404765	17.657204
	ผู้ชาย 60 วินาที	22.588559	23.782297
ระดับที่ 2 หรือ Order 20	ผู้หญิง 5 วินาที	30.536701	14.82256
	ผู้ชาย 5 วินาที	24.893672	20.100045
	ผู้หญิง 60 วินาที	27.309833	17.807994
	ผู้ชาย 60 วินาที	26.265939	23.250548
ระดับที่ 1 หรือ Order 30	ผู้หญิง 5 วินาที	34.548318	13.704442
	ผู้ชาย 5 วินาที	32.612891	19.78766
	ผู้หญิง 60 วินาที	33.886936	17.548176
	ผู้ชาย 60 วินาที	31.965845	23.639487

สรุปผล

บทความนี้เป็นการศึกษาและการใช้ประโยชน์การแปลงเวฟเล็ตสำหรับการบีบอัดสัญญาณเสียงพูด โดยใช้เทคนิคการบีบอัดสัญญาณเสียงด้วย DWT และ CELP เพื่อเปรียบเทียบคุณภาพหลังการคืนกลับสัญญาณ ซึ่งปกติแล้วในกระบวนการส่งสัญญาณเสียงที่มีการบีบอัดสัญญาณเสียงนั้น เพื่อต้องการลดขนาดของสัญญาณเสียงพูดจากสัญญาณเสียงต้นฉบับ และต้องการให้สัญญาณเสียงพูดมีคุณสมบัติใกล้เคียงกับสัญญาณเสียงพูดต้นฉบับมากที่สุดหรือขึ้นอยู่กับการนำเสียงนั้นๆ ไปประยุกต์ใช้งาน

ผลการทดลองพบว่าเวฟเล็ตที่เหมาะสมและผ่านการคัดเลือกคือ Biorthogonal เวฟเล็ต เนื่องจากให้ประสิทธิภาพในการบีบอัดสัญญาณที่สูงกว่าเวฟเล็ตทั้ง 2 ตระกูล ดังนั้นจึงนำ Biorthogonal เวฟเล็ต เข้าสู่กระบวนการบีบอัดสัญญาณเสียงพูดโดยใช้เทคนิคของ DWT ระดับที่ 1 – 3 และคืนกลับสัญญาณเสียงด้วย IDWT จากผลการทดลองสามารถวิเคราะห์ได้ว่าเสียงพูดที่มีความยาว 5 วินาที จะมีประสิทธิภาพในการคืนกลับสัญญาณมากกว่าเสียงพูดที่มีความยาว 60 วินาที โดยเฉพาะสัญญาณเสียงพูดของผู้หญิง

เมื่อนำสัญญาณเสียงพูดหลังจากการคืนกลับสัญญาณเสียงไปเปรียบเทียบคุณภาพของสัญญาณเสียงกับเทคนิคการบีบอัดสัญญาณของ CELP ที่ Order 10, 20 และ 30 เนื่องจากเทคนิคการบีบอัดสัญญาณเสียงของ CELP นั้นเป็นมาตรฐานการบีบอัดที่นิยมใช้กันมากในปัจจุบัน ซึ่งหลักการที่ใช้สำหรับเปรียบเทียบประสิทธิภาพของสัญญาณระหว่างการบีบอัดสัญญาณเสียงพูดด้วย DWT กับ CELP นั้น จะใช้ค่า MSE และ PSNR มาวิเคราะห์ผล ซึ่งปรากฏว่าการบีบอัดสัญญาณเสียงด้วย DWT ให้ประสิทธิภาพในการบีบอัดสัญญาณเสียงที่ดีกว่า CELP และทั้งนี้ยังมีค่าผิดพลาดที่น้อยกว่าเมื่อนำสัญญาณไปเปรียบเทียบกับสัญญาณต้นฉบับ และยังสามารถนำไปประยุกต์ใช้บีบอัดข้อมูลชนิดอื่นได้

เอกสารอ้างอิง

[1] วีระยุทธ คุณรัตนสิริ และจักรี ศรีนนท์ฉัตร .“การเปรียบเทียบประสิทธิภาพการบีบอัดเสียงพูดโดยใช้เทคนิคเวฟเล็ต”. การประชุมเครือข่ายวิชาการด้านวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 3(EENET), 2011, หน้า 339-342

[2] B.Dennis. “Details to assist in implementation of Federal Standard 1016 CELP," National Communication Bulletin ,1-92The Manager National Communications system,1992 .

[3] A.N.Suen, J.F. Wang and T. C. Yao. “Dynamic partial search scheme for stochastic codebook of FS1016 CELP coder,” IEEE Proc. – Image Signal Process, vol. 142, China, February 1, 1995, pp. 52-58.

[4] M.A. Osman and Hussein, N. “Speech compression using LPC and wavelet,” 2nd International Conference on Computer Engineering and Technology (ICCET), April Chengdu China, 2010, pp. 92-97.

[5] S. M. Joseph and P. Babu Anto, “Speech Compression Using Wavelet Transform,” IEEE-International Conference on Recent Trends in Information Technology (ICRTIT), Anna University, Chennai. June 5-3, 2011, pp. 754-758.

[6] Z. Dan and M.A. Sheng-qian, “Speech Compression With Best Wavelet Packet Transform And SPIHT

Algorithm," Second International Conference on Computer Modeling and Simulation, 2010, pp. 360-363.

[7] S. Ranjan, "A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition," International Conference on Signal Acquisition and Processing (ICSAP) ,Bangalore India, 2010 , pp. 345-348.

[8] S. Kornsing and J. Srinonchat. "Exploring Dynamic Time Warping Techniques to Wavelet Speech Compressing," 9th International Conference on Electrical Engineering/ Electronics Computer Telecommunications and Information Technology , Hua Hin, Thailand, 2012, pp. 1-4.

[9] S. Kornsing and J. Srinonchat, "Enhancement Speech Compression Technique Using Modern Wavelet Transforms," International Symposium on Computer Consumer and Control, Hua Hin, Thailand, 2012, pp. 393-396.