

การศึกษาและการเปรียบเทียบเวฟเล็ตฟิลเตอร์สำหรับการบีบอัดสัญญาณเสียง

**STUDYING AND COMPARISON OF WAVELET FILTER  
FOR SPEECH COMPRESSION**

สุภาวณิ กรสิงห์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2555

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

# การศึกษาและการเปรียบเทียบเวฟเล็ตฟิลเตอร์สำหรับการบีบอัดสัญญาณเสียง

สุภาวณิ กรสิงห์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2555

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

หัวข้อวิทยานิพนธ์	การศึกษาและการเปรียบเทียบเวฟเล็ตฟิลเตอร์สำหรับการบีบอัดสัญญาณเสียง
ชื่อ - นามสกุล	นางสาวสุภาวชิณี กรสิงห์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	อาจารย์จักรี ศรีนนท์ฉัตร, Ph.D.
ปีการศึกษา	2555

---

คณะกรรมการสอบวิทยานิพนธ์

.....	ประธานกรรมการ
(อาจารย์อำนวยการ เรืองวารี, Dr.-Ing)	
.....	กรรมการ
(รองศาสตราจารย์ ดร.สมเกียรติ อุดมหารธาดา)	
.....	กรรมการ
(อาจารย์สุรินทร์ แห่งงาม, Ph.D.)	
.....	กรรมการ
(อาจารย์จักรี ศรีนนท์ฉัตร, Ph.D.)	

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี อนุมัติวิทยานิพนธ์ฉบับนี้  
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโท

.....คณบดีคณะวิศวกรรมศาสตร์  
(ผู้ช่วยศาสตราจารย์สมหมาย ผิวสอาด, Ph.D.)  
วันที่ 7 เดือน ตุลาคม พ.ศ. 2555

หัวข้อวิทยานิพนธ์	การศึกษาและการเปรียบเทียบเวฟเลตฟิลเตอร์สำหรับการบีบอัดสัญญาณเสียง
ชื่อ - นามสกุล	นางสาวสุภาวชิณี กรสิงห์
สาขาวิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	อาจารย์จักรี ศรีนนท์ฉัตร, Ph.D.
ปีการศึกษา	2555

## บทคัดย่อ

ปัจจุบันงานวิจัยทางการบีบอัดสัญญาณเสียงมีวัตถุประสงค์คือการทำให้อัตราสัญญาณเสียงมีขนาดเล็กลง แต่ยังคงรักษาคุณภาพของสัญญาณเสียงให้มีคุณภาพใกล้เคียงกับต้นฉบับให้มากที่สุด วิทยานิพนธ์นี้นำเสนอการศึกษาและการเปรียบเทียบเวฟเลตฟิลเตอร์สำหรับการบีบอัดสัญญาณเสียง

การทดลองนี้ได้ใช้สัญญาณเสียงทั้งหมด 80 เสียง โดยแบ่งเป็น 4 กลุ่ม ดังนี้ เสียงพูดของผู้หญิงและผู้ชายที่มีความยาว 5 วินาที และ 60 วินาที อย่างละ 20 เสียง จากนั้นนำสัญญาณเสียงไปเข้ากระบวนการคัดเลือกเวฟเลตจากเวฟเลต 3 ตระกูล คือ ฮาร์เวฟเลต ไบออร์ทอลโกนอลเวฟเลตและการประมาณค่าไม่ต่อเนื่องของเมเยอร์เวฟเลต เพื่อหาเวฟเลตที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ ทั้งนี้วิธีการคัดเลือกใช้หลักการหาค่าพลังงานเฉลี่ย การเปรียบเทียบความถี่สเปกโตรแกรม และหลักการของ Dynamic Time Warping (DTW) เป็นตัววัดผล หลังจากนั้นนำเวฟเลตที่ผ่านการคัดเลือกไปทำการบีบอัดสัญญาณเสียงใน ระดับที่ 1-3 ซึ่งผลลัพธ์ที่ได้จะถูกนำไปเปรียบเทียบกับเทคนิคการบีบอัดสัญญาณด้วย Federal Standard 1016 Code Excite Linear Prediction (FS 1016 CELP) โดยใช้หลักการของค่าเฉลี่ยผิดพลาดกำลังสองและอัตราส่วนของสัญญาณสูงสุดเป็นตัววัดคุณภาพของการบีบอัดสัญญาณ

ผลการทดสอบในวิทยานิพนธ์นี้พบว่า เวฟเลตตระกูลไบออร์ทอลโกนอลให้ประสิทธิภาพในการบีบอัดสูงที่สุด และในการสังเคราะห์เสียงด้วยการคืนกลับเวฟเลตแบบไม่ต่อเนื่อง นั้นสัญญาณเสียงของผู้หญิงที่มีความยาว 5 วินาทีให้ประสิทธิภาพเฉลี่ยดีที่สุด ในการเปรียบเทียบค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงพูดทั้งหมดด้วยการแปลงเวฟเลตแบบไม่ต่อเนื่อง และ CELP ผลปรากฏว่าการแปลงเวฟเลตแบบไม่ต่อเนื่องให้ประสิทธิภาพในการบีบอัดสัญญาณเสียงดีกว่า CELP และยังมีค่าผิดพลาดน้อยกว่าเมื่อนำสัญญาณไปเปรียบเทียบกับสัญญาณต้นฉบับ

คำสำคัญ: ฮาร์เวฟเลต ไบออร์ทอลโกนอลเวฟเลต การประมาณค่าไม่ต่อเนื่องของเมเยอร์เวฟเลต

<b>Thesis Title</b>	Studying and Comparison of Wavelet Filter for Speech Compression
<b>Name - Surname</b>	Miss Supatinee Kornsing
<b>Program</b>	Electrical Engineering
<b>Thesis Advisor</b>	Mr. Jakkree Srinonchat, Ph.D.
<b>Academic Year</b>	2012

## 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

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ ดร. จักรี ศรีนนท์ฉัตร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งได้ให้คำแนะนำข้อคิดเห็นและสนับสนุนในการทำวิจัย มาด้วยดีตลอด ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

กราบขอบพระคุณ คณะกรรมการสอบวิทยานิพนธ์ ดร. อำนวย เรืองวารี ดร. สุรินทร์ แหงมงาม ดร. อภिरดา นามแสง และผู้ทรงคุณวุฒิ รองศาสตราจารย์ ดร. สมเกียรติ อุดมหารธากุล ที่ให้คำแนะนำ ความรู้และประสบการณ์ที่เป็นแนวทางในการทำวิทยานิพนธ์นี้ให้สำเร็จลุล่วงด้วยดี และเป็นประโยชน์ในการทำวิจัยครั้งต่อไป

ขอขอบคุณพี่ๆ เพื่อนๆ ที่ห้องปฏิบัติการและวิจัยทางด้านการประมวลผลสัญญาณ ที่ได้ช่วยเหลือเกี่ยวกับข้อมูล รวมถึงคำแนะนำต่างๆ ตลอดเวลาที่ทำวิจัยอย่างดียิ่ง

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดามารดาที่ให้การสนับสนุนแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

สุภาวณิ กรสิงห์

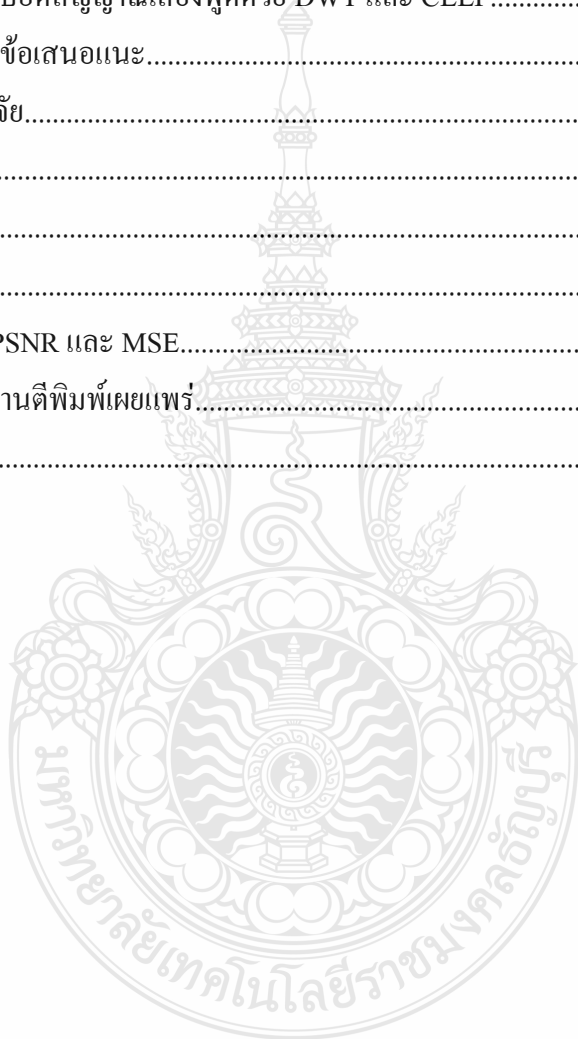


## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ฌ
คำอธิบายสัญลักษณ์และคำย่อ.....	ฎ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ขอบเขตของการศึกษา.....	2
1.4 ข้อยกเว้นของการศึกษา.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	3
2.1 การบีบอัดสัญญาณเสียงพูด.....	3
2.2 การเข้ารหัสสัญญาณเสียงพูดด้วย CELP.....	5
2.3 การเข้ารหัสการทำนายพัลส์เชิงเส้น.....	6
2.4 คู่เส้นสเปกตรัม.....	12
2.5 การแปลงเวฟเล็ต.....	14
2.6 งานวิจัยที่เกี่ยวข้อง.....	25
3 วิธีการดำเนินงานวิจัย.....	28
3.1 การเลือกตระกูลเวฟเล็ต.....	29
3.2 ขั้นตอนการบีบอัดสัญญาณเสียงพูด.....	39
3.3 เทคนิคการเปรียบเทียบคุณภาพของสัญญาณเสียง.....	40
4 ผลของงานวิจัย.....	43
4.1 ผลการบีบอัดสัญญาณเสียงด้วย DWT.....	43

## สารบัญ (ต่อ)

บทที่	หน้า
4.2 ผลการคืนกลับสัญญาณเสียงด้วย IDWT.....	46
4.3 ผลการบีบอัดสัญญาณเสียงด้วยเทคนิค CELP.....	63
4.4 การเปรียบเทียบคุณภาพของการบีบอัดระหว่าง DWT กับ CELP.....	68
4.5 สรุปผลการบีบอัดสัญญาณเสียงพูดด้วย DWT และ CELP.....	81
5 สรุปผลการวิจัยและข้อเสนอแนะ.....	82
5.1 สรุปผลการวิจัย.....	82
5.2 ข้อเสนอแนะ.....	83
รายการอ้างอิง.....	84
ภาคผนวก.....	87
ภาคผนวก ก ค่า PSNR และ MSE.....	88
ภาคผนวก ข ผลงานตีพิมพ์เผยแพร่.....	101
ประวัติผู้เขียน.....	133





## สารบัญตาราง

ตารางที่	หน้า
2.1 พารามิเตอร์และอัตราบิตที่ใช้ส่งข้อมูลของ FS 1016 CELP.....	5
3.1 Sample Rate และอัตราการบีบอัดสัญญาณเสียงของอัตราส่วน $\frac{p}{q}$ แต่ละค่า.....	32
3.2 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 5 วินาที.....	39
3.3 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 5 วินาที.....	39
3.4 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 60 วินาที...	39
3.5 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 60 วินาที.....	40
4.1 สูตรค่าเฉลี่ยของ MSE และ PSNR.....	81
5.1 ประสิทธิภาพของสัญญาณเสียง PSNR และค่าผิดพลาด MSE.....	83



## สารบัญญภาพ

ภาพที่	หน้า
2.1 การส่งสัญญาณเสียงพูดในระบบสื่อสาร.....	3
2.2 บล็อกไดอะแกรมการเข้ารหัส FS 1016 CELP.....	6
2.3 การวางเรียงสลับของรากของพหุนามคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ .....	13
2.4 ความสัมพันธ์ระหว่างรากของ $A(z)$ กับรากของคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ .....	14
2.5 ลักษณะของระนาบเวลาและความถี่สำหรับการวิเคราะห์สัญญาณ.....	15
2.6 ลักษณะสัญญาณชาวน์ที่มีตัวประกอบแตกต่างกัน.....	16
2.7 ลักษณะเวฟเล็ตที่มีตัวประกอบขนาดแตกต่างกัน.....	17
2.8 ลักษณะของเวฟเล็ตที่มีการเลื่อนตำแหน่ง.....	17
2.9 ลักษณะของเวฟเล็ตแม่ที่ถูกสเกลและเลื่อนตำแหน่งไปที่ค่า $a, b$ ต่างๆ กัน.....	18
2.10 ลักษณะของเวฟเล็ต Haar.....	18
2.11 ลักษณะของ เวฟเล็ต Daubechies.....	19
2.12 ลักษณะของเวฟเล็ต Symlet.....	19
2.13 ลักษณะของเวฟเล็ต Coiflet.....	20
2.14 ลักษณะของเวฟเล็ต Morlet.....	20
2.15 ลักษณะของเวฟเล็ต Maxican Hat.....	20
2.16 ลักษณะของเวฟเล็ต Meyer.....	21
2.17 ลักษณะของเวฟเล็ต Biorthogonal.....	21
2.18 การแปลงฟูเรียร์ที่ประกอบด้วยสัญญาณชาวน์ที่มีความถี่แตกต่างกัน.....	22
2.19 ลักษณะของผลการ Correlation ระหว่าง $f(t), \psi\left(\frac{t-b}{a}\right)$ .....	23
2.20 การสร้างสัญญาณการแปลงสัญญาณเวฟเล็ตแบบแบ่งช่วงในขั้นตอนเดียว.....	24
3.1 ขั้นตอนการทำงานของงานวิจัย.....	28
3.2 ขั้นตอนการคัดเลือกตระกูลเวฟเล็ต.....	29
3.3 ตัวอย่างสัญญาณเสียงพูด ที่เวลา 5 วินาที.....	30
3.4 ตัวอย่างสัญญาณเสียงพูด ที่เวลา 60 วินาที.....	30
3.5 การแยกองค์ประกอบของความถี่ต่ำและความถี่สูงของการแปลงเวฟเล็ต.....	31
3.6 ขั้นตอนและองค์ประกอบของการแปลงวิธีเวฟเล็ต.....	31

## สารบัญภาพ (ต่อ)

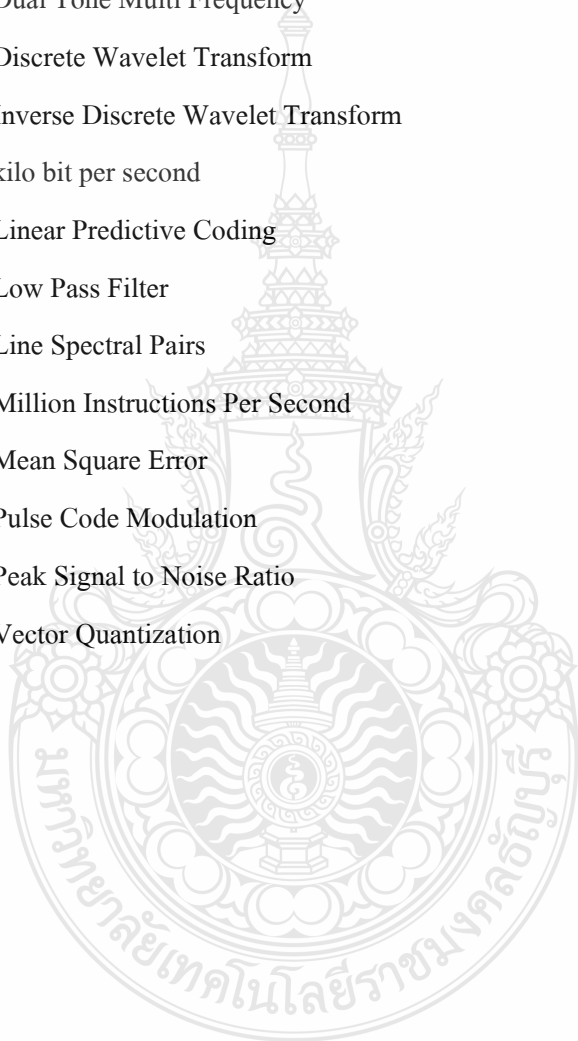
ภาพที่	หน้า
3.7 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 5 วินาที.....	33
3.8 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 60 วินาที.....	33
3.9 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 5 วินาที.....	34
3.10 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 60 วินาที.....	34
3.11 ตัวอย่างการเปรียบเทียบความถี่สเปกตรัมของสัญญาณเสียง.....	36
3.12 ตัวอย่างการเปรียบเทียบความคล้ายกันระหว่างสัญญาณเสียงต้นฉบับและเสียงใหม่.....	38
3.13 กระบวนการบีบอัดสัญญาณเสียงพูดของเวฟเล็ต.....	41
3.14 กระบวนการบีบอัดสัญญาณเสียงพูดของ CELP.....	41
4.1 การบีบอัดสัญญาณเสียงพูดด้วย DWT.....	43
4.2 การแบ่งช่วงของสัญญาณเสียงพูดผู้หญิงที่เวลา 5 วินาที.....	44
4.3 การแบ่งช่วงของสัญญาณเสียงพูดผู้ชายที่เวลา 5 วินาที.....	45
4.4 การแบ่งช่วงของสัญญาณเสียงพูดผู้หญิงที่เวลา 60 วินาที.....	45
4.5 การแบ่งช่วงของสัญญาณเสียงพูดผู้ชายที่เวลา 60 วินาที.....	46
4.6 การแปลงกลับสัญญาณด้วยตัวกรองแบบ 2 ช่องสัญญาณ.....	47
4.7 การคืนกลับสัญญาณเสียงพูดด้วย IDWT.....	47
4.8 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 5 ใน 1 เฟรม.....	48
4.9 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 5 ใน 1 เฟรม.....	50
4.10 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 4.....	51
4.11 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 4 ใน 1 เฟรม.....	53
4.12 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 3.....	54
4.13 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 3 ใน 1 เฟรม.....	56
4.14 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 2.....	57
4.15 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 2 ใน 1 เฟรม.....	59

## สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.16 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 1.....	60
4.17 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 1 ใน 1 เฟรม.....	62
4.18 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้ชายที่ 5 วินาที ด้วย CELP ใน 1 เฟรม.....	64
4.19 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้ชายที่ 5 วินาที ด้วย CELP ใน 1 เฟรม.....	65
4.20 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้หญิงที่ 60 วินาที ด้วย CELP ใน 1 เฟรม.....	66
4.21 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้หญิงที่ 60 วินาที ด้วย CELP ใน 1 เฟรม.....	67
4.22 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้หญิงที่เวลา 5 วินาที.....	69
4.23 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้ชายที่เวลา 5 วินาที.....	70
4.24 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้หญิงที่เวลา 60 วินาที.....	71
4.25 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้ชายที่เวลา 60 วินาที.....	72
4.26 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้หญิงที่เวลา 5 วินาที.....	73
4.27 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้ชายที่เวลา 5 วินาที.....	74
4.28 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้หญิงที่เวลา 60 วินาที.....	75
4.29 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้ชายที่เวลา 60 วินาที.....	76
4.30 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้หญิงที่เวลา 5 วินาที.....	77
4.31 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้ชายที่เวลา 5 วินาที.....	78
4.32 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้หญิงที่เวลา 60 วินาที.....	79
4.33 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้ชายที่เวลา 60 วินาที.....	80

## คำอธิบายสัญลักษณ์และคำย่อ

cA	Coefficient Approximation
cD	Coefficient Detail
CELP	Code Excite Linear Prediction
DTMF	Dual Tone Multi Frequency
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
kbps	kilo bit per second
LPC	Linear Predictive Coding
LPF	Low Pass Filter
LSP	Line Spectral Pairs
MIPS	Million Instructions Per Second
MSE	Mean Square Error
PCM	Pulse Code Modulation
PSNR	Peak Signal to Noise Ratio
VQ	Vector Quantization



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

สัญญาณเสียงพูด (Speech Signal) [1] นั้น เป็นสัญญาณที่มีแถบความถี่ต่ำที่สุด จึงเป็นที่สนใจในการนำมาประมวลผลสัญญาณแบบเวลาจริง (Real Time) เพื่อนำไปพัฒนาต่อเทคโนโลยีต่างๆ และในงานทางการสื่อสารอีกเป็นจำนวนมาก แต่ปัญหาที่สำคัญของสื่อสารด้วยสัญญาณเสียงพูดคือ การที่ระบบสื่อสารนั้นมีแบนด์วิดท์ที่มีจำกัด (Limited Bandwidth) ดังนั้นการพัฒนากิจการประมวลผลสัญญาณเสียงพูดมีความเร็วแบบเวลาจริงนั้นจึงจำเป็นต้องลดขนาดของสัญญาณลงหรือเรียกว่าการบีบอัดสัญญาณเสียงพูด (Speech Compression) เพื่อให้เหมาะสมกับแบนด์วิดท์ที่มีอย่างจำกัด และเมื่อนำเสียงพูดที่ถูกบีบอัดไปแล้วนั้น มาทำการคืนกลับสัญญาณเสียงหรือการสังเคราะห์สัญญาณเสียงขึ้นมาใหม่ อาจทำให้คุณภาพของสัญญาณเสียงพูดที่ได้มีคุณภาพที่ต่ำลงมาก เนื่องจากการสูญเสียข้อมูลจำนวนมากขณะบีบอัดสัญญาณ ซึ่งในปัจจุบันนี้เทคนิคที่ใช้สำหรับการลดขนาดหรือการบีบอัดสัญญาณเสียงพูดมีหลากหลายเทคนิค ซึ่งแต่ละเทคนิคนั้นจะมีข้อดีและข้อเสียที่แตกต่างกัน และจากการศึกษาพบว่า การบีบอัดสัญญาณด้วยการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform; DWT) เป็นเทคนิคที่ได้รับความนิยมมากในปัจจุบัน

เทคนิค DWT คือกระบวนการบีบอัดสัญญาณหรือการวิเคราะห์สัญญาณ โดยใช้ตัวกรองความถี่ ซึ่งโดยปกติสัญญาณเสียงมนุษย์จะมีทั้งความถี่ต่ำและความถี่สูง ดังนั้นเมื่อนำสัญญาณเสียงเข้าสู่กระบวนการบีบอัดด้วยเวฟเล็ตสัญญาณจะถูกแบ่งออกเป็น 2 ส่วน คือ การประมาณค่า (Approximation) หรือ ส่วนของความถี่ต่ำ และรายละเอียด (Detail) หรือ ส่วนของความถี่สูง แต่สำหรับการกรองความถี่ของสัญญาณเสียงพูดนั้น ส่วนที่สำคัญที่สุดก็คือ ส่วนที่เป็นความถี่ต่ำ เนื่องจากเสียงมนุษย์นั้นเมื่อนำส่วนความถี่สูงออกเสียงจะเปลี่ยนแปลงไป แต่ยังคงรู้ว่าผู้พูดต้องการสื่อถึงอะไร แต่ถ้าหากนำส่วนความถี่ต่ำออกไปนั้น จะกลายเป็นเสียงพิมพ์พาดำนั้น

งานวิจัยนี้ได้เลือกเทคนิคการบีบอัดสัญญาณเสียงด้วย DWT เพื่อนำไปประยุกต์ใช้กับเสียงทั้งหมด 80 เสียง และได้นำเทคนิค Code Excite Linear Prediction (CELP) มาเปรียบเทียบกับเพื่อหาคุณภาพของการบีบอัดสัญญาณเสียงพูดหลังจากการคืนกลับสัญญาณหรือการสังเคราะห์สัญญาณนั้นๆ 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) แต่สำหรับงานวิจัยนี้ได้เลือกใช้การบีบอัดสัญญาณเสียงด้วย CELP ในมาตรฐานของ FS1016 CELP เพื่อนำไปเทียบคุณภาพกับเทคนิค DWT

## 1.2 วัตถุประสงค์ของการศึกษา

1.2.1 ศึกษาและเปรียบเทียบคุณสมบัติของเวฟเล็ตสำหรับการบีบอัดสัญญาณเสียง

1.2.2 เปรียบเทียบคุณภาพของสัญญาณเสียงพูดหลังจากผ่านกระบวนการบีบอัดสัญญาณเสียงโดยใช้เทคนิค DWT และ CELP

1.2.3 วิเคราะห์คุณภาพของสัญญาณเสียงพูดหลังการบีบอัดสัญญาณเสียงพูด

## 1.3 ขอบเขตของการศึกษา

1.3.1 คัดเลือกชนิดของเวฟเล็ตที่เหมาะสมสำหรับการบีบอัดสัญญาณเสียงพูด

1.3.2 ออกแบบเทคนิคการบีบอัดสัญญาณเสียงพูดให้สามารถลดขนาดของสัญญาณเสียงได้ไม่น้อยกว่า 15 เปอร์เซ็นต์ เมื่อเทียบกับสัญญาณเสียงต้นฉบับ

1.3.3 วิเคราะห์คุณภาพของสัญญาณเสียงที่สังเคราะห์ได้ด้วยวิธีการ MSE และ PSNR

## 1.4 ข้อยกเว้นของการศึกษา

1.4.1 เสียงที่นำวิจัยต้องเป็นสัญญาณเสียงพูดโดยแบ่งเป็น 4 กลุ่ม คือ เสียงผู้ชาย 5 วินาที 20 เสียง เสียงผู้หญิง 5 วินาที 20 เสียงผู้ชาย 60 วินาที 20 เสียง และ เสียงผู้หญิง 60 วินาที 20 เสียง

1.4.2 ประมวลผลโดยใช้โปรแกรม MATLAB

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 พัฒนางานวิจัยที่เกี่ยวข้องกับการบีบอัดสัญญาณเสียงพูด โดยใช้เทคนิค DWT และ CELP

1.5.2 เป็นแนวทางในการพัฒนาการบีบอัดสัญญาณเสียง เพื่อลดขนาดของข้อมูลและสามารถนำไปประยุกต์ใช้ในระบบการสื่อสารในอนาคต

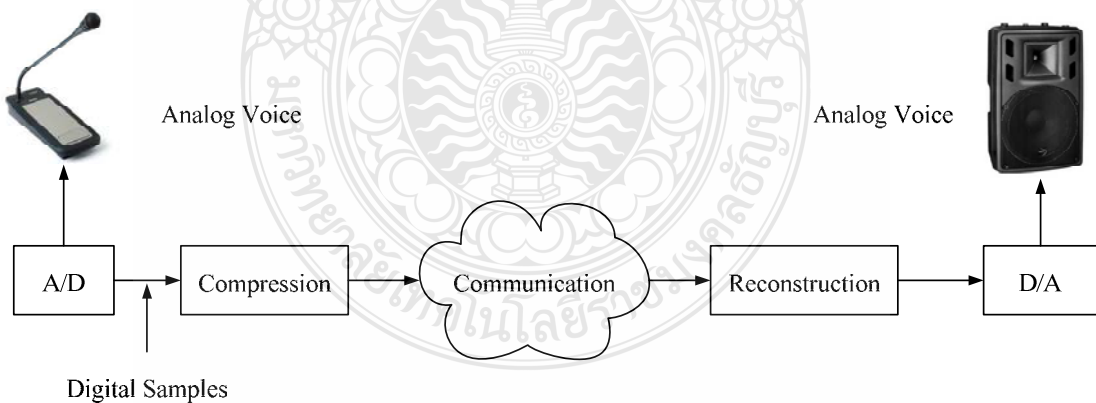
## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและหลักการบีบอัดสัญญาณเสียงพูด ซึ่งประกอบด้วย การบีบอัดสัญญาณเสียงโดยใช้ CELP การประมาณพหุเชิงเส้น (Linear Predictive Coding; LPC) คู่เส้นสเปกตรัม (Line Spectrum Pairs; LSP) การแปลงเวฟเล็ตไม่ต่อเนื่อง (Discrete Wavelet Transform; DWT) และงานวิจัยที่เกี่ยวข้องกับการบีบอัดสัญญาณเสียงพูด

#### 2.1 การบีบอัดสัญญาณเสียงพูด

สัญญาณเสียงพูดเป็นพื้นฐานการสื่อสารที่สำคัญของมนุษย์ เช่น การสื่อสารทางโทรศัพท์ และการสื่อสารบนเครือข่ายอินเทอร์เน็ต เป็นต้น ในการส่งสัญญาณเสียงพูดนั้นผู้ส่งจะต้องเปลี่ยนสัญญาณเสียงจากไมโครโฟนที่อยู่ในรูปแบบของสัญญาณอนาล็อก (Analog) ให้เป็นสัญญาณดิจิทัล (Digital) ก่อน ซึ่งโดยปกติแล้วเสียงพูดของมนุษย์จะมีความถี่ไม่เกิน 4 กิโลเฮิร์ตซ์ และจะมีอัตราการสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 8 กิโลเฮิร์ตซ์ (2 เท่าของความถี่สูงสุด) ตามทฤษฎีของไนควิสต์ (Nyquist Theory) ซึ่งในระบบสื่อสารจะเรียกเทคนิคนี้ว่าการมอดูเลตแบบเข้ารหัสพัลส์ (Pulse Code Modulation; PCM)



ภาพที่ 2.1 การส่งสัญญาณเสียงพูดในระบบสื่อสาร

จากภาพที่ 2.1 เป็นการส่งสัญญาณเสียงพูดในระบบสื่อสาร ในขั้นตอนแรกเมื่อมีสัญญาณเสียงผ่านไมโครโฟน สัญญาณเสียงนั้นจะเป็นสัญญาณแบบอนาล็อก ดังนั้นจึงจำเป็นต้อง



มีการแปลงสัญญาณเสียงจากอนาล็อกให้เป็นดิจิทัลก่อน ซึ่งจะเรียกสัญญาณนี้ว่า ค่าตัวอย่าง เมื่อได้ข้อมูลเสียงในรูปแบบของดิจิทัลแล้วก็สามารถส่งข้อมูลเสียงผ่านระบบสื่อสารไปให้กับผู้รับได้ แต่จะสังเกตได้ว่าการแปลงสัญญาณเสียงจากอนาล็อกเป็นดิจิทัลนั้นจะใช้อัตราการสุ่มตัวอย่าง 8000 เฮิรตซ์ (ใน 1 วินาที มีการสุ่มตัวอย่าง 8000 ครั้ง) ถ้าหากใช้ข้อมูลดิจิทัล 8 บิตต่อหนึ่งค่าตัวอย่าง แบนด์วิดท์ที่ใช้ในการส่งข้อมูลก็จะเป็น 64 กิโลบิตต่อวินาที (Kbps) และถ้าใช้ 16 บิตต่อค่าตัวอย่าง แบนด์วิดท์ที่เพิ่มเป็น 128 กิโลบิตต่อวินาที ซึ่งถือว่าเป็นค่าที่สูงมาก ยิ่งไปกว่านั้น ถ้าหากมีผู้ต้องการสื่อสารเสียงพร้อมๆ กันหลายคู่สนทนา ระบบที่มีแบนด์วิดท์อย่างจำกัดก็จะไม่สามารถรองรับได้ ดังนั้นต้องมีการลดขนาดของสัญญาณเสียงก่อนที่จะส่งไปยังระบบสื่อสาร ซึ่งเทคนิคนี้เรียกว่า การบีบอัดสัญญาณเสียงพูดสำหรับเทคนิคนี้แบนด์วิดท์จะเป็นส่วนที่สำคัญมากในการสื่อสารสัญญาณเสียง และปริมาณการใช้แบนด์วิดท์นั้นก็ขึ้นอยู่กับอัลกอริทึมที่ใช้ในการบีบอัดสัญญาณเสียง ถ้าแบนด์วิดท์ที่ใช้ในขณะนั้นไม่เพียงพอจะทำให้เกิดปัญหาต่างๆ ตามมา และปัจจัยเหล่านี้จะส่งผลให้คุณภาพของเสียงที่ผู้รับจะต่ำกว่าคุณภาพของเสียงต้นฉบับได้ ดังนั้นจึงต้องเลือกอัลกอริทึมหรือเทคนิคที่เหมาะสมกับงานนั้นๆ เพื่อให้ได้คุณภาพการบีบอัดสัญญาณที่สูงสุด โดยการเปรียบเทียบกับประสิทธิภาพของการเข้ารหัสเสียงพูด

ประสิทธิภาพของการเข้ารหัสเสียงพูดจะพิจารณาจากคุณสมบัติต่างๆ เช่น อัตราการเข้ารหัส (Bit Rate) คุณภาพเสียงที่ผ่านการเข้ารหัส (Speech Quality) ความซับซ้อนของการเข้ารหัส (Complexity มีหน่วยเป็น MIPS ย่อมาจาก Million Instructions Per Second) ค่าหน่วงเวลา (Delay Time) ความทนทานต่อความผิดพลาดที่เกิดภายในช่องสัญญาณ (Robustness) หรือการแทรกสอดที่เกิดจากเสียงอื่นๆ (Acoustic Interferences) เช่น เสียงรบกวน สัญญาณ DTMF ในระบบโทรศัพท์ สัญญาณของโมเด็ม เป็นต้น [2]

ในการสื่อสารข้อมูลแบบดิจิทัล คุณภาพของเสียงพูดถูกแบ่งออกเป็น 4 ระดับ ได้แก่

- 1) ระดับกระจายเสียง (Broadcast) เสียงพูดในระดับนี้จะอ้างถึงเสียงพูดบรรยายที่มีคุณภาพสูง โดยปกติจะมีอัตราการเข้ารหัสตั้งแต่ 64 กิโลบิตต่อวินาที ขึ้นไป
- 2) ระดับเครือข่าย (Toll หรือ Network) คุณภาพเสียงจะสามารถเทียบได้กับเสียงพูดในระบบอนาล็อก ในช่วงความถี่ 200-320 Hz โดยปกติจะมีอัตราการเข้ารหัสตั้งแต่ 16 กิโลบิตต่อวินาที ขึ้นไป
- 3) ระดับสื่อสาร (Communication) ขอมให้คุณภาพเสียงลดลงได้บ้างแต่ยังคงความเป็นธรรมชาติของเสียงอยู่ มีคุณภาพเพียงพอที่จะใช้ในการสื่อสาร สามารถสร้างได้โดยอัตราการเข้ารหัสตั้งแต่ 4.8 กิโลบิตต่อวินาที ขึ้นไป แต่ปัจจุบันมีเป้าหมายให้ลดลงมาที่ 4.0 กิโลบิตต่อวินาที

4) ระดับสังเคราะห์ (Synthetic) สามารถรับฟังได้เข้าใจแต่ไม่เป็นธรรมชาติ และสูญเสียคุณสมบัติในการรู้จำเจ้าของเสียงพูด มีอัตราการเข้ารหัสต่ำกว่า 4.0 กิโลบิตต่อวินาที

ซึ่งในปัจจุบันนี้อัลกอริทึมที่ใช้สำหรับการบีบอัดสัญญาณเสียงมีหลากหลายอัลกอริทึมและแต่ละประเภทจะมีสมรรถนะที่แตกต่างกัน ดังนั้นในงานวิจัยนี้ได้ทำการศึกษาเทคนิคการบีบอัด 2 เทคนิค คือ CELP และการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง

## 2.2 การเข้ารหัสสัญญาณเสียงพูดด้วย CELP

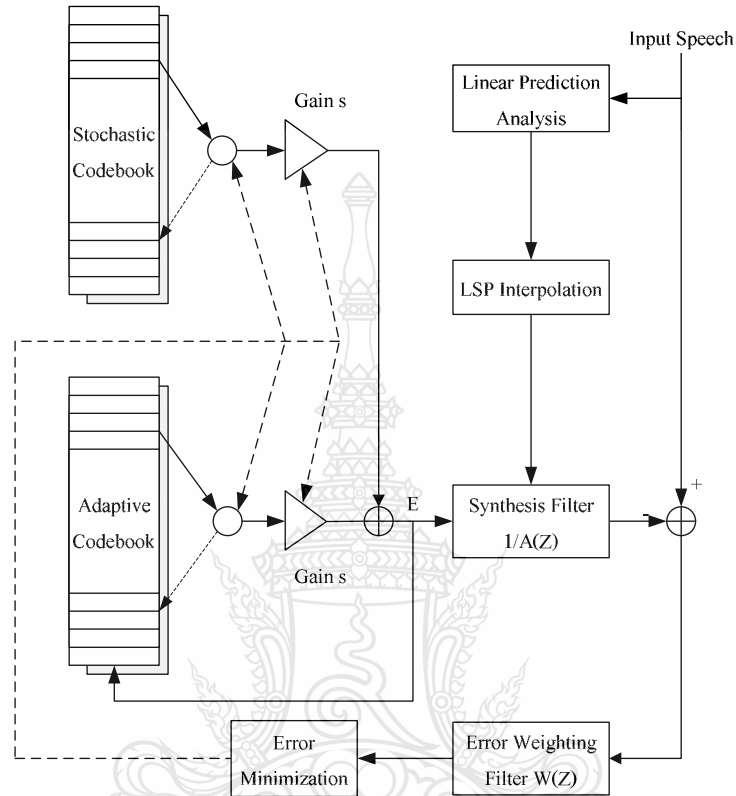
CELP ย่อมาจาก Code Excite Linear Prediction [3] เป็นเทคโนโลยีที่ใช้สำหรับการบีบอัดสัญญาณเสียงซึ่งได้รับมาตรฐานมากในระดับสากล และได้ถูกนำไปประยุกต์ใช้งานทางด้านงานประมวลผลมากมาย ซึ่งมาตรฐานของ CELP ยังสามารถแยกออกเป็นมาตรฐานย่อยอื่นๆ อีก แต่สำหรับในงานวิจัยนี้จะศึกษาเฉพาะมาตรฐานของสถาบันการวิจัยมาตรฐานและเทคโนโลยีแห่งชาติ Federal Standard 1016 CELP (FS1016 CELP) [4] ซึ่งมาตรฐานนี้ได้มีการพัฒนาร่วมกันระหว่างสหรัฐอเมริกาและห้องปฏิบัติการของ AT & Bell ซึ่งมาตรฐานนี้สามารถบีบอัดสัญญาณเสียงพูดที่ 4.8 กิโลบิตต่อวินาที ซึ่งจะเข้ารหัสโดยการแบ่งสัญญาณเสียงพูดเป็นเฟรม (Frame) ซึ่งในหนึ่งเฟรมจะใช้ อัตราการสุ่มตัวอย่าง (Sample Rate) 8 kHz และขนาดเฟรม (Frame Size) 30 วินาที (ประมาณ 240 ตัวอย่างต่อ 1 เฟรม) ซึ่งพารามิเตอร์และความต้องการอัตราบิตในการส่งข้อมูลแสดงดังตารางที่ 2.1

ตารางที่ 2.1 พารามิเตอร์และอัตราบิตที่ใช้ส่งข้อมูลของ FS1016 CELP [4]

พารามิเตอร์	บิตต่อเฟรม	บิตต่อวินาที
Adaptive Codebook	32	1066.67
Adaptive Codebook gain	20	666.67
Stochastic Codebook gain	36	1200
พารามิเตอร์ LSP	34	1133.33
รวม	142	4733.33 (4800)

การเข้ารหัสของ FS1016 CELP เริ่มจากการนำสัญญาณเสียงแปลงจากสัญญาณอนาล็อกเป็นดิจิทัล หลังจากนั้นสัญญาณจะถูกนำไปเข้ากระบวนการเข้ารหัส CELP บนพื้นฐานการวิเคราะห์ด้วยการสังเคราะห์ค่าถ่วงน้ำหนักของเวกเตอร์คอนไคซ์เซชัน (Vector Quantization; VQ) และจาก

การทำนายพัลลภเชิงเส้น (Linear Prediction) ซึ่งมีการกระตุ้นของสัญญาณด้วย Codebook ที่ใช้งาน 2 ส่วน คือ Adaptive Codebook และ Stochastic Codebook ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 บล็อกไดอะแกรมการเข้ารหัส FS 1016 CELP [5]

Adaptive Codebook จะถูกนำมาสร้างสัญญาณกระตุ้นความถี่ต่ำและ Stochastic Codebook จะถูกนำมาเพื่อแสดงความแตกต่างระหว่างรูปแบบของคลื่นที่เกิดขึ้นจริงและส่วนขยายระยะที่เหมาะสมของการกระตุ้น

จากที่กล่าวข้างต้นนั้นการทำงานของ CELP จะต้องอาศัยพารามิเตอร์จากการเข้ารหัสการทำนายพัลลภเชิงเส้น เพื่อนำค่าสัมประสิทธิ์ของวงจรรองสัญญาณที่ได้ไปใช้สังเคราะห์สัญญาณเสียงใน Adaptive Codebook ต่อไป

### 2.3 การเข้ารหัสการทำนายพัลลภเชิงเส้น

การทำนายพัลลภเชิงเส้น (Linear Predictive) [6-8] เป็นเทคนิคที่สำคัญทางด้านการวิเคราะห์เสียงเนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของเสียงพูดเมื่อเทียบกับ

ความเร็วในการประมวลผล หลักการพื้นฐานของการทำนายพัลส์เชิงเส้นอาศัยแนวความคิดว่า ตัวอย่างสัญญาณเสียงพูดสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณเสียงพูดจากอดีต การวิเคราะห์หาพารามิเตอร์เพื่อใช้ในการทำนายโดยทั่วไปเรียกว่าการเข้ารหัสการทำนายพัลส์เชิงเส้น (Linear Predictive Coding; LPC) ในด้านการประมวลผลสัญญาณเสียง การเข้ารหัสการทำนายพัลส์เชิงเส้นถูกนำไปใช้ในสองแนวทาง ได้แก่

### 2.3.1 การเข้ารหัสสัญญาณเสียง

การเข้ารหัสสัญญาณเสียงจะถูกนำไปใช้เป็นวงจรกรองวิเคราะห์การทำนายพัลส์เชิงเส้น (LP Analysis Filter) เพื่อแยกส่วนซ้ำซ้อน (Redundancy) ของสัญญาณเสียงออก ส่วนที่เหลือเรียกว่าสัญญาณตกค้าง (Residual Signal)

### 2.3.2 การสังเคราะห์สัญญาณเสียงพูด

การสังเคราะห์สัญญาณเสียงพูดจะถูกนำไปใช้เป็นวงจรกรองการทำนายพัลส์เชิงเส้นผกผัน (Inverse LP Filter) หรือวงจรกรองสังเคราะห์การทำนายพัลส์เชิงเส้น (LP Synthesis Filter) โดยที่ฟังก์ชันถ่ายโอนของวงจรกรองดังกล่าวแสดงกรอบสเปกตรัมของสัญญาณเสียงพูด วงจรกรองสังเคราะห์การทำนายพัลส์เชิงเส้นแสดงช่องทางเดินเสียงของมนุษย์และใช้หาสัญญาณกระตุ้นที่เหมาะสม

ในการวิเคราะห์การเข้ารหัสการทำนายพัลส์เชิงเส้นเริ่มต้นจากพิจารณากรอบสัญญาณเสียงที่มีตัวอย่าง  $N$  ตัวอย่าง คือ  $s_1, s_2, \dots, s_N$  โดยอ้างว่าตัวอย่างสัญญาณปัจจุบันสามารถทำนายได้จากผลรวมของตัวอย่างสัญญาณในอดีต  $p$  ตัวอย่าง ดังสมการที่ 2.1

$$\tilde{s}_n = -a_1s_{n-1} - a_2s_{n-2} - a_3s_{n-3} - \dots - a_p s_{n-p} = -\sum_{k=1}^p a_k s_{n-k} \quad (2.1)$$

เมื่อ  $p$  คืออันดับของการวิเคราะห์การทำนายพัลส์เชิงเส้น และ  $a_1, a_2, \dots, a_p$  คือสัมประสิทธิ์การเข้ารหัสการทำนายพัลส์เชิงเส้น กำหนด  $e_n$  แทนค่าผิดพลาดระหว่างค่าจริงและค่าที่ทำนายได้ จะได้ตามสมการที่ 2.2 และสมการที่ 2.3

$$e_n = s_n - \tilde{s}_n \quad (2.2)$$

$$e_n = s_n + \sum_{k=1}^p a_k s_{n-k} \quad (2.3)$$

สัญญาณ  $e_n$  เรียกว่าสัญญาณตกค้าง เนื่องจากสัญญาณ  $e_n$  ได้จากผลการลบสัญญาณ  $s_n$  ด้วย  $\hat{s}_n$  และเนื่องด้วยค่าสหสัมพันธ์ช่วงสั้น (Short-Term Correlation) ระหว่างตัวอย่างของสัญญาณตกค้างมีค่าต่ำ ดังนั้นประมาณได้ว่ากรอบสเปกตรัมกำลังของสัญญาณตกค้างมีลักษณะเรียบ เมื่อทำการแปลงแซด (Z-Transform) ของสมการที่ 2.4 ได้ค่าดังสมการที่ 2.5

$$E(z) = A(z) \cdot S(z) \quad (2.4)$$

โดยที่  $S(z)$  เป็นผลการแปลงแซดของสัญญาณเสียงและ  $E(z)$  และเป็นผลการแปลงแซดของสัญญาณตกค้างตามลำดับ

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.5)$$

โดยที่  $A(z)$  หรือวงจรกรองไวเทนนิ่ง (Whitening Filter) มีหน้าที่แยกความสหสัมพันธ์ช่วงสั้นที่ปรากฏในสัญญาณเสียงพูด ซึ่งก็คือการทำให้สเปกตรัมเรียบ เนื่องจาก  $E(z)$  ประมาณได้ว่ามีสเปกตรัมเรียบ ดังนั้นสามารถออกแบบกรอบสเปกตรัมของสัญญาณช่วงสั้นได้จากการวิเคราะห์การทำนายพันธะเชิงเส้นในรูปแบบออลโพล (All-Pole Model) หรือรูปแบบออโตรีเกรสซีฟ (Autoregressive Model) ดังสมการที่ 2.6

$$H(z) = \frac{1}{A(z)} \quad (2.6)$$

วงจรกรอง  $A(z)$  เรียกได้อีกชื่อหนึ่งว่าวงจรกรองผกผัน (Inverse Filter) เนื่องจากเป็นส่วนผกผันของรูปแบบออลโพล  $H(z)$  ของสัญญาณเสียงพูด และรากของ  $A(z)$  ทำให้เกิดโพลใน  $H(z)$  นั่นคือตำแหน่งฟอร์แมนต์ของเสียงที่ได้จากช่องทางเดินเสียงที่มีฟังก์ชันถ่ายโอน  $H(z)$

การหากรอบสเปกตรัมกำลังช่วงสั้นของเสียงพูดด้วยวิธีการวิเคราะห์การทำนายพันธะเชิงเส้น คำนวณได้จาก  $H(z)$  บนวงกลมหนึ่งหน่วย (Unit Circle) โดยในขั้นแรกต้องหาสัมประสิทธิ์การทำนายพันธะเชิงเส้นของสัญญาณเสียงพูดก่อน โดยปกติหาได้จากการทำให้ค่าผิดพลาดการทำนายพันธะเชิงเส้นทั้งหมดยกกำลังสอง ดังสมการที่ 2.7 มีค่าต่ำที่สุด

$$E = \sum_{n=n_1}^{n_2} e_n^2 \quad (2.7)$$

โดยที่ผลรวมของช่วง  $n_1$  ถึง  $n_2$  ที่คำนวณได้ขึ้นอยู่กับวิธีการที่ใช้ ซึ่งมีอยู่ 2 วิธี [1] ได้แก่

1) วิธีอัตโนมัติสัมพันธ์ (Autocorrelation) ในการวิเคราะห์การทำนายพันธะเชิงเส้นช่วงสั้นสามารถหาได้โดยใช้การวิเคราะห์แบบหน้าต่างสัญญาณเสียงพูดและอ้างว่าตัวอย่างสัญญาณภายนอกหน้าต่างนี้มีค่าเท่ากับศูนย์ ตามสมการที่ 2.8 แล้วจึงทำให้ได้ค่าผิดพลาดตามสมการที่ 2.7 มีค่าต่ำสุด

$$\sum_{k=1}^p r_{|i-k|} a_k = -r_i \quad \text{เมื่อ} \quad 1 \leq i \leq p \quad (2.8)$$

โดยที่  $r_k$  คือ ค่าสัมประสิทธิ์อัตโนมัติสัมพันธ์อันดับที่  $k$  ของหน้าต่างสัญญาณเสียง โดยที่

$$r_k = \frac{1}{N} \sum_{n=k}^N w_n s_n w_{n-k} s_{n-k} \quad (2.9)$$

เมื่อ  $w_n$  คือ ฟังก์ชันหน้าต่างที่มีระยะเวลา  $N$  ตัวอย่าง

การหาค่าสัมประสิทธิ์การเข้ารหัสการทำนายพันธะเชิงเส้นสามารถหาได้จากการแก้สมการที่ 2.8 ซึ่งมีจำนวน  $p$  สมการ สมการดังกล่าวเรียกว่าสมการยูล-วอล์กเกอร์ (Yule-Walker) สมการทั้งหมดสามารถเขียนในรูปของเมทริกได้ดังนี้

$$Ra = -r \quad (2.10)$$

โดยที่

$$R = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix} \quad (2.11)$$

$$a = [a_1, a_2, \dots, a_p]^T \quad (2.12)$$

$$r = [r_1, r_2, \dots, r_p]^T \quad (2.13)$$

เมตริก  $R$  เรียกว่า เมตริกอัตโนมัติสัมพันธ์ (Autocorrelation Matrix) ซึ่งมีลักษณะโครงสร้างแบบโทพลิตซ์ (Toeplitz) โครงสร้างนี้รับรองว่าโพลของวงจรรองส่งเคราะห์การทำนายพัลชะเชิงเส้น  $H(z)$  จะอยู่ภายในวงกลมหนึ่งหน่วย (Unit Circle) ดังนั้นจึงรับรองได้ว่าวงจรรองส่งเคราะห์  $H(z)$  ที่ได้จากวิธีอัตโนมัติสัมพันธ์นี้จะเสถียรเสมอ

สำหรับวิธีการคำนวณหาสัมประสิทธิ์การทำนายพัลชะเชิงเส้นในสมการที่ 2.10 มีอยู่หลายวิธี และวิธีหนึ่งที่น่าใช้คือ วิธีการวนซ้ำของเลวินสัน-เดอบิน (Levinson-Durbin Algorithm)

ขั้นตอนวิธีการวนซ้ำของเลวินสัน-เดอบิน แบ่งออกเป็น 4 ขั้นตอน ดังนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้น :  $E_0 = R(0)$  และ  $a_0 = 0$

ขั้นที่ 2 คำนวณค่าสัมประสิทธิ์การสะท้อน (Reflection Coefficient)

$$k_m = \frac{R(m) - \sum_{i=1}^{m-1} a_{m-1} R(m-i)}{E_{m-1}} ; \quad m = 1, 2, 3, \dots, p$$

เมื่อ  $R(m)$  และ  $R(m-i)$  คำนวณได้จากสมการ

$$R(m) = \sum_{n=m}^{N-1} x(n)x(n-m)$$

ขั้นที่ 3 คำนวณค่าสัมประสิทธิ์ของการทำนายพัลชะเชิงเส้น

ให้  $a_m(m) = k_m$

และ  $a_m(i) = a_{m-1}(i) - k_m a_{m-1}(m-i) ; 1 \leq i \leq m$

ขั้นที่ 4 คำนวณค่าผิดพลาดใหม่

$$E_m = (1 - k_m^2) E_{m-1}$$

$$m = m + 1$$

วนซ้ำขั้นที่ 2 ถึง 4 เมื่อ  $m < p$  เมื่อ  $m = p$  แล้ว  $a_i = a_p(i)$  โดยที่  $p$  คืออันดับของค่าสัมประสิทธิ์การทำนายพันธะเชิงเส้น

2) วิธีโคเวเรียนซ์ (Covariance Method) ในการวิเคราะห์การทำนายพันธะเชิงเส้น ช่วงของการรวมอยู่ในช่วง  $(p+1, N)$  ดังนั้นจึงไม่จำเป็นต้องใช้หน้าต่าง การทำให้ค่าผิดพลาดทั้งหมดยกกำลังสองมีค่าต่ำที่สุดหาได้จากสมการจำนวน  $p$  ดังสมการ 2.14

$$\sum_{k=1}^p c_{ik} a_k = -c_{i0} \quad \text{เมื่อ } 1 \leq i \leq p \quad (2.14)$$

โดยที่

$$c_{ik} = \sum_{n=p+1}^N s_{n-i} s_{n-k} \quad (2.15)$$

สมการจำนวน  $p$  สมการ สามารถเขียนในรูปของเมตริก ได้ดังนี้

$$Ca = -c \quad (2.16)$$

โดยที่

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2p} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & c_{p3} & \cdots & c_{pp} \end{bmatrix} \quad (2.17)$$

$$a = [a_1, a_2, \dots, a_p]^T \quad (2.18)$$

$$c = [c_{10}, c_{20}, \dots, c_{p0}]^T \quad (2.19)$$

เมตริก  $C$  เรียกว่า เมตริกโคเวเรียนซ์ (Covariance Matrix) และเป็นเมตริกสมมาตร นั่นคือ  $C_{1p} = C_{p1}$  แต่เมตริกไม่ได้มีโครงสร้างเป็นแบบโทพลิซ ดังนั้นการหาสัมประสิทธิ์การเข้ารหัสการทำนายพันธะเชิงเส้นด้วยวิธีนี้จึงมีประสิทธิภาพต่ำกว่าวิธีออสสัมพันธ์ และสัมประสิทธิ์การเข้ารหัสการทำนายพันธะเชิงเส้นที่ได้จากวิธีนี้จึงรับรองไม่ได้ว่าจะให้วงจรกรองสัณฐานที่มีความเสถียร



เสมอ นอกจากนี้ โครงสร้างที่สมมาตรทำให้การคำนวณบางส่วนสามารถใช้เทคนิคการคำนวณให้เร็วขึ้นได้ แต่ยังไม่เร็วเท่าวิธี Levinson-Debin

เนื่องจากพารามิเตอร์ LPC มีความเสถียรของสัญญาณที่ต่ำ ซึ่งอาจส่งผลกระทบต่อ การไปคืนกลับสัญญาณเสียงได้ ดังนั้นจึงได้นำค่าพารามิเตอร์จาก LPC ไปพัฒนาต่อด้วย LSP

## 2.4 คู่เส้นสเปกตรัม

คู่เส้นสเปกตรัม (Line Spectral Pairs; LSP) [6-8] หรือความถี่เส้นสเปกตรัม (Line Spectral Frequency; LSF) เป็นพารามิเตอร์รูปแบบหนึ่ง ที่พัฒนามาจากพารามิเตอร์การทำนายพันธะเชิงเส้น เนื่องจากพารามิเตอร์การทำนายพันธะเชิงเส้น ในขั้นตอนการประมาณค่าพารามิเตอร์อาจทำให้เกิดความไม่เสถียรของสัญญาณได้ ซึ่งส่งผลกระทบต่อคุณภาพของเสียง ในขณะที่พารามิเตอร์คู่เส้นสเปกตรัมมีคุณสมบัติที่เด่นคือ ค่าพารามิเตอร์อยู่ภายในขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์ และสามารถตรวจสอบเสถียรภาพของวงจรกรองได้ง่าย นอกจากนี้คู่เส้นสเปกตรัมยังแสดงในรูปเชิงความถี่จึงสามารถนำไปใช้ในการหาค่าคุณสมบัติที่แน่นอนในระบบการรับรู้ของคนได้

ในการคำนวณหาคู่เส้นสเปกตรัมเริ่มต้นจากพหุนามอันดับ  $M$  ของวงจรกรองผกผันในเชิงซัดตรงสมการที่ 2.5 โดยทำการแยกส่วนสมการดังกล่าวเป็นพหุนามอันดับ  $M+1$  จำนวน 2 พหุนามดังสมการที่ 2.20 และสมการที่ 2.21

$$P(z) = A(z) + z^{-(M+1)}A(z^{-1}) \quad (2.20)$$

$$Q(z) = A(z) - z^{-(M+1)}A(z^{-1}) \quad (2.21)$$

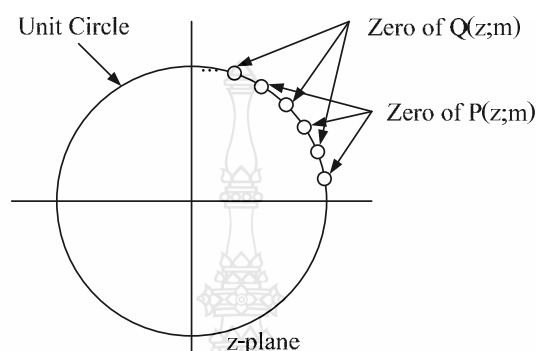
โดยพหุนาม  $P(z)$  และ  $Q(z)$  มีความสัมพันธ์กับ  $A(z)$  ตามสมการที่ 2.22

$$A(z) = \frac{P(z) + Q(z)}{2} \quad (2.22)$$

พหุนาม  $P(z)$  และ  $Q(z)$  สอดคล้องกับรูปแบบช่องทางเดินเสียงที่ไร้การสูญเสียขณะที่ช่องระหว่างเส้นเสียง (Glottis) ปิดและเปิดตามลำดับ และรากของพหุนาม  $P(z)$  และ  $Q(z)$  เรียกว่าความถี่เส้นสเปกตรัม โดยพหุนามทั้งสองมีคุณสมบัติดังต่อไปนี้

1) ราก (Zeroes) ทั้งหมดของพหุนาม  $P(z)$  และ  $Q(z)$  นั้นจะตั้งอยู่บนวงกลมหนึ่งหน่วยเสมอ

2) ราก (Zeroes) ของพหุนาม  $P(z)$  และ  $Q(z)$  จะวางเรียงสลับกันตามลำดับจากน้อยไปหามาก ดังแสดงในภาพที่ 2.3

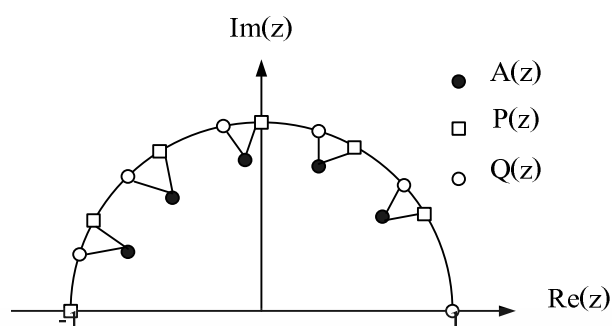


ภาพที่ 2.3 การวางเรียงสลับของรากของพหุนามคู่เส้นสเปกตรัม  $P(z)$  และ  $Q(z)$  [8]

ภาพที่ 2.3 แสดงให้เห็นได้ว่าวงจรรองพหุนาม  $A(z)$  จะมีเฟสต่ำสุด ถ้าคู่เส้นสเปกตรัมของ  $A(z)$  สอดคล้องกับคุณสมบัติทั้งสองนี้ ดังนั้นวงจรรังเคราะห์การเข้ารหัสการทำนายพันธะเชิงเส้นสามารถยืนยันได้ว่ามีเสถียรภาพ โดยการทำการประมวลผลพารามิเตอร์การเข้ารหัสการทำนายพันธะเชิงเส้นในรูปแบบคู่เส้นสเปกตรัม [1]

เมื่อพิจารณารากของพหุนามทั้งสองพบว่าพหุนาม  $P(z)$  และ  $Q(z)$  มีรากจริงอยู่ที่  $-1$  และ  $1$  ตามลำดับ สำหรับรากอื่นๆ อยู่บนวงกลมหนึ่งหน่วยโดยวางเรียงสลับกันตามคุณสมบัติของพหุนามทั้งสอง และรากทั้งสองของพหุนามมีลักษณะเป็นคู่เชิงซ้อนสังยุค ดังนั้นในการเก็บรากของพหุนามเพื่อใช้เป็นพารามิเตอร์จึงเก็บเพียง  $M$  ค่า

เนื่องจากรากของพหุนาม  $A(z)$  แสดงตำแหน่งฟอร์แมนต์ของเสียงพูด และพหุนาม  $P(z)$  และ  $Q(z)$  สัมพันธ์กับ  $A(z)$  ตามสมการที่ 2.20 - 2.22 ดังนั้นรากของพหุนามทั้งสองจึงสัมพันธ์กับฟอร์แมนต์ด้วย โดยรากของพหุนาม  $A(z)$  แต่ละอันจะจับคู่กับรากของพหุนาม  $P(z)$  และ  $Q(z)$  อย่างละหนึ่งราก ดังแสดงในภาพที่ 2.4



ภาพที่ 2.4 ความสัมพันธ์ระหว่างรากของ  $A(z)$  กับรากของกลุ่มสเปกตรัม  $P(z)$  และ  $Q(z)$  [8]

ภาพที่ 2.4 รูปวงกลมสีดำจะเป็นรากของพหุนาม  $A(z)$  ส่วนรูปสี่เหลี่ยมและรูปวงกลมสีขาวเป็นรากของพหุนาม  $P(z)$  และ  $Q(z)$  ตามลำดับ

การเกาะกลุ่มของรากหรือความถี่เส้นสเปกตรัมจำนวน 2 หรือ 3 รากจะกำหนดลักษณะความถี่ฟอร์แมนต์และขนาดแบนด์วิดท์ของฟอร์แมนต์ โดยขึ้นกับความชิดของกลุ่มรากดังกล่าว ถ้ารากชิดกันมากจะทำให้เกิดฟอร์แมนต์และมีแบนด์วิดท์แคบ ในทางตรงข้ามกลุ่มรากจะแสดงถึงสเปกตรัมที่มีแบนด์วิดท์กว้างคือไม่เกิดฟอร์แมนต์ นอกจากนี้คุณสมบัติอีกอย่างหนึ่งของกลุ่มเส้นสเปกตรัมคือความไวทางสเปกตรัมของกลุ่มเส้นสเปกตรัมมีลักษณะเฉพาะที่ คือเมื่อมีการเปลี่ยนแปลงกลุ่มเส้นสเปกตรัมใดๆ จะทำให้เกิดการเปลี่ยนแปลงของสเปกตรัมกำลังของการเข้ารหัสการทำนายพันธะเชิงเส้นเฉพาะบริเวณรอบๆ เท่านั้น ทำให้สามารถทำการประเมินค่า (Quantization) ได้อย่างอิสระโดยไม่มีผลกระทบจากการลดทอนเนื่องจากการประเมินค่าจากสเปกตรัมหนึ่งไปสู่สเปกตรัมอื่นๆ

งานวิจัยนี้ได้ศึกษาเกี่ยวกับการบีบอัดสัญญาณเสียงพูดโดยใช้หลักการของ DWT เพื่อหาประสิทธิภาพที่ดีที่สุดสำหรับการบีบอัดสัญญาณเสียงเมื่อเปรียบเทียบกับวิธีการบีบอัดสัญญาณเสียงพูดด้วย CELP

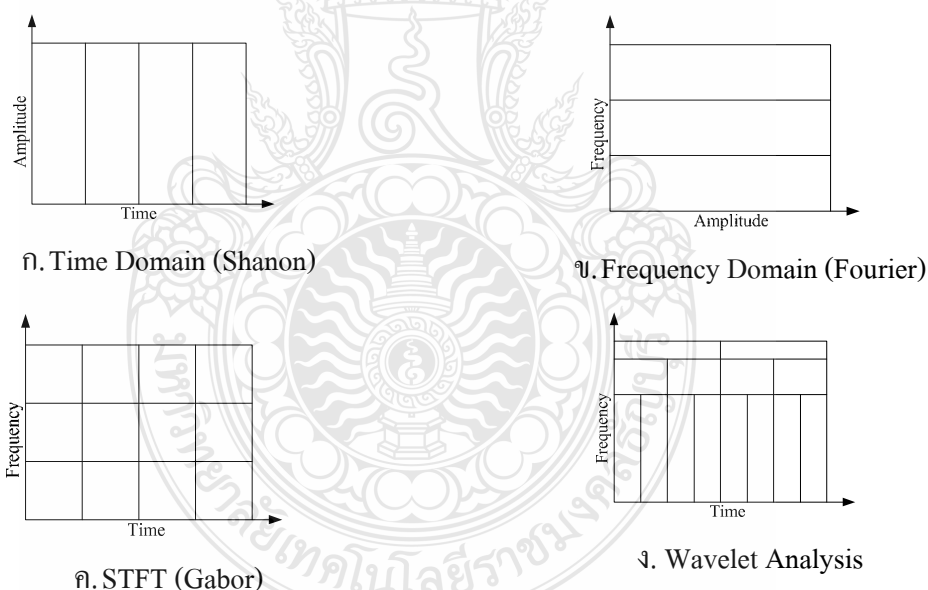
## 2.5 การแปลงเวฟเล็ต

การแปลงเวฟเล็ต (Wavelet Transform) [9-11] เป็นคณิตศาสตร์ที่ใช้ในการวิเคราะห์และสังเคราะห์ลักษณะของสัญญาณซึ่งมีประโยชน์มากในงานทางด้านการประมวลผลสัญญาณ (Signal Processing) ทฤษฎีเวฟเล็ตสามารถนำมาประยุกต์เพื่ออธิบายลักษณะของสิ่งต่างๆ หรือระบบใดๆ ได้ เช่น การหาขอบภาพ การลดขนาดข้อมูลภาพหรือสัญญาณ การรู้จำภาพหรือเสียง เป็นต้น การแปลงเวฟเล็ตมีพื้นฐานมาจากการแปลงฟูเรียร์ ซึ่งสามารถวิเคราะห์สัญญาณเฉพาะบางช่วงเวลาและความถี่เท่านั้นและจะเสียเวลาในการคำนวณมากเพราะต้องคำนวณใหม่ตลอดย่าน ดังนั้นจึงได้มีการ

พัฒนาการเปลี่ยนแปลงที่สามารถวิเคราะห์สัญญาณได้เฉพาะช่วงเวลาและช่วงความถี่ที่สนใจเท่านั้น ทำให้สามารถพิจารณาผลกระทบของการเปลี่ยนแปลงของสัญญาณเฉพาะช่วงได้อย่างสะดวกและรวดเร็ว การแปลงที่พัฒนาขึ้นนี้เรียกว่า “การแปลงเวฟเล็ต” ที่สามารถกระทำการแปลงผ่านฟังก์ชันหน้าต่างโดยสามารถเลื่อนตำแหน่งและปรับเปลี่ยนช่วงเวลาให้เหมาะสมกับความถี่ที่จะวิเคราะห์ได้ [12]

### 2.5.1 การวิเคราะห์เวฟเล็ต (Wavelet Analysis)

การวิเคราะห์เวฟเล็ต [9-14] เป็นระดับการพัฒนาต่อมาจากการแปลงฟูเรียร์และการวิเคราะห์ฟูเรียร์ในช่วงเวลาสั้นๆ ที่สามารถปรับเปลี่ยนขนาดของวินโดว์ได้ กล่าวคือสามารถปรับเปลี่ยนช่วงเวลาให้เหมาะสมกับช่วงความถี่ที่จะวิเคราะห์ได้ โดยที่สัญญาณความถี่สูงจะมีช่วงคาบเวลาในการวิเคราะห์ที่น้อยลง ในขณะที่สัญญาณความถี่ต่ำลงจะมีช่วงคาบเวลาในการวิเคราะห์ที่กว้างขึ้น ดังภาพที่ 2.5 จะเป็นการเปรียบเทียบเวลาและความถี่ของสัญญาณของแต่ละรูปแบบการวิเคราะห์



ภาพที่ 2.5 ลักษณะของระนาบเวลาและความถี่สำหรับการวิเคราะห์สัญญาณ [11]

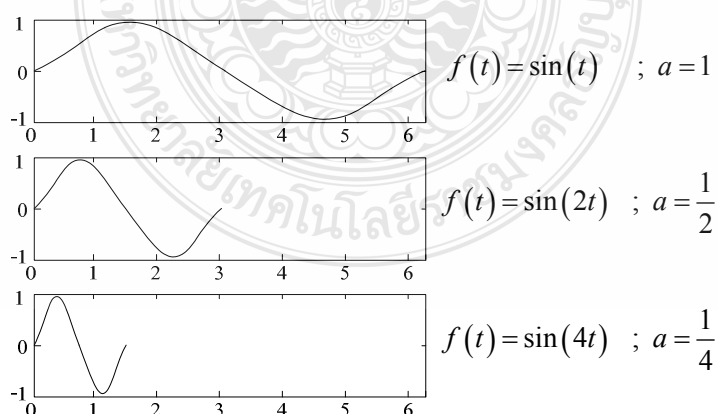
ลักษณะพิเศษของฟังก์ชันเวฟเล็ตจะมีลักษณะของการออสซิลเลต (Oscillation) ตามแนวแกนนอนและมีแอมพลิจูด (Amplitude) ที่ลดลงสู่ศูนย์ทั้งทางด้านบวกและด้านลบซึ่งแตกต่างกับฟังก์ชันไซน์และโคไซน์ซึ่งแผ่ไปตามแกนแนวนอนไปสู่อนันต์

การนำเวฟเล็ตหลายๆ อันมารวมกันเป็นเซตเพื่ออธิบายสัญญาณใดๆ โดยที่คลื่นเวฟเล็ตแต่ละตัวจะมีโครงสร้างมาจากฟังก์ชันเดียวกัน ฟังก์ชันนี้จะเป็นเวฟเล็ตต้นแบบที่เรียกว่า “เวฟเล็ตแม่” (Mother Wavelet) คลื่นเวฟเล็ตแต่ละอันจะอยู่ภายในเซตของเวฟเล็ตนี้ โดยคลื่นแต่ละคลื่นจะเกิดจากการสเกลและการเลื่อนตำแหน่ง ของเวฟเล็ตแม่ การสเกลจะหมายถึงการหดเข้าหรือการขยายออก ในที่นี้จะให้ “a” แทนพารามิเตอร์ของการหดหรือขยาย ซึ่งจะเป็นการเปลี่ยนความถี่ของเวฟเล็ตแม่ นั่นเอง ส่วนการเลื่อนตำแหน่งกำหนดโดยพารามิเตอร์ “b” จะหมายถึงการเลื่อนตำแหน่งการเกิดของคลื่นเวฟเล็ตบนแกนเวลา ซึ่งจะแสดงเวฟเล็ตที่ถูกสเกลและถูกเลื่อนตำแหน่งไปที่ค่า “a” และ “b” ต่างๆ กัน และเวฟเล็ตที่ถูกสเกลจะต้องทำการ Normalize ด้วย  $1/\sqrt{a}$  เสมอ เพื่อให้เวฟเล็ตที่ถูกสเกลไปแล้วยังคงมีพลังงานเท่ากับ เวฟเล็ตแม่ ดังนั้นถ้าให้  $\psi(t)$  เป็นฟังก์ชันเวฟเล็ตแม่ จะสามารถเขียนเป็นสมการทั่วไปของเวฟเล็ตที่ตำแหน่ง a, b ใดๆ ซึ่งสัมพันธ์กับเวฟเล็ตแม่ได้ดังนี้

$$\psi_{b,a}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) = \psi(\text{scale}, \text{position}, t) \quad (2.1)$$

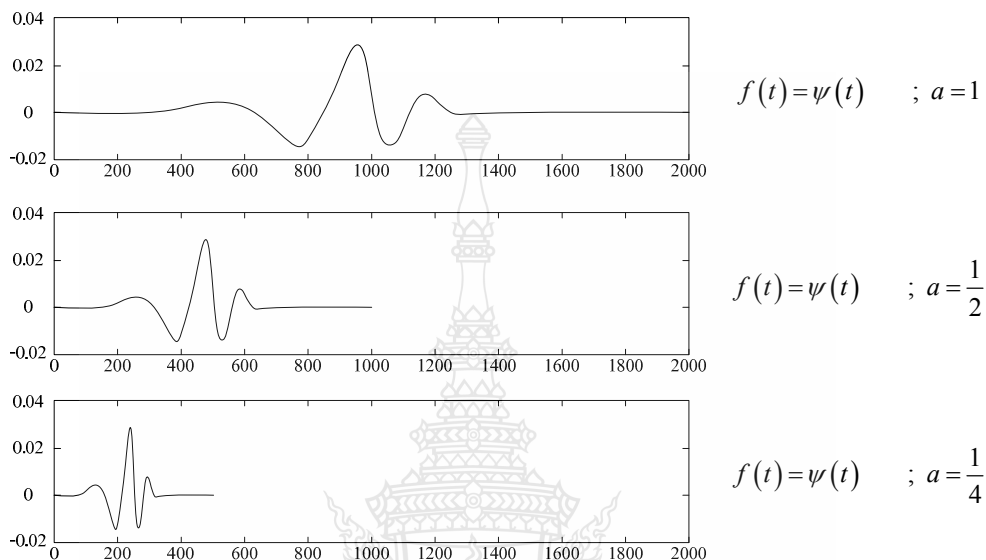
เพื่อให้เห็นภาพได้ชัดเจนยิ่งขึ้น จะขออธิบายถึงเวฟเล็ตที่มีการสเกลและเลื่อนตำแหน่งดังนี้

1) ขนาด (Scaling) จะหมายถึงการหดเข้า (Compressing) หรือการขยายออก (Dilation) ซึ่งเราจะให้ “a” แทนด้วยพารามิเตอร์ของการหดหรือขยายซึ่งจะเป็นการเปลี่ยนความถี่ของเวฟเล็ตแม่ นั่นเอง หรืออาจเรียก “a” ว่าเป็นตัวประกอบขนาด (Scale Factor) และถ้าพิจารณาสัญญาณไซน์ที่มีตัวประกอบขนาดแตกต่างกันจะมีลักษณะดังภาพที่ 2.6



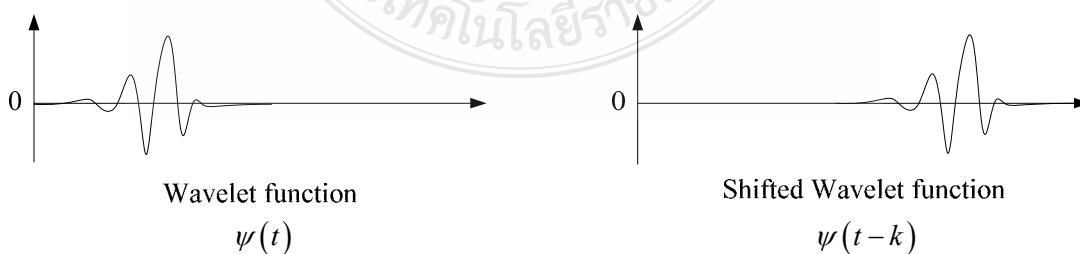
ภาพที่ 2.6 ลักษณะสัญญาณไซน์ที่มีตัวประกอบแตกต่างกัน [11]

ตัวประกอบขนาด ( $a$ ) จะมีผลในทางเดียวกันกับเวฟเล็ต ถ้าตัวประกอบขนาดน้อยลงเท่าใด เวฟเล็ตจะถูกบีบอัดมากขึ้นเท่านั้น เมื่อพิจารณาเวฟเล็ตที่มีตัวประกอบแตกต่างกันจะปรากฏลักษณะดังภาพที่ 2.7



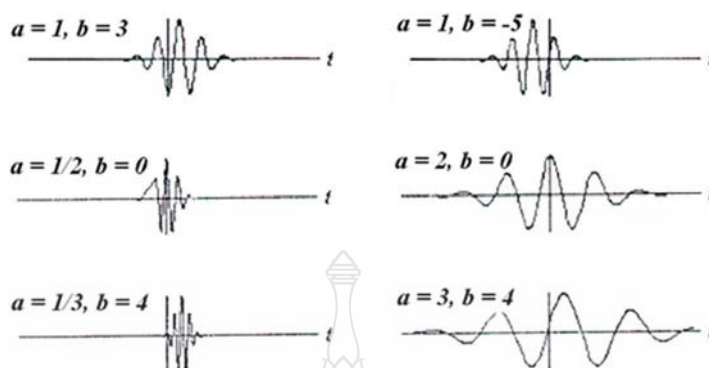
ภาพที่ 2.7 ลักษณะเวฟเล็ตที่มีตัวประกอบขนาดแตกต่างกัน [11]

2) การเลื่อน (Shifting) ของเวฟเล็ตจะมีความหมายง่ายๆ คือ การเลื่อนสัญญาณในทางคณิตศาสตร์ซึ่งการเลื่อนของฟังก์ชัน  $\psi(t)$  ด้วย  $k$  จะแสดงได้โดย  $\psi(t-k)$  การเลื่อนตำแหน่งจะกำหนดโดยพารามิเตอร์ “ $k$ ” ซึ่งจะหมายถึงการเลื่อนตำแหน่งการเกิดคลื่นเวฟเล็ตบนแกนเวลา ในตำราบางเล่มจะแทนพารามิเตอร์ “ $b$ ” เป็นการเลื่อนตำแหน่ง เมื่อพิจารณาเวฟเล็ตที่มีการเลื่อนที่ตำแหน่งใดๆ สามารถพิจารณาได้ดังภาพที่ 2.8



ภาพที่ 2.8 ลักษณะของเวฟเล็ตที่มีการเลื่อนตำแหน่ง [11]

การแสดงคลื่นเวฟเล็ตที่มีการสเกลและการเลื่อนตำแหน่งได้ดังภาพที่ 2.9

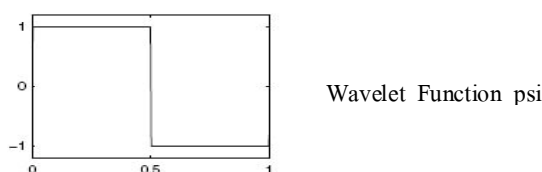


ภาพที่ 2.9 ลักษณะของเวฟเล็ตแม่ที่ถูกสเกลและเลื่อนตำแหน่งไปที่ค่า  $a, b$  ต่างๆ กัน [11]

### 2.5.2 ตระกูลของเวฟเล็ตแม่แบบต่างๆ

ปัจจุบันนี้เวฟเล็ตที่นำมาใช้งานทางด้านการประมวลผลสัญญาณมีหลายตระกูล [9-14] ที่สามารถเลือกใช้ได้เหมาะสมสำหรับแต่ละงาน โดยสามารถแบ่งเป็นเวฟเล็ต Biorthogonal และเวฟเล็ตเชิงตั้งฉากปกติ [14] เช่น Daubechies, Symlet และ Coiflet เป็นต้น แต่ละตระกูลจะเป็นฟังก์ชันพื้นฐานที่มีรูปร่างลักษณะที่แตกต่างกันไป ซึ่งแต่ละตระกูลจะมีค่า Number of Vanishing Moments (NVM) ที่ายชื่อตระกูลเช่น Daubechies 4, 8, ..., 20, Symlet 4, 5, 6, ..., 10 และ Coiflet 1, 2, ..., 5 เป็นต้น ถ้าค่า NVM มีค่ามากขึ้นลักษณะของฟังก์ชันพื้นฐานที่เลือกจะมีความราบเรียบ (Smooth) มากขึ้น ประโยชน์ของค่า NVM นี้ก็คือสามารถเลือกชนิดของเวฟเล็ตมาประยุกต์ใช้กับงานที่ต้องการได้อย่างหลากหลายและเหมาะสมมากขึ้น ตระกูลของเวฟเล็ตแม่ที่สำคัญและนิยมใช้กันอย่างแพร่หลายในปัจจุบัน ได้แก่

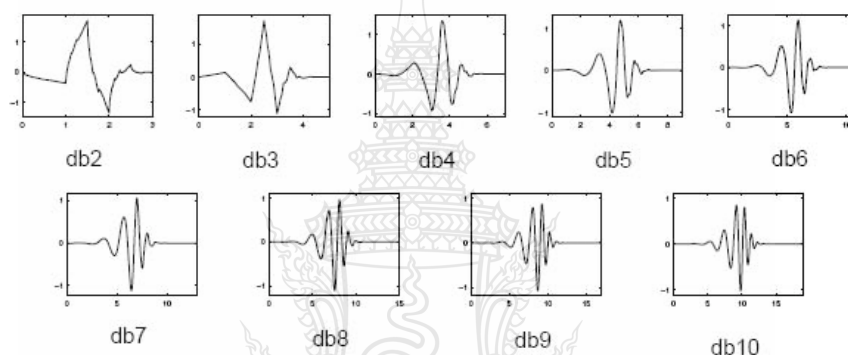
1) เวฟเล็ต Haar เป็นเวฟเล็ตชนิดแรกและสามารถประมวลผลง่าย ดังนั้นจะถูกนำไปวิเคราะห์สัญญาณในช่วงแรกๆ โดยลักษณะของเวฟเล็ตชนิดนี้จะคล้ายกับฟังก์ชันขั้นบันได ซึ่งแสดงดังภาพที่ 2.10



ภาพที่ 2.10 ลักษณะของเวฟเล็ต Haar [11]

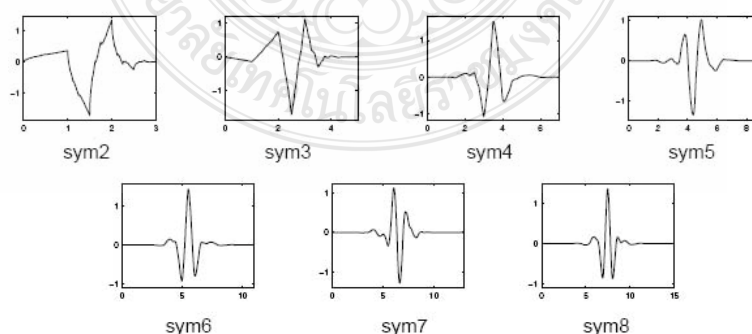
2) เวฟเลต Daubechies เป็นเวฟเลตที่ถูกคิดค้นโดย Ingrid Daubechies ซึ่งเป็นนักวิจัยทางด้านเวฟเลตที่ประสบผลสำเร็จและสำคัญคนหนึ่งของโลก เขาได้สร้างสิ่งที่เรียกว่า “ขอบเขตที่แน่นอนของเวฟเลตเชิงตั้งฉากปกติ” (Compactly - Subport Orthonormal Wavelet) ซึ่งเป็นการสร้างการวิเคราะห์เวฟเลตแบบไม่ต่อเนื่องในทางปฏิบัติขึ้นมาได้สำเร็จ และได้กำหนดค่า NVM ของตระกูลเวฟเลตขึ้นมาเพื่อสร้างความหลากหลายในการเลือกใช้งานได้อย่างเหมาะสม ค่า NVM นี้จะกำกับไว้หลังชื่อตระกูลของเวฟเลตเช่น Daubechies 4 ก็คือมีค่า NVM เท่ากับ 4

สำหรับเวฟเลตตระกูล Daubechies นี้จะมีลักษณะดังภาพที่ 2.11 ซึ่งมีหลายหลายรูปแบบตามค่า NVM



ภาพที่ 2.11 ลักษณะของเวฟเลต Daubechies [11]

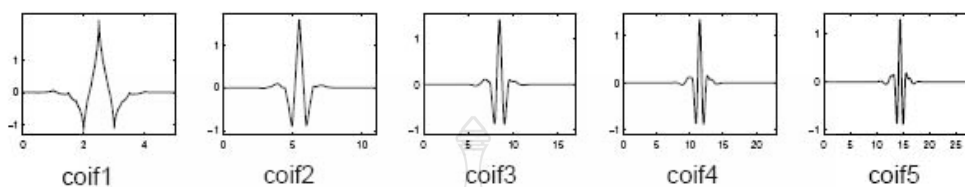
3) เวฟเลต Symlets จะมีสัดส่วนใกล้เคียงกับเวฟเลตที่ถูกเสนอโดย Ingrid Daubechies ซึ่งได้ปรับปรุงมาจากเวฟเลตตระกูล Daubechies คุณสมบัติของเวฟเลตทั้งสองตระกูลนี้จะคล้ายคลึงกัน ลักษณะของเวฟเลตชนิดนี้จะแสดงดังภาพที่ 2.12



ภาพที่ 2.12 ลักษณะของเวฟเลต Symlet [11]

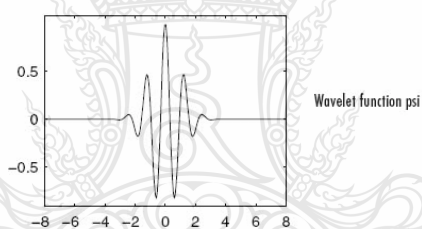


4) เวฟเลต Coiflets เป็นเวฟเลตที่ถูกสร้างขึ้น โดย Ingrid Daubechies โดยการเรียกร่องจาก R. Coifman ภายหลังจากที่ Ingrid Daubechies ประสบความสำเร็จจากการสร้างเวฟเลตชนิด Daubechies มาแล้ว เวฟเลตชนิดนี้จะมีลักษณะดังภาพที่ 2.13



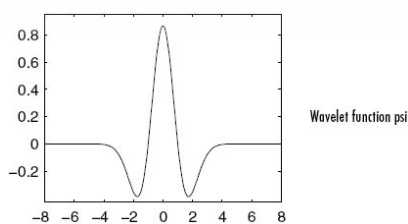
ภาพที่ 2.13 ลักษณะของเวฟเลต Coiflets [11]

5) เวฟเลต Morlet เป็นเวฟเลตที่ไม่มีฟังก์ชันสเกลลิง (Scaling Function) แต่จะมีความแน่นอน สามารถพิจารณาลักษณะของเวฟเลตชนิดนี้ได้ภาพที่ 2.14



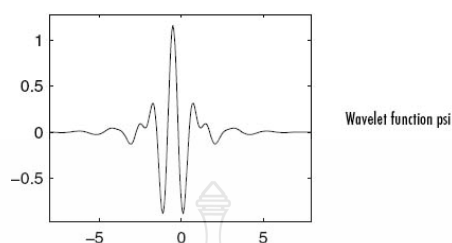
ภาพที่ 2.14 ลักษณะของเวฟเลต Morlet [11]

6) เวฟเลต Mexican Hat เป็นเวฟเลตที่ไม่เป็นฟังก์ชันสเกลลิง (Scaling Function) และมีที่มาจากฟังก์ชันอนุพันธ์อันดับที่สองของฟังก์ชันความหนาแน่นที่มีความน่าจะเป็นของเกาส์เซียน (The Second Derivative Function of Gaussian Probability Density Function) ลักษณะของเวฟเลตชนิดนี้แสดงดังภาพที่ 2.15



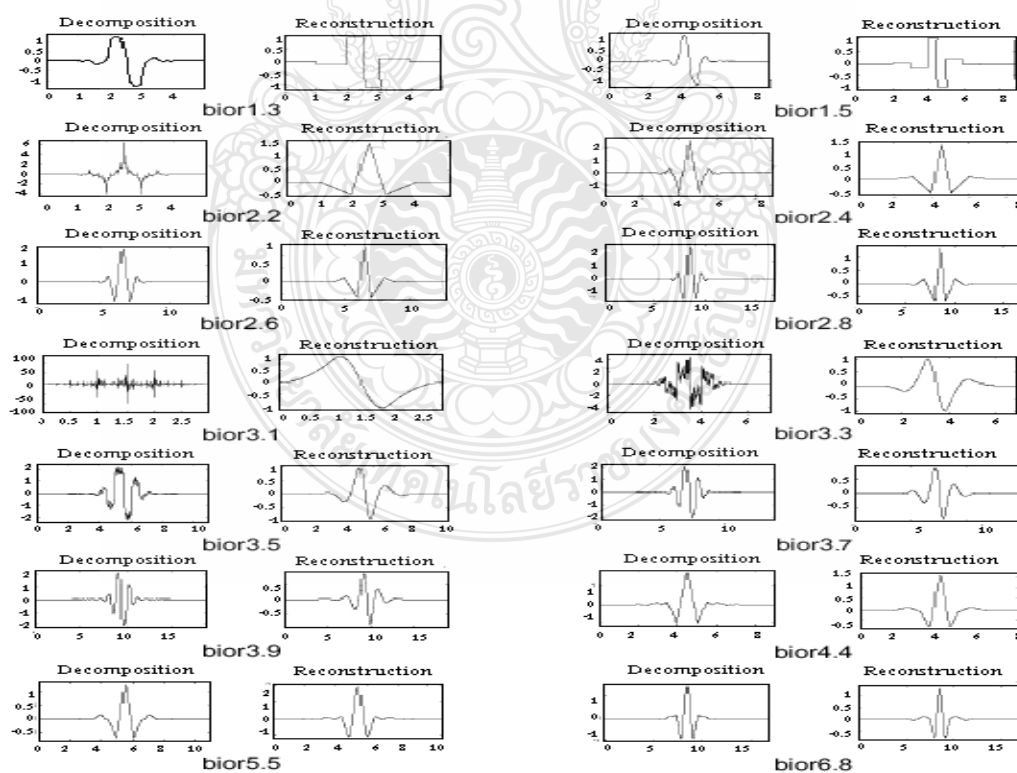
ภาพที่ 2.15 ลักษณะของเวฟเลต Mexican Hat [11]

7) เวฟเลต Meyer เป็นเวฟเลตที่มีฟังก์ชันการสเกลลิงกำหนดไว้ในโดเมนความถี่ ซึ่งสามารถพิจารณาลักษณะของเวฟเลตชนิดนี้ได้จากภาพที่ 2.16



ภาพที่ 2.16 ลักษณะของ Meyer Wavelet [11]

8) เวฟเลต Biorthogonal [11, 14] กลุ่มของเวฟเลตชนิดนี้จะแสดงคุณสมบัติของเฟสเชิงเส้น ซึ่งจำเป็นสำหรับการสร้างสัญญาณและภาพขึ้นใหม่ โดยการใช้เวฟเลต 2 ตัว ตัวหนึ่งสำหรับแยกตัวประกอบ อีกตัวหนึ่งสำหรับการสร้างใหม่แทนอีกตัวหนึ่งที่เหมือนกัน ที่มีคุณสมบัติที่น่าสนใจ ลักษณะของเวฟเลต Biorthogonal นี้แสดงดังภาพที่ 2.17



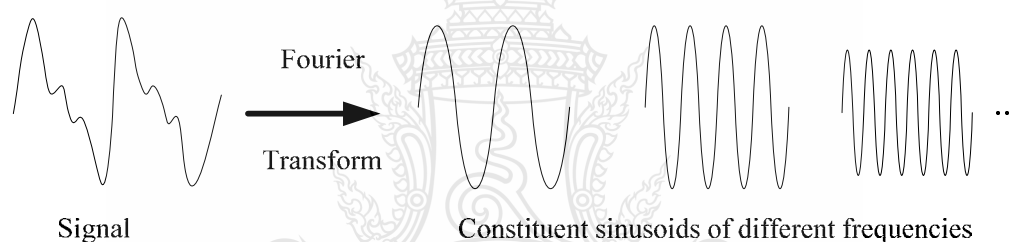
ภาพที่ 2.17 ลักษณะของเวฟเลต Biorthogonal [11]

### 2.5.3 การแปลงเวฟเลตแบบต่อเนื่อง (Continuous Wavelet Transform: CWT) [11]

ในทางคณิตศาสตร์กระบวนการวิเคราะห์ฟูเรียร์ จะถูกแสดงโดยการแปลงฟูเรียร์ (Fourier Transform)

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (2.23)$$

ซึ่งก็คือผลบวกตลอดทั้งช่วงเวลาของสัญญาณ  $f(t)$  คูณด้วยเอ็กโพเนนเชียลเชิงซ้อน (เรียกว่าเอ็กโพเนนเชียลเชิงซ้อนนั้นสามารถแตกออกมาได้เป็นส่วนไซน์ค่าจริงและเชิงซ้อน) ผลของการแปลงสัมประสิทธิ์ฟูเรียร์  $F(\omega)$  (Fourier Coefficient) ซึ่งเมื่อถูกคูณด้วยสัญญาณ Sine ที่มีความเร็วเชิงมุม  $\omega$  ที่เหมาะสมแล้ว จะได้ส่วนประกอบสัญญาณ Sine ของสัญญาณต้นแบบ ซึ่งกระบวนการจะเห็นได้ดังภาพที่ 2.18



ภาพที่ 2.18 การแปลงฟูเรียร์ที่ประกอบด้วยสัญญาณไซน์ที่มีความถี่แตกต่างกัน [11]

การแปลงเวฟเลตที่มีความต่อเนื่อง กำหนดได้ว่าเป็นผลบวกตลอดช่วงเวลาของสัญญาณ คูณด้วยขนาดและการเลื่อนของฟังก์ชันเวฟเลต  $\psi$

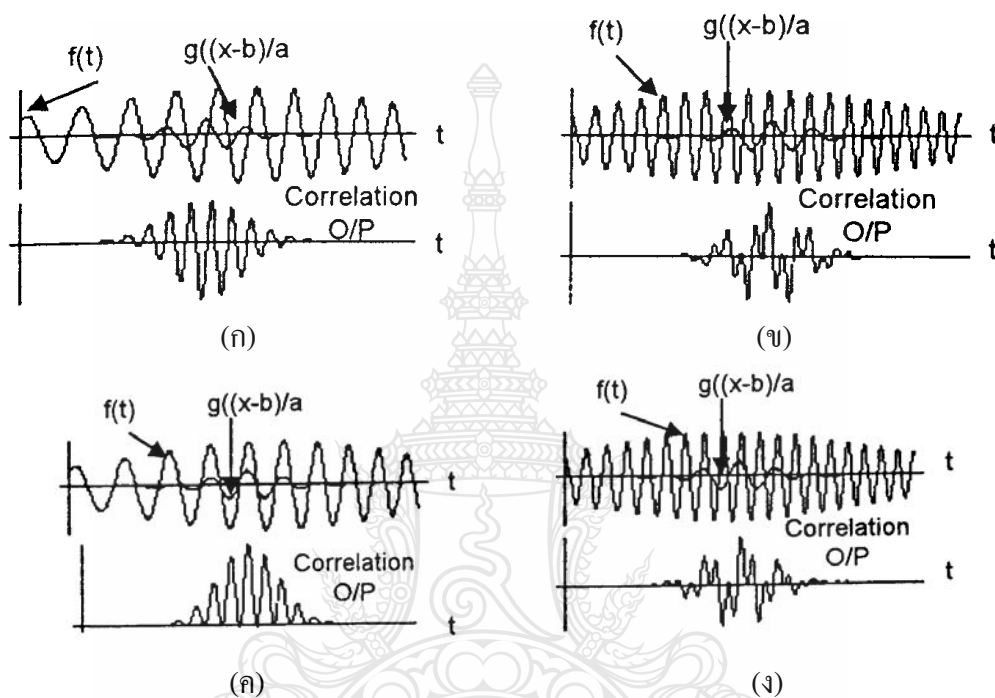
$$C(\text{scale}, \text{position}) = \int_{-\infty}^{\infty} f(t) \psi(\text{scale}, \text{position}, t) dt \quad (2.24)$$

หรือ

$$C(a, b) = \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

โดยที่ “ $a$ ” เป็นพารามิเตอร์ของการสเกลและ “ $b$ ” เป็นพารามิเตอร์ของการเลื่อนตำแหน่ง ส่วน  $c(a, b)$  จะเป็นสัญญาณอินพุตที่ขึ้นอยู่กับตัวแปรอิสระ  $t$  ซึ่งจะถูกลส่ง (Map) ไปเป็นฟังก์ชัน

2 มิติที่ขึ้นอยู่กับตัวแปรอิสระ  $a$  และ  $b$  โดยที่  $c(a,b)$  จะเป็นค่าสัมประสิทธิ์ของเวฟเล็ต (Wavelet Coefficient) เวฟเล็ตแม่ที่ถูกสเกลและเลื่อนตำแหน่งจะใช้เป็นตัวแทนสัญญาณได้ สัมประสิทธิ์ของเวฟเล็ตที่ตำแหน่ง  $a, b$  จะเป็นค่าแสดงความเหมือนกัน (Degree of Correlation) ระหว่างเวฟเล็ตแม่ที่ถูกสเกล ( $a$ ) กับสัญญาณอินพุตที่ตำแหน่งเวลานั้น ( $b$ ) ดังภาพที่ 2.19



ภาพที่ 2.19 ลักษณะของการ Correlation ระหว่าง  $f(t), \psi\left(\frac{t-b}{a}\right)$  [11]

จากภาพที่ 2.19 (ก) จะพบว่าคลื่นเวฟเล็ตมีความถี่และเฟสใกล้เคียงกับสัญญาณอินพุตมาก ทำให้ปรากฏคลื่นเอาต์พุตออกมามีค่าสูง แต่ในภาพที่ 2.19 (ข) คลื่นอินพุตจะมีค่าความถี่สูงกว่าคลื่นเวฟเล็ตทำให้เอาต์พุตมีค่าต่ำกว่ามาก และในขณะเดียวกันเมื่อลองพิจารณาเปรียบเทียบภาพที่ 2.19 (ก) กับภาพที่ 2.19 (ค) ซึ่งสัญญาณอินพุตมีความถี่เดียวกันแต่ทำการเปรียบเทียบที่ตำแหน่งต่างกัน โดยในภาพที่ 2.19 (ก) จะทำที่ตำแหน่งเฟสตรงกัน ซึ่งก็จะให้ผลการ Correlation ที่สูงกว่าภาพที่ 2.19 (ค) ซึ่งกระทำที่ตำแหน่งเฟสไม่ตรงกัน ส่วนในภาพที่ 2.15 (ง) ก็คล้ายกับภาพที่ 2.19 (ข) เพียงแต่เลื่อนตำแหน่งการเปรียบเทียบออกไป 180 องศา ซึ่งจะสังเกตเห็นว่าจะให้ผลลัพธ์ต่างกัน ดังนั้นจึงสามารถสรุปได้ว่า ถ้าคลื่นเวฟเล็ตมีความคล้ายกับสัญญาณอินพุตเท่าไรทั้งในด้านความถี่ ลักษณะและตำแหน่งเวลาในการเปรียบเทียบ ก็จะทำให้ได้ค่าของสัมประสิทธิ์ของการทรานส์ฟอร์มที่มาก

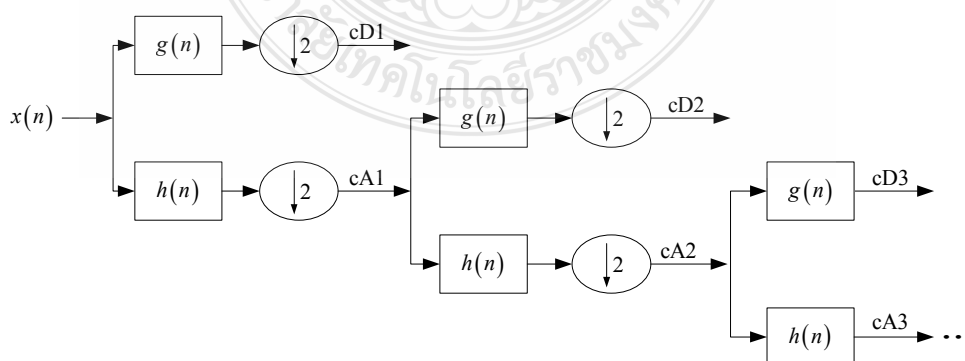
เพราะฉะนั้นการทำการแปลงเวฟเลตจะเป็นการอธิบายสัญญาณ  $f(t)$  ในรูปแบบของเวฟเลตโดเมน (Wavelet Domain) เทียบกับคลื่นเวฟเลตที่ค่า  $a, b$  ใดๆ ซึ่งผลลัพธ์ของการแปลงเวฟเลตแบบต่อเนื่อง (CWT) นี้ จะได้เป็นสัมประสิทธิ์เวฟเลต C หลายตัวซึ่งเป็นฟังก์ชันของขนาดและตำแหน่ง

#### 2.5.4 การแปลงเวฟเลตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform)

จากหัวข้อการแปลงเวฟเลตแบบต่อเนื่องที่ได้กล่าวมาข้างต้นจะพบว่าในทางปฏิบัติแล้วสัญญาณที่นำมาวิเคราะห์โดยคอมพิวเตอร์จะมีลักษณะแบบแบ่งช่วง หรือเป็นสัญญาณที่ถูกแซมปลิง (Sampling) เข้ามา ดังนั้นจึงได้มีพัฒนาการวิเคราะห์การแปลงเวฟเลตแบบไม่ต่อเนื่องขึ้น (Discrete Wavelet Transform: DWT) โดยจะใช้ตัวกรองในกระบวนการแปลง ซึ่งถูกพัฒนาขึ้นโดย Mallat ในปี 1988 [13] ระเบียบวิธีการของ Mallat เป็นแบบแผนที่รู้จักกันในกลุ่มผู้ที่ทำการแปลงสัญญาณว่าเป็นการเข้ารหัสแบบ 2 ช่องสัญญาณย่อย (Two-channel Subband Coder) [9-14]

สำหรับหลายๆ สัญญาณจะมีส่วนความถี่ต่ำเป็นส่วนสำคัญมากที่สุด ซึ่งจะให้ลักษณะของสัญญาณ ในทางตรงกันข้าม ส่วนความถี่สูงจะบอกความแตกต่างกัน เมื่อพิจารณาเสียงมนุษย์ ถ้าคุณย้ายส่วนความถี่สูงออก เสียงจะต่างออกไป แต่ยังคงรู้ว่าพูดอะไร อย่างไรก็ตามถ้าหากคุณย้ายส่วนความถี่ต่ำออกไปมากพอแล้วจะกลายเป็นเสียงพิมพ์ทำเท่านั้น ดังนั้นการวิเคราะห์เวฟเลตจึงมักพูดถึงการประมาณค่า (Approximation) และรายละเอียด (Detail) ซึ่งค่าการประมาณจะเป็นส่วนของความถี่ต่ำ และค่ารายละเอียดจะเป็นส่วนของความถี่สูง ถ้าสร้างสัญญาณการแปลงสัญญาณเวฟเลตแบบแบ่งช่วงในขั้นตอนเดียว (1 ระดับการแปลง) เมื่อสัญญาณอินพุต 1000 แซมเปิ้ล โดย  $cA$  จะเป็นค่าการประมาณ และ  $cD$  จะเป็นค่ารายละเอียด

เมื่อพิจารณาการแยกส่วนประกอบหลายๆ ระดับโดยการแยกส่วนประกอบที่สามารถกระทำซ้ำด้วยการประมาณค่าที่ต่อเนื่องตามระดับที่ต้องการ เช่น การแปลงเวฟเลตแบบแบ่งช่วงที่ 3 ระดับการแปลงดังภาพที่ 2.20



ภาพที่ 2.20 การสร้างสัญญาณการแปลงสัญญาณเวฟเลตแบบแบ่งช่วงในขั้นตอนเดียว [11]

## 2.6 งานวิจัยที่เกี่ยวข้อง

จากทฤษฎีทั้งหมดที่กล่าวมานี้ จะนำไปประยุกต์ใช้กับการศึกษาเทคนิคการบีบอัดสัญญาณเสียงพูดในงานวิจัยนี้ เพื่อลดขนาดของสัญญาณเสียงพูด ซึ่งงานวิจัยทางการประมวลผลสัญญาณเสียงพูดมีการพัฒนาอย่างหลากหลายแนวทาง และมีความแตกต่างกันออกไปตามวัตถุประสงค์ของการวิจัยนั้นๆ ดังนั้นในการวิจัยครั้งนี้จึงได้คัดเลือกงานวิจัยที่เกี่ยวข้องและมีความใกล้เคียงกับงานวิจัยที่กำลังศึกษา ดังตัวอย่างต่อไปนี้

2.6.1 Shijo M Joseph และ Babu P Anto [15] นำเสนอในหัวข้อ “Speech Compression Using Wavelet Transform” 2011 เป็นการอธิบายเทคนิคใหม่สำหรับการบีบอัดเสียงพูด โดยขบวนการของการแปลงสัญญาณเสียงพูดของมนุษย์ให้กลายเป็นรูปแบบกระชับมากขึ้นสำหรับการสื่อสารและการเก็บข้อมูลโดยที่ไม่มีการสูญเสียที่สามารถเข้าใจได้ของสัญญาณต้นฉบับ การแปลงเวฟเล็ตแบบไม่ต่อเนื่องและการสลายตัวแพ็คเก็ตเวฟเล็ตเป็นเทคนิคเมื่อไม่นานมานี้นำมาใช้ในรูปแบบเป็นร่างในการบีบอัด Malayalam เป็นภาษาหนึ่งของภาษาอินเดียใต้ได้ถูกใช้เพื่อการทดลองนี้ การบีบอัดเป็นผลสำเร็จและการสร้างใหม่ของคำพูดของภาษา Malayalam ถูกแยกออกโดยการใช้การสลายตัวแพ็คเก็ตเวฟเล็ต

2.6.2 Zhao Dan และ MA Sheng-Qian [16] นำเสนอในหัวข้อ “Speech Compression With Best Wavelet Packet Transform and SPIHT Algorithm” เป็นการเสนออัลกอริทึมการเข้ารหัสแบบใหม่ของการบีบอัดสัญญาณเสียงพูด ซึ่งอยู่บนพื้นฐานการแปลงแพ็คเก็ตเวฟเล็ตที่ดีที่สุดและอัลกอริทึมการเข้ารหัส SPIHT วิธีใหม่ได้นำมาใช้กับการแปลงแพ็คเก็ตเวฟเล็ตเพื่อเข้ากระบวนการสัญญาณเสียงพูดที่ได้รับจากทรีเวฟเล็ตที่ดีที่สุด การหาตำแหน่งบิต Dynamic และหลังจากนั้นใช้ปรับปรุงอัลกอริทึมการเข้ารหัส SPIHT เพื่อบีบอัดค่าสัมประสิทธิ์จากการแปลงแพ็คเก็ตเวฟเล็ต การเข้ารหัสเอ็นโทรปีถูกใช้ร่วมกับการเพิ่มอัตราการบีบอัดเสียงพูด จากการทดลองนี้ทำให้รู้ว่าวิธีการใหม่สามารถเพิ่มคุณภาพการสร้างสัญญาณที่อัตราการบีบอัดสูงได้ดีกว่า มีความซับซ้อนน้อย และมี delay น้อย

2.6.3 Shivesh Ranjan [17] นำเสนอในหัวข้อ “A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition” เป็นการแปลงเวฟเล็ตเพื่อช่วยลดการประมวลผลในขั้นตอนการรู้จำเสียงพูดภาษาฮินดี โดยใช้การบีบอัดสัญญาณเสียงโดยใช้เวฟเล็ตชนิดต่างๆ เช่น Daubechies, Coiflet และ Mayer Wavelet เป็นต้น เพื่อลดขนาดข้อมูลเสียงก่อนนำเข้าสู่กระบวนการสกัดค่าคุณลักษณะ โดยใช้ Linear Predictive Coding (LPC) ร่วมกับการอัลกอริทึม K-Mean เพื่อใช้ในการรู้จำเสียงพูด โดยใช้หมวดค่าในการทดสอบคือ ตัวเลขศูนย์ ถึง เก้า ในภาษาฮินดี (Hindi Language) ซึ่งใช้วัดประสิทธิภาพของเวฟเล็ตชนิดต่างๆ จากค่าเฉลี่ยของเปอร์เซ็นต์ความสำเร็จของการรู้จำเสียงพูดโดย

แยกเป็นแต่ละคำและแยกตามชนิดของเวฟเล็ต ในการหาค่าประสิทธิภาพในงานวิจัยนี้จะพิจารณาจากเปอร์เซ็นต์ความสำเร็จที่เกิดจากกระบวนการรู้จำ ซึ่งเลือกใช้กลุ่มคำ “One” ถึง “Ten” ในภาษาฮินดี

2.6.4 Yan Long Liu Gang Guo Jun [18] นำเสนอในหัวข้อ “Selection of The Best Wavelet Base for Speech Signal” เป็นการศึกษาเกี่ยวกับการเลือกเวฟเล็ตที่เหมาะสมในการประมวลผลสัญญาณเสียงพูด ซึ่งงานวิจัยนี้ได้ทำการเลือกเปรียบเทียบประสิทธิภาพของเวฟเล็ต 6 ชนิด ได้แก่ Haar, Daubechies, Discrete Meyer, Biorthogonal, Coiflets, Symlets ซึ่งใช้การหาประสิทธิภาพจากค่าผิดพลาดเฉลี่ย และค่าผิดพลาดสูงสุดในการเปรียบเทียบ ซึ่งจากผลการวิจัยนี้พบว่า ค่าผิดพลาดที่ได้จากเวฟเล็ตชนิด Dmey มีค่าประสิทธิภาพที่ดีที่สุด เนื่องจากมีค่าผิดพลาดที่น้อยที่สุด ซึ่งเหมาะสมสำหรับการวิจัยด้านการวิเคราะห์สัญญาณเสียงพูด นอกจากนี้จากผลการวิจัยยังพบว่าเวฟเล็ตชนิด Haar ไม่เหมาะสำหรับการวิเคราะห์สัญญาณเสียงพูด เนื่องจากมีที่ผิดพลาดสูง ในแต่ละระดับของเวฟเล็ตชนิด Haar

2.6.5 Noman Junejo และคณะ [19] นำเสนอในหัวข้อ “Speech and Image Compression Using Discrete Wavelet Transform” เป็นการศึกษาหาประสิทธิภาพบีบอัดสัญญาณเสียงพูดและสัญญาณภาพ โดยใช้ค่า SNR, ค่า PSNR, ค่า Retained Energy: RSE และค่าอัตราการบีบอัดสัญญาณ (Compression Ratio) ซึ่งจากผลงานวิจัยพบว่าเวฟเล็ตแบบไม่ต่อเนื่องเหมาะสำหรับการประมวลผลสัญญาณภาพและเสียงที่เป็นแบบ Non-Stationary โดยเวฟเล็ตชนิด Haar มีประสิทธิภาพการบีบอัดสัญญาณเสียงที่ดีที่สุด เนื่องจากให้ค่า SNR ที่มีค่ามากที่สุดเมื่อเปรียบเทียบกับเวฟเล็ตชนิดอื่นๆ คือ ค่า SNR = 2.31 ในการประมวลผลสัญญาณเสียงพูด และ SNR = 2.67 ในการประมวลผลสัญญาณภาพ

2.6.6 Abdul Mawla M A Najih และคณะ [20] นำเสนอในหัวข้อ “Speech Compression using Discrete Wavelet Transform” เป็นวิธีเกี่ยวกับการหาประสิทธิภาพการบีบอัดสัญญาณเสียงพูดโดยใช้เวฟเล็ตแบบไม่ต่อเนื่อง โดยทำการหาค่าประสิทธิภาพจาก SNR: Signal-to-Noise Ratio, PSNR: Peak Signal-to-Noise Ratio และ NRMSE: Normalized Root Mean Square Error โดยทำการทดสอบกับเสียงพูดขนาด 8 บิต ที่ความถี่ 8 kHz จากผลการวิจัยพบว่าเวฟเล็ตชนิด Daubechies 10 (dB10) ให้ค่า SNR และ PSNR ที่สูงที่สุดทั้งเสียงพูดของเพศชายและเพศหญิง นอกจากนี้ยังให้อัตราการบีบอัดของเพศชายสูงถึง 4.31

2.6.7 Jakkree Srinonchat [21] นำเสนอในหัวข้อ “Double Clustering Algorithm applied to Speaker Dependent Information” ได้ศึกษาเกี่ยวกับโครงสร้างการบีบอัดข้อมูลให้มีค่า Bit Rate ต่ำลงด้วยการใช้ LPC-10 แบบใหม่ โดยพบว่าค่าพารามิเตอร์ LPC ที่ใช้ทั่วไปจะให้ค่าการควอนไทซ์ไม่ค่อยมีประสิทธิภาพ เนื่องจากจะเกิดค่าผิดพลาดขึ้นในกรณีที่ระดับความแตกต่างของสัญญาณเสียงพูดมี

ระดับต่ำ เพื่อแก้ปัญหาเรื่องความผิดพลาดของการควอนไทซ์ จึงได้นำเสนอวิธีการโดยเปลี่ยนค่าพารามิเตอร์ LPC เป็นพารามิเตอร์ LSP เพื่อจะสร้าง Codevector ของการแบ่งลำดับชั้นใน Vector Quantization เพื่อที่จะใช้เข้ารหัสสัญญาณเสียงพูดใหม่ เรียกว่า LPC-10 จากผลทดลองพบว่าสามารถลดจำนวนบิตของสัญญาณเสียงพูดในบิตสัญญาณเสียงพูด p1-p4 ได้ 4 บิต และอัตราการส่งข้อมูลลดลง 2-66%

2.6.8 Somchai. Jitapunkul และคณะ [22] นำเสนอในหัวข้อ “Novel Technique for Tonal Language Speech Compression Based on Bitrate Scalable MP-CELP Coder” ได้นำเสนอการเข้ารหัสเสียงพูดมาตรฐาน MPEG-4 ด้วยการจำลองตัวเข้ารหัสและตัวถอดรหัส MP-CELP โดยใช้เสียงพูดภาษาอังกฤษและเสียงพูดภาษาไทย ผลการทดลองแสดงให้เห็นว่าเสียงพูดภาษาไทยที่ผ่านการเข้ารหัสและถอดรหัสที่จำลองขึ้น มีคุณภาพด้อยกว่าเสียงพูดภาษาอังกฤษ ซึ่งใช้ค่าอัตราส่วนกำลังของสัญญาณต่อกำลังของสัญญาณรบกวนในการวัดผล

2.6.9 วีระยุทธ คุณรัตนสิริ และ จักริ ศรีนนท์ฉัตร [23] นำเสนอในหัวข้อ “การเปรียบเทียบประสิทธิภาพการบีบอัดสัญญาณเสียงพูดโดยใช้เทคนิคเวฟเล็ต” ได้นำเสนอเกี่ยวกับการเปรียบเทียบเวฟเล็ต 3 ตระกูล คือ Daubechies Symlet และ Coiflets สำหรับการบีบอัดสัญญาณ 5 ระดับ โดยวิเคราะห์ทั้งโดเมนเวลาและโดเมนความถี่ จากการทดลองสรุปได้ว่า Daubechies3 ในระดับที่ 2 มีค่า SNR และค่า PSNR ดีที่สุดคือ 51.2578

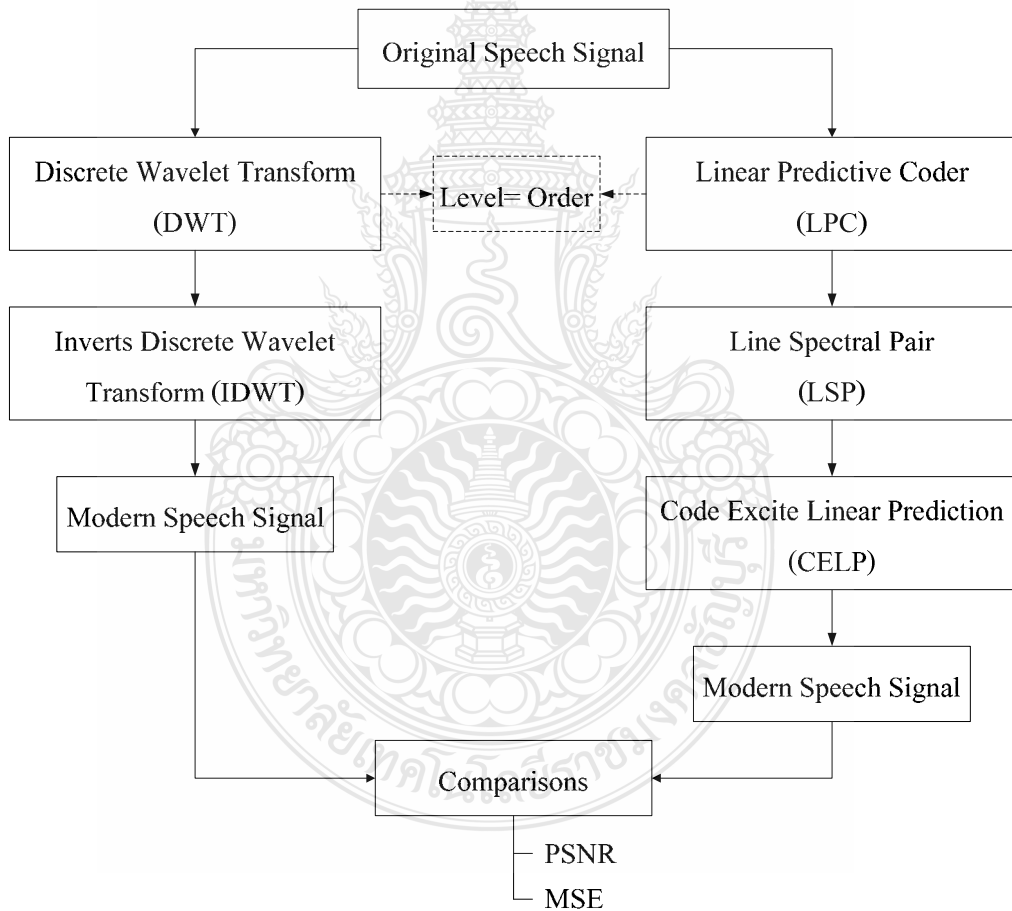




### บทที่ 3

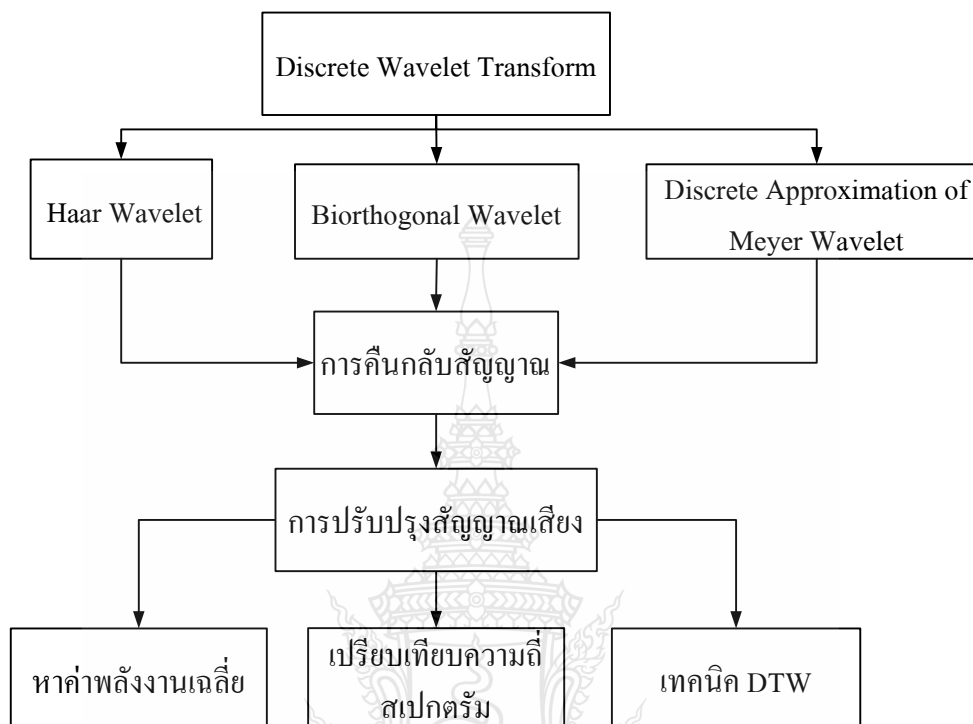
#### วิธีการดำเนินงานวิจัย

บทนี้นำเสนอเทคนิคการบีบอัดสัญญาณเสียงพูด 2 เทคนิค ซึ่งได้ดำเนินการในงานวิจัยนี้ โดยขั้นตอนต่างๆ จะประกอบด้วย การเตรียมสัญญาณเสียงพูดก่อนการบีบอัดสัญญาณ กระบวนการบีบอัดเสียง กระบวนการคืนกลับสัญญาณจากการบีบอัดสัญญาณเสียงพูด และการเปรียบเทียบคุณภาพของสัญญาณเสียงใหม่ที่ใกล้เคียงกับสัญญาณต้นฉบับโดยเปรียบเทียบกันทั้ง 2 เทคนิค ซึ่งขั้นตอนต่างๆ จะแสดงในภาพที่ 3.1



ภาพที่ 3.1 ขั้นตอนการทำงานของงานวิจัย

### 3.1 การเลือกตระกูลเวฟเล็ต



ภาพที่ 3.2 ขั้นตอนการคัดเลือกตระกูลเวฟเล็ต

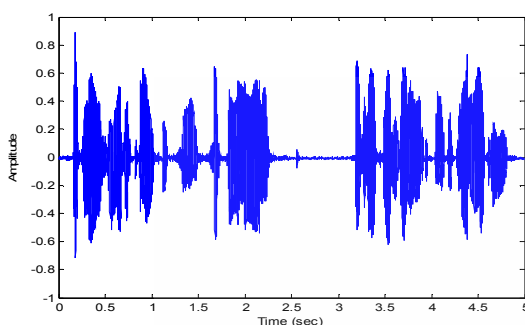
จากภาพที่ 3.2 เป็นกระบวนการการคัดเลือกเวฟเล็ต เพื่อนำไปใช้ในการบีบอัดสัญญาณเสียง เพื่อเลือกเวฟเล็ตที่เหมาะสมที่สุดสำหรับงานวิจัยนี้ โดยจะนำเวฟเล็ตที่ผ่านการคัดเลือกจะถูกบีบอัดสัญญาณเสียงและคืนกลับสัญญาณตามกระบวนการในภาพที่ 3.1 เพื่อเปรียบเทียบกับเทคนิค LPC ซึ่งขั้นตอนในการดำเนินการมีดังนี้

#### 3.1.1 เตรียมสัญญาณเสียงพูด

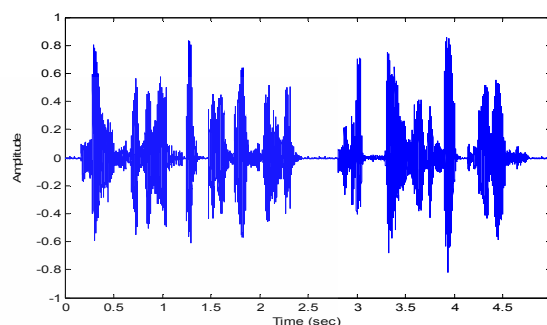
สัญญาณเสียงพูดที่ใช้สำหรับงานวิจัยนี้มีทั้งหมด 80 เสียง ซึ่งเป็นสัญญาณเสียงภาษาอังกฤษและแต่ละสัญญาณเสียงเป็นประโยคคำพูดที่แตกต่างกัน โดยสามารถแบ่งกลุ่มตามเวลาและแยกเป็นสัญญาณเสียงของผู้ชายและผู้หญิงได้ดังนี้

- 1) เสียงผู้ชายที่เวลา 5 วินาที และ 60 วินาที อย่างละ 20 เสียง
- 2) เสียงผู้หญิงที่เวลา 5 วินาที และ 60 วินาที อย่างละ 20 เสียง

3) สัญญาณเสียงที่นำมาทดลอง จะมีอัตราการสุ่มตัวอย่าง (Sampling Rate) อยู่ที่ 8000 Hz โดยแต่ละข้อมูลมีขนาด 8 บิต ใช้ช่องสัญญาณเดียว (Mono) โดยรูปแบบจะเป็นมาตรฐานของระบบพีซีเอ็ม (Pulse-Code Modulation: PCM)

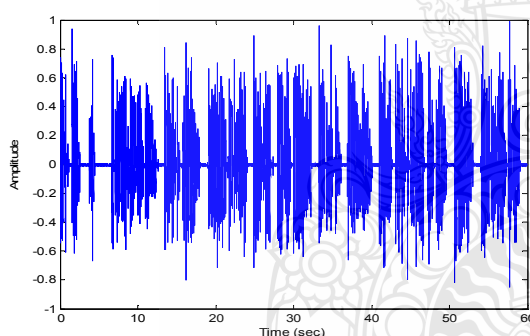


(ก) เสียงพูดของผู้หญิง

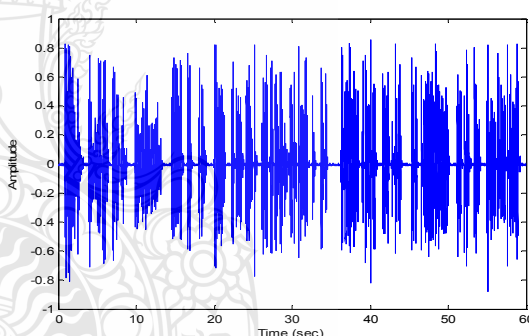


(ข) เสียงพูดของผู้ชาย

ภาพที่ 3.3 ตัวอย่างสัญญาณเสียงพูด ที่เวลา 5 วินาที



(ก) เสียงพูดของผู้หญิง

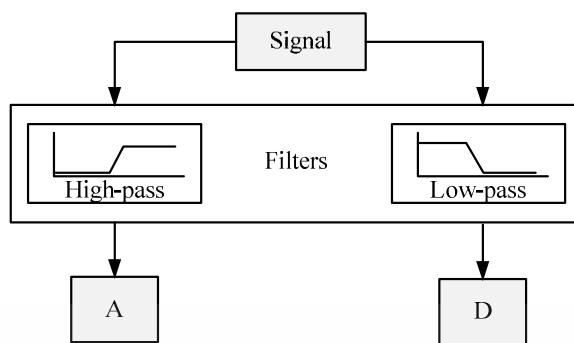


(ข) เสียงพูดของผู้ชาย

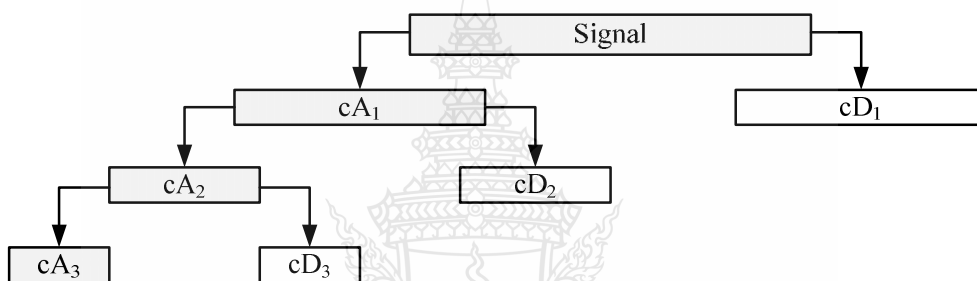
ภาพที่ 3.4 ตัวอย่างสัญญาณเสียงพูด ที่เวลา 60 วินาที

### 3.1.2 การบีบอัดสัญญาณเสียงพูดเบื้องต้นเพื่อคัดเลือกเวฟเล็ตที่เหมาะสม

งานวิจัยนี้ใช้การบีบอัดสัญญาณเสียงด้วยวิธี DWT ระดับที่ 2 ซึ่งเวฟเล็ตที่นำมาทำการวิจัยทั้งหมด 3 ตระกูล คือ เวฟเล็ต Haar เวฟเล็ต Biorthogonal และ เวฟเล็ต Discrete Approximation of Meyer ซึ่งในกระบวนการบีบอัดสัญญาณเสียงโดยใช้เวฟเล็ตนั้น จะทำการแยกสัญญาณออกเป็นค่าสัมประสิทธิ์ 2 ค่า คือ ค่าสัมประสิทธิ์การประมาณ (Approximate Coefficient) และค่าสัมประสิทธิ์รายละเอียด (Detail Coefficient) โดยนำข้อมูลเหล่านี้ไปประมวลผลในกระบวนการอื่นต่อไปดังภาพที่ 3.5 และ 3.6



ภาพที่ 3.5 การแยกองค์ประกอบของความถี่ต่ำและความถี่สูงของการแปลงเวฟเล็ด [8]



ภาพที่ 3.6 ขั้นตอนและองค์ประกอบของการแปลงวิธีเวฟเล็ด [8]

### 3.1.3 การปรับปรุงสัญญาณเสียงพูดใหม่

นำสัญญาณเสียงพูดใหม่ที่ได้มาปรับปรุงอัตราการสุ่มตัวอย่าง (Sample Rate) เพื่อให้ได้คุณภาพของสัญญาณเสียงที่ดีขึ้นดังสมการที่ 3.1 [24]

$$x_{new} = \frac{p}{q} \times sampling\ rate_{new} \times t_{sec} \quad (3.1)$$

โดยที่  $x_{new}$  คือ สัญญาณเสียงพูดใหม่

$t_{sec}$  คือ เวลาในหน่วยของวินาที (ในการทดลองนี้คือ 5 วินาที และ 60 วินาที)

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

ในงานวิจัยนี้จะใช้อัตราส่วน  $\left(\frac{p}{q}\right)$  ทั้งหมด 5 อัตราส่วน คือ 3/2, 4/2, 5/2, 3 และ 4 ซึ่งค่า  $\frac{p}{q}$  ทั้งหมดนี้ได้ผ่านการทดลองเพื่อหาค่าที่เหมาะสมแล้วโดยอาศัยเวลาเป็นเกณฑ์อ้างอิง ซึ่งค่าที่ได้เมื่อนำไปทดสอบแล้วจะทำให้ได้ค่าสัญญาณเสียงพูดที่มีเวลาเท่าเดิมดังนั้นจากการคำนวณจึงทำให้ได้ค่า Sample Rate และค่าอัตราการบีบอัดสัญญาณเสียงพูดดังตารางที่ 3.1

ตารางที่ 3.1 อัตราการสุ่มตัวอย่างและอัตราการบีบอัดสัญญาณเสียงพูดของอัตราส่วน  $\frac{p}{q}$  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

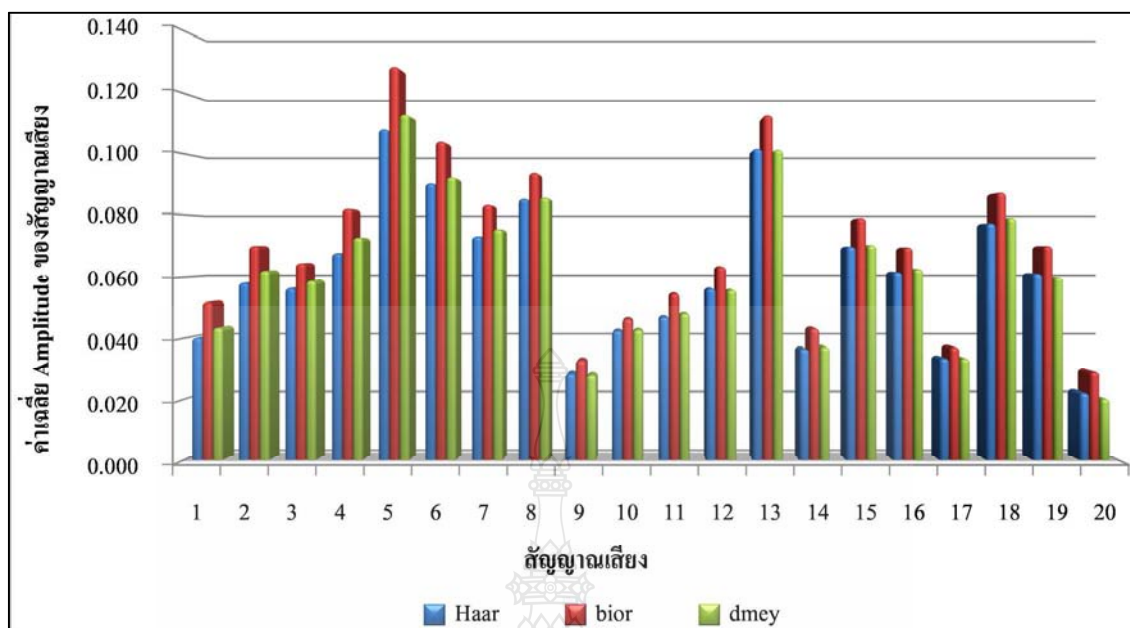
จากตารางที่ 3.1 จะสังเกตเห็นว่าที่ค่าอัตราส่วน  $p/q$  ที่ 3/2 หรือ 1.5 จะสามารถลดค่า Sampling Rate เท่ากับ 3000 Hz จาก 8000 Hz หรือลดลง 2.67 เท่าของสัญญาณเสียงพูดต้นฉบับ [24]

#### 3.1.4 เปรียบเทียบคุณภาพของการบีบอัดสัญญาณเสียงของเวฟเล็ดแต่ละตระกูล

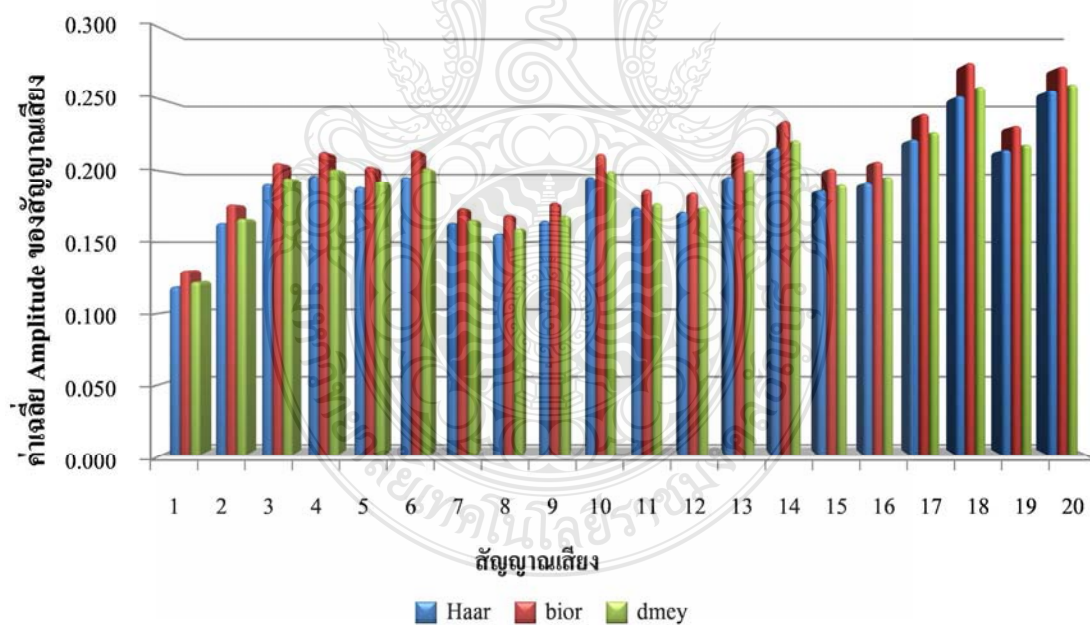
ในการบีบอัดสัญญาณเสียงครั้งนี้ เพื่อหาเวฟเล็ดที่ดีที่สุดจากทั้ง 3 ตระกูล โดยวิธีทดสอบและตัดสินใจดังนี้

1) การหาค่าพลังงานเฉลี่ยของสัญญาณเสียง โดยนำสัญญาณเสียงที่ผ่านการปรับปรุงสัญญาณเสียงทั้ง 80 เสียง คำนวณหาค่าพลังงานเฉลี่ย เพื่อเปรียบเทียบคุณภาพของสัญญาณเสียงพูดหลังจากการบีบอัดของเวฟเล็ดทั้ง 3 ตระกูล โดยวิธีการหาค่าพลังงานเฉลี่ยตามสมการที่ 3.2 ซึ่งถ้าเวฟเล็ดตระกูลใดมีค่าพลังงานเฉลี่ยที่มากที่สุดจะเป็นเวฟเล็ดที่ถูกเลือกมาใช้งานเนื่องจากให้ค่าความผิดพลาดน้อยที่สุด และแสดงกราฟค่าพลังงานเฉลี่ยในภาพที่ 3.7 – 3.10

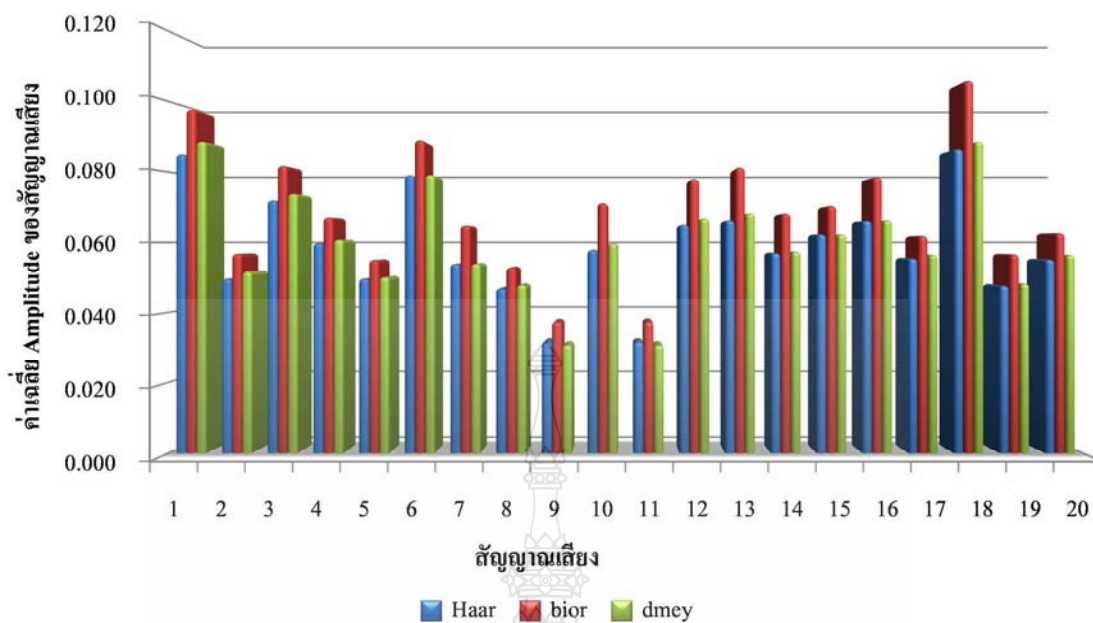
$$x = \frac{1}{n} \sum_{i=1}^n |S_i| \quad (3.2)$$



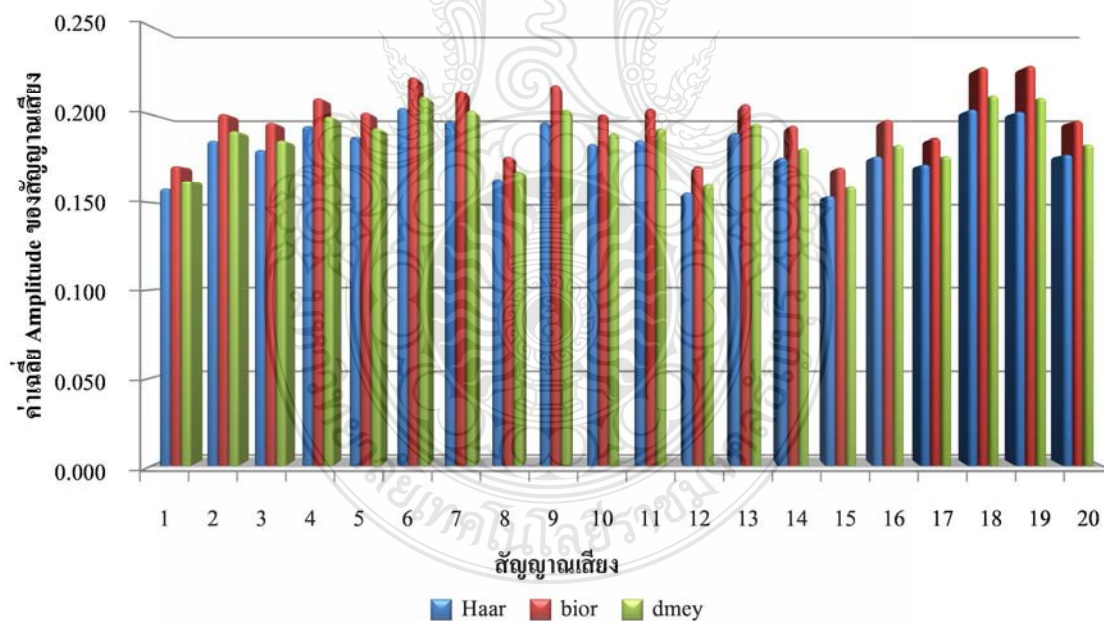
ภาพที่ 3.7 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 5 วินาที



ภาพที่ 3.8 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 60 วินาที



ภาพที่ 3.9 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 5 วินาที

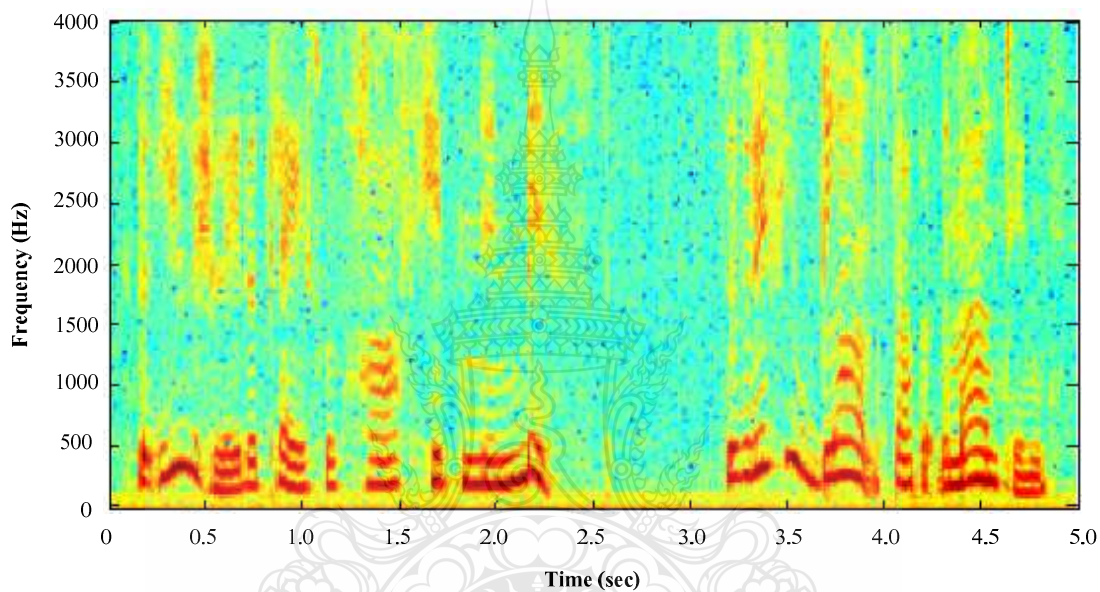


ภาพที่ 3.10 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 60 วินาที

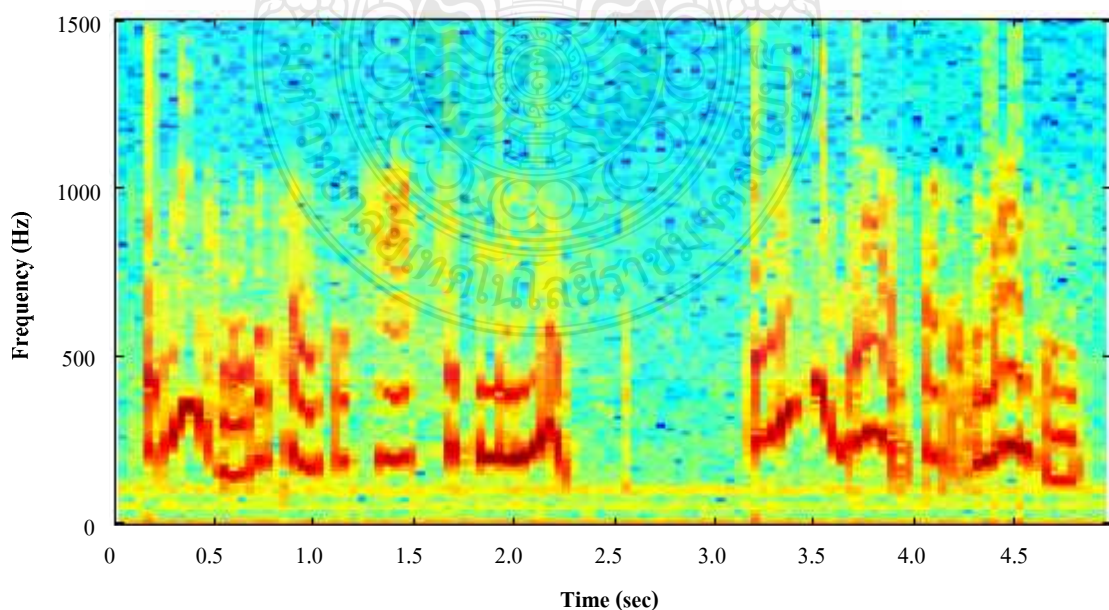
2) สืบเนื่องจากความถี่สเปกตรัมโดยใช้สเปกโตรแกรม (Spectrogram) ซึ่งโดยทั่วไปจะแบ่งเป็น 2 แบบคือ

2.1) สเปกโตรแกรมแถบกว้าง (Wideband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดชั่วขณะในเชิงเวลาได้ดี แต่มีความละเอียดทางแกนความถี่ต่ำ ซึ่งส่วนมากจะถูกนำไปวิเคราะห์การเปลี่ยนแปลงของฟอร์แมนต์และแสดงคาบพิตซ์

2.2) สเปกโตรแกรมแถบแคบ (Narrowband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดทางด้านฮาร์โมนิกได้ดี แต่แสดงรายละเอียดชั่วขณะในเชิงเวลาไม่ชัดเจน ส่วนมากจะถูกนำไปประยุกต์ใช้ในการแยกช่วงที่เป็นเสียงก้องออกจากเสียงไม่ก้อง

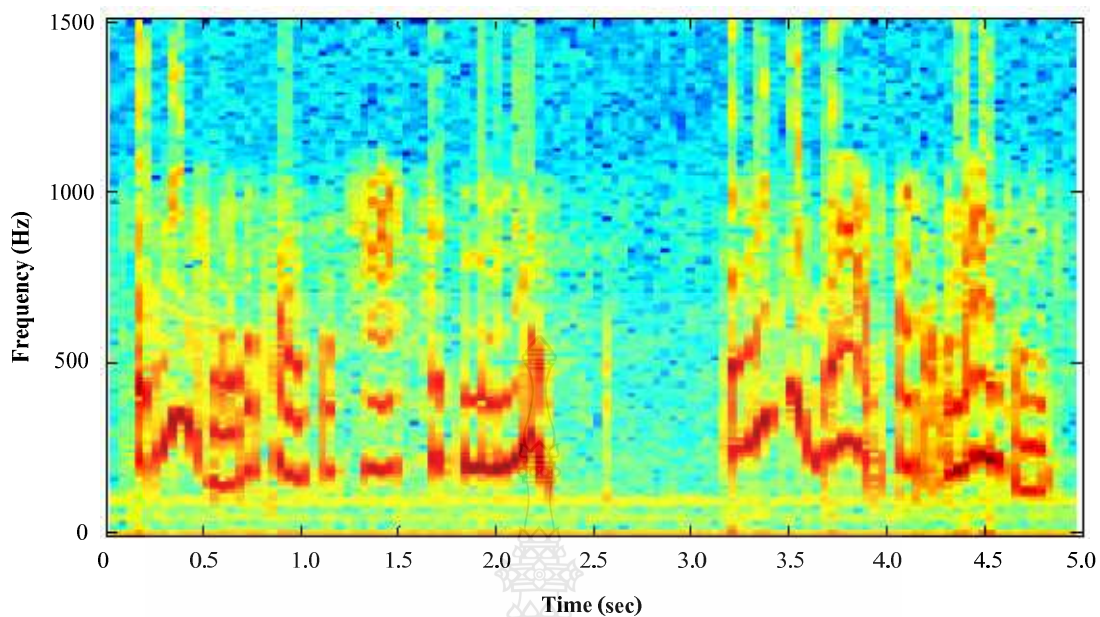


ก. Original

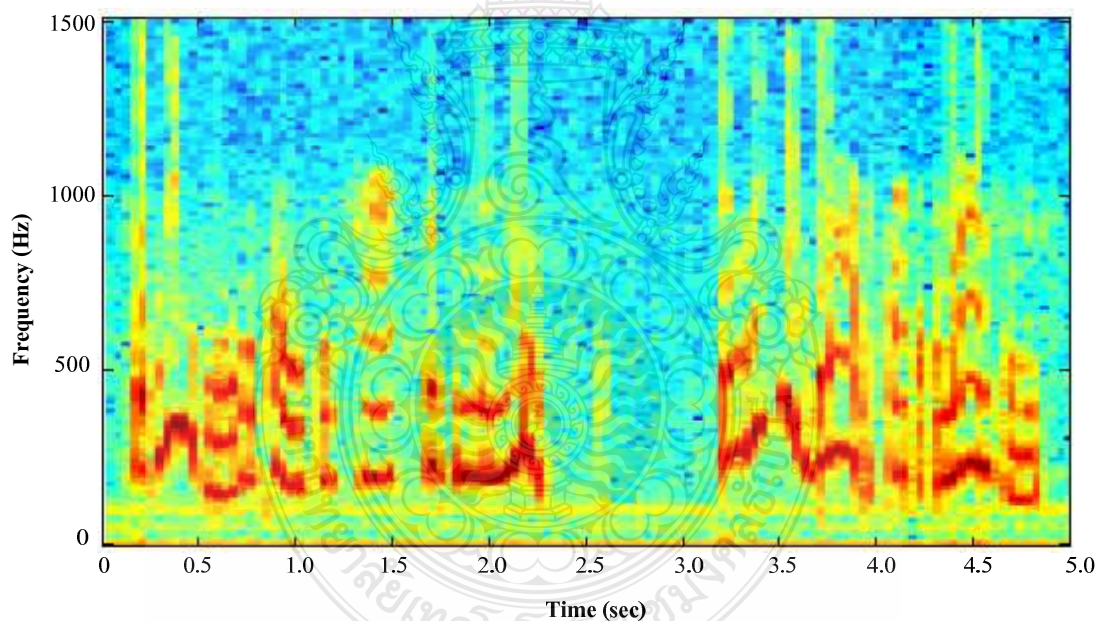


ข. เวฟเล็ต Haar





ค. เวก์เล็ท Biorthogonal



ง. เวก์เล็ท Discrete Approximation of Meyer

ภาพที่ 3.11 ตัวอย่างการเปรียบเทียบความถี่สเปกตรัมของสัญญาณเสียง

จากภาพที่ 3.11 เป็นตัวอย่างของการเปรียบเทียบคุณภาพของสัญญาณเสียงพูดหลังจากการบีบอัดด้วยของเวก์เล็ททั้ง 3 ตระกูล โดยใช้เทคนิคสเปกโตรแกรมมาช่วยในการเปรียบเทียบค่าความถี่

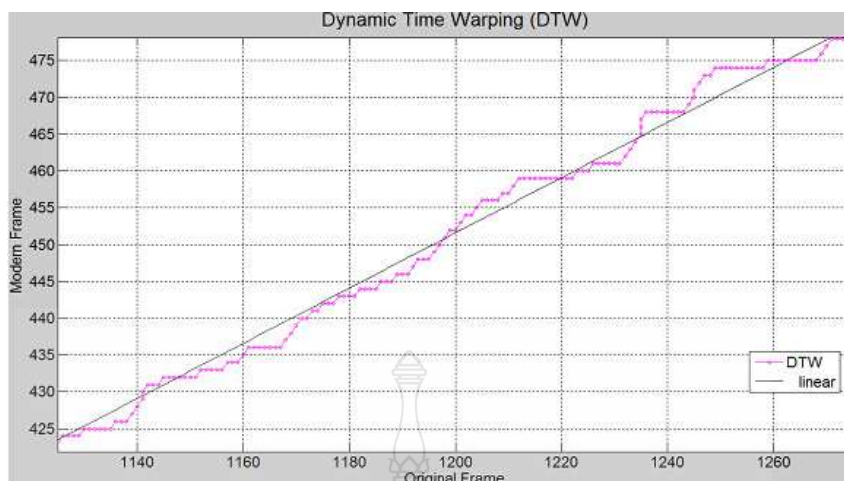
ของสัญญาณเสียงหลังจากการปรับปรุงสัญญาณเสียงพูดแล้ว โดยวิเคราะห์จากค่าความเข้มของความถี่ ถ้าเวฟเล็ดตระกูลใดมีลักษณะของกราฟความถี่ที่มีค่าผิบน้อยที่สุดหรือมีกราฟความถี่ที่คล้ายกับ ต้นฉบับมากที่สุดจะถือว่าเวฟเล็ดตระกูลนั้นดีที่สุดสำหรับบทความนี้ โดยเปรียบเทียบกับ สัญญาณเสียงต้นฉบับ

3) ใช้ Dynamic Time Warping (DTW) เป็นขั้นตอนวิธีสำหรับการเปรียบเทียบความ คล้ายกันระหว่างข้อมูล 2 ชุด โดยผลลัพธ์ที่ได้จะให้ค่าระยะทางและวิธีการปรับแนว (Alignment) ที่ดี ที่สุดระหว่างข้อมูลทั้งสอง ซึ่งสามารถยืดหรือหดให้รองรับความแปรผันในแกนเวลาได้ ซึ่ง DTW สามารถนำไปประยุกต์ได้กับวิดีโอ เสียง และภาพ รวมไปถึงการประมวลผลสัญญาณต่างๆ ที่สามารถ แปลงให้อยู่ในรูปของข้อมูลเชิงเส้นได้ ดังนั้นงานวิจัยนี้จึงนำเทคนิคของ DTW มาประยุกต์ใช้ เพื่อ จัดการกับคำพูดที่มีความเร็วและการสุมข้อมูลที่ไม่เท่ากันแม้จะสื่อความหมายเดียวกัน หรือเพื่อให้ จังหวะทางเวลาของสัญญาณตัวอย่างมีค่าตรงกับสัญญาณต้นแบบ เพื่อลดข้อผิดพลาดที่จะเกิดขึ้นใน การเปรียบเทียบสเปกตรัมของสัญญาณ โดยกำหนดให้ลำดับของเวกเตอร์ 2 เวกเตอร์ คือ เวกเตอร์  $A$  และเวกเตอร์  $B$  ดังสมการที่ 3.3 และ 3.4 การเปรียบเทียบระยะห่างระหว่างเวกเตอร์  $A$  และเวกเตอร์  $B$  แสดงในสมการที่ 3.5 เมื่อ  $d(i, j)$  เป็นระยะห่างระหว่างคู่จุดที่จุด  $i$  ของเฟรม  $A$  และ  $j$  ของ เฟรม  $B$  โดยที่  $a_{in}$  เป็นค่าลำดับที่  $i$  ของเวกเตอร์  $A$  และ  $b_{jn}$  เป็นค่าลำดับที่  $j$  ของเวกเตอร์  $B$  และ  $K$  เป็นจำนวนของลักษณะของเวกเตอร์ ซึ่งค่าที่ได้จะเป็นค่าระยะทางที่มีค่าน้อยที่สุด

$$A = a_1, a_2, \dots, a_I \quad (3.3)$$

$$B = b_1, b_2, \dots, b_J \quad (3.4)$$

$$d(i, j) = \sum_{n=1}^K (a_{in} - b_{jn})^2 \quad (3.5)$$



ภาพที่ 3.12 ตัวอย่างการเปรียบเทียบความคล้ายกันระหว่างสัญญาณเสียงต้นฉบับและเสียงใหม่

จากภาพที่ 3.12 จะเห็นได้ว่าเส้น DTW จะมีลักษณะใกล้เคียงกับเส้นตรงที่เป็นเส้นอ้างอิงมาก ซึ่งเทคนิคถ้าเส้น DTW ใกล้เคียงเส้นตรงมากเท่าไร แสดงให้เห็นว่าสัญญาณเสียงมีความใกล้เคียงกับต้นฉบับมากเท่านั้น ดังนั้นการเปรียบเทียบลักษณะเด่นของสัญญาณเสียงจึงจะใช้ตัวตรวจจับแบบสหพันธ์ (Correlation Detector) โดยการคำนวณหาค่าสัมประสิทธิ์สหพันธ์ (Correlation Coefficient) ระหว่างเส้น DTW และ เส้นอ้างอิง ตามสมการที่ 3.6

$$r = \frac{N \sum XY - ((\sum X)(\sum Y))}{\sqrt{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)}} \quad (3.6)$$

เมื่อ  $X$  คือ สัญญาณเสียงต้นฉบับ

$Y$  คือ สัญญาณเสียงที่ถูกคืนกลับ

จากการสมการที่ 3.6 เป็นการคำนวณหาค่าสัมประสิทธิ์สหพันธ์ ( $r$ ) เพื่อเปรียบเทียบประสิทธิภาพของเวฟเล็ททั้ง 3 ตระกูล ซึ่งผลลัพธ์ของสัญญาณเสียงพูดแต่ละกลุ่มได้แสดงดังตารางที่ 3.2-3.5

ตารางที่ 3.2 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 5 วินาที

อัตราการสุ่มตัวอย่าง ของสัญญาณเสียงพูด	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 3.3 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 5 วินาที

อัตราการสุ่มตัวอย่าง ของสัญญาณเสียงพูด	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 3.4 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้หญิงที่เวลา 60 วินาที

อัตราการสุ่มตัวอย่าง ของสัญญาณเสียงพูด	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 3.5 การเปรียบเทียบค่าสัมประสิทธิ์สหพันธ์ของสัญญาณเสียงของผู้ชายที่เวลา 60 วินาที

อัตราการสุ่มตัวอย่าง ของสัญญาณเสียงพูด	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 3.2 – 3.5 เป็นการหาค่าสัมประสิทธิ์สหพันธ์ ( $r$ ) โดยผลการทดลองแสดงให้เห็นว่าเวฟเล็ตทั้ง 3 ตระกูล มีค่าสัมประสิทธิ์สหพันธ์ที่ใกล้เคียงกันมาก โดยเฉพาะ Haar เวฟเล็ต ให้ประสิทธิภาพการบีบอัดสัญญาณเสียงพูดและคืนกลับสัญญาณได้ดีกว่า Biorthogonal เวฟเล็ต และ Discrete Approximation of Meyer เวฟเล็ต เนื่องจากมีค่าสัมประสิทธิ์สหพันธ์ ที่มีค่าเข้าใกล้ 1 มากที่สุด

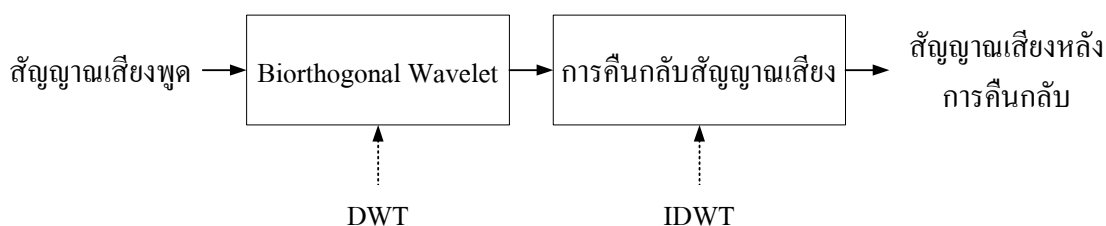
ดังนั้นจากการเปรียบเทียบเพื่อหาตระกูลเวฟเล็ตทั้ง 3 วิธีนี้ สามารถสรุปได้ว่า Biorthogonal เวฟเล็ต ดีกว่าตระกูลเวฟเล็ตชนิดอื่น ดังนั้นในงานวิจัยนี้จึงเลือกใช้ Biorthogonal เวฟเล็ตเพียงตระกูลเดียวเพื่อวิจัยตามขั้นตอนตามภาพที่ 3.1 [25]

### 3.2 ขั้นตอนการบีบอัดสัญญาณเสียงพูด

จากการคัดเลือกตระกูลเวฟเล็ตทั้ง 3 ตระกูล เพื่อหาเวฟเล็ตที่เหมาะสมสำหรับงานวิจัยนี้ได้ข้อสรุปเป็นตระกูลเวฟเล็ต Biorthogonal เนื่องจากเวฟเล็ตตระกูลนี้มีฟังก์ชันพื้นฐานที่เหมาะสมกับการวิเคราะห์สัญญาณที่มีลักษณะไม่คงที่ ดังเช่นสัญญาณเสียงพูดนั่นเอง ในขั้นตอนต่อไปนี้จะกล่าวถึงการบีบอัดด้วยเวฟเล็ตเปรียบเทียบกับวิธีการบีบอัดด้วย CELP โดยการบีบอัดสัญญาณเสียงพูดนั้นจะใช้ Order สำหรับการบีบอัดที่เท่ากัน คือ Order 10, 20 และ 40

#### 3.2.1 การบีบอัดสัญญาณเสียงพูดโดยใช้เวฟเล็ตตระกูล Biorthogonal

ในขั้นตอนนี้จะนำสัญญาณเสียงที่ได้เตรียมไว้เข้าสู่กระบวนการบีบอัดสัญญาณเสียงพูดด้วยเวฟเล็ตดังภาพที่ 3.13

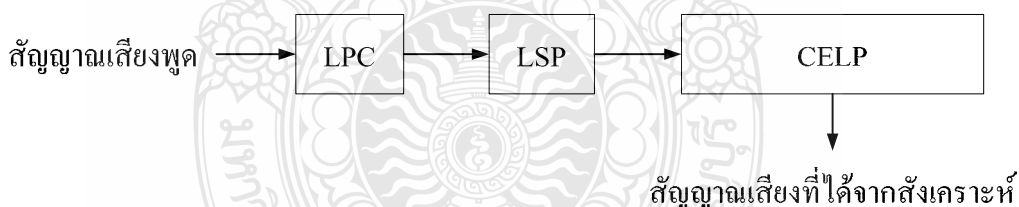


ภาพที่ 3.13 กระบวนการบีบอัดสัญญาณเสียงพูดของเวฟเล็ต

จากภาพที่ 3.13 สัญญาณเสียงทั้งหมดจะถูกบีบอัดสัญญาณเสียงพูดด้วยเวฟเล็ต Biorthogonal โดยใช้เทคนิคของการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (DWT) และทำการคืนกลับสัญญาณเสียงโดยใช้เทคนิคการแปลงกลับเวฟเล็ตแบบไม่ต่อเนื่อง (Inverse Discrete Wavelet Transform; IDWT)

### 3.2.2 การสกัดลักษณะของสัญญาณเสียงพูดโดยใช้การเข้ารหัสแบบเชิงเส้น (LPC)

ในขั้นตอนนี้จะนำสัญญาณเสียงพูดเดียวกันกับการบีบอัดสัญญาณเสียงด้วยเวฟเล็ต เพื่อเข้าสู่กระบวนการสกัดลักษณะเด่นของสัญญาณเสียงด้วยเทคนิค LPC และใช้หลักการของ CELP เพื่อบีบอัดสัญญาณและสังเคราะห์สัญญาณเสียงขึ้นมาใหม่ดังภาพที่ 3.14



ภาพที่ 3.14 กระบวนการบีบอัดสัญญาณเสียงพูดของ CELP

จากภาพที่ 3.14 เสียงทั้งหมดจะถูกนำไปสกัดลักษณะเด่นของสัญญาณเสียงด้วย LPC ที่ Oder 10, 20 และ 40 แต่เนื่องจากเทคนิคของ LPC จะมีประสิทธิภาพไม่ค่อยดี จึงส่งค่าพารามิเตอร์การควอนไทซ์เซชันของ LPC ไปแปลงเป็นค่าพารามิเตอร์ของ LSP เพื่อปรับปรุงการเข้ารหัสสัญญาณเสียงพูดให้มีประสิทธิภาพดีขึ้น หลังจากนั้นนำไปบีบอัดสัญญาณเสียงด้วย CELP และสังเคราะห์สัญญาณเสียงขึ้นมาใหม่โดยใช้ค่าพารามิเตอร์ของ LSP ในการสังเคราะห์ ซึ่ง CELP เป็นเทคโนโลยีในการบีบอัดสัญญาณเสียงที่ได้รับความนิยมและเป็นที่ยอมรับและน่าจับตามองในปัจจุบัน

### 3.3 เทคนิคการเปรียบเทียบคุณภาพของสัญญาณเสียง

หลังจากการบีบอัดสัญญาณเสียงจาก DWT และ CELP นั้น ได้ทำการคืนกลับสัญญาณเสียงด้วยเทคนิคที่แตกต่างกันคือ DWT จะคืนกลับเสียงด้วย IDWT และ CELP สังเคราะห์สัญญาณเสียงขึ้นมาใหม่โดยใช้พารามิเตอร์ของ LSP ซึ่งเสียงที่ได้จะมีความแตกต่างกัน ดังนั้นในงานวิจัยนี้จึงใช้เทคนิคค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Mean Square Error; MSE) และอัตราส่วนของสัญญาณต่อสัญญาณรบกวนสูงสุด (Peak Signal to Noise Ratio; PSNR) ในการเปรียบเทียบคุณภาพของสัญญาณเสียง เพื่อหาเทคนิคการบีบอัดสัญญาณเสียงที่เหมาะสมที่สุดสำหรับงานวิจัยนี้

#### 3.3.1 การหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

MSE เป็นการหาค่าความผิดพลาดระหว่างสัญญาณเสียงต้นฉบับกับสัญญาณเสียงที่ถูกคืนกลับว่ามีความคลาดเคลื่อนเท่าไร โดยสามารถหาค่าได้จากสมการที่ 3.8

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} (s_i - p_i)^2 \quad (3.7)$$

โดยที่  $s_i$  คือ สัญญาณเสียงต้นฉบับ

$p_i$  คือ สัญญาณเสียงที่ถูกคืนกลับ

#### 3.3.2 อัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงสุด

PSNR เป็นค่าที่นิยมใช้กันอย่างแพร่หลายในงานทางด้านการประมวลผลสัญญาณทั้งภาพและเสียง ซึ่งเป็นการพิจารณาจากค่าอัตราส่วนระหว่างค่ากำลังสูงสุดของสัญญาณ และความแตกต่างของสัญญาณเสียงที่ถูกรบกวนด้วยสัญญาณรบกวน เมื่อเทียบกับสัญญาณเสียงต้นฉบับ มีหน่วยเป็นเดซิเบล (dB) ถ้าค่าของ PSNR ที่มีค่าสูงจะบ่งบอกถึงคุณภาพการบีบอัดที่ดี โดยสามารถหาค่าได้จากสมการที่ 3.8

$$PSNR = 10 \log_{10} \frac{NX^2}{\|x-r\|^2} \quad (3.8)$$

โดยที่  $N$  คือ ความยาวของสัญญาณเสียงที่ถูกคืนกลับ

$X$  คือ ค่าสูงสุดของค่าเฉลี่ยกำลังสองของสัญญาณเสียงต้นฉบับ

$\|x-r\|^2$  คือ ความแตกต่างของระดับพลังงานระหว่างสัญญาณเสียงต้นฉบับกับสัญญาณเสียงคืนกลับ

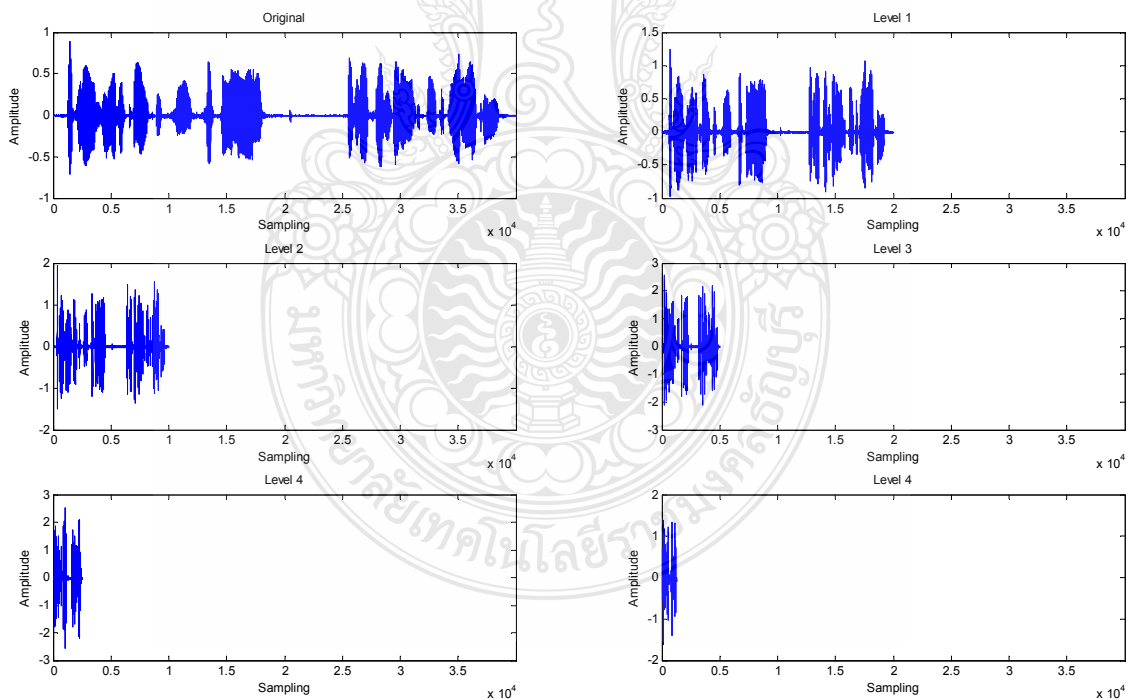
## บทที่ 4

### ผลของงานวิจัย

บทนี้เป็นนำเสนอผลการเปรียบเทียบการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค DWT (ซึ่งเลือกใช้เวฟเล็ตตระกูล Biorthogonal) และเทคนิค CELP เพื่อหาคุณภาพของสัญญาณเสียงพูดหลังจากการคืนกลับสัญญาณเสียง ซึ่งผลการดำเนินการวิจัยและการทดสอบการบีบอัดสัญญาณเสียงพูด มีดังต่อไปนี้

#### 4.1 ผลการบีบอัดสัญญาณเสียงด้วย DWT

ในงานวิจัยนี้ใช้เสียงพูดของผู้หญิงและผู้ชาย ซึ่งมีความยาว 5 วินาที และ 60 วินาที มาทดสอบบีบอัดสัญญาณเสียงด้วยเวฟเล็ต Biorthogonal ระดับที่ 1-5 โดยใช้โปรแกรม MATLAB ในการวิเคราะห์สัญญาณ ซึ่งผลของการบีบอัดสัญญาณเสียงพูดเบื้องต้นแสดงดังภาพที่ 4.1

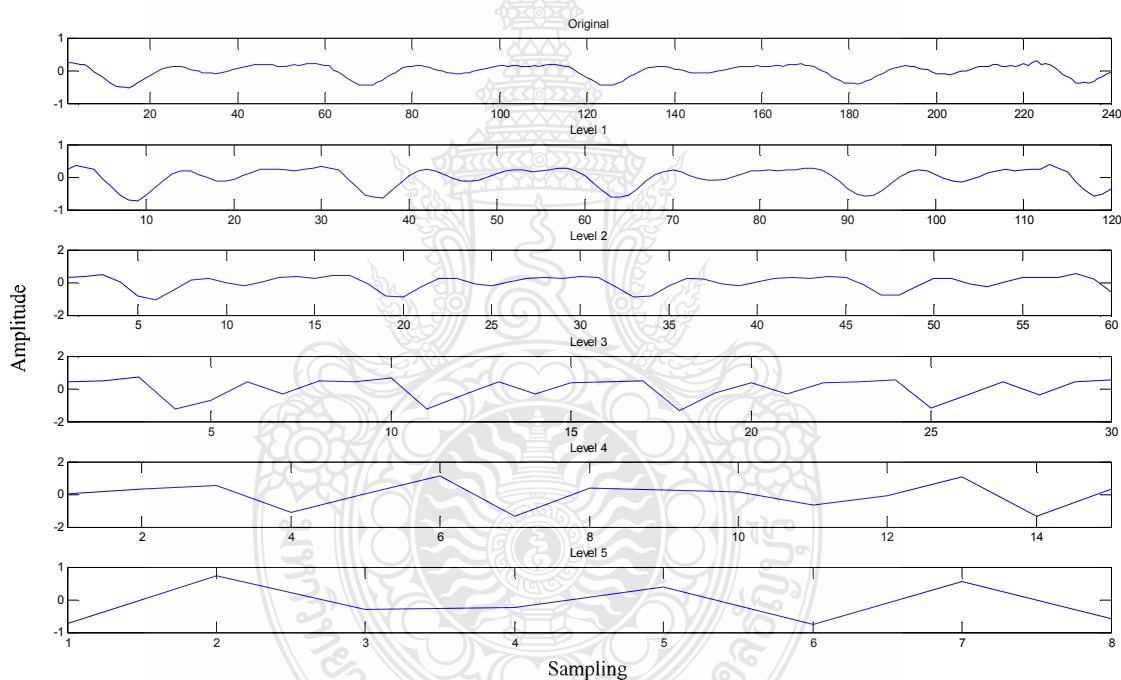


ภาพที่ 4.1 การบีบอัดสัญญาณเสียงพูดด้วย DWT

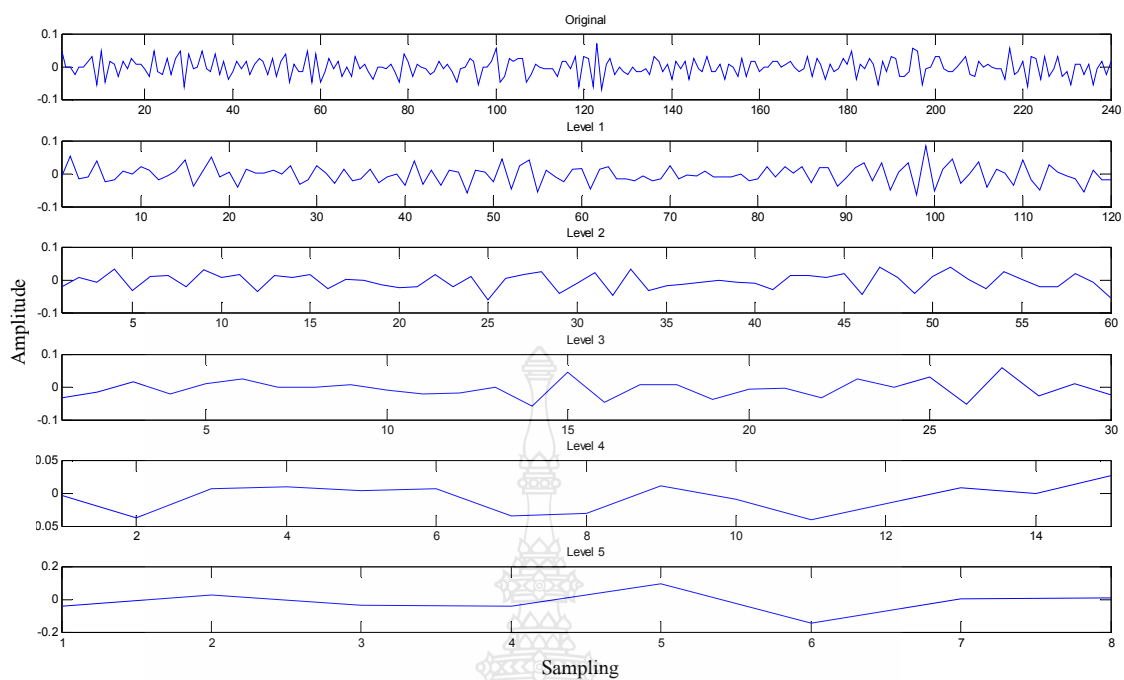


จากภาพที่ 4.1 เป็นการบีบอัดสัญญาณเสียงตั้งแต่ระดับที่ 1 ถึง ระดับที่ 5 โดยเทียบกับสัญญาณเสียงต้นฉบับ ซึ่งผลที่ได้คือ หลังจากการบีบอัดสัญญาณเสียงพูดแต่ละระดับสัญญาณเสียงยังคงมีลักษณะเหมือนเดิมกับต้นฉบับ แต่ความยาวของสัญญาณหรือจำนวนสุ่มตัวอย่างจะถูกลดลงทีละครึ่งหนึ่งจากสัญญาณต้นฉบับหรือสัญญาณที่ได้จากระดับก่อนหน้า

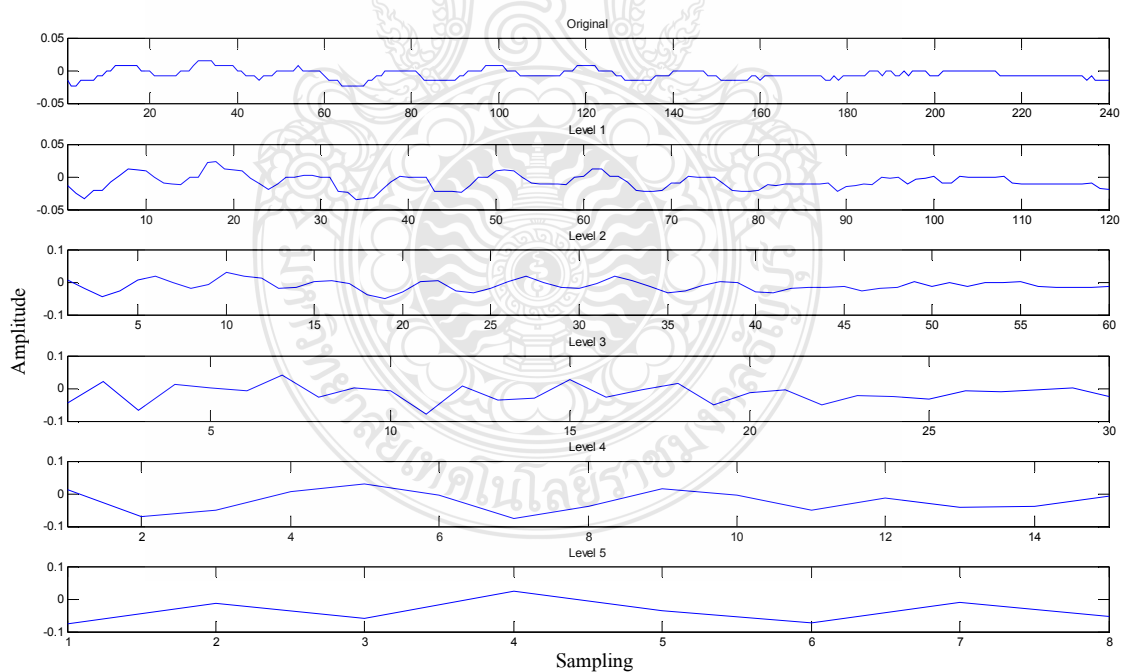
ซึ่งหากจะนำสัญญาณมาวิเคราะห์ทางเทคนิคการมองเห็น จะไม่สามารถบอกได้ว่าแต่ละระดับแตกต่างกันอย่างไร เพื่อให้สังเกตเห็นถึงความแตกต่างได้ง่ายขึ้น ดังนั้นจึงต้องนำสัญญาณเสียงแต่ละระดับทำการแบ่งเฟรม โดยกำหนดให้จำนวนเฟรมในแต่ละระดับเท่ากับจำนวนเฟรมของสัญญาณเสียงพูดต้นฉบับ และในส่วนนี้จะแสดงตัวอย่างเพียง 1 เฟรมเท่านั้น โดยได้แสดงดังภาพที่ 4.2 - 4.5



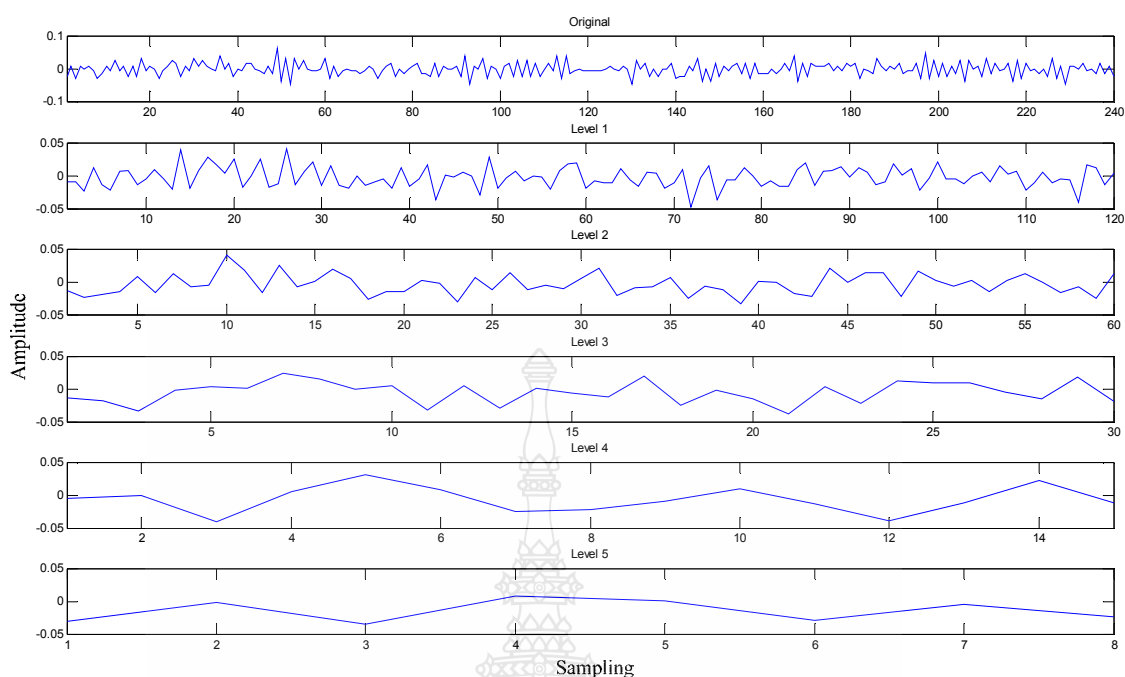
ภาพที่ 4.2 การแบ่งช่วงของสัญญาณเสียงพูดผู้หญิงที่เวลา 5 วินาที



ภาพที่ 4.3 การแบ่งช่วงของสัญญาณเสียงพูดผู้ชายที่เวลา 5 วินาที



ภาพที่ 4.4 การแบ่งช่วงของสัญญาณเสียงพูดผู้หญิงที่เวลา 60 วินาที



ภาพที่ 4.5 การแบ่งช่วงของสัญญาณเสียงพูดผู้ชายเป็นเวลา 60 วินาที

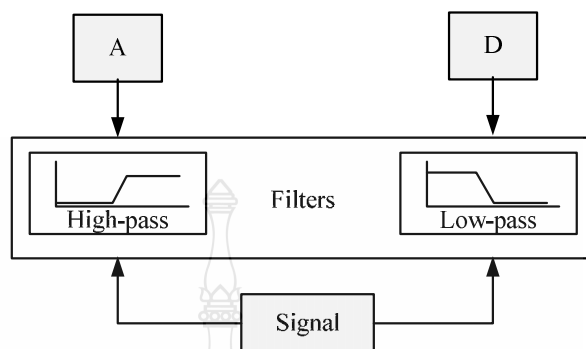
จากภาพที่ 4.2 – 4.5 จะสังเกตเห็นว่าสัญญาณเสียงที่ถูกบีบอัดด้วย DWT ในระดับที่ 1 - 3 นั้น ยังมีลักษณะคล้ายกับสัญญาณเสียงต้นฉบับค่อนข้างมาก โดยเฉพาะระดับที่ 1 ซึ่งมีความผิดเพี้ยนของสัญญาณน้อยมากแต่ในระดับที่ 4 และ ระดับที่ 5 สัญญาณมีความผิดเพี้ยนค่อนข้างมาก

ซึ่งการบีบอัดสัญญาณด้วยเวฟเลตนั้นเป็นการลดขนาดของสัญญาณเสียงหรือลดอัตราการสุ่มตัวอย่างให้มีขนาดลดลงสู่ศูนย์ โดยสังเกตจากค่าการสุ่มตัวอย่างที่ลดลงในแต่ละเฟรม จากเฟรมต้นฉบับมี 240 การสุ่มตัวอย่าง จะถูกลดลงครึ่งหนึ่งในทุกๆ ระดับการบีบอัดสัญญาณ เพื่อช่วยเพิ่มความเร็วในการส่งข้อมูลหรือสัญญาณเสียง ดังนั้นการบีบอัดด้วยระดับที่สูงๆ จะสูญเสียข้อมูลมากกว่าการบีบอัดด้วยระดับต่ำๆ และการทำให้สัญญาณเสียงกลับมาเหมือนเดิมนั้นจะต้องทำการคืนกลับสัญญาณเสียง ซึ่งในงานวิจัยนี้ได้เลือกใช้เทคนิค IDWT ในการคืนกลับสัญญาณเสียงพูด

#### 4.2 ผลการคืนกลับสัญญาณเสียงด้วย IDWT

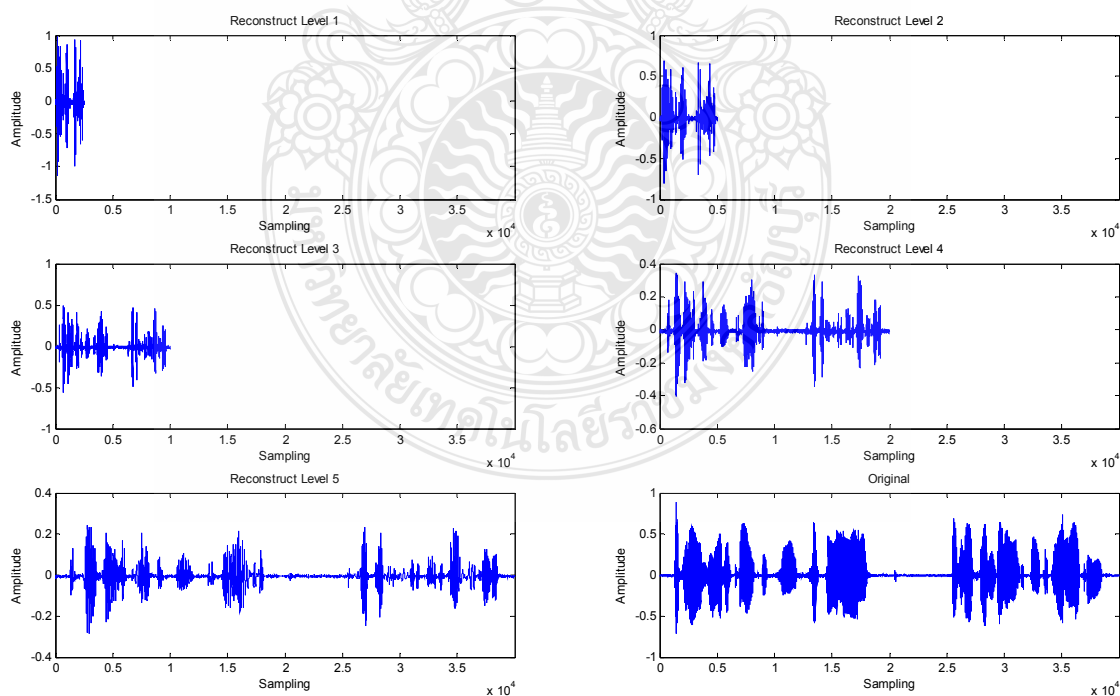
ในโครงสร้างที่กลับกัน การวิเคราะห์แบบสร้างกลับสองช่องสัญญาณ (Reconstruction 2 Channel Synthesis Filter Bank) จะนำส่วนของสัญญาณความถี่สูงและความถี่ต่ำมารวมกันเป็น

สัญญาณต้นแบบได้ ซึ่งกระบวนการนี้เรียกว่า การสังเคราะห์สัญญาณ (Synthesis) โดยได้แสดงกระบวนการไว้ดังภาพที่ 4.6



ภาพที่ 4.6 การแปลงกลับสัญญาณด้วยตัวกรองแบบ 2 ช่องสัญญาณ [8]

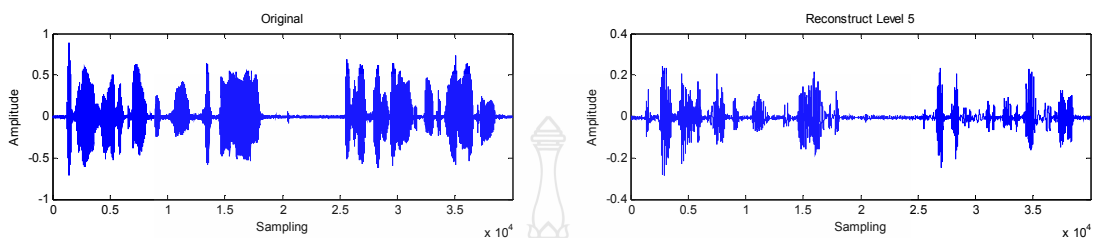
แต่เนื่องจากงานวิจัยนี้เป็นการศึกษาการลดขนาดสัญญาณ ดังนั้นจึงนำเฉพาะส่วนของสัญญาณความถี่ต่ำมาสร้างกลับเพื่อให้ใกล้เคียงสัญญาณต้นฉบับมากที่สุด ซึ่งตัวอย่างการคืนกลับสัญญาณเสียงในแต่ละระดับได้แสดงไว้ดังภาพที่ 4.7



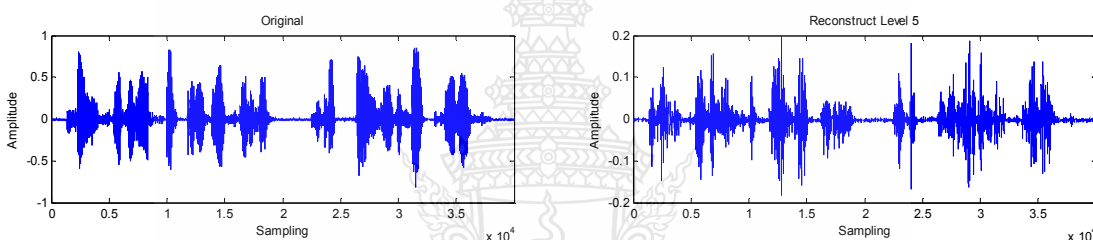
ภาพที่ 4.7 การคืนกลับสัญญาณเสียงพูดด้วย IDWT

#### 4.2.1 การคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 5

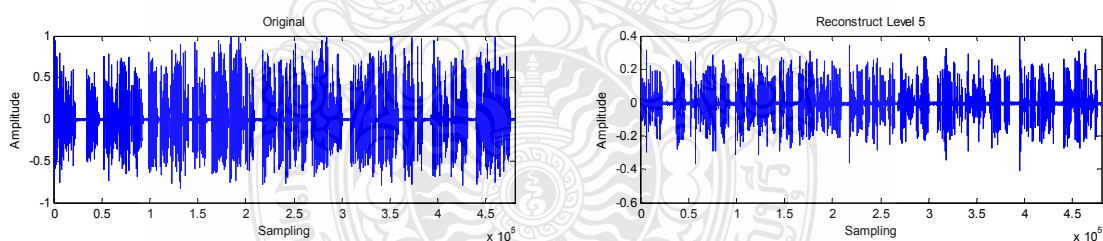
การนำสัญญาณเสียงที่ผ่านการบีบอัดจากระดับที่ 5 มาเข้ากระบวนการคืนกลับสัญญาณ ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับสัญญาณต้นฉบับดังภาพที่ 4.8



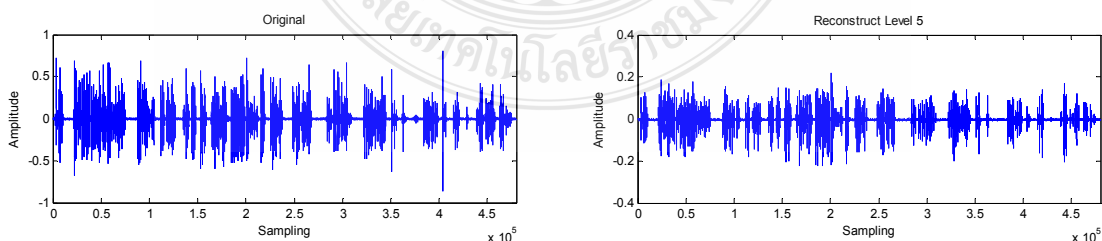
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



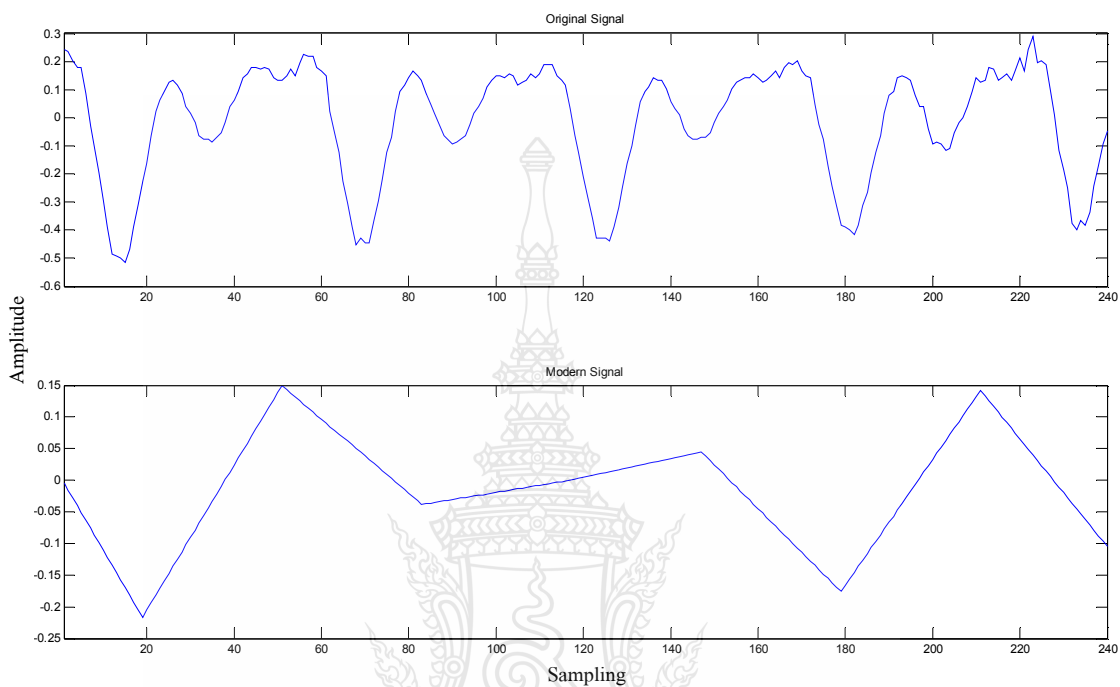
ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



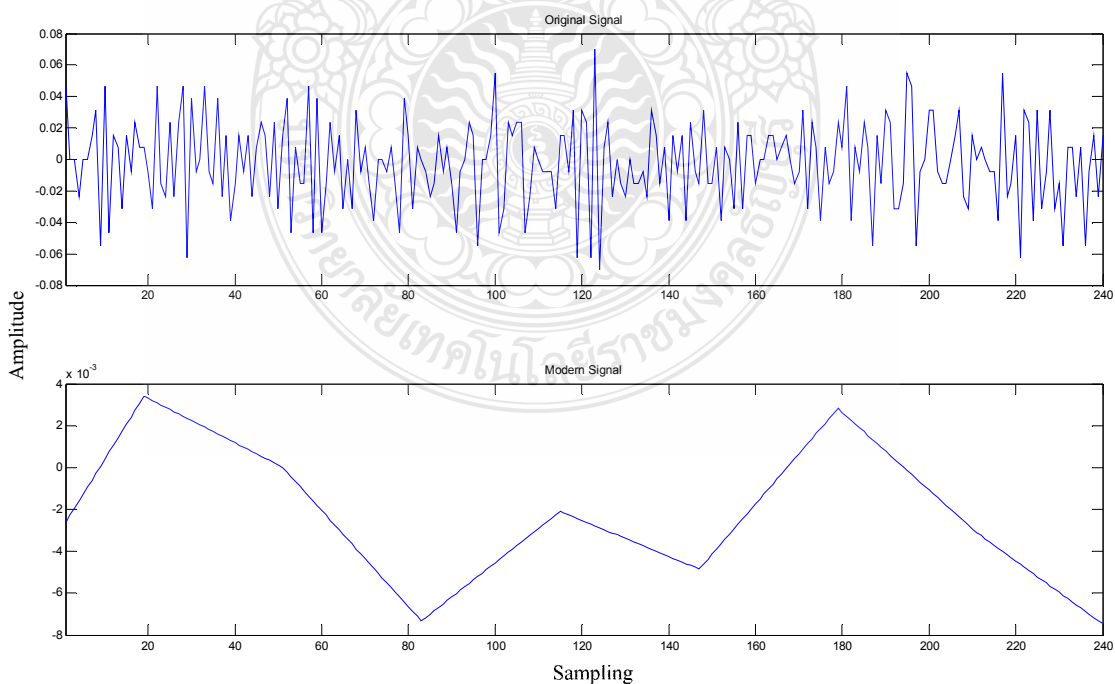
ง) เสียงพูดของผู้หญิงที่เวลา 60 วินาที

ภาพที่ 4.8 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 5 ใน 1 เฟรม

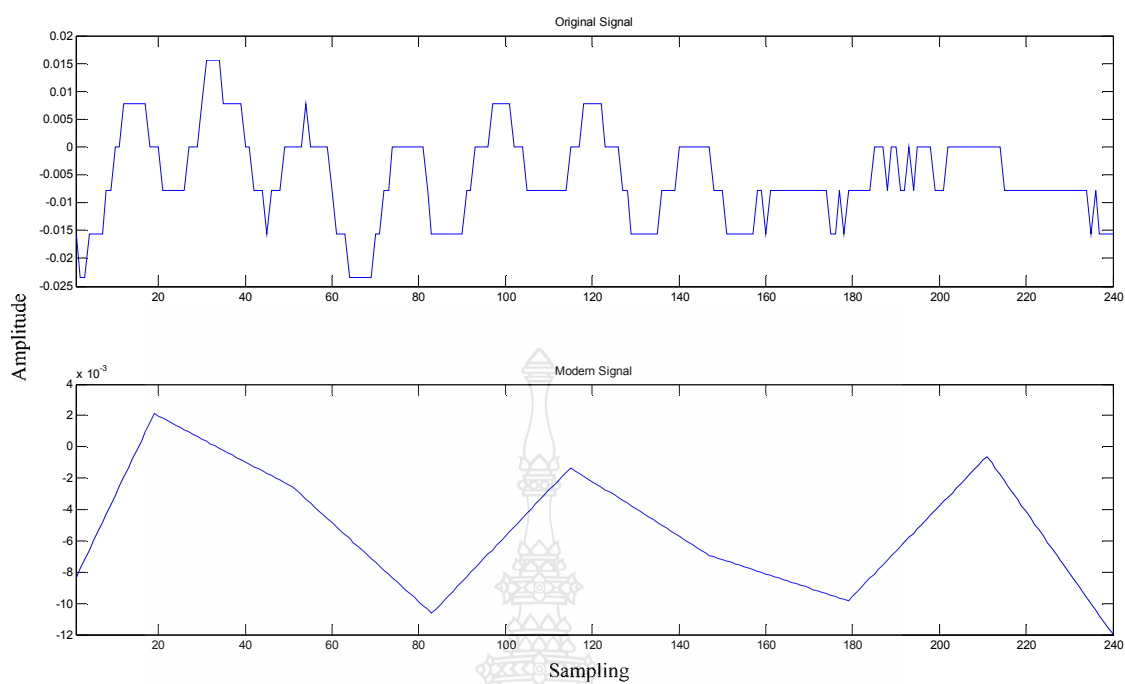
จากภาพที่ 4.8 จะสังเกตว่าภาพในส่วนที่ถูกคืนกลับสัญญาณนั้น จะมีความผิดเพี้ยนจากต้นฉบับมาก เนื่องจากเกิดสูญเสียในการบีบอัดด้วยระดับที่สูง และสามารถวิเคราะห์สัญญาณได้แม่นยำมากขึ้นและถูกต้องมากขึ้น โดยการแบ่งเฟรม ซึ่งได้แสดงในภาพที่ 4.9



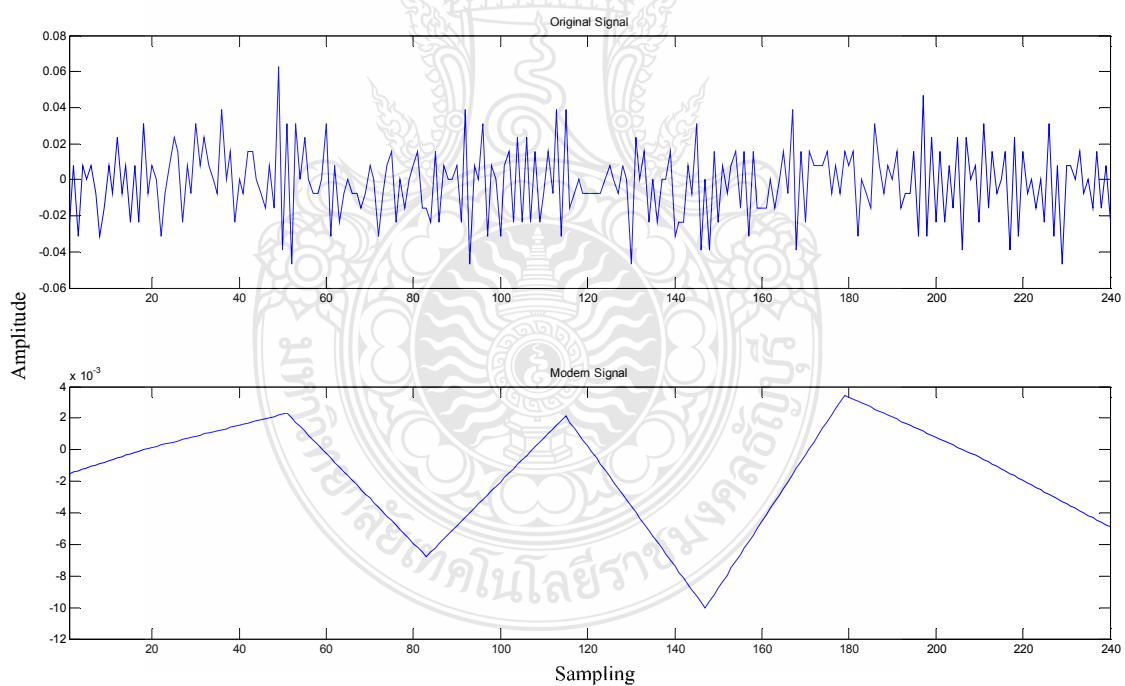
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที

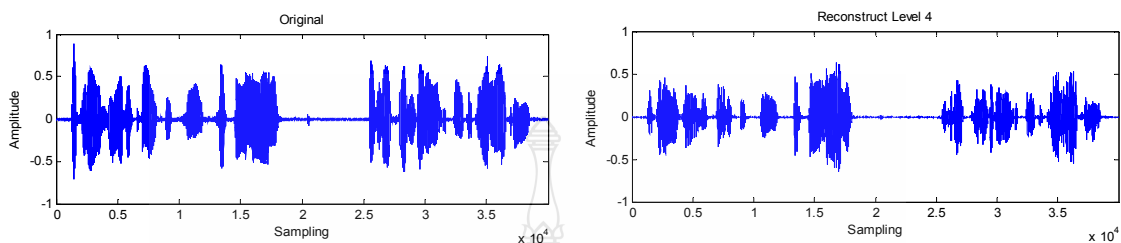


ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

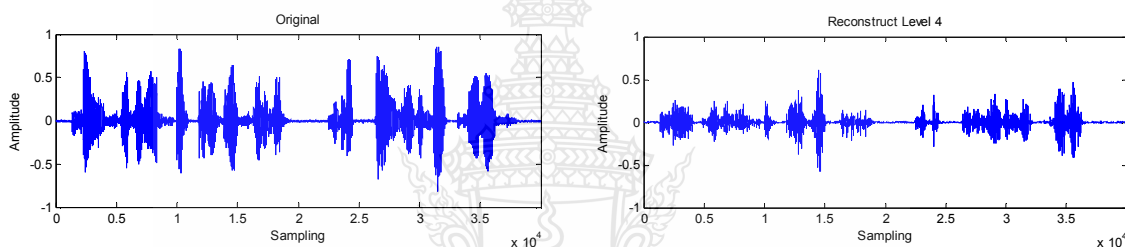
ภาพที่ 4.9 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 5 ใน 1 เฟรม

#### 4.2.2 การคืนกลับสัญญาณเสียงพูดจากการบีบอัดระดับที่ 4

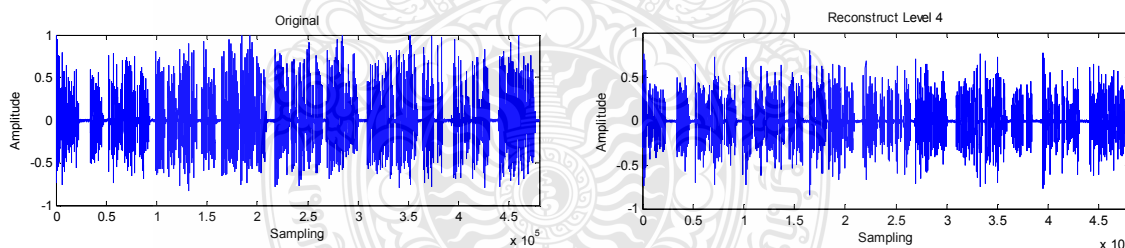
เป็นการนำสัญญาณเสียงที่ผ่านการบีบอัดระดับที่ 4 มาเข้ากระบวนการคืนกลับสัญญาณ ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับสัญญาณต้นฉบับดังภาพที่ 4.10



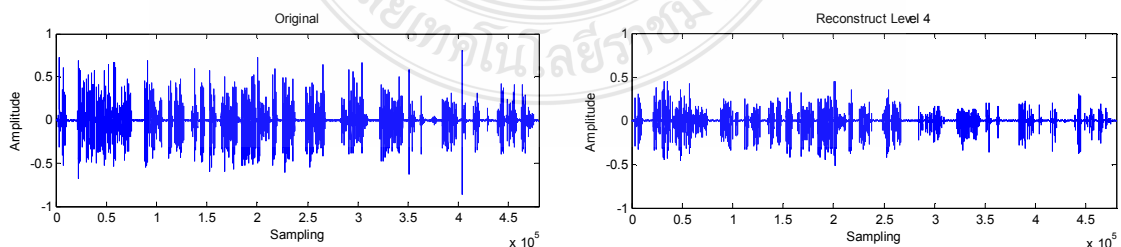
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที

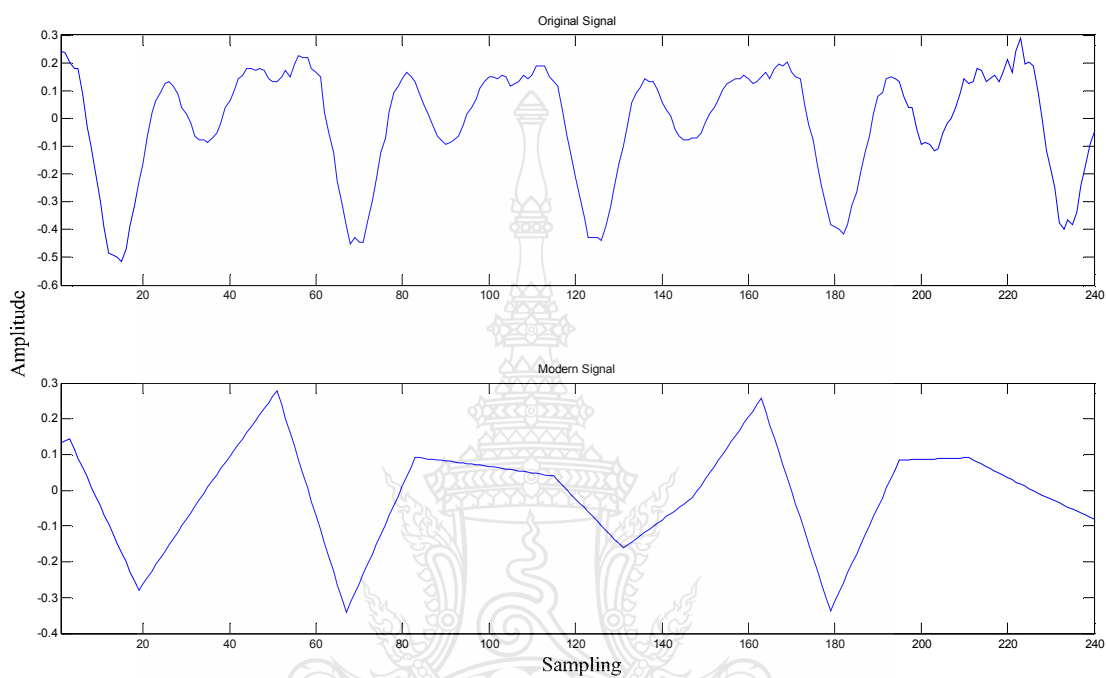


ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

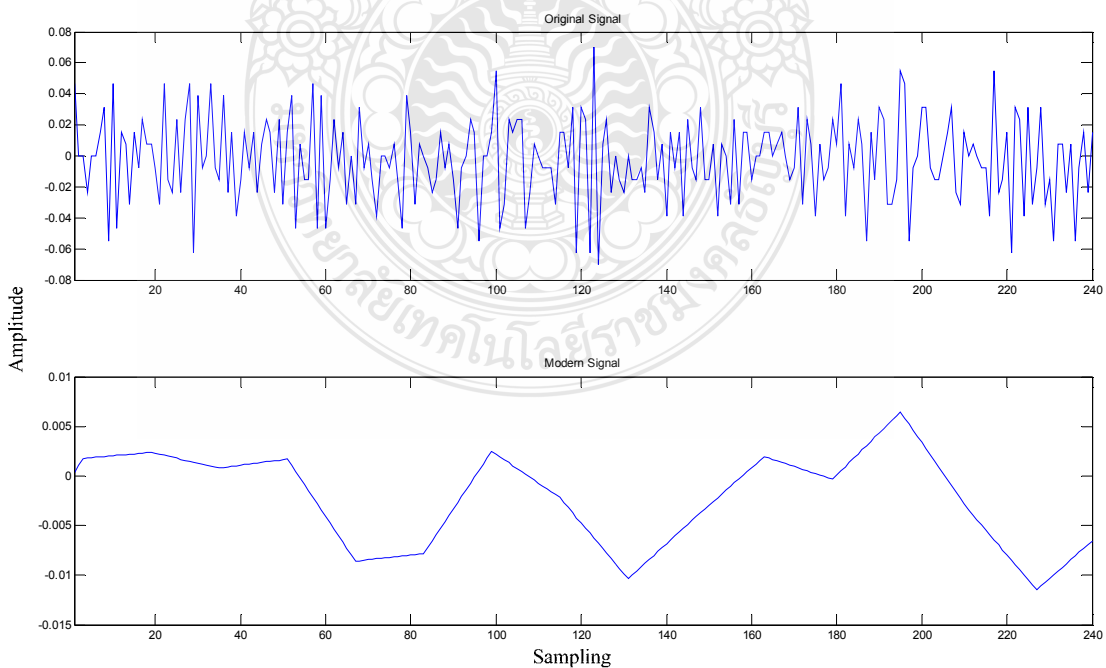
ภาพที่ 4.10 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 4



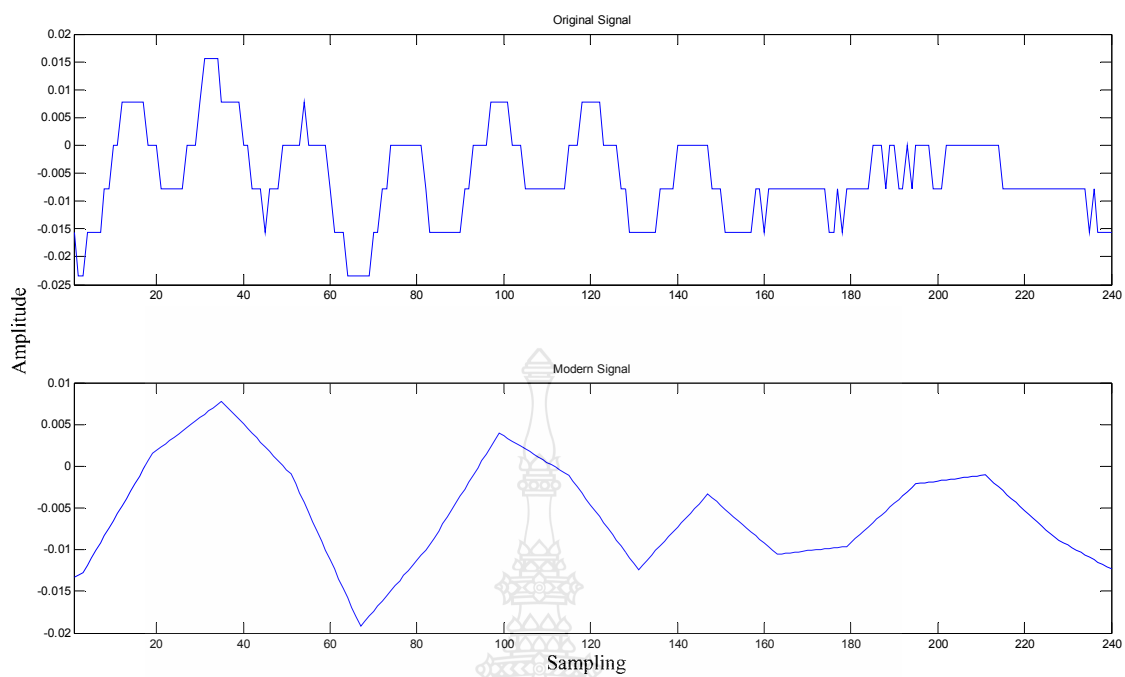
จากภาพที่ 4.10 จะสังเกตว่าภาพในส่วนที่ถูกคืนกลับสัญญาณนั้น จะมีความผิดเพี้ยนจากต้นฉบับมากแต่คุณภาพของสัญญาณเสียงจะดีกว่าการบีบอัดระดับที่ 5 ดังนั้นเพื่อให้สามารถวิเคราะห์สัญญาณได้แม่นยำมากขึ้นและถูกต้องมากขึ้น จึงจำเป็นต้องวิเคราะห์โดยการแบ่งเฟรมซึ่งได้แสดงในภาพที่ 4.11



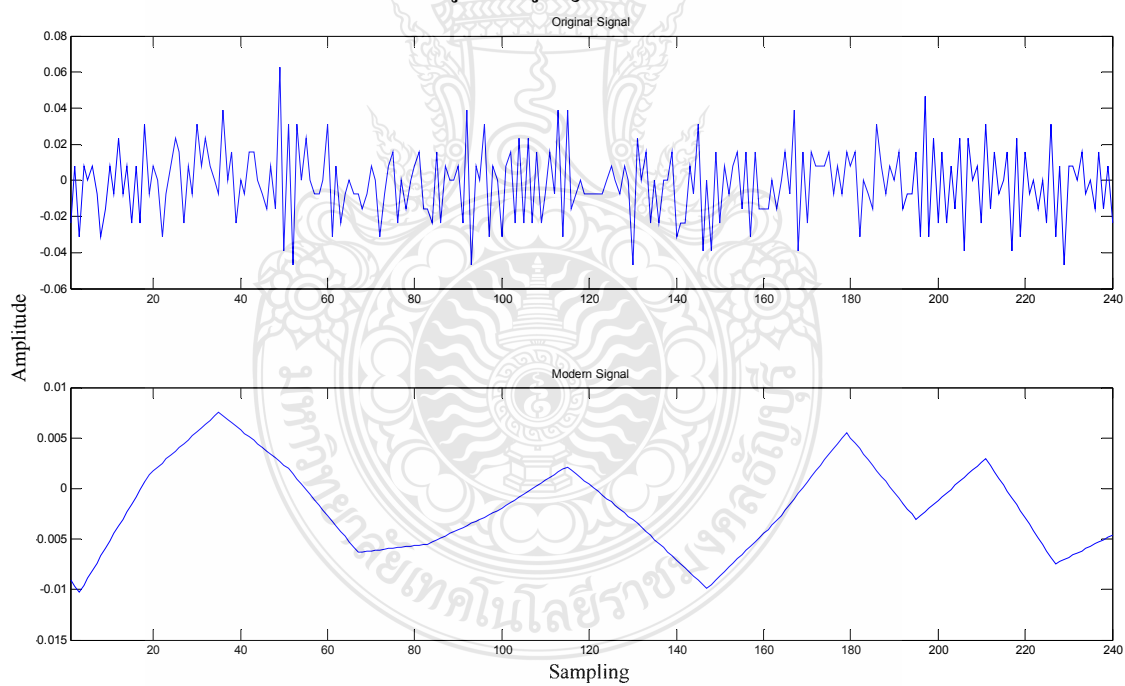
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที

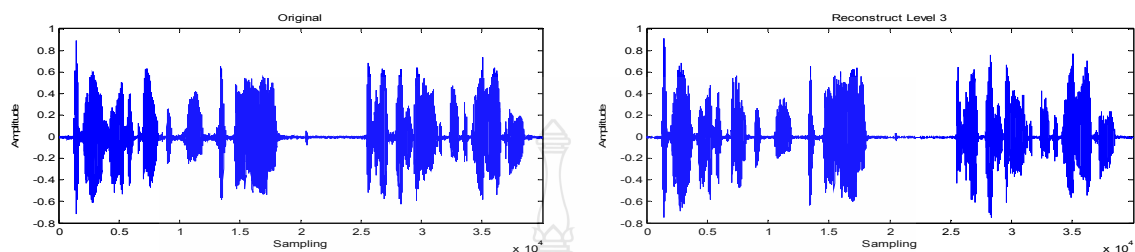


ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

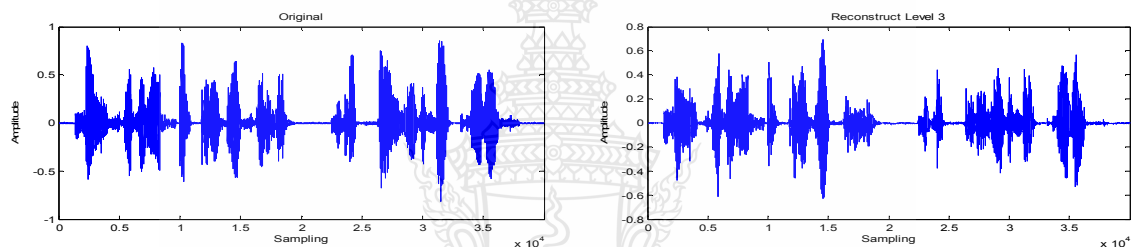
ภาพที่ 4.11 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 4 ใน 1 เฟรม

### 4.2.3 การคืนกลับสัญญาณเสียงพูดจากการบีบอัดระดับที่ 3

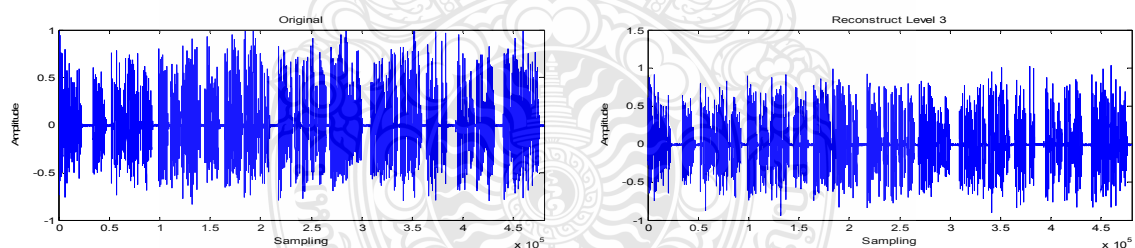
เป็นการนำสัญญาณเสียงที่ผ่านการบีบอัดระดับที่ 3 มาเข้ากระบวนการคืนกลับสัญญาณ ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับสัญญาณต้นฉบับดังภาพที่ 4.12



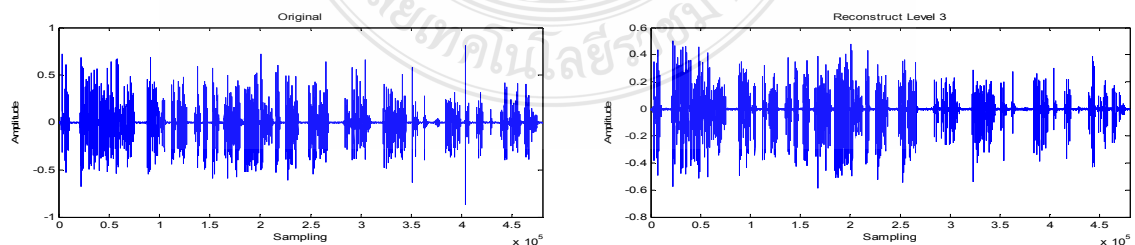
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



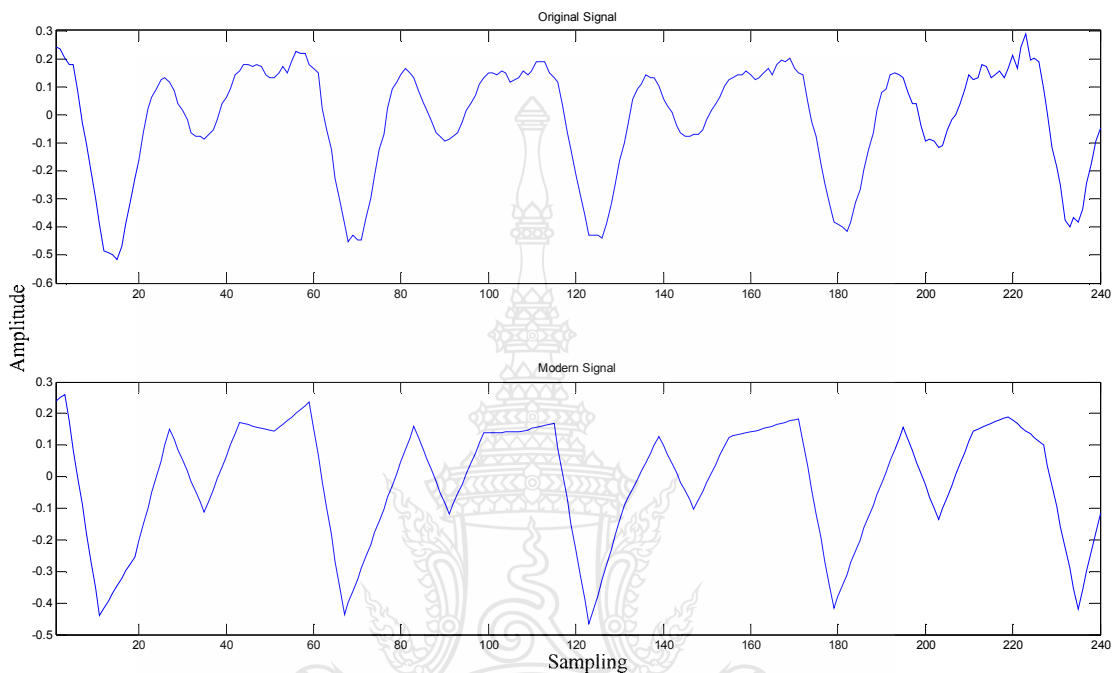
ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



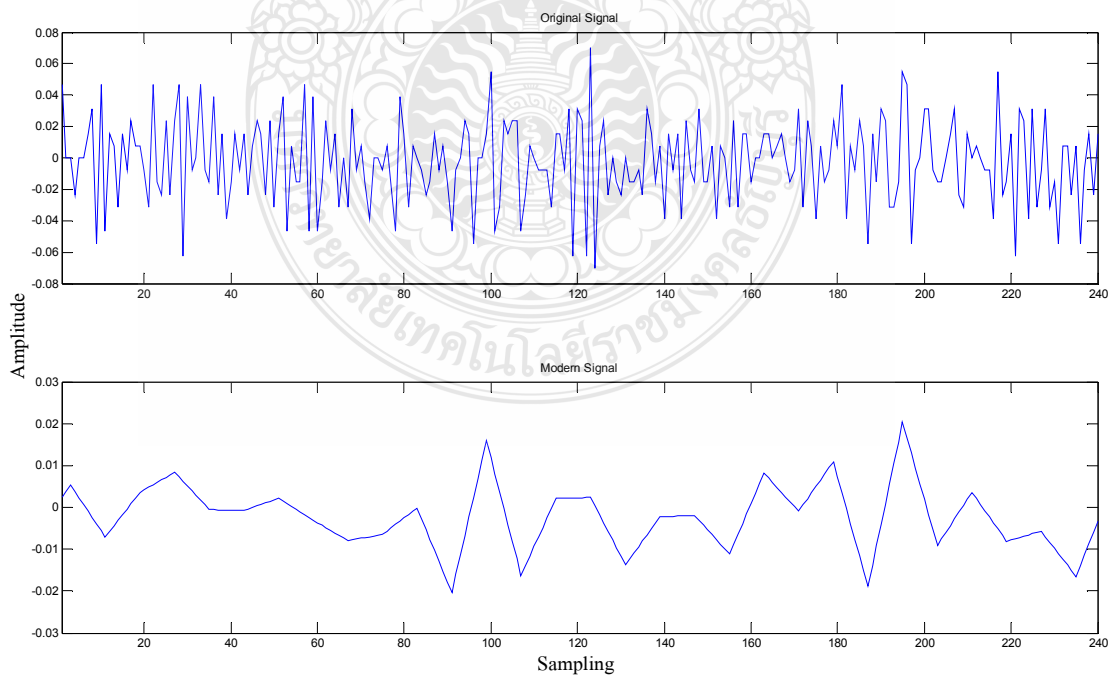
ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

ภาพที่ 4.12 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 3

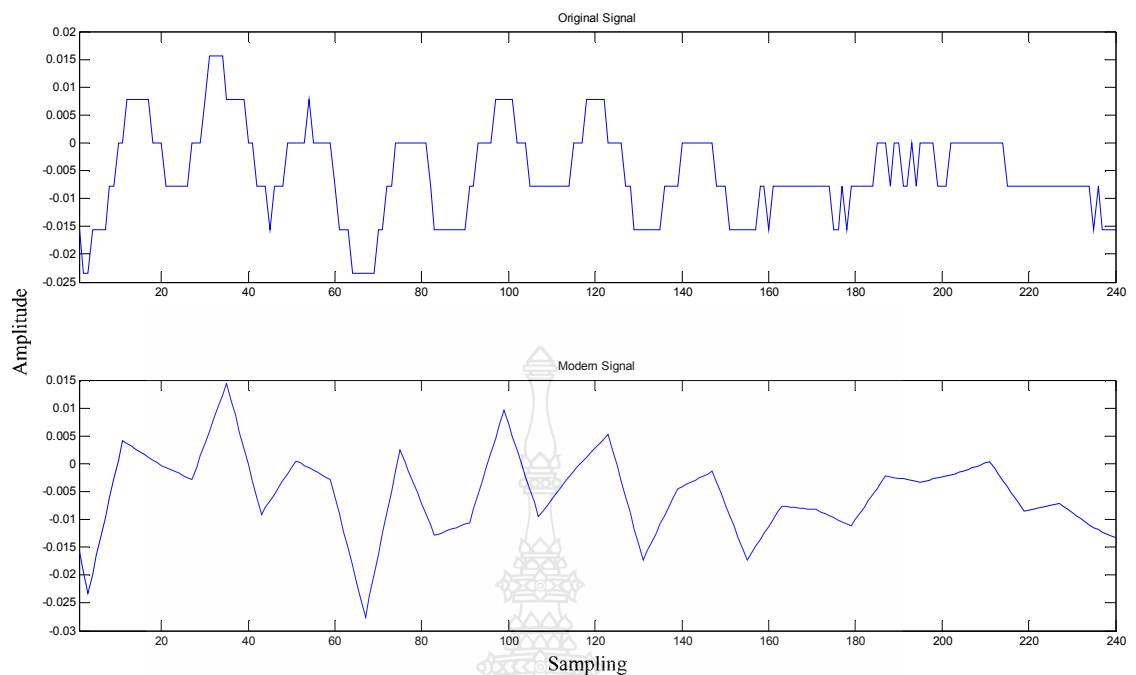
จากภาพที่ 4.12 จะสังเกตว่าภาพในส่วนที่ถูกคืนกลับสัญญาณนั้น จะมีความผิดเพี้ยนจากต้นฉบับไม่มากและคุณภาพของสัญญาณเสียงจะดีกว่าการบีบอัดที่ระดับที่ 4 และ 5 ดังนั้นเพื่อให้สามารถวิเคราะห์สัญญาณได้แม่นยำมากขึ้นและถูกต้องมากขึ้น จึงจำเป็นต้องวิเคราะห์โดยการแบ่งเฟรม ซึ่งได้แสดงในภาพที่ 4.13



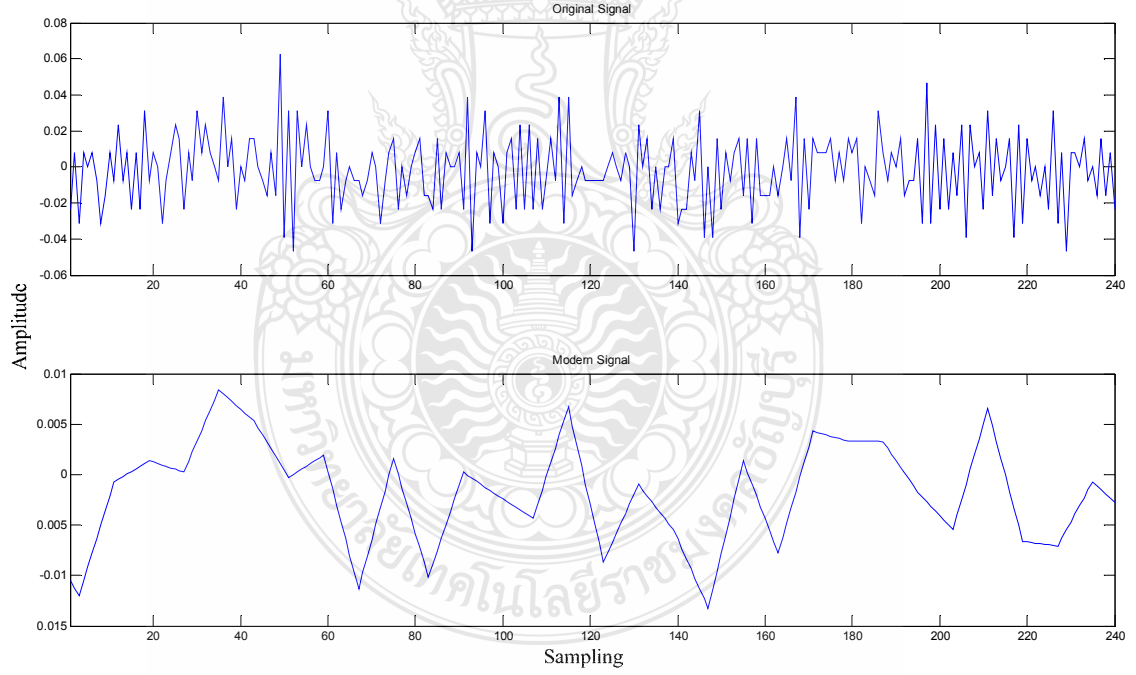
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที

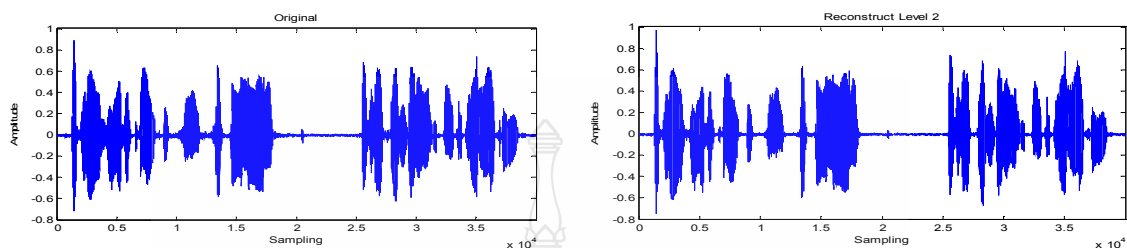


ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

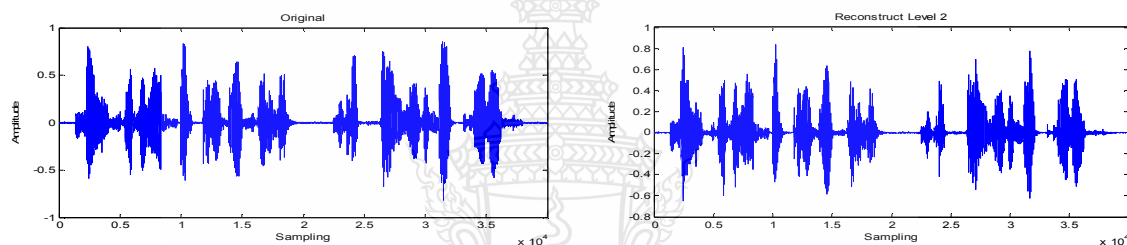
ภาพที่ 4.13 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 3 ใน 1 เฟรม

#### 4.2.4 การคืนกลับสัญญาณเสียงพูดจากการบีบอัดระดับที่ 2

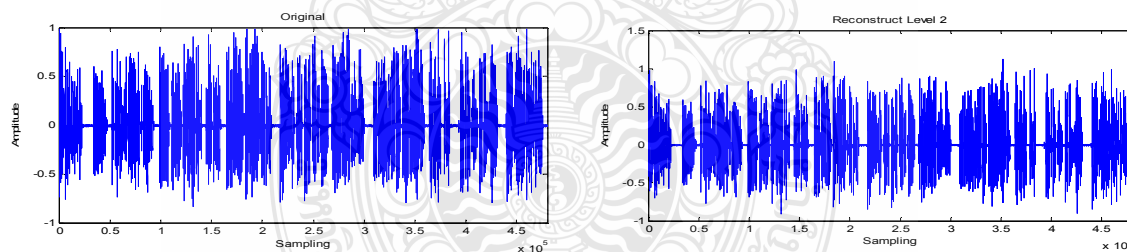
เป็นการนำสัญญาณเสียงที่ผ่านการบีบอัดระดับที่ 2 มาเข้ากระบวนการคืนกลับสัญญาณ ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับสัญญาณต้นฉบับดังภาพที่ 4.14



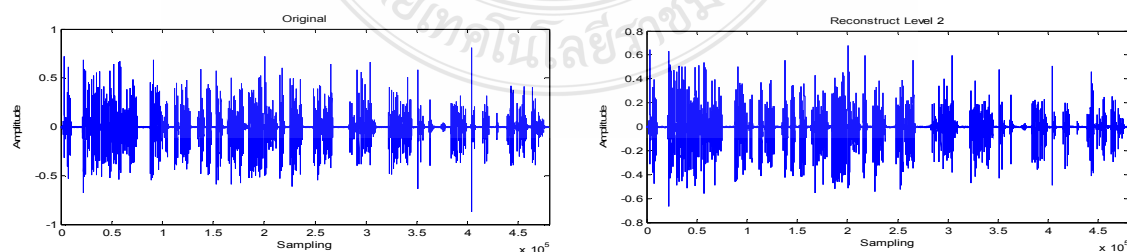
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



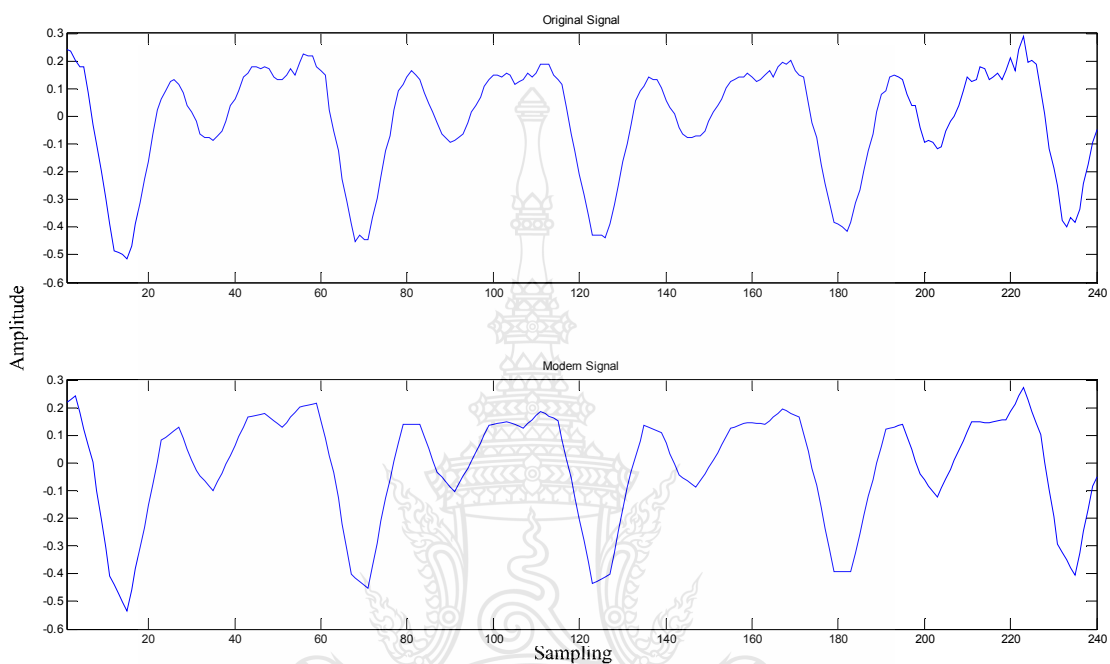
ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



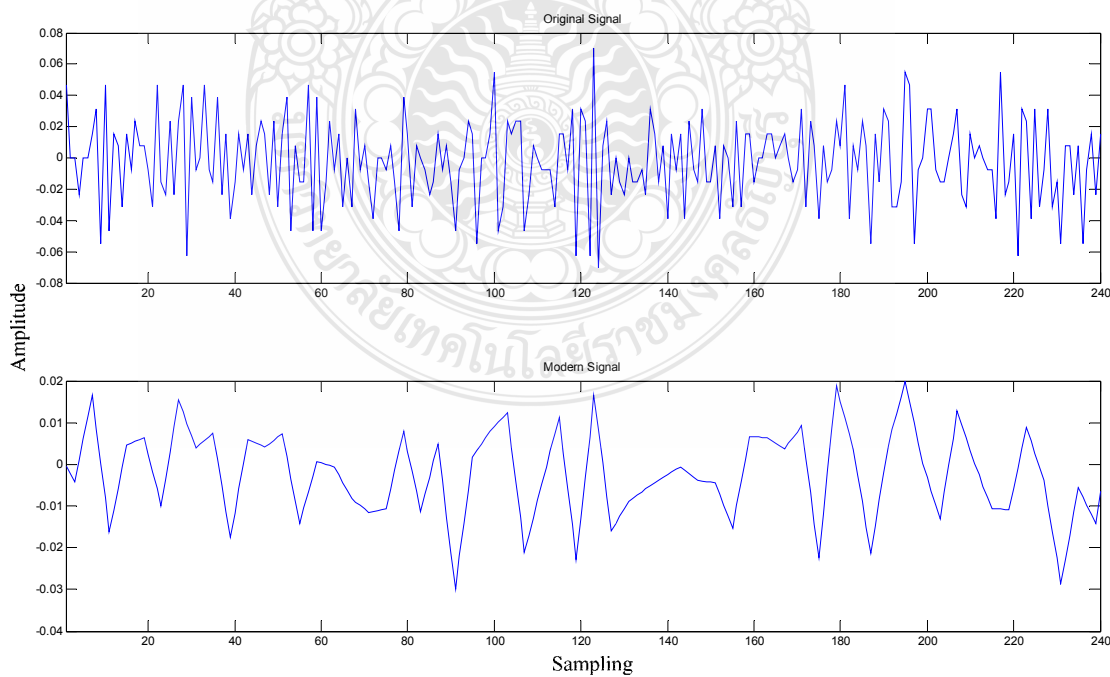
ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

ภาพที่ 4.14 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 2

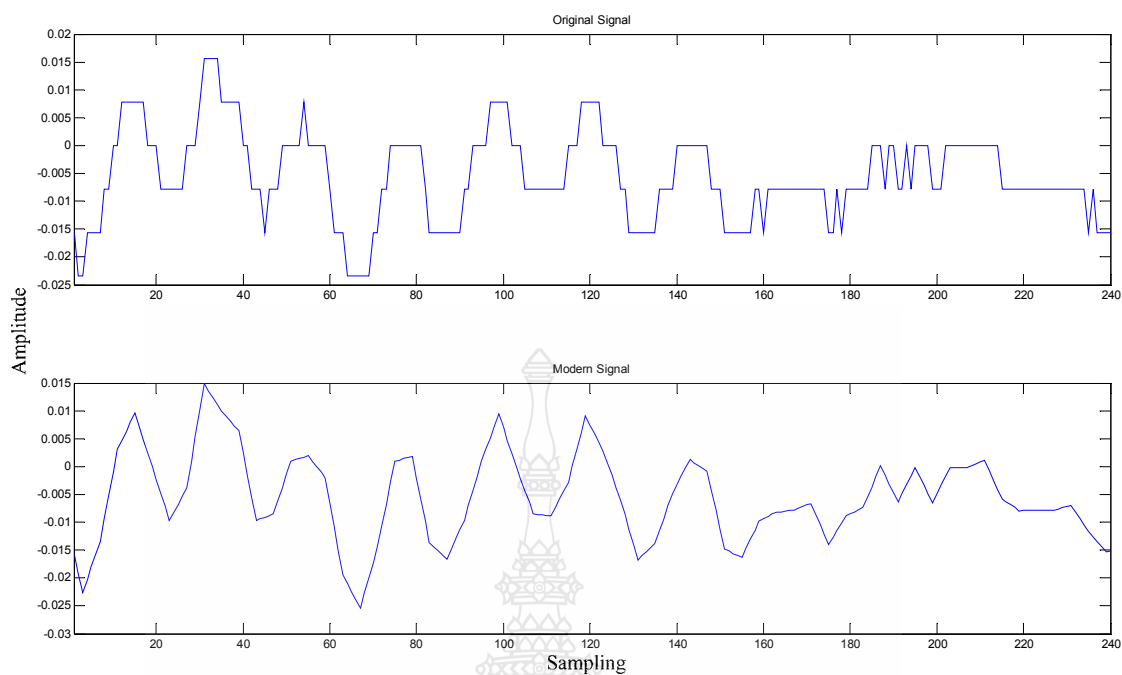
จากภาพที่ 4.14 จะสังเกตว่าภาพในส่วนที่ถูกคืนกลับสัญญาณนั้น จะมีความผิดเพี้ยนจากต้นฉบับเพียงเล็กน้อยและคุณภาพของสัญญาณเสียงจะดีกว่าการบีบอัดระดับที่ 3, 4 และ 5 ดังนั้นเพื่อให้สามารถวิเคราะห์สัญญาณได้แม่นยำมากขึ้นและถูกต้องมากขึ้น จึงจำเป็นต้องวิเคราะห์โดยการแบ่งเฟรม ซึ่งได้แสดงในภาพที่ 4.15



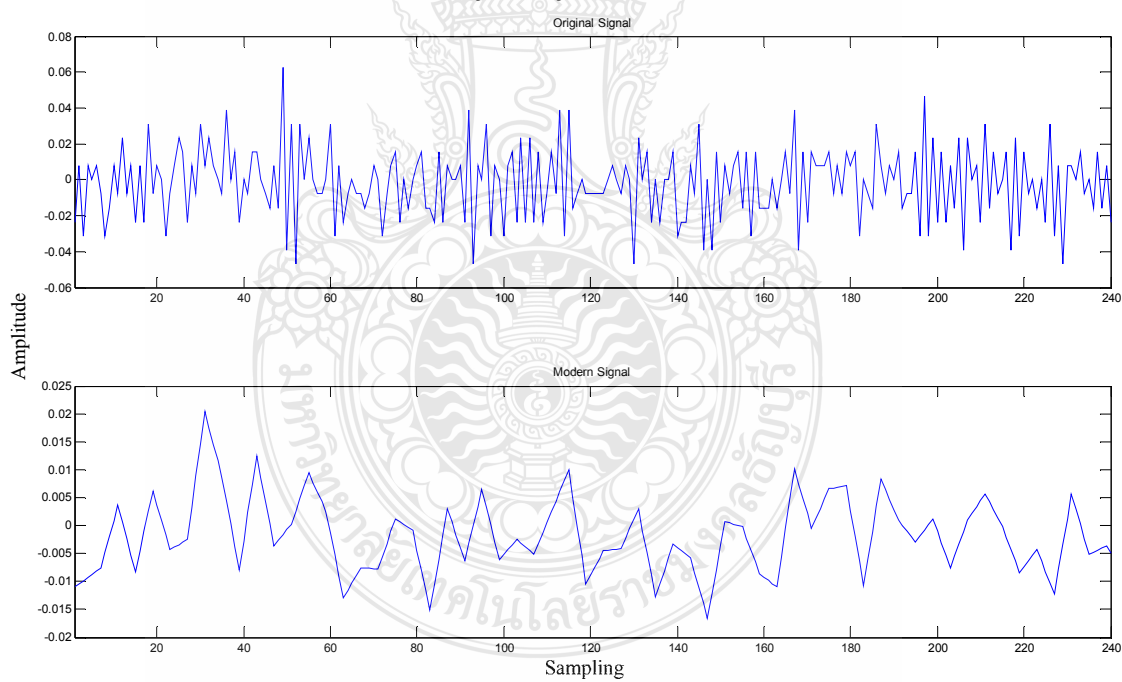
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



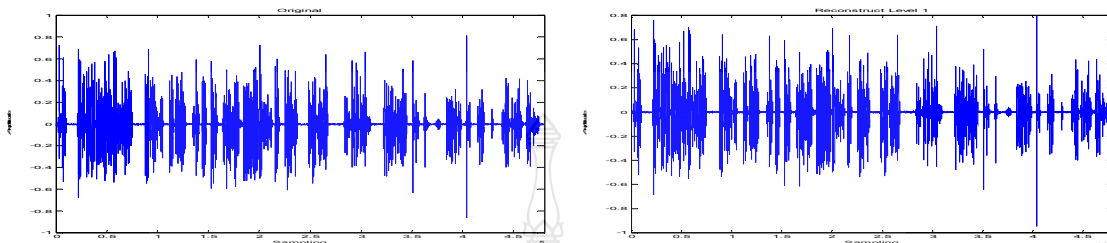
ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

ภาพที่ 4.15 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 2 ใน 1 เฟรม

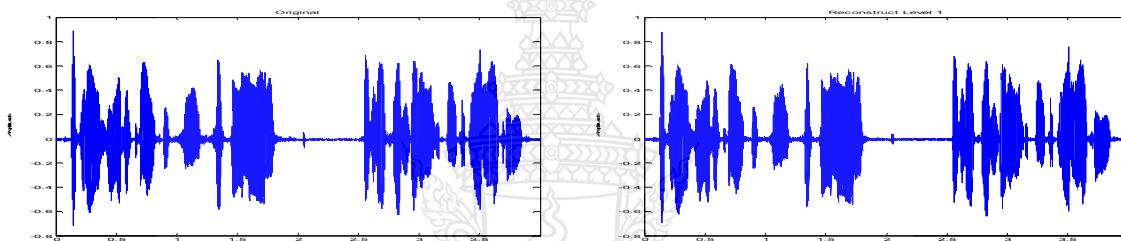


#### 4.2.5 การคืนกลับสัญญาณเสียงพูดจากการบีบอัดระดับที่ 1

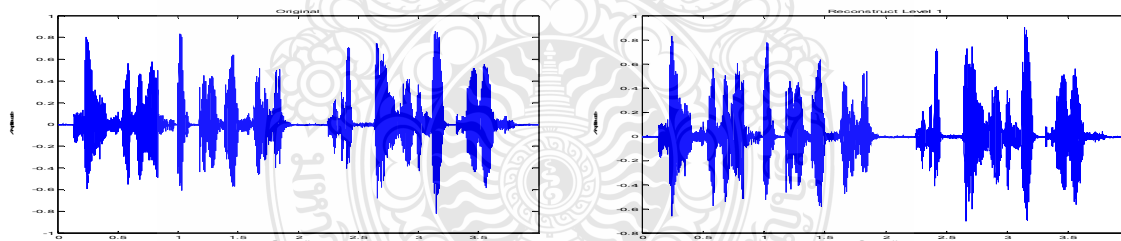
เป็นการนำสัญญาณเสียงที่ผ่านการบีบอัดระดับที่ 1 มาเข้ากระบวนการคืนกลับสัญญาณ ซึ่งผลที่ได้จะนำไปเปรียบเทียบกับสัญญาณต้นฉบับดังภาพที่ 4.14



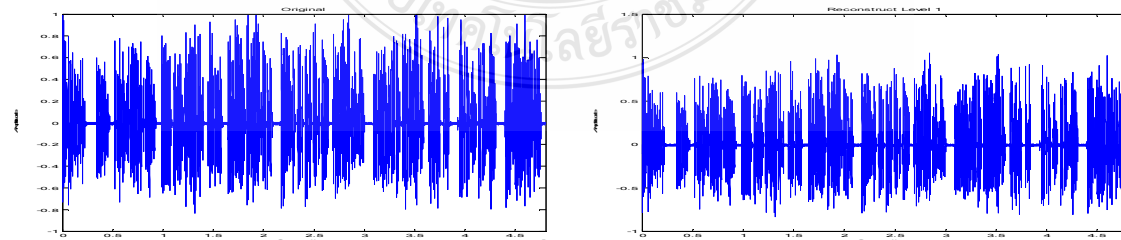
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



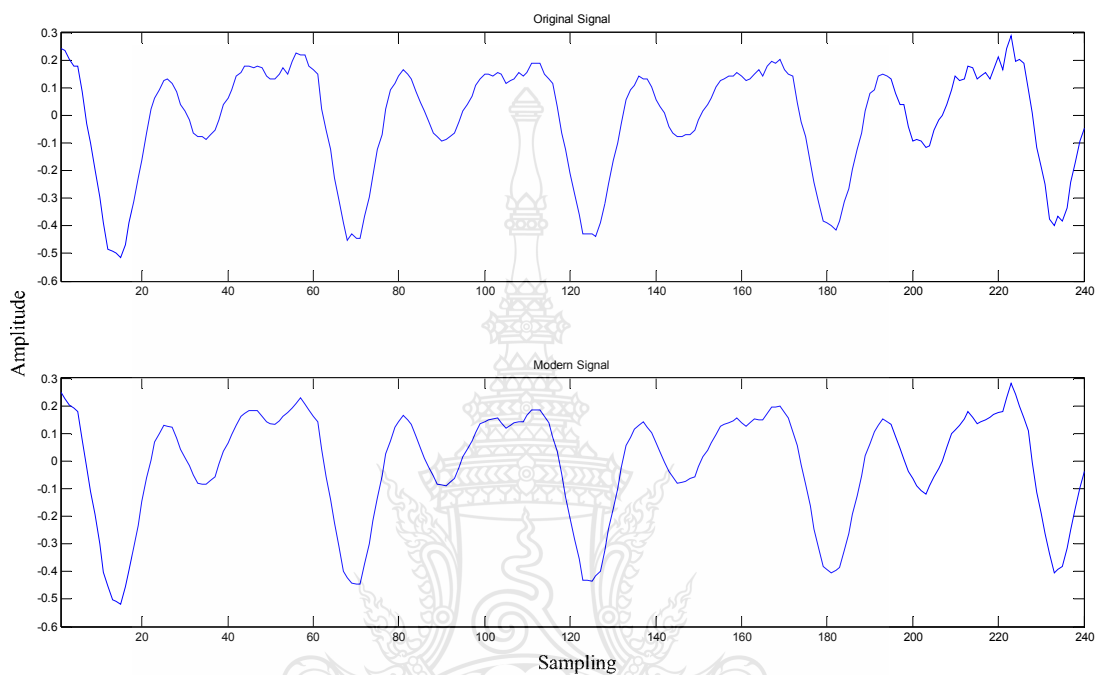
ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



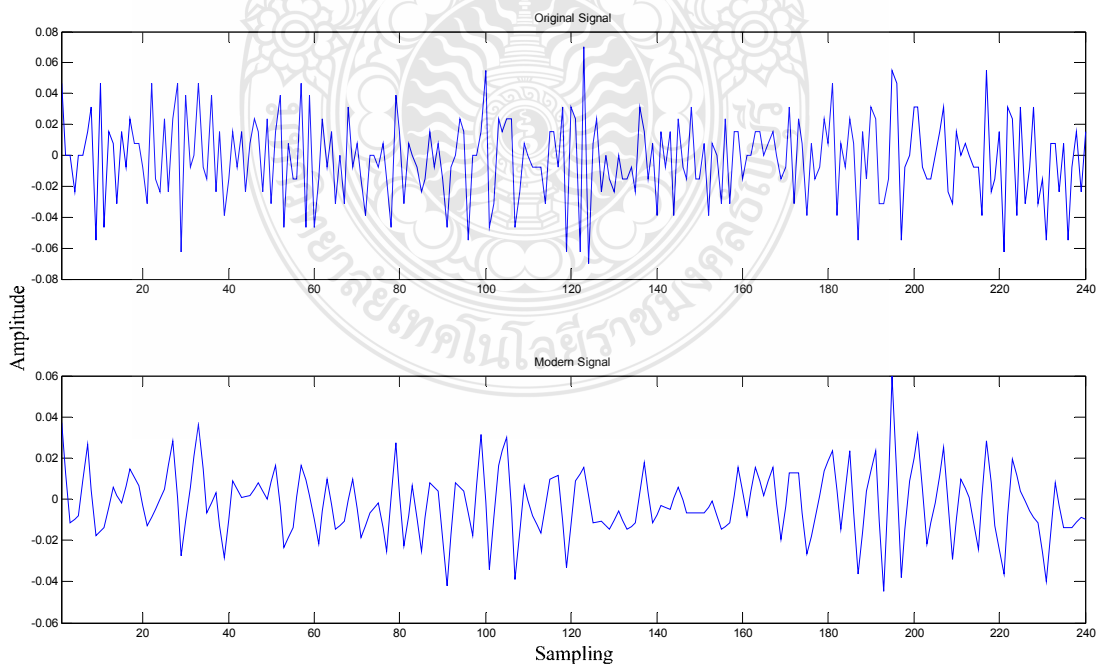
ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

ภาพที่ 4.16 การเปรียบเทียบการคืนกลับสัญญาณเสียงจากการบีบอัดระดับที่ 1

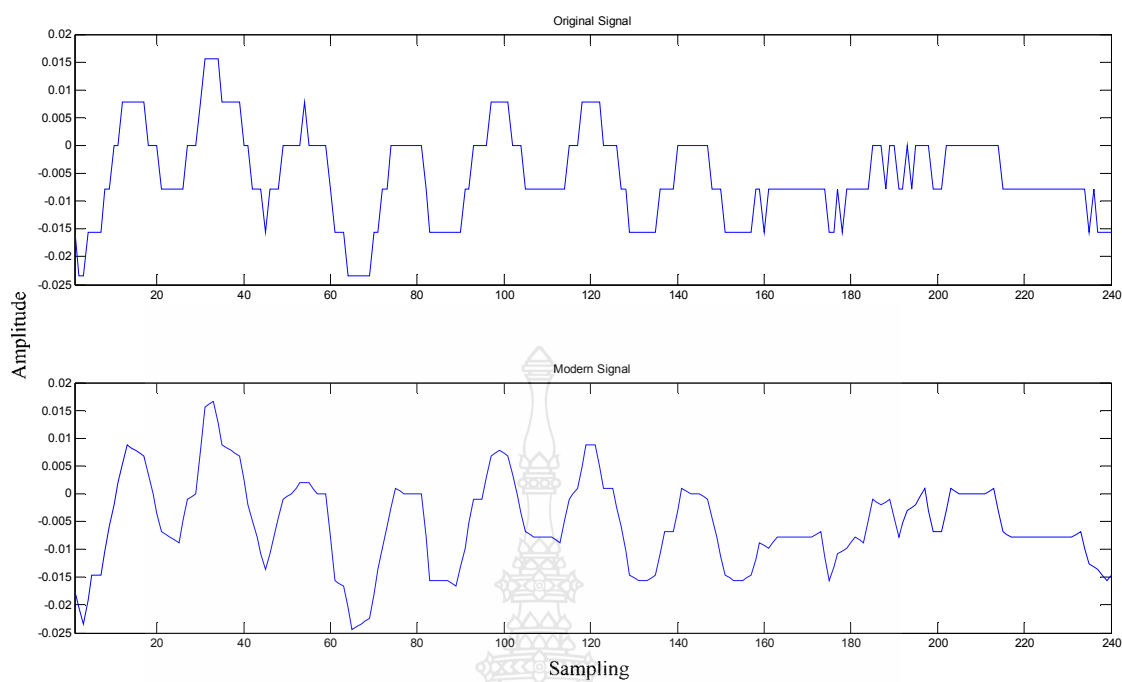
จากภาพที่ 4.16 จะสังเกตว่าภาพในส่วนที่ถูกคืนกลับสัญญาณนั้น จะมีความผิดเพี้ยนจากต้นฉบับเพียงน้อยมากและคุณภาพของสัญญาณเสียงจะใกล้เคียงกับต้นฉบับมากที่สุด ดังนั้นเพื่อให้สามารถวิเคราะห์สัญญาณได้แม่นยำมากขึ้นและถูกต้องมากขึ้น จึงจำเป็นต้องวิเคราะห์โดยการแบ่งเฟรม ซึ่งได้แสดงในภาพที่ 4.17



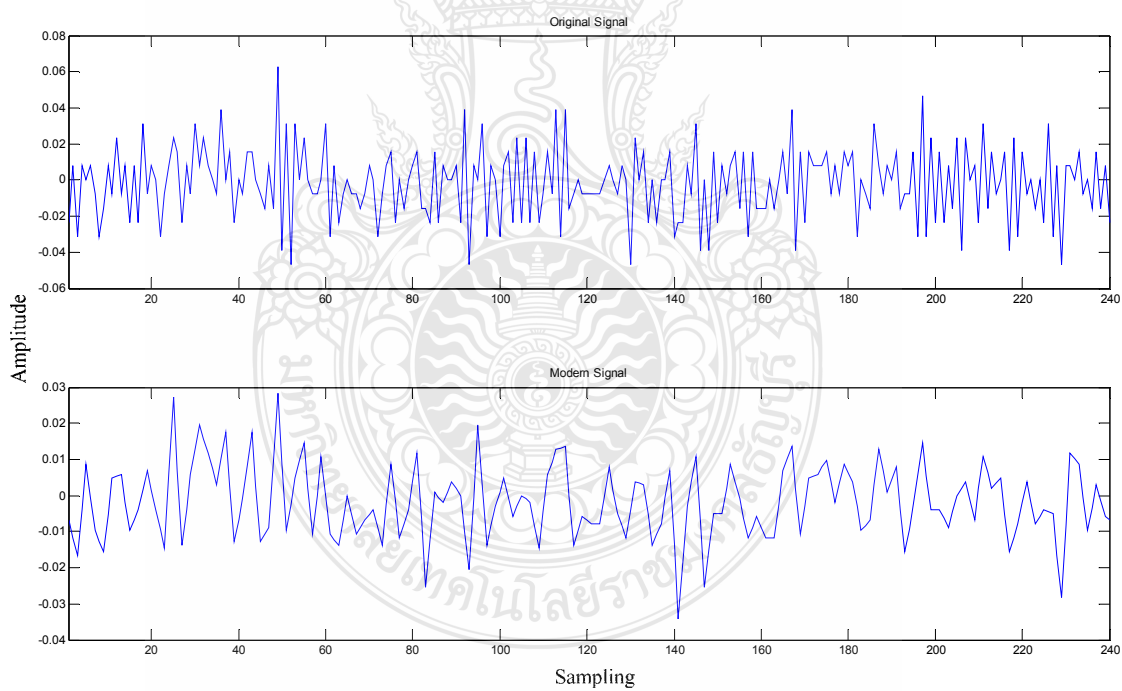
ก) เสียงพูดของผู้หญิงที่เวลา 5 วินาที



ข) เสียงพูดของผู้ชายที่เวลา 5 วินาที



ค) เสียงพูดของผู้หญิงที่เวลา 60 วินาที



ง) เสียงพูดของผู้ชายที่เวลา 60 วินาที

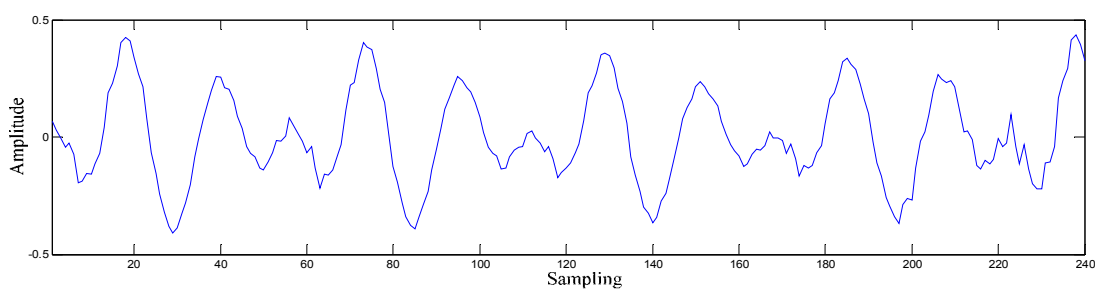
ภาพที่ 4.17 การแบ่งช่วงของสัญญาณเสียงที่ผ่านกระบวนการคืนกลับสัญญาณระดับที่ 1 ใน 1 เฟรม

จากภาพที่ 4.9, 4.11, 4.13, 4.15 และ 4.17 เป็นการแบ่งช่วงของสัญญาณเสียงพูดที่ได้จากการคืนกลับสัญญาณเสียงด้วย IDWT โดยแบ่งเป็น 240 การสุ่มตัวอย่างต่อ 1 เฟรม จะเห็นว่าการบีบอัดสัญญาณเสียงระดับที่ 4 และระดับที่ 5 จะสูญเสียการสุ่มตัวอย่างมาก เนื่องจากสัญญาณต้นฉบับจะมีการสุ่มตัวอย่างที่ 8000 ครั้งต่อวินาที ดังนั้นเมื่อบีบอัดสัญญาณเสียงระดับที่ 4 และ ระดับที่ 5 สัญญาณเสียงใหม่ที่ได้จะมีการสุ่มตัวอย่างเหลือเพียง 500 ครั้งต่อวินาทีของการบีบอัดระดับที่ 4 และ 250 ครั้งต่อวินาทีของการบีบอัดระดับที่ 5 โดยถ้านำไปคืนกลับสัญญาณเสียงคุณภาพของสัญญาณเสียงจะต่ำมาก ดังที่แสดงในภาพที่ 4.9 และ 4.11 จะเห็นถึงความแตกต่างของสัญญาณเสียงที่ถูกคืนกลับเมื่อเปรียบเทียบกับสัญญาณเสียงพูดต้นฉบับ

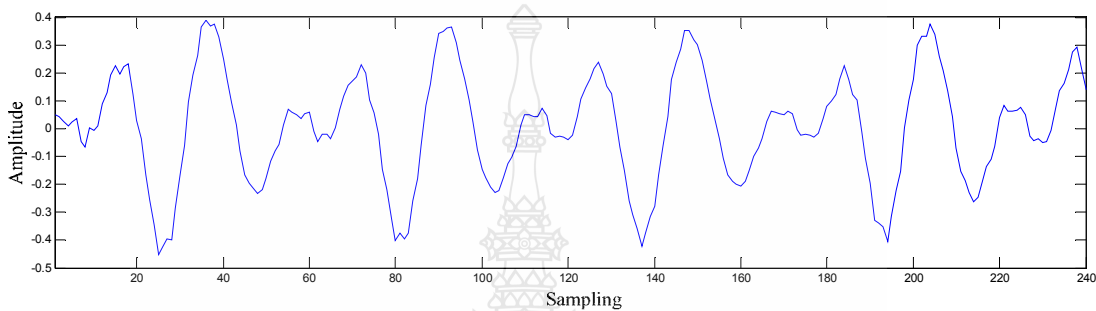
ดังนั้นในงานวิจัยนี้จึงเลือกเฉพาะการบีบอัดระดับที่ 1, 2 และ 3 เท่านั้น เนื่องจากผลของสัญญาณเสียงใหม่ที่ได้จากการคืนกลับสัญญาณเสียงในภาพที่ 4.13, 4.15 และ 4.17 มีความใกล้เคียงกับต้นฉบับค่อนข้างมาก โดยเฉพาะในการบีบอัดสัญญาณเสียงระดับที่ 1 ภาพของสัญญาณมีความคล้ายกับต้นฉบับมากที่สุด ถึงแม้จะมีการสูญเสียสัญญาณส่วนที่เป็นความถี่สูง (Detail) ออกไป

#### 4.3 ผลการบีบอัดสัญญาณเสียงด้วยเทคนิค CELP

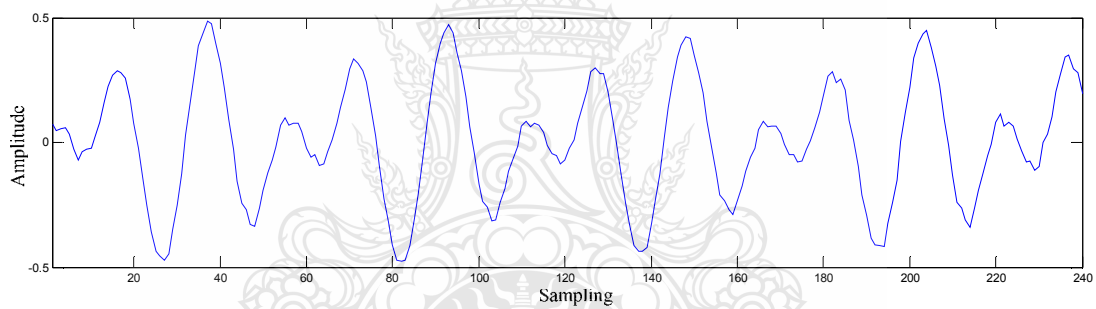
ในส่วนงานวิจัยนี้ได้นำเทคนิค CELP มาวิเคราะห์เพื่อเปรียบเทียบกับเวฟเล็ต เนื่องจาก CELP เป็นเทคโนโลยีในการบีบอัดสัญญาณเสียงที่ได้รับมาตรฐาน ซึ่งการทำงานของ CELP จะนำมาวิเคราะห์โดยใช้การสังเคราะห์จาก Linear Prediction และมีการกระตุ้นของสัญญาณด้วย Codebook ซึ่งลำดับการกระตุ้นในระบบ CELP สามารถแบ่งออกเป็น 2 ประเภท คือ Adaptive Codebook และ Stochastic Codebook ซึ่งในส่วนของ Adaptive Codebook จะถูกนำมาเพื่อสร้างสัญญาณกระตุ้นความถี่ต่ำ และ Stochastic Codebook จะถูกนำมาเพื่อแสดงความแตกต่างระหว่างรูปแบบของคลื่นที่เกิดขึ้นจริงและส่วนขยายระยะที่เหมาะสมของการกระตุ้น จากการทดลองนี้สามารถเปรียบเทียบให้เห็นถึงความแตกต่างระหว่างสัญญาณเสียงที่ถูกสังเคราะห์ขึ้นมาใหม่จากเทคนิค CELP กับ สัญญาณเสียงต้นฉบับดังภาพที่ 4.18 – 4.21



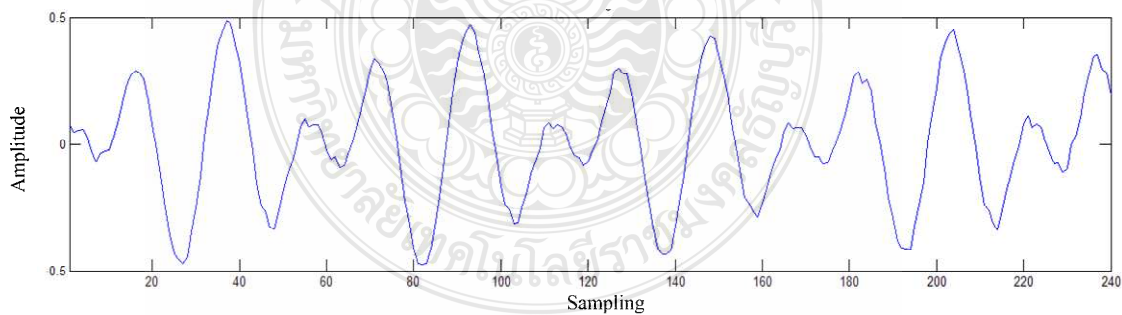
ก) สัญญาณเสียงต้นฉบับ



ข) Order 10

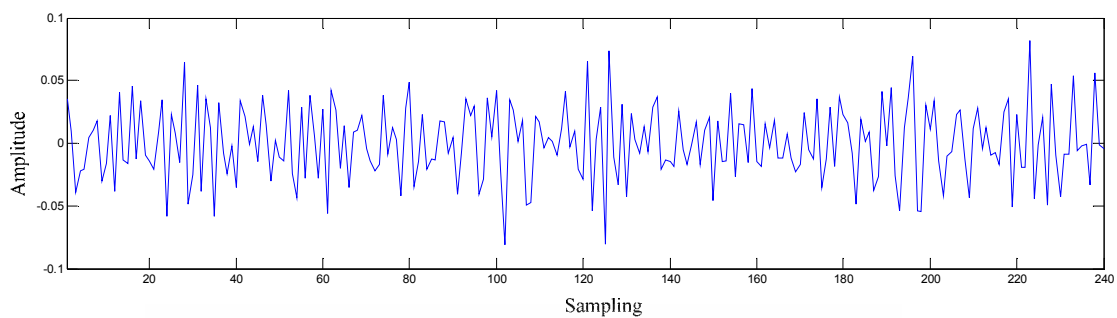


ค) Order 20

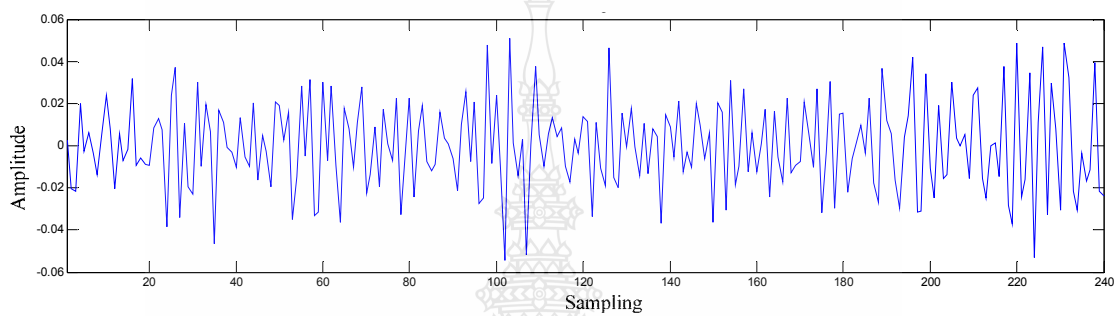


ง) Order 40

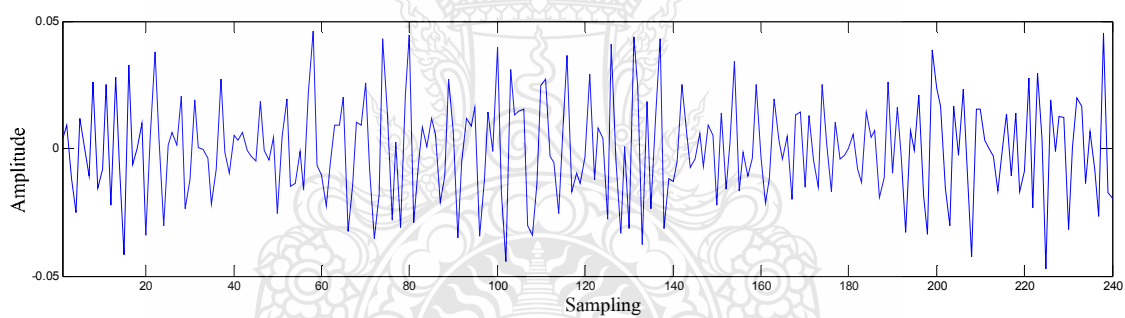
ภาพที่ 4.18 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้ชายที่ 5 วินาที ด้วย CELP ใน 1 เฟรม



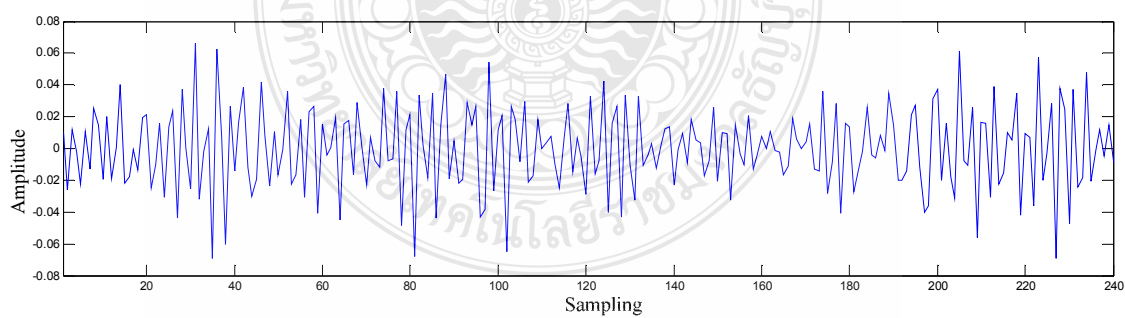
ก) สัญญาณเสียงต้นฉบับ



ข) Order 10

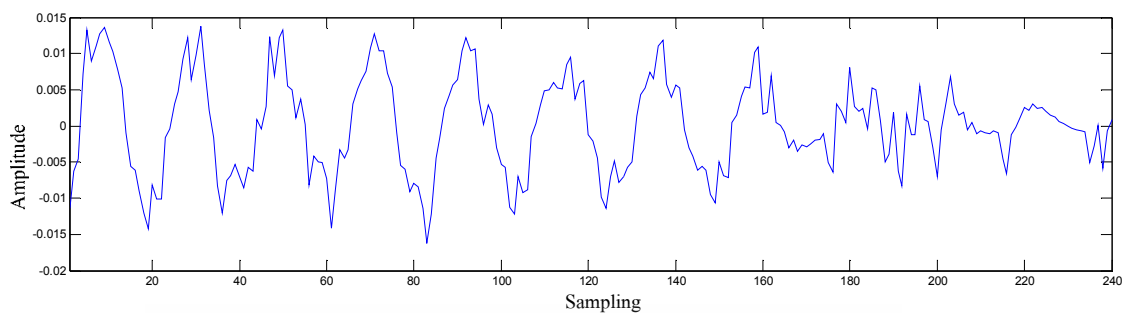


ค) Order 20

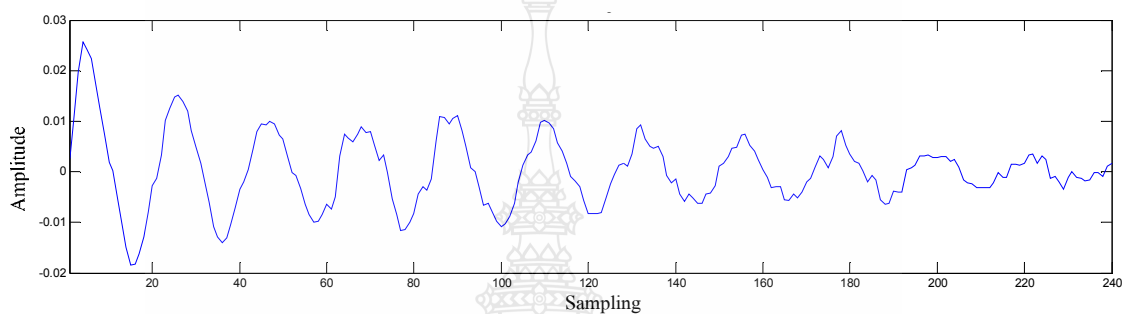


ง) Oder 40

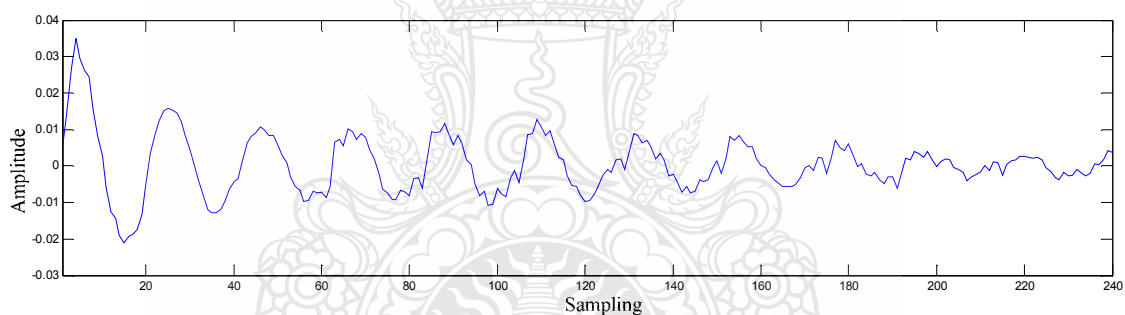
ภาพที่ 4.19 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้ชายที่ 5 วินาที ด้วย CELP ใน 1 เฟรม



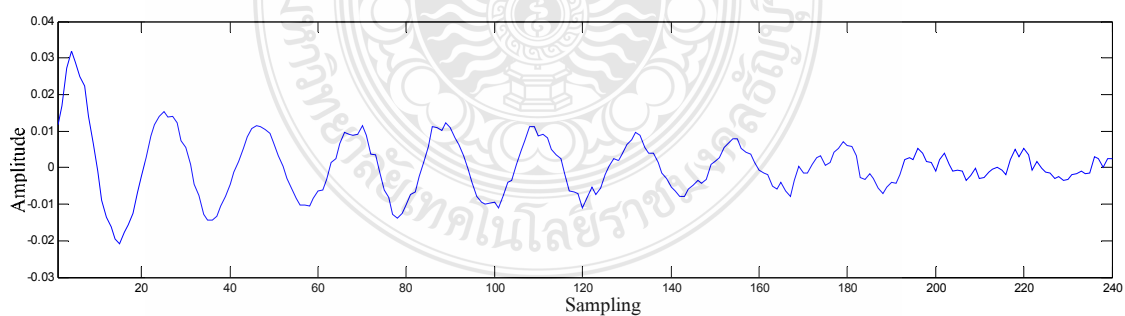
ก) สัญญาณเสียงต้นฉบับ



ข) Order 10

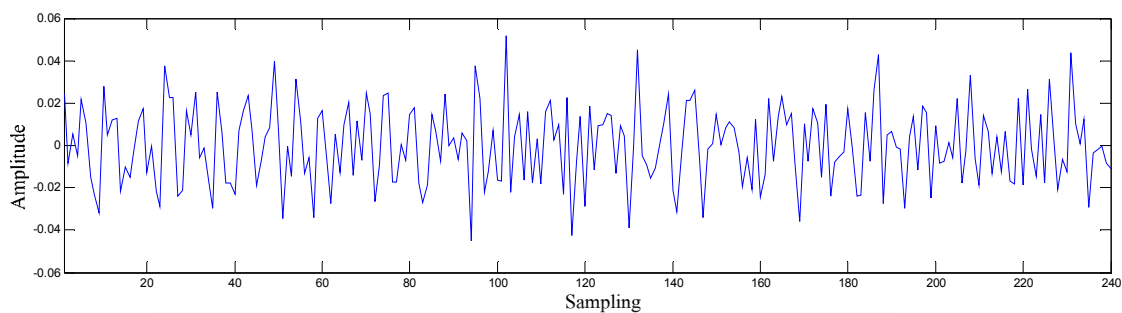


ค) Order 20

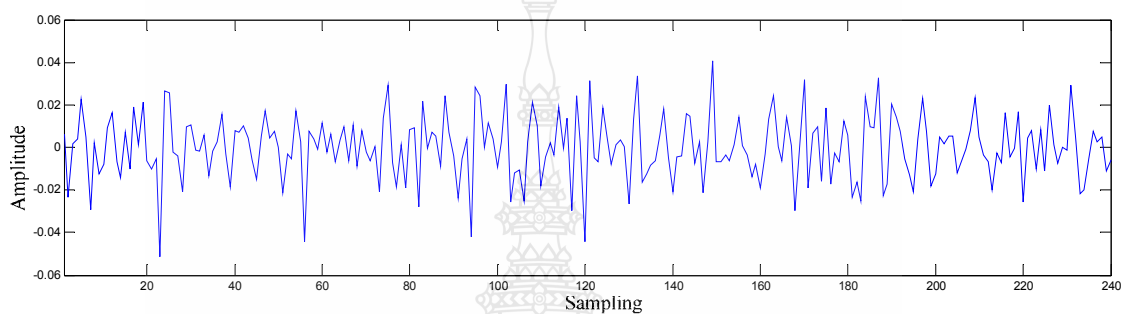


ง) Order 40

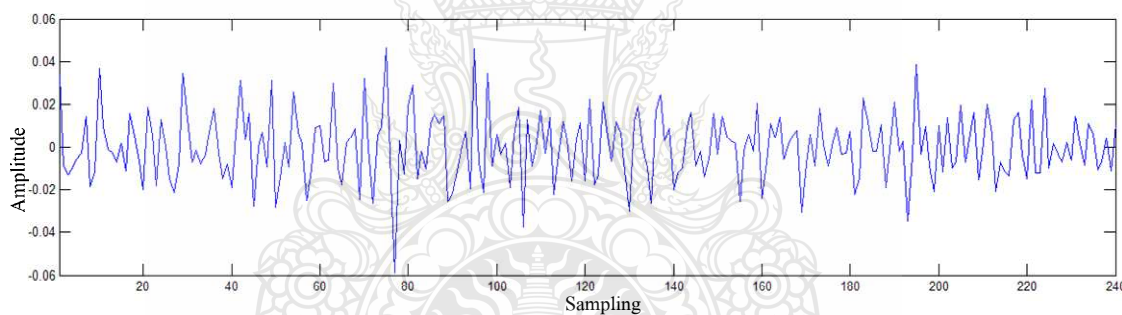
ภาพที่ 4.20 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้หญิงที่ 60 วินาที ด้วย CELP ใน 1 เฟรม



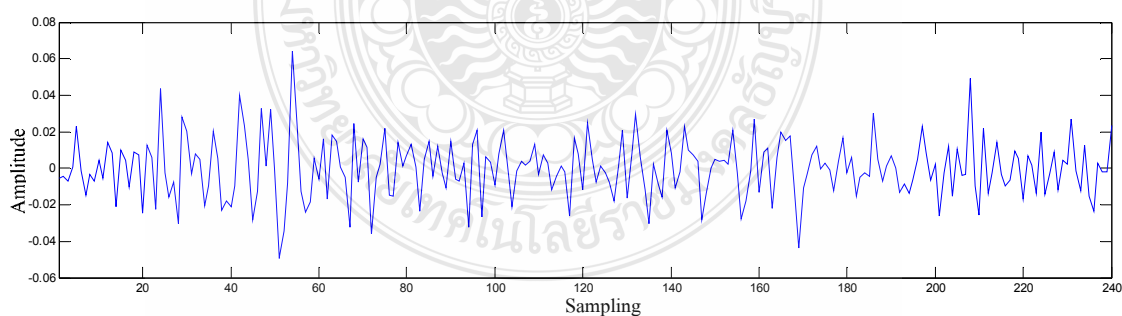
ก) สัญญาณเสียงต้นฉบับ



ข) Order 10



ค) Order 20



ง) Order 40

ภาพที่ 4.21 การแบ่งช่วงของการสังเคราะห์สัญญาณเสียงพูดผู้หญิงที่ 60 วินาที ด้วย CELP ใน 1 เฟรม



จากภาพที่ 4.18 – 4.21 เป็นการแบ่งสัญญาณเสียงหลังการสังเคราะห์ด้วย CELP ด้วย Order 10, 20 และ 40 ของเสียงผู้ชายและผู้หญิง ซึ่งผลปรากฏว่าทุกๆ Order ของแต่ละเสียงมีความแตกต่างกันเล็กน้อย เนื่องจาก CELP ที่ใช้สำหรับการทดลองนี้ได้ใช้ Codebook ตัวเดียวกันคือ 1024 ดังนั้น สรุปได้ว่าการเปลี่ยน Order ของ CELP ที่ใช้ Codebook ตัวเดียวกัน จะมีผลกระทบกับสัญญาณเสียงน้อยมาก

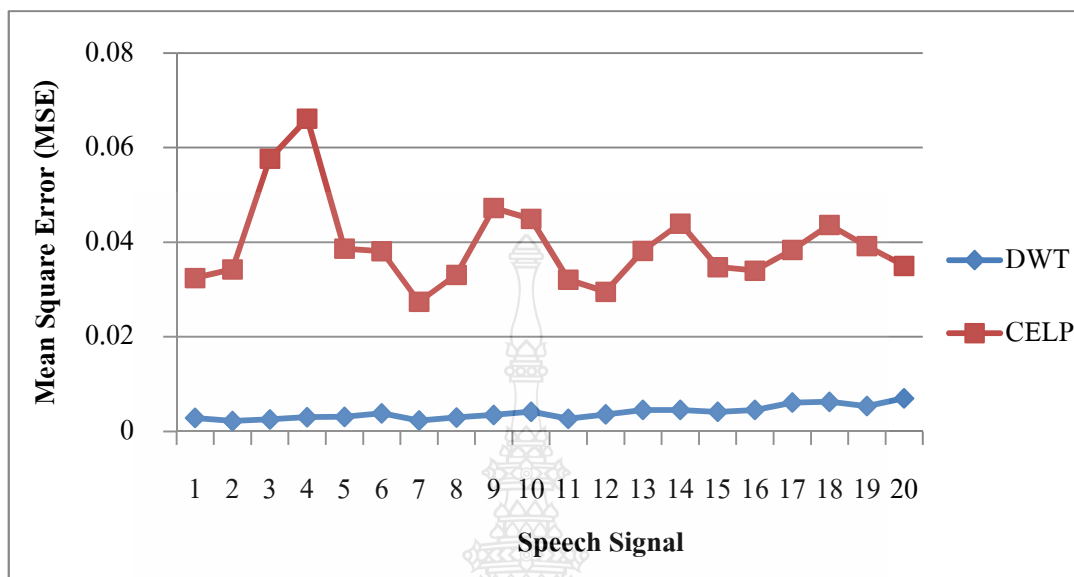
#### 4.4 การเปรียบเทียบคุณภาพของการบีบอัดระหว่าง DWT กับ CELP

ในงานวิจัยนี้ได้ทำการทดลอง 3 ระดับ โดยนำระดับของเวฟเล็ตไปเทียบกับจำนวนพารามิเตอร์หรือ Order ของ LSP โดยใช้หลักการแบ่งเฟรมให้แต่ละเฟรมมีค่าพารามิเตอร์ที่เท่ากันทั้งของการบีบอัดด้วย DWT และ CELP ซึ่งสามารถแบ่งได้ดังนี้

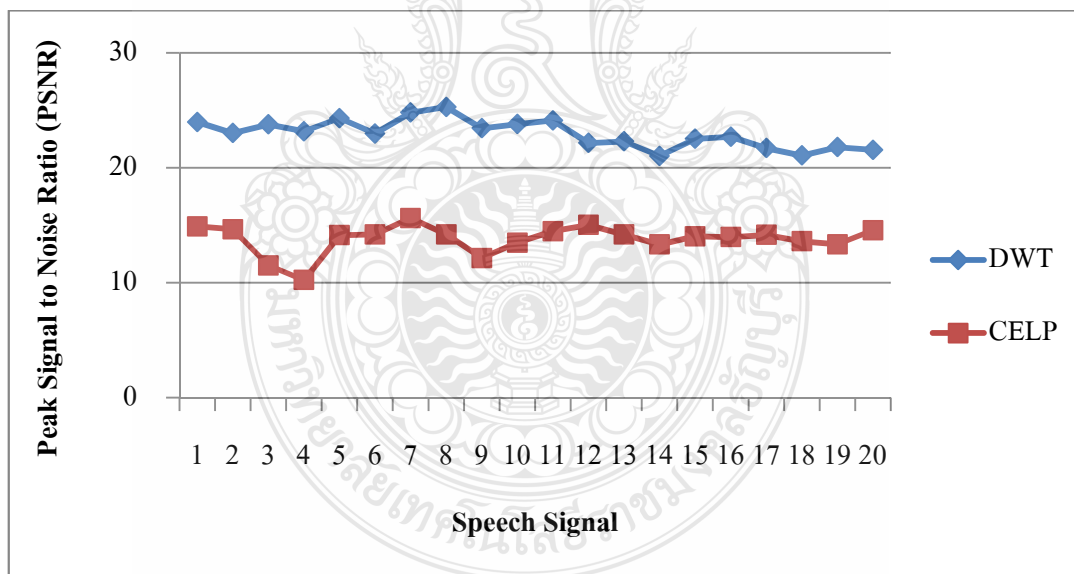
- ระดับที่ 1 ของ DWT เทียบกับ Order 40 ของ LPC
- ระดับที่ 2 ของ DWT เทียบกับ Order 20 ของ LPC
- ระดับที่ 3 ของ DWT เทียบกับ Order 10 ของ LPC

การบีบอัดด้วยเวฟเล็ตจะทำงานโดยกรองความถี่ออกไประดับละครึ่งหนึ่งของสัญญาณก่อนหน้า ดังนั้นในงานวิจัยนี้จึงได้นำ Order ของ CELP ที่เพิ่มขึ้น 2 เท่าของ Order ก่อนหน้า หลังจากนั้นจึงนำสัญญาณจากการคืนกลับสัญญาณเสียงทั้งสองเทคนิคมาเปรียบเทียบประสิทธิภาพของสัญญาณ โดยหลักการของ MSE และ PSNR

#### 4.4.1 การเปรียบเทียบค่า MSE และ PSNR ของการบีบอัดสัญญาณระดับที่ 3

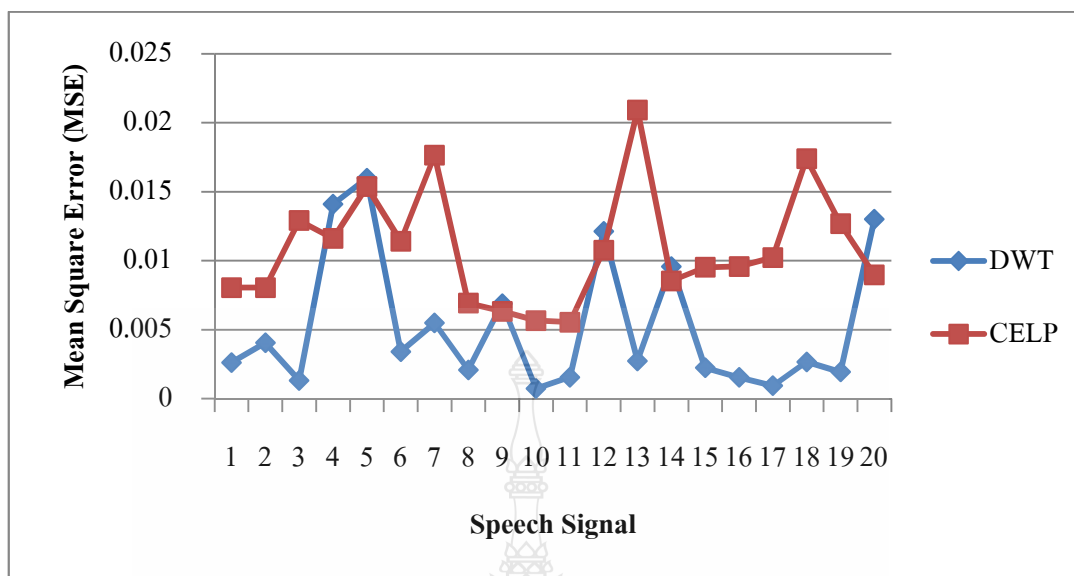


ก) ค่า MSE

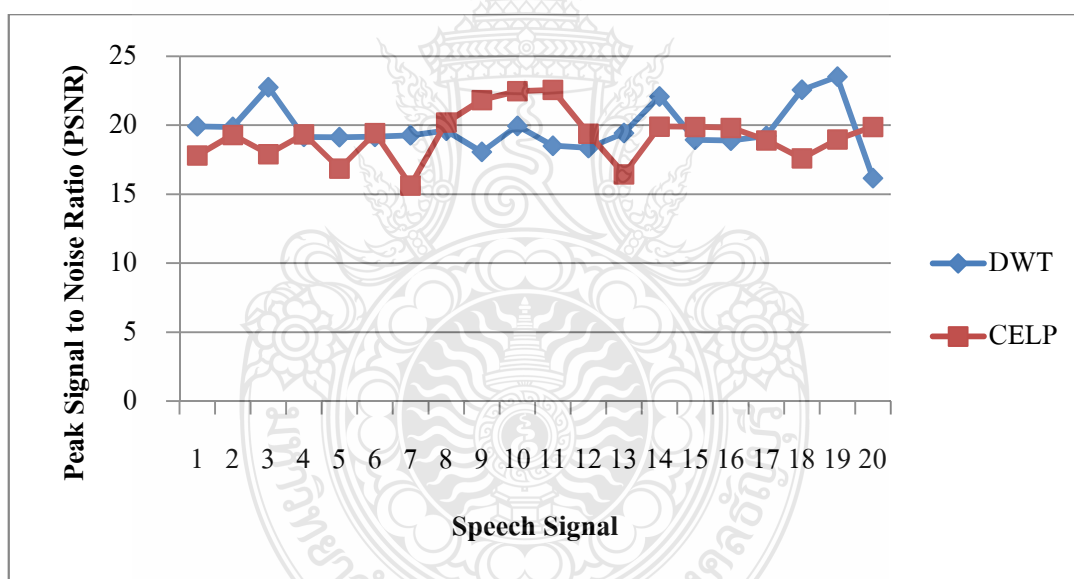


ข) ค่า PSNR

ภาพที่ 4.22 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้หญิงที่เวลา 5 วินาที

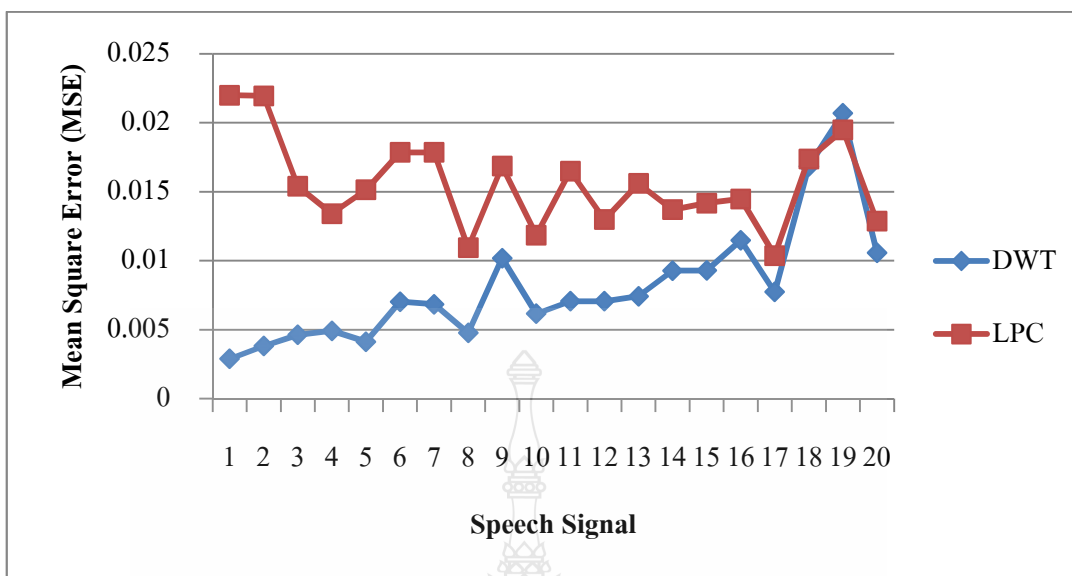


ก) ค่า MSE

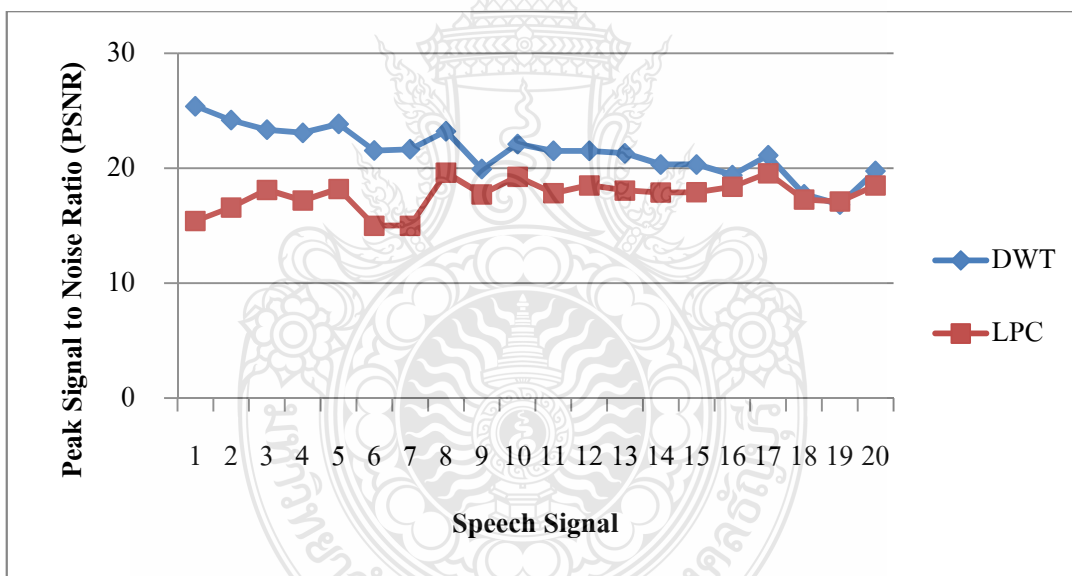


ข) ค่า PSNR

ภาพที่ 4.23 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้ชายที่เวลา 5 วินาที

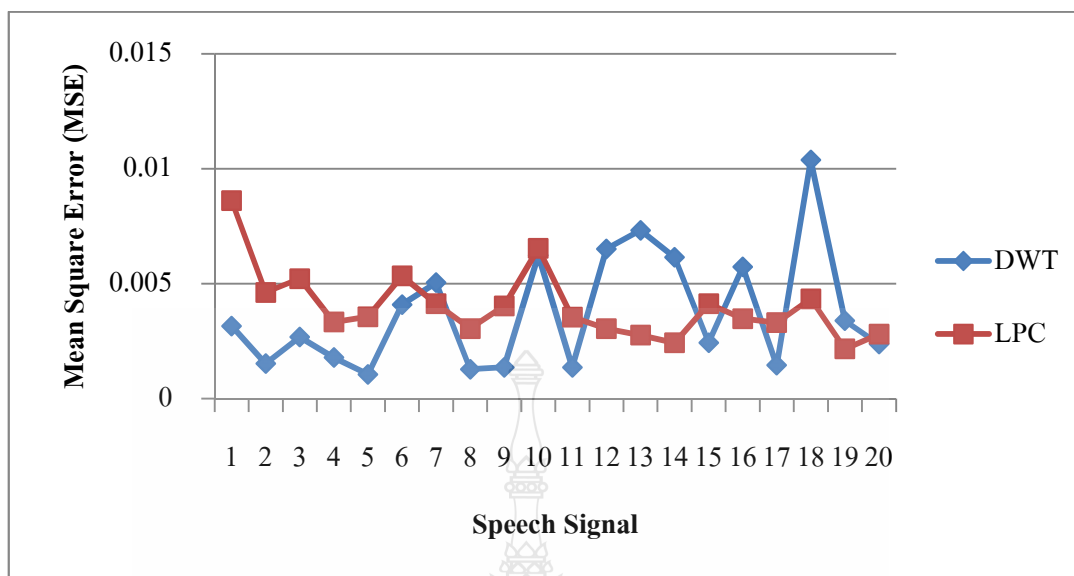


ก) ค่า MSE

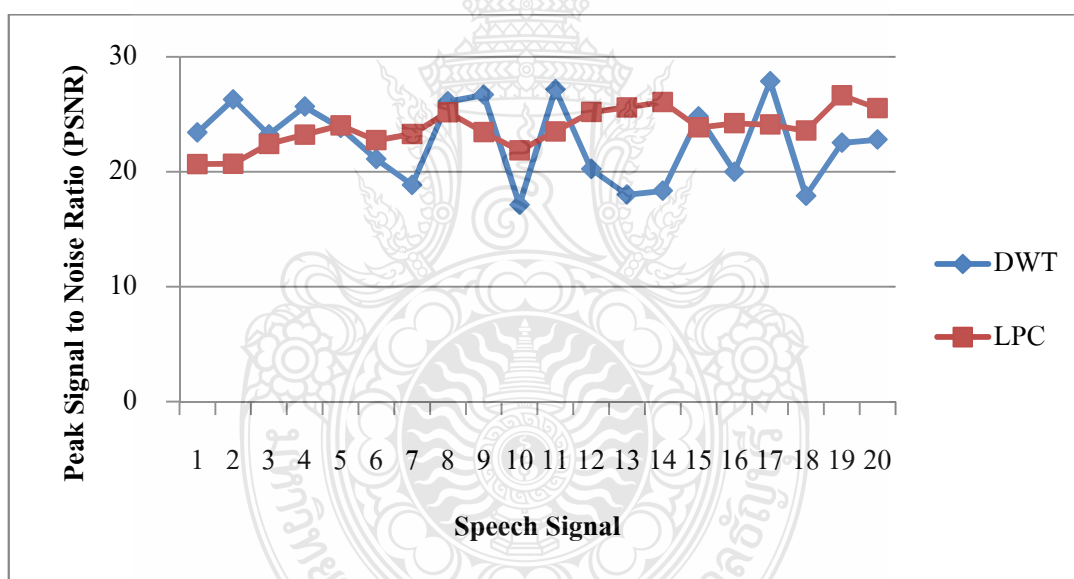


ข) ค่า PSNR

ภาพที่ 4.24 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้หญิงที่เวลา 60 วินาที



ก) ค่า MSE



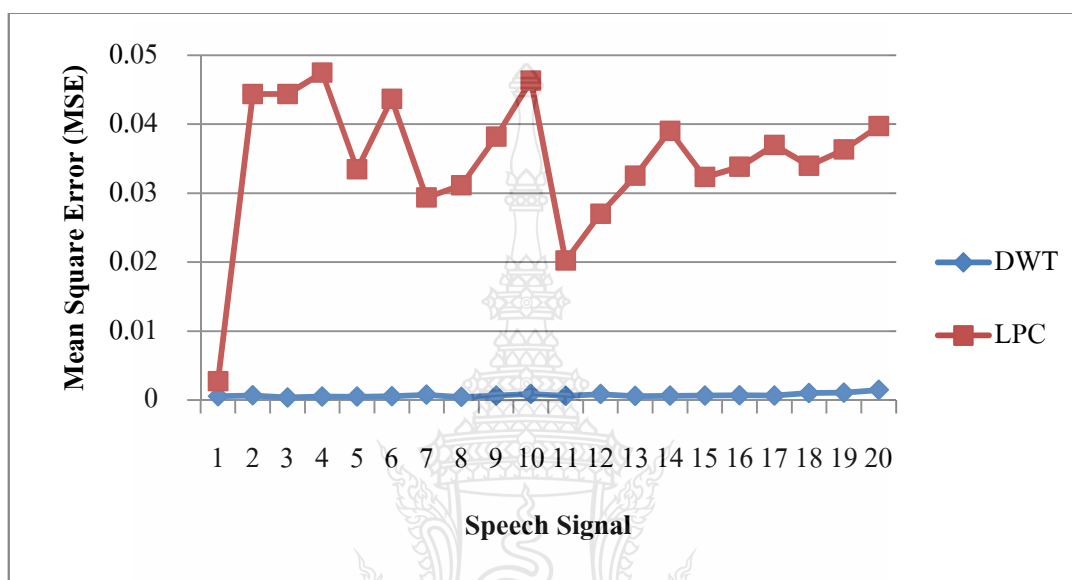
ข) ค่า PSNR

ภาพที่ 4.25 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 3 ของเสียงผู้ชายที่เวลา 60 วินาที

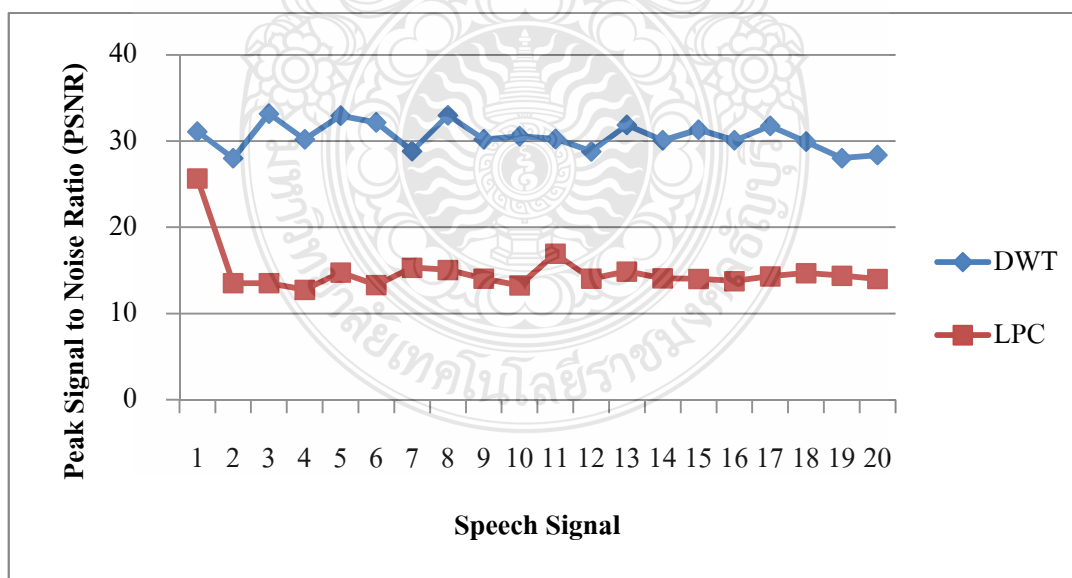
จากภาพที่ 4.22-4.25 เป็นการแสดงผลจากการหาค่า MSE และ PSNR ของเสียงพูดที่ผ่านการบีบอัดสัญญาณด้วย DWT ระดับที่ 3 และ CELP Order 10 จะสังเกตเห็นว่าการบีบอัดสัญญาณเสียง DWT มีค่า MSE น้อยกว่า CELP แสดงให้เห็นว่าค่าผิดพลาดที่เกิดขึ้นระหว่างการบีบอัดนั้น DWT มีค่าผิดพลาดน้อยมาก และในส่วนของประสิทธิภาพของสัญญาณเสียงที่ได้จากการหาค่า

PSNR พบการบีบอัดสัญญาณเสียงในระดับที่ 3 ด้วย DWT จะมีประสิทธิภาพของสัญญาณเสียงที่ดีกว่า CELP เฉพาะเสียงพูดของผู้หญิงเท่านั้น ส่วนสัญญาณเสียงของชายการบีบอัดด้วย CELP ที่ Order 10 จะมีประสิทธิภาพที่ดีกว่า

#### 4.4.2 การเปรียบเทียบค่า MSE และ PSNR ของการบีบอัดสัญญาณที่ระดับ 2

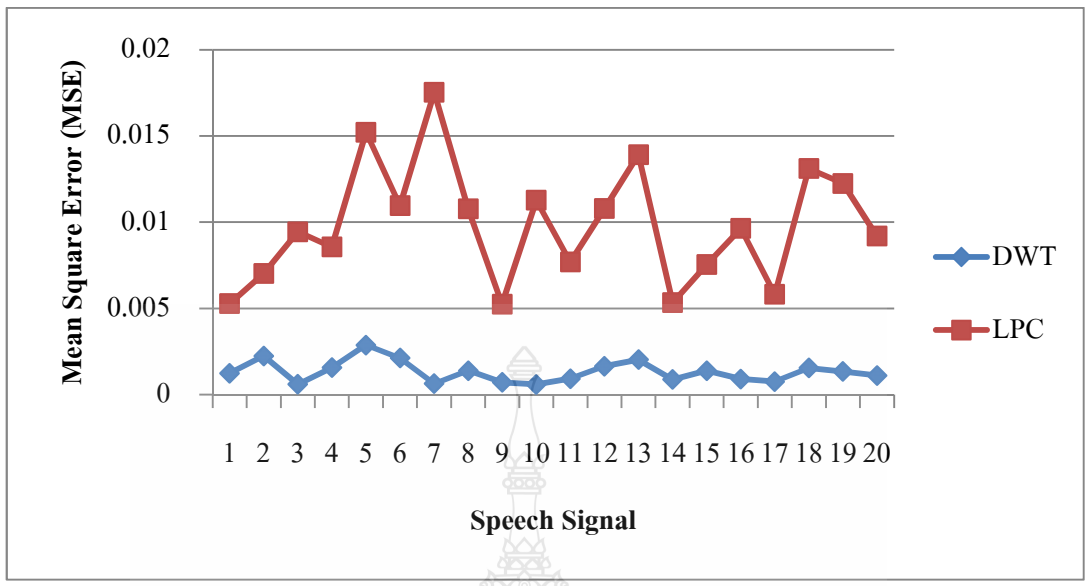


ก) ค่า MSE

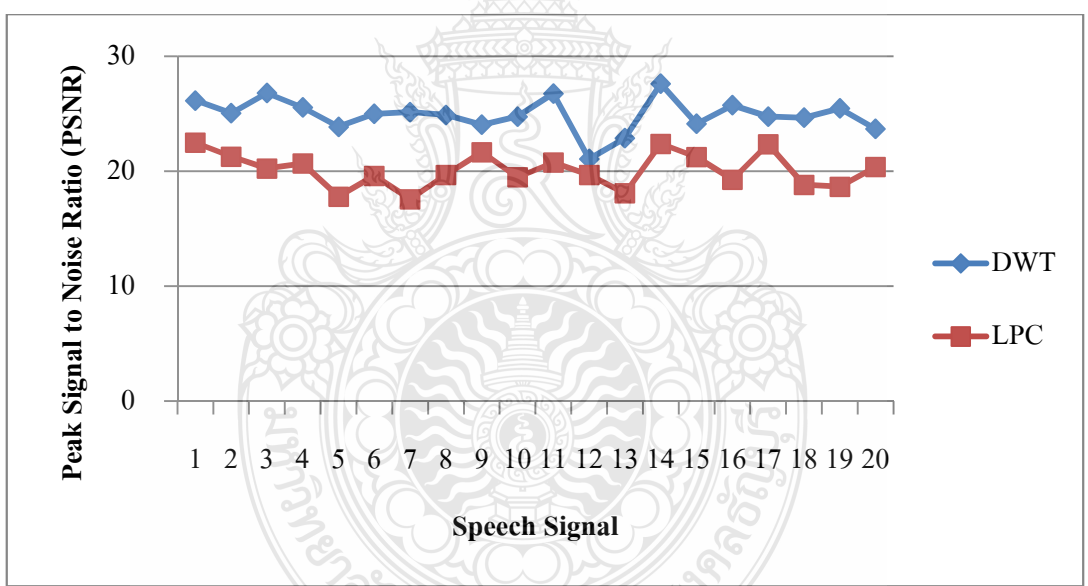


ข) ค่า PSNR

ภาพที่ 4.26 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้หญิงที่เวลา 5 วินาที

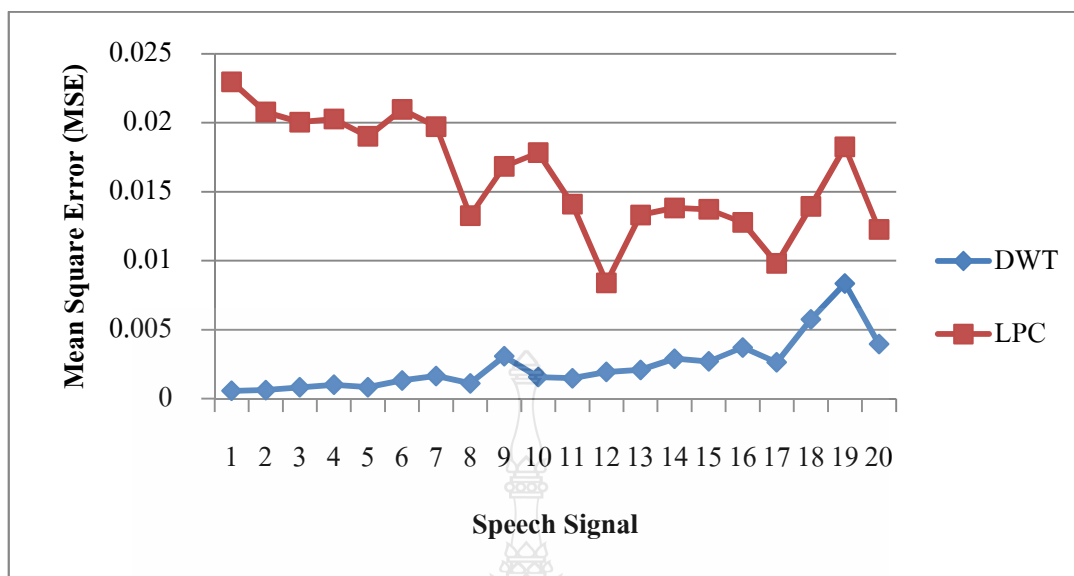


ก) ค่า MSE

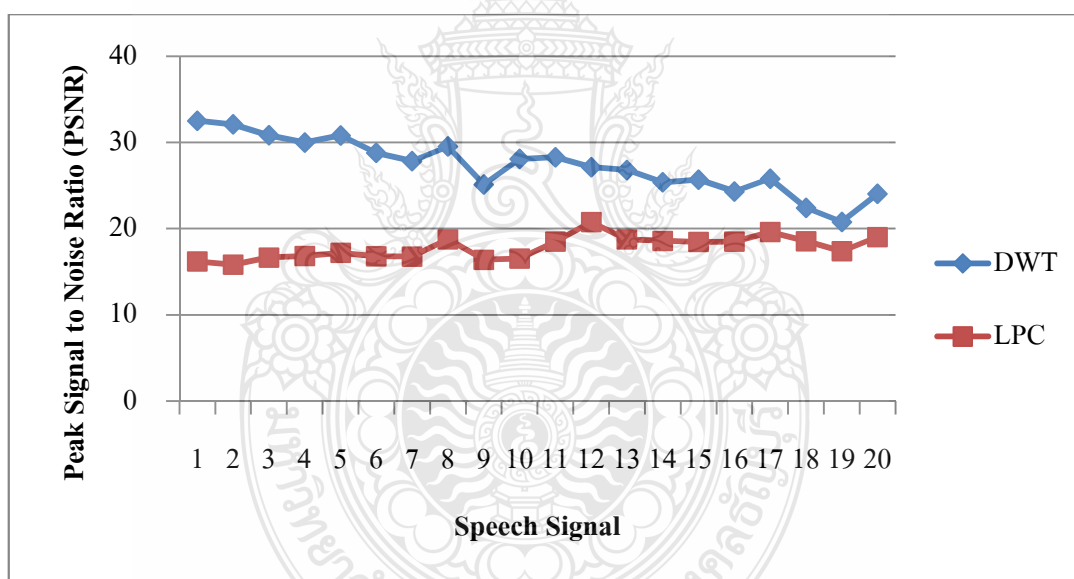


ข) ค่า PSNR

ภาพที่ 4.27 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้ชายที่เวลา 5 วินาที



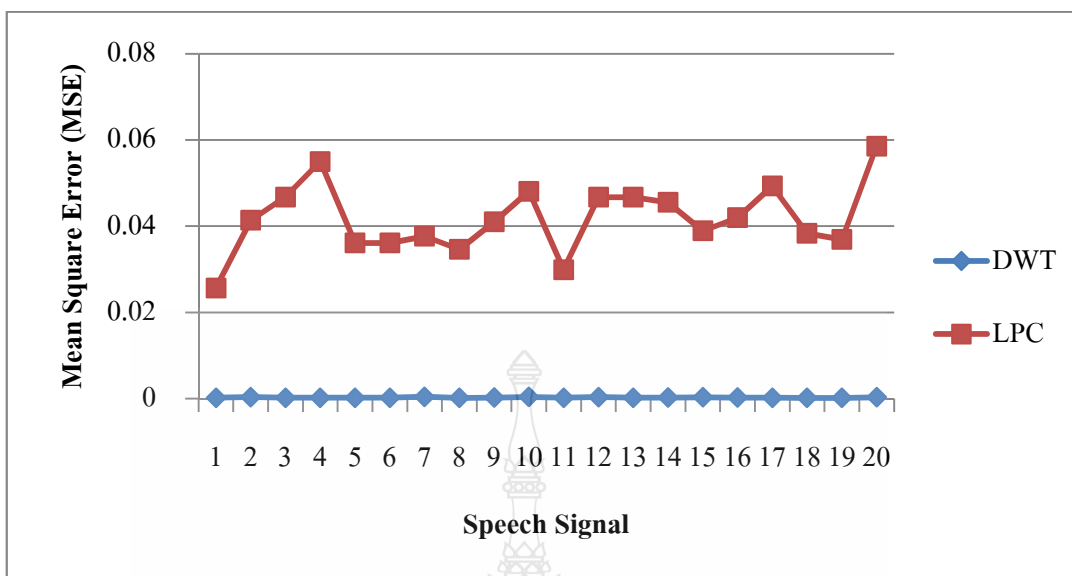
ก) ค่า MSE



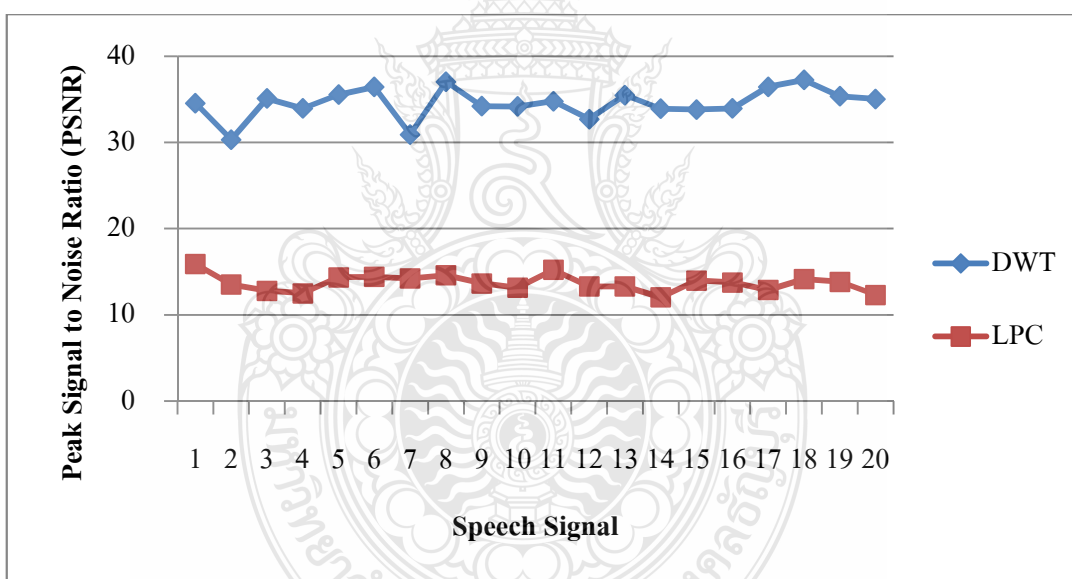
ข) ค่า PSNR

ภาพที่ 4.28 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้หญิงที่เวลา 60 วินาที





ก) ค่า MSE



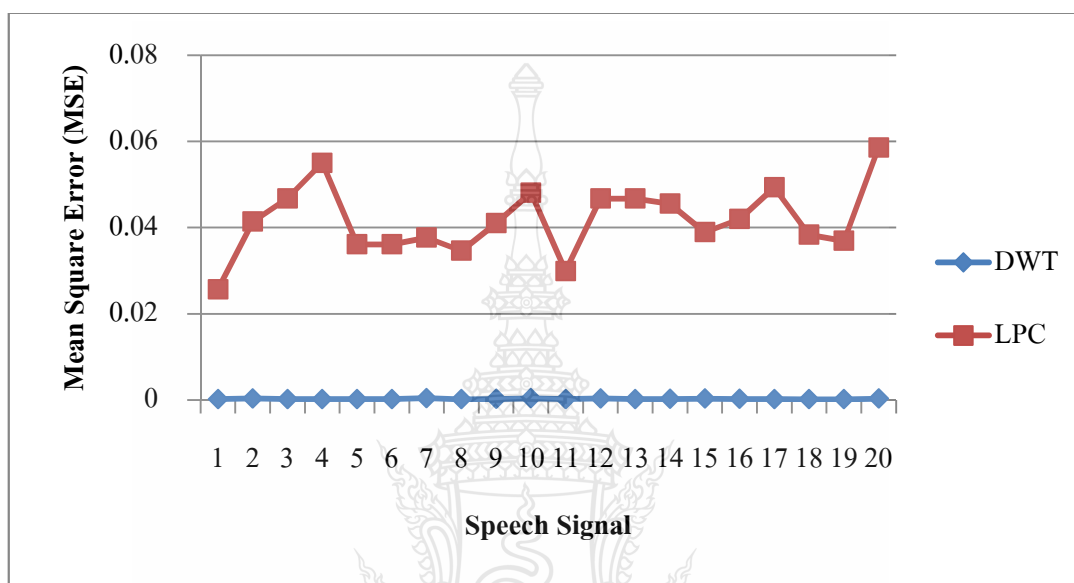
ข) ค่า PSNR

ภาพที่ 4.29 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 2 ของเสียงผู้ชายที่เวลา 60 วินาที

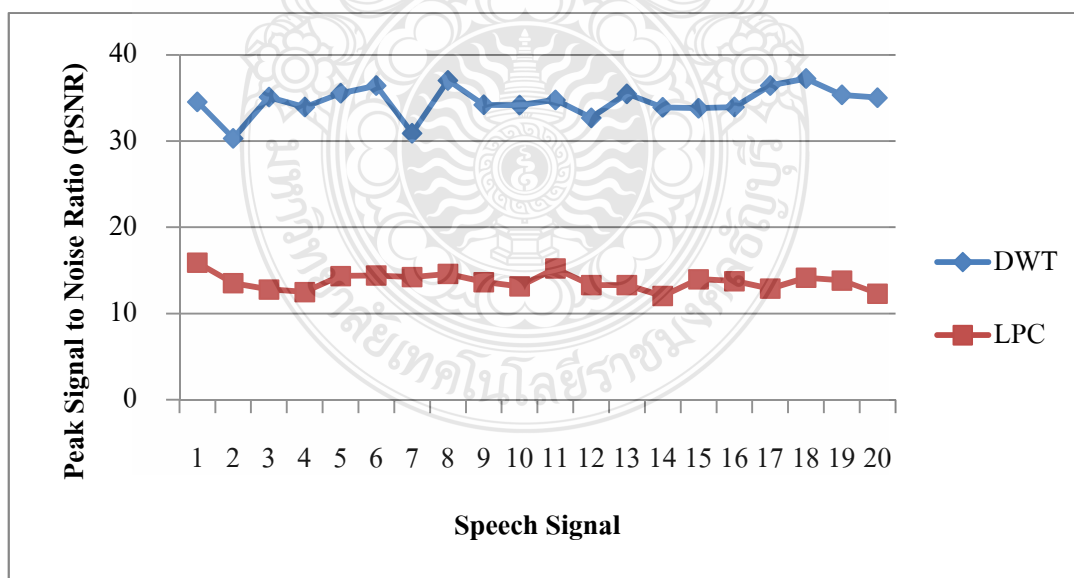
จากภาพที่ 4.26-4.29 เป็นการแสดงผลที่จากการหาค่า MSE และ PSNR ของเสียงพูดที่ผ่านการบีบอัดสัญญาณด้วย DWT ระดับที่ 2 และ CELP Order 20 จะสังเกตเห็นว่าการบีบอัดสัญญาณเสียง DWT มีค่า MSE น้อยกว่า CELP แสดงให้เห็นว่าค่าผิดพลาดที่เกิดขึ้นระหว่างการบีบอัดนั้น

DWT มีค่าผิดพลาดน้อยมาก และในส่วนของประสิทธิภาพของสัญญาณเสียงที่ได้จากการหาค่า PSNR พบว่าการบีบอัดสัญญาณเสียงในระดับที่ 2 ด้วย DWT จะมีประสิทธิภาพของสัญญาณเสียงที่ดีกว่า CELP ทั้งสัญญาณเสียงพูดของผู้หญิงและผู้ชาย

#### 4.4.3 การเปรียบเทียบค่า MSE และ PSNR ของการบีบอัดสัญญาณระดับที่ 1

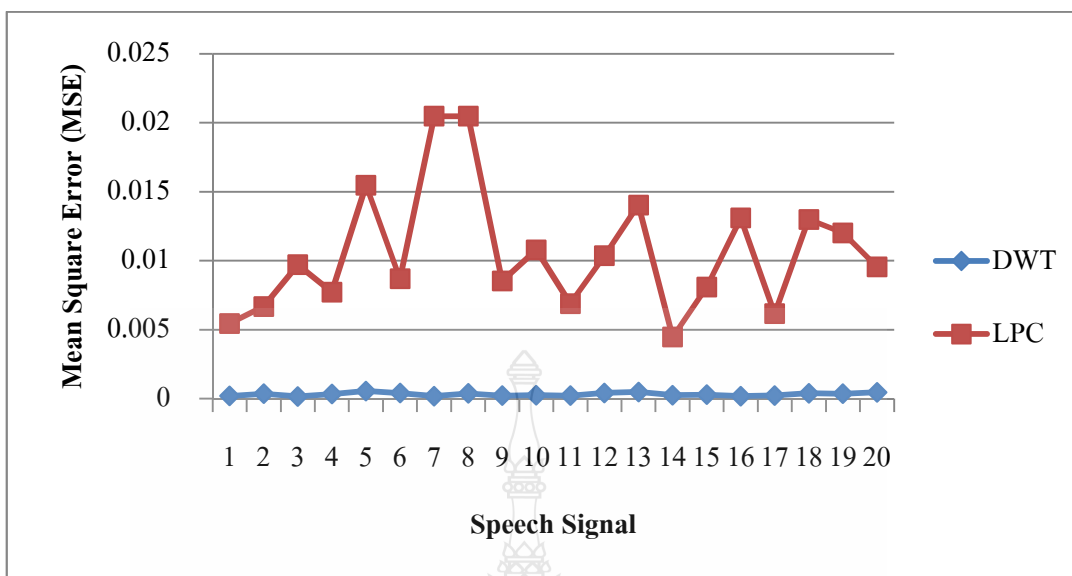


ก) ค่า MSE

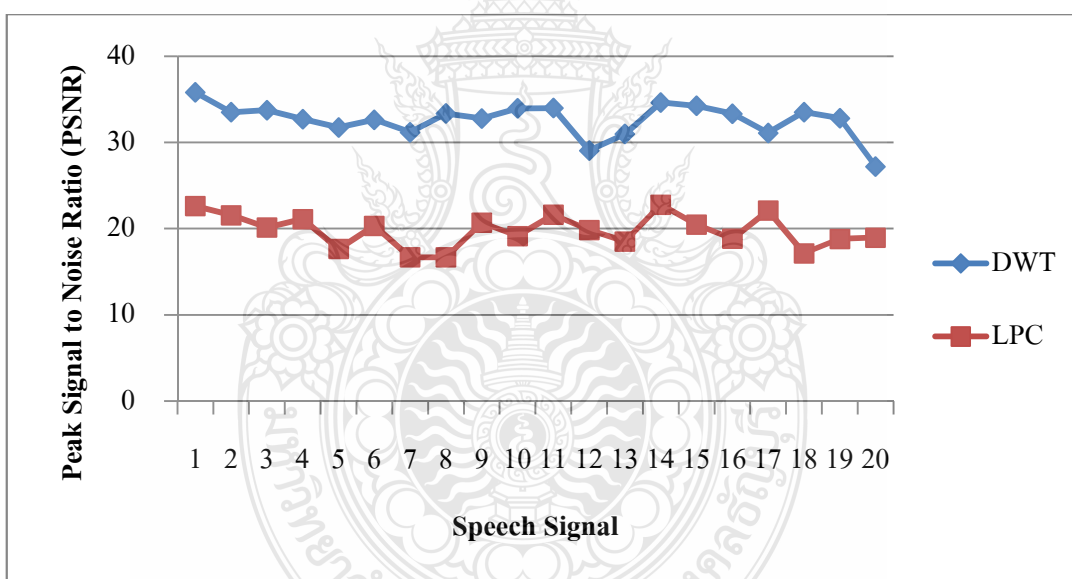


ข) ค่า PSNR

ภาพที่ 4.30 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้หญิงที่เวลา 5 วินาที

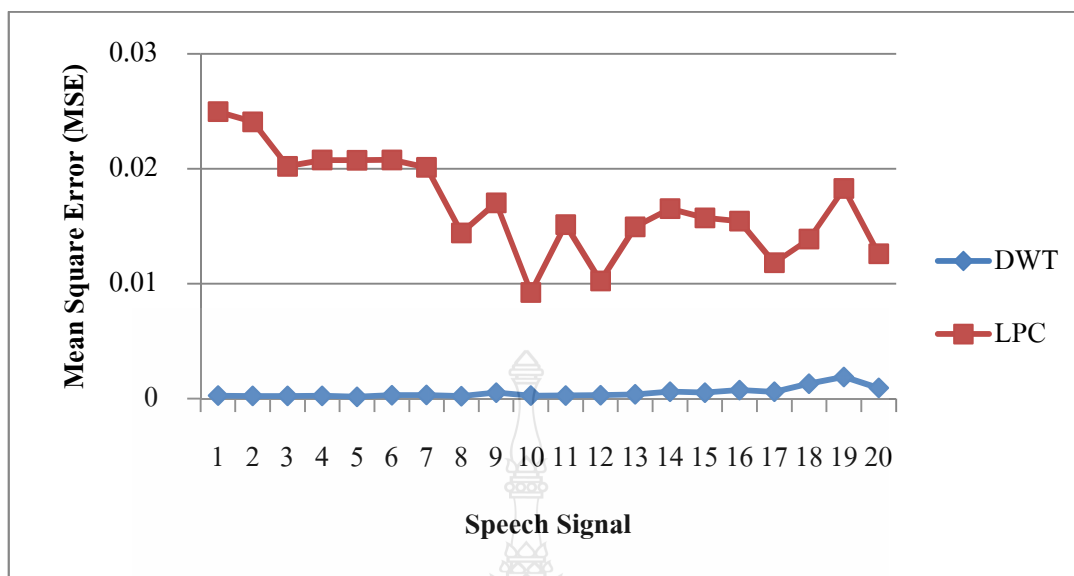


ก) ค่า MSE

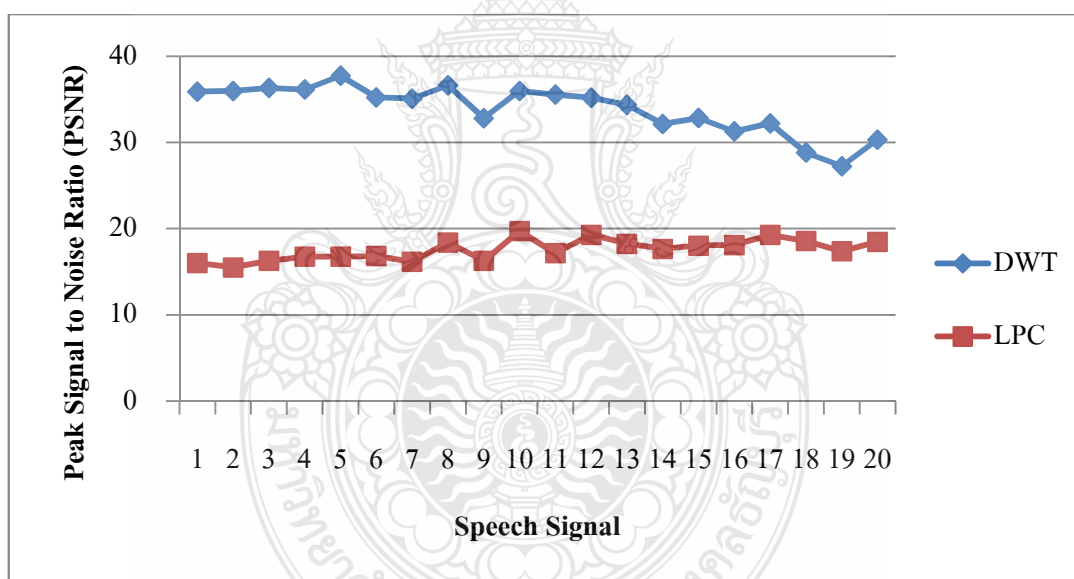


ข) ค่า PSNR

ภาพที่ 4.31 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้ชายที่เวลา 5 วินาที

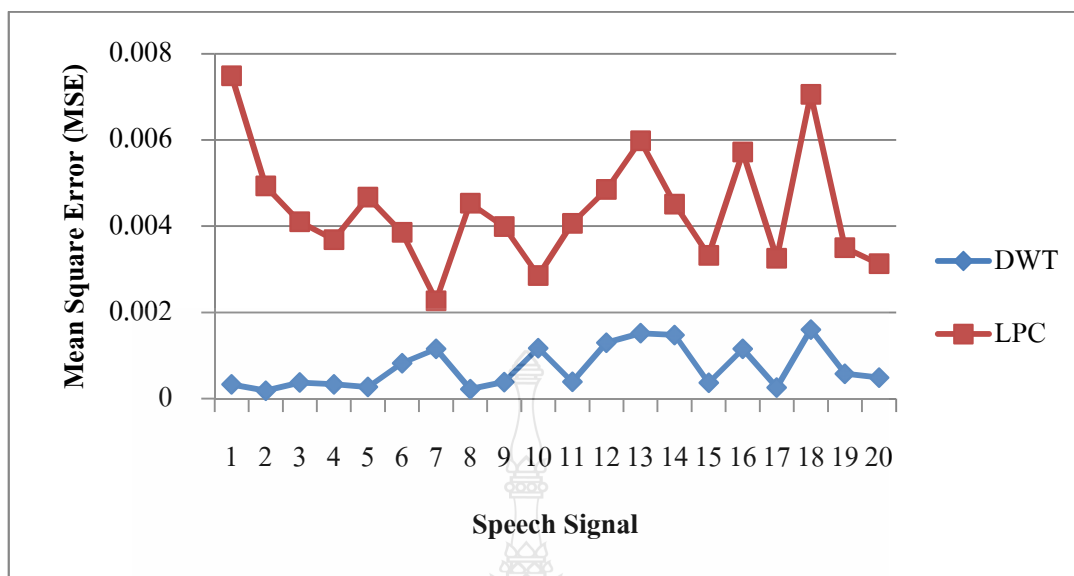


ก) ค่า MSE

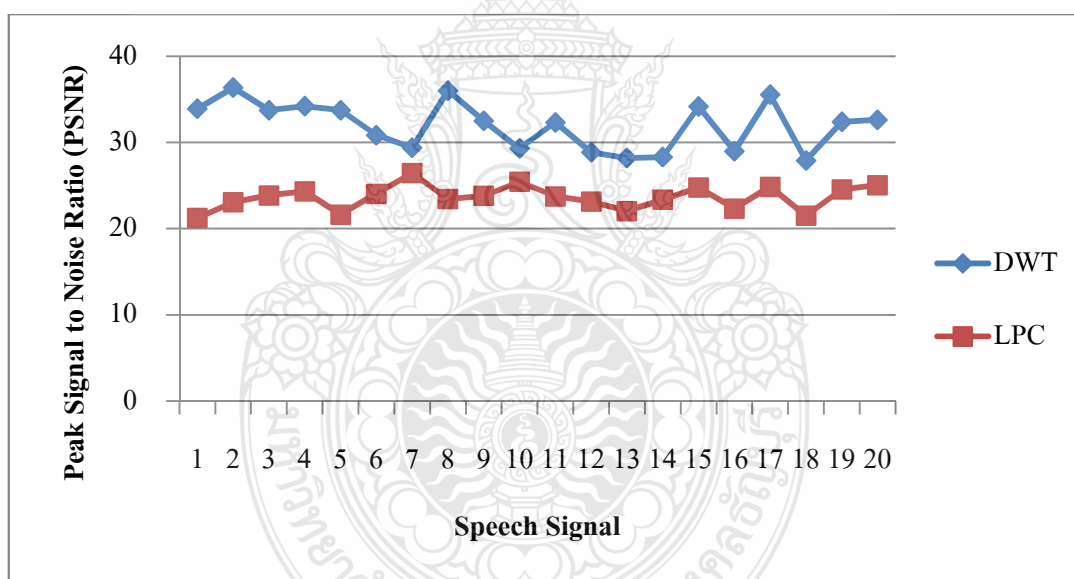


ข) ค่า PSNR

ภาพที่ 4.32 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้หญิงที่เวลา 60 วินาที



ก) ค่า MSE



ข) ค่า PSNR

ภาพที่ 4.33 ค่า MSE และ PSNR ที่ผ่านการบีบอัดระดับที่ 1 ของเสียงผู้ชายที่เวลา 60 วินาที

จากภาพที่ 4.30-4.33 เป็นการแสดงผลจากการหาค่า MSE และ PSNR ของเสียงพูดที่ผ่านการบีบอัดสัญญาณด้วย DWT ระดับที่ 1 และ CELP Order 40 จะสังเกตเห็นว่าการบีบอัดสัญญาณเสียง DWT มีค่า MSE น้อยกว่า CELP ซึ่งแสดงให้เห็นว่าค่าผิดพลาดที่เกิดขึ้นระหว่างการบีบอัดนั้น

DWT มีค่าผิดพลาดน้อยมาก และในส่วนของประสิทธิภาพของสัญญาณเสียงที่ได้จากการหาค่า PSNR พบว่าการบีบอัดสัญญาณเสียงในระดับที่ 1 ด้วย DWT จะมีประสิทธิภาพของสัญญาณเสียงที่ดีกว่า CELP และมีประสิทธิภาพสูงกว่าทุกระดับ

#### 4.5 สรุปผลการบีบอัดสัญญาณเสียงพูดด้วย DWT และ CELP

จากการบีบอัดสัญญาณเสียงทั้งหมด 80 เสียง โดยแบ่งเป็นเสียงของผู้หญิงและผู้ชายที่มีความยาว 5 วินาที และ 60 วินาที อย่างละ 20 เสียง เพื่อบีบอัดสัญญาณเสียงด้วย DWT ระดับที่ 1-3 และ CELP Order ที่ 10, 20 และ 40 โดยเปรียบเทียบผลการทดลองจากค่าความผิดพลาดของ MSE และการหาประสิทธิภาพของ PSNR ตามภาพที่ 4.22 – 4.33 แสดงให้เห็นว่าคุณภาพของสัญญาณเสียงที่ผ่านการบีบอัดด้วย DWT จะเห็นว่าคุณภาพของสัญญาณเสียงจะดีขึ้นเป็นลำดับ โดยเฉพาะอย่างยิ่งการบีบอัดด้วยระดับ 1 เมื่อเปรียบเทียบกับ การบีบอัดด้วย CELP ซึ่งสามารถสรุปเป็นค่าเฉลี่ยของเสียงแต่ละประเภทดังตารางที่ 4.1 และผลลัพธ์อื่นๆ จะแสดงไว้ในภาคผนวก ก

ตารางที่ 4.1 สรุปค่าเฉลี่ยของ MSE และ PSNR

ระดับการบีบอัด	เสียง	ค่าเฉลี่ย MSE		ค่าเฉลี่ย PSNR	
		DWT	CELP	DWT	CELP
ระดับที่ 3 หรือ Order 10	เสียงผู้หญิง 5 วินาที	0.0039684	0.0394481	22.964986	13.781483
	เสียงผู้ชาย 5 วินาที	0.00524721	0.01090414	19.7262602	19.2066053
	เสียงผู้หญิง 60 วินาที	0.0081388	0.015538	21.404765	17.657204
	เสียงผู้ชาย 60 วินาที	0.0037642	0.0040201	22.588559	23.782297
ระดับที่ 2 หรือ Order 20	เสียงผู้หญิง 5 วินาที	0.000699	0.0346344	30.536701	14.82256
	เสียงผู้ชาย 5 วินาที	0.001325	0.0098298	24.893672	20.100045
	เสียงผู้หญิง 60 วินาที	0.0024015	0.0161028	27.309833	17.807994
	เสียงผู้ชาย 60 วินาที	0.0025496	0.0045219	26.265939	23.250548
ระดับที่ 1 หรือ Order 40	เสียงผู้หญิง 5 วินาที	0.0002561	0.041775	34.548318	13.704442
	เสียงผู้ชาย 5 วินาที	0.0003154	0.0105815	32.612891	19.78766
	เสียงผู้หญิง 60 วินาที	0.0005144	0.0168456	33.886936	17.548176
	เสียงผู้ชาย 60 วินาที	0.0007185	0.00439	31.965845	23.639487

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงสรุปผลการวิจัยการบีบอัดสัญญาณเสียงพูดระหว่าง DWT กับ CELP และข้อเสนอแนะสำหรับแนวทางการพัฒนางานวิจัยต่อไป

#### 5.1 สรุปผลการวิจัยการบีบอัดสัญญาณเสียงพูด

วิทยานิพนธ์ฉบับนี้เป็นการศึกษาและเปรียบเทียบเวฟเล็ตฟิวดอร์สำหรับการบีบอัดสัญญาณเสียงพูด โดยใช้เทคนิคการบีบอัดสัญญาณด้วย DWT และ CELP เพื่อเปรียบเทียบคุณภาพหลังการคืนกลับสัญญาณ ซึ่งปกติแล้วในกระบวนการส่งสัญญาณเสียงที่มีการบีบอัดสัญญาณเสียงนั้น เพื่อต้องการลดขนาดของสัญญาณเสียงพูดจากสัญญาณเสียงต้นฉบับ และต้องการให้สัญญาณเสียงพูดมีคุณสมบัติใกล้เคียงกับสัญญาณเสียงพูดต้นฉบับมากที่สุดหรือขึ้นอยู่กับการนำเสียงนั้นๆ ไปประยุกต์ใช้งาน ซึ่งสัญญาณเสียงพูดที่นำมาวิเคราะห์ในงานวิจัยนี้ได้แก่ เสียงของผู้หญิงและผู้ชายที่มีความยาว 5 วินาที และ 60 วินาที

ผลการวิจัยพบว่าเวฟเล็ตที่เหมาะสมในงานวิจัยนี้และผ่านการคัดเลือก คือ เวฟเล็ต Biorthogonal เนื่องจากให้ประสิทธิภาพในการบีบอัดสัญญาณที่สูงกว่าเวฟเล็ตทั้ง 2 ตระกูล ดังนั้นจึงนำเวฟเล็ต Biorthogonal เข้าสู่กระบวนการบีบอัดสัญญาณเสียงพูดโดยใช้เทคนิคของ DWT ระดับที่ 1 – 3 และคืนกลับสัญญาณเสียงด้วย IDWT จากผลการทดลองสามารถวิเคราะห์ได้ว่าเสียงพูดที่มีความยาว 5 วินาที จะมีประสิทธิภาพในการคืนกลับสัญญาณมากกว่าเสียงพูดที่มีความยาว 60 วินาที โดยเฉพาะสัญญาณเสียงพูดของผู้หญิง

งานวิจัยนี้ได้นำสัญญาณเสียงพูดหลังจากการคืนกลับสัญญาณเสียงไปเปรียบเทียบคุณภาพของสัญญาณเสียงกับเทคนิคการบีบอัดสัญญาณของ CELP ที่ Order 10, 20 และ 40 เนื่องจากเทคนิคการบีบอัดสัญญาณเสียงของ CELP นั้นเป็นมาตรฐานการบีบอัดที่นิยมใช้กันมากในปัจจุบัน ซึ่งหลักการที่ใช้สำหรับเปรียบเทียบประสิทธิภาพของสัญญาณระหว่างการบีบอัดสัญญาณเสียงพูดด้วย DWT กับ CELP นั้น จะใช้ค่า MSE และ PSNR มาวิเคราะห์ผล สามารถสรุปตามคุณสมบัติของสัญญาณเสียงที่นำมาใช้ได้ดังตารางที่ 5.1

ตารางที่ 5.1 ประสิทธิภาพของสัญญาณเสียง PSNR และค่าผิดพลาด MSE

ระดับการบีบอัด	เสียงผู้หญิง 5 วินาที		เสียงผู้ชาย 5 วินาที		เสียงผู้หญิง 60 วินาที		เสียงผู้ชาย 60 วินาที	
	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR
ระดับที่ 3 หรือ Oder 10	100%	100%	70%	50%	95%	95%	65%	45%
ระดับที่ 2 หรือ Oder 20	100%	100%	100%	100%	100%	100%	90%	90%
ระดับที่ 1 หรือ Oder 40	100%	100%	100	100%	100%	100%	100%	100%

จากตารางที่ 5.1 เป็นผลสรุปของประสิทธิภาพการบีบอัดสัญญาณเสียง ตัวอย่างเช่น การบีบอัดระดับที่ 3 ผลปรากฏว่าจาก 20 เสียง ของเสียงผู้หญิง 5 วินาที ซึ่งมีค่าความผิดพลาด MSE ของสัญญาณเสียงที่บีบอัดด้วย DWT น้อยกว่าการบีบอัดสัญญาณเสียงด้วย CELP และคิดเป็น 100 เปอร์เซ็นต์ ซึ่งหมายความว่าใน 20 สัญญาณเสียงผลเป็น DWT ดีกว่า CELP ทั้ง 20 เสียง เช่นเดียวกันกับการหาประสิทธิภาพของสัญญาณเสียงด้วยค่า PSNR

ดังนั้นจากตารางที่ 5.1 ได้แสดงค่า PSNR และค่า MSE ของการบีบอัดสัญญาณเสียงพูดทั้ง 3 ระดับ จาก 80 สัญญาณเสียงมาเปรียบเทียบระหว่างการบีบอัดสัญญาณเสียงด้วย DWT และ CELP ซึ่งผลปรากฏว่าการบีบอัดของ DWT ให้ประสิทธิภาพในการบีบอัดสัญญาณเสียงที่ดีกว่า CELP และทั้งนี้ยังมีค่าผิดพลาดที่น้อยกว่าเมื่อนำสัญญาณไปเปรียบเทียบกับสัญญาณต้นฉบับ ซึ่งจากงานวิจัยนี้สามารถลดขนาดของสัญญาณเสียงได้มากกว่า 15 เปอร์เซ็นต์ เมื่อเทียบกับสัญญาณเสียงต้นฉบับ

## 5.2 ข้อเสนอแนะและแนวทางในการพัฒนา

5.2.1 ควรมีการปรับปรุงสัญญาณเสียงพูดหลังจากการบีบอัดสัญญาณ เช่น การใช้ตัวกรองคาลมาน (Kalman Filter) เพื่อเพิ่มคุณภาพของสัญญาณให้ดีขึ้น

5.2.2 ควรนำสัญญาณเสียงพูดไปทดลองกับการบีบอัดสัญญาณเทคนิคอื่นๆ เช่น MFCC โดยเปรียบเทียบกับเวฟเลต เพื่อหาว่าวิธีใดสามารถบีบอัดสัญญาณเสียงให้มีประสิทธิภาพที่ดีที่สุดในงานด้านการประมวลผลสัญญาณเสียงพูด

5.2.3 ควรนำหลักการบีบอัดเสียงพูดโดยการแปลงเวฟเลตไปประยุกต์ใช้ในตัวประมวลผลสัญญาณ เช่น บอร์ดเอฟพีจีเอ (FPGA) บอร์ด ARM หรือบอร์ดที่ใช้สำหรับการประมวลผลสัญญาณดิจิทัล (DSP) เพื่อเป็นการทดสอบการอุปกรณ์จริงและเป็นการประมวลผลแบบตามเวลาจริง (Real Time)



## รายการอ้างอิง

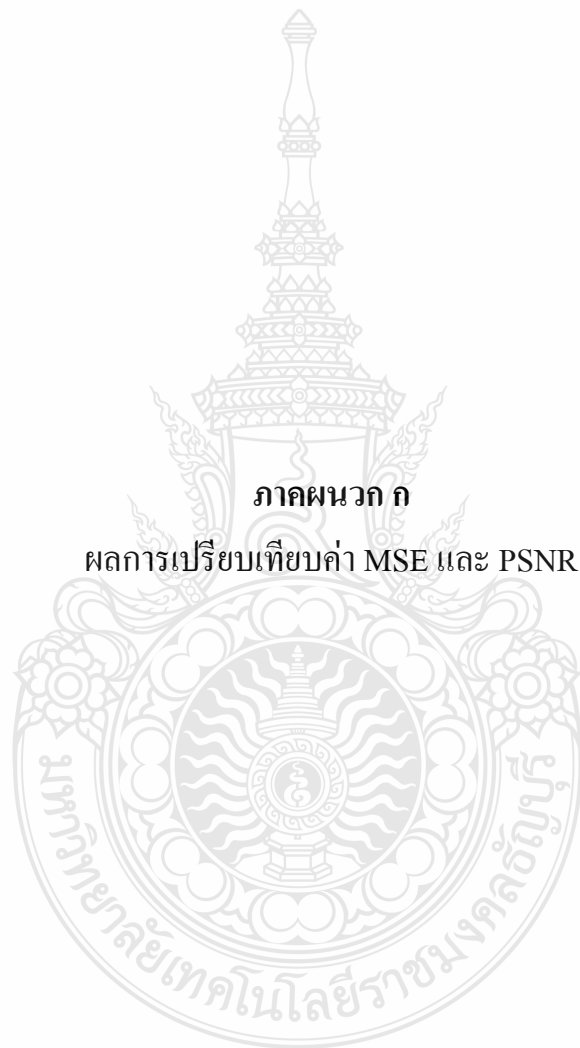
- [1] เอกชัย เนาวนิช. โปรแกรมฝึกออกเสียงพยัญชนะไทยสำหรับผู้บกพร่องทางการได้ยินโดยใช้  
โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเทคโนโลยีพระ  
จอมเกล้าพระนครเหนือ, 2551.
- [2] นลินี ชนะมุล และ เบญญาดา อุณหเลขจิตฺร. ซอฟต์แวร์สำหรับวิเคราะห์การออกเสียงภาษาไทย  
ด้วยการประมวลผลสัญญาณเสียงพูด. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัย  
บูรพา, 2551.
- [3] สุภัทรชัย ชมพันธุ์. การบีบอัดเสียงพูดภาษาไทยโดยใช้การเข้ารหัส MP-CELP. วิทยานิพนธ์  
ปริญญาโทบริหารธุรกิจ. จุฬาลงกรณ์มหาวิทยาลัย, 2543.
- [4] B.Dennis. **Details to assist in implementation of Federal Standard 1016 CELP.** National  
Communication Bulletin 92-1. The Manager National Communications system, 1992.
- [5] 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.
- [6] สุทธิ ทับทองดี. การพัฒนาเทคนิคสำหรับการบีบอัดสัญญาณเสียงพูดภาษาไทย ในมาตรฐาน  
**LPC-10.** วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2552.
- [7] Osman M.A. and Hussein, N. "Speech compression using LPC and wavelet," **2nd  
International Conference on Computer Engineering and Technology (ICCET)**,  
April 2010, Chengdu China, 2010. pp. V7-92-V7-99
- [8] จักรกฤษ อ่อนชื่นจิตฺร. การวิเคราะห์แนวทางเดินเสียงพูดในรูปแบบของสัมประสิทธิ์คู่เส้น  
สเปกตรัมร่วมกับ **Double Clustering.** วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัย  
เทคโนโลยีราชมงคลธัญบุรี, 2552.
- [9] วีระยุทธ คุณรัตน์ศิริ. การเพิ่มประสิทธิภาพการบีบอัดสัญญาณเสียงด้วยเทคนิคเวฟเล็ตร่วมกับ  
**ไวเนอร์ฟิลเตอร์.** วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเทคโนโลยีราชมงคล  
ธัญบุรี, 2554.
- [10] นพรัตน์ ศรียิ่งยงค์. การสร้างลายน้ำสำหรับสัญญาณเสียงดิจิทัลโดยใช้การแปลงเวฟเล็ตและ  
การค้นหาแบบตามูเชิงปรับตัว. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ. มหาวิทยาลัยเทคโนโลยี  
สุรนารี, 2548

- [11] ภาณี น้อยยิ่ง. การสร้างแบบจำลองความสัมพันธ์ทางเวลาระหว่างสัญญาณ ECG และ PCG โดยใช้โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, 2549.
- [12] รศ.ดร. มนัส สังวรศิลป์ และ วรรัตน์ ภัทรอมรกุล. คู่มือการใช้งาน MATLAB ฉบับสมบูรณ์, พิมพ์ครั้งที่ 1 สำนักพิมพ์อินโฟเพรส, กรุงเทพฯ 2543
- [13] ปิยสวัสดิ์ นวรัตน์ ณ อยุธา. การดึงคุณลักษณะและขนาดข้อมูลโดยอาศัยแบบจำลองของระบบโสตรับเสียงและเวฟเลตสำหรับการรู้จำเสียงภาษาไทย. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี, 2542.
- [14] อรรถจักร จิตต์โสภักดิ์. **Digital Image Processing**, พิมพ์ครั้งที่ 1, บริษัท สงวนกิจ พรินท์ แอนด์ มีเดีย, 2552: หน้า 200-222.
- [15] Shijo M Joseph and Babu Anto P, “Speech Compression Using Wavelet Transform,” **IEEE-International Conference on Recent Trends in Information Technology (ICRTIT 2011)**, MIT, Anna University, Chennai. June 3-5, 2011, pp. 754-758.
- [16] ZHAO Dan and MA Sheng-qian, “Speech Compression With Best Wavelet Packet Transform And SPIHT Algorithm,” **2010 Second International Conference on Computer Modeling and Simulation**, pp. 360-363.
- [17] Shivesh Ranjan, “A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition,” **International Conference on Signal Acquisition and Processing (ICSAP 2010)**, Feb 2010, Bangalore India, 2010: pp. 345-348
- [18] Yan Long Liu and Gang Guo Jun, “Selection of The Best Wavelet Base for Speech Signal,” **Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing**, Oct 2004, Beijing China, 2004, pp. 218-221.
- [19] Noman Junejo, Ahmed, N. Unar, M.A. Rajput and A.Q.K. “Speech and Image Compression Using Discrete Wavelet Transform,” **IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication**, April 2005, Princeton NJ, 2005: pp. 45-48.
- [20] Abdul Mawla M A Najih, et al., “Speech Compression using Discrete Wavelet Transform,” **Proceedings. 4th National Conference on Telecommunication Technology (NCTT 2003)**, Jan 2003, Serdang Malaysia, 2003. pp. 1-4

- [21] Srinonchat, J., et al. "Double Clustering Algorithm applied to Speaker Dependent Information," **9<sup>th</sup> International conference on Speech and Computer (SPECOM 2004)**, 2004, pp. 37-376
- [22] Somchai. Jitapunkul, S. Chompun, and D. Tancharoen. "Novel Technique For Tonal Language Speech Compression Based on Bitrate Scalable MP-CELP Coder," **International Conference on Coding and Computing, 2001**, April 2001, pp. 461-464
- [23] วีระยุทธ คุณรัตน์ศิริ และจักรี ศรีนนท์นัทร. "การเปรียบเทียบประสิทธิภาพการบีบอัดเสียงพูด โดยใช้เทคนิคเวฟเล็ต". การประชุมเครือข่ายวิชาการด้านวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 3 (EENET 2011). 9-11 มีนาคม 2554, กรุงเทพฯ: หน้า 339-342.
- [24] Supatinee Kornsing and Jakkree Srinonchat. "Exploring Dynamic Time Warping Techniquw to Wavelet Speech Compressing," **9<sup>th</sup> International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology 2012 (ECTI-CON 2012)**, 16-18 May 2012, Hua Hin, Thailand, pp.1-4.
- [25] Supatinee Kornsing and Jakkree Srinonchat. "Enhancement Speech Compression Technique Using Modern Wavelet Tranforms," **2012 International Symposium on Computer, Consumer and Control**, 16-18 May 2012, Hua Hin, Thailand, pp. 393-396.



ภาคผนวก



ภาคผนวก ก

ผลการเปรียบเทียบค่า MSE และ PSNR

ตารางที่ ก.1 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงที่เวลา 5 วินาที

เสียงที่	DWT ระดับที่ 3		CELP ที่ Oder 10	
	MSE	PSNR	MSE	PSNR
1	0.0028748	23.97022	0.0324608	14.886143
2	0.0022127	23.012422	0.0342991	14.646906
3	0.0025393	23.763364	0.0576671	11.491106
4	0.0030224	23.159794	0.0661155	10.227434
5	0.0030829	24.303797	0.0386544	14.127749
6	0.0038266	22.966173	0.0381055	14.189861
7	0.0023237	24.788017	0.0274359	15.616539
8	0.0029671	25.276481	0.0331252	14.194184
9	0.0034997	23.423705	0.0472434	12.133331
10	0.0041761	23.792002	0.0449538	13.472075
11	0.0027042	24.114896	0.0320957	14.475029
12	0.0036028	22.146111	0.0295448	15.024892
13	0.004547	22.283524	0.0382226	14.176528
14	0.0045648	21.016978	0.0439744	13.320371
15	0.0041308	22.511003	0.0347751	14.034367
16	0.0045517	22.683727	0.0340293	13.965413
17	0.0060976	21.695152	0.0383886	14.157708
18	0.0062703	21.06468	0.0436595	13.598947
19	0.0053645	21.784233	0.0391855	13.33512
20	0.0070086	21.543442	0.0350249	14.555965

ตารางที่ ก.2 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายที่เวลา 5 วินาที

เสียงที่	DWT ระดับที่ 3		CELP ที่ Oder 10	
	MSE	PSNR	MSE	PSNR
1	0.0026153	17.922604	0.0080534	19.805501
2	0.0040416	19.375243	0.0080534	19.805501
3	0.00132	22.761335	0.0129103	17.908324
4	0.0040954	19.154955	0.0116077	19.352259
5	0.0059752	19.129907	0.0153652	16.85767
6	0.0034073	19.159426	0.0114129	19.425771
7	0.0014847	19.263656	0.0176416	15.640579
8	0.0020805	19.599626	0.006923	20.19701
9	0.0008925	18.05563	0.0063298	21.833968
10	0.0007272	19.947053	0.0056623	22.469798
11	0.0015456	18.513348	0.0055406	22.564202
12	0.002136	18.347009	0.0107575	19.38738
13	0.0027298	19.427813	0.0209147	16.445522
14	0.001584	22.089982	0.0085368	19.902794
15	0.0022302	18.94442	0.0095225	19.87797
16	0.0015407	18.879948	0.0095804	19.80803
17	0.0009355	18.226283	0.0102253	19.90297
18	0.0026682	22.559609	0.0173879	17.597261
19	0.00193	23.511258	0.0126787	18.968972
20	0.0013004	16.156099	0.0089788	19.880626

ตารางที่ ก.3 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงที่เวลา 60 วินาที

เสียงที่	DWT ระดับที่ 3		CELP ที่ Oder 10	
	MSE	PSNR	MSE	PSNR
1	0.0028881	25.393601	0.022	15.4092
2	0.0038161	24.183502	0.0219431	16.586764
3	0.0046281	23.345686	0.0153928	18.126569
4	0.0049199	23.080162	0.0134076	17.183265
5	0.0041229	23.847702	0.0151328	18.200532
6	0.0070284	21.531171	0.0178518	15.001009
7	0.0068391	21.649727	0.0178518	15.001009
8	0.0047649	23.219155	0.0109473	19.606662
9	0.0101814	19.921666	0.0168535	17.732832
10	0.0061512	22.110155	0.0118593	19.259159
11	0.0070606	21.511289	0.0164978	17.825477
12	0.007062	21.51048	0.0129803	18.516091
13	0.0074172	21.297357	0.0155963	18.069525
14	0.0092776	20.325366	0.0136984	17.885817
15	0.0092927	20.318326	0.0141755	17.910452
16	0.0114822	19.399468	0.0144807	18.391836
17	0.0077297	21.118128	0.0103663	19.564434
18	0.0168477	17.734339	0.0173823	17.271298
19	0.020689	16.842334	0.0194717	17.105705
20	0.010578	19.755696	0.0128715	18.496443



ตารางที่ ก.4 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายที่เวลา 60 วินาที

เสียงที่	DWT ระดับที่ 3		CELP ที่ Oder 10	
	MSE	PSNR	MSE	PSNR
1	0.0031574	23.40308	0.008609	20.65021
2	0.0015206	26.275351	0.0046216	20.689929
3	0.0026766	23.229676	0.005219	22.410451
4	0.0017846	25.649991	0.0033351	23.212055
5	0.0010553	23.804905	0.0035585	24.026801
6	0.004088	21.087727	0.0053349	22.728471
7	0.0050452	18.847163	0.0041231	23.277849
8	0.0012771	26.096963	0.0030441	25.165094
9	0.0013611	26.687807	0.004031	23.429848
10	0.0062324	17.126783	0.0065368	21.846061
11	0.0013616	27.171917	0.0035419	23.491152
12	0.0065117	20.225269	0.0030368	25.175514
13	0.0073189	17.992727	0.0027592	25.59184
14	0.0061476	18.327816	0.0024305	26.056956
15	0.0024327	24.789368	0.0041242	23.846305
16	0.0057325	19.98606	0.003477	24.206809
17	0.0014493	27.873579	0.0033074	24.09833
18	0.0103739	17.899945	0.0043448	23.572965
19	0.0033868	22.507603	0.0021639	26.647361
20	0.0023716	22.787445	0.002804	25.521934

ตารางที่ ก.5 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงที่เวลา 5 วินาที

เสียงที่	DWT ระดับที่ 2		CELP ที่ Oder 20	
	MSE	PSNR	MSE	PSNR
1	0.0005529	31.107511	0.002696	25.692502
2	0.0006697	28.006242	0.0443531	13.530493
3	0.0003715	33.193913	0.0443531	13.530493
4	0.0005045	30.203972	0.0474502	12.760899
5	0.0004786	32.939783	0.0334419	14.75682
6	0.000557	32.164677	0.0436645	13.309254
7	0.0007597	28.840684	0.0293374	15.325517
8	0.0004316	32.992252	0.0311255	15.068578
9	0.0006529	30.19193	0.0381754	14.046719
10	0.0008756	30.576496	0.0462981	13.260467
11	0.0006051	30.252044	0.0202068	16.944753
12	0.0008536	28.811976	0.0270052	14.038802
13	0.0005782	31.859607	0.0325079	14.879851
14	0.00062	30.082635	0.0390177	14.08712
15	0.0006315	31.334409	0.0323082	14.020481
16	0.0006904	30.077977	0.0338112	13.77532
17	0.0006339	31.761009	0.036959	14.32253
18	0.0010047	29.944598	0.0339599	14.690076
19	0.0010545	28.02079	0.0363162	14.398732
20	0.0014549	28.371512	0.0397004	14.011785

ตารางที่ ก.6 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายที่เวลา 5 วินาที

เสียงที่	DWT ระดับที่ 2		CELP ที่ Oder 20	
	MSE	PSNR	MSE	PSNR
1	0.0012443	26.136326	0.0052763	22.478755
2	0.0022403	25.043865	0.0070211	21.264277
3	0.000602	26.801058	0.00945	20.245411
4	0.0015629	25.547727	0.0085618	20.674094
5	0.0028698	23.861225	0.0152227	17.794717
6	0.0021228	24.985964	0.0109601	19.601573
7	0.000633	25.13399	0.0175316	17.561531
8	0.0013927	24.886573	0.010769	19.677965
9	0.0007094	24.040034	0.0052474	21.656799
10	0.0005905	24.753329	0.0112706	19.480265
11	0.0009261	26.756598	0.0076929	20.753491
12	0.0016458	21.056031	0.0107884	19.670151
13	0.0020264	22.867249	0.0139194	18.086237
14	0.0008788	27.600783	0.0053296	22.37761
15	0.0013887	24.122585	0.0075445	21.223446
16	0.0009019	25.736274	0.0096426	19.260482
17	0.0007584	24.75008	0.0058153	22.354007
18	0.0015511	24.656619	0.0131125	18.822868
19	0.001344	25.464527	0.0122464	18.652149
20	0.0011118	23.672604	0.0091932	20.365064

ตารางที่ ก.7 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงเวลา 60 วินาที

เสียงที่	DWT ระดับที่ 2		CELP ที่ Oder 20	
	MSE	PSNR	MSE	PSNR
1	0.0005601	32.516923	0.0229674	16.195087
2	0.0006167	32.084655	0.020773	15.825155
3	0.0008253	30.833536	0.0200482	16.663988
4	0.0010036	29.984288	0.0202611	16.824006
5	0.000831	30.803636	0.0190176	17.208175
6	0.0013242	28.780155	0.020958	16.786228
7	0.0016474	27.831723	0.0197169	16.772026
8	0.0011141	29.530353	0.0132571	18.775263
9	0.0030652	25.135152	0.0168355	16.39513
10	0.0015577	28.074837	0.0178333	16.54327
11	0.0014818	28.291806	0.0141074	18.505263
12	0.0019348	27.13331	0.0083963	20.75886
13	0.0020859	26.806898	0.013311	18.757643
14	0.0029012	25.373912	0.0138279	18.59218
15	0.0027016	25.683589	0.0137186	18.460164
16	0.0037101	24.305897	0.0127738	18.496391
17	0.0026307	25.799018	0.0097978	19.615065
18	0.0057385	22.411784	0.0139338	18.555586
19	0.0083491	20.783353	0.0182502	17.387066
20	0.0039518	24.031828	0.012271	19.043338

ตารางที่ ก.8 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายเวลา 60 วินาที

เสียงที่	DWT ระดับที่ 2		CELP ที่ Oder 20	
	MSE	PSNR	MSE	PSNR
1	0.001506	27.525524	0.007379	21.319754
2	0.0007126	30.843898	0.0043422	23.006446
3	0.0015526	27.545657	0.0042501	23.715711
4	0.001182	29.213167	0.0041147	23.856393
5	0.0007733	27.77027	0.0048069	21.81303
6	0.0030727	23.697984	0.0040561	23.918661
7	0.0038862	23.701072	0.0022648	26.449515
8	0.0008069	29.583542	0.0047785	23.206821
9	0.0011021	28.05185	0.0034298	23.363912
10	0.0040554	23.919451	0.0056786	22.457355
11	0.0010932	27.901698	0.0040694	23.904436
12	0.0042995	23.559157	0.0047803	22.996189
13	0.0049279	23.073135	0.0057734	21.888492
14	0.0045925	23.332718	0.0045019	22.96462
15	0.0014255	28.183329	0.0035031	24.455172
16	0.0041591	22.647763	0.0058631	22.086323
17	0.0008825	30.326147	0.0032044	24.676853
18	0.0068971	21.613101	0.0069897	21.447687
19	0.0024696	25.443849	0.003511	24.545432
20	0.0015964	27.385471	0.0031406	22.938156

ตารางที่ ก.9 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงเวลา 5 วินาที

เสียงที่	DWT ระดับที่ 1		CELP ที่ Oder 40	
	MSE	PSNR	MSE	PSNR
1	0.0002208	34.556237	0.0256867	15.902649
2	0.0003573	30.318581	0.0413773	13.531126
3	0.000212	35.099575	0.0467729	12.802588
4	0.0002032	33.945003	0.0549935	12.511112
5	0.0002152	35.540572	0.0361193	14.368634
6	0.0001983	36.430003	0.0360951	14.425248
7	0.0004384	30.922765	0.0376727	14.239472
8	0.0001541	37.039505	0.0346263	14.605676
9	0.0002409	34.2094	0.0410259	13.662874
10	0.0003792	34.16888	0.0480885	13.153839
11	0.0002181	34.778548	0.0299273	15.239058
12	0.0003604	32.701333	0.0467411	13.302746
13	0.0002269	35.4926	0.0467411	13.302746
14	0.0002391	33.90546	0.0455252	12.048539
15	0.0003278	33.809396	0.03893	13.992084
16	0.000259	33.94805	0.0419845	13.768847
17	0.0001946	36.457826	0.0493275	12.898326
18	0.0001884	37.248961	0.0383398	14.163239
19	0.0001763	35.35046	0.0369373	13.848344
20	0.0003131	35.043215	0.0585873	12.321697

ตารางที่ ก.10 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายเวลา 5 วินาที

เสียงที่	DWT ระดับที่ 1		CELP ที่ Oder 40	
	MSE	PSNR	MSE	PSNR
1	0.0001981	35.802786	22.632297	0.0054544
2	0.0003579	33.494248	21.559234	0.0066838
3	0.0001639	33.743386	20.128468	0.0097079
4	0.0003347	32.695044	21.123147	0.0077207
5	0.0005485	31.743878	17.645408	0.0154681
6	0.0004072	32.624549	20.334115	0.0086887
7	0.0001822	31.186348	16.681037	0.0204771
8	0.0003771	33.361893	16.681037	0.0204771
9	0.0002218	32.777354	20.692839	0.0085249
10	0.0002557	33.950615	19.154245	0.0107902
11	0.000216	33.984687	21.62254	0.0068821
12	0.0004166	29.057391	19.849428	0.0103522
13	0.0004792	30.973879	18.527205	0.0140363
14	0.0002485	34.63038	22.784583	0.0044609
15	0.0002819	34.241042	20.461068	0.0080782
16	0.0001828	33.338075	18.827228	0.0130994
17	0.0002315	31.112245	22.105347	0.006158
18	0.0003898	33.512791	17.150457	0.0129962
19	0.000355	32.818387	18.818864	0.0120148
20	0.0004588	27.208848	18.974656	0.0095589

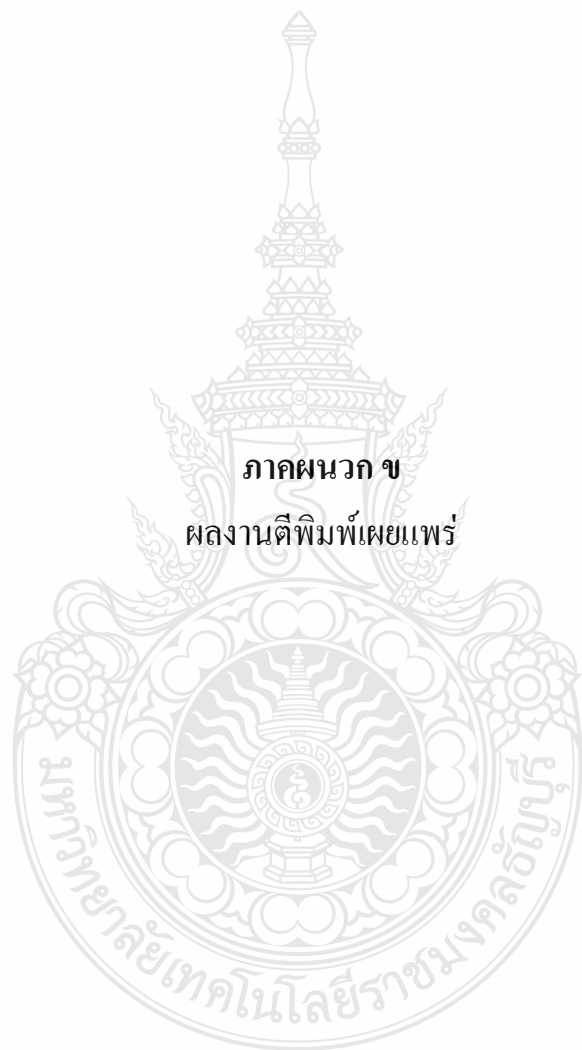
ตารางที่ ก.11 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้หญิงเวลา 60 วินาที

เสียงที่	DWT ระดับที่ 1		CELP ที่ Oder 40	
	MSE	PSNR	MSE	PSNR
1	0.0002549	35.893426	0.0249576	16.021077
2	0.0002365	35.977468	0.0240848	15.511309
3	0.000233	36.326672	0.0202093	16.278948
4	0.0002429	36.144633	0.0207571	16.754861
5	0.0001685	37.734764	0.0207435	16.764664
6	0.0002998	35.232075	0.0207701	16.825356
7	0.0003106	35.077413	0.0201129	16.169093
8	0.000218	36.614328	0.0144159	18.411326
9	0.000524	32.806543	0.0170393	16.270492
10	0.000253	35.968101	0.0092351	19.723136
11	0.0002795	35.535806	0.0151445	17.162423
12	0.000302	35.200187	0.0102343	19.300017
13	0.0003673	34.348944	0.0149539	18.252198
14	0.0006105	32.142987	0.0165259	17.658095
15	0.0005199	32.84065	0.0157318	18.031956
16	0.0007465	31.269162	0.0154317	18.115611
17	0.0006006	32.214162	0.0118139	19.275817
18	0.0013094	28.829069	0.0138766	18.576899
19	0.0018845	27.247679	0.0182829	17.379295
20	0.0009258	30.334643	0.0125915	18.480946



ตารางที่ ก.11 ค่า MSE และ PSNR ของการบีบอัดสัญญาณเสียงสัญญาณเสียงผู้ชายเวลา 60 วินาที

เสียงที่	DWT ระดับที่ 1		CELP ที่ Oder 40	
	MSE	PSNR	MSE	PSNR
1	0.0003282	33.9183	0.0074887	21.25566
2	0.000186	36.355948	0.0049332	23.068459
3	0.0003743	33.742817	0.0041015	23.870315
4	0.000333	34.215062	0.0036891	24.330574
5	0.0002682	33.742489	0.0046732	21.646832
6	0.0008213	30.837941	0.0038572	24.058228
7	0.0011532	29.380748	0.0022703	26.438928
8	0.0002199	35.999833	0.0045334	23.435469
9	0.0003831	32.497935	0.0039907	23.820399
10	0.0011717	29.311608	0.0028542	25.444938
11	0.0003893	32.325213	0.0040673	23.760214
12	0.0012981	28.866694	0.0048528	23.139827
13	0.0015168	28.190444	0.0059837	22.042748
14	0.0014781	28.3026	0.0045145	23.361389
15	0.0003666	34.170014	0.003327	24.779239
16	0.0011551	28.992474	0.0057197	22.345873
17	0.0002568	35.539948	0.0032564	24.872391
18	0.0016038	27.914598	0.0070589	21.51238
19	0.0005771	32.386991	0.0035014	24.557332
20	0.0004892	32.625249	0.0031269	25.048555



**ภาคผนวก ข**  
**ผลงานตีพิมพ์เผยแพร่**

# **Electrical Engineering Network 2012**

## **of Rajamangala University of Technology (EENET 2012)**



### **CONFERENCE TOPICS**

#### **GROUP 1 (PE)**

Power Electronics, Electric Machines, Motor Control and Drive, Measurement, Control and Robotics.

#### **GROUP 2 (PW)**

Power System, Transmission and Distribution, High Voltage and Electrical Energy, Generating Systems.

#### **GROUP 3 (RE)**

Renewable Energy, Energy Saving Technologies, Industry Specific Energy Conversion and Conditioning Technologies, Materials for Energy and Environment.

#### **GROUP 4 (TE)**

Telecommunication, Electronics, Information and Communication Technologies, Antennas, Microwave Theory and Techniques.

#### **GROUP 5 (CP)**

Computer Technologies and Network, Computer Graphics, Machine Learning and Human-Computer Interaction.

#### **GROUP 6 (GN)**

Education in Electrical Engineering, Simulation Software and Design tools, Related Topics in Electrical Engineering.



**GRAND PARADISE HOTEL**  
Nong Khai, THAILAND  
*April 3-5, 2012*

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

**คณะกรรมการเครือข่ายวิชาการวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล**

รองศาสตราจารย์ ดร.โกศล โอบารโพรจน์	ประธานกรรมการ
นายชาญฤทธิ์ ธาราสันติสุข	กรรมการ
ดร.กฤษณ์ชนม์ ภูมิภิตติพิชญ์	กรรมการ
นายประหยัด กองสุข	กรรมการ
ดร.นัฐโชติ รักไทยเจริญชีพ	กรรมการ
ผู้ช่วยศาสตราจารย์ ดร.เจนศักดิ์ เอกบุรณวัฒน์	กรรมการ
ดร.อุเทน คำน่าน	กรรมการ
ผู้ช่วยศาสตราจารย์วิสุทธิ์ พงศ์พิทยภา	กรรมการ
รองศาสตราจารย์ ดร.ธวัช เกิดชื่น	กรรมการ
ผู้ช่วยศาสตราจารย์ ดร.ศักดิ์ระวี รัชวิกุล	กรรมการ
ดร.สุรียา แก้วอาษา	กรรมการ
ผู้ช่วยศาสตราจารย์ ดร.ประมุข อุณหเลขกะ	กรรมการและเลขานุการ



การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

รายชื่อผู้พิจารณาบทความการประชุมเครือข่ายวิชาการวิศวกรรมไฟฟ้า  
มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4

รศ. ดร. วีระเชษฐ ชันเงิน

รศ. ดร. วิจิตร กิณเรศ

รศ. ดร. มนตรี ศิริปรัชญานันท์

รศ. ดร. เวทิน ปิยะรัตน์

รศ. ดร. เศชา พวงดาวเรือง

รศ. ดร. วิบูลย์ ชื่นแขก

รศ. ดร. เสถียร ธัญญศรีรัตน์

ผศ. ดร. แนบบุญ หุนเจริญ

ผศ. ดร. อภินันท์ อูร์โสภณ

ผศ. ดร. วรวัฒน์ เสี่ยมวิบูล

ผศ. ดร. อาทิตย์ โสทรโยม

ผศ. ดร. เผด็จ เผ่าละออ

ผศ. ดร. กองพล อารีรักษ์

ดร. ยุทธนา จำสุวรรณ์

ดร. นิวัฒน์ อังควิสิษฐพันธ์

รศ. ดร. โกศล โอฬารไพโรจน์

ดร. อุเทน คำน่าน

นายณรงค์ นันทกุล

ผศ. กลยุศยา ยิ่งขยัน

นายเอกทัศน์ พฤษวรรณ

ดร. จัตตุฤทธิ์ ทองปรอน

ผศ. ชาญชัย เศรษฐธรรมรงค์

ดร. จักรกฤษณ์ เคลือบวัง

นายสมนึก เครือสอน

นายทัศนะ ถมทอง

นายณรงค์ฤทธิ์ พิมพ์คำวงศ์

ผศ. อภิศักดิ์ ชันแก้วหล้า

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพ เชียงใหม่

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

ศส.สุรสิทธิ์ แสนทอน	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก
นายเอกสิทธิ์ สุนนพันธุ์	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก
นายปรีชา มหาไม้	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก
นายก่อเกียรติ อี๊ดทรัพย์	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก
รศ.ดร.ชวัล เกิดชื่น	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ดร.พินิจ ศรีธร	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.ประเสริฐ เพื่อนหมื่น ไวย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.พันธ์พงศ์ อภิชาติกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.สุธาธินันท์ ต้นโพธิ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.ศิริชัย ลาภาสระน้อย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.กฤตวิทย์ บัวใหญ่	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ศส.วุฒิชัย สว่างงาม	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นายกิตติวุฒิ จีนะบุตร	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นายรุ่งเพชร ก่องนอก	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นายเอกจิต สุ่มวงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นางอุษา กองเมือง	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นายชิตีสรศักดิ์ วิชิโต	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
ดร.วรรณรัชย์ วงศ์ไตรรัตน์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา
นายบุญช่วย เจริญผล	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ
นายวุฒิวัดน์ กงรัตนประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ
นายชูศักดิ์กู กมลขันดิธร	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ
นายวินัย เมธาวิฑิต	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ
นายชาญฤทธิ์ ธาราสันติสุข	มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ
นายประหยัด กองสุข	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา
นายภัทรพงศ์ อัญชันภาติ	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา
นายสมพล โคศรี	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา
นายทักษ์พันธ์ สุวรรณทัต	มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา
นายจตุรงค์ จตุรเชิดชัยสกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายพูนศรี วรรณการ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ดร.ณัฐพงศ์ พันธุนะ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

ดร.นัฐโชติ รักรไทยเจริญชีพ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายนิติพันธ์ คุณประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายชนารัตน์ ตันมณีประเสริฐ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายสุภาวุฒิ เนตรโพธิ์แก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายทอง ถานธารทอง	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายสมเกียรติ ทองแก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.พิชญ์ ดาราพงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายมนัส บุญเพียรทอง	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายพนา ดุสิตาการ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.กิจจา ถักยงอำนาจการ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.โกศล นิธิโสภာ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.จรินทร์ จุลวานิช	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายนิลमित นิลาส	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายเกรียงไกร เหลืองอำพล	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายวัฒนศัพท์ วัลย์วุฒิ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
นายฉกัฏฐ์วรินทร์ ทองรักย์	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.ศรีศักดิ์ น้อยไร่ภูมิ	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ
ศส.ดร.ประมุข อุณหเลขกะ	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
นางสาวพัชรนันท์ ศรีธนาอุทัยกร	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ศส.สรายุทธ ทองกุลภัทร์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ศส.วารุณี ศรีสงคราม	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ดร.ยุพธนา กัณฑ์พะเยา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ศส.เฉลิมพล เรื่องพัฒนาวิวัฒน์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ศส.กระจำง พิทักษ์วงศ์วิทยา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
ดร.ศรีสุดา ไชยทองสูง	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
รศ.นภัทร วัฒนเทพินทร์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี
รศ.สมพันธ์ ย่ำพาวัน	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา
ดร.ชูวงศ์ วัฒนศักดิ์ภูบาล	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา
ศส.วิสุทธิ์ พงศ์พิฤกษ์ชาติ	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา
ดร.สุริยา แก้วอ้อษา	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

ดร.นิธิโรจน์ พรสุวรรณเจริญ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
ดร.มธนา ทิศคร	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
ดร.นำพน พิพัฒน์ไพบูลย์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
ผศ.วิชัย ครอบกิจศิริ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายเอกวิทย์ หายักวงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายวีระ ชันยาภิรักษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายรัก สกุลพงษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายนครินทร์ ศรีปัญญา	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายกฤตยา สมสัย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายจงเจริญ คุ้มบุญ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายกฤษฏา บุญมีวิเศษ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายจิระพงษ์ ศรีวิชัย	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายวีระชัย จรบูรมย์	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร
นายปฏิวัติ บุญมา	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี
นายเจสสิว เกตุแก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี
ผศ.ดร.ปรีชา สาकरังค์	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี
ดร.ภักวัฒน์ จันทร์ศรี	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี
นายไพบูลย์ เกียรติสุขเกษมธร	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี
รศ.พันธ์ พิริยวรรณ	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ผศ.ประวิช เปรียบเหมือน	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ผศ.ดร.ศักดิ์ระวี วัชรวิกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ผศ.ประยงค์ เสาร์แก้ว	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ผศ.จักรวัฒน์ บุตรบุญชู	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ดร.สุระ ตันดี	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ดร.ณรงค์ สีหาจ่อง	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น
ผศ.ดร.สมชัย หิรัญวโรคม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผศ.ดร.วันชัย ทรัพย์สิงห์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผศ.วิชัย ผดุงศิลป์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผศ.ศิริชัย แดงเอม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผศ.ณัฐวุฒิ ไส้มะเกษตรินทร์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี



การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

ดร.ฉัตรชัย	ศุภพิทักษ์สกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.บุญยัง	ปลั่งกลาง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.สุรินทร์	แหงมงาม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.กฤษณ์ชนม์	ภูมิภักดีพชญ์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.ณัฐภัทร	พันธ์คง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายพีนิจ	จิตจริง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายสมชาย	เขียนสูงเนิน	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายพร้อมศักดิ์	คือภักดีกุล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายนิติพงศ์	ปานกลาง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายธีระพล	เหมือนขาว	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายณัฐพล	หาอุปละ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ศส.จินตนา	นาคะสุวรรณ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ศส.ปราชญ์	ดาบบัง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.จักรี	ศรีรินทร์ฉัตร	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.อำนาจ	เรืองอำนาจ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.ไพฑูรย์	รักเหลือ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายวิโรจน์	พิราจนเนนชัย	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายพงษ์ศักดิ์	อำภา	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายบุญยัง	นบนอบ	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายอภิรดา	นามแสง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.สุทินัน	พรอนุรักษ์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.วินัย	วิชัยพาณิชย์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ดร.กิตติวัฒน์	นัมเกิดผล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ศส.อิฐอรัญ	ปีติมล	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
นายฉัตรดิพงษ์	อุทอง	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

TE05	การทดลองออกอากาศแพร่ภาพระบบโทรทัศน์ดิจิทัลแบบเคลื่อนที่ตามมาตรฐาน DVB-H ในพื้นที่กรุงเทพมหานคร	489
TE06	อัลกอริทึมสำหรับการฝังและการถอดรายน้ําดิจิตอลสำหรับภาพสีบนพื้นฐานการแปลงเวฟเล็ตแพกเก็ตส์กับแบนด์ย่อยที่ดีที่สุด	493
TE07	การออกแบบสายอากาศความถี่กว้างมากแบบสวมใส่ได้สำหรับอุปกรณ์พกพา	497
TE08	การออกแบบสายอากาศแบบโมโนโพลรูปสี่เหลี่ยมผืนผ้า สำหรับประยุกต์ใช้งานย่านความถี่กว้างยิ่งด้วยเทคนิคการเซาะร่อง	501
TE09	Realization of Electronic Tunable Current-mode Quadrature Oscillator Based on Third Order Technique	505
TE10	โมดูลวงจรมอดูเลตทางขนาดด้วยชิพวงจรรวมคุณสมบัติสัญญาณ	509
TE11	การสังเคราะห์ตัวควบคุมแบบ PI, PD และ PID โหมดกระแสโดยใช้วงจรรายพานกระแส	513
TE12	ระบบสังเคราะห์ความถี่ด้วยเฟสล็อกคูล	517
TE13	การขยายแบนด์วิดท์ของสายอากาศของเป็กรูปสี่เหลี่ยมผืนผ้าที่มีสัณฐานรูปตัวไอแบบแถวลำดับสำหรับการสื่อสารไร้สาย	521
TE14	สายอากาศโมโนโพลรูปตัววายสำหรับการใช้งานย่านความถี่แถบคู่	525
TE15	การออกแบบติดตั้งแอดเซสท้อยท์มาตรฐาน IEEE 802.11b/g ย่านความถี่ 2.4 GHz ระบบเครือข่ายแลนไร้สายพื้นที่ภายนอกอาคาร	529
TF16	การออกแบบสายอากาศหลายความถี่ที่โหนดด้วยโครงสร้างระนาบร่วมรูปร่างรีซัน	533
TE17	เทคนิคการเปรียบเทียบเวฟเล็ตสำหรับยัตราการบีบอัดใหม่ของสัญญาณเสียง	537
TE18	การพัฒนาเทคนิคในการแยกแยะเสียงโฆษะและเสียงโฆษะบนพื้นฐาน LSP Coefficients	541
TE19	ยูนิเวอร์แซลฟิลเตอร์สามอินพุตหนึ่งเอาต์พุตรูปแบบกระแสปรับค่าได้ทางอิเล็กทรอนิกส์ด้วยโครงสร้างของ CCDDCCs	545

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

## เทคนิคการเปรียบเทียบเวฟเล็ตสำหรับอัตราการบีบอัดใหม่ของสัญญาณเสียง Wavelet Comparison Technique for New Bit Rate Speech Compression

ศุภาธิณี กรสิงห์ และ จักรี ศรีนนท์ถัศคร

ห้องปฏิบัติการวิจัยการประมวลผลสัญญาณ สาขาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

39 ม.1 ค.คลองหก อ.ธัญบุรี จ.ปทุมธานี 12110 โทรศัพท์: 02-549-3588 E-mail: supatinee\_k@hotmail.com, jakkree.s@en.rmutt.ac.th

### บทคัดย่อ

บทความนี้ได้ทำการทดลองบีบอัดสัญญาณเสียงพูดด้วยเวฟเล็ต 3 ตระกูล ได้แก่ ฮาร์ (Haar Wavelet :Haar) ไบออร์ทอโกลนอล (Biorthogonal Wavelet: Bior) และการประมาณไม่ต่อเนื่องของเมเยอร์เวฟเล็ต (Discrete Approximation of Meyer Wavelet: Dmey) และใช้เสียงสำหรับการทดลองทั้งหมด 80 เสียง โดยแบ่งเป็นเสียงของผู้ชาย 40 เสียง และเสียงของผู้หญิง 40 เสียง ในบทความนี้ได้นำเทคนิคคุณสมบัติของการหาค่าพลังงานเฉลี่ย และการเปรียบเทียบความดีสเปกตรัมของสัญญาณเสียงใหม่กับสเปกตรัมของสัญญาณเสียงต้นฉบับ จากผลการทดลองสรุปได้ว่า การบีบอัดสัญญาณเสียงด้วย Bior มีคุณภาพดีกว่า Haar และ Dmey ทั้งนี้เทคนิคนี้ยังสามารถลดอัตราการบีบอัดสัญญาณเสียงลงมาถึง 24 kHz

คำสำคัญ: การบีบอัดสัญญาณเสียง ฮาร์ ไบออร์ทอโกลนอล การประมาณไม่ต่อเนื่องของเมเยอร์เวฟเล็ต

### Abstract

This paper presents the analysis and comparison of speech compression using Haar wavelet (Haar), Biorthogonal wavelet (Bior) and Discrete approximation of Meyer wavelet (Dmey). There are 80 speech signal input which are collected from 40 male and 40 female. The physical properties of the average energy and the comparison of the frequency spectrum of speech signal are also used in this research. The results show that the compressed speech signal with Bior provides better quality than the Haar and Dmey. This technique also provides the speech compression rate at 24 kHz which is less than the original.

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

### 1. บทนำ

ปัจจุบันมีการบีบอัดข้อมูลมีความจำเป็นอย่างมาก เช่น สำหรับการส่งข้อมูล สัญญาณเสียง หรือสัญญาณภาพ เป็นต้น ส่วนเทคนิคที่ใช้ในการบีบอัดที่พบนั้นมีหลายเทคนิค เช่น การบีบอัดด้วย Linear Predictive Coder (LPC), Line Spectrum Pairs (LSP) เพื่อสร้างอัตราบิต (Bit Rate) [1], การใช้ Artificial Neural Network (ANNs) สร้างสัญญาณเสียงให้มีลักษณะคล้ายกับเสียงต้นฉบับ [2] และการบีบอัดสัญญาณเสียงพูดโดยใช้การแปลงเวฟเล็ต (Wavelet Transform) [3-5] เป็นต้น ซึ่งแต่ละเทคนิคนี้มีวิธีการบีบอัดและคุณสมบัติที่ต่างกัน การแปลงเวฟเล็ตเป็นเทคนิคหนึ่งที่มีนิยามใช้กันอย่างแพร่หลาย ซึ่งเป็นการวิเคราะห์สัญญาณแบบ Multi-Resolution โดยการแปลงเวฟเล็ตจะให้ผลลัพธ์เป็นค่าสัมประสิทธิ์ของเวฟเล็ต (Wavelet Coefficient) การแปลงสัมประสิทธิ์เวฟเล็ตสามารถทำได้ 2 แบบ คือ แบบต่อเนื่อง (Continuous Wavelet Transform: CWT) และแบบไม่ต่อเนื่อง (Discrete Wavelet Transform: DWT) ซึ่งจากงานวิจัยที่ผ่านมาการบีบอัดสัญญาณเสียงด้วย DWT ถูกนำมาใช้จนค่อนข้างมา

บทความนี้ได้ใช้การแปลงเวฟเล็ตแบบไม่ต่อเนื่อง 3 ตระกูล ได้แก่ ฮาร์ ไบออร์ทอโกลนอล และการประมาณไม่ต่อเนื่องของเมเยอร์เวฟเล็ต เพื่อบีบอัดสัญญาณเสียงพูด จากนั้นนำสัญญาณเสียงใหม่ที่ผ่านกระบวนการบีบอัดแล้วมาเปรียบเทียบกับสัญญาณเสียงต้นฉบับ โดยใช้หลักการหาค่าพลังงานเฉลี่ยและการเปรียบเทียบความดีสเปกตรัม

### 2. การแปลงสัมประสิทธิ์เวฟเล็ตแบบไม่ต่อเนื่อง (Discrete Wavelet Transform: DWT)

ทั้งขั้นตอนการแปลงเวฟเล็ตแบบไม่ต่อเนื่อง (DWT) จะมีการแบ่งค่าสัมประสิทธิ์ออกเป็น 2 ชนิด คือ สัมประสิทธิ์การประมาณ (Approximation Coefficients) และสัมประสิทธิ์รายละเอียด (Detail Coefficients) โดยการแปลงสัมประสิทธิ์เวฟเล็ตจะทำการคำนวณค่าสัมประสิทธิ์การแปลง  $DWT_{j,k}$  จากการหาความสัมพันธ์ของสัญญาณอินพุต  $(f(x))$  กับฟังก์ชันเวฟเล็ตแม่  $(M(x))$  ที่เลือกใช้ในแต่ละสเกล  $(j)$  และการเลื่อนตำแหน่ง  $(k)$  ดังสมการที่ (1)

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

$$DWT_{a,k} = \int_R s(x) \cdot \frac{1}{\sqrt{a}} \psi\left(\frac{x-k}{a}\right) dx \quad (1)$$

บทความนี้ใช้เวฟเล็ต 3 ตระกูลเพื่อการวิเคราะห์การบีบอัดสัญญาณเสียงชุดนี้

**2.1 Haar Wavelet (Haar)**

ฟังก์ชันเวฟเล็ตแม่แบบ Haar เป็นฟังก์ชันเวฟเล็ตที่นำมาประยุกต์ใช้ได้ง่ายที่สุดใช้เวลานในการประมวลผลสั้นและใช้หน่วยความจำในการคำนวณอย่างมีประสิทธิภาพ นอกจากนี้ฟังก์ชันเวฟเล็ตแบบ Haar ยังมีคุณสมบัติตั้งฉาก (Orthogonal)

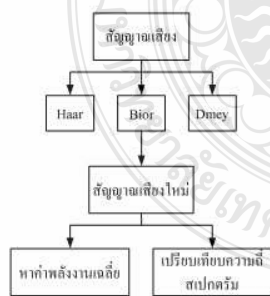
**2.2 Biorthogonal Wavelet (Bior)**

ฟังก์ชันเวฟเล็ตแม่แบบที่มีคุณสมบัติ Biorthogonal มีความสามารถในการรวบรวม (Capturing) ค่ากำลังของสัญญาณที่ต้องการได้อย่างรวดเร็วด้วยการใช้จำนวนของค่าสัมประสิทธิ์ที่น้อย ดังนั้นเทคนิคของการแปลงเวฟเล็ตแบบนี้ถูกนำมาประยุกต์ใช้กับงานทางด้านการประมาณสัญญาณ (Signal Approximation) นอกจากนี้ฟังก์ชันเวฟเล็ตแม่แบบ Biorthogonal ยังมีคุณสมบัติที่สามารถและกระชับ คุณสมบัติสมมาตรนี้เป็นคุณสมบัติที่ต้องการในการสร้างฟังก์ชันการกรองเนื่องจากให้ผลลัพธ์ของฟังก์ชันการแปลงที่ให้ความสัมพันธ์ของเฟสเชิงเส้น (Linear Phase)

**2.3 Discrete Approximation of Meyer Wavelet (Dmey)**

เวฟเล็ต Dmey คือรูปแบบที่ต่อเนื่องของฟังก์ชันเวฟเล็ตเมเยอร์ และสามารถจำลองเวฟเล็ตเมเยอร์ผ่านการสร้างเมทริกซ์กรองย้อนกลับได้ ซึ่งการแปลงเวฟเล็ตแบบเร็วสามารถประมาณค่าการแปลงเวฟเล็ตเมเยอร์ได้

**3. ขั้นตอนการทดลอง**



รูปที่ 1 ขั้นตอนการทดลองการบีบอัดสัญญาณเสียง

ก.) เตรียมสัญญาณเสียงที่มีคุณสมบัติดังที่กล่าวไว้แล้วข้างต้น 80 เสียง

- เสียงผู้ชายที่เวลา 5 วินาที 20 เสียง
- เสียงผู้ชายที่เวลา 60 วินาที 20 เสียง
- เสียงผู้หญิงที่เวลา 5 วินาที 20 เสียง
- เสียงผู้หญิงที่เวลา 60 วินาที 20 เสียง
- สัญญาณเสียงที่นำมาทดลองมีคุณสมบัติดังนี้
  - Bit Rate = 64 kbps
  - Sample size = 8 bit
  - Sample rate = 8 kHz

ข.) นำสัญญาณเสียงเข้าสู่กระบวนการบีบอัดสัญญาณเสียงด้วยวิธี Discrete Wavelet Transform ที่ Level 2 และ Wavelet ที่ใช้ทั้งหมด 3 ตระกูล คือ Haar Wavelet, Biorthogonal Wavelet และ Discrete Approximation of Meyer Wavelet

ค.) หลังจากผ่านขั้นตอนการบีบอัดสัญญาณแล้ว จะได้สัญญาณเสียงใหม่ที่มี Sample rate = 2 kHz ซึ่งต่างจากสัญญาณเสียงต้นฉบับอยู่ 4 เท่า

ง.) นำสัญญาณเสียงใหม่ที่ได้มาปรับปรุง Sample Rate เพื่อให้ได้คุณภาพของสัญญาณเสียงที่ดีขึ้นดังสมการ (2)

$$x_{new} = \frac{p}{q} \times sampling\ rate_{new} \times t_{sec} \quad (2)$$

โดยที่  $x_{new}$  คือ สัญญาณใหม่

$t_{sec}$  คือ เวลาในหน่วยของวินาที (ในการทดลองนี้คือ 5 วินาที และ 60 วินาที)

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

จ.) เปรียบเทียบ ตระกูลของ Wavelet ที่นำมาใช้ในการบีบอัดสัญญาณเสียงครั้งนี้ เพื่อหา Wavelet ที่ดีที่สุดจากทั้ง 3 ตระกูล โดยวิธีทดสอบและสังเกตมีดังนี้

1) หาค่าพลังงานเฉลี่ยของแต่ละสัญญาณเสียงโดยที่ 1 สัญญาณเสียงจะทดสอบค่าอัตราส่วน  $\frac{p}{q}$  ในการปรับปรุงสัญญาณเสียงทั้งหมด 5 อัตราส่วน โดยใช้สมการที่ (3)

$$x = \frac{1}{n} \sum_{i=1}^n |S_i| \quad (3)$$

2) สังเกตจากความถี่สเปกตรัมที่ได้ค่าพลังงานคิดเป็นน้อยที่สุดโดยเปรียบเทียบกับสัญญาณเสียงต้นฉบับ

4. ผลการทดลอง

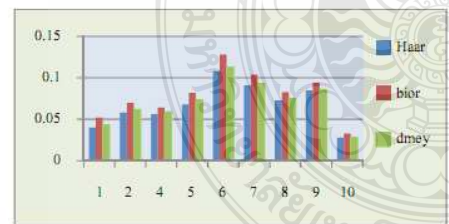
ในการทดลองนี้ใช้อัตราส่วนทั้งหมด 5 อัตราส่วน คือ  $\frac{3}{2}, \frac{4}{2}, \frac{5}{2}, \frac{3}{1}$  และ  $\frac{4}{1}$  ซึ่งค่า  $\frac{p}{q}$  ทั้งหมดนี้ได้ผ่านการทดลองเพื่อหาค่าที่เหมาะสมแล้วโดยอาศัยเวลาเป็นเกณฑ์อ้างอิง ซึ่งค่าที่ได้เมื่อนำไปทดสอบแล้วจะทำให้ได้ค่าสัญญาณเสียงที่มีเวลาเท่าเดิมดังนั้นจากการคำนวณจึงทำให้ได้ค่า Sample Rate และค่าอัตราการบีบอัดสัญญาณเสียงดังตารางที่ 1

ตารางที่ 1 Sample Rate และอัตราการบีบอัดสัญญาณเสียงของอัตราส่วน

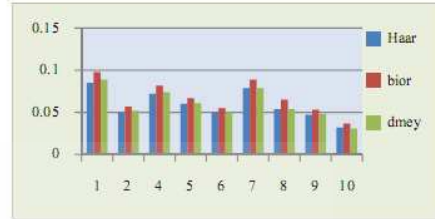
ครั้งที่	P	q	p/q	Sampling rate	Compression rate
1	3	2	1.5	3000	2.67
2	4	2	2	4000	2.00
3	5	2	2.5	5000	1.60
4	3	1	3	6000	1.33
5	4	1	4	8000	1.00

จากผลการทดลองทั้งหมดมี 80 เสียง โดยแบ่งตามทีกล่าวไว้ข้างต้นนั้น ซึ่งมีจำนวนข้อมูลค่อนข้างมาก ดังนั้นจึงยกตัวอย่างจากเสียงทั้งหมดเหลือเพียงแค่ 4 เสียง โดยแบ่งดังนี้ เสียงผู้ชาย 5 วินาที 1 เสียง เสียงผู้ชาย 60 วินาที 1 เสียง เสียงผู้หญิง 5 วินาที 1 เสียง และเสียงผู้หญิง 60 วินาที 1 เสียง และการแสดงกราฟทางความถี่ของสัญญาณในส่วนของการบีบอัด  $\frac{p}{q} = \frac{3}{2}$  เท่านั้น เนื่องจากเป็นอัตราส่วนที่เหมาะสมสำหรับบทความนี้

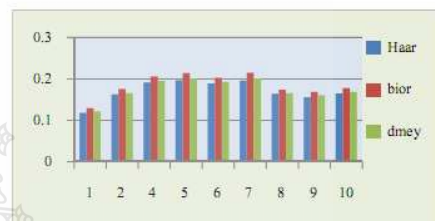
4.1 ผลการทดลองจากวิธีการหาค่าพลังงานเฉลี่ยของสัญญาณเสียง



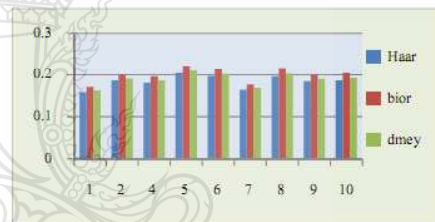
รูปที่ 2 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 5 วินาที



รูปที่ 3 ค่าพลังงานเฉลี่ยของเสียงผู้ชายที่เวลา 60 วินาที



รูปที่ 4 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 5 วินาที



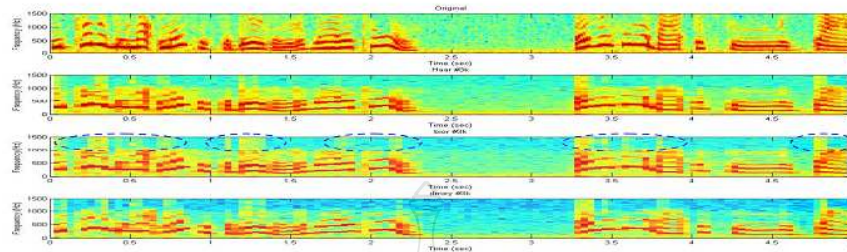
รูปที่ 5 ค่าพลังงานเฉลี่ยของเสียงผู้หญิงที่เวลา 60 วินาที

จากรูปที่ 2 ถึง รูปที่ 5 นั้นเป็นการแสดงการเปรียบเทียบคุณภาพของสัญญาณเสียงหลังจากการบีบอัดของเวฟเลตทั้ง 3 ตระกูล โดยวิธีการหาค่าพลังงานเฉลี่ยตามสมการที่ (3) ด้านเวฟเลตตระกูลใดมีค่าพลังงานเฉลี่ยที่มากที่สุดจะเป็นเวฟเลตที่ถูกเลือกมาใช้งานเนื่องจากให้ค่าความผิดพลาดน้อยที่สุด

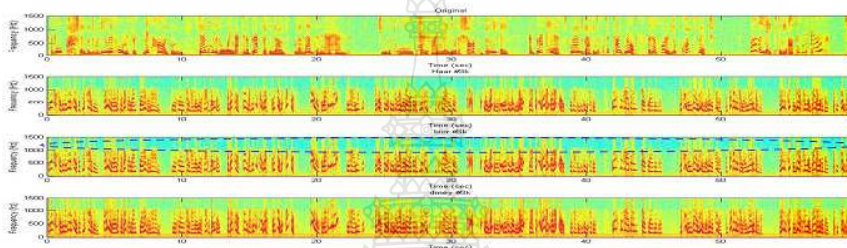
4.2 ผลการทดลองจากวิธีการเปรียบเทียบความถี่สเปกตรัม

การเปรียบเทียบคุณภาพของสัญญาณเสียงหลังจากการบีบอัดด้วยของเวฟเลตทั้ง 3 ตระกูล โดยใช้วิธีการเปรียบเทียบความถี่สเปกตรัมกับความถี่ของสัญญาณเสียงต้นฉบับ ดังภาพที่ 6 และ 7 ด้านเวฟเลตตระกูลใดมีลักษณะของกราฟความถี่ที่เกิดการสูญเสียที่น้อยที่สุดหรือมีกราฟความถี่ที่คล้ายกับต้นฉบับมากที่สุดจะถือว่าเวฟเลตตระกูลนั้นดีที่สุดสำหรับบทความนี้

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555



รูปที่ 6 ความถี่สเปกตรัมของเสียงที่เวลา 5 วินาที



รูปที่ 7 ความถี่สเปกตรัมของเสียงที่เวลา 5 วินาที

5. สรุป

จากการผลการทดลองนี้สรุปได้ว่าแนวไฟต์ทั้งสามตระกูลให้คุณภาพการบีบอัดสัญญาณเสียงที่ใกล้เคียง แต่จากการทดลองทั้งสองวิธีได้ให้ข้อสรุปที่ตรงกันทั้งทางด้านการหาค่าพหุนามเดลีซ การเปรียบเทียบทางด้านความถี่สเปกตรัม โดยเวฟเลตแบบไบออร์ทอโกลนอลจะสามารถบีบอัดสัญญาณเสียงได้ดีที่สุดในบทความนี้ ส่วนเวฟเลตแบบฮาร์ และ เวฟเลตแบบการประมาณไปต่อเนื่องของเมเยอร์ จะมีคุณภาพของสัญญาณเสียงที่ใกล้เคียงกัน และอัตราบีบอัดใหม่ที่ได้เท่ากับ 24 kHz จากบทความนี้เป็นการทดลองบีบอัดสัญญาณเบื้องต้น เพื่อนำไปสู่พัฒนาคุณภาพของสัญญาณเสียงต่อไป เช่น การใช้ OTW เพื่อหาความถี่ใกล้เคียงสัญญาณเสียงใหม่เมื่อเทียบกับเสียงต้นฉบับ

เวฟเลตที่สามารถบีบอัดได้คุณภาพที่ดีที่สุดไปพัฒนาในด้าน การสื่อสารและการบีบอัดข้อมูลต่อไป

เอกสารอ้างอิง

[1] O.Wiriyarunakorn; J.Srinochat. A Finite State Vector Quantizer for New Bit Rate Speech Compression. International Conference on Signal Processing, Communications and Networking, 2008.  
 [2] J.Srinochat. Enhancement Artificial Neural Networks for Low-Bit Rate Speech Compression system. International Symposium on Communications and Information Technologies, 2006.

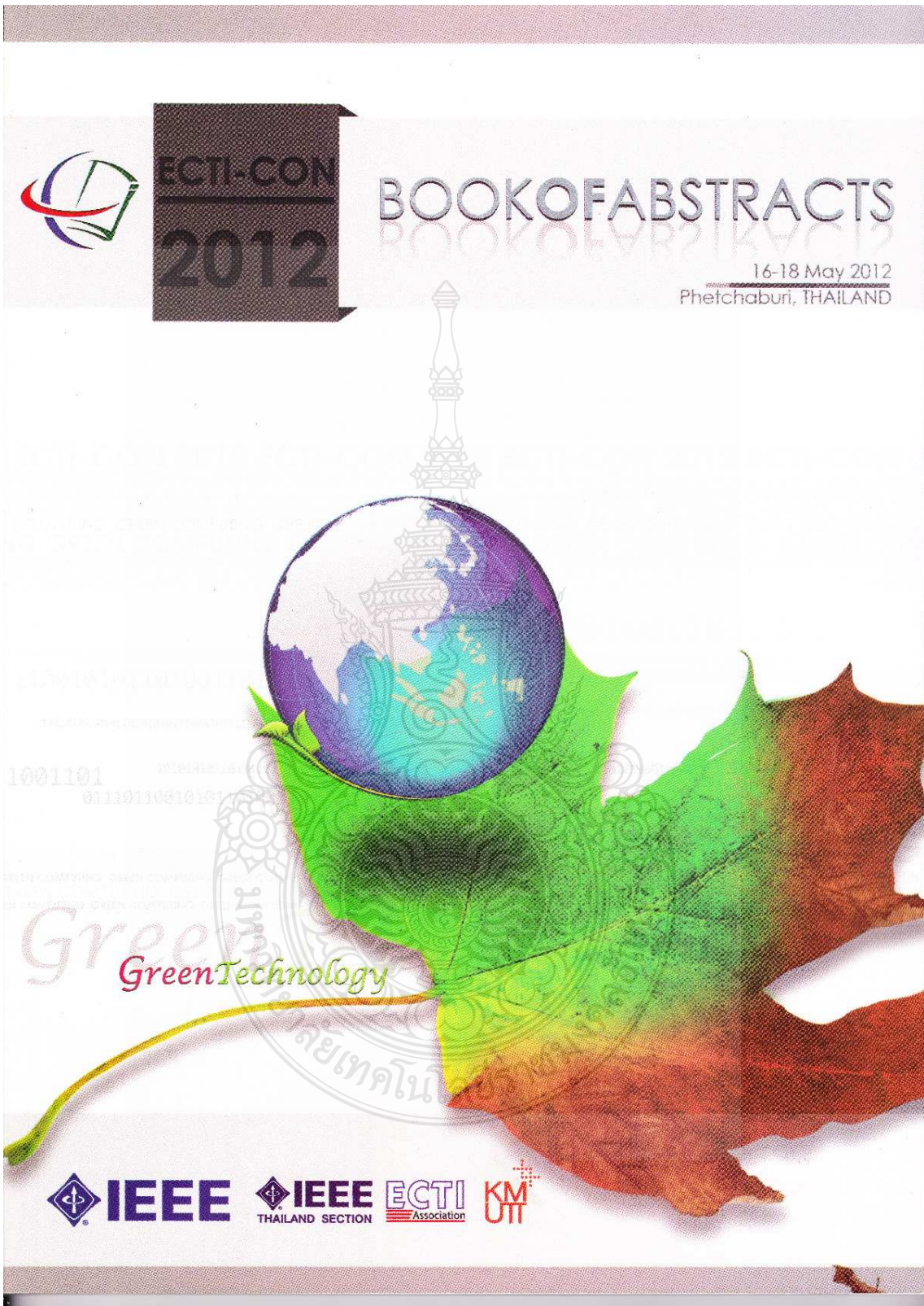
[3] Dr.V.Radha, Vimala.C, M.Krishnaveni. Comparative Analysis of Compression Techniques for Tamil Speech Datasets. International Conference on Recent Trends in Information Technology, 2011.  
 [4] Mahmoud.A.Osman; Nasser Al; Hussein M. Maboud; S.A.Alfandi. Speech Compression Using LPC and Wavelet. 2<sup>nd</sup> International Conference on Computer Engineering and Technology, 2010.  
 [5] Shivesh Ranjan. Exploring the Discrete Wavelet Transform as a Tool for Hindi Speech Recognition. International Journal of Computer Theory and Engineering, 2010.



ดร.วิมล กรสิงห์ สำเร็จการศึกษาระดับปริญญาตรี ด้านวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม จากมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ในปี พ.ศ. 2552 ปัจจุบันกำลังศึกษาาระดับปริญญาโท ด้านวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม



จักรี ศรีนนท์พัชร สำเร็จการศึกษาระดับปริญญาเอก จาก Northumbria University, UK. ในปี พ.ศ. 2548 ในสาขาวิศวกรรมไฟฟ้า สาขาเอกวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคมปัจจุบันดำรงตำแหน่งอาจารย์ผู้สอนที่ภาควิชาอิเล็กทรอนิกส์และโทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี



ECTI-CON  
2012

# BOOK OF ABSTRACTS

16-18 May 2012  
Phetchaburi, THAILAND



GreenTechnology

IEEE THAILAND SECTION ECTI Association KM

## *Committees*

### *Advisory Committee*

Monai Krairiksh  
Booncharoen Sirinaovakul  
Kosin Chamnongthai  
Anawach Sangswang  
Pakorn Kaewtrakulpong

### *General Chair*

Prabhas Chongstitvatana (CU)  
Monai Krairiksh (KMITL)

### *General Co-Chair*

Tiranee Achalakul (KMUTT)

### *General Secretary*

Nuttanart Facundes (KMUTT)  
Marong Phadoongsidhi (KMUTT)

### *Local Arrangement Chairs*

Jumpol Polvichai (KMUTT)  
Jaturon Harnsomburana (KMUTT)

### *Registration & Finance Chairs*

Pceरण Siripongwutikorn (KMUTT)  
Songrit Maneewongvatana (KMUTT)  
Rujipan Sampanna (BU)  
Pairin Kaewkuay (ECTI)

### *Steering Committee*

Banlue Srisuchinwong (12) (SIIT)  
Prayoot Akkaraekthalin (12) (KMUTNB)  
Jitkasame Ngarmnil (12) (MUT)  
Vara Varavithya (12) (KMUTNB)  
Vutipong Areekul (12) (KU)  
Wanlop Surakampontorn (12) (KMITL)  
Tuptim Angkaew (13) (CU)  
Chaiwut Chat-uthai (13) (KMITL)  
Monai Krairiksh (13) (KMITL)  
Tiranee Achalakul (13) (KMUTT)  
Nipon Theera-Umpon (13) (CMU)  
Sansanee Auephanwiriyaikul (13) (CMU)  
Anan Phonphoem (13) (KU)  
Putchong Uthayopas (13) (KU)  
Werachet Khan-ngen (13) (KMITL)  
Thumrongrat Amornraksa (13) (KMUTT)  
Kanda Runapongsasaikaew (14) (KKU)

### *Publicity Chair*

Suthep Madarasmi (KMUTT)

### *Publication Chairs*

Natasha Dejdumrong (KMUTT)  
Naruemon Wattanapongsakorn (KMUTT)



## Committees

### Technical Program Committees

Prayoot Akkaraekthalin (KMUTNB) : TPC Chair  
Kosin Chamnongthai (KMUTT) : TPC Co-chair

#### Area: Circuits and Systems

Apinunt Thanachayanont (KMITL) : Area Chair  
Songphol Kanjanachuchai (CU)  
Pasin Isarasena (NECTEC)  
Jirayuth Mahattanakul (MUT)

#### Area: Computers and Information Technology

Anan Phonphoem (KU) : Area Chair  
Sansanee Auephanwiryakul (CMU)  
Chotipat Pornavalai (KMITL)  
Chalermek Intanagonwiwat (CU)

#### Area: Communication Systems

Chuwong Phongcharoenpanich (KMITL) : Area Chair  
Kamol Kaemarungsi (NECTEC)  
Titipong Lertwiryaprapa (KMUTNB)  
Phumin Kirawanich (MU)

#### Area: Controls

David Banjerdpongchai (CU) : Area Chair  
Diew Koolpiruck (KMUTT)  
Kasemsak Uthaichana (CMU)  
Ithiack Nilkhamhang (SIIT)

#### Area: Electrical Power Systems

Issarachai Ngamroo (KMITL) : Area Chair  
Komsan Hongesombut (KU)  
Surapong Suwankawin (CU)  
Panida Jirutitijaroen (NUS)

#### Area: Signal Processing

Matthew Dailey (AIT) : Area Chair  
Sanparith Marukatai (NECTEC)  
Wuttipong Kumwilaisak (KMUTT)

### Website & Information System Chair

Rajchawit Sarochawikasit

1350	Performance Analysis of SVD and K-means Clustering for Optimization of Fuzzy outputs in Classification of Epilepsy Risk Level from EEG Signals <i>Dr. R. Harikumar, T. Vijayakumar, M.G. Sreejith</i>	137
1417	Tooth Recognition in Dental Radiographs via Hu's Moment Invariants <i>Nakintorn Pattanachai, Nongluk Covavisaruch, Chanjira Suthanayothin</i>	137
1425	Iris Recognition based on Dynamic Radius Matching of Iris Image <i>P. Thumwarin, N. Chitanont, T. Matsuura</i>	138
1435	A Survey on Hemorrhage Detection in Diabetic Retinopathy Retinal Images <i>Parisut Jitpakdee, Pakinee Aimmance, Bunyarit Uyyanonvara</i>	138
<b>Signal Processing</b>		139
1021	Frequency Modulation Using Digital Sinusoidal Oscillator <i>R. Panchalard and A. Nosan</i>	140
1097	Channel Response of HAMR with Linear Temperature-Dependent Coercivity and Remanent Magnetization <i>Rati Wongsathan, Pornchai Supnithi</i>	140
1290	Smart Antenna Using a Combination of D3LS and NCLS Algorithms <i>Phichet Nilsuwan, Akkurat Boonpoonga, Jukkrit Tagapanij, Phaophak Sirisuk</i>	141
1295	Distributed Arithmetic LMS Adaptive Filter Implementation without Look-Up Table <i>Anirut Trakultrirang, Ekkawin Thonuechusin, Sorawat Chivapreecha</i>	141
1402	Generating Transcriptions For Romanized Thai Persons' Names <i>Atiwong Suchato, Chutleekorn Kittikool, and Proadpran Panyabukkana</i>	142
1454	Sub-block Encoder of High-Rate LDPC Code over Fading Channel for Image Transmission <i>Tanaporn Payommai, Werapon Chiracharit and Kosin Chumrongthai</i>	142
1053	MuralCut: Automatic Character Segmentation from Mural Images <i>Thanapong Intharath, Noppadon Khiripet</i>	143
1073	Multi-band FIR Filter Design Using Particle Swarm Optimization with Minimax Initialization <i>Ruksaphol Ngamtawee and Paramote Wardkein</i>	143
1145	Fast Computation of Look-Ahead Unscented Rao-Blackwellised Particle Filters <i>Peerapol Yivapongsitanon</i>	144
1302	Integrity Verification Method of Thai Content for Faxed Document <i>Khavittha Thongkor, Thitporn Pransom, Chavinee Chaisri and Thunrongrat Amornraksa</i>	144
1405	Exploring Dynamic Time Warping Technique to Wavelet Speech Compression <i>Supatinee Korsing, Jakkree Srinonchat</i>	145
1066	Fire Detection for Early Fire Alarm Based on Optical Flow Video Processing <i>Suchet Rinsurongkawongl, Mongkol Ekpanyapong, and Matthew N. Dailey</i>	145
<b>Image Processing</b>		147
1022	Absolute Gradient Algorithm for Adaptive IIR Notch Filterm <i>A. Nosan and R. Panchalard</i>	148

# Exploring Dynamic Time Warping Technique to Wavelet Speech Compression

Supatinee Korsing

Electronics and Telecommunication Engineering Department  
Rajamangala University of Technology Thanyaburi  
Pathumthani, Thailand  
supatinee\_k@hotmail.com

Jakkree Srinonchat

Electronics and Telecommunication Engineering Department  
Rajamangala University of Technology Thanyaburi  
Pathumthani, Thailand  
jakkrit.o@en.rmutt.ac.th, jakkree.s@en.rmutt.ac.th

**ABSTRACT**—Wavelet analysis is a significant tool that has been used for speech and image processing. Three wavelet families, Haar wavelet (Haar), Biorthogonal wavelet (Bior) and Discrete approximation of Meyer wavelet (Dmey) are used in this experiment. Eighty speech signal, which are selected from male and female, are used as input signal. Dynamic Time Warping (DTW) technique is applied to determine and adjust speech feature for compression. The results shown that the Haar wavelet provides good performance in compression condition and also keep the quality of speech close to the original speech.

**Keywords**—Speech signal compression; Haar wavelet; Biorthogonal wavelet; Discrete approximation of Meyer wavelet; Dynamic Time Warping

## I. INTRODUCTION

Speech compression normally uses the bandwidth not over 4 kHz, which this present applications of speech compression have very numerous, in most applications involve the real time coding of speech signal for use in mobile satellite communications, cellular telephony and audio for video phones or video teleconferencing systems etc. [1, 2] There are many research technique which involves to speech compression such as Linear Predictive Coder (LPC), Line Spectrum Pairs (LSP) [3], Artificial Neural Network (ANNs) [4] and Wavelet Transform.

Wavelet transform is a fundamental mathematical tool which can analyze signals both frequency domain and time domain. In order to the wavelet transform is used within an interval, resolution in time and frequency domains, there are quick look at a particular signal interval efficiently. Which the results of wavelet transform is called wavelet coefficients, the coefficients conversion of wavelet able to analyze the 2 types, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) [6, 7], which there are two parameters as a control the length and location of the signal, scale and shifting. The scale used to look at data determines the resolution. In the [5], are studied about recognition of isolated words in Hindi Language speech, based on the DWT by found the LPC coefficients of DWT coefficient in a manner very similar to that use in finding the LPC coefficients of an actual speech signal. The mother wavelets are selected to use 3 families, Daubechies (db), Coiflets (coif) and Discrete Meyer Wavelet (dmey). Found that Daubechies 10, 5-level decomposition and

the DWT gave comparable performance, The Speech Compression based on DWT using Dynamic Time Warping (DTW) Technique for comparison. While the Daubechies 8, 3-level decomposition gave the poorest performance.

Thus this paper selected the DWT to use with speech compression and compare properties of each mother wavelet. The detail of DTW is discussed in the II section, which used the principle of R-square theory into calculate. The mother wavelets and process of the experiment are discussed in the III section. The experiment results and conclusion are explained in IV and V section respectively

## II. DYNAMIC TIME WARPING

The simplest way to recognize an isolated word sample is to compare it against a number of stored word templates and determine is the best match. This goal is complicated by a number of factors. First, different samples of a given word will have somewhat different durations. This problem can be eliminated by simply normalizing the templates and the unknown speech so that they all have an equal duration. However, another problem is that the rate of speech may not be constant throughout the word; in other words, the optimal alignment between a template and the speech sample may be nonlinear. The DTW is an efficient method for finding this optimal nonlinear alignment. The DTW is an instance of the general class of algorithms known as dynamic programming. Its time and space complexity is merely linear in the duration of the speech sample and the vocabulary size.

## III. DISCRETE WAVELET TRANSFORM

The DWT is the most powerful and new signal compression technique which use multi-resolution analysis for analyzing speech signals. Function of this DWT is a frequency scale adjustments and values shifting position discrete, which able to calculate coefficients as the equation (1)

$$C_{a,k} = \int_{\mathbb{R}} s(x) \cdot \frac{1}{\sqrt{a}} \psi\left(\frac{x-k}{a}\right) dx \quad (1)$$

Where  $C_{a,k}$  is the coefficient,  $a$  is scale,  $k$  is shifting by  $\psi(x)$  is the mother wavelet function, which this research is

used to 3 family for compression speech signal analysis follows.

#### A. Haar Wavelet (Haar)

Haar is wavelet function which applies the easiest, using the least time and it has orthogonal property. Limitations of a Haar wavelet are unable to include a value changes immediately and conversion coefficients of Haar wavelet functions can't be separate the data between low frequency and high frequency from each other effectively. Thus it isn't suitable for the application on a compression and the noise removal. The mother wavelet of Haar  $\psi(x)$  can be described as

$$\psi(x) = \begin{cases} 1 & 0 < t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

#### B. Biorthogonal wavelet (Bior)

Bior or semi-orthogonal wavelet. It is only orthogonal to the shifted base function under different scale factor, but has no orthogonality in the same scale factor [6]. by the following equation (3)

$$C_{a,k} = \int_R s(x) \tilde{\psi}_{a,k}(x) dx \quad (3)$$

#### C. Discrete approximation of Meyer wavelet (Dmey)

Dmey wavelet is the discrete format of meyer wavelet function, but it can simulate meyer wavelet through constructing reversible filter matrix, as a result, the fast wavelet transform can approximate meyer wavelet transform [6].

#### D. Process of the speech signal compression

All speech signals which is selected from 40 male and female, is tested this technique. The speech signal can be categorized as following

- 20 male speech signals which each speech has 5 seconds length.
- 20 female speech signals which each speech has 5 seconds length.
- 20 male speech signals which each speech has 60 seconds length.
- 20 female speech signals which each speech has 60 seconds length.
  - Bit Rate = 64 kbps
  - Sample size = 8 bit
  - Sample rate = 8 kHz

The compression speech signal has two processes. Firstly compression by DWT at 2<sup>nd</sup> level and finally de-sample of the new speech signal is show in Fig.1

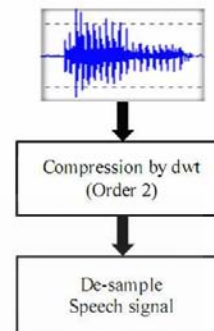


Figure 1. The process compressed speech signal

Firstly, Speech signal is passing through the process of compressing speech signals using 2<sup>nd</sup> level of DWT and all wavelet (Haar, Bior and Dmey). Then it will provide the new speech signal with new sample rate of 2 kHz, that different from original speech signal 4 times. After that the new speech signal is improved the speech quality by using (1) as following formula.

$$x_{new} = \frac{p}{q} \times \text{sampling rate}_{new} \times t_{sec} \quad (4)$$

In which,  $x_{new}$  is the new speech signal,  $t_{sec}$  is the time in units of seconds and  $\frac{p}{q}$  is the ratio used to resample.

At this point, the new speech signal from all wavelet techniques is compared to determine the best technique in condition of quality

- a. Find the average energy of each speech signal.

$$x = \frac{\sum_{i=1}^n |S_i|}{n} \quad (5)$$

In which,  $x$  is the average energy of speech signal,  $n$  is all sampling number and  $S_i$  is the sampling.

- b. Using the spectrum analyzer to analysis the speech signal by comparing with the original speech signal.

#### IV. EXPERIMENTAL RESULTS

This experiment used 5 ratios of 3/2, 4/2, 5/2, 3/1 and 4/1 which are the  $p/q$  to determine the sampling rate appropriate. The sample rate and the speech signal compressed rate are summarized in table 1.

TABLE I. THE SAMPLE RATE AND SPEECH SIGNAL COMPRESSION RATE OF EACH  $\frac{p}{q}$  RATIO.

No.	p	q	p/q	Sample 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

The result shown that, this technique used the  $p/q$  ratio as  $3/2$  or sampling rate as 3000 Hz. It provides the high compression rate as 2.67. Thus the  $3/2$  ratio is used to be the sampling rate for the following experiment

Using principle based on DTW for considered speech compression quality that was used R-square theory shows as the equation (6)

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

Where  $X$  is the frame of modern speech signal,  $Y$  is the original speech signal  $N$  is of all number frame.

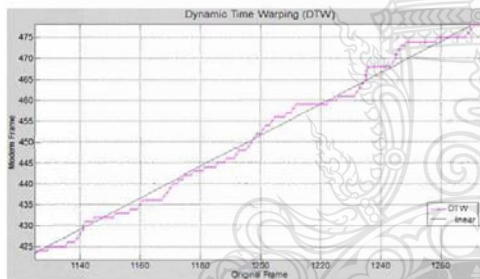


Figure 2. The comparison DTW based on Haar wavelet

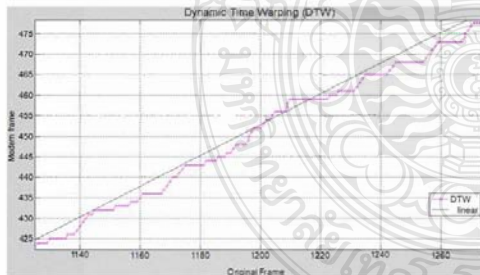


Figure 3. The comparison DTW based on Bior wavelet

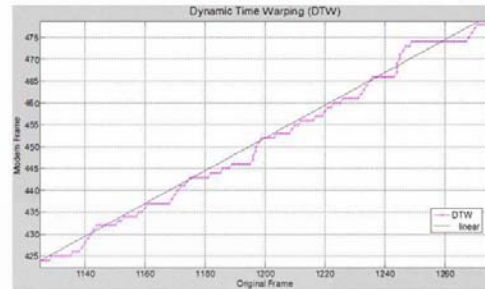


Figure 4. The comparison DTW based on Dmey wavelet

Fig 2-4 show the example of comparison speech signal between DTW and linear line. The R-square technique is used to calculate the error of those signal if the result of R-square equals to one, that means the technique provide the best performance. From the experiments show that Haar wavelet provide the good result as show in table II-V

TABLE II. R-SQUARE VALUE OF MALE SPEECH AT 5 SECONDS.

Speech sample	Haar	Bior	Dmey
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

TABLE III. R-SQUARE VALUE OF MALE SPEECH AT 60 SECONDS.

Speech sample	Haar	Bior	Dmey
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

TABLE IV. R-SQUARE VALUE OF FEMALE SPEECH AT 5 SECONDS.

Speech sample	Haar	Bior	Dmey
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

TABLE V. EXAMPLE OF THE FEMALE SPEECH SIGNAL AT 60 SECONDS

Speech sample	Haar	Bior	Dmey
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	0.99967	0.99986

Table II-V shows the result of three wavelet. It can notice that those wavelets provide similar performance. However, when the system is used to all speech input, it shows the different of compression performing. It shows in table VI.

TABLE VI. PERCENT OF SPEECH QUALITY

Speech	Haar	Bior	Dmey
Female 5 sec	46%	18%	36%
Male 5 sec	69%	12%	19%
Female 60 sec	59%	10%	31%
Male 60 sec	91%	1%	8%

Table VI shows the quality of wavelet compression which used DTW and R-square to measure. It shows that the Haar provided the best performance specially male speech signal.

#### V. CONCLUSION

This experiment explores DTW technique to the wavelet speech compression. Haar, Bior and Dmey are used as compression techniques with eighty speech signal input.

The results show that using ratio at  $\frac{3}{2}$ , sampling rate at 3000 Hz, Provided the compression rate as 2.67 when it is compared to the original speech signal. Also exploring DTW technique to wavelet shows that Haar wavelet give the best compression specially in male speech. This technique can be applied to speech compression area and also image compression.

#### REFERENCES

- [1] Shijo M Joseph. Spoken Digit Compression Using Wavelet Packet. International Conference on Signal and Image Processing, 2010, PP: 255-259.
- [2] Mahmoud.A.Osman; Nasser Al; Hussein M. Maboud; S.A.Alfandi. Speech Compression Using LPC and Wavelet. 2<sup>nd</sup> International Conference on Computer Engineering and Technology, 2010, PP: V7-92 – V7-99.
- [3] O.Wiriyaturakorn; J.Srinonchat. A Finite State Vector Quantizer for New Bit Rate Speech Compression. International Conference on Signal Processing, Communications and Networking, 2008, PP: 255-259.
- [4] J.Srinonchat. Enhancement Artificial Neural Networks for Low-Bit Rate Speech Compression system. International Symposium on Communications and Information Technologies, 2006, PP: 118-121
- [5] ShiveshRanjan. Exploring the Discrete Wavelet Transform as a Tool for Hindi Speech Recognition, International Journal of Computer Theory and Engineering, 2010, PP: 642-646.
- [6] Jian-Yu Zhang; Ling-Li Cui; Gui-Yan Yao; Li-XinGao. Research on The Selection of Wavelet Function for The Feature Extraction of Shock Fault in The Bearing Diagnosis. International Conference on Wavelet Analysis and Pattern Recognition, 2007, PP: 1630-1634.
- [7] R.s.Cai; Y.T.Zhu; Y.M.Guo. Wavelet-Based Multi-Feature Voiced/Unvoiced Speech Classification Algorithm. Conference on Wireless Mobile and Sensor Networks, 2007, PP: 897 – 900.

The cover features a dark background with a grid of blue and green squares. A large, faint watermark of a traditional Thai temple structure is visible in the center. The text is white and positioned in the upper left quadrant.

**IS3C 2012**  
International Symposium on  
**Computer, Consumer and Control**  
4-6 June 2012 • Taichung, Taiwan

[CONFERENCE INFORMATION](#)[PAPERS BY SESSION](#)[PAPERS BY AUTHOR](#)[GETTING STARTED](#)[TRADEMARKS](#)[SEARCH](#)

Published by



## Organizing Committee

### Honorary Chairs

Kuen-Suan Chen, *National Chin-Yi University of Technology, Taiwan*  
Guanrong Chen, *City University of Hong Kong, China*

### Honorary Co-chair

Meng-Hui Wang, *National Chin-Yi University of Technology, Taiwan*

### General Chair

Wen-Yuan Chen, *National Chin-Yi University of Technology, Taiwan*

### General Co-chairs

Her-Terng Yau, *National Chin-Yi University of Technology, Taiwan*  
Chuin-Mu Wang, *National Chin-Yi University of Technology, Taiwan*

### Publication Chair

Chen-Chung Liu, *National Chin-Yi University of Technology, Taiwan*

### Organizing Committee Chair

Jzau-Sheng Lin, *National Chin-Yi University of Technology, Taiwan*

### Program Committee Chairs

Yeo Kiat Seng, *NANYANG Technological University, Singapore*  
Hsiung-Cheng Lin, *National Chin-Yi University of Technology, Taiwan*

### Best Paper Award Chair

Pau-Choo Chung, *National Cheng Kung University, Taiwan*

### Registration and Finance Chair

Yun-Long Lay, *National Chin-Yi University of Technology, Taiwan*

### Local Arrangement Chair

Chiou-Kou Tung, *National Chin-Yi University of Technology, Taiwan*

### Invited Session Organizer

Cheng-Jian Lin, *National Chin-Yi University of Technology, Taiwan*

### Conference Secretary

Ruey-Maw Chen, *National Chin-Yi University of Technology, Taiwan*

### Contact

Hsin-Chiang You, *National Chin-Yi University of Technology, Taiwan*



## Reviewers

Akash Singh, *IT Architect, IBM, USA*  
 Anthony Y. Chang, *Overseas Chinese University, Taiwan*  
 Bor-Chen Kuo, *National Taichung University, Taiwan*  
 C.-C. Kuo, *Saint John's University, Taiwan*  
 C.-M. Huang, *Kun Shan University, Taiwan*  
 C.-Y. Lee, *Chung Yuan Christian University, Taiwan*  
 Chang-Pin Lin, *National Taiwan Ocean University, Taiwan*  
 Chao-Ching Ho, *National Yunlin University of Science and Technology, Taiwan*  
 Chao-Lung Chiang, *Nan Kai University of Technology, Taiwan*  
 Chao-Tang Yu, *Southern Taiwan University, Taiwan*  
 Chao-Wang Young, *National Chiayi University, Taiwan*  
 Chao-Yin Hsiao, *Feng Chia University*  
 Chen-Chung Liu, *National Chin-Yi University of Technology, Taiwan*  
 Cheng-Chien Kuo, *Saint John's University*  
 Cheng-Chien Lin, *National Chin-Yi University of Technology, Taiwan*  
 Cheng-Jian Lin, *National Chin-Yi University of Technology, Taiwan*  
 Cheng-Kang Wen, *Tainan University of Technology, Taiwan*  
 Cheng-Shion Shieh, *Far-East University, Taiwan*  
 Cheng-Tao Tsai, *National Chin-Yi University of Technology, Taiwan*  
 Cheng-Yi Chen, *Cheng Shiu University, Taiwan*  
 Cheng-Yu Yeh, *National Chin-Yi University of Technology, Taiwan*  
 Cheng-Yuan Chang, *National United University, Taiwan*  
 Chia-Ching Lin, *Far-East University, Taiwan*  
 Chia-Nan Ko, *Nan Kai University of Technology, Taiwan*  
 Chieh-Tsung Chi, *Lunghwa University of Science and Technology, Taiwan*  
 Chien-Chuan Ko, *Chiayi University, Taiwan*  
 Chien-Jen Wang, *Kun Shan University, Taiwan*  
 Chih-Chyau Yang, *National Chip Implementation Center, Taiwan*  
 Chih-Hong Lin, *National United University, Taiwan*  
 Chih-Hung Wu, *National University of Kaohsiung, Taiwan*  
 Chih-Peng Li, *National Sun Yat-Sen University, Taiwan*  
 Chi-Jie Lu, *Ching Yun University, Taiwan*  
 Ching-Hua Ting, *National Chiayi University, Taiwan*  
 Ching-Hwa Cheng, *Feng-Chia University, Taiwan*  
 Ching-Lin Wang, *National Chin-Yi University of Technology, Taiwan*  
 Ching-Min Lee, *I-Shou University, Taiwan*  
 Ching-San Lai, *National Taipei University of Education, Taiwan*  
 Ching-Wen Hsue, *National Taiwan University of Technology, Taiwan*  
 Ching-Yen Ho, *Hwa Hsia Institute of Technology, Taiwan*  
 Chin-Pao Huang, *National Chin-Yi University of Technology, Taiwan*  
 Chiou-Kou Tung, *National Chin-Yi University of Technology, Taiwan*  
 Chiu-Keng Lai, *National Chin-Yi University of Technology, Taiwan*  
 Chi-Ying Lin, *National Taiwan University of Science and Technology, Taiwan*  
 Chi-Yuan Lin, *National Chin Yi University of Technology, Taiwan*  
 Chi-Yung Lee, *Nan Kai University of Technology, Taiwan*  
 Chuin-Mu Wang, *National Chin-Yi University of Technology, Taiwan*  
 Chun-Chieh Chu, *National Chip Implementation Center, Taiwan*  
 Chun-Feng Lu, *Chung Chou Institute of Technology, Taiwan*  
 Chung-Chi Huang, *Far East University, Taiwan*  
 Chung-Ho Tsai, *National Chin-Yi University of Technology, Taiwan*  
 Chung-Liang Chang, *National Pingtung University of Science and Technology, Taiwan*  
 Chun-Liang Chen, *Nan Kai University of Technology, Taiwan*  
 Chwan Lu Tseng, *National Taipei University of Technology, Taiwan*

Chyi-Wen Hwang, *Lan Yang Institute of Technology, Taiwan*  
 Cong-Hui Huang, *Far East University, Taiwan*  
 Der-Chen Huang, *National Chung-Hsing University, Taiwan*  
 Der-Fang Shiau, *Fooyin University, Taiwan*  
 Don-Lin Yang, *Feng Chia University, Taiwan*  
 Fan-Chih Chung, *Far East University, Taiwan*  
 Fuh-Gwo Chen, *Hungkuang University, Taiwan*  
 Gary Chang, *National Chung Cheng University, Taiwan*  
 Gowrishankar Kasilingam, *Anna University Chennai, India*  
 Guo-Shiang Lin, *Da-Yeh University, Taiwan*  
 H.Y. Lin, *National Chin-Yi University of Technology, Taiwan*  
 Hao-Chiao Hong, *National Chiao Tung University, Taiwan*  
 Hee-Cheol Kim, *Inje University, Korea*  
 Hiroyuki Kasai, *University of Electro-Communications, Japan*  
 Ho-Nien Shou, *Air Force Institute of Technology, Taiwan*  
 Hsei-Yi Lin, *National Chin-Yi University of Technology, Taiwan*  
 Hsiang-Yueh Lai, *National Chin-Yi University of Technology, Taiwan*  
 Hsin-Chiang You, *National Chin-Yi University of Technology, Taiwan*  
 Hsing Chia Kuo, *Far East University, Taiwan*  
 Hsinying Liang, *Chaoyang University of Technology, Taiwan*  
 Hsu-Chia Chang, *National Chiao Tung University, Taiwan*  
 Hsueh-Hsien Chang, *Jin-Wen University of Science and Technology, Taiwan*  
 Hui-Chieh Li, *Ta Hwa Institute of Technology, Taiwan*  
 Hung-Cheng Chen, *National Chin-Yi University of Technology, Taiwan*  
 Hung-Wei Wu, *Kun Shan University, Taiwan*  
 Hung-Yi Chen, *Ming Chi University of Technology, Taiwan*  
 Hung-Yu Wang, *National Kaohsiung University of Applied Sciences, Taiwan*  
 Ing-Jr Ding, *National Formosa University, Taiwan*  
 Jakkree Srinonchat, *Rajamangala University, Thailand*  
 Jeng-Shyang Pan, *National Kaohsiung University of Applied Sciences, Taiwan*  
 Jian-Liung Chen, *Kao Yuan University, Taiwan*  
 Jiann-Der Lee, *Chang Gung University, Taiwan*  
 Jiann-Shu Lee, *National University of Tainan, Taiwan*  
 Jia-Sheng Hu, *National University of Tainan, Taiwan*  
 Jin-Fu Li, *National Central University, Taiwan*  
 Jin-Shiuh Taur, *National Chung Hsing University, Taiwan*  
 Jong Yih Kuo, *National Taipei University of Technology, Taiwan*  
 Juei-Nan Chen, *Hsiuping University of Science & Technology, Taiwan*  
 Jui-Sheng Lin, *Far East University, Taiwan*  
 Jung-Chieh Chen, *National Kaohsiung Normal University, Taiwan*  
 Jung-Ruey Tsai, *Asia University, Taiwan*  
 Jun-Juh Yan, *Shu-Te University, Taiwan*  
 Jyh-Chyang Renn, *National Yunlin University of Science and Technology, Taiwan*  
 Jyun-Jie Wang, *National Chung Hsing University, Taiwan*  
 Jzau-Sheng Lin, *Department of CSE/National Chin-Yi Institute of Technology, Taiwan*  
 Jzau-Sheng Lin, *National Chin-Yi University of Technology, Taiwan*  
 Kreangkri Ratchagit, *Maejo University, Thailand*  
 Kuan-Chen Lin, *Far East University, Taiwan*  
 Kuang-Hao Lin, *National Chin-Yi University of Technology, Taiwan*  
 Kun-Li Wen, *Chienkuo Technology University, Taiwan*  
 Kuo-Hsien Hsia, *Far East University, Taiwan*  
 Kuo-Hsun Hsu, *National Taichung University of Education, Taiwan*  
 Kuo-Hua Huang, *National Chin-Yi University of Technology, Taiwan*  
 Kuo-Jung Lin, *Fortune Institute of Technology, Taiwan*  
 Kuo-Lan Su, *National Yunlin University of Science & Technology, Taiwan*  
 Liang-Tsung Huang, *Mingdao University, Taiwan*

- Liang-Tu Chen, *National Pingtung Institute of Commerce, Taiwan*  
 Li-Chun Lai, *Chung Chou University of Science and Technology, Taiwan*  
 Lien-Fu Lai, *National Changhua University of Education, Taiwan*  
 Linhong Zhou, *Shanghai University of Finance and Economics, China*  
 Lung-Sheng Yang, *Far East University, Taiwan*  
 Mao-Jiun J. Wang, *National Tsing Hua University, Taiwan*  
 Meei-Ling Huang, *Far East University, Taiwan*  
 Meng-Hui Wang, *National Chin-Yi University of Technology, Taiwan*  
 Ming-Chang Pai, *Nan Kai University of Technology, Taiwan*  
 Ming-Shyan Wang, *Southern Taiwan University, Taiwan*  
 Ming-Tsung Tsai, *Southern Taiwan University, Taiwan*  
 Ming-Yuan Shieh, *Southern Taiwan University, Taiwan*  
 Mirosław Trzupek, *AGH University of Science and Technology, Poland*  
 Mu-Yu Ting, *National Taichung University of Education, Taiwan*  
 Naofumi Uchihara, *University of Electro-Communications, Japan*  
 Neng-Sheng Pai, *National Chin-Yi University of Technology, Taiwan*  
 Po-Hung Chen, *St. John's University, Taiwan*  
 Po-Ming Lee, *Southern Taiwan University, Taiwan*  
 Quang-Cherng Hsu, *National Kaohsiung University of Applied Sciences, Taiwan*  
 Rui-Hua Yang, *Jiangnan University, China*  
 Sanjay Sharma, *Thapar University, India*  
 Seiichi Serikawa, *Kyushu Institute of Technology, Japan*  
 Sendren Xu, *National Taiwan University of Science and Technology, Taiwan*  
 Shang-Pin Ma, *National Taiwan Ocean University, Taiwan*  
 Shann-Chyi Mou, *Ching Yun University, Taiwan*  
 Shao-Ming Yang, *Asia University, Taiwan*  
 Sheng-Fuu Lin, *National Chiao Tung University, Taiwan*  
 Sheng-Jen Hsiao, *National Kaohsiung University, Taiwan*  
 Shih-Cheng Horng, *Chaoyang University of Technology, Taiwan*  
 Shih-Ching Yeh, *National Central University, Taiwan*  
 Shih-Tang Lo, *Kun Shan University, Taiwan*  
 Shing-Jen Wu, *Da-Yeh University, Taiwan*  
 Shun-Jih Wang, *Far East University, Taiwan*  
 Shyam Akashe, *Thapar University, India*  
 Shyi-Kae Yang, *Far East University, Taiwan*  
 Songdong Xue, *Taiyuan University of Science and Technology, China*  
 Tai-Yu Wang, *Nan-Kai University of Technology, Taiwan*  
 Tao Chen, *Key Laboratory of Modern Measurement & Control Technology, China*  
 Tomas Duranik, *Technical University of Kosice, Slovakia*  
 Trong-Yen Lee, *National Taipei University of Technology, Taiwan*  
 Tsai-Rong Chang, *Southern Taiwan University, Taiwan*  
 Tse-Ming Kuo, *National Cheng Kung University, Taiwan*  
 Tsung-Chu Huang, *National Changhua University of Education, Taiwan*  
 Tsung-Hung Lin, *National Chin-Yi University of Technology, Taiwan*  
 Wei-Bin Yang, *Tamkang University, Taiwan*  
 Wei-Chih Hsu, *National Kaohsiung First University of Science and Technology, Taiwan*  
 Wen-Cheng Pu, *National Chin-Yi University of Technology, Taiwan*  
 Wen-Jan Chen, *Da-Yeh University Kaohsiung Normal*  
 Wen-Jeng Liu, *Nan Kai University of Technology, Taiwan*  
 Wen-Tin Lee, *National Kaohsiung Normal University, Taiwan*  
 Wen-Tsai Sung, *National Chin-Yi University of Technology, Taiwan*  
 Xiaodong Sun, *Jiangsu University, China*  
 Xue-Mei Liu, *North China Institute of Water Conservancy and Hydroelectric Power, China*  
 Y.-J. Wang, *National Yunlin University of Science & Technology, Taiwan*  
 Yaw-Wen Kuo, *National Chi Nan University, Taiwan*  
 Yen-Lin Chen, *National Taipei University of Technology, Taiwan*

Yeong-Kang Lai, *National Chung Hsing University, Taiwan*  
 Yi-Cheng Huang, *National Changhua University of Education, Taiwan*  
 Yih-Feng Su, *Far East University, Taiwan*  
 Yih-Lang Li, *National Chiao-Tung University, Taiwan*  
 Ying-Piao Kuo, *National Chin-Yi University of Technology, Taiwan*  
 Yin-Mou Shen, *Kun Shan University, Taiwan*  
 Yin-Tien Wang, *Tamkang University, Taiwan*  
 Yi-Pin Kuo, *Far East University, Taiwan*  
 Yi-Ting Ma, *Hsiuping University of Science & Technology, Taiwan*  
 Yi-Ying Chang, *National Chin-Yi University of Technology, Taiwan*  
 Yi-You Hou, *Far East University, Taiwan*  
 Yongji Wang, *Huazhong University of Science and Technology, China*  
 Yong-Lin Kuo, *National Taiwan University of Science and Technology, Taiwan*  
 Yu-Cheng Fan, *National Taipei University of Technology, Taiwan*  
 Yu-Cheng Huang, *National Chin-Yi University of Technology, Taiwan*  
 Yu-Chi Wu, *National United University, Taiwan*  
 Yu-Hsien Lin, *National United University, Taiwan*  
 Yu-Hsuan Lee, *Yuan Ze University, Taiwan*  
 Yung-Chang Luo, *National Chin-Yi University of Technology, Taiwan*  
 Yung-Chun Wu, *National Tsing Hua University, Taiwan*  
 Yung-Kuan Chan, *National Chung Hsing University, Taiwan*  
 Yun-Long Lay, *National Chin Yi University of Technology, Taiwan*  
 Yu-Sheng Lu, *Nation Cheng Kung University, Taiwan*  
 Yu-Yi Fu, *Nan Kai University of Technology, Taiwan*  
 Yu-Yi Liao, *National Cheng Kung University, Taiwan*  
 Zebin Yang, *Jiangsu University, China*  
 Zhaoqiang Ge, *Xi'an Jiaotong University, China*  
 Zitian Li, *Qijing Normal University, China*  
 Zong-Mu Yeh, *National Taiwan Normal University, Taiwan*



### **C1: Image and Audio Processing**

Induction Motor Bearing Fault Diagnosis Using Hilbert-Based Bispectral Analysis .....	385
<i>D.M. Yang</i>	
Enhancement Pattern Analysis Technique for Voiced/Unvoiced Classification .....	389
<i>Kreangsak Pattanaburi, Jakkrit Onshaunjit, and Jakkree Srinonchat</i>	
Enhancement Speech Compression Technique Using Modern Wavelet Transforms .....	393
<i>Supatinee Kornsing and Jakkree Srinonchat</i>	
Striking Panel Touch Location Detect .....	397
<i>Chin-Yu Wang, Hon-Ta Liu, and Shih-Yu Shen</i>	
Ellipsoidal Function Modulated ART Neural Networks for Pattern Recognition .....	401
<i>Chao-Yin Hsiao, Chin Kun Teng, and Po Shih Hsu</i>	
A New Technique for Text Data Compression .....	405
<i>Udita N. Katugampola</i>	

### **C2: Digital Signal Processing IV**

An Optimal Selection Approach for a Multi-tenancy Service Based on a SLA Utility .....	410
<i>Chih-Kun Ke, Zheng-Hua Lin, Mei-Yu Wu, and Shih-Fang Chang</i>	
Energy Efficient and QoS Based Multi-path Hierarchical Routing Protocol in WSNs .....	414
<i>Maryam Dehnavi, Mohammad Reza Mazaheri, Behzad Homayounfar, and Sayyed Majid Mazinani</i>	
An RFID Authentication and Anti-counterfeit Transaction Protocol .....	419
<i>Chin-Ling Chen, Yu-Yi Chen, Tzay-Farn Shih, and Tsungm-Min Kuo</i>	
A Petri Net Modeling Approach Based on Boolean Function Transition .....	423
<i>Hsing-Chung Chen, Jia-Rong Sun, Yung-Fa Huang, and Zhen-Dong Wu</i>	
A Performance Study for Power Management Schemes in WPANs .....	427
<i>Neng-Chung Wang and Chih-Cheng Wen</i>	
Analysis of Vocational High School Websites in Taiwan .....	431
<i>Hsiu-Te Sung and Yi-Chuan Ho</i>	

### **C3: System and Control I**

Computational and Experimental Study of a Coaxial Pipe Flow Generator .....	435
<i>Songhao Wang and Ronald Doblado</i>	
Bifurcation Analysis of Trailing Edge Angle for Aeroelastic System .....	440
<i>Cheng-Chi Wang, David T-W Lin, and Her-Terng Yau</i>	

## Enhancement Speech Compression Technique Using Modern Wavelet Transforms

Supatinee Korsing /1st

Electronics and Telecommunication Engineering  
Department Rajamangala University of Technology  
Thanyaburi Pathumthani, Thailand  
supatinee\_k@hotmail.com

Jakkree Srinonchat/2nd

Electronics and Telecommunication Engineering  
Department Rajamangala University of Technology  
Thanyaburi Pathumthani, Thailand  
jakkree\_s@hotmail.com

**Abstract**— The Discrete Wavelet Transform is the most powerful and new signal compression technique which uses multi-resolution analysis for analyzing speech signal. This article proposes the frequency lossless comparison of the modern wavelets based on speech compression technique. Haar wavelet (Haar), Biorthogonal wavelet (Bior) and Discrete approximation of Meyer wavelet (Dmey) are investigated in this experiment with eighty speech signal from male and female. The experiments indicated that the Bior provides the best performance in time and frequency domain. Also it can also gain the compression rate as 2.67 to the original speech signal.

**Keywords**- Speech signal compression; Haar wavelet; Biorthogonal wavelet; Discrete approximation of Meyer wavelet;

### I. INTRODUCTION

Since speech signal contains a large number of redundant information, how to compress high quality speech maintaining transparent quality at low bit rates is still a very important topic. To reduce redundancy and make full use of the human's auditory masking effect by using a variety of source coding techniques, not only can compress the coding rate by many times, but also has the ability to regain high intelligibility and acceptability of speech signals. Therefore, a speech compression system focuses on reducing the amount of redundant data while preserving the integrity of signals. The different transformation of speech signals to the time-frequency and time-scale domains for the purpose of compression aim at representing them with the minimum number of coding parameters. [1, 2]

Speech is non-linear random process. Wavelet transform devotes a lot to deal with time-varying, non-stationary signal, for it has excellent resolution in both time and frequency domain. Wavelet transform with detail of signal, decomposes the high-frequency, and the signal was decomposed to the time-frequency space which has a certain correspondence with critical band of speech. The result of wavelet transform is called wavelet coefficients. The coefficients conversion of wavelet can classify into two types, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) [3, 4]. This article investigates the improvement speech compression technique based on the model wavelet transform which is focused in the frequency lossless. There are some related researches works which used wavelet transform for speech compression. In [5], it explores the DWT as a tool for Hindi speech

recognition. It studies the recognition of isolated words in Hindi Language speech. The mother wavelets are selected to use 3 families, Daubechies (db), Coiflets (coif) and Discrete Meyer Wavelet (dmey). It is found that Daubechies 10, 5-level decomposition and the Discrete Meyer wavelet give comparable performance, while the Daubechies 8, 3-level decomposition provides the poorest performance. In [6] presents reliable measures of speech compression by detecting the end points of the speech signals prior to compressing them. Two different compression schemes used are the Global threshold and the Level method is tested with the Signal to Noise Ratios (SNR), Peak Signal to Noise Ratios (PSNR) and Normalized Root Mean Square Error (NRMSE) parameter measures. In [7] applies the Wavelet Packet Transform (WPT) to process speech signal to obtain optimal wavelet tree to allocate the dynamic bits, and then uses the modified Set Partitioning in Hierarchical Trees (SPIHT) coding algorithm to compress the coefficients from the wavelet packet transform. It indicates that it can gain better high compression. However the quality signal reconstruction is still not perfect according to loss of frequency. Thus this article is concerned to compare the frequency lossless of the new model wavelet for speech compression.

This article is organized as follows. Section II describes the Discrete Wavelet Transform while Section III details the speech compression process. Section IV shows its simulations results. Finally, Section V concludes this work.

### II. DISCRETE WAVELET TRANSFORM

The DWT is the most powerful and new signal compression technique which uses multi-resolution analysis for analyzing speech signals. Function of the DWT is a frequency scale adjustments and values shifting position discrete which able to calculate coefficients as the equation (1)

$$C_{a,k} = \int_{-\infty}^{\infty} s(x) \frac{1}{\sqrt{a}} \psi\left(\frac{x-k}{a}\right) dx \quad (1)$$

Where  $C_{a,k}$  is the coefficient which  $a$  is scale,  $k$  is shifting.

$\psi(x)$  is the mother wavelet function, which three families wavelet have selected for speech compression. These details are followed.

A. Haar wavelet (Haar)

Haar [3] is wavelet function which applies the easiest, using the least time and it has orthogonal property. The mother wavelet of Haar  $\psi(x)$  can be described as the equation (2).

$$\psi(x) = \begin{cases} 1 & 0 < t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

B. Biorthogonal wavelet (Bior)

Bior [3] or semi-orthogonal wavelet is only orthogonal to the shifted base function under different scale factor, but has no orthogonality in the same scale factor. The mother wavelet of Bior can be described as equation (3).

$$C_{a,k} = \int_R s(x) \tilde{\psi}_{a,k}(x) dx \quad (3)$$

C. Discrete approximation of Meyer wavelet (Dmey)

Dmey [3] wavelet is the discrete format of meyer wavelet function. Mayer's wavelet as shown in equation (4), is fundamentally a solvent method for solving the two-scale equation. Given a basis  $\phi$  for the approximation space, Mayer employed Fourier techniques to derive the DTF of the two-scale education coefficients.

$$G_0(e^{i\omega}) = \sqrt{2} \sum_k \phi(2\omega + 4k\pi) \quad (4)$$

III. SPEECH COMPRESSION PROCESS

All speech signals are selected from 40 male and female to be input which can be categorized as following.

- 5 seconds length of each 20 male and female speech signals.
- 60 seconds length of each 20 male and female speech signals.

The sampling rate of speech in put is 8 kHz and each sampling uses 8 bits. The speech compression process is shown in Fig 1.

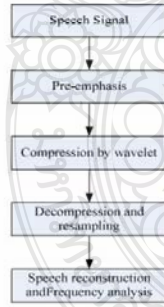


Figure 1. speech compression process

Firstly, speech signal is passed through the pre-emphasis technique to filter the back ground noise. The speech signal is then extracted the speech feature by using 2<sup>nd</sup> level of DWT which contains of Haar, Bior and Dmey. At this point the speech input is changed the sampling rate as 2 kHz. It means it reduces the original sampling rate to 4 times. The decompression speech is then process. However the new sampling rate is directly effect to the speech quality. It provides the poor quality. Therefore, it needs to re-sampling and analysis the losing frequency to improve the quality of speech using the equation (5) as followed.

$$x_{new} = \frac{p}{q} \times \text{sampling rate}_{new} \times t_{sec} \quad (5)$$

which,  $x_{new}$  is the reconstruct speech signal

$t_{sec}$  is the time

$\frac{p}{q}$  is the ratio of each sampling rate

The result of this technique is shown in Table I. To analysis the quality of reconstruction speech, the average energy is used to analyze in time-domain and frequency spectrum is used to analyze in frequency-domain.

A. The average energy

$$x = \frac{\sum_{i=1}^n |S_i|}{n} \quad (6)$$

which,  $x$  is the average energy of speech signal

$n$  is the number of signal

$S_i$  is the sampling rate

The results of the average energy are shown in Table 2-5.

B. Frequency spectrum

The frequency spectrum is used to analyze the losing frequency of reconstruction speech. At this point, the three wavelet technique was compared in the frequency domain, to find the best technique for lossless frequency. The results are shown in Fig 2.

IV. EXPERIMENT RESULT

From (6), the experiment uses 5 ratios as  $\frac{3}{2}, \frac{4}{2}, \frac{5}{2}, \frac{3}{1}$  and  $\frac{4}{1}$  to be the value of  $\frac{p}{q}$  to determine the appropriate value of sampling rate.

TABLE I. THE SPEECH SIGNAL COMPRESSION RATE

No.	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

The result shows that this technique provides the high compression rate at 2.67 when uses the  $p/q$  ratio as  $3/2$  and sampling rate at 3000 Hz. Thus the following experiment is chosen these parameters to be the primary parameters to test the performance of three wavelets.

TABLE II. EXAMPLE OF THE AVERAGE ENERGY VALUE OF 5 SECONDS MALE SPEECH SIGNAL

No.	Haar	Bior	Dmey
1	0.040	0.052	0.043
2	0.058	0.070	0.062
3	0.056	0.064	0.059
4	0.068	0.082	0.073
5	0.098	0.107	0.088
6	0.108	0.128	0.113
7	0.091	0.104	0.093
8	0.073	0.083	0.075
9	0.085	0.094	0.086
10	0.028	0.033	0.028

TABLE III. EXAMPLE OF THE AVERAGE ENERGY VALUE OF 60 SECONDS MALE SPEECH SIGNAL

No.	Haar	Bior	Dmey
1	0.085	0.098	0.089
2	0.050	0.057	0.052
3	0.072	0.082	0.074
4	0.055	0.061	0.072
5	0.060	0.067	0.061
6	0.050	0.055	0.050
7	0.079	0.089	0.079
8	0.054	0.065	0.054
9	0.047	0.053	0.048
10	0.032	0.037	0.031

TABLE IV. EXAMPLE OF THE AVERAGE ENERGY VALUE OF 5 SECONDS FEMALE SPEECH SIGNAL

No.	Haar	Bior	Dmey
1	0.118	0.130	0.122
2	0.163	0.176	0.166
3	0.176	0.189	0.164
4	0.191	0.206	0.195
5	0.197	0.214	0.202
6	0.189	0.203	0.193
7	0.196	0.215	0.202
8	0.164	0.174	0.166
9	0.156	0.169	0.160
10	0.165	0.178	0.169

TABLE V. EXAMPLE OF THE AVERAGE ENERGY VALUE OF 60 SECONDS FEMALE SPEECH SIGNAL

No.	Haar	Bior	Dmey
1	0.158	0.171	0.163
2	0.186	0.201	0.191
3	0.183	0.198	0.188
4	0.181	0.196	0.186
5	0.204	0.222	0.210
6	0.197	0.214	0.202
7	0.164	0.177	0.168
8	0.196	0.217	0.203
9	0.184	0.200	0.190
10	0.186	0.204	0.192

It can be seen that the Bior wavelet provides the best performance in time domain when is compared to the all wavelet. However, it is also necessary to analyze the quality of speech signal in frequency domain. The results of frequency analysis are shown in Fig 2.

Fig.2 shows the comparison of the quality speech signal based on frequency domain. The results show that the Bior wavelet technique provides the best performance in frequency domain when was compared to the all wavelet. It provides the lossless frequency speech signal when is compared to the original speech signal.

## V. CONCLUSION

This article presents the frequency lossless comparison of the modern wavelets based on speech compression technique. The Haar wavelet (Haar), Biorthogonal wavelet (Bior) and discrete approximation of Meyer wavelet (Dmey), are investigated to test with eighty speech signal. The experiments show that the Bior provides the best performance in time and frequency domain when is compared to the all wavelet. This technique can also gain the compression rate as 2.67 to the original speech signal. However Haar and Dmey do not provide the good performance because Haar and Dmey are unable to include a value changes immediately. Also the conversion coefficients of Haar and Dmey wavelet functions cannot be separate the data between low frequency and high frequency from each other effectively. Thus it wasn't suitable for the application on a compression and the noise removal.

## ACKNOWLEDGMENT

We would like to thank the Office of National Research Council of Thailand for financial support (2012) in this research.

## REFERENCES

- [1] S. M. Joseph, "Spoken digit compression using wavelet packet," International conference on signal and image processing, PP: 255-259, 2010.
- [2] M. A. Osman and N. Al, "Speech compression using LPC and wavelet," 2<sup>nd</sup> International conference on computer engineering and technology, PP: V7-92 - V7-99, 2010.
- [3] J. Y. Zhang and L. X. Gao, "Research on the selection of wavelet function for the feature extraction of shock fault in the bearing diagnosis," International conference on wavelet analysis and pattern recognition, PP: 1630-1634, 2007.
- [4] R.s.Cai and Y.M.Guo, "Wavelet based multi feature V/UV speech classification algorithm," International conference wireless mobile and sensor networks, PP: 897 - 900, 2007.
- [5] S. Ranjan, "Exploring the DWT as a tool for Hindi speech recognition," International journal of computer theory and engineering, PP: 642-646, 2010.
- [6] J. Karam, "End point detection for wavelet based speech compression," International journal of biological and life sciences, PP: 167-170, 2008.
- [7] Z. Dan and M. S. Qian, "Speech compression with best wavelet package transform and SPIHT algorithm," 2<sup>nd</sup> International conference on computer modeling and simulation, PP: 360-363, 2010.



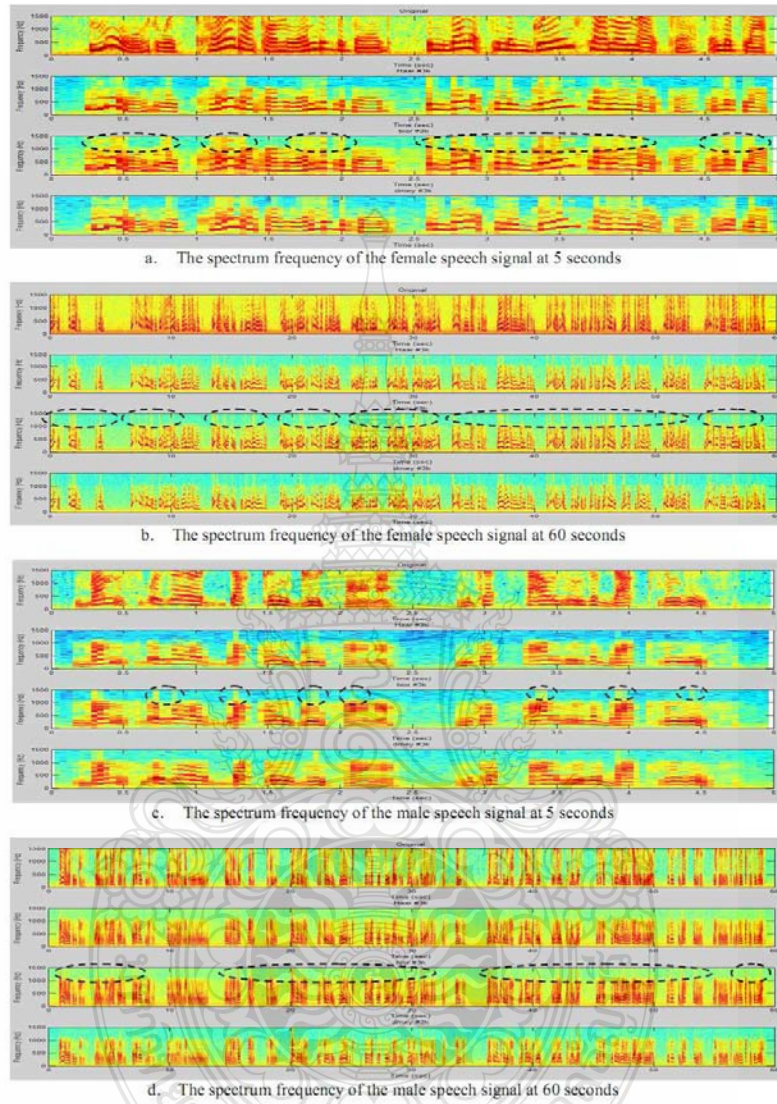


Figure 2 The frequency spectrum speech signal

## ประวัติผู้เขียน

ชื่อ – นามสกุล	นางสาวสุภาวณิ กรสิงห์
วัน เดือน ปีเกิด	13 พฤษภาคม 2529
ที่อยู่	63 หมู่ 16 ตำบล นิคมทุ่งโพธิ์ทะเลพิทยา อำเภอ เมือง จังหวัด กำแพงเพชร 62000
การศึกษา	
2553 – ปัจจุบัน	กำลังศึกษาปริญญาโท คณะวิศวกรรมศาสตร์ สาขาวิชา วิศวกรรมไฟฟ้า แขนง วิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคล ธัญบุรี
2549 – 2552	ปริญญาตรี คณะวิศวกรรมศาสตร์ สาขาวิชา วิศวกรรมอิเล็กทรอนิกส์และ โทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
2547 – 2549	ประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาอิเล็กทรอนิกส์ วิทยาลัยเทคนิค กำแพงเพชร
2544 – 2547	ประกาศนียบัตรวิชาชีพ สาขาวิชาอิเล็กทรอนิกส์ วิทยาลัยเทคนิคกำแพงเพชร

