

การพัฒนา LVO สำหรับการบีบอัดสัญญาณเสียง

ENHANCEMENT LVO FOR SPEECH COMPRESSION



เกรียงศักดิ์ พัฒนบุรี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2555

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

การพัฒนา LVO สำหรับการบีบอัดสัญญาณเสียง



เกรียงศักดิ์ พัฒนบุรี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

ปีการศึกษา 2555

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

หัวข้อวิทยานิพนธ์ การพัฒนา LVQ สำหรับการบีบอัดสัญญาณเสียง
Enhancement LVQ for Speech Compression
ชื่อ - นามสกุล นายเกรียงศักดิ์ พัฒนบุรี
สาขาวิชา วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา อาจารย์จักรี ศรีนนท์ฉัตร, Ph.D.
ปีการศึกษา 2555

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการ

.....
(อาจารย์อำนวยการ เรืองวารีย์, Dr.-Ing)

กรรมการ

.....
(รองศาสตราจารย์ ดร.สมเกียรติ อุดมพระยากุล)

กรรมการ

.....
(อาจารย์สุรินทร์ แห่งงาม, Ph.D.)

กรรมการ

.....
(อาจารย์จักรี ศรีนนท์ฉัตร, Ph.D.)

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี อนุมัติวิทยานิพนธ์ฉบับนี้
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(ผู้ช่วยศาสตราจารย์สมหมาย ผิวสอาด, Ph.D.)

วันที่ 7 เดือน ตุลาคม พ.ศ. 2555

| | |
|-------------------|--|
| หัวข้อวิทยานิพนธ์ | การพัฒนา LVO สำหรับการบีบอัดสัญญาณเสียงพูด |
| ชื่อ - นามสกุล | นายเกรียงศักดิ์ พัฒนบุรี |
| สาขาวิชา | วิศวกรรมไฟฟ้า |
| อาจารย์ที่ปรึกษา | อาจารย์จักรี ศรีนนท์นัทร, Ph.D. |
| ปีการศึกษา | 2555 |

บทคัดย่อ

จุดประสงค์ของการบีบอัดสัญญาณเสียงคือการลดขนาดของสัญญาณเสียงให้มีขนาดเล็กที่สุด และขณะเดียวกันจะต้องรักษาคุณภาพของสัญญาณเสียงให้มีคุณภาพสูงสุด วิทยานิพนธ์ฉบับนี้ได้การนำเสนอการเพิ่มประสิทธิภาพการบีบอัดเสียงด้วยการนำเทคนิคเครือข่ายการเรียนรู้เวกเตอร์ควอนไทซ์เซชัน มาประยุกต์ใช้ในกระบวนการเข้ารหัสเสียงพูด เช่น LPC-10 และ LSP-10 เพื่อเพิ่มอัตราการบีบอัดสัญญาณเสียงพูด

การทดลองเริ่มจากการเตรียมสัญญาณเสียงพูดที่มาวิเคราะห์ ซึ่งใช้ตัวอย่างเสียงพูด จำนวน 10 เสียงพูด ผู้ชาย 5 คน ผู้หญิง 5 คน โดยการบันทึกสัญญาณเสียงผ่านไมโครโฟนด้วยอัตราการสุ่มตัวอย่าง 8 กิโลเฮิรตซ์ต่อวินาที จากนั้นนำสัญญาณเสียงพูดวิเคราะห์หาค่าสัมประสิทธิ์ LPC-10 และ LSP-10 ตามลำดับ เมื่อได้ค่าสัมประสิทธิ์ LPC-10 และ LSP-10 นำค่าสัมประสิทธิ์ทั้งสองมาบีบอัดด้วยเทคนิคเครือข่ายการเรียนรู้เวกเตอร์ควอนไทซ์เซชันเพื่อเปรียบเทียบประสิทธิภาพของการบีบอัดสัญญาณเสียงโดยใช้หลักการของ อัตราส่วนของการแทนค่าข้อมูลที่ขาดหายไป อัตราส่วนสัญญาณต่อสัญญาณรบกวน ต่อสัญญาณรบกวนสูงสุด และ อัตราส่วนสัญญาณต่อสัญญาณรบกวน เป็นตัววัดผลการทดสอบ

ผลทดสอบการบีบอัดค่าสัมประสิทธิ์ของ LPC-10 และ LSP-10 โดยการวัดประสิทธิภาพด้วยหลักการของอัตราส่วนของการแทนค่าข้อมูลที่ขาดหายไป อัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงสุด และ อัตราส่วนสัญญาณต่อสัญญาณรบกวน พบว่า การบีบอัดค่าสัมประสิทธิ์ LSP-10 เป็นการบีบอัดที่ดีที่สุด โดยรายละเอียดของการบีบอัดค่าสัมประสิทธิ์ LSP-10 จะมีค่า NRMSE ต่ำสุดเท่ากับ 0.0111 ค่า PSNR สูงสุดเท่ากับ 41.0208 dB ค่า SNR สูงสุดเท่ากับ 36.6372 dB

คำสำคัญ : การประมวลสัญญาณเสียง การบีบอัดสัญญาณเสียงพูด เครือข่ายการเรียนรู้เวกเตอร์ควอนไทซ์เซชัน

| | |
|----------------|--|
| Thesis Title | Enhancement LVQ for speech compression |
| Name - Surname | Mr. Kreangsak Pattanaburi |
| Program | Electrical Engineering |
| Thesis Advisor | Mr. Jakkree Srinonchat, Ph.D. |
| Academic Year | 2012 |

ABSTRACT

The speech compression aims to compress speech signal into as small an amount of information as possible while maintaining the speech quality in as high a level as possible. Therefore, this thesis presents the enhancement efficiency of speech compression using Learning Vector Quantization (LVQ) technique, which it is applied to speech compression such as Linear Predictive Coefficient order 10 (LPC-10) and Linear Spectral Pairs order 10 (LSP-10) for increasing compression rate of speech signal.

In the experiment, 10 speech signals which corrected from 5 male and 5 female, are used as the input speech signal. These speech signals are record using microphone with sampling rate at 8 kHz. These speech signal are then analyzed and calculated the LPC-10 and LSP-10 respectively. The LVQ is then used to compress both coefficients. Also the Normalized Root Mean Squared Error (NRMSE) is used to measure the error coding. Moreover, Peak Signal to Noise Ratio (PSNR) and Signal to Noise Ratio (SNR) are used to measure the synthesis speech quality.

The results of LPC-10 and LSP-10 coefficients compression in the term of NRMSE, PSNR and SNR show that LVQ-LSP-10 provides the best compression. It can notice that the minimum NRMSE is 0.0111 and the maximum PSNR is equal to 41.0208 dB. Also the maximum SNR is 36.6372 dB

Keywords : speech processing, speech compression, learning vector quantization

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ ดร.จักรี ศรีนนท์ฉัตร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งได้ให้คำแนะนำข้อคิดเห็นและสนับสนุนในการทำวิจัย มาด้วยดีตลอด ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

กราบขอบพระคุณ คณะกรรมการสอบวิทยานิพนธ์ ดร.อำนาจ เรืองวาริ ดร.สุรินทร์ แหงมงาม ดร.อภิรดา นามแสง และผู้ทรงคุณวุฒิ รองศาสตราจารย์ ดร.สมเกียรติ อุดมพระยากุล ที่ให้คำแนะนำความรู้และประสบการณ์ที่เป็นแนวทางในการทำวิทยานิพนธ์นี้จนสำเร็จลุล่วงด้วยดี และเป็นประโยชน์ในทำการวิจัยครั้งต่อไป

ขอขอบคุณรุ่นพี่ เพื่อน น้องนิสิตที่ห้องปฏิบัติการและวิจัยทางด้านการประมวลผลสัญญาณ ที่ได้ช่วยเหลือเกี่ยวกับข้อมูล รวมถึงคำแนะนำต่างๆ ตลอดเวลาที่ทำวิจัยอย่างดียิ่ง

ท้ายนี้ผู้วิจัยกราบขอบพระคุณ บิดามารดาที่ให้การสนับสนุนแก่ผู้วิจัยเสมอมาจนสำเร็จ การศึกษา

เกรียงศักดิ์ พัฒนบุรี



สารบัญ

| | หน้า |
|---|------|
| บทคัดย่อภาษาไทย..... | ค |
| บทคัดย่อภาษาอังกฤษ..... | ง |
| กิตติกรรมประกาศ..... | จ |
| สารบัญ | ฉ |
| สารบัญตาราง..... | ช |
| สารบัญภาพ..... | ฉ |
| บทที่ | |
| 1 บทนำ | |
| 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 การเข้ารหัสสัญญาณเสียงพูด..... | 3 |
| 2.2 พารามิเตอร์จำลองสัญญาณเสียง..... | 4 |
| 2.3 เวกเตอร์คอสไทซ์เซชัน..... | 13 |
| 2.4 เครื่องมือการเรียนรู้การคอสไทซ์เซชัน..... | 15 |
| 2.5 การวัดความผิดพลาด..... | 17 |
| 2.6 การวัดค่าประสิทธิภาพของสัญญาณ..... | 18 |
| 2.7 การประเมินคุณภาพของข้อมูลเสียงพูด..... | 19 |
| 2.8 งานวิจัยที่เกี่ยวข้อง..... | 20 |
| 3 วิธีการดำเนินงานวิจัย | 23 |
| 3.1 การเก็บบันทึกสัญญาณเสียงพูด..... | 24 |
| 3.2 การแบ่งสัญญาณเสียง..... | 25 |
| 3.3 การสกัดพารามิเตอร์จำลองสัญญาณเสียง..... | 26 |
| 3.4 การออกแบบโครงข่าย LVO..... | 28 |

สารบัญ (ต่อ)

| บทที่ | หน้า |
|---|------|
| 3.5 การวัดค่าผิดพลาด..... | 37 |
| 3.6 การวัดค่าประสิทธิภาพของสัญญาณ..... | 37 |
| 4 ผลการวิจัย..... | 39 |
| 4.1 ขนาดการสร้างโค้ดบิต..... | 39 |
| 4.2 ผลทดสอบการหาค่าผิดพลาด..... | 40 |
| 4.3 การหาประสิทธิภาพของการบีบอัดสัญญาณเสียงพูด..... | 44 |
| 4.4 ระยะเวลาในการฝึกสอน (Training Time)..... | 52 |
| 4.5 สเปกโตรแกรม (Spectrogram)..... | 56 |
| 4.6 ผลการบีบอัดตามขนาดของชุดโค้ดบิต..... | 60 |
| 4.6 ผลการประเมินคุณภาพเสียงพูด (MOS)..... | 58 |
| 5 สรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะ..... | 63 |
| 5.1 สรุปผลการวิจัย..... | 63 |
| 5.2 ข้อเสนอแนะและแนวทางในการพัฒนา..... | 64 |
| รายการอ้างอิง..... | 65 |
| ภาคผนวก..... | 66 |
| ภาคผนวก ก ผลงานตีพิมพ์เผยแพร่..... | 67 |
| ประวัติผู้เขียน..... | 97 |

สารบัญตาราง

| ตารางที่ | หน้า |
|--|------|
| 2.1 ค่า MOS ที่เหมาะสมกับการใช้งานในระบบต่างๆ..... | 19 |
| 2.2 รายละเอียดวิธีการให้คะแนนในการวัด MOS..... | 19 |
| 4.1 การแบ่งระดับของโค้ดบู้ค..... | 39 |
| 4.2 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 40 |
| 4.3 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 41 |
| 4.4 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 42 |
| 4.5 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 43 |
| 4.6 ค่า PSNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 44 |
| 4.7 ค่า PSNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 45 |
| 4.8 ค่า PSNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 46 |
| 4.9 ค่า PSNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 47 |
| 4.10 ค่า SNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 48 |
| 4.11 ค่า SNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 49 |
| 4.12 ค่า SNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 50 |
| 4.13 ค่า SNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 51 |
| 4.14 ผลการบีบอัดตามระดับขนาดของโค้ดบู้ค..... | 60 |
| 4.15 การประเมินคุณภาพเสียงจากการรับฟัง..... | 61 |

สารบัญภาพ

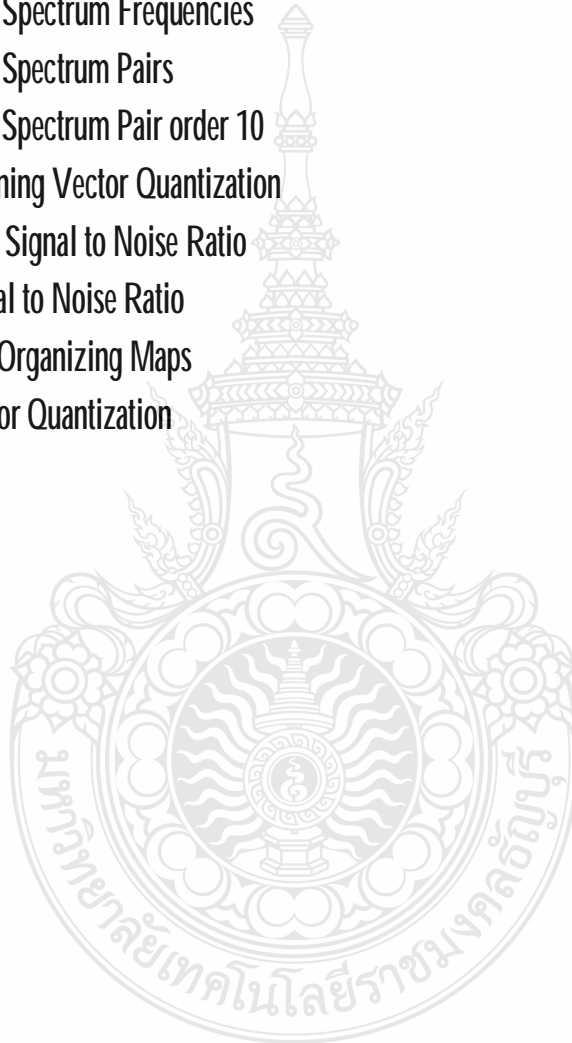
| ภาพที่ | หน้า | |
|--------|--|----|
| 2.1 | คุณภาพของการเข้ารหัสสัญญาณเสียงพูดในแต่ละวิธี..... | 4 |
| 2.2 | การวางเรียงสลับของรากของพหุนามคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ | 11 |
| 2.3 | ความสัมพันธ์ระหว่างรากของ $A(z)$ กับรากของคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$.. | 12 |
| 2.4 | การทำงานของเวกเตอร์ควอนไทซ์เซชัน..... | 13 |
| 2.5 | ตัวอย่างของการควอนไทซ์ในระนาบ 2 มิติ..... | 14 |
| 2.6 | โครงสร้างเครือข่าย LVO..... | 16 |
| 2.7 | การเรียนรู้ของเครือข่าย LVO..... | 17 |
| 3.1 | แผนผังขั้นตอนการดำเนินการวิจัย..... | 23 |
| 3.2 | การตั้งค่าเริ่มต้นบันทึกเสียงในโปรแกรม GoldWave..... | 24 |
| 3.3 | ตัวอย่างสัญญาณเสียงพูดในกลุ่มผู้พูดที่ใช้ในการวิเคราะห์..... | 25 |
| 3.4 | การแบ่งสัญญาณเสียงออกเป็น Frame..... | 25 |
| 3.5 | การหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้นและค่าสัมประสิทธิ์คู่เส้นสเปกตรัม..... | 26 |
| 3.6 | พารามิเตอร์ LPC-10..... | 27 |
| 3.7 | พารามิเตอร์ LSP-10..... | 28 |
| 3.8 | การแบ่งกลุ่มของค่าสัมประสิทธิ์ LPC-10..... | 29 |
| 3.9 | Hidden layer ในการบีบอัดพารามิเตอร์ LPC-10..... | 30 |
| 3.10 | การดำเนินการฝึกฝนข้อมูลของโครงข่าย LVO สำหรับการบีบอัดพารามิเตอร์ LPC-10..... | 31 |
| 3.11 | จุดสิ้นสุดของกระบวนการฝึกฝนข้อมูลโครงข่าย LVO ในการบีบอัดพารามิเตอร์ LPC-10..... | 32 |
| 3.12 | ชุดสมาชิกรหัสของการบีบอัดพารามิเตอร์ LPC-10 ในไฟล์ Excel..... | 32 |
| 3.13 | การแบ่งกลุ่มของค่าสัมประสิทธิ์ LSP-10..... | 33 |
| 3.14 | Hidden layer ของการบีบอัดพารามิเตอร์ LSP-10..... | 34 |
| 3.15 | การดำเนินการฝึกฝนข้อมูลของโครงข่าย LVO..... | 35 |
| 3.16 | จุดสิ้นสุดของกระบวนการฝึกฝนข้อมูลโครงข่าย LVO ในการบีบอัดพารามิเตอร์ LSP-10..... | 36 |

สารบัญภาพ (ต่อ)

| ภาพที่ | หน้า |
|--|------|
| 3.17 ชุดสมาชิกรหัสของการบีบอัดพารามิเตอร์ LSP-10 ในไฟล์ Excel..... | 36 |
| 3.18 การวัดค่าผิดพลาดด้วยวิธี NRMSE..... | 37 |
| 3.19 การหาค่าประสิทธิภาพของสัญญาณ PSNR และ SNR..... | 38 |
| 4.1 ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 52 |
| 4.2 ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 53 |
| 4.3 ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน..... | 54 |
| 4.4 ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน..... | 55 |
| 4.5 สเปคโตรแกรมของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย..... | 56 |
| 4.6 สเปคโตรแกรมของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง..... | 57 |
| 4.7 สเปคโตรแกรมของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย..... | 58 |
| 4.8 สเปคโตรแกรมของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง..... | 59 |

คำอธิบายสัญลักษณ์และคำย่อ

| | |
|--------|--|
| KSOFM | Kohonen Self Organizing Feature Maps |
| LPC | Linear Predictive Coefficient |
| LPC-10 | Linear Predictive Coefficient order 10 |
| LSF | Line Spectrum Frequencies |
| LSP | Line Spectrum Pairs |
| LSP-10 | Line Spectrum Pair order 10 |
| LVO | Learning Vector Quantization |
| PSNR | Peak Signal to Noise Ratio |
| SNR | Signal to Noise Ratio |
| SOM | Self Organizing Maps |
| VQ | Vector Quantization |



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันในการสื่อสารสัญญาณเสียงพูด (Speech Signal) [1] ของระบบโทรคมนาคมได้ให้ความสำคัญกับการส่งสัญญาณเสียงพูดผ่านช่องสัญญาณที่มีขนาดจำกัด ส่งผลให้เกิดเทคโนโลยีดิจิทัลใหม่ๆ ขึ้นมาเพื่อตอบสนองความต้องการ ดังนั้น เมื่อต้องการจะส่ง (Transmitted) สัญญาณเสียงพูด สิ่งที่จะต้องคำนึงถึงคือความกว้างแบนด์วิดท์ (Bandwidth) ของช่องสัญญาณเสียงพูดและการลดอัตราการส่งข้อมูล (Bit Rate) ของสัญญาณเสียงพูด ด้วยเหตุนี้จึงต้องทำการลดขนาดข้อมูลของสัญญาณเสียงพูดลง (Reduce Bit Rate) [2] โดยการบีบอัดสัญญาณเสียงพูด (Speech Compression) สิ่งจำเป็นที่ต้องใช้ในการบีบอัดสัญญาณเสียงพูดมักพบได้ในงานวิจัยการเข้ารหัสสัญญาณเสียงพูด (Speech Coding) ซึ่งได้มีการพัฒนาเทคนิคต่างๆ ขึ้นมาให้เป็นมาตรฐาน ซึ่งแต่ละมาตรฐานจะมีข้อดีและข้อเสียแตกต่างกัน โดยในงานวิจัยเกี่ยวข้องกับการเข้ารหัสสัญญาณเสียงพูดที่มีอัตราการส่งข้อมูลต่ำๆ ส่วนใหญ่นิยมใช้รูปแบบการเข้ารหัสแบบเชิงเส้น (Linear Predictive Coder: LPC) และการเข้ารหัสแบบคู่เส้นสเปกตรัม (Line Spectrum Pairs: LSP) จากศึกษาพบว่า (Learning Vector Quantization: LVQ) สามารถบีบอัดสัญญาณของการเข้ารหัสแบบเชิงเส้น และ เข้ารหัสแบบคู่เส้นสเปกตรัม เพิ่มเติมได้

LVQ คือเทคนิคเวกเตอร์ควอนไทซ์เซชันรูปแบบหนึ่ง จากงานวิจัยที่เคยพัฒนาผ่านมาพบว่า LVQ เหมาะสำหรับการนำไปประยุกต์ใช้กับบีบอัดสัญญาณหรือคัดแยกสัญญาณ สำหรับ LVQ มีหลักการสำคัญสองส่วนด้วยกัน ส่วนแรก คือ เวกเตอร์อินพุตที่ใช้สำหรับการบีบอัด และส่วนที่สองคือ การกำหนดกลุ่มที่ใช้สัญญาณเพื่อใช้เป็นส่วนหนึ่งในการฝึกสอน โดยที่ก่อนการบีบอัดสัญญาณหรือคัดแยกสัญญาณใดๆ ก็ตาม จะต้องมีการจัดกลุ่มของสัญญาณก่อน สำหรับในงานวิจัยนี้ก็เช่นกันต้องนำพารามิเตอร์ LPC-10 และพารามิเตอร์ LSP-10 มาจัดกลุ่มข้อมูลก่อน แล้วจึงนำพารามิเตอร์ไปฝึกสอนด้วย LVQ เพื่อบีบอัดสัญญาณเสียง

ดังนั้นผู้วิจัยจึงได้ศึกษาเทคนิควิธีการบีบอัดสัญญาณเสียงพูด โดยนำสัญญาณเสียงพูดวิเคราะห์หาพารามิเตอร์ LPC-10 และ LSP-10 ตามลำดับ แล้วใช้ LVQ ทำการลดค่าอัตราการส่งของพารามิเตอร์ทั้งสอง โดยในลดค่าอัตราการส่งพารามิเตอร์ที่สร้างขึ้นยังคงคุณลักษณะที่ใกล้เคียงกับพารามิเตอร์ต้นฉบับไว้

1.2 วัตถุประสงค์ของการศึกษา

1.2.1 ศึกษาวิธีการบีบอัดเสียงโดยใช้ LVO

1.2.2 พัฒนาเทคนิคในการเข้ารหัสเพื่อลดบิตในการส่งสัญญาณ

1.3 ขอบเขตของการศึกษา

1.3.1 ศึกษาและพัฒนากการเข้ารหัสสัญญาณเสียงบนพื้นฐานของ LVO โดยลดอัตราการบีบอัดให้น้อยลงไปอีก 15 % เมื่อเทียบกับพารามิเตอร์ LPC-10

1.3.2 ศึกษาและพัฒนากการเข้ารหัสสัญญาณเสียงบนพื้นฐานของ LVO โดยลดอัตราการบีบอัดให้น้อยลงไปอีก 15 % เมื่อเทียบกับพารามิเตอร์ LSP-10

1.3.3 วิเคราะห์คุณภาพของสัญญาณเสียงพูดหลังบีบอัด

1.4 ข้อจำกัดของการศึกษา

1.4.1 ประมวลผลโดยใช้โปรแกรม MATLAB

1.4.2 เสียงที่นำมาวิจัยต้องเป็นสัญญาณเสียงพูดเท่านั้น

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้เรียนรู้การบีบอัดเสียงพูดโดยใช้ LVO

1.5.2 เป็นแนวทางในการพัฒนากการบีบอัดสัญญาณเสียง เพื่อลดขนาดของข้อมูลและสามารถนำไปประยุกต์ใช้ในระบบการสื่อสารในอนาคต

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดของทฤษฎีและความรู้ต่างๆ ที่ได้ศึกษาและนำมาใช้ในวิทยานิพนธ์ ประกอบด้วยในเรื่องของการเข้ารหัสสัญญาณเสียง (Speech Coding) การวิเคราะห์สัญญาณเสียงพูด เวกเตอร์ควอนไทซ์เซชัน (Vector Quantization) เครื่องมือการเรียนรู้เวกเตอร์ควอนไทซ์เซชัน (Learning Vector Quantization) การวัดค่าผิดพลาด (Error) ของการเข้ารหัส การวัดค่าประสิทธิภาพของสัญญาณ และสุดท้ายจะกล่าวถึงงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์นี้

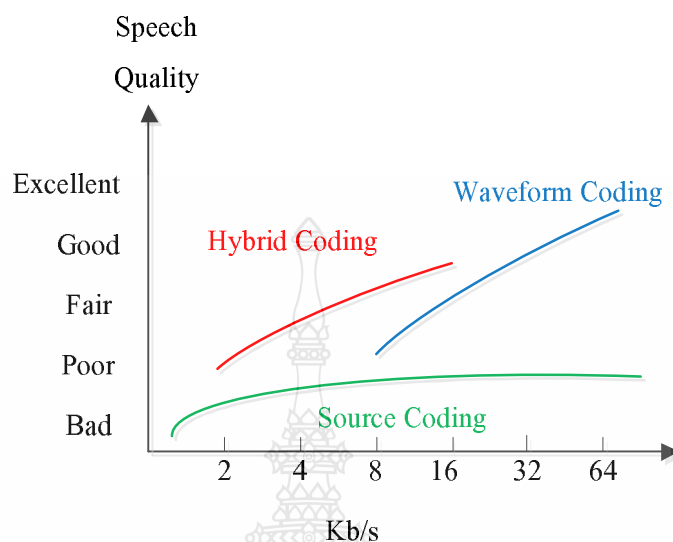
2.1 การเข้ารหัสสัญญาณเสียงพูด

การรับส่งสัญญาณข้อมูลข่าวสารในการสื่อสารโทรคมนาคมได้ให้ความสำคัญกับการส่งข่าวสารผ่านช่องสัญญาณที่มีขนาดจำกัดอย่างมีประสิทธิภาพ เนื่องจากในปัจจุบันและอนาคตอันใกล้ปริมาณความต้องการเข้าใช้งานของผู้บริโภคมีแนวโน้มสูงขึ้น เพื่อตอบสนองความต้องการดังกล่าวจึงจำเป็นต้องมีการเพิ่มความจุของช่องสัญญาณ ซึ่งวิธีการเข้ารหัสหรือการบีบอัดสัญญาณเสียงเป็นวิธีการในการช่วยเพิ่มความจุของช่องสัญญาณให้สามารถใช้งานช่องสัญญาณที่มีแบนด์วิดท์จำกัดได้อย่างมีประสิทธิภาพ

การบีบอัดสัญญาณเสียงพูดได้มีการวิจัยและพัฒนาเทคนิคต่างๆ เพื่อที่ทำให้สัญญาณเสียงพูดนั้นมีขนาดที่เล็กที่สุด สำหรับการลดขนาดพื้นที่ในการเก็บและลด แบนด์วิดท์ในการรับ-ส่งสัญญาณ โดยจะรักษาคุณภาพของสัญญาณเสียงพูดให้มีประสิทธิภาพใกล้เคียงกับต้นฉบับให้มากที่สุด กรรมวิธีส่วนใหญ่ที่ใช้กับการบีบอัดเสียงจะอยู่ในรูปแบบของการเข้ารหัสสัญญาณเสียงพูด ซึ่งได้มีการพัฒนาเทคนิคต่างๆ ขึ้นให้เป็นมาตรฐานเพื่อใช้ในการสื่อสาร ซึ่งการเข้ารหัสสัญญาณเสียงพูดโดยทั่วไปนั้นแบ่งออกได้เป็น 3 วิธี [3] ดังนี้

- 1) **Waveform Coding** เป็นวิธีที่เหมาะสมสำหรับบิตเรตสูงและให้คุณภาพของเสียงที่ดี
- 2) **Source Coding** จะเป็นวิธีที่มีอัตราบิตเรตต่ำกว่า 8 กิโลบิตต่อวินาที ใช้สำหรับเพื่อสร้างสัญญาณเสียงพูดเชิงสังเคราะห์
- 3) **Hybrid Coding** จะเป็นการนำเทคนิคทั้งสองวิธีของ **Waveform Coding** และ **Source Coding** มาใช้ร่วมกัน และให้คุณภาพสัญญาณที่ดีในระดับที่มีอัตราบิตเรตระดับกลาง 8-16 กิโลบิตต่อวินาที

ซึ่งรายละเอียดของคุณภาพแต่ละวิธีแสดงดังในภาพที่ 2.1



ภาพที่ 2.1 คุณภาพของการเข้ารหัสสัญญาณเสียงพูดในแต่ละวิธี [3]

การเข้ารหัสสัญญาณเสียงพูดในแต่ละวิธีมีความเหมาะสมที่แตกต่างกัน แต่สำหรับการบีบอัดสัญญาณเสียงพูดที่มีบิตเรตต่างๆ นิยมใช้วิธีสัมประสิทธิ์การทำนายพัลซเซิงเส้น และสัมประสิทธิ์คู่เส้นสเปกตรัม (Line Spectrum Pair: LSP) ซึ่งเป็นการเข้ารหัสสัญญาณเสียงพูดในรูปแบบของ Source Coding [4]

2.2 พารามิเตอร์จำลองสัญญาณเสียง

2.2.1 สัมประสิทธิ์การทำนายพัลซเซิงเส้น

การทำนายพัลซเซิงเส้น (Linear Predictive) เป็นเทคนิคที่สำคัญทางการวิเคราะห์สัญญาณเสียงพูด เนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของสัญญาณเสียงพูดเมื่อเทียบกับความเร็วในการประมวลผล หลักการพื้นฐานของการทำนายพัลซเซิงเส้นอาศัยแนวความคิดว่าตัวอย่างสัญญาณเสียงพูดสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณเสียงพูดจากอดีต การวิเคราะห์หาพารามิเตอร์เพื่อใช้ในการทำนายโดยทั่วไปเรียกว่าการเข้ารหัสการทำนายพัลซเซิงเส้น (LPC) ในด้านการประมวลผลสัญญาณเสียง การเข้ารหัสการทำนายพัลซเซิงเส้นนี้ถูกนำไปใช้ในสองแนวทาง [4] ได้แก่

1) การเข้ารหัสสัญญาณเสียง โดยถูกนำไปใช้เป็นวงจรกรองวิเคราะห์การทำนายพันธะเชิงเส้น (LP Analysis Filter) เพื่อแยกส่วนซ้ำซ้อนของสัญญาณเสียงออก ส่วนที่เหลือเรียกว่าสัญญาณตกค้าง (Residual Signal)

2) การสังเคราะห์สัญญาณเสียงพูด โดยถูกนำไปใช้เป็นวงจรกรองการทำนายพันธะเชิงเส้นผกผัน (Inverse LP Filter) หรือวงจรกรองสังเคราะห์การทำนายพันธะเชิงเส้น (LP Synthesis Filter) โดยที่ฟังก์ชันถ่ายโอนของวงจรกรองดังกล่าวแสดงกรอบสเปกตรัมของสัญญาณเสียงพูด วงจรกรองสังเคราะห์การทำนายพันธะเชิงเส้นถูกใช้แสดงแทนช่องทางเดินเสียงของมนุษย์และใช้หาสัญญาณกระตุ้นที่เหมาะสม

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

$$\tilde{s}_n = -a_1 s_{n-1} - a_2 s_{n-2} - a_3 s_{n-3} - \dots - a_p s_{n-p} = -\sum_{k=1}^p a_k s_{n-k} \quad (2.1)$$

เมื่อ \tilde{s}_n คือ สัญญาณเสียงตัวอย่างปัจจุบัน
 s_n คือ สัญญาณเสียงเอาต์พุตตัวอย่างในช่วงเวลานั้นๆ
 n คือ จำนวนตัวอย่างที่อยู่ในแต่ละเฟรม โดยที่ $n = 0, 1, \dots, N-1$
 p คือ อันดับของการวิเคราะห์การทำนายพันธะเชิงเส้น
 a_1, a_2, \dots, a_p คือ สัมประสิทธิ์การเข้ารหัสการทำนายพันธะเชิงเส้น

กำหนดให้ e_n แทนค่าผิดพลาดระหว่างค่าจริงและค่าที่ทำนายได้ จะได้ตามสมการที่ 2.2 และสมการที่ 2.3

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

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

จากสมการที่ 2.3 จะเห็นว่าค่าสัญญาณ e_n เป็นผลมาจากการลบสัญญาณ \tilde{s}_n ด้วยสัญญาณ s_n จึงอาจเรียกสัญญาณ e_n นี้ว่าเป็นสัญญาณตกค้างได้ด้วยเช่นกัน และเนื่องด้วยค่าสหสัมพันธ์

ช่วงสั้น (Short-Term Correlation) ระหว่างตัวอย่างของสัญญาณตกค้างมีค่าต่ำ ดังนั้นประมาณได้ว่า กรอบสเปกตรัมกำลังของสัญญาณตกค้างมีลักษณะเรียบ เมื่อทำการแปลงแซด (Z Transform) สมการที่ 2.3 จากจะได้ค่าดังสมการที่ 2.4

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

เมื่อ $S(z)$ คือ ผลการแปลงแซดของสัญญาณเสียงพูด

$E(z)$ คือ ผลการแปลงแซดของสัญญาณตกค้าง

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

เมื่อ $A(z)$ คือ วงจรกรองแยกค่าสหสัมพันธ์ช่วงสั้นที่ปรากฏในสัญญาณเสียงพูด

ผลจากสมการที่ 2.5 วงจรกรอง $A(z)$ จะได้ค่าสเปกตรัมที่เรียบ เนื่องจาก $E(z)$ ประมาณได้ว่ามีสเปกตรัมเรียบ ดังนั้นสามารถออกแบบกรอบสเปกตรัมของสัญญาณช่วงสั้นได้จากการวิเคราะห์การทำนายพันธะเชิงเส้นในรูปแบบออลโพล (All-Pole Model) หรือรูปแบบออโตรีเกรสซีฟ (Autoregressive Model) ดังสมการที่ 2.6

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

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

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

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

โดยผลรวมของช่วง n_1 ถึง n_2 ที่คำนวณได้ขึ้นอยู่กับวิธีการที่ใช้ ซึ่งมีอยู่ 2 วิธี ได้แก่

1) วิธีอัตโนมัติสัมพันธ์ (Autocorrelation) ในการวิเคราะห์การทำนายพันธะเชิงเส้นช่วงสั้นสามารถหาได้โดยใช้การวิเคราะห์แบบหน้าต่างสัญญาณเสียงพูดโดยกำหนดให้ตัวอย่างสัญญาณภายนอกหน้าต่างนี้มีค่าเท่ากับศูนย์ ดังสมการที่ 2.8 จึงทำให้ได้ค่าผิดพลาดตามสมการที่ 2.7 มีค่าต่ำสุด

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

เมื่อ r_k คือ ค่าสัมประสิทธิ์อัตโนมัติสัมพันธ์อันดับที่ k ของหน้าต่างสัญญาณเสียง

หาค่า r_k ได้ดังสมการที่ 2.9

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

เมื่อ w_n คือ ฟังก์ชันหน้าต่างที่มีระยะเวลา N ตัวอย่าง

การหาค่าสัมประสิทธิ์การเข้ารหัสการทำนายพันธะเชิงเส้นสามารถหาได้จากการแก้สมการที่ 2.9 ซึ่งมีจำนวน p สมการ สมการดังกล่าวเรียกว่าสมการยูล-วอล์กเกอร์ (Yule-Walker) สมการทั้งหมดสามารถเขียนในรูปของเมทริกซ์ได้ดังนี้

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

โดยที่

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

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

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

เมตริกซ์อัตโนมัติสหสัมพันธ์ (Autocorrelation Matrix) หรือเมตริกซ์ R นี้ มีลักษณะโครงสร้างที่กำหนดให้โพลของวงจรรองส่งเคราะห์การนำพาพันธะเชิงเส้น $H(z)$ อยู่ภายในวงกลมหนึ่งหน่วย (Unit Circle) โครงสร้างลักษณะเช่นนี้จะเรียกว่าเป็นโครงสร้างแบบโทพลิตซ์ (Toeplitz) ดังนั้นจึงรับรองได้ว่าวงจรรองส่งเคราะห์ $H(z)$ ที่ได้จากวิธีอัตโนมัติสหสัมพันธ์นี้จะเสถียรเสมอ

สำหรับวิธีการคำนวณหาสัมประสิทธิ์การนำพาพันธะเชิงเส้นในสมการที่ 2.10 มีอยู่หลายวิธี และวิธีหนึ่งที่น่าสนใจคือวิธีการวนซ้ำของเลวินสัน-เดอบิน (Levinson-Durbin Algorithm) ซึ่งแบ่งออกเป็น 4 ขั้นตอน ดังนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้น

$$E_0 = R(0) \text{ และ } a_0 = 0$$

ขั้นที่ 2 คำนวณหาค่าสัมประสิทธิ์การสะท้อน (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 คำนวณค่าสัมประสิทธิ์ของการนำพาพันธะเชิงเส้น

ให้

$$a_m(m) = k_m$$

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

ขั้นที่ 4 คำนวณค่าผิดพลาดใหม่

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

$$m = m + 1$$

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

โดยที่ p คือ อันดับของค่าสัมประสิทธิ์การทำนายพันธะเชิงเส้น

2) วิธีโคเวเรียนซ์ (Covariance Method) ในการวิเคราะห์การทำนายพันธะเชิงเส้น ช่วงของการรวมอยู่ในช่วง $(p+1, N)$ ดังนั้นจึงไม่จำเป็นต้องใช้หน้าต่าง การทำให้ค่าผิดพลาดทั้งหมด ยกกำลังสองมีค่าต่ำที่สุดหาได้จากสมการจำนวน p สมการ ดังต่อไปนี้

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

โดยที่

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

สมการจำนวน p สมการ สามารถเขียนในรูปของเมตริกซ์ได้ดังนี้

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

โดยที่

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

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

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

เมตริกซ์ C เรียกว่า เมตริกซ์โคเวเรียนซ์ (Covariance Matrix) และเป็นเมตริกซ์สมมาตร นั่นคือ $C_{1p} = C_{p1}$ แต่เมตริกซ์ไม่ได้มีโครงสร้างเป็นแบบโทพลิตซ์ ดังนั้นการหาสัมประสิทธิ์การเข้ารหัสการทำนายพัลซิงเส้นด้วยวิธีนี้จึงมีประสิทธิภาพต่ำกว่าวิธีออสทสสัมพันธ์ และสัมประสิทธิ์การเข้ารหัสการทำนายพัลซิงเส้นที่ได้จากวิธีนี้จึงรับรองไม่ได้ว่าจะให้วงจรรองส่งเคราะห์ที่มีความเสถียรเสมอ นอกจากนี้โครงสร้างที่สมมาตรทำให้การคำนวณบางส่วนสามารถใช้เทคนิคการคำนวณให้เร็วขึ้นได้ แต่ยังไม่เร็วเทียบเท่าวิธีเลวินสัน-โคบิน

2.2.2 สัมประสิทธิ์คู่เส้นสเปกตรัม

คู่เส้นสเปกตรัม (Line Spectral Pairs: LSP) หรือความถี่เส้นสเปกตรัม (Line Spectral Frequencies : LSF) เป็นพารามิเตอร์รูปแบบหนึ่งที่พัฒนามาจากพารามิเตอร์การทำนายพัลซิงเส้น เนื่องจากพารามิเตอร์การทำนายพัลซิงเส้นในขั้นตอนการประมาณค่าพารามิเตอร์อาจทำให้เกิดความไม่เสถียรของสัญญาณได้ ซึ่งส่งผลต่อคุณภาพของเสียง ในขณะที่พารามิเตอร์คู่เส้นสเปกตรัมมีคุณสมบัติที่เด่นคือ ค่าพารามิเตอร์อยู่ภายในขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์ และสามารถตรวจสอบเสถียรภาพของวงจรรองส่งได้ง่าย นอกจากนี้คู่เส้นสเปกตรัมยังแสดงในรูปเชิงความถี่จึงสามารถนำไปใช้ในการหาคุณสมบัติที่แน่นอนในระบบการรับรู้ของคนได้

ในการคำนวณหาคู่เส้นสเปกตรัมเริ่มต้นจากพหุนามอันดับ M ของวงจรรองส่งผกผันในเชิงแซดดังสมการ 2.10 โดยทำการแยกส่วนสมการดังกล่าวเป็นพหุนามอันดับ $M+1$ จำนวน 2 พหุนาม ดังสมการที่ 2.20 และสมการที่ 2.21

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

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

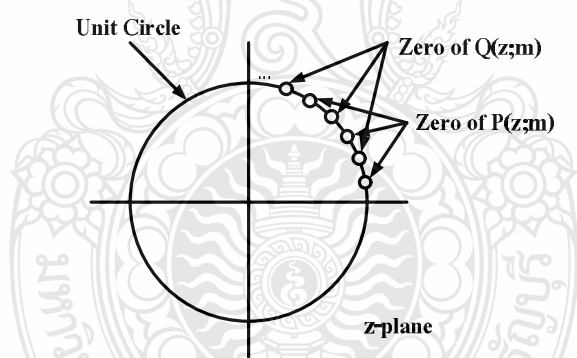
โดยพหุนาม $P(z)$ และ $Q(z)$ มีความสัมพันธ์กับ $A(z)$ ตามสมการที่ 2.27

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

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

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

2) ราก (Zeroes) ของพหุนาม $P(z)$ และ $Q(z)$ จะวางเรียงสลับกันตามลำดับจากน้อยไปหามาก ดังแสดงในภาพที่ 2.2

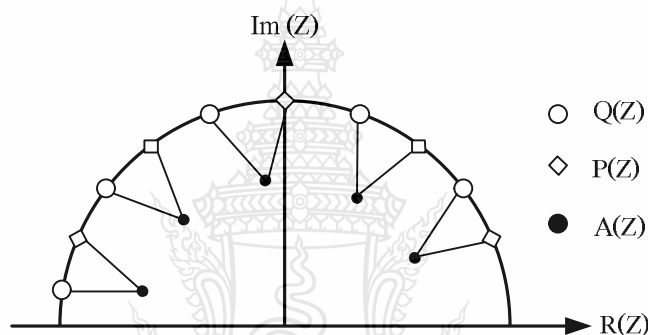


ภาพที่ 2.2 การวางเรียงสลับของรากของพหุนามคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ [4]

โดยสามารถแสดงให้เห็นได้ว่าวงจรรองผกผัน $A(z)$ จะมีเฟสต่ำสุด ถ้าคู่เส้นสเปกตรัมของ $A(z)$ สอดคล้องกับคุณสมบัติทั้งสองนี้ ดังนั้นวงจรสังเคราะห์การเข้ารหัสการทำนายพัลส์เชิงเส้นสามารถยืนยันได้ว่ามีเสถียรภาพ โดยการทำการประมวลผลพารามิเตอร์การเข้ารหัสการทำนายพัลส์เชิงเส้นในรูปแบบคู่เส้นสเปกตรัม

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 2.3



ภาพที่ 2.3 ความสัมพันธ์ระหว่างรากของ $A(z)$ กับรากของกลุ่มเส้นสเปกตรัม $P(z)$ และ $Q(z)$ [4]

จากภาพที่ 2.3 รูปวงกลมสีดำจะเป็นรากของพหุนาม $A(z)$ ส่วนรูปสี่เหลี่ยมและรูปวงกลมสีขาวเป็นรากของพหุนาม $P(z)$ และ $Q(z)$ ตามลำดับ

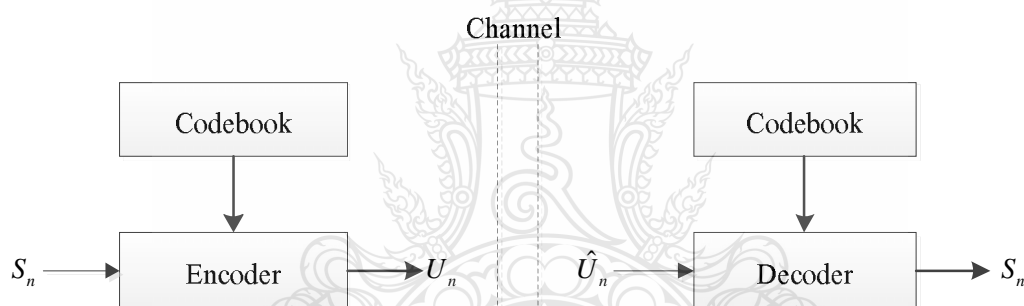
การเกาะกลุ่มของรากหรือความถี่เส้นสเปกตรัมจำนวน 2 หรือ 3 รากจะกำหนดลักษณะความถี่ฟอร์แมนต์และขนาดแบนด์วิดท์ของฟอร์แมนต์ โดยขึ้นกับความชิดของกลุ่มรากดังกล่าว ถ้ารากชิดกันมากจะทำให้เกิดฟอร์แมนต์และมีแบนด์วิดท์แคบ ในทางตรงข้ามกลุ่มรากจะแสดงถึงสเปกตรัมที่มีแบนด์วิดท์กว้างคือไม่เกิดฟอร์แมนต์ นอกจากนี้คุณสมบัติอีกอย่างหนึ่งของเส้นสเปกตรัมคือความไวทางสเปกตรัมของกลุ่มเส้นสเปกตรัมมีลักษณะเฉพาะที่ คือเมื่อมีการเปลี่ยนแปลงกลุ่มเส้นสเปกตรัมใดๆ จะทำให้เกิดการเปลี่ยนแปลงของสเปกตรัมกำลังของการเข้ารหัสการทำนายพัลเซเชิงเส้นเฉพาะบริเวณรอบๆ เท่านั้น ทำให้สามารถทำการประเมินค่า (Quantization) ได้อย่างอิสระโดยไม่มีผล กระทบจากการลดทอนเนื่องจากการประเมินค่าจากสเปกตรัมหนึ่งไปสู่สเปกตรัมอื่นๆ

จากการวิเคราะห์สัญญาณเสียงทั้งของสัมประสิทธิ์การทำนายพัลเซเชิงเส้นและสัมประสิทธิ์กลุ่มเส้นสเปกตรัม ถึงจะเป็นการเข้ารหัสที่มีบิตเรตต่ำแล้วก็ตาม แต่การเกิดขึ้นของ

พารามิเตอร์ทั้งสองยังคงมีค่าสัมประสิทธิ์ความใกล้เคียงกันหรือซ้ำกัน ทำให้ยังคงปรับปรุงการเข้ารหัสของค่าสัมประสิทธิ์ให้ลดลงได้ ซึ่งเทคนิคเวกเตอร์ควอนไทซ์เซชันเป็นเทคนิคที่นิยมใช้ในการบีบอัดสัมประสิทธิ์การทำนายพันธะเชิงเส้นและสัมประสิทธิ์คู่เส้นสเปกตรัม เพราะเป็นการเข้ารหัสแบบพารามเมตริก ทำให้สามารถปรับปรุงการเข้ารหัสสัญญาณเสียงพูดให้มีประสิทธิภาพดีขึ้นได้อีก ซึ่งรายละเอียดของทฤษฎีจะกล่าวในหัวข้อถัดไปจากนี้

2.3 เวกเตอร์ควอนไทซ์เซชัน

เวกเตอร์ควอนไทซ์เซชัน (Vector Quantization: VQ) เป็นการเข้ารหัสข้อมูลที่พัฒนามาจากสเกลาร์ควอนไทซ์เซชัน โดยที่เวกเตอร์ควอนไทซ์เซชันเป็นการเข้ารหัสข้อมูลแบบพารามเมตริกเป็นการนำสัญญาณโบนารีข้อมูลมาแทนที่โมเดลของเสียง (Model Speech) หรือพารามิเตอร์ทางความถี่ของเสียง

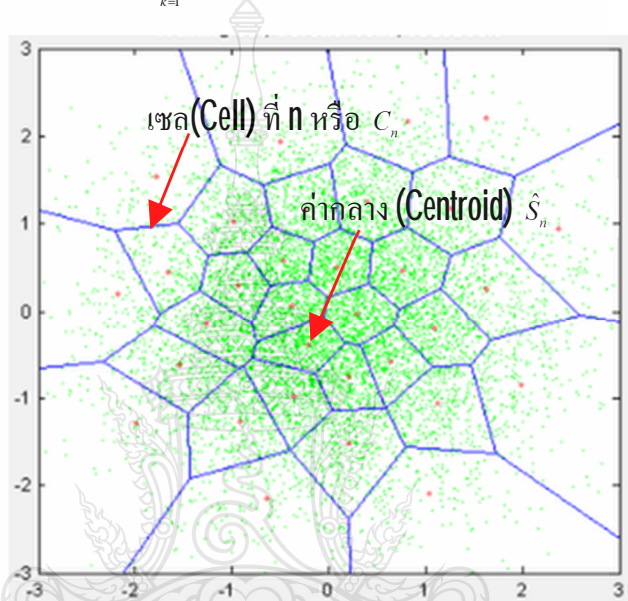


ภาพที่ 2.4 การทำงานของเวกเตอร์ควอนไทซ์เซชัน [3]

การทำงานของเวกเตอร์ควอนไทซ์เซชันประกอบด้วยตัวควอนไทซ์แบบ N มิติ และชุดของสมาชิกเวกเตอร์รหัสหรือเรียกว่าโค้ดบุ๊ก (Codebook) เวกเตอร์ที่เข้ามาถูกสร้างขึ้นจากการสุ่มตัวอย่างของสัญญาณที่ติดกันหรือจากค่าพารามิเตอร์ของสัญญาณเสียงนั้น ตัวควอนไทซ์จะทำการจับคู่เวกเตอร์ขนาด $N \times 1$ ที่เข้ามาตัวที่ i คือ $S_i = [s_i(1) \ s_i(2) \ s_i(3) \ \dots \ s_i(N)]$ เข้ากับสัญลักษณ์ที่ถูกส่งไปในช่องสัญญาณ (Channel Symbol) U_n ($n=1,2,3,\dots,L$) สำหรับกรณีทั่วไปเราถือว่าในช่องสัญญาณไม่มีเสียงรบกวนดังนั้น $U_n = \hat{U}_n$ ในชุดรหัสจะประกอบด้วยเวกเตอร์รหัส (Codevector) L ตัว คือ $S_i = [\hat{s}_i(1) \ \hat{s}_i(2) \ \hat{s}_i(3) \ \dots \ \hat{s}_i(N)]$; $n=1,2,3,\dots,L$ ซึ่งอยู่ในหน่วยความจำของทั้งตัวรับและตัวส่ง ดังแสดงในภาพที่ 2.4 ส่วนวิธีการค้นหาเวกเตอร์ที่ต้องการของ VQ นั้น เวกเตอร์ S_i ที่เข้ามาจะถูก

เปรียบเทียบกับเวกเตอร์รหัสที่ละตัวทุกๆ ตัว และค่าดัชนีที่ใช้แทนเวกเตอร์ที่ใกล้เคียงกับ s_i ที่สุดจะส่งไปในช่องสัญญาณ การพิจารณาความใกล้เคียงของเวกเตอร์รหัสจะดูจากการวัดความเพี้ยน (Distortion Measure) e_n วิธีการที่ใช้มากที่สุด และพื้นฐานที่สุดในการวัดความเพี้ยน คือการหาค่าผลรวมของความผิดพลาดยกกำลังสอง ดังในสมการที่ 2.23

$$e_n = \sum_{k=1}^N [S_i(k) - \hat{S}_n(k)]^2 \quad (2.23)$$



ภาพที่ 2.5 ตัวอย่างของการควอนไทซ์ในระนาบ 2 มิติ [5]

เวกเตอร์รหัสจำนวน L ตัวที่อยู่ในโค้ดบุ๊ก คือ เวกเตอร์ค่าจริงขนาด $N \times 1$ จำนวน L ตัว ถูกออกแบบโดยการแบ่งปริภูมิเวกเตอร์ออกเป็นเซลล์ที่ไม่ซ้อนทับกันจำนวน L และเซลล์ C_n แต่ละเซลล์จะถูกเชื่อมโยงกับเวกเตอร์ \hat{S}_n แสดงดังในภาพที่ 2.5 ตัวอย่างการควอนไทซ์จะระบุสัญลักษณ์ของช่องสัญญาณ U_n ให้ s_i เมื่อ s_i อยู่ใน C_n นั่นคือมันจะถูกเสนอโดยให้ค่าเป็น \hat{S}_n ซึ่งเป็นจุดศูนย์กลาง (Centroid) ของเซลล์ C_n สัญลักษณ์ของช่องสัญญาณมักจะใช้เป็นสัญญาณไบนารีของดัชนีหรือตำแหน่งของ \hat{S}_n รูปแบบที่ง่ายที่สุดของ VQ คือเวกเตอร์ PCM (VPCM) ซึ่งใช้การค้นหาโค้ดบุ๊กแบบเต็มรูปแบบ (Full search VQ : F-VQ) โดยสามารถหาค่าจำนวนบิตต่อสัญญาณจากสมการ 2.24

$$B = \frac{\log_2 L}{N} \quad (2.24)$$

โดย B คือ อัตราจำนวนบิตต่อสัญญาณ
 N คือ มิติของตัวควอนไทซ์
 L คือ จำนวนเวกเตอร์รหัส

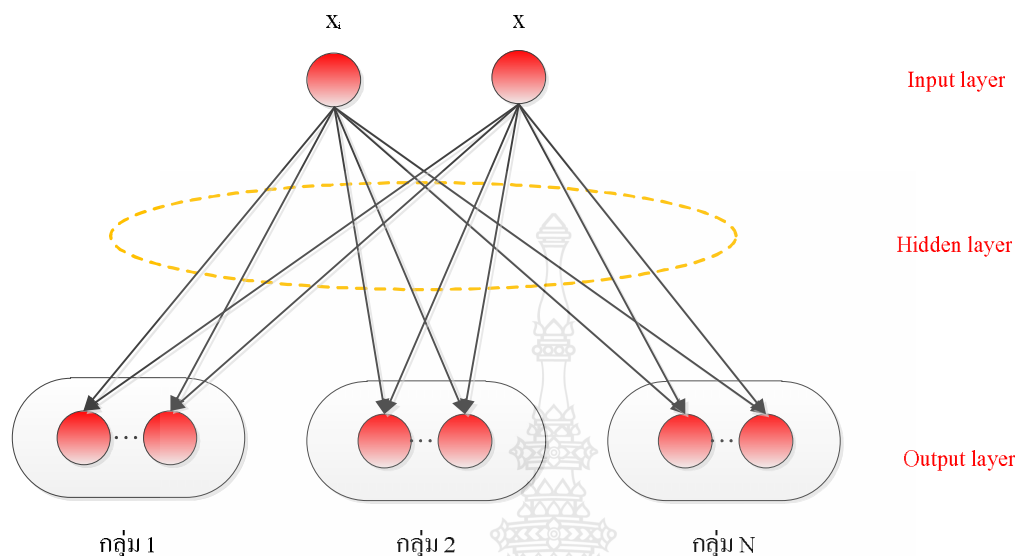
VQ มีข้อเสียเปรียบในเรื่องของความซับซ้อนสูงในการค้นหาโค้ดบุ๊ก เทคนิคบางเทคนิคให้คุณภาพในการลดความซับซ้อนของสัญญาณเสียงที่แตกต่างกัน เช่น ในงานวิจัยเวกเตอร์ควอนไทซ์เซชัน ด้วยเทคนิค KSOFM [6] ในการบีบอัดพารามิเตอร์ LPC และ LSP โดยมีค่าความผิดพลาดสูงสุดที่ 40 % และสามารถลดค่า bit rate ได้ 33% เมื่อเปรียบเทียบกับพารามิเตอร์เดิม ในงานวิจัยเวกเตอร์ควอนไทซ์เซชัน ด้วยเทคนิค Local Cosine Transform Coefficient (LPT) [7] สำหรับการบีบอัดสัญญาณเสียงบิตต่ำที่ 2.4 kb/s โดยใช้เวกเตอร์ควอนไทซ์แบบแยกชุดโค้ดบุ๊กเป็น 4 ส่วน มีคุณภาพ PSNR มีค่าเท่ากับ 20.54 dB แต่ในงานวิจัยนี้ได้เลือกใช้เวกเตอร์ควอนไทซ์เซชัน ด้วยเทคนิค LVQ มาประยุกต์กับการบีบอัดสัญญาณเสียง เพราะจากการศึกษาค้นคว้า LVQ สามารถจัดกลุ่มข้อมูลและบีบอัดข้อมูลได้อย่างมีประสิทธิภาพ [8] ซึ่งรายละเอียดทฤษฎีจะอธิบายในหัวข้อถัดไป

2.4 เครื่องมือการเรียนรู้การควอนไทซ์ไทเซชัน

เครื่องมือการเรียนรู้เวกเตอร์ควอนไทซ์เซชัน (Learning Vector Quantization : LVQ) เป็นการขยายเครือข่ายโคโฮเนน (Kohonen Network) หรือเรียกอีกแบบว่าการเรียนรู้ที่ไม่มีผู้ฝึกสอน เครื่องมือ LVQ ใช้ในการคัดแยกเวกเตอร์อินพุตด้วยการกำหนดคุณลักษณะของกลุ่มสำหรับคัดแยก จะมีการกำหนดกลุ่มที่ใช้การคัดแยกและใช้เป็นส่วนหนึ่งในการฝึกสอน โครงสร้างของเครือข่าย LVQ มีลักษณะเช่นเดียวกับเครือข่ายการจัดการตนเอง (Self-Organizing Feature Maps : SOFM) [4] โดยมีข้อแตกต่างคือเครือข่าย LVQ แต่ละนิวรอลเอาต์พุตจะถูกกำหนดให้เป็นกลุ่มของการคัดแยก ดังที่แสดงในภาพที่ 2.6 โดยปกติแล้วกลุ่มการคัดแยกหนึ่งๆ จะประกอบด้วยนิวรอลเอาต์พุตมากกว่าหนึ่งนิวรอล โดยที่เวกเตอร์น้ำหนักประสาทของนิวรอลเอาต์พุตจะทำการโต้ตอบกันกับเวกเตอร์รูปแบบอินพุตที่มีความสัมพันธ์กัน คือเมื่ออินพุต X ถูกป้อนให้กับเครือข่าย นิวรอลเอาต์พุตที่ใกล้เคียงกับ X ที่สุดจะเป็นนิวรอลผู้ชนะ เครื่องมือจะทำการปรับค่าของเวกเตอร์น้ำหนักประสาทของนิวรอลผู้ชนะและนิวรอลอื่นๆ ตามความสัมพันธ์ดังสมการที่ 2.25 และ 2.26

$$w_{ij}^{new} = w_{ij}^{old} + \alpha (x_i - w_{ij}^{old}) \quad \text{ถ้ากลุ่มการคัดแยกถูกต้อง} \quad (2.25)$$

$$w_{ij}^{new} = w_{ij}^{old} - \alpha (x_i - w_{ij}^{old}) \quad \text{ถ้ากลุ่มการคัดแยกไม่ถูกต้อง} \quad (2.26)$$



ภาพที่ 2.6 โครงสร้างเครือข่าย LVQ

สำหรับการฝึกฝนหรือฝึกสอนของการเรียนรู้ของ LVQ จะเห็นได้ว่าการฝึกสอนถ้า นิวรอลใดที่ได้ถูกเลือกเป็นนิวรอลผู้ชนะของกลุ่มของเวกเตอร์ จะถูกการปรับค่าให้เวกเตอร์น้ำหนัก ประสาทของนิวรอลนั้นให้มีค่าเข้าใกล้รูปแบบอินพุตเวกเตอร์ และในทางตรงกันข้ามถ้านิวรอลใดที่ไม่ได้ถูกเลือกเป็นนิวรอลผู้ชนะ จะถูกการปรับค่าให้เวกเตอร์น้ำหนักประสาทของนิวรอลนั้นให้มีค่า ห่างจากรูปแบบอินพุตเวกเตอร์ ดังภาพที่ 2.7

โดยที่โครงสร้างของเครือข่าย LVQ จะมีการกำหนดให้ x^p คือเวกเตอร์รูปแบบอินพุตตัวที่ p สำหรับใช้ในการฝึกสอน T^p แทนกลุ่มการคัดแยกที่ต้องการของ x^p และท้ายสุด C_j แทนกลุ่มการคัดแยกของนิวรอลเอาต์พุตตัวที่ j อัลกอริทึมการเรียนรู้ของเครือข่าย LVQ สามารถสรุปได้ดังนี้

- 1) กำหนดค่าเริ่มต้นให้กับเวกเตอร์น้ำหนักประสาทและค่าคงที่ของการเรียนรู้
- 2) แต่ละเวกเตอร์รูปแบบอินพุต x^p คำนวณหานิวรอลผู้ชนะ k จากความสัมพันธ์ดังนี้

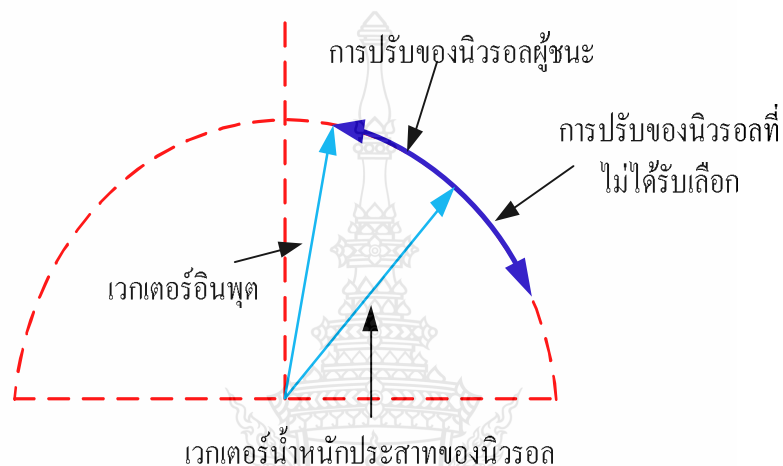
$$i(x^p) = k, \text{ โดยที่ } \|x^p - w_k\| < \|x^p - w_j\|, j = 1, 2, \dots, n \quad (2.27)$$

- 3) ทำการปรับค่า w_k ดังนี้

$$w_{ij}^{new} = w_{ij}^{old} + \alpha(x_i - w_{ij}^{old}) \text{ ถ้า } T^P = C_j \quad (2.28)$$

$$w_{ij}^{new} = w_{ij}^{old} - \alpha(x_i - w_{ij}^{old}) \text{ ถ้า } T^P \neq C_j \quad (2.29)$$

4) ปรับค่าคงที่การเรียนรู้ให้มีค่าลดลงตามจำนวนรอบที่เพิ่มขึ้น



ภาพที่ 2.7 การเรียนรู้ของเครือข่าย LVO [8]

2.5 การวัดความผิดพลาด

การวัดค่าความผิดพลาด (**Error**) ของสัญญาณจะใช้การวัดค่าเชิงวัตถุ เนื่องจากเป็นวิธีการที่ทำได้ง่ายไม่ยุ่งยากใช้เวลาน้อย การหาค่าความผิดพลาดระหว่างสัญญาณก่อนบัพอัดและสัญญาณที่ผ่านการบัพอัดแล้ว โดยงานวิจัยนี้ได้ใช้การวัดความผิดพลาดด้วยวิธี **Normalized Root Mean Squared Error (NRMSE)** ซึ่งเป็นวิธีที่ใช้ในการคำนวณหาประสิทธิภาพของการแทนค่าข้อมูลที่ขาดหายไป โดยมีสมการดังต่อไปนี้

$$NRMSE = \sqrt{\frac{(p(n) - r(n))^2}{(p(n) - \mu p(n))^2}} \quad (2.30)$$

โดยที่ $p(x)$ คือ สัญญาณต้นฉบับ

$r(n)$ คือ สัญญาณที่สร้างขึ้นใหม่

$\mu p(n)$ คือ ค่าเฉลี่ยของสัญญาณต้นฉบับ

2.6 การวัดค่าประสิทธิภาพของสัญญาณ

การหาค่าประสิทธิภาพของสัญญาณจะใช้การวัดค่าเชิงวัตถุ ซึ่งสามารถวัดคุณภาพของอัลกอริทึมในการบีบอัดข้อมูลได้ [9] โดยการหาค่าประสิทธิภาพของสัญญาณแบ่งการหาค่าเป็น 2 เทคนิค ดังนี้ เทคนิคแรก คือ Peak Signal to Noise Ratio (PSNR) เป็นการพิจารณาในรูปของค่าอัตราส่วนระหว่างค่ากำลังงานสูงของสัญญาณกับความแตกต่างของสัญญาณที่ถูกรบกวนเทียบกับสัญญาณต้นฉบับ และเทคนิคที่สอง คือ Signal to Noise Ratio (SNR) เป็นการพิจารณาหาค่าอัตราส่วนสัญญาณเสียงต้นฉบับต่อสัญญาณรบกวนแล้วทำการเปรียบเทียบค่าระหว่างสัญญาณเสียงต้นฉบับกับสัญญาณเสียงที่ผ่านการบีบอัด โดยที่การหาค่าประสิทธิภาพของสัญญาณของ PSNR และ SNR มีสมการดังต่อไปนี้

$$PSNR = 10 \log_{10} \frac{NP^2}{\|p-r\|^2} \quad (2.31)$$

โดยที่ N คือ จำนวนความยาวของสัญญาณที่ถูกคืนกลับ

P^2 คือ ค่าสูงสุดของสัญญาณต้นฉบับ

$\|p-r\|^2$ คือ ค่าความแตกต่างระดับพลังงานระหว่างสัญญาณต้นฉบับกับสัญญาณคืนกลับ

$$SNR = 10 \log_{10} \left| \frac{\sigma_x^2}{\sigma_e^2} \right| \quad (2.32)$$

โดยที่ σ_x^2 คือ ค่าเฉลี่ยกำลังสองของสัญญาณ

σ_e^2 คือ ค่าเฉลี่ยระหว่างของความแตกต่างระหว่างสัญญาณต้นฉบับเทียบกับสัญญาณคืนกลับ

จากทฤษฎีทั้งหมดที่ได้กล่าวมา จะนำไปประยุกต์ใช้กับการศึกษาเทคนิคการบีบอัดสัญญาณเสียงพูดในงานวิจัยนี้ เพื่อลดขนาดของสัญญาณเสียงพูด ซึ่งงานวิจัยทางการประมวลผลสัญญาณเสียงพูดมีการพัฒนาอย่างหลากหลายแนวทาง และมีความแตกต่างกันออกไปตามวัตถุประสงค์ของการวิจัยนั้นๆ ดังนั้นในการวิจัยครั้งนี้จึงได้คัดเลือกงานวิจัยที่เกี่ยวข้องและมีความใกล้เคียงกับงานวิจัยที่กำลังศึกษา ดังตัวอย่างต่อไปนี้

2.7 การประเมินคุณภาพของข้อมูลเสียงพูด

การประเมินคุณภาพของข้อมูลเสียงพูดโดยใช้วิธีการของ MOS (Mean Opinion Score) เป็นวิธีการหนึ่งที่มีนิยมนำมาใช้เปรียบเทียบระหว่างเสียงพูดต้นแบบกับเสียงพูดที่ผ่านการบีบอัดข้อมูล โดยใช้การรับรู้และความรู้สึกของมนุษย์เป็นเกณฑ์ในการตัดสิน (Subjective Measure)

ตารางที่ 2.1 ค่า MOS ที่เหมาะสมกับการใช้งานในระบบต่างๆ

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

วิธีการประเมินหรือวัดคุณภาพเสียงนั้นจะใช้คนประมาณ 12-24 คน ทดสอบคุณภาพเสียงด้วยการฟัง โดยที่แต่ละคนจะให้คะแนนที่มีค่าอยู่ระหว่าง 1-5 ตามคุณภาพของสัญญาณที่ตัวเองรู้สึก จากนั้นหาค่าเฉลี่ยแต่ละเสียงพูดว่าอยู่ในระดับใด

ตารางที่ 2.2 รายละเอียดวิธีการให้คะแนนในการวัด MOS

| คะแนน | คุณภาพของเสียง |
|-------|--|
| 5 | ดีมาก (คุณภาพเสียงชัดเจนและเข้าใจง่าย) |
| 4 | ดี (คุณภาพเสียงดีและเข้าใจง่าย แต่อาจได้ยินเสียงรบกวนบ้าง) |
| 3 | พอใช้ (คุณภาพเสียงดีและเข้าใจ ได้แต่อาจต้องการอาศัยความตั้งใจ หรือบางที่ต้องขอให้พูดซ้ำ) |
| 2 | เลว (คุณภาพเสียงดีและเข้าใจได้ก็ต่อเมื่อมีความตั้งใจมากๆ และบ่อยครั้งที่ต้องขอให้พูดซ้ำ) |
| 1 | เลวมาก (ฟังไม่รู้เรื่องเลย) |

2.8 งานวิจัยที่เกี่ยวข้อง

2.8.1 Dong.E, Guizhong.L และ Hongliang [7] นำเสนอในหัวข้อ "2.4 kb/s low bit rate speech coding based on local cosine transform" เป็นการอธิบายการเข้ารหัสอัตราบิตต่ำ 2.4 kb/s คุณภาพเสียงระดับสังเคราะห์ โดยใช้ LPT (Local Cosine Transform Coefficient) ใช้เทคนิคเวกเตอร์ควอนไทซ์แยกชุดของสมาชิกเวกเตอร์รหัส (Codebook) เป็น 4 ส่วน โดยมีค่า PSNR 20.54 dB

2.8.2 Ornanong Wiriyannuruknakorn และ Jakkree Srinonchat [10] นำเสนอในหัวข้อ "A Finite State Vector Quantizer for New Bit Rate Speech Compression" เป็นการอธิบายใช้เวกเตอร์ควอนไทซ์เซชัน (VQ) ในการบีบอัดสัญญาณเสียงพูดด้วยการใช้ LPC-10 แบบใหม่ พบว่าการบีบอัดในวิธีการของ LPC-10 แบบเดิมและแบบใหม่เปรียบเทียบกัน จะเห็นได้ว่าการบีบอัดด้วยวิธีของ LPC-10 แบบใหม่สามารถลดบิตเรทลงได้ 5.5-37%

2.8.3 Jakkree Srinonchat [11] นำเสนอในหัวข้อ "Address Vector Quantisation applied to Speech Coding" เป็นการอธิบายการใช้ตำแหน่งเวกเตอร์ควอนไทซ์เซชัน (VQ) ในการบีบอัดสัญญาณเสียงพูด โดยการแบ่งเสียงพูดออกเป็นเฟรม แต่ละเฟรมมีค่า 30 ms. แล้วนำมาสกัดค่าเป็นสัมประสิทธิ์ LPC และ LSP เพื่อใช้เป็นพารามิเตอร์ในการสร้าง Codebook ด้วย KSOFM ซึ่งทำหน้าที่คำนวณค่าและแทนค่าสัมประสิทธิ์สัญญาณเสียงพูดลงใน Codebook ให้มีความผิดพลาดน้อยที่สุด ผลจากการทดลองเสียงพูดที่เป็นชายจำนวน 2 คนและหญิง 2 คน พบว่า ค่าผิดพลาดสูงสุดของการบีบอัดค่า LSP ในเสียงพูดชาย 35% และในเสียงพูดหญิง 40% ส่วนค่าผิดพลาดสูงสุดของการบีบอัดค่า LPC ในเสียงพูดชาย 35% และในเสียงพูดหญิง 30% สรุปแล้วการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค LPC-VQ จะใช้ KSOFM และการใช้ Address-Codebook จะลดค่า Bit Rate ได้ 33%

2.8.4 Jakkree Srinonchat [12] นำเสนอในหัวข้อ "New Technique to Reduce Bit Rate of LPC-10 Speech Coder" เป็นการอธิบายเกี่ยวกับโครงสร้างการบีบอัดข้อมูลให้มีค่า bit rate ต่ำลงด้วยการใช้ LPC-10 แบบใหม่โดยพบว่าค่าพารามิเตอร์ LPC ที่ใช้ทั่วไปจะให้ค่าการควอนไทซ์ไม่ค่อยมีประสิทธิภาพ เนื่องจากจะเกิดค่าผิดพลาดขึ้นในกรณีที่ระดับความแตกต่างของสัญญาณเสียงพูดมีระดับต่ำ เพื่อแก้ปัญหาเรื่องความผิดพลาดของการควอนไทซ์ จึงได้นำเสนอวิธีการโดยเปลี่ยนค่าพารามิเตอร์ LPC เป็นพารามิเตอร์ LSP เพื่อจะสร้าง Codevector ของการแบ่งลำดับชั้นใน Vector Quantization เพื่อที่จะใช้เข้ารหัสสัญญาณเสียงพูดใหม่ เรียกว่า LPC-10 จากผลทดลองพบว่าสามารถลดจำนวนบิตของสัญญาณเสียงพูดในบิตสัญญาณเสียงพูด p1-p4 ได้ 4 บิต และอัตราการส่งข้อมูลลดลง 2-66%

2.8.5 Song Lin, Dimitrios Gunopulos, Vana Kalogeraki และ Stefano Lonardi [13] นำเสนอในหัวข้อ "A Data Compression Technique for Sensor Networks with Dynamic Bandwidths Allocation" เป็นการอธิบายการใช้ LVO ทำการควอนไทซ์เพื่อสร้างโค้ดบुकเพื่อลดขนาดแบนด์วิดท์ การสื่อสารข้อมูลของ Sensor Networks แล้วการเปรียบเทียบค่าความผิดพลาด (Error) กับข้อมูลต้นฉบับ พบว่ามี Less Bandwidth ที่ 25 %

2.8.6 Anand D Subramaniam และ Bhaskar D. Rao [14] นำเสนอในหัวข้อ "Low Complexity Recursive Coding of Spectrum Parameters" เป็นการอธิบาย Gaussian Mixture Model (GMM) ภายใต้ขอบเขตการควอนไทซ์การเข้ารหัสแบบหลายมิติ [9] โดยมีการพัฒนาเป็น 2 รูปแบบ รูปแบบแรกคือการควอนไทซ์ในขอบเขตการเข้ารหัสแบบแทนที่โดยการควอนไทซ์เวกเตอร์ที่เสริมประสิทธิภาพทำให้ bit rate ลดลง ในส่วนรูปแบบที่สองคือ ผู้วิจัยสรุปขอบเขตของการจัดให้เหมาะสมกับการเข้ารหัสแบบย้อนกลับ ในการเข้าใกล้ร่วมกับ Product Vector Quantizes (PVQ) ของตัวแปรเวกเตอร์ของเฟรมเสียงต่อเนื่องเป็นการสร้างแบบทดสอบใช้ GMM ความหนาแน่นแบบมีเงื่อนไขของตัวแปรเวกเตอร์ของเฟรมเสียงบนฐานค่าควอนไทซ์ของตัวแปรเวกเตอร์ของเฟรมเสียงแรกใช้การสร้าง Codebook และจากการทดลองสรุปให้เห็นว่ารูปแบบที่ใช้ของการควอนไทซ์ Spectrum เสียงพูดจะได้ประสิทธิภาพ

2.8.7 Jakkree Srinonchat [15] นำเสนอในหัวข้อ "New Technique to Reduce Bit Rate of LPC-10 Speech Coder" เป็นการอธิบายเกี่ยวกับการลด Bit Rate ของระบบการบีบอัดเสียงพูดเสียงโดยใช้ Artificial Neural Networks ซึ่งในงานวิจัยนี้มีด้วยกันสองแบบคือ KSOFM และ PNNs [11] ซึ่งเป็นการตรวจสอบที่ใช้ในการเข้ารหัสเสียงพูดแบบ CELP ในส่วน KSOFM เป็นการใช้อัดกลุ่มของการซ้ำของสัญญาณเสียงและสร้าง Codebook ที่เหมาะสม และ PNNs จะถูกนำไปใช้ในการทำนายของ Codebook โดยใช้กับความรู้ในการฝึกฝน จากผลการทดลองทำให้ Bit Rate ลดลงประมาณ 25% เหมือนเปรียบเทียบกับแบบเดิมของสัญญาณเสียง

2.8.8 Gao Li'ai และ Yongjie Li Lihua [16] นำเสนอในหัวข้อ "A New Codebook Design Method Based on Genetic Programming" เป็นการอธิบายเกี่ยวกับ GP-L algorithm และ LBG algorithm [12] ในกระบวนการออกแบบ Codebook และผู้วิจัยได้ประยุกต์พัฒนาโดยนำ GL-P Algorithm บนพื้นฐานของ Genetic Programming มาออกแบบ Codebook พร้อมเปรียบเทียบกับ LBG-Algorithm จากผลการทดลองแสดงให้เห็นว่าค่าเปอร์เซ็นต์ identify-rate ของ GP-L Algorithm จะดีกว่า LBG-Algorithm

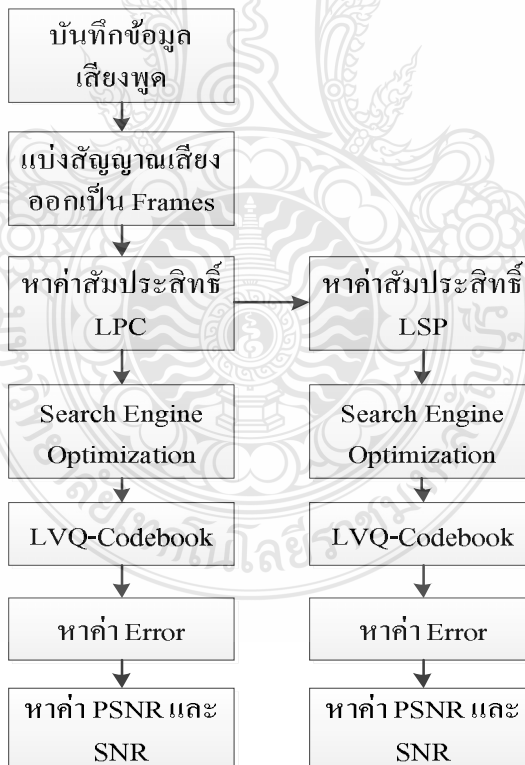
2.8.9 Gao Li'ai และ Yongjie Li Lihua [17] นำเสนอในหัวข้อ “Combined Compression and Classification with Learning Vector Quantization” เป็นการอธิบายการพัฒนาและวิเคราะห์บนพื้นฐาน Learning Vector Quantization (LVQ) [13] สำหรับการบีบอัดแบบรวมและการจัดกลุ่ม จากผลการทดลองกับตัวอย่างข้อมูลแบบหลายมิติ พบว่า LVQ ด้วยวิธีที่ใช้ ODE จากการประมาณการสุ่มมีประสิทธิภาพในการบีบอัดและแยกแยะข้อมูล โดยที่ยังคงมีความใกล้เคียงกับตัวอย่างข้อมูลต้นฉบับ



บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยนี้ได้ศึกษาการบีบอัดสัญญาณเสียง โดยนำ LVQ มาประยุกต์ใช้การเข้ารหัสในมาตรฐาน LPC-10 และ LSP-10 เพื่อลดขนาดในการส่งสัญญาณเสียง แล้วทำการทดสอบประสิทธิภาพของบีบอัดสัญญาณเสียงพูด เมื่อมีขนาดของโค้ดบุ๊กที่ระดับต่างๆกัน โดยในขั้นตอนแรก จะทำการบันทึกสัญญาณเสียงของผู้พูด 10 คน (ชาย 5 คน หญิง 5 คน) ทำการแบ่งสัญญาณเสียงเป็นช่วงเวลา (Frame) ละ 200 ตัวอย่าง แล้วหาพารามิเตอร์ตัวแทนความถี่ของสัญญาณเสียง ด้วยการหาพารามิเตอร์ LPC และ LSP จากนั้นทำการสร้างโค้ดบุ๊กเพื่อลดขนาดในการส่งค่าพารามิเตอร์ทางความถี่ของสัญญาณเสียง แล้วทำการวัดค่าความผิดพลาดที่เกิดจากการบีบอัดพารามิเตอร์ พร้อมกับเปรียบเทียบประสิทธิภาพของการบีบอัดพารามิเตอร์ด้วยหลักการของ PSNR และ SNR ซึ่งมีขั้นตอนในการดำเนินงานวิจัยดังภาพที่ 3.1



ภาพที่ 3.1 แผนผังขั้นตอนการดำเนินการวิจัย

3.1 การเก็บบันทึกสัญญาณเสียงพูด

การบันทึกสัญญาณเสียงพูดใช้การบันทึกเสียงผ่านไมโครโฟนชนิดคอนเดนเซอร์ลงในคอมพิวเตอร์ส่วนบุคคลในห้องทำงานที่มีสภาพแวดล้อมปกติ (สัญญาณรบกวนโดยรวมไม่เกิน 0 - 0.5 dB) โดยบันทึกเสียงแบบดิจิทัลในระบบโมโน กำหนดให้มีอัตราการสุ่มตัวอย่าง 8,000 Hz/s และมีการแบ่งระดับที่ 8 บิต บันทึกข้อมูลอยู่ในรูปของไฟล์ *.Wav โดยมีขั้นตอนการบันทึกเสียงพูด ดังนี้

- เตรียมบทความที่ให้ผู้พูดอ่าน
- เตรียมอาสาสมัครผู้พูด 10 คน แบ่งเป็นชาย 5 คน หญิง 5 คน
- ให้ผู้พูดทดลองอ่านบทความเพื่อปรับระดับสัญญาณเสียงอยู่ในช่วง 0.7-1.0 โวลต์
- ให้ผู้พูดแต่ละคนอ่านบทความที่เตรียมไว้ บันทึกลงในคอมพิวเตอร์

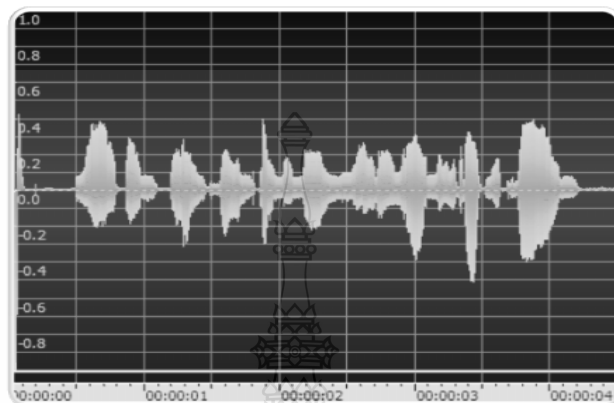
ขั้นตอนการบันทึกสัญญาณเสียงพูดด้วยโปรแกรม GoldWave Version 5.5 มีดังนี้

1) กำหนดค่าการบันทึกสัญญาณเสียงพูด กำหนดให้ความแรงของสัญญาณเสียงพูดมีขนาด -1.0 ถึง 1.0 โดยเลือกเมนูคำสั่ง File>New จะปรากฏหน้าต่าง New Sound เป็นการกำหนดคุณภาพเสียงและระยะเวลาการบันทึก กำหนดค่าที่จำนวนช่องสัญญาณเสียงเป็น 1 (Mono) จำนวนของการสุ่มค่าตัวอย่าง ตั้งค่าเป็น 8,000



ภาพที่ 3.2 การตั้งค่าเริ่มต้นบันทึกเสียงในโปรแกรม GoldWave

2) การบันทึกสัญญาณเสียงพูดนำมาคลิกที่ปุ่มการเริ่มต้นบันทึก **Starts Recording** เพื่อทำการบันทึกไฟล์เสียงใน **Directory** ที่กำหนดโดยใช้ผู้ทดสอบอายุระหว่าง **18 ถึง 40** ปี เป็นผู้ชาย **5** คน ผู้หญิง **5** คน



ภาพที่ 3.3 ตัวอย่างสัญญาณเสียงพูดในกลุ่มผู้พูดที่ใช้ในการวิเคราะห์

3.2 การแบ่งสัญญาณเสียง

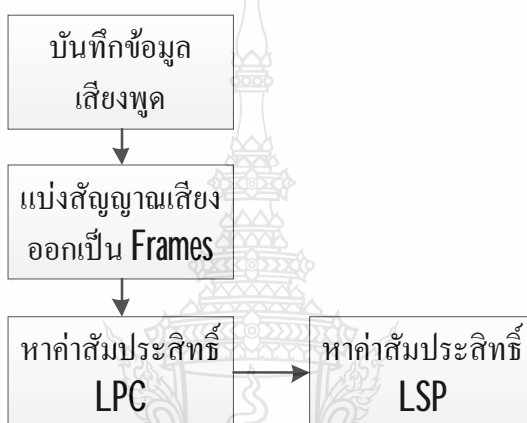
สัญญาณเสียงพูดที่เป็นสัญญาณอนาล็อกจะถูกเปลี่ยนให้เป็นสัญญาณดิจิทัลในขั้นตอนบันทึกเสียงนั้น โดยมีอัตราการสุ่มตัวอย่างที่ **8000 Hz** จะถูกกรองความถี่ให้อยู่ในช่วงช่องสัญญาณเสียงโทรศัพท์ความถี่ **300 ถึง 3,400 Hz** ตามระบบมาตรฐานสัญญาณเสียงโทรศัพท์ และทำการแบ่งสัญญาณเป็นช่วงเวลา (**Frame**) ละ **200** ตัวอย่าง ตามขั้นตอนดังภาพที่ 3.4



ภาพที่ 3.4 การแบ่งสัญญาณเสียงออกเป็น Frame

3.3 การสกัดพารามิเตอร์จำลองสัญญาณเสียง

เมื่อสัญญาณเสียงพูดได้ผ่านการแบ่งเป็น Frame ตามขั้นตอนในภาพที่ 3.4 แล้ว จึงทำการวิเคราะห์หาค่าพารามิเตอร์จำลองสัญญาณเสียง โดยแบ่งเป็น 2 เทคนิค คือ เทคนิคของการหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้น LPC และเทคนิคการหาค่าสัมประสิทธิ์คู่เส้นสเปกตรัม LSP โดยในขั้นตอนจะหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้น LPC ก่อนแล้วจึงมาหาการหาค่าสัมประสิทธิ์คู่เส้นสเปกตรัม LSP ในการหาสัมประสิทธิ์แทนพารามิเตอร์ทางความถี่ โดยมีขั้นตอนดังภาพที่ 3.5



ภาพที่ 3.5 การหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้น และค่าสัมประสิทธิ์คู่เส้นสเปกตรัม

3.3.1 ส่วนการหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้น (Linear Predictive)

การหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้นเป็นเทคนิคที่สำคัญทางด้านกรวิเคราะห์เสียง เนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของเสียงพูดเมื่อเทียบกับความเร็วในการประมวลผล ในการวิเคราะห์การเข้ารหัสการทำนายพันธะเชิงเส้นเริ่มต้นจากพิจารณาจากกรอบสัญญาณเสียงที่มีตัวอย่างตัวอย่าง N ตัวอย่าง คือ $s_1, s_2, s_3, \dots, s_N$ โดยอ้างอิงว่าตัวอย่างสัญญาณปัจจุบันจะสามารถทำนายได้จากผลรวมของตัวอย่างสัญญาณในอดีต p ตัวอย่าง ในงานวิจัยนี้จะใช้วิธีอิตสหสัมพันธ์ของวิธีการ Least-Squares และจะใช้ขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin ในการคำนวณหาค่าสัมประสิทธิ์ a_k โดยจะใช้สัมประสิทธิ์ทำนายพันธะเชิงเส้นที่ 10 อันดับ (LPC-10) ซึ่งผลการหาพารามิเตอร์ด้วยเทคนิคสัมประสิทธิ์ทำนายพันธะเชิงเส้นแสดงตัวอย่างดังภาพที่ 3.6

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----|---|---------|--------|---------|--------|---------|--------|---------|--------|---------|--------|
| 1 | 1 | -0.7959 | 0.8748 | -0.8201 | 0.6999 | -0.6339 | 0.5407 | -0.3944 | 0.3380 | -0.1710 | 0.1297 |
| 2 | 1 | -0.5036 | 0.5097 | -0.4893 | 0.6475 | -0.3808 | 0.4089 | -0.2296 | 0.2223 | -0.0837 | 0.0751 |
| 3 | 1 | -0.4137 | 0.4993 | -0.3805 | 0.5379 | -0.3155 | 0.4303 | -0.2583 | 0.3132 | -0.1843 | 0.1358 |
| 4 | 1 | -0.2485 | 0.3678 | -0.4770 | 0.2877 | -0.4181 | 0.2741 | -0.2160 | 0.2664 | -0.0929 | 0.1443 |
| 5 | 1 | -0.4092 | 0.3181 | -0.6083 | 0.4440 | -0.2906 | 0.4415 | -0.2012 | 0.1596 | -0.1142 | 0.1819 |
| 6 | 1 | -0.3611 | 0.7293 | -0.4862 | 0.6663 | -0.4994 | 0.5051 | -0.2688 | 0.2662 | -0.1012 | 0.0989 |
| 7 | 1 | -0.4423 | 0.3891 | -0.3872 | 0.3912 | -0.3880 | 0.3917 | -0.3337 | 0.2681 | -0.0757 | 0.1484 |
| 8 | 1 | -0.4813 | 0.6008 | -0.3934 | 0.5291 | -0.4822 | 0.3243 | -0.2299 | 0.2831 | -0.1721 | 0.1522 |
| 9 | 1 | -0.1575 | 0.5126 | -0.3549 | 0.4844 | -0.1815 | 0.4040 | -0.2666 | 0.4460 | -0.0643 | 0.2483 |
| 10 | 1 | -0.4087 | 0.5667 | -0.4638 | 0.6342 | -0.3223 | 0.4451 | -0.3232 | 0.2124 | -0.1173 | 0.1791 |
| 11 | 1 | -0.3494 | 0.4941 | -0.3519 | 0.4081 | -0.4892 | 0.3201 | -0.2123 | 0.2723 | -0.0993 | 0.1887 |
| 12 | 1 | -0.4639 | 0.5439 | -0.5632 | 0.5581 | -0.3556 | 0.4029 | -0.1548 | 0.2479 | -0.1769 | 0.1344 |
| 13 | 1 | -0.6034 | 0.7825 | -0.7409 | 0.7048 | -0.4965 | 0.5737 | -0.4001 | 0.3777 | -0.1843 | 0.2141 |
| 14 | 1 | -0.3209 | 0.5707 | -0.4528 | 0.4951 | -0.3577 | 0.3086 | -0.2491 | 0.2881 | -0.0354 | 0.1417 |
| 15 | 1 | -0.2510 | 0.6299 | -0.5541 | 0.3824 | -0.3669 | 0.3584 | -0.1426 | 0.2054 | 0.0271 | 0.1291 |
| 16 | 1 | -0.4944 | 0.4888 | -0.2139 | 0.4000 | -0.3083 | 0.3787 | -0.1268 | 0.3178 | -0.1537 | 0.2151 |
| 17 | 1 | -0.6161 | 0.5535 | -0.5121 | 0.6159 | -0.3124 | 0.3993 | -0.3404 | 0.2770 | -0.1211 | 0.1450 |
| 18 | 1 | -0.6028 | 0.7135 | -0.6712 | 0.6381 | -0.5628 | 0.5470 | -0.4090 | 0.3694 | -0.1308 | 0.1784 |
| 19 | 1 | -0.2593 | 0.2957 | -0.5365 | 0.4491 | -0.2894 | 0.3779 | -0.1678 | 0.2083 | -0.0244 | 0.1367 |
| 20 | 1 | -0.3621 | 0.5946 | -0.3418 | 0.5493 | -0.4202 | 0.4045 | -0.1538 | 0.2479 | -0.0329 | 0.1334 |

ภาพที่ 3.6 พารามิเตอร์ LPC-10

3.3.2 ส่วนการหาค่าสัมประสิทธิ์คู่เส้นสเปกตรัม

สัมประสิทธิ์คู่เส้นสเปกตรัม เป็นพารามิเตอร์รูปแบบหนึ่ง que พัฒนาจากสัมประสิทธิ์ทำนาย พ้นระยะเชิงเส้น (LPC) เนื่องจากการประมาณค่าพารามิเตอร์ของ LPC อาจมีไม่สม่ำเสมอของสัญญาณ ส่งผลให้คุณภาพเสียงต่ำลงได้ ในขณะที่พารามิเตอร์ที่ได้จาก LSP มีคุณสมบัติที่ชัดเจนชัดกว่าคือ ค่าพารามิเตอร์อยู่ภายใต้ขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์และสามารถตรวจสอบ เสถียรภาพของวงจรกรองได้ง่าย อีกทั้งยังแสดงอยู่ในรูปเชิงความถี่จึงสามารถนำไปใช้ในการหา คุณสมบัติที่แน่นอนในระบบการรับรู้ของเสียงได้ ดังกล่าวไว้แล้วในบทที่ 2

ค่าพารามิเตอร์ LPC-10 ที่คำนวณไว้ในขั้นตอนที่ 3.3.1 จะถูกนำมาคำนวณหาสัมประสิทธิ์ คู่เส้นสเปกตรัม โดยในแต่ละเฟรมของสัญญาณเสียงพูดจะประกอบด้วยค่าสัมประสิทธิ์คู่เส้น สเปกตรัม จำนวน 10 ค่า (LSP-10) ซึ่งผลการหาพารามิเตอร์ด้วยเทคนิคสัมประสิทธิ์คู่เส้น สเปกตรัม แสดงตัวอย่างดังภาพที่ 3.7

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 0.0460 | 0.0817 | 0.1335 | 0.1741 | 0.2163 | 0.2531 | 0.2957 | 0.3326 | 0.3780 | 0.4122 |
| 2 | 0.0563 | 0.0894 | 0.1288 | 0.1701 | 0.2183 | 0.2654 | 0.3102 | 0.3481 | 0.3831 | 0.4159 |
| 3 | 0.0580 | 0.0920 | 0.1319 | 0.1752 | 0.2210 | 0.2638 | 0.3102 | 0.3444 | 0.3932 | 0.4180 |
| 4 | 0.0421 | 0.0794 | 0.1420 | 0.1842 | 0.2279 | 0.2693 | 0.3149 | 0.3497 | 0.3931 | 0.4200 |
| 5 | 0.0482 | 0.0745 | 0.1236 | 0.1741 | 0.2338 | 0.2657 | 0.3154 | 0.3490 | 0.3864 | 0.4194 |
| 6 | 0.0578 | 0.0971 | 0.1433 | 0.1834 | 0.2235 | 0.2621 | 0.3081 | 0.3475 | 0.3811 | 0.4102 |
| 7 | 0.0471 | 0.0820 | 0.1365 | 0.1714 | 0.2229 | 0.2626 | 0.3054 | 0.3522 | 0.3955 | 0.4181 |
| 8 | 0.0506 | 0.0924 | 0.1432 | 0.1768 | 0.2130 | 0.2562 | 0.3115 | 0.3430 | 0.3876 | 0.4189 |
| 9 | 0.0617 | 0.0880 | 0.1457 | 0.1775 | 0.2340 | 0.2778 | 0.3114 | 0.3412 | 0.3988 | 0.4187 |
| 10 | 0.0568 | 0.0967 | 0.1342 | 0.1674 | 0.2291 | 0.2649 | 0.3094 | 0.3456 | 0.3858 | 0.4160 |
| 11 | 0.0486 | 0.0862 | 0.1474 | 0.1802 | 0.2196 | 0.2581 | 0.3136 | 0.3500 | 0.3911 | 0.4191 |
| 12 | 0.0548 | 0.0855 | 0.1287 | 0.1793 | 0.2235 | 0.2629 | 0.3141 | 0.3400 | 0.3828 | 0.4190 |
| 13 | 0.0546 | 0.0850 | 0.1344 | 0.1731 | 0.2276 | 0.2608 | 0.3013 | 0.3329 | 0.3814 | 0.4121 |
| 14 | 0.0529 | 0.0889 | 0.1459 | 0.1773 | 0.2254 | 0.2706 | 0.3089 | 0.3440 | 0.3865 | 0.4185 |
| 15 | 0.0520 | 0.0840 | 0.1457 | 0.1884 | 0.2375 | 0.2678 | 0.3059 | 0.3456 | 0.3802 | 0.4193 |
| 16 | 0.0599 | 0.0858 | 0.1378 | 0.1749 | 0.2142 | 0.2505 | 0.3094 | 0.3453 | 0.3967 | 0.4250 |
| 17 | 0.0548 | 0.0882 | 0.1279 | 0.1615 | 0.2197 | 0.2642 | 0.3035 | 0.3401 | 0.3873 | 0.4186 |
| 18 | 0.0505 | 0.0820 | 0.1380 | 0.1733 | 0.2225 | 0.2596 | 0.3004 | 0.3398 | 0.3839 | 0.4111 |
| 19 | 0.0516 | 0.0773 | 0.1315 | 0.1772 | 0.2324 | 0.2735 | 0.3184 | 0.3545 | 0.3869 | 0.4180 |
| 20 | 0.0586 | 0.0916 | 0.1453 | 0.1795 | 0.2191 | 0.2593 | 0.3108 | 0.3512 | 0.3849 | 0.4174 |

ภาพที่ 3.7 พารามิเตอร์ LSP-10

3.4 การออกแบบโครงข่าย LVO

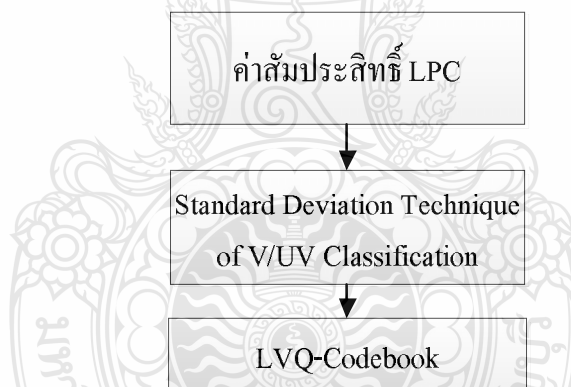
การออกแบบโครงข่ายประสาท LVO สำหรับการบีบอัดสัญญาณสัญญาณเสียง ในวิจัยนี้ได้ ออกแบบโครงข่ายประสาทเป็น 2 แบบ แบบแรก คือ การออกแบบโครงข่ายประสาทสำหรับการบีบอัดพารามิเตอร์ LPC-10 โดยจะแสดงในขั้นตอนที่ 3.4.1 และแบบที่สอง คือ การออกแบบโครงข่ายประสาท LVO สำหรับการบีบอัดพารามิเตอร์ LSP-10 ซึ่งจะแสดงในขั้นตอนที่ 3.4.2

3.4.1 การออกแบบโครงข่าย LVO สำหรับการบีบอัดพารามิเตอร์ LPC-10 มีขั้นตอนดังนี้

การออกแบบโครงข่ายประสาท LVO สำหรับการบีบอัดสัญญาณเสียงนั้น ก่อนเข้ากระบวนการฝึกสอน LVO ต้องมีอินพุตเวกเตอร์ เมื่อได้ค่าอินพุตเวกเตอร์แล้วก็ทำการกำหนดกลุ่มให้กับอินพุตเวกเตอร์ หลังจากนั้นนำข้อมูลอินพุตเวกเตอร์ที่แบ่งกลุ่มเสร็จสมบูรณ์เข้ากระบวนการฝึกสอนด้วยโครงข่าย LVO ซึ่งมีลำดับขั้นตอนการทดลองในหัวข้อถัดไป

1) นำพารามิเตอร์ LPC-10 ที่คำนวณไว้ในขั้นตอน 3.3.1 มาสร้างเป็นอินพุตเวกเตอร์ จากนั้นนำพารามิเตอร์ LPC-10 มาทำการแบ่งกลุ่มให้กับอินพุตเวกเตอร์หรือเรียกว่าการแบ่งกลุ่มให้กับพารามิเตอร์ LPC-10 โดยงานวิจัยนี้ได้ทำศึกษาและการวิเคราะห์หาเทคนิคเพื่อใช้ในการกำหนด

คุณลักษณะของพารามิเตอร์ LPC-10 ซึ่งมีเทคนิคด้วยกัน 4 เทคนิค เทคนิคที่ 1 คือ 1thAverang Technique of V/UV Classification เทคนิคที่ 2 คือ 2ndAverang Technique of V/UV Classification เทคนิคที่ 3 คือ Value Range Technique of V/UV Classification และเทคนิคที่ 4 คือ Standard Deviation Technique of V/UV Classification ผลจากการทดลองการแบ่งกลุ่มคำศัพท์ LPC-10 พบว่าเทคนิคค่าความเบี่ยงเบนมาตรฐาน (Standard Deviation Technique of V/UV Classification) [18] เป็นเทคนิคที่ดีที่สุด เมื่อเปรียบเทียบกับเทคนิค Prosodic Feature Technique มีคุณภาพความถูกต้องที่ 89.29 % ในการแบ่งกลุ่มคุณลักษณะของพารามิเตอร์ LPC-10 โดยเทคนิคนี้จะแบ่งกลุ่มคุณลักษณะของพารามิเตอร์ LPC-10 ออกเป็น 2 กลุ่ม กลุ่มแรก คือ กลุ่มของเสียงโฆษะ (Voiced) และกลุ่มของเสียงอโฆษะ (Unvoiced) และจากผลการวิเคราะห์ดังกล่าว จึงได้นำเทคนิคค่าความเบี่ยงเบนมาตรฐานมาใช้ในการแบ่งกลุ่มของพารามิเตอร์ LPC-10 เพื่อใช้เป็นส่วนหนึ่งก่อนได้รับการฝึกสอนในโครงข่าย LVQ ดังภาพที่ 3.8 และมีรายละเอียดของเทคนิคค่าความเบี่ยงเบนมาตรฐานแสดงดังสมการที่ 3.1 และ 3.2



ภาพที่ 3.8 การแบ่งกลุ่มของคำศัพท์ LPC-10

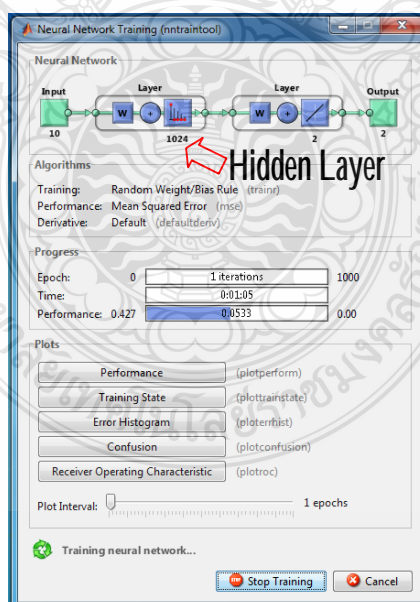
- เทคนิคค่าความเบี่ยงเบนมาตรฐาน (Standard Deviation Technique of V/UV Classification)

$$SD_n = \sqrt{\frac{\sum_{n=1}^w (\Psi(n) - \bar{X}(n))^2}{w-1}} \quad (3.1)$$

$$Th = max_{SD} \times ratio \quad (3.2)$$

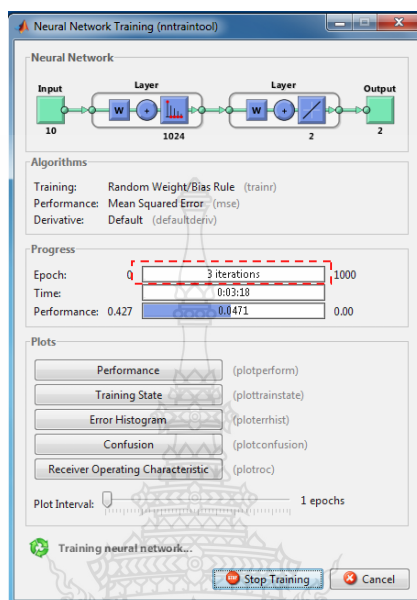
- โดยที่ Ψ คือ พารามิเตอร์ LPC-10
- \bar{X}_n คือ ค่าเฉลี่ยของพลังงานทั้งหมดใน 1 เฟรม
- n คือ จำนวนครั้งของการชักตัวอย่าง
- SD_n คือ ค่าเบี่ยงมาตรฐานทั้งหมดใน 1 เฟรม
- max_{SD} คือ ค่าเบี่ยงมาตรฐานของค่าพารามิเตอร์ LPC-10 ของเฟรมที่มีค่าสูงสุด
- SD คือ ค่าเบี่ยงมาตรฐานในแต่ละเฟรม
- w คือ จำนวนการชักตัวอย่างทั้งหมดใน 1 เฟรม
- Th คือ ค่าการตัดสินใจ (Threshold)
- $ratio$ คือ เปอร์เซนต์ของการตัดสินใจ

2) นำพารามิเตอร์ LPC-10 ที่แบ่งกลุ่มสั้นสมบูรณ์เข้ากระบวนการฝึกสอน พร้อมกับกำหนดจำนวน Hidden layer ของโครงข่าย LVQ เพื่อการบีบอัดสัญญาณตามความสัมพันธ์ของ 2" โดยการกำหนดขนาดของจำนวนตั้งแต่ 1024 512 256 128 และ 64 โดยจำนวนของ Hidden Layer แสดงดังภาพที่ 3.9



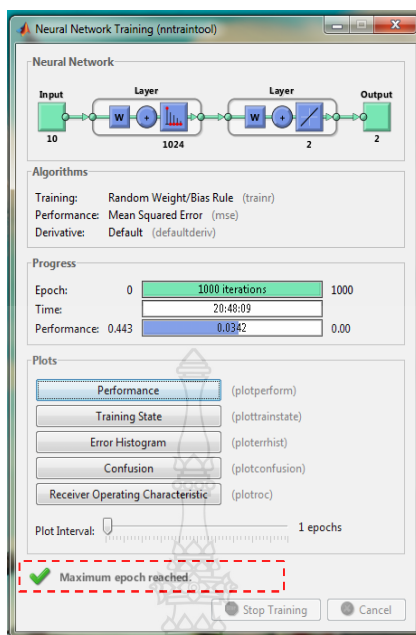
ภาพที่ 3.9 Hidden layer ในการบีบอัดพารามิเตอร์ LPC-10

3) กำหนดจำนวนการเรียนรู้โครงข่าย LVO จากคำสั่ง **train** ในโปรแกรม โดยจะดำเนินการฝึกฝนจำนวน 1000 ครั้ง เพื่อปรับค่าน้ำหนักและหาระยะทางระหว่างปมประสาท ดังภาพที่ 3.10



ภาพที่ 3.10 การดำเนินการฝึกฝนข้อมูลของโครงข่าย LVO สำหรับการบีบอัดพารามิเตอร์ LPC-10

4) การแสดงสถานะการเรียนรู้โครงข่ายโดยเปิดหน้าต่าง Neural Network Training โปรแกรมจะดำเนินการฝึกฝน โครงข่ายรอนจนจบวนการสิ้นสุดโดยสังเกตจากตำแหน่งล่างจะปรากฏข้อความว่า **Maximum Epoch Reached** ดังภาพที่ 3.11



ภาพที่ 3.11 จุดสิ้นสุดของกระบวนการฝึกฝนข้อมูลโครงข่าย LVQ ในการบีบอัดพารามิเตอร์ LPC-10

5) โปรแกรมจะดำเนินการฝึกฝนเสร็จสมบูรณ์ จะแสดงค่าผลลัพธ์โค้ดบู้คของการบีบอัดพารามิเตอร์ LPC-10 โดยค่าสัมประสิทธิ์จะถูกจะเก็บในไฟล์ Excel ดังภาพที่ 3.12

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | -0.81745 | -0.82657 | -1.61857 | -0.82182 | -0.81676 | -1.29864 | -0.82486 | -0.81695 | -0.81799 | -0.8191 | -0.8181 | -0.81808 | -0.8211 |
| 2 | 1.716338 | 1.715763 | 1.560466 | 1.719377 | 1.717385 | 0.84021 | 1.720088 | 1.713868 | 1.716725 | 1.717376 | 1.716114 | 1.715425 | 1.71637 |
| 3 | -2.03128 | -2.02699 | -1.50016 | -2.03372 | -2.03419 | -0.75769 | -2.03738 | -2.03088 | -2.03321 | -2.03378 | -2.03101 | -2.03349 | -2.03246 |
| 4 | 2.264286 | 2.257591 | 1.290798 | 2.267761 | 2.267669 | 0.808309 | 2.26813 | 2.263739 | 2.265593 | 2.266303 | 2.2649 | 2.265307 | 2.264245 |
| 5 | -1.90075 | -1.89656 | -0.88361 | -1.90325 | -1.90216 | -0.56578 | -1.90614 | -1.89963 | -1.90235 | -1.90172 | -1.90114 | -1.90013 | -1.90201 |
| 6 | 1.625051 | 1.621469 | 0.591383 | 1.626689 | 1.625257 | 0.441611 | 1.62865 | 1.623445 | 1.626104 | 1.625367 | 1.624316 | 1.624783 | 1.624953 |
| 7 | -1.08132 | -1.08113 | -0.53115 | -1.08118 | -1.08074 | -0.50253 | -1.0825 | -1.08023 | -1.08142 | -1.08197 | -1.08089 | -1.08014 | -1.08102 |
| 8 | 0.70505 | 0.706907 | 0.445838 | 0.706172 | 0.70528 | 0.519098 | 0.705463 | 0.704795 | 0.705623 | 0.705799 | 0.705094 | 0.704094 | 0.705732 |
| 9 | -0.3126 | -0.31259 | -0.17402 | -0.31392 | -0.31275 | -0.25436 | -0.31426 | -0.31238 | -0.31226 | -0.31236 | -0.31239 | -0.3119 | -0.3132 |
| 10 | 0.123108 | 0.123189 | 0.094304 | 0.122809 | 0.123121 | 0.12164 | 0.123004 | 0.123066 | 0.121928 | 0.122522 | 0.122954 | 0.122771 | 0.123172 |

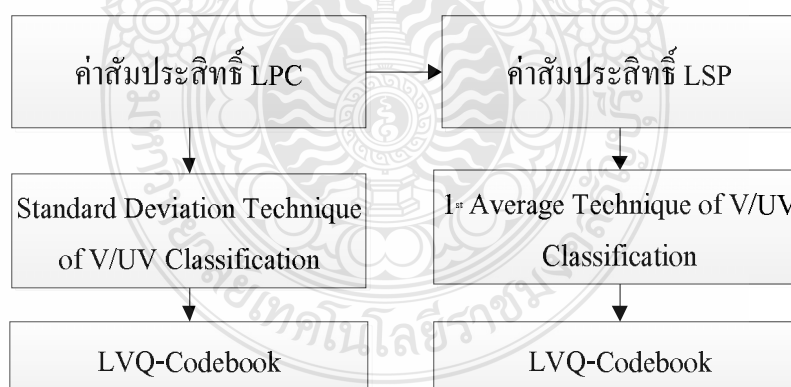
ภาพที่ 3.12 ชุดโค้ดบู้คของการบีบอัดพารามิเตอร์ LPC-10 ในไฟล์ Excel

3.4.2 การออกแบบโครงข่าย LVQ สำหรับการบีบอัดพารามิเตอร์ LSP-10 มีขั้นตอนการดังนี้

การออกแบบโครงข่ายประสาท LVQ สำหรับการบีบอัดสัญญาณเสียงนั้น ก่อนเข้ากระบวนการฝึกสอน LVQ ต้องมีอินพุตเวกเตอร์ เมื่อได้ค่าอินพุตเวกเตอร์แล้วก็ทำการกำหนดคกลุ่มให้กับอินพุต

เวกเตอร์ หลังจากนั้นนำข้อมูลอินพุตเวกเตอร์ที่แบ่งกลุ่มเสร็จสมบูรณ์เข้ากระบวนการฝึกสอนด้วยโครงข่าย LVQ ซึ่งมีลำดับขั้นตอนการทดลองดังนี้

1) นำพารามิเตอร์ LSP-10 ที่คำนวณไว้ในขั้นตอน 3.3.2 มาสร้างเป็นอินพุตเวกเตอร์ จากนั้นนำพารามิเตอร์ LSP-10 มาทำการแบ่งกลุ่มให้กับอินพุตเวกเตอร์ หรือ เรียกว่าการแบ่งกลุ่มให้กับค่าสัมประสิทธิ์ LPC-10 โดยงานวิจัยนี้ได้ทำศึกษาและการวิเคราะห์หาเทคนิคในการกำหนดคุณลักษณะของค่าสัมประสิทธิ์ LSP-10 ซึ่งมีเทคนิคด้วยกัน 4 เทคนิค เทคนิคที่ 1 คือ 1st Averang Technique of V/UV Classification เทคนิคที่ 2 คือ 2nd Averang Technique of V/UV Classification เทคนิคที่ 3 คือ Value Range Technique of V/UV Classification และเทคนิคที่ 4 คือ Standard Deviation Technique of V/UV Classification ผลจากการทดลองการแบ่งกลุ่มค่าสัมประสิทธิ์ LSP-10 พบว่าเทคนิคค่าเฉลี่ย (1st Average Technique of V/UV Classification) [19] เป็นเทคนิคที่ดีที่สุดในการแบ่งกลุ่มคุณลักษณะของพารามิเตอร์ LSP-10 เมื่อเปรียบเทียบกับเทคนิค Prosodic Feature Technique โดยมีคุณภาพความถูกต้อง 83.05 % โดยเทคนิคนี้จะแบ่งกลุ่มคุณลักษณะของพารามิเตอร์ LSP-10 ออกเป็น 2 กลุ่ม กลุ่มแรก คือ กลุ่มของเสียงโฆษะและกลุ่มของเสียงอโฆษะและจากผลการวิเคราะห์ดังกล่าว จึงได้นำเทคนิคค่าเฉลี่ย มาใช้ในการแบ่งกลุ่มของพารามิเตอร์ LSP-10 เพื่อใช้เป็นส่วนหนึ่งก่อนได้รับการฝึกสอนในโครงข่าย LVQ ดังภาพที่ 3.8 และมีรายละเอียดของเทคนิคค่าเฉลี่ยจะแสดงดังสมการที่ 3.3 และ 3.4



ภาพที่ 3.13 การแบ่งกลุ่มของค่าสัมประสิทธิ์ LSP-10

- เทคนิคค่าเฉลี่ย (1st Average Technique of V/UV Classification)

$$\bar{X}_n = \frac{1}{w} \sum_{n=1}^w \Psi(n) \quad (3.3)$$

$$Th = \max_{\bar{X}} \times ratio \quad (3.4)$$

โดยที่ \bar{X}_n คือ ค่าเฉลี่ยของพลังงานทั้งหมดใน 1 เฟรม

Ψ คือ ค่าพารามิเตอร์ LSP-10

n คือ จำนวนครั้งของการชักตัวอย่าง

\bar{X} คือ ค่าเฉลี่ยในแต่ละเฟรม (Average Frame)

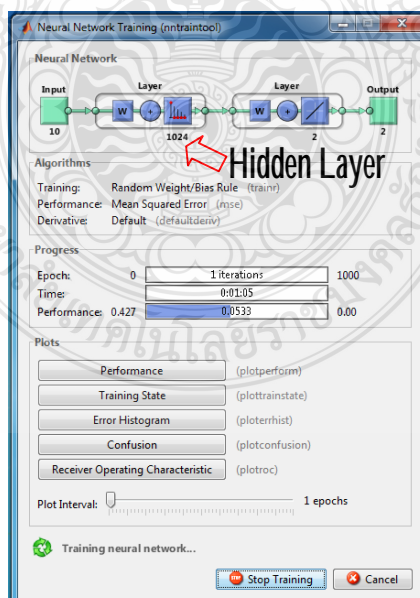
w คือ จำนวนการชักตัวอย่างทั้งหมดใน 1 เฟรม

$\max_{\bar{X}}$ คือ ค่าสูงสุดของค่าเฉลี่ยของพารามิเตอร์ LSP-10

Th คือ ค่าการตัดสินใจ (Threshold)

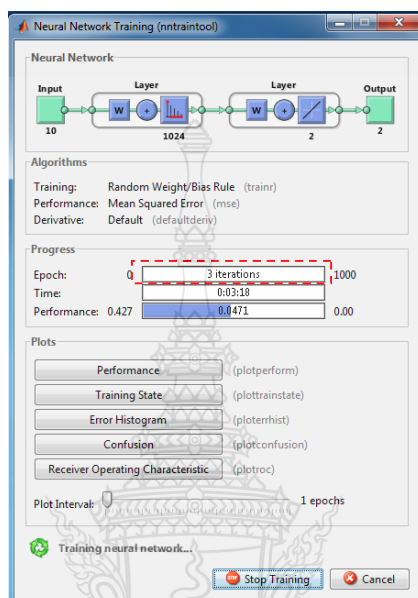
$ratio$ คือ เปรอเซ็นต์ของการตัดสินใจ

2) นำพารามิเตอร์ LSP-10 ที่แบ่งกลุ่มคลื่นสมบรูณ์เข้ากระบวนการฝึกสอน พร้อมกับกำหนดจำนวน Hidden Layer ของโครงข่าย LVQ เพื่อการบีบอัดสัญญาณตามความสัมพันธ์ของ 2" โดยการกำหนดขนาดของจำนวนตั้งแต่ 1024 512 256 128 และ 64 โดยจำนวนของ Hidden Layer แสดงดังภาพที่ 3.9



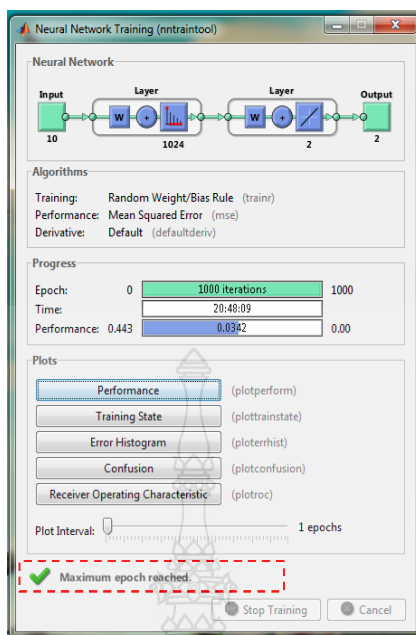
ภาพที่ 3.14 Hidden layer ของการบีบอัดพารามิเตอร์ LSP-10

3) กำหนดจำนวนการเรียนรู้โครงข่ายจัดการตนเองจากคำสั่ง **Train** ในโปรแกรม โดยจะดำเนินการฝึกฝนจำนวน **1000** ครั้ง เพื่อปรับค่าน้ำหนักและหาระยะทางระหว่างปมประสาท ดังภาพที่ 3.15



ภาพที่ 3.15 การดำเนินการฝึกฝนข้อมูลของโครงข่าย LVQ สำหรับการบีบอัดพารามิเตอร์ LSP-10

4) การแสดงสถานการณ์เรียนรู้โครงข่ายโดยเปิดหน้าต่าง **Neural Network Training** โปรแกรมจะดำเนินการฝึกฝน โครงข่ายรอนขบวนการสิ้นสุดโดยสังเกตจากตำแหน่งล่างจะปรากฏข้อความว่า **Maximum Epoch Reached** ดังภาพที่ 3.16



ภาพที่ 3.16 จุดสิ้นสุดของกระบวนการฝึกฝนข้อมูลโครงข่าย LVQ ในการบีบอัดพารามิเตอร์ LSP-10

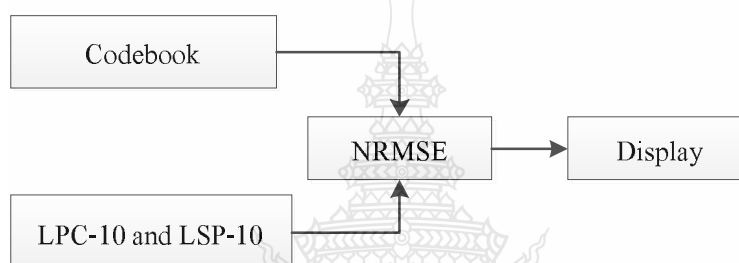
5) โปรแกรมจะดำเนินการฝึกฝนเสร็จสมบูรณ์ จะแสดงค่าผลลัพธ์ชุดโค้ดบुक (Codebook) ของการบีบอัด LSP-10 โดยค่าสัมประสิทธิ์จะถูกเก็บในไฟล์ Excel ดังภาพที่ 3.17

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 0.060152 | 0.062637 | 0.06265 | 0.062408 | 0.062515 | 0.062556 | 0.062425 | 0.062577 | 0.062596 | 0.062515 | 0.062627 | 0.062585 | 0.062593 |
| 2 | 0.100901 | 0.093276 | 0.093313 | 0.093263 | 0.09327 | 0.093262 | 0.093214 | 0.093226 | 0.093355 | 0.09327 | 0.093147 | 0.093246 | 0.093459 |
| 3 | 0.156423 | 0.13742 | 0.137265 | 0.137256 | 0.13717 | 0.13736 | 0.137362 | 0.137208 | 0.137274 | 0.13717 | 0.137244 | 0.137323 | 0.13735 |
| 4 | 0.18486 | 0.162108 | 0.162194 | 0.161933 | 0.161867 | 0.162108 | 0.162143 | 0.162234 | 0.162212 | 0.161867 | 0.16191 | 0.162043 | 0.162129 |
| 5 | 0.246957 | 0.212107 | 0.212037 | 0.211997 | 0.212028 | 0.21208 | 0.212185 | 0.211866 | 0.212016 | 0.212028 | 0.212129 | 0.212105 | 0.212061 |
| 6 | 0.294673 | 0.244736 | 0.244815 | 0.244717 | 0.244844 | 0.244687 | 0.244795 | 0.244712 | 0.244808 | 0.244844 | 0.244884 | 0.244997 | 0.244734 |
| 7 | 0.338015 | 0.289023 | 0.288915 | 0.288897 | 0.289201 | 0.289029 | 0.28902 | 0.288902 | 0.288997 | 0.288992 | 0.289201 | 0.289162 | 0.288941 |
| 8 | 0.376118 | 0.337059 | 0.337024 | 0.336927 | 0.336824 | 0.33703 | 0.337069 | 0.33696 | 0.33683 | 0.336824 | 0.33711 | 0.337042 | 0.337109 |
| 9 | 0.408437 | 0.371168 | 0.371028 | 0.3713 | 0.37102 | 0.37133 | 0.371151 | 0.371156 | 0.371149 | 0.37102 | 0.371133 | 0.371224 | 0.371324 |
| 10 | 0.431324 | 0.405788 | 0.405832 | 0.405856 | 0.4059 | 0.405852 | 0.405865 | 0.405862 | 0.405855 | 0.4059 | 0.405916 | 0.40593 | 0.405984 |

ภาพที่ 3.17 ชุดโค้ดบुकของการบีบอัดพารามิเตอร์ LSP-10 ในไฟล์ Excel

3.5 การวัดค่าผิดพลาด (Error)

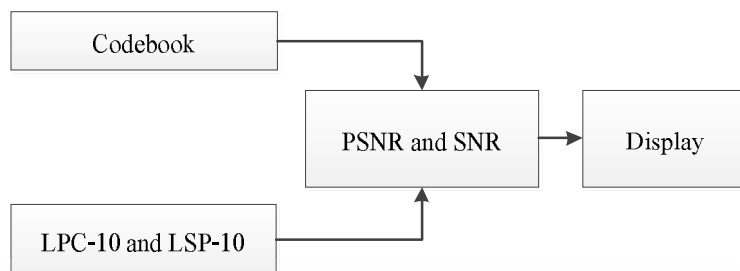
งานวิจัยนี้ใช้เทคนิค **NRMSE** ในการหาค่าผิดพลาดระหว่างสัญญาณเสียงก่อนการบีบอัดกับสัญญาณเสียงที่ผ่านการบีบอัดแล้ว ซึ่งเทคนิค **NRMSE** เป็นวิธีที่ใช้ในการคำนวณหาประสิทธิภาพของการแทนค่าข้อมูลที่ขาดหายไป โดยรายละเอียดของสมการได้ถูกกล่าวไว้ในบทที่ 2 สำหรับการวัดค่าความผิดพลาดนั้น ถ้าค่าของ **NRMSE** ที่มีค่าต่ำจะบ่งบอกถึงคุณภาพของการบีบอัดที่ดี โดยในงานวิจัยเป็นการวัดความผิดพลาดระหว่างการบีบอัดสัญญาณเสียงในระดับโค้ดบุ๊กในระดับต่างๆ กับพารามิเตอร์ **LPC-10** และ **LSP-10** ต้นฉบับ โดยมีขั้นตอนแสดงดังภาพที่ 3.18



ภาพที่ 3.18 การวัดค่าผิดพลาดด้วยวิธี **NRMSE**

3.6 การวัดค่าประสิทธิภาพของสัญญาณ

สำหรับการหาค่าประสิทธิภาพของสัญญาณเสียงก่อนการบีบอัดและสัญญาณเสียงที่ผ่านการบีบอัดแล้ว โดยงานวิจัยนี้ใช้เทคนิคในการหาค่าประสิทธิภาพ คือเทคนิค **PSNR** และ **SNR** สำหรับ **PSNR** เป็นเทคนิคที่ใช้พิจารณาค่าอัตราส่วนระหว่างค่าสูงสุดของสัญญาณ และความแตกต่างของด้วยสัญญาณรบกวนเมื่อเทียบกับสัญญาณเสียงต้นฉบับ และ เทคนิค **SNR** เป็นเทคนิคการพิจารณาค่าอัตราส่วนสัญญาณเสียงต้นฉบับต่อสัญญาณรบกวน โดยที่เทคนิคทั้งสอง มีหน่วยเป็นเดซิเบล (**dB**) สำหรับการวัดค่าประสิทธิภาพของสัญญาณเสียง ถ้าค่าของ **PSNR** และ **SNR** มีค่าเดซิเบลที่สูง จะเป็นการบ่งบอกคุณภาพในการบีบอัดที่มีประสิทธิภาพที่ดี โดยในงานวิจัยเป็นการวัดค่าประสิทธิภาพระหว่างการบีบอัดสัญญาณเสียงในระดับโค้ดบุ๊กในระดับต่างๆ กับพารามิเตอร์ **LPC-10** และ **LSP-10** ต้นฉบับและมีขั้นตอนแสดงดังภาพที่ 3.19



ภาพที่ 3.19 การหาค่าประสิทธิภาพของสัญญาณ PSNR และ SNR



บทที่ 4

ผลการวิจัย

การวิจัยเรื่องการพัฒนาเทคนิคการบีบอัดสัญญาณเสียงพูดเสียงพูดในมาตรฐาน LPC-10 และ LSP-10 ได้นำเทคนิค LVQ มาประยุกต์ใช้ในการเข้ารหัสสัญญาณเสียงพูดชนิด LPC-10 และ LSP-10 โดยทำการทดสอบการบีบอัดกับสัญญาณเสียงพูด ที่มีรูปประโยคการพูดแบบเล่าเรื่องในสถานการณ์ที่แตกต่างกัน จำนวน 10 คน ประกอบด้วยผู้ชาย (M) จำนวน 5 คน และผู้หญิง (FM) จำนวน 5 คน โดยระดับการบีบอัดสัญญาณจะขึ้นอยู่กับโค้ดบุ๊กที่แตกต่างกัน ซึ่งรายละเอียดของผลการวิจัยในส่วนถัดไป

4.1 ขนาดการสร้างโค้ดบุ๊ก

การทดลองในการบีบอัดสัญญาณเสียงพูด โดยใช้ LVQ ได้สร้างโค้ดบุ๊กตามความสัมพันธ์ของ 2^n ซึ่งระดับโค้ดบุ๊กจะมีขนาด 5 ระดับ ซึ่งรายละเอียดแสดงดังตารางที่ 4.1

ตารางที่ 4.1 การแบ่งระดับของโค้ดบุ๊ก

| ระดับที่ | จำนวนโค้ดบุ๊ก (Codebook) |
|----------|--------------------------|
| 1 | $2^{10}=1024$ |
| 2 | $2^9=512$ |
| 3 | $2^8=256$ |
| 4 | $2^7=128$ |
| 5 | $2^6=64$ |

จากตารางที่ 4.1 เมื่อพารามิเตอร์ LPC-10 และ พารามิเตอร์ LSP-10 ผ่านการบีบอัดสัญญาณเสียงเสรีจสมบูรณ์ ตัวแทนของตำแหน่งพารามิเตอร์ในการบีบอัดสัญญาณเสียงจะเท่ากับจำนวน โค้ดบุ๊ก จากนั้นนำตัวแทนใหม่คืนกลับเพื่อเปรียบเทียบกับพารามิเตอร์ LPC-10 และ พารามิเตอร์ LSP-10 ดั้งเดิม ด้วยทดสอบการหาค่าความผิดพลาดด้วย NRMSE และวิเคราะห์หาประสิทธิภาพของการบีบอัดสัญญาณเสียงด้วย PSNR และ SNR ซึ่งรายละเอียดจะกล่าวถึงในส่วนถัดไป

4.2 ผลทดสอบการหาค่าผิดพลาด

การทดลองการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค LVQ ใช้จำนวนเสียงพูดจำนวน 10 คน ประกอบด้วยผู้ชาย (M) จำนวน 5 คน และผู้หญิง (FM) จำนวน 5 คน แล้วทำการบันทึกเสียงพูดจะใช้สถานที่บันทึกเสียงที่เดียวกัน และความถี่ในการบันทึกเสียงพูด 8 กิโลเฮิรตซ์ ขนาด 8 บิต แบบโมโน โดยทำการจัดเก็บไฟล์เสียงในรูปแบบ WAV ไฟล์ นำสัญญาณที่ได้มาแบ่งเฟรมข้อมูล โดยในงานวิจัยนี้เลือกใช้การแบ่งเฟรมข้อมูล 200 ตัวอย่าง (Sample) ต่อหน่วยเฟรม นำเฟรมข้อมูลที่ได้เข้าสู่กระบวนการวิเคราะห์หาพารามิเตอร์ LPC-10 และ LSP-10 ตามลำดับ จากนั้นนำพารามิเตอร์ LPC-10 และพารามิเตอร์ LSP-10 มาทำการบีบอัดพารามิเตอร์โดยใช้ LVQ หลังจากทำการบีบอัดสัญญาณเสียงเสร็จสมบูรณ์แล้ว ทำการคืนกลับสัญญาณด้วยตัวแทน โค้ดบุ๊กทั้ง 5 ระดับ จากนั้นทำการหาค่าความผิดพลาดของการบีบอัดสัญญาณเสียงเปรียบเทียบกับพารามิเตอร์ LPC-10 และ LSP-10 ต้นฉบับ โดยใช้เทคนิคนำ NRMSE มาใช้ในการหาค่าผิดพลาด

ตารางที่ 4.2 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน

| Normalized Root Mean Squared Error | | | | | |
|------------------------------------|---------------------|---------|---------|---------|--------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 0.2000 | 0.2000 | 0.2001 | 0.2007 | 0.2041 |
| เสียงผู้ชายที่ 2 (M2) | 0.1931 | 0.1954 | 0.1945 | 0.1953 | 0.1979 |
| เสียงผู้ชายที่ 3 (M3) | 0.1842 | 0.1857 | 0.1868 | 0.1883 | 0.1897 |
| เสียงผู้ชายที่ 4 (M4) | 0.2023 | 0.2034 | 0.2035 | 0.2040 | 0.2065 |
| เสียงผู้ชายที่ 5 (M5) | 0.2110 | 0.2113 | 0.2113 | 0.2140 | 0.2178 |
| ค่าเฉลี่ย (Average) | 0.19812 | 0.19916 | 0.19924 | 0.20046 | 0.2032 |

จากตารางที่ 4.2 แสดงผลลัพธ์ค่าความผิดพลาดของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่าความผิดพลาดที่แตกต่างกัน จากการทดลองของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.19812 ซึ่งเป็นค่าเฉลี่ยผิดพลาดที่ต่ำที่สุดของทุกๆ ระดับ ที่โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.19916 ซึ่งเป็นค่าเฉลี่ยความผิดพลาดที่เพิ่มมากขึ้นจากโค้ดบุ๊ก ที่ 1024 ตำแหน่ง และโค้ดบุ๊ก ที่ 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าความผิดพลาดที่เพิ่มมากขึ้นตามลำดับของโค้ดบุ๊ก

จากผลการทดลองนี้สัญญาณเสียงผู้ชายที่ 3 (M3) มีค่าความผิดพลาดที่น้อยที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าความผิดพลาดต่ำที่สุดมีค่าเท่ากับ 0.1842 ซึ่งน้อยกว่าค่าเฉลี่ยความผิดพลาดของสัญญาณเสียงพูดที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดความผิดพลาดที่สูงขึ้น และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความผิดพลาดที่ต่ำ

ตารางที่ 4.3 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน

| Normalized Root Mean Squared Error | | | | | |
|------------------------------------|---------------------|---------|--------|---------|---------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 0.1952 | 0.1954 | 0.1956 | 0.1970 | 0.1979 |
| เสียงผู้หญิงที่ 2 (FM2) | 0.1578 | 0.1611 | 0.1609 | 0.1616 | 0.1626 |
| เสียงผู้หญิงที่ 3 (FM3) | 0.1747 | 0.1753 | 0.1754 | 0.1754 | 0.1764 |
| เสียงผู้หญิงที่ 4 (FM4) | 0.1595 | 0.1605 | 0.1619 | 0.1629 | 0.1635 |
| เสียงผู้หญิงที่ 5 (FM5) | 0.2335 | 0.2375 | 0.2377 | 0.2387 | 0.2393 |
| ค่าเฉลี่ย (Average) | 0.18414 | 0.18596 | 0.1863 | 0.18712 | 0.18794 |

จากตารางที่ 4.3 แสดงผลลัพธ์ค่าความผิดพลาดของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่าความผิดพลาดที่ต่างกัน จากการทดลองของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้หญิง จำนวน 5 คน โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.18414 ซึ่งเป็นค่าเฉลี่ยความผิดพลาดที่ต่ำที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.18596 ซึ่งเป็นค่าเฉลี่ยผิดพลาดที่เพิ่มมากขึ้นจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าความผิดพลาดที่เพิ่มมากขึ้นตามลำดับของโค้ดบุ๊ก จากผลการทดลองนี้สัญญาณเสียงผู้หญิงที่ 4 (FM4) มีค่าความผิดพลาดที่น้อยที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าความผิดพลาดต่ำที่สุดมีค่าเท่ากับ 0.1595 ซึ่งน้อยกว่าค่าเฉลี่ยความผิดพลาดของสัญญาณเสียงพูดที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับ

ตำแหน่งมีจำนวนน้อยทำให้เกิดความผิดพลาดที่สูงขึ้น และถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความผิดพลาดที่ต่ำ

จากตารางที่ 4.2 และ 4.3 แสดงการวัดค่า NRMSE สรุปได้ว่าการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิงมีค่าความผิดพลาดที่น้อยกว่าเสียงผู้ชาย เพราะพารามิเตอร์ LPC-10 ที่ผ่านการบีบอัดเสร็จสมบูรณ์แล้วของผู้หญิงมีความใกล้เคียงกับต้นฉบับมากกว่าผู้ชาย

ตารางที่ 4.4 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน

| Normalized Root Mean Squared Error | | | | | |
|------------------------------------|---------------------|----------|--------|--------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 0.0189 | 0.0190 | 0.0190 | 0.0190 | 0.0192 |
| เสียงผู้ชายที่ 2 (M2) | 0.0201 | 0.0204 | 0.0204 | 0.0204 | 0.0205 |
| เสียงผู้ชายที่ 3 (M3) | 0.0181 | 0.0182 | 0.0182 | 0.0182 | 0.0184 |
| เสียงผู้ชายที่ 4 (M4) | 0.0194 | 0.0194 | 0.0195 | 0.0195 | 0.0195 |
| เสียงผู้ชายที่ 5 (M5) | 0.0205 | 0.0223 | 0.0227 | 0.0227 | 0.0227 |
| ค่าเฉลี่ย (Average) | 0.019525 | 0.020075 | 0.0202 | 0.0202 | 0.020275 |

จากตารางที่ 4.4 แสดงผลลัพธ์ค่าความผิดพลาดของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่าความผิดพลาดที่แตกต่างกัน จากการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้ชาย จำนวน 5 คน โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.019525 ซึ่งเป็นค่าเฉลี่ยความผิดพลาดที่ต่ำที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.020075 ซึ่งเป็นค่าเฉลี่ยผิดพลาดที่เพิ่มมากขึ้นจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าความผิดพลาดที่เพิ่มมากขึ้นตามลำดับของโค้ดบุ๊ก จากผลการทดลองนี้สัญญาณเสียงผู้ชายที่ 3 (M3) มีค่าความผิดพลาดที่น้อยที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าความผิดพลาดต่ำที่สุดมีค่าเท่ากับ 0.0181 ซึ่งน้อยกว่าค่าเฉลี่ยความผิดพลาดของสัญญาณเสียงพูดที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดความผิดพลาดที่สูงขึ้น และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผล

ให้ค่าความผิดพลาดที่ต่ำ ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความผิดพลาดที่ต่ำ

ตารางที่ 4.5 ค่า NRMSE ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน

| Normalized Root Mean Squared Error | | | | | |
|------------------------------------|---------------------|---------|---------|--------|---------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 0.0190 | 0.0190 | 0.0191 | 0.0193 | 0.0194 |
| เสียงผู้หญิงที่ 2 (FM2) | 0.0175 | 0.0177 | 0.0177 | 0.0175 | 0.0176 |
| เสียงผู้หญิงที่ 3 (FM3) | 0.0181 | 0.0182 | 0.0182 | 0.0182 | 0.0184 |
| เสียงผู้หญิงที่ 4 (FM4) | 0.0111 | 0.0112 | 0.0112 | 0.0113 | 0.0118 |
| เสียงผู้หญิงที่ 5 (FM5) | 0.0181 | 0.0180 | 0.0181 | 0.0182 | 0.0182 |
| ค่าเฉลี่ย (Average) | 0.01676 | 0.01682 | 0.01686 | 0.0169 | 0.01708 |

จากตารางที่ 4.5 แสดงผลลัพธ์ค่าความผิดพลาดของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่าความผิดพลาดที่ต่างกัน จากการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้หญิง จำนวน 5 คน โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.01676 ซึ่งเป็นค่าเฉลี่ยผิดพลาดที่ต่ำที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยผิดพลาดของสัญญาณเท่ากับ 0.01682 ซึ่งเป็นค่าเฉลี่ยผิดพลาดที่เพิ่มมากขึ้นจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าความผิดพลาดที่เพิ่มมากขึ้นตามลำดับของโค้ดบุ๊ก จากผลการทดลองนี้สัญญาณเสียงผู้หญิงที่ 4 (FM4) มีค่าความผิดพลาดที่น้อยที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าความผิดพลาดต่ำที่สุดมีค่าเท่ากับ 0.0111 ซึ่งน้อยกว่าค่าเฉลี่ยความผิดพลาดของสัญญาณเสียงพูดที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดความผิดพลาดที่สูงขึ้น และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความผิดพลาดที่ต่ำ

จากตารางที่ 4.4 และ 4.5 แสดงการวัดค่า NRMSE สรุปได้ว่า การบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิงมีค่าความผิดพลาดที่น้อยกว่าเสียงผู้ชาย เพราะพารามิเตอร์ LSP-10 ที่ผ่านการบีบอัดเสร็จสมบูรณ์แล้วของผู้หญิงมีความใกล้เคียงกับต้นฉบับมากกว่าผู้ชาย

4.3 การหาประสิทธิภาพของการบีบอัดสัญญาณเสียงพูด

การทดลองได้ใช้เสียงพูดสำหรับการทดสอบหาประสิทธิภาพการบีบอัดสัญญาณเสียงพูด โดยใช้ LVQ จำนวน 10 คน ประกอบด้วยผู้ชาย (M) จำนวน 5 คน และผู้หญิง (FM) จำนวน 5 คน แล้วทำการบันทึกเสียงพูดจะใช้สถานที่บันทึกเสียงที่เดียวกัน และความถี่ในการบันทึกเสียงพูด 8 กิโลเฮิรตซ์ ขนาด 8 บิต แบบโมโน โดยทำการจัดเก็บไฟล์เสียงในรูปแบบ WAV ไฟล์ นำสัญญาณที่ได้มาแบ่งเฟรมข้อมูล โดยในงานวิจัยนี้เลือกใช้การแบ่งเฟรมข้อมูล 200 ตัวอย่าง (Sample) ต่อหน่วยเฟรม นำเฟรมข้อมูลที่ได้เข้าสู่กระบวนการวิเคราะห์หาพารามิเตอร์ LPC-10 และ พารามิเตอร์ LSP-10 ตามลำดับ จากนั้นนำพารามิเตอร์ LPC-10 และ LSP-10 มาทำการบีบอัดสัญญาณโดยใช้ LVQ หลังจากการบีบอัดสัญญาณเสียงเสร็จสมบูรณ์แล้ว ทำการคืนกลับสัญญาณด้วยตัวแทนโค้ดบุ๊กทั้ง 5 ระดับ จากนั้นทำวิเคราะห์หาประสิทธิภาพของการบีบอัดสัญญาณเปรียบเทียบกับ LPC-10 และ LSP-10 ต้นฉบับ โดยการหาประสิทธิภาพของสัญญาณด้วยเทคนิค SNR และ PSNR

ตารางที่ 4.6 ค่า PSNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน

| Peak signal to noise ratio | | | | | |
|----------------------------|---------------------|----------|----------|----------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 26.1162 | 26.0485 | 25.9966 | 25.9272 | 25.8613 |
| เสียงผู้ชายที่ 2 (M2) | 24.3278 | 24.3280 | 24.2865 | 24.2376 | 24.1985 |
| เสียงผู้ชายที่ 3 (M3) | 28.9432 | 28.8432 | 28.5604 | 28.4131 | 28.2671 |
| เสียงผู้ชายที่ 4 (M4) | 14.8912 | 14.8414 | 14.8373 | 14.8151 | 14.7111 |
| เสียงผู้ชายที่ 5 (M5) | 14.1414 | 14.1286 | 14.1286 | 14.0184 | 13.8649 |
| ค่าเฉลี่ย (Average) | 21.68396 | 21.63794 | 21.56188 | 21.48228 | 21.38058 |

จากตารางที่ 4.6 แสดงผลลัพธ์ค่า PSNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า PSNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้ชาย

จำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค PSNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 21.68396 dB ซึ่งเป็นค่าเฉลี่ยของ PSNR ที่มากที่สุด โค้ดบุ๊ก ที่ 512 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 21.63794 dB ซึ่งเป็นค่า PSNR น้อยลงจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า PSNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก โดยในการทดลองนี้สัญญาณเสียงผู้พูดผู้ชายที่ 3 (M3) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 28.9432 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ PSNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่าถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ PSNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนค่าตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ PSNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนค่าตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.7 ค่า PSNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน

| Peak signal to noise ratio | | | | | |
|----------------------------|---------------------|---------|----------|----------|---------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 16.9510 | 16.8964 | 16.8203 | 16.7629 | 16.7345 |
| เสียงผู้หญิงที่ 2 (FM2) | 27.8795 | 27.6994 | 27.7460 | 27.5003 | 27.3852 |
| เสียงผู้หญิงที่ 3 (FM3) | 16.4618 | 16.4297 | 16.4246 | 16.4239 | 16.3740 |
| เสียงผู้หญิงที่ 4 (FM4) | 28.1904 | 28.1891 | 28.1448 | 28.1385 | 28.0505 |
| เสียงผู้หญิงที่ 5 (FM5) | 13.4585 | 13.3134 | 13.3051 | 13.2681 | 13.2478 |
| ค่าเฉลี่ย (Average) | 20.58824 | 20.5056 | 20.48816 | 20.41874 | 20.3584 |

จากตารางที่ 4.7 แสดงผลลัพธ์ค่า PSNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า PSNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้หญิงจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค PSNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 20.58824 dB ซึ่งเป็นค่าเฉลี่ยของ PSNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 20.5056 dB ซึ่งเป็นค่า PSNR น้อยลงจากโค้ดบุ๊ก ที่ 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า PSNR ที่น้อยลง

ตามลำดับของโค้ดบุ๊ก โดยในการทดลองนี้สัญญาณเสียงผู้หญิงที่ 4 (FM4) มีประสิทธิภาพมากที่สุด ในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 28.1904 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ PSNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวสรุปได้ว่า ถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ PSNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ PSNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.8 ค่า PSNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน

| Peak signal to noise ratio | | | | | |
|----------------------------|---------------------|----------|----------|---------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 36.3107 | 36.2704 | 36.2358 | 36.2650 | 36.1713 |
| เสียงผู้ชายที่ 2 (M2) | 35.6218 | 35.5005 | 35.4856 | 35.5049 | 35.4462 |
| เสียงผู้ชายที่ 3 (M3) | 36.6292 | 36.5702 | 36.6013 | 36.5814 | 36.5014 |
| เสียงผู้ชายที่ 4 (M4) | 35.8325 | 35.8208 | 35.8054 | 35.7957 | 35.7777 |
| เสียงผู้ชายที่ 5 (M5) | 35.4409 | 34.6373 | 34.6575 | 34.6495 | 34.6730 |
| ค่าเฉลี่ย (Average) | 35.96702 | 35.75984 | 35.75712 | 35.7593 | 35.71392 |

จากตารางที่ 4.8 แสดงผลลัพธ์ค่า PSNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า PSNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้ชายจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค PSNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 35.96702 dB ซึ่งเป็นค่าเฉลี่ยของ PSNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 35.75984 dB ซึ่งเป็นค่า PSNR น้อยลงจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า PSNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก ในการทดลองนี้สัญญาณเสียงผู้ชายที่ 3 (M3) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 36.6292 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ PSNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ PSNR ที่สูง ซึ่ง

เหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ PSNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.9 ค่า PSNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน

| Peak signal to noise ratio | | | | | |
|----------------------------|---------------------|---------|----------|----------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 36.2349 | 36.2480 | 36.2281 | 36.1302 | 36.0360 |
| เสียงผู้หญิงที่ 2 (FM2) | 37.0569 | 36.9419 | 36.9458 | 37.0353 | 36.9986 |
| เสียงผู้หญิงที่ 3 (FM3) | 36.6292 | 36.6013 | 36.5702 | 36.5814 | 36.5014 |
| เสียงผู้หญิงที่ 4 (FM4) | 41.0208 | 40.9618 | 40.9408 | 40.8723 | 40.5206 |
| เสียงผู้หญิงที่ 5 (FM5) | 36.6771 | 36.6935 | 36.6317 | 36.5997 | 36.5878 |
| ค่าเฉลี่ย (Average) | 37.52378 | 37.4893 | 37.46332 | 37.44378 | 37.32888 |

จากตารางที่ 4.9 แสดงผลลัพธ์ค่า PSNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า PSNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้พูดผู้หญิง จำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค PSNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ PSNR เท่ากับ 37.52378 dB ซึ่งเป็นค่าเฉลี่ยของ PSNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ PSNR ซึ่งเป็นค่า PSNR น้อยลงจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก ที่ 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า PSNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก ในการทดลองนี้สัญญาณเสียงผู้หญิงที่ 4 (FM4) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 41.0208 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ PSNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ PSNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีความละเอียดของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ PSNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.10 ค่า SNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน

| Signal to noise ratio | | | | | |
|-----------------------|---------------------|----------|----------|----------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 14.1268 | 14.1263 | 14.1263 | 14.0962 | 13.9502 |
| เสียงผู้ชายที่ 2 (M2) | 13.0010 | 13.0012 | 12.9365 | 12.8776 | 12.7829 |
| เสียงผู้ชายที่ 3 (M3) | 13.2802 | 13.2802 | 13.1607 | 13.0913 | 13.0253 |
| เสียงผู้ชายที่ 4 (M4) | 12.8822 | 12.8324 | 12.8284 | 12.7021 | 12.7021 |
| เสียงผู้ชายที่ 5 (M5) | 12.8983 | 12.8855 | 12.7754 | 12.7754 | 12.6218 |
| ค่าเฉลี่ย (Average) | 13.2377 | 13.22512 | 13.16546 | 13.10852 | 13.01646 |

จากตารางที่ 4.10 แสดงผลลัพธ์ค่า SNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า SNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้ชายจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค SNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 13.2377 dB ซึ่งเป็นค่าเฉลี่ยของ SNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 13.22512 dB ซึ่งเป็นค่า SNR น้อยลงจากโค้ดบุ๊ก ที่ 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า SNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก โดยในการทดลองนี้สัญญาณเสียงผู้พูดผู้ชายที่ 1 (M1) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 14.1268 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ SNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ SNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ SNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.11 ค่า SNR ของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน

| Signal to noise ratio | | | | | |
|-------------------------|---------------------|----------|----------|----------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 12.8215 | 12.8189 | 12.8070 | 12.7461 | 12.7093 |
| เสียงผู้หญิงที่ 2 (FM2) | 14.9431 | 14.8885 | 14.7550 | 14.7550 | 14.7266 |
| เสียงผู้หญิงที่ 3 (FM3) | 13.8176 | 13.8125 | 13.8125 | 13.8118 | 13.7619 |
| เสียงผู้หญิงที่ 4 (FM4) | 15.0333 | 14.8532 | 14.8734 | 14.8195 | 14.6889 |
| เสียงผู้หญิงที่ 5 (FM5) | 11.8074 | 11.6170 | 11.6623 | 11.6623 | 11.5967 |
| ค่าเฉลี่ย (Average) | 13.68458 | 13.59802 | 13.58204 | 13.55894 | 13.49668 |

จากตารางที่ 4.11 แสดงผลลัพธ์ค่า SNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า SNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้หญิงจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค SNR โดยที่โค้ดบุ๊กที่ 1024 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 13.68458 dB ซึ่งเป็นค่าเฉลี่ยของ SNR ที่มากที่สุด โค้ดบุ๊กที่ 512 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 13.59802 dB ซึ่งเป็นค่า SNR น้อยลงจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า SNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก โดยในการทดลองนี้สัญญาณเสียงผู้หญิงที่ 4 (FM4) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 15.0333 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ SNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ SNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ SNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

จากตารางที่ 4.10 และ 4.11 แสดงการวัดค่า SNR สรุปได้ว่า การบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิงมีประสิทธิภาพที่มากกว่าเสียงผู้ชาย เพราะพารามิเตอร์ LSP-10 ที่ผ่านการบีบอัดสัญญาณเสียงเสร็จสมบูรณ์แล้วของผู้หญิงมีความใกล้เคียงกับต้นฉบับมากกว่าผู้ชาย

ตารางที่ 4.12 ค่า SNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน

| Signal to noise ratio | | | | | |
|-----------------------|---------------------|----------|----------|----------|----------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 31.6112 | 31.5917 | 31.5531 | 31.5823 | 31.4707 |
| เสียงผู้ชายที่ 2 (M2) | 31.1275 | 31.0062 | 30.9973 | 30.9925 | 30.9438 |
| เสียงผู้ชายที่ 3 (M3) | 32.2904 | 32.2239 | 32.2514 | 32.2383 | 32.1536 |
| เสียงผู้ชายที่ 4 (M4) | 31.4592 | 31.4321 | 31.4056 | 31.4093 | 31.3926 |
| เสียงผู้ชายที่ 5 (M5) | 30.9432 | 30.1795 | 30.0453 | 30.0310 | 30.0541 |
| ค่าเฉลี่ย (Average) | 31.4863 | 31.28668 | 31.25054 | 31.25068 | 31.20296 |

จากตารางที่ 4.12 แสดงผลลัพธ์ค่า SNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า SNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้ชายจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค SNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 31.4863 dB ซึ่งเป็นค่าเฉลี่ยของ SNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 31.28668 dB ซึ่งเป็นค่า SNR น้อยลงจากโค้ดบุ๊ก ที่ 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า SNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก ในการทดลองนี้สัญญาณเสียงผู้ชายที่ 3 (M3) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 32.2904 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ SNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ SNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ SNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

ตารางที่ 4.13 ค่า SNR ของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน

| Signal to noise ratio | | | | | |
|-------------------------|---------------------|----------|---------|----------|---------|
| เสียงพูด (Speech) | โค้ดบุ๊ก (Codebook) | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้หญิงที่ 1 (FM1) | 31.8940 | 31.8931 | 31.8664 | 31.7697 | 31.6873 |
| เสียงผู้หญิงที่ 2 (FM2) | 32.7204 | 32.6243 | 32.6271 | 32.7191 | 32.6510 |
| เสียงผู้หญิงที่ 3 (FM3) | 32.2904 | 32.2514 | 32.2239 | 32.2383 | 32.1536 |
| เสียงผู้หญิงที่ 4 (FM4) | 36.6372 | 36.5820 | 36.5722 | 36.4572 | 36.1373 |
| เสียงผู้หญิงที่ 5 (FM5) | 32.4145 | 32.4533 | 32.3952 | 32.3632 | 32.3513 |
| ค่าเฉลี่ย (Average) | 33.51563 | 33.47775 | 33.4546 | 33.44445 | 33.3233 |

จากตารางที่ 4.13 แสดงผลลัพธ์ค่า SNR ของโค้ดบุ๊กทั้ง 5 ระดับ ซึ่งแต่ละระดับโค้ดบุ๊กมีผลต่อค่า SNR ที่แตกต่างกัน จากผลการทดลองของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้ชายจำนวน 5 คน ด้วยการวัดประสิทธิภาพของการบีบอัดสัญญาณเสียงพูดด้วยเทคนิค SNR โดยที่โค้ดบุ๊ก 1024 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 33.51563 dB ซึ่งเป็นค่าเฉลี่ยของ SNR ที่มากที่สุด โค้ดบุ๊ก 512 ตำแหน่ง มีค่าเฉลี่ยของ SNR เท่ากับ 33.47775 dB ซึ่งเป็นค่า SNR น้อยลงจากโค้ดบุ๊ก 1024 ตำแหน่ง และโค้ดบุ๊ก 256 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่า SNR ที่น้อยลงตามลำดับของโค้ดบุ๊ก ในการทดลองนี้สัญญาณเสียงผู้ชายที่ 3 (M3) มีประสิทธิภาพมากที่สุดในทุกๆ ของโค้ดบุ๊ก โดยที่โค้ดบุ๊ก 1024 ตำแหน่งของสัญญาณเสียงผู้พูดมีค่าประสิทธิภาพมากที่สุดมีค่าเท่ากับ 36.6372 dB ซึ่งมากกว่าค่าเฉลี่ยประสิทธิภาพของ SNR ที่ตำแหน่งเดียวกัน ทำให้กล่าวได้ว่า ถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าประสิทธิภาพของ SNR ที่สูง ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนน้อยทำให้เกิดประสิทธิภาพต่ำ และถ้าโค้ดบุ๊กมีของจำนวนตำแหน่งที่มากจะส่งผลให้ประสิทธิภาพของ SNR สูงขึ้น ซึ่งเหตุผลเกิดจากตัวแทนในการคืนกลับตำแหน่งมีจำนวนมากทำให้เกิดความประสิทธิภาพที่สูง

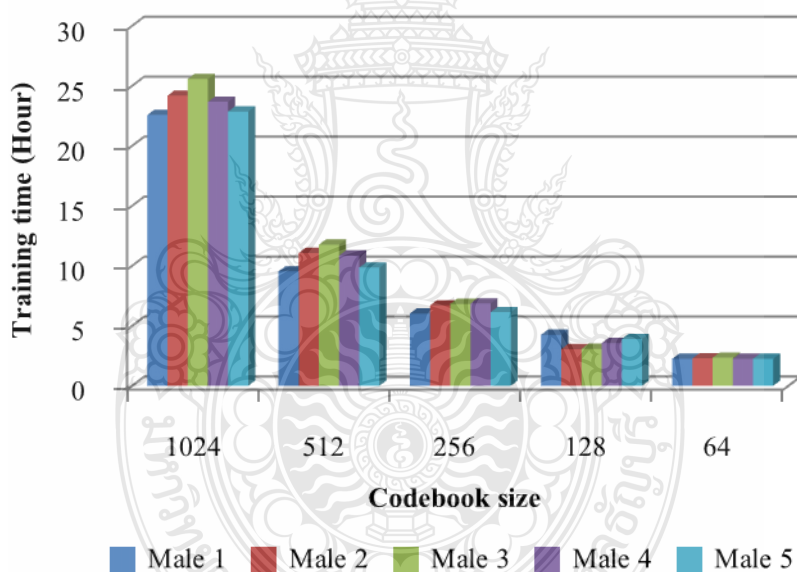
จากตารางที่ 4.12 และ 4.13 แสดงการวัดค่า SNR สรุปได้ว่า การบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิงมีประสิทธิภาพที่มากกว่าเสียงผู้ชาย เพราะพารามิเตอร์ LSP-10 ที่ผ่านการบีบอัดเสร็จสมบูรณ์แล้วของผู้หญิงมีความใกล้เคียงกับต้นฉบับมากกว่าผู้ชาย

4.4 ระยะเวลาในการฝึกสอน (Training Time)

สำหรับโค้ดบุ๊กที่มีจำนวนตำแหน่งที่มาก จะมีค่าความผิดพลาดของสัญญาณที่มีค่าต่ำ และมีค่าประสิทธิภาพของสัญญาณที่ดีก็ตาม แต่ในเรื่องระยะเวลาในการฝึกสอนสำหรับโค้ดบุ๊กที่มีจำนวนตำแหน่งที่มากก็ต้องเวลาที่ยาวนานเช่นกัน โดยในหัวข้อนี้แสดงระยะเวลาในการฝึกสอนของ LVO สำหรับการบีบอัดสัญญาณของโค้ดบุ๊กทั้ง 5 ระดับ

4.4.1 ระยะเวลาในการฝึกสอนของการบีบอัดสัญญาณเสียง LPC-10

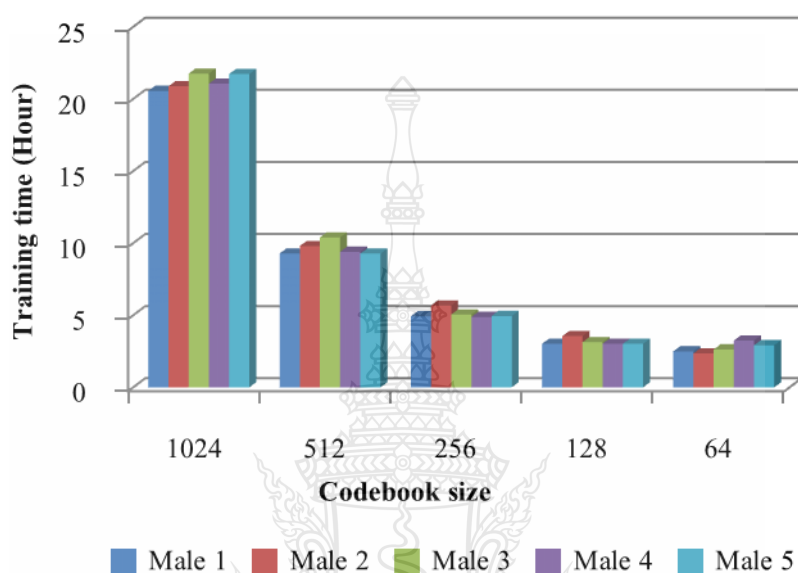
ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ทั้ง 5 ระดับ จะแบ่งเป็น 2 แบบ คือ ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน ดังภาพที่ 4.1 และระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน ดังภาพที่ 4.2



ภาพที่ 4.1 ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชาย จำนวน 5 คน

จากภาพที่ 4.2 แสดงระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้ชายจำนวน 5 คน จากผลลัพธ์จะทราบว่าในแต่ละระดับของโค้ดบุ๊กจะมีระยะเวลาในการฝึกสอนที่แตกต่างกัน โดยโค้ดบุ๊ก 1024 ตำแหน่ง มีระยะเวลาในการฝึกสอนที่ยาวนานกว่าโค้ดบุ๊ก อย่างเห็นได้ชัดเจน และโค้ดบุ๊ก 512 ตำแหน่ง ถึง 64 ตำแหน่ง จะใช้ระยะเวลาในการฝึกสอนที่สั้นลงตามลำดับ จากผลดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่มากจะส่งผลโดยตรงให้ระยะเวลาใน

การฝึกสอนมากขึ้น ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดสูงในการการฝึกสอน แต่ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่น้อยจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนน้อย ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดต่ำในการการฝึกสอน



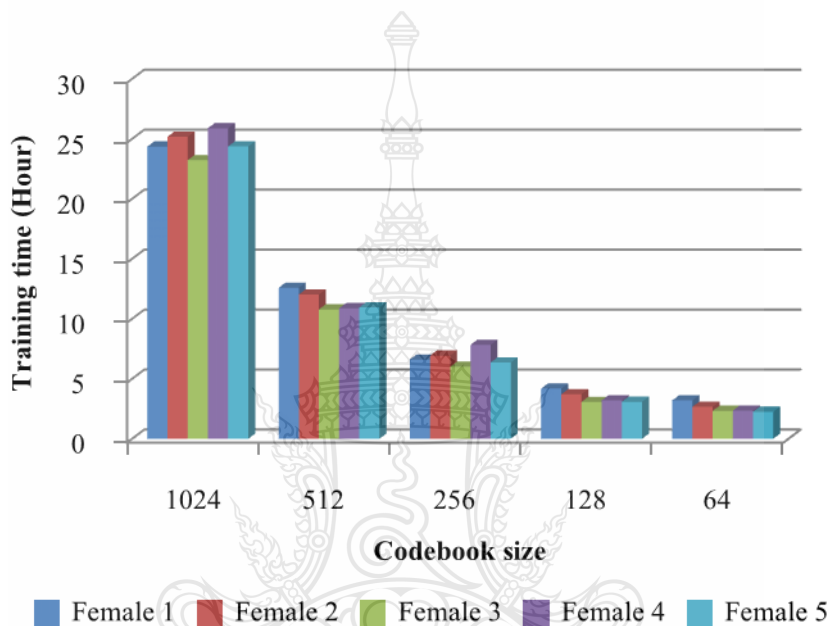
ภาพที่ 4.2 ระยะเวลาการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน

จากภาพที่ 4.2 แสดงระยะเวลาในการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง จำนวน 5 คน จากผลลัพธ์จะทราบว่าในแต่ละระดับของโค้ดบุ๊กจะมีระยะเวลาในการฝึกสอนที่แตกต่างกัน โดยโค้ดบุ๊ก 1024 ตำแหน่ง มีระยะเวลาในการฝึกสอนที่ยาวนานกว่า โค้ดบุ๊กอย่างเห็นได้ชัดเจน และโค้ดบุ๊ก 512 ตำแหน่ง ถึง 64 ตำแหน่ง จะใช้ระยะเวลาในการฝึกสอนที่สั้นลงตามลำดับ จากผลดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่มากจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนมากขึ้น ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดสูงในการฝึกสอน แต่ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่น้อยจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนน้อย ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดต่ำในการการฝึกสอน

จากภาพที่ 4.1 และ 4.2 แสดงระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LPC-10 สรุปได้ว่า การบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิงใช้ระยะเวลาในการฝึกสอนมากกว่าเสียงผู้ชาย เพราะมีความใกล้เคียงของกลุ่มตัวแทนพารามิเตอร์ของผู้หญิงมากกว่าผู้ชาย

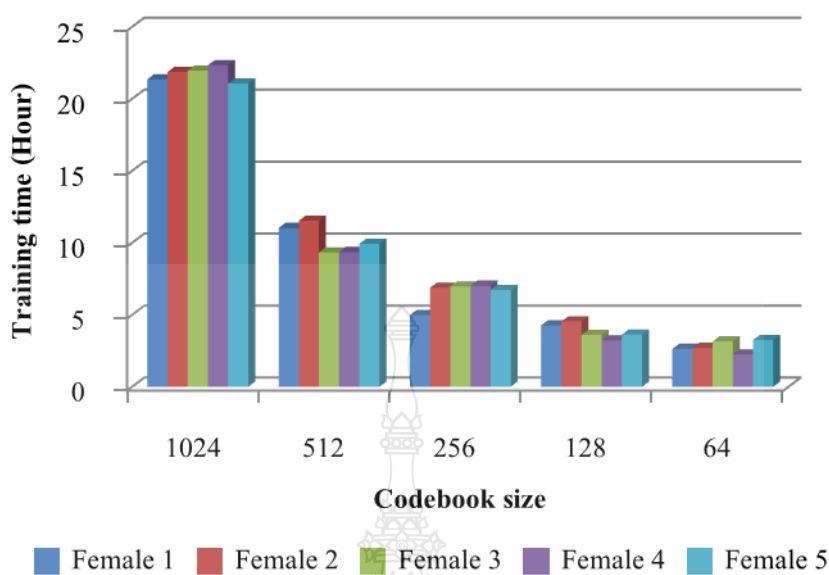
4.4.2 ระยะเวลาในการฝึกสอนของการบีบอัดสัญญาณเสียง LSP-10

ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ทั้ง 5 ระดับ จะแบ่งเป็น 2 แบบ คือ ระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน ดังภาพที่ 4.3 และระยะเวลาในการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง จำนวน 5 คน ดังภาพที่ 4.2



ภาพที่ 4.3 ระยะเวลาการฝึกสอนของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย จำนวน 5 คน

จากภาพที่ 4.3 แสดงระยะเวลาในการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้ชายจำนวน 5 คน จากผลลัพธ์จะทราบว่าในแต่ละระดับของโค้ดบุ๊กจะมีระยะเวลาในการฝึกสอนที่แตกต่างกัน โดยโค้ดบุ๊ก 1024 ตำแหน่ง มีระยะเวลาในการฝึกสอนที่ยาวนานกว่าโค้ดบุ๊กอย่างเห็นได้ชัดเจน และโค้ดบุ๊ก ที่ 512 ตำแหน่ง ถึง 64 ตำแหน่ง จะใช้ระยะเวลาในการฝึกสอนที่สั้นลงตามลำดับ จากผลดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่มากจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนมากขึ้น ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดสูงในการการฝึกสอน แต่ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่น้อยจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนน้อย ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดต่ำในการฝึกสอน



ภาพที่ 4.4 ระยะเวลาการฝึกสอนของการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้หญิง จำนวน 5 คน

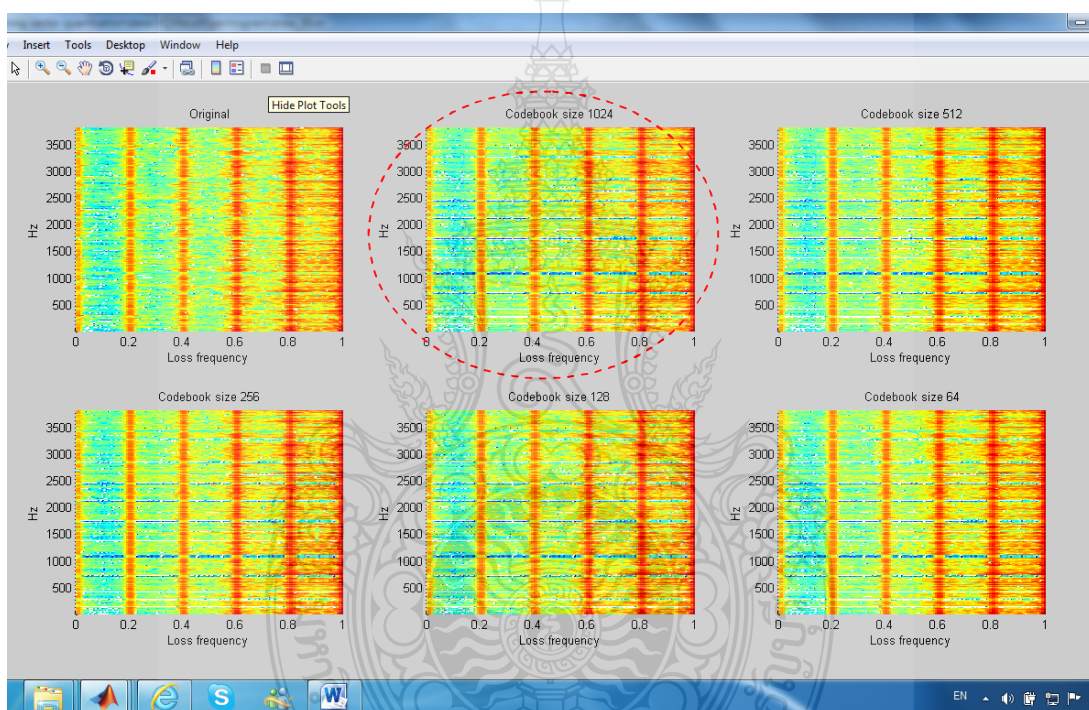
จากภาพที่ 4.3 แสดงระยะเวลาในการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้หญิง จำนวน 5 คน จากผลลัพธ์จะทราบว่าในแต่ละระดับของโค้ดบุ๊กจะมีระยะเวลาในการฝึกสอนที่แตกต่างกัน โดยโค้ดบุ๊ก 1024 ตำแหน่ง มีระยะเวลาในการฝึกสอนที่ยาวนานกว่าโค้ดบุ๊กอย่างเห็นได้ชัดเจน และโค้ดบุ๊ก 512 ตำแหน่ง ถึง 64 ตำแหน่ง จะใช้ระยะเวลาในการฝึกสอนที่สั้นลงตามลำดับ จากผลดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่มากจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนมากขึ้น ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดสูงในการฝึกสอน แต่ถ้าชุดสมาชิกมีจำนวนของตำแหน่งที่น้อยจะส่งผลโดยตรงให้ระยะเวลาในการฝึกสอนน้อย ซึ่งเหตุเกิดจากจำนวนตัวแทนที่มีจำนวนมากทำให้เกิดความละเอียดต่ำในการการฝึกสอน

จากภาพที่ 4.1 และ 4.2 แสดงระยะเวลาในการฝึกสอนของการบีบพารามิเตอร์ LSP-10 สรุปได้ว่า การบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิงใช้ระยะเวลาในการฝึกสอนมากกว่าเสียงผู้ชาย เพราะมีความใกล้เคียงของกลุ่มตัวแทนพารามิเตอร์ของผู้หญิงมากกว่าผู้ชาย

4.5 สเปกโตรแกรม (Spectrogram)

สเปกโตรแกรม คือ การอ่านสัญญาณเสียงพูดโดยที่ผู้อ่านไม่ต้องฟังเสียงพูดนั้นจริงๆ ซึ่งการอ่านสเปกโตรแกรมนั้นทำได้โดยวิเคราะห์ลักษณะต่างๆ ของสัญญาณเสียงทั้งทางเวลา และความถี่ หลังจากนั้นจึงจะคาดเดารูปร่างของทางเสียง (Vocal Tract) ตลอดจนแหล่งกำเนิดสัญญาณ ในตำแหน่งเวลาที่สอดคล้องกับลักษณะทางเวลา หรือทางความถี่นั้นๆ ซึ่งการอ่านสเปกโตรแกรมเป็นที่นิยมในการอ่านสัญญาณเสียงพูดที่ผ่านการบีบอัดสัญญาณเสียง และในงานวิจัยนี้ก็ได้นำสเปกโตรแกรมมาอ่านเพื่อทดสอบคุณภาพของการบีบอัดสัญญาณเสียง

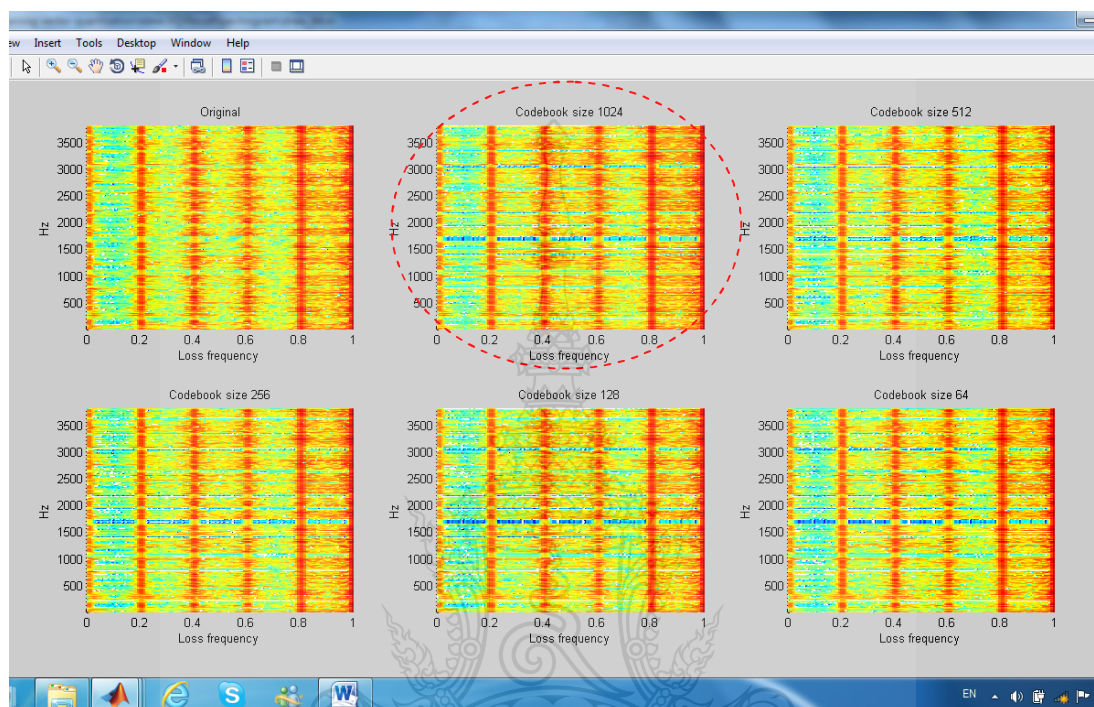
4.5.1 สเปกโตรแกรมของการบีบอัดพารามิเตอร์ LPC-10



ภาพที่ 4.5 สเปกโตรแกรมของการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้ชาย

จากภาพที่ 4.5 เป็นการสังเกตผลลัพธ์ของสเปกโตรแกรมในการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้พูด ผู้ชายจำนวน 1 ตัวอย่าง ของโค้ดบุ๊กทั้ง 5 ระดับ โดยโค้ดบุ๊ก ที่ 1024 ตำแหน่ง จะมีภาพสเปกโตรแกรมที่ใกล้เคียงกับสเปกโตรแกรมพารามิเตอร์ LPC-10 เดิมมากที่สุด และ โค้ดบุ๊ก ที่ 512 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมี ความใกล้เคียงลงมาตามลำดับของระดับโค้ดบุ๊ก จากผลของสเปกโตรแกรมดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่มาก จะส่งผลให้ผลการอ่านด้วยสเปกโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่ต่ำ และ ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่น้อยน้อยลง

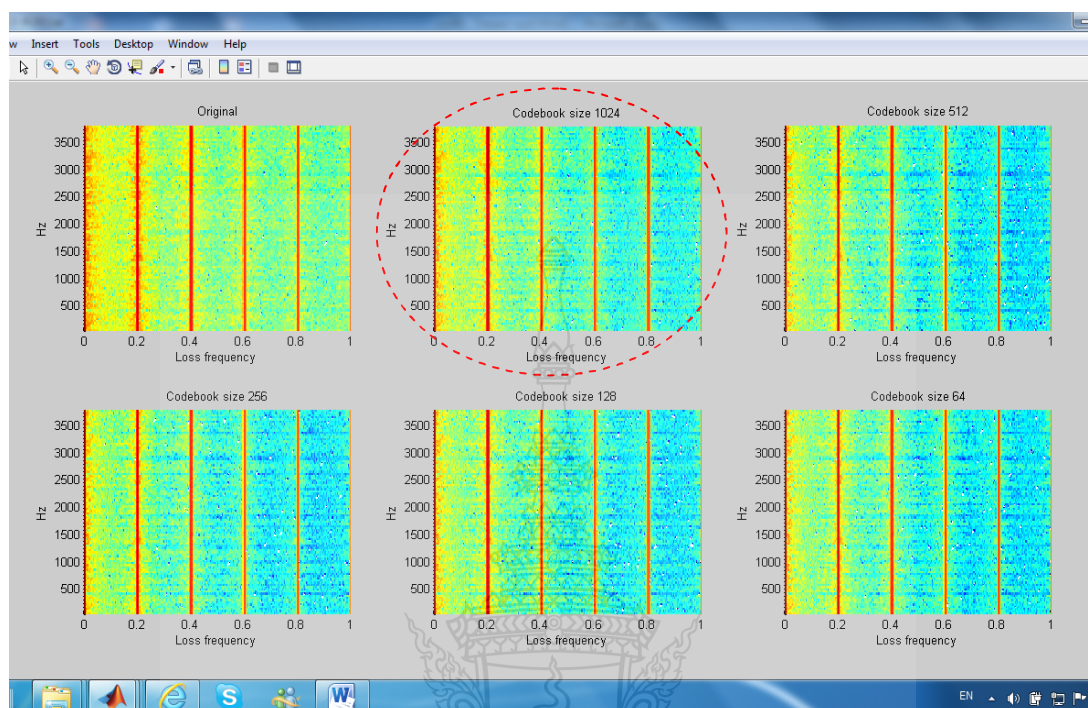
จะส่งผลให้ผลการอ่านด้วยสเปคโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่มากขึ้น ซึ่งสอดคล้องกับการวัดค่าความผิดพลาด คือ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ



ภาพที่ 4.6 สเปคโตรแกรมของการบีบอัดพารามิเตอร์ LPC-10 ของเสียงผู้หญิง

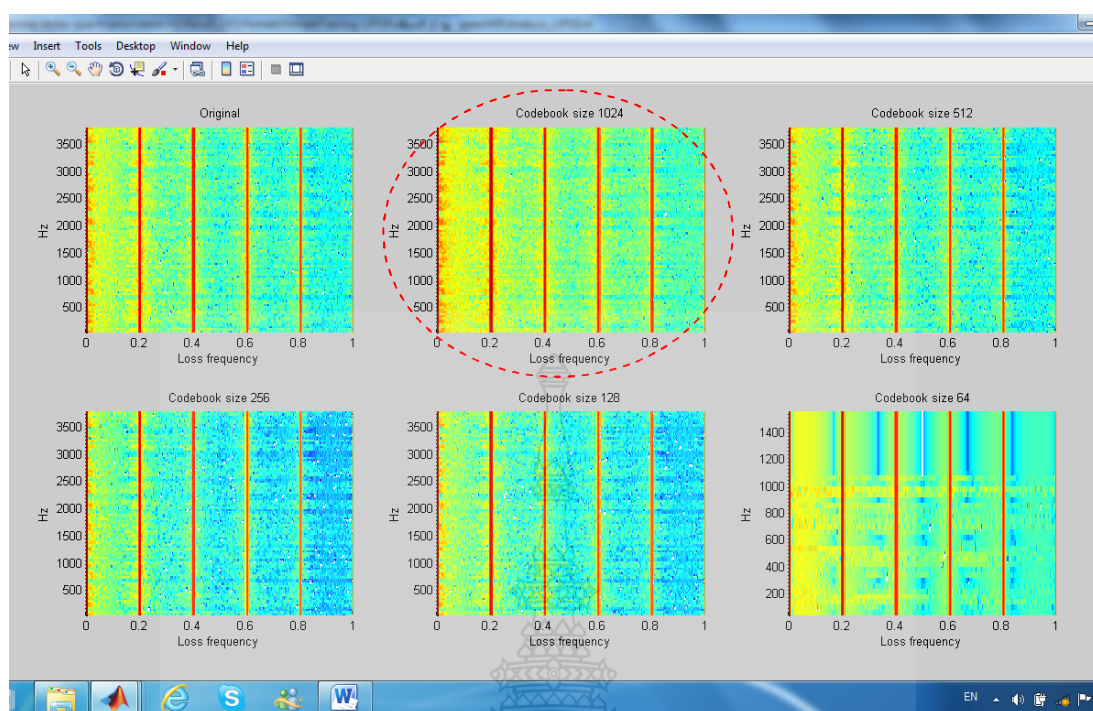
จากภาพที่ 4.6 เป็นการสังเกตผลลัพธ์ของสเปคโตรแกรมในการบีบอัดสัญญาณเสียง LPC-10 ของเสียงผู้พูด ผู้หญิง จำนวน 1 ตัวอย่าง ของโค้ดบุ๊กทั้ง 5 ระดับ โดยโค้ดบุ๊ก ที่ 1024 ตำแหน่ง จะมีภาพสเปคโตรแกรมที่ใกล้เคียงกับสเปคโตรแกรมพารามิเตอร์ LPC-10 เดิมมากที่สุด และโค้ดบุ๊ก ที่ 512 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าใกล้เคียงลงมาตามลำดับของระดับโค้ดบุ๊ก จากผลของสเปคโตรแกรมดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่มาก จะส่งผลให้ผลการอ่านด้วยสเปคโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่ต่ำ และ ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่น้อยน้อยลง จะส่งผลให้ผลการอ่านด้วยสเปคโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่มากขึ้น ซึ่งสอดคล้องกับการวัดค่าความผิดพลาด คือ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง และถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ

4.5.2 สเปกโตรแกรมของการบีบอัดสัญญาณเสียง LSP-10



ภาพที่ 4.7 สเปกโตรแกรมของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้ชาย

จากภาพที่ 4.7 เป็นการสังเกตผลลัพธ์ของสเปกโตรแกรมในการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้พูด ผู้ชายจำนวน 1 ตัวอย่าง ของโค้ดบุ๊กทั้ง 5 ระดับ โดยโค้ดบุ๊ก ที่ 1024 ตำแหน่ง จะมีภาพสเปกโตรแกรมที่ใกล้เคียงกับสเปกโตรแกรมพารามิเตอร์ LSP-10 เดิมมากที่สุด และ โค้ดบุ๊ก ที่ 512 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าใกล้เคียงลงมาตามลำดับของระดับโค้ดบุ๊ก จากผลของสเปกโตรแกรมดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่มาก จะส่งผลให้ผลการอ่านด้วยสเปกโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่ต่ำ และ ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่น้อยน้อยลง จะส่งผลให้ผลการอ่านด้วยสเปกโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่มากขึ้น ซึ่งสอดคล้องกับการวัดค่าความผิดพลาด คือ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง และ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ



ภาพที่ 4.8 สเปกโตรแกรมของการบีบอัดพารามิเตอร์ LSP-10 ของเสียงผู้หญิง

จากภาพที่ 4.8 เป็นการสังเกตผลลัพธ์ของสเปกโตรแกรมในการบีบอัดสัญญาณเสียง LSP-10 ของเสียงผู้พูด ผู้ชายจำนวน 1 ตัวอย่าง ของโค้ดบุ๊กทั้ง 5 ระดับ โดยโค้ดบุ๊ก ที่ 1024 ตำแหน่ง จะมีภาพสเปกโตรแกรมที่ใกล้เคียงกับสเปกโตรแกรมพารามิเตอร์ LSP-10 เดิมมากที่สุด และ โค้ดบุ๊ก ที่ 512 ตำแหน่ง ถึง 64 ตำแหน่ง ก็จะมีค่าใกล้เคียงลงมาตามลำดับของระดับโค้ดบุ๊ก จากผลของสเปกโตรแกรมดังกล่าวสรุปได้ว่า ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่มาก จะส่งผลให้ผลการอ่านด้วยสเปกโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่ต่ำ และ ถ้าชุดสมาชิกมีจำนวนตำแหน่งที่น้อยลง จะส่งผลให้ผลการอ่านด้วยสเปกโตรแกรมมีความผิดเพี้ยนไปจากต้นฉบับที่มากขึ้น ซึ่งสอดคล้องกับการวัดค่าความผิดพลาด คือ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่น้อยจะส่งผลให้ค่าความผิดพลาดที่สูง และ ถ้าโค้ดบุ๊กมีจำนวนตำแหน่งที่มากจะส่งผลให้ค่าความผิดพลาดที่ต่ำ

4.6 ผลการบีบอัดตามขนาดของโค้ดบุ๊ก

ในงานวิจัยนี้ใช้สัญญาณเสียงพูดตัวอย่างที่มีความยาว 1 นาที ที่มีรูปประโยคการพูดแบบเล่าเรื่องในสถานการณ์ที่แตกต่างกัน โดยที่นำสัญญาณเสียงพูดตัวอย่างดังกล่าวเข้าสู่กระบวนการหาค่าสัมประสิทธิ์ทำนายพันธะเชิงเส้นที่ 10 อันดับ (LPC-10) และ สัมประสิทธิ์คู่เส้นสเปกตรัม (LSP-10)

โดยที่สัญญาณเสียงพูดที่มีความยาว 1 นาที มีจำนวนเฟรมข้อมูลของพารามิเตอร์ LPC-10 และ LSP-10 เท่ากับ 2400 เฟรม จากนั้นการนำเฟรมข้อมูลของพารามิเตอร์ทั้งสองมาทำการบีบอัดเฟรมสัญญาณ โดยใช้เทคนิค LVO โดยมีระดับการบีบอัดตามระดับขนาดของโค้ดบุ๊ก ดังในตารางที่ 4.1 เพื่อที่จะลดจำนวนบิตที่ใช้ในการส่งสัญญาณเสียงพูดของพารามิเตอร์ต้นฉบับของ LPC-10 และ LSP-10 ซึ่งวิธีการลดจำนวนบิตของการบีบอัดสัญญาณสามารถหาได้จากสมการต่อไปนี้

$$Z_c = \left| \frac{Z_{Original} - Z_{Compression}}{Z_{Original}} \right| \times 100 \quad (3.5)$$

โดยที่ $Z_{Original}$ คือ จำนวนสัญญาณต้นฉบับ
 $Z_{Compression}$ คือ จำนวนสัญญาณที่ผ่านการบีบอัด
 Z_c คือ เปอร์เซ็นต์การบีบอัดสัญญาณที่ลดลง

ตารางที่ 4.14 ผลการบีบอัดตามระดับขนาดของโค้ดบุ๊ก

| ระดับที่ | ค่าจำนวนตัวอย่าง (เฟรม) | | % การบีบอัด |
|----------|--|----------|-------------|
| | ค่าต้นฉบับของพารามิเตอร์ LPC-10 และ LSP-10 | โค้ดบุ๊ก | |
| 1 | 2400 | 1024 | 81.4815 % |
| 2 | 2400 | 512 | 83.3333 % |
| 3 | 2400 | 256 | 85.1852 % |
| 4 | 2400 | 128 | 87.0370 % |
| 5 | 2400 | 64 | 88.8889 % |

จากตารางที่ 4.14 จะสังเกตผลการบีบอัด พบว่าโค้ดบุ๊กขนาด 64 มีเปอร์เซ็นต์การบีบอัดสัญญาณที่ลดลง มากที่สุดคือ 88.8889 % รองลงมาได้แก่โค้ดบุ๊กขนาด 128 โค้ดบุ๊กขนาด 256 โค้ดบุ๊กขนาด 512 และโค้ดบุ๊กขนาด 1024 โดยมีเปอร์เซ็นต์การบีบอัดสัญญาณที่ลดลง คือ 87.0370 % 85.1852 % 83.3333 % และ 81.4815 % ตามลำดับ

สรุปผลการทดลอง ได้ว่า เฟอร์เซ็นต์การบีบอัดสัญญาณที่ลดลง จะแปรผันตามขนาดของโค้ดบู้ค นั่นคือโค้ดบู้คที่มีขนาดใหญ่จะเก็บรักษาคุณลักษณะเสียงพูดต้นฉบับไว้ได้มากส่งผลให้การบีบอัดสัญญาณมีเฟอร์เซ็นต์ที่น้อยกว่าโค้ดบู้คที่มีขนาดขนาดเล็ก

4.7 ผลการประเมินคุณภาพเสียงพูด (MOS)

การประเมินคุณภาพเสียงพูดจะใช้การรับรู้และความรู้สึกของผู้ฟังเป็นเกณฑ์ในการตัดสินใจ โดยการฟังเสียงพูดที่ผ่านการบีบอัดด้วย LVQ แล้วเปรียบเทียบกับความชัดเจนกับเสียงต้นฉบับแล้วทำการประเมินคุณภาพด้วยวิธีการของ MOS

ตารางที่ 4.15 การประเมินคุณภาพเสียงจากการรับฟัง

| เสียงพูด (speech) | ค่าระดับการประเมินคุณภาพเสียงจากการฟัง (MOS) | | | | | |
|-------------------------|--|-------|-------|-------|-----|----|
| | ต้นฉบับ | 1024 | 512 | 256 | 128 | 64 |
| เสียงผู้ชายที่ 1 (M1) | 4.76 | 3.88 | 3.15 | 2.56 | 1 | 1 |
| เสียงผู้ชายที่ 2 (M2) | 4.81 | 4.15 | 3.34 | 2.78 | 1 | 1 |
| เสียงผู้ชายที่ 3 (M3) | 4.75 | 3.87 | 3.21 | 2.72 | 1 | 1 |
| เสียงผู้ชายที่ 4 (M4) | 4.80 | 4.04 | 3.22 | 2.72 | 1 | 1 |
| เสียงผู้ชายที่ 5 (M5) | 4.78 | 3.91 | 3.18 | 2.60 | 1 | 1 |
| เสียงผู้หญิงที่ 1 (FM1) | 4.82 | 4.11 | 3.22 | 2.62 | 1 | 1 |
| เสียงผู้หญิงที่ 2 (FM2) | 4.80 | 4.02 | 3.18 | 2.70 | 1 | 1 |
| เสียงผู้หญิงที่ 3 (FM3) | 4.78 | 3.98 | 3.16 | 2.70 | 1 | 1 |
| เสียงผู้หญิงที่ 4 (FM4) | 4.88 | 4.20 | 3.42 | 2.82 | 1 | 1 |
| เสียงผู้หญิงที่ 5 (FM5) | 4.83 | 3.86 | 3.24 | 2.80 | 1 | 1 |
| ค่าเฉลี่ย (Average) | 4.801 | 4.002 | 3.232 | 2.702 | 1 | 1 |

จากผลการประเมินคุณภาพเสียงพูดด้วย MOS พบว่าโค้ดบู้คขนาด 1024 มีค่าเฉลี่ย MOS มากที่สุดคือ 4.002 รองลงมาได้แก่โค้ดบู้คขนาด 512 โค้ดบู้คขนาด 256 โค้ดบู้คขนาด 128 และโค้ดบู้คขนาด 64 โดยมีค่าเฉลี่ย MOS คือ 3.232 2.702 และ 1 ตามลำดับ

สรุปผลการทดลองการประเมินเสียงพูดด้วยวิธี MOS ได้ว่า คุณภาพเสียงของการบีบอัดสัญญาณเสียงจะแปรผันตามขนาดของโค้ดบุ๊ก นั่นคือ โค้ดบุ๊กที่มีขนาดใหญ่จะเก็บรักษาคุณลักษณะเสียงพูดต้นฉบับไว้ได้มากกว่าโค้ดบุ๊กที่มีขนาดเล็ก คุณภาพเสียงจึงใกล้เคียงกับเสียงต้นฉบับ



บทที่ 5

สรุปผลการวิจัย

5.1 สรุปผลการทดลอง

การพัฒนา LVQ สำหรับการบีบอัดสัญญาณเสียงในมาตรฐานของ LPC-10 และ LSP-10 ได้ดำเนินการนำสัญญาณเสียงพูดวิเคราะห์หาพารามิเตอร์ LPC-10 และ LSP-10 จากนั้นนำพารามิเตอร์ทั้งสอง ทำการบีบอัดพารามิเตอร์เพื่อลดจำนวนบิตการส่งสัญญาณค่าพารามิเตอร์ โดยใช้ LVQ ในออกแบบโค้ดบุ๊กเพื่อจัดกลุ่มของลักษณะสัญญาณที่มีความใกล้เคียงกันทางคุณสมบัติของสัญญาณนั้นๆ โดยสร้างระดับโค้ดบุ๊กตั้งแต่ 1024 ตำแหน่ง ถึง 64 ตำแหน่ง

ผลการทดลองการประยุกต์ใช้ LVQ สำหรับการบีบอัดพารามิเตอร์ LPC-10 และ LSP-10 โดยการสร้างโค้ดบุ๊กตั้งแต่ 1024 ตำแหน่ง ถึง 64 ตำแหน่งกับการวัดประสิทธิภาพด้วย NRMSE พบว่าการบีบอัดพารามิเตอร์ LPC-10 มีค่า NRMSE ต่ำที่สุด เท่ากับ 0.1578 การบีบอัดพารามิเตอร์ LSP-10 มีค่า NRMSE ต่ำที่สุด เท่ากับ 0.0111 การวัดประสิทธิภาพด้วย PSNR พบว่า การบีบอัดพารามิเตอร์ LPC-10 มีค่า PSNR สูงที่สุด เท่ากับ 28.1904 dB การบีบอัดพารามิเตอร์ LSP-10 มีค่า PSNR สูงที่สุด เท่ากับ 41.0208 dB และในการวัดประสิทธิภาพด้วย SNR การบีบอัดพารามิเตอร์ LPC-10 มี SNR สูงสุดที่ 15.0333 dB การบีบอัดพารามิเตอร์ LSP-10 มี SNR สูงสุดที่ 36.6372 dB จากผลการทดลองด้วยการวัดประสิทธิภาพแบบเชิงวัตถุ (Objective Test) ทั้ง 3 วิธี สรุปได้ว่า การใช้ LVQ ในการบีบอัดพารามิเตอร์ LPC-10 และ LSP-10 ยังคงรักษาความใกล้เคียงของพารามิเตอร์ LPC-10 และ LSP-10 ต้นฉบับเดิมไว้ได้ และจากผลการประเมินคุณภาพแบบฟังเสียงพูด (Subjective Test) ด้วยวิธี MOS สรุปได้ว่า การใช้ LVQ ในการบีบอัดพารามิเตอร์ LPC-10 และ LSP-10 ก็ยังคงสามารถรับฟังเสียงพูดได้อย่างเข้าใจ โดยมีค่า MOS เฉลี่ยสูงสุด เท่ากับ 4.002

สรุปผลการงานวิจัย ได้ว่า การใช้ LVQ สำหรับการบีบอัดพารามิเตอร์ LPC-10 และ LSP-10 สามารถบีบอัดพารามิเตอร์ LPC-10 และ LSP-10 ลดลงได้ไม่น้อยกว่า 81.4815 % จากพารามิเตอร์ LPC-10 และ LSP-10 ต้นฉบับเดิม ดังตารางที่ 4.14 และจากการวัดประสิทธิภาพด้วยหลักการวัดประสิทธิภาพแบบเชิงวัตถุและหลักการประเมินคุณภาพแบบฟังเสียงพูด แสดงให้เห็นว่าในการบีบอัดพารามิเตอร์ยังคงรักษาความใกล้เคียงของพารามิเตอร์ต้นฉบับไว้ได้อย่างมีคุณภาพ

5.2 ข้อเสนอแนะในการวิจัย

ในการทำงานวิจัยในครั้งนี้ พบว่าผลการวิจัยที่ได้นั้นเป็นที่ยอมรับได้ในระดับหนึ่ง แต่ควรมีการศึกษาในมุมมองด้านอื่นๆ บ้าง ซึ่งมีหัวข้อในการพัฒนาดังต่อไปนี้

5.2.1 ในการบันทึกเสียงพูดให้สัญญาณเสียงอยู่ในช่วงที่ต้องการนั้นสามารถควบคุมได้ยาก และการบันทึกเสียงเป็นเวลานานจะทำให้ระดับเสียงไม่คงที่ อันเกิดจากการต้องเปล่งเสียงดังทำให้ผู้พูดเกิดอาการเหนื่อยล้า จึงควรระมัดระวังในการเก็บข้อมูล และควรควบคุมสภาพแวดล้อมในระหว่างทำการบันทึกเสียงของผู้พูดเพื่อลดสัญญาณรบกวนให้มากที่สุด

5.2.2 เทคนิคในการวิเคราะห์สัญญาณเสียงพูดนั้นมีอยู่หลากหลายเทคนิค ในงานวิจัยนี้ได้เพียง 2 เทคนิค คือ เทคนิคการทำนายพัลซาร์เชิงเส้น (LPC) และ เทคนิคกลุ่มเส้นสเปกตรัม ควรทดลองด้วยเทคนิคอื่นๆหลายๆ เทคนิค เพื่อเปรียบเทียบกับผลการวิจัย

5.2.3 ควรมีการวิจัยหากกลุ่มของเสียงที่เหมาะสมสำหรับการฝึกสอนโครงข่าย LVO ซึ่งไม่ควรมีจำนวนมากนักเพราะจะทำให้การฝึกสอนช้า



รายการอ้างอิง

- [1] J.R. Deller, J.G. Proakis and J.H. Hansen, "Discrete-Time Processing of Speech Signals," , 2000.
- [2] W.B.Kleijin and K.K.Paliwal, "Speech Coding and Synthesis," , 1995.
- [3] สุทธิ ทับทองดี. การพัฒนาเทคนิคสำหรับการบีบอัดสัญญาณเสียงพูดภาษาไทย ในมาตรฐาน LPC-10. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2552.
- [4] จักรกฤษ อ่อนชื่นจิตร. การวิเคราะห์แนวทางเดินเสียงพูดในรูปแบบของสัมประสิทธิ์คู่เส้นสเปกตรัมร่วมกับ Double Clustering. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2552.
- [5] MATLAB Manual and MATLAB help
Available : <http://www.mathworks.com/help/toolbox/nnet/>, 2012
- [6] Jakkree Srinonchat, "New Technique to Reduce Bit Rate of LPC-10 Speech Coder," TENCON, 2006. 2006 IEEE Region 10, Nov. 2006, p. 1-4
- [7] Dong,E.,Guizhong,L.and Hongliang, L., "2.4 kb/s low bit rate speech coding based on local cosine transform," Signal Processing, 2002 6th International Conference (Electronic),Vol.1,2002,pp.429-429,IEEE organization/IEEEXploer (7 July 2008)
- [8] Phattarasuda Witchayaphong, Napat Lakhawattana and Khedsadaporn Chaiwong, "Comparing the Performance of Wearing Helmet Behavior Model While Driving Motorcycle: By Binary Logistic Regression Analysis Method and Learning Vector Quantization," 3rd Atrans Symposium Student Chapter Session, August 27 ,2010
- [9] วีระยุทธ คุณรัตนศิริ. การเพิ่มประสิทธิภาพการบีบอัดสัญญาณเสียงด้วยเทคนิคเวฟเล็ตร่วมกับไวเนอร์ฟิลเตอร์. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2554.
- [10] Oranong Wiriyaruruknakorn and Jakkree Srinonchat, "A Finite State Vector Quantizer for New Bit Rate Speech Compression," IEEE International on Signal Processing, Communications and Networking 2008, ICSCN '08, Jan. 2008, p.255-259.

- [11] Dong,E.,Guizhong,L.and Hongliang, L., "2.4 kb/s low bit rate speech coding based on local cosine transform," Signal Processing, 2002 6th International Conference (Electronic),Vol.1,2002,pp.429-429,IEEE organization/IEEEXploer (7 July 2012)
- [12] Jakkree Srinonchat et al., "Address Vector Quantisation applied to Speech Coding," IEEE International Symposium on Signal Processing and Information Technology, 2003, pp.745-748.
- [13] Song Lin, Dimitrios Gunopulos, Vana Kalogeraki, Stefano Lonardi, "A Data Compression Technique for Sensor Networks with Dynamic Bandwidths Allocation," IEEE- Proceedings of the 12th International Symposium on Temporal Representation and Reasoning (TIME'05)
- [14] Anand D. Subramaniam, WilliamR. Gardner, and Bhaskar D. Rao, "Low Complexity Recursive Coding of Spectrum Parameters," IEEE-International Conference on Acoustics, Speech, and Signal Processing (ICASSP),2002,pp.637-640
- [15] Jakkree Srinonchat,J., Danaher,S.and Murry,A., " An Efficient of Codebook Design for Speaker Dependent Coding System," International Symposium Communication System, Networks and Digital Signal Processing (Electronic), Vol.1,2004,pp.484-486
- [16] Gao Li'ai Zhang Shuguang Zhou Yongjie Li Lihua , "A New Codebook Design Method Based on Genetic Programming," International Conference on Electronic Measurement and Instruments (ICEMI'2007),2007,pp.250-253
- [17] John S. Baras, Fellow, IEEE, and Subhrakanti Dey, Member, IEEE , "Combined compression and classification with learning vector quantization," IEEE Transactions on Information Theory, Vol. 45, No. 6, september 1999, pp. 1911-1920
- [18] Kreangsak Pattanaburi and jakkree Srinonchat, "Enhancement pattern analysis technique for voiced/unvoiced classification," IEEE- 2012 International Symposium on Computer, Consumer and Control (IS3C)I, 2012, pp 389-392
- [19] Kreangsak Pattanaburi and Jakkree Srinonchat, "Devlopment Voiced-Unvoiced Classification Technique Based On LSP Coefficients," Electical engineering network 2012 of rajamangala university of technology (EENET 2012), April, 3-5 2012, p. 541-54

ภาคผนวก



ภาคผนวก ก
ผลงานตีพิมพ์เผยแพร่



Electrical Engineering Network 2012

of Rajamangala University of Technology (EENET 2012)



ดอกแฉะสด ดอก Hibiscus ราชภัฏวชิราวุฒวิทยาลัย

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.นิธิโชติ รักไทยเจริญชีพ | กรรมการ |
| ผู้ช่วยศาสตราจารย์ ดร.เจนศักดิ์ เอกบุรณะวัฒน์ | กรรมการ |
| ดร.อุเทน คำน่าน | กรรมการ |
| ผู้ช่วยศาสตราจารย์วิสุทธิ์ พงศ์พฤกษธาต | กรรมการ |
| รองศาสตราจารย์ ดร.ธวัช เกิดชื่น | กรรมการ |
| ผู้ช่วยศาสตราจารย์ ดร.ศักดิ์ระวี ระวีกุล | 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 ขันเงิน

รศ. ดร. วิจิตร กิณเรศ

รศ. ดร. มนต์รี ศิริปรัชญานันท์

รศ. ดร. เวทิน ปิยะรัตน์

รศ. ดร. เดชา พวงดาวเรือง

รศ. ดร. วิบูลย์ ชื่นแขก

รศ. ดร. เสถียร ชาญศรีรัตน์

ผศ. ดร. แนบบุญ หุนเจริญ

ผศ. ดร. อภินันท์ อูโรโสภณ

ผศ. ดร. วรวัฒน์ เสี่ยมวิบูล

ผศ. ดร. อาทิตย์ โสทรโยม

ผศ. ดร. เสด็จ เผ่าละออ

ผศ. ดร. กองพล อารีรักษ์

ดร. ยุทธนา ขำสุวรรณ

ดร. นิวัฒน์ อังควิสิษฐพันธ์

รศ. ดร. โกศล โอฬารไพโรจน์

ดร. อุเทน คำน่าน

นายณรงค์ นันทกุล

ผศ. กฤษดา ยิ่งขันธ์

นายเอกทัศน์ พงษ์วารณ

ดร. จัตุฤทธิ์ ทองปรอน

ผศ. ชาลุชัย เดชธรรมรงค์

ดร. จักรกฤษณ์ เคลือบวัง

นายสมนึก เครือสอน

นายทัศนะ ถมทอง

นายณรงค์ฤทธิ์ พิมพ์คำวงศ์

ผศ. อภิศักดิ์ ขันแก้วหล้า

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 แสทนอน | มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก |
| นายเอกลักษณ์ สุมนพันธุ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก |
| นายปรีชา มหาไม้ | มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก |
| นายก่อเกียรติ อีอัครทรัพย์ | มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เขตพื้นที่ตาก |
| รศ.ดร.ชวัล เกิดชื่น | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ดร.พินิจ ศรีธร | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.ประเสริฐ เตือนหมื่น ไวย | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.พันธ์พงศ์ อภิชาติกุล | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.สุทธินันท์ ต้นโพธิ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.ศิริชัย ลาภาสระน้อย | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.กฤตวิทย์ บัวใหญ่ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ศส.วุฒิชัย สง่างาม | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นายกิตติวุฒิ จินนะบุตร | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นายรุ่งเพชร ก่องนอก | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นายเอกจิต คุ้มวงศ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นางอุษา คงเมือง | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นายชิติศรศักดิ์ วิชิโต | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| ดร.วรรณรีย์ วงศ์ไตรรัตน์ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จ.นครราชสีมา |
| นายบุญช่วย เจริญผล | มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ |
| นายวุฒิวัดน์ กงรัตนประเสริฐ | มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ |
| นายชูศักดิ์ชู กมลขันดิสร | มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ |
| นายวินัย เมธาวิท | มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ |
| นายชาญฤทธิ์ ธาราสันติสุข | มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ.กรุงเทพฯ |
| นายประหัด กองสุข | มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา |
| นายภัทรพงศ์ อัญชันภาติ | มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา |
| นายสมพล โคศรี | มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา |
| นายทัศนีย์ สุวรรณศักดิ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก จ.ฉะเชิงเทรา |
| นายจตุรงค์ จตุเรชชัยสกุล | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายพูนศรี วรรณการ | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| ดร.ณัฐพงษ์ พันธนะ | 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 |
| นายสุภาภูมิ เนตรโพธิ์แก้ว | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายทอง ลานธารทอง | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายสมเกียรติ ทองแก้ว | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| ศส.พิชญ์ ดาราพงษ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายมนัส บุญเพียรทอง | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายพนา ดุสิตากร | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| ศส.กิจจา ลักยณ์อำนวยการ | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| ศส.โกศล นิธิโสภา | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| ศส.จรินทร์ จุลวานิช | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายนิลमित นิลาศ | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายเกรียงไกร เหลืองอำพล | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายวัฒนพันธ์ วัชวุฒิ | มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร จ.กรุงเทพฯ |
| นายณัฏฐ์วรินทร์ ทองรักษ์ | 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 กันทะพะเยา | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี |
| ศส.เฉลิมพล เรื่องพัฒนาวิวัฒน์ | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี |
| ศส.กระจ่าง พิทักษ์วงษ์วิทยา | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี |
| ดร.ศรีสุดา ไชยทองสูง | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี |
| รศ.นภัทร วัจนเทพินทร์ | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.นนทบุรี |
| รศ.สมพันธ์ อำพาวิน | มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา |
| ดร.ชูวงศ์ วัฒนศักดิ์ภูบาล | มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา |
| ศส.วิสุทธิ์ พงศ์พิฤกษ์ชาติ | มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย จ.สงขลา |
| ดร.สุริยา แก้วอามา | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

| | |
|-----------------------------|--|
| ดร.นิธิโรจน์ พรสุวรรณเจริญ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |
| ดร.เมธา ทัศกร | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |
| ดร.นำพน พิพัฒน์ไพบูลย์ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |
| ศส.วิชชัย ครอบกิจศิริ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |
| นายเอกวิทย์ หายักวงษ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตสกลนคร |
| นายวีระ ฐานยาภิรักษ์ | 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 |
| นายเฉลียว เกตุแก้ว | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี |
| ศส.ดร.ปรีชา สาकरังค์ | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี |
| ดร.ภักพัฒน์ จันทร์ตรี | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี |
| นายไพบูลย์ เกียรติสุขกณธร | มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จ.สุพรรณบุรี |
| รศ.พันธ์ พิริยะวรรณ | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ศส.ประวิช เปรียบเหมือน | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ศส.ดร.ศักดิ์ระวี ระวังกุล | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ศส.ประยงค์ เสาร์แก้ว | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ศส.จักรวัฒน์ บุตรบุญชู | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ดร.สุระ ตันดี | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ดร.ณรงค์ สีหาจ้อง | มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขตขอนแก่น |
| ศส.ดร.สมชัย หิรัญวโรดม | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.ดร.วันชัย ทรัพย์สิงห์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.วิชชัย ผลุงศิลป์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.ศิริชัย แดงอม | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.ณัฐวุฒิ โสมะเกษศรีรินทร์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

| | | |
|----------------|---------------|------------------------------------|
| ดร.ฉัตรชัย | ศุภทิพย์สกุล | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.บุญยัง | ปลั่งกลาง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.สุรินทร์ | แหงมงาม | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.กฤษณ์ชนม์ | ภูมิภคิตพิชญ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.ณัฐภัทร | พันธ์คง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายพีนิจ | จิตจริง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายสมชาย | เขียนสูงเนิน | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายพร้อมศักดิ์ | อภิตกุล | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายนิติพงศ์ | ปานกลาง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายธีระพล | เหมือนขาว | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายณัฐพล | หาอุปละ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.จินตนา | นาคะสุวรรณ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.ปราชญ์ | ดาบบัง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.จักรี | ศรีนนท์ฉัตร | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.อำนาจ | เรืองอำนาจ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.ไพฑูรย์ | รักเหลือ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายวิโรจน์ | พิราจเนนชัย | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายพงษ์ศักดิ์ | อำภา | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายบุญยิ่ง | นบนอบ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายอภิรดา | นามแสง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.สุทินัน | พรอนุรักษ์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.วินัย | วิชัยพาณิชย์ | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ดร.กิตติวัฒน์ | นัมเกิดผล | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| ศส.อิฐอาร์จ | ปิติมล | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |
| นายฉัตรพิงศ์ | อุทอง | มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี |

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

| | | |
|------|--|-----|
| TE05 | การทดลองออกอากาศแพร่ภาพระบบโทรทัศน์ดิจิทัลแบบเคลื่อนที่ตามมาตรฐาน DVB-H ในพื้นที่กรุงเทพมหานคร | 489 |
| TE06 | อัลกอริทึมสำหรับการฝังและการถอดลายน้ำดิจิทัลสำหรับภาพสีบนพื้นฐานการแปลงเวฟเล็ตแพกเกตส์กับแบนด์ย่อยที่ดีที่สุด | 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 ใหม่ดกระแสโดยใช้วงจรรายพานกระแส | 513 |
| TE12 | ระบบสังเคราะห์ความถี่ด้วยเฟสล็อกคูลูป | 517 |
| TE13 | การขยายแบนด์วิดท์ของสายอากาศของเบิครูปสี่เหลี่ยมผืนผ้าที่มีสตัดรูปตัวโอแบบแถวลำดับ สำหรับการสื่อสารไร้สาย | 521 |
| TE14 | สายอากาศโมโนโพลรูปตัววายสำหรับการใช้งานย่านความถี่แถบคู่ | 525 |
| TE15 | การออกแบบติดตั้งแอดเซสท้อยที่มาตรฐาน IEEE 802.11b/g ย่านความถี่ 2.4 GHz ระบบเครือข่ายแลนไร้สายพื้นที่ภายนอกอาคาร | 529 |
| TF16 | การออกแบบสายอากาศหลายความถี่ที่โค่นด้วยโครงสร้างระนาบร่วมรูปร่างรีซัคน | 533 |
| TE17 | เทคนิคการเปรียบเทียบเวฟเล็ตสำหรับอัตราการบีบอัดใหม่ของสัญญาณเสียง | 537 |
| TE18 | การพัฒนาเทคนิคในการแยกแยะเสียงโฆษะและเสียงโฆษะบนพื้นฐาน LSP Coefficients | 541 |
| TE19 | ยูนิเวอร์แซลฟิลเตอร์สามอินพุตหนึ่งเอาต์พุตรูปแบบกระแสปรับค่าได้ทางอิเล็กทรอนิกส์ด้วยโครงสร้างของ CCDDCCs | 545 |

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

การพัฒนาเทคนิคในการแยกแยะเสียงโงะและเสียงอโงะบนพื้นฐาน LSP Coefficients Development Voiced-Unvoiced Classification Technique Based On LSP Coefficients

เกรียงศักดิ์ พัฒนบุรี และ จักร ศรีนนท์ฉัตร

ภาควิชาวิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
39 ม.1 ด.คลองหก อ.ธัญบุรี จ.ปทุมธานี 12110 โทรศัพท์: 02-549-3588 E-mail: kreangsak_p@hotmail.com, jakkree@hotmail.ac.th

บทคัดย่อ

การแยกแยะเสียงโงะและอโงะโดยอาศัยคุณลักษณะในแต่ละเฟรมได้มีการวิจัยมากกว่า 5 ปี [1] ในงานวิจัยนี้ได้เสนอ 4 เทคนิคในการวิเคราะห์การแยกแยะเสียงโงะหรือเสียงอโงะบนพื้นฐานของสัมประสิทธิ์คู่เส้นสเปกตรัม (LSP) โดยเทคนิคทั้งหมดจะถูกเปรียบเทียบประสิทธิภาพกับเทคนิคนัลทอน (Prosodic Technique) งานวิจัยนี้ได้นำสัญญาณเสียงพูดความยาว 1 นาที จากผู้ชาย 20 คน และผู้หญิง 20 คน มาเข้ากระบวนการเทคนิคการตัดสินใจทั้ง 4 เทคนิคเพื่อแยกแยะเสียงโงะหรือเสียงอโงะ ผลการทดลองพบว่าเทคนิคที่ 1 เป็นเทคนิคที่มีประสิทธิภาพที่ดีในการหาคุณภาพการแยกแยะเสียงโงะและเสียงอโงะที่ 83.05% ที่จำนวนเสียงโงะเท่ากับ 1,322 เฟรม โดยเทคนิคนี้สามารถประยุกต์ใช้กับเทคนิคการเวกเตอร์ควอนไทซ์เชิงชั้นสำหรับการบีบอัดสัญญาณเสียงและการเรียนรู้จำของสัญญาณเสียง ซึ่งโดยปกติ LSP Coefficients จะถูกใช้ในส่วนของคุณลักษณะเสียง (Speech Feature)

คำสำคัญ: เสียงโงะ เสียงอโงะ สัมประสิทธิ์คู่เส้นสเปกตรัม

Abstract

Many researches have been studied to voiced and unvoiced classification technique using frame characteristic. This research presents four enhancement pattern analysis techniques to classify voiced and unvoiced based on the Line Spectrum Pairs (LSP) coefficients. Those techniques are also compared the performance with the prosodic technique. One minutes of speech signal are collected from each 20 male and female to be input speech. The results show that the 1st technique provides the best performance of the quality of V/UV classification at 83.05% with 1,322 voiced frame. This technique can be apply to vector quantization technique for speech compression and speech recognition which usually uses the LSP coefficients as the speech feature.

Keywords: voiced, unvoiced, line spectrum pairs

1. บทนำ

เทคนิคการสกัดคุณลักษณะของสัญญาณที่เป็นเสียงโงะและเสียงอโงะ [2] นั้นมีประโยชน์และจำเป็นอย่างยิ่งสำหรับการประมวลผลสัญญาณเสียง ระบบการรู้จำเสียงพูด การบีบอัดสัญญาณเสียง และสังเคราะห์เสียง และเป็นส่วนหนึ่งของการประยุกต์ใช้การประมวลผลสัญญาณทางด้านจิตอลเข้ากับสัญญาณเสียง ในงานวิจัยนี้จะมุ่งเน้นทางด้านคุณลักษณะของเสียง เพื่อบ่งบอกที่เป็นเสียงโงะและอโงะที่มีค่าพารามิเตอร์ LSP-10 [3] ในการแยกแยะคุณลักษณะของสัญญาณเสียงนั้นมีการดำเนินการวิจัยในหลายๆวิธีเช่น [4] ได้นำเสนอเทคนิคที่ชื่อว่า Instantaneous Frequency Amplitude Spectrum (IFAS) ในการแยกแยะสัญญาณเสียงโงะและเสียงอโงะ ซึ่งจากผลการทดลองพบว่าเทคนิคนี้สามารถแยกแยะสัญญาณเสียงของผู้ชายได้ดีกว่าเสียงของผู้หญิงประมาณ 5% ในงานวิจัย [5] ได้ประยุกต์ใช้เทคนิคเวฟเล็ตในการแยกแยะสัญญาณเสียงโงะและเสียงอโงะ ผลการวิจัยพบว่าเทคนิคที่ใช้ในการวิจัยนี้สามารถแยกแยะสัญญาณเสียงโงะและเสียงอโงะของผู้ชายได้ดีกว่าเสียงของผู้หญิงอันเนื่องมาจากความถี่เสียงของผู้ชายที่มีความถี่ต่ำกว่าเสียงผู้หญิง ในงานวิจัย [6] ได้นำเสนอการสกัดคุณลักษณะของสัญญาณที่เป็นเสียงโงะและเสียงอโงะรวมถึงการนำสัญญาณเหล่านั้นมาทำการสังเคราะห์ขึ้นใหม่โดยอาศัยหลักการของ Gaussian Mixer โดยนำไปประยุกต์ใช้กับการบีบอัดสัญญาณเสียง ซึ่งจากผลการวิจัยพบว่าสามารถนำไปเพิ่มประสิทธิภาพของการบีบอัดสัญญาณเสียงแบบ MELP ได้ และในงานวิจัย [7] ได้ใช้เทคนิคที่นำเอาอาศัยคุณสมบัติทางกรกระจ่ายน้ำหนักของพลังงานเสียง (Weight Distribution) และการกระจายของควมถี่ (Frequency Distribution) ซึ่งมีด้วยกัน 5 วิธี เปรียบเทียบกัน พบว่าวิธีที่ 5 ของการวิจัยเป็นวิธีที่ดีที่สุดในการแยกแยะเสียงโงะและเสียงอโงะ

งานวิจัยนี้ได้เสนอเทคนิคใหม่ในการแยกเสียงโงะและสัญญาณอโงะจากค่าพารามิเตอร์ LSP-10 โดยมีด้วยกัน 4 เทคนิค ซึ่งจะอธิบายถึงทฤษฎีไว้ในส่วนที่ 3 ของงานวิจัย ผลการทดลองจะถูกนำเสนอในส่วนที่ 4 และสรุปผลการทดลองจะถูกกล่าวถึงในลำดับต่อไป

2. สัมประสิทธิ์คู่เส้นสเปกตรัม (LSP)

สัมประสิทธิ์คู่เส้นสเปกตรัม เป็นพารามิเตอร์รูปแบบหนึ่งที่พัฒนาจากสัมประสิทธิ์ทำนายขั้นเชิงเส้น (LPC) เนื่องจากการประมาณค่าพารามิเตอร์ของ LPC อาจไม่สม่ำเสมอของสัญญาณส่งผลให้คุณภาพเสียงต่ำลง ในขณะที่พารามิเตอร์ที่ได้จาก LSP มีคุณสมบัติที่ดีเด่นชัดกว่าคือ ค่าพารามิเตอร์อยู่ภายใต้ขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์และสามารถตรวจสอบเสถียรภาพของวงจรกรองได้ง่าย อีกทั้งยังแสดงอยู่ในรูปเชิงความถี่ซึ่งสามารถนำไปใช้ในการหาคุณสมบัติที่แน่นอนในระบบการรับรู้ของเสียงได้

ในการคำนวณสามารถออกแบบกรอบสเปกตรัมของสัญญาณช่วงสั้นได้จากการวิเคราะห์การทำนายแบบเชิงเส้นในรูปแบบของ All Pole Model ดังสมการที่ (1)

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

โดยวงจรกรอง $A(z)$ ได้ถูกเรียกชื่ออีกหนึ่งว่าวงจรกรองผกผัน (Inverse Filter) เนื่องจากเป็นส่วนผกผันของรูปแบบ All Pole $H(z)$ ของสัญญาณเสียงพูด และรากของ $A(z)$ ทำให้เกิดโพลใน $H(z)$ นั้น คือตำแหน่งของเสียงที่ได้จากช่องทางเดินเสียง ในการคำนวณหาสัมประสิทธิ์คู่เส้นสเปกตรัมเริ่มต้นจากพหุนามอันดับ M ของวงจรกรองในเชิงซัดดังสมการที่ (1) โดยทำการแยกส่วนสมการดังกล่าวเป็นพหุนามอันดับ $M+1$ เป็นจำนวนสองพหุนามดังสมการที่ (2) และ (3)

$$P(z) = A(z) + z^{-(M+1)}A(z^{-1}) \quad (2)$$

$$Q(z) = A(z) - z^{-(M+1)}A(z^{-1}) \quad (3)$$

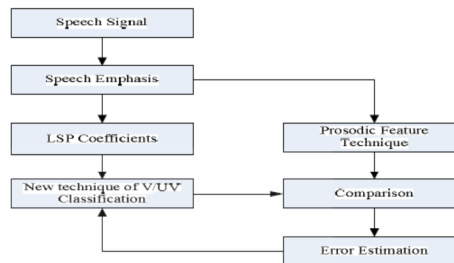
ทำให้ความสัมพันธ์ของ $A(z)$ เป็นไปดังสมการที่ (4)

$$A(z) = \frac{P(z) + Q(z)}{2} \quad (4)$$

พหุนามของ $P(z)$ และ $Q(z)$ สอดคล้องกับรูปแบบช่องทางเดินเสียงที่ไร้การสูญเสียขณะที่ช่องระหว่างเส้นเสียง ปิดเปิดตามลำดับ และรากของพหุนาม $P(z)$ และ $Q(z)$ เรียกว่า ความถี่สเปกตรัม [8]

3. เทคนิคการวิเคราะห์เสียงจากสัมประสิทธิ์คู่เส้นสเปกตรัม

งานวิจัยนี้ได้นำเสนอเทคนิคใหม่ในการวิเคราะห์แยกแยะเสียงโห่และเสียงโห้ (Voiced/Unvoiced) จากค่าพารามิเตอร์ LSP-10 โดยมีรายละเอียดแสดงดังบล็อกไดอะแกรมในรูปที่ 1



รูปที่ 1 บล็อกไดอะแกรมเทคนิคการวิเคราะห์เสียง

จากบล็อกไดอะแกรมในรูปที่ 1 โดยมีกระบวนการดังนี้ เริ่มต้นที่สัญญาณเสียงพูดจะเข้ากระบวนการ Speech Emphasis เพื่อลดสัญญาณรบกวน จากนั้นนำสัญญาณเสียงพูดที่ได้ไปเข้ากระบวนการเพื่อหาค่าสัมประสิทธิ์คู่เส้นสเปกตรัม LSP Coefficients เมื่อได้ LSP Coefficients แล้ว นำค่าเข้าสู่เทคนิคใหม่ในการแยกแยะเสียงโห่และเสียงโห้ ซึ่งมีด้วยกัน 4 วิธี พร้อมกับเปรียบเทียบกับ Prosodic Feature Technique เพื่อหาความผิดพลาดของการวิเคราะห์ (Error Estimation) โดยรายละเอียดเทคนิค 4 วิธี จะแสดงในลำดับต่อไป

3.1 เทคนิคที่ 1

เทคนิคนี้เป็นการตัดสินใจเพื่อหาเสียงโห่จากค่าพารามิเตอร์ LSP-10 โดยการกำหนดค่าการตัดสินใจ (Threshold) ของค่าเฉลี่ยของเฟรมที่มีค่าสูงสุดเทียบกับค่าเฉลี่ยของแต่ละเฟรมดังสมการที่ (5) และ (6)

$$\bar{X}_n = \frac{1}{N} \sum_{m=1}^N \Psi(m) \quad (5)$$

$$Th = \max_x \times ratio \quad (6)$$

โดย \bar{X}_n คือ ค่าเฉลี่ยของพลังงานทั้งหมดใน 1 เฟรม, Ψ คือ ค่าพารามิเตอร์ LSP-10, m คือ จำนวนครั้งของการซัดตัวอย่าง, \bar{X} คือ ค่าเฉลี่ยในแต่ละเฟรม (Average Frame), N คือ จำนวนการซัดตัวอย่างทั้งหมดใน 1 เฟรม, \max_x คือ ค่าสูงสุดของค่าเฉลี่ยของพารามิเตอร์ LSP-10 และ Th คือ ค่าการตัดสินใจ (Threshold)

3.2 เทคนิคที่ 2

เทคนิคนี้เป็นการตัดสินใจเพื่อหาเสียงโห่จากค่าพารามิเตอร์ LSP-10 โดยการกำหนดค่าการตัดสินใจ (Threshold) ของค่าสูงสุด (Max Frame) ในแต่ละเฟรมโดยเทียบกับค่าเฉลี่ยของแต่ละเฟรม ดังสมการที่ (7) และ (8)

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

$$\bar{X}_n = \frac{1}{N} \sum_{m=1}^N \Psi(m) \quad (7)$$

$$Th = max_m \times ratio \quad (8)$$

โดย m คือ เฟรม (Frame), max_m คือ ค่าสูงสุดของค่าพารามิเตอร์ LSP-10 เสียในแต่ละเฟรม

3.3 เทคนิคที่ 3

เทคนิคนี้เป็นการตัดสินใจเพื่อหาเสียง โฆษะ จากค่าพารามิเตอร์ LSP-10 โดยการกำหนดค่าการตัดสินใจ (Threshold) ของค่าพิสัยของเฟรมที่มีค่าสูงสุด (ค่าพิสัย = ค่าข้อมูลสูงสุด - ค่าข้อมูลต่ำสุด) โดยเทียบกับค่าพิสัยของแต่ละเฟรม ดังสมการที่ (9) และ (10)

$$R_n = max \Psi(n) - min \Psi(n) \quad (9)$$

$$Th = max_R \times ratio \quad (10)$$

โดย R_n คือ ค่าพิสัยใน 1 เฟรม, R คือ ค่าพิสัย ในแต่ละเฟรม, max_R คือ ค่าพิสัยของค่าพารามิเตอร์ LSP-10 ของเฟรมที่มีค่าสูงสุด

3.4 เทคนิคที่ 4

เทคนิคนี้เป็นการตัดสินใจเพื่อหาเสียง โฆษะ จากค่าพารามิเตอร์ LSP-10 โดยการกำหนดค่าการตัดสินใจ (Threshold) ของค่าเบี่ยงเบนมาตรฐานเฟรมที่มีค่าสูงสุด (Standard Deviation Frame) โดยเทียบกับค่าเบี่ยงเบนมาตรฐานของแต่ละเฟรม ดังสมการที่ (11) และ (12)

$$SD_n = \sqrt{\frac{\sum_{m=1}^N (\Psi(m) - \bar{X}(n))^2}{N-1}} \quad (11)$$

$$Th = max_{SD} \times ratio \quad (12)$$

โดย SD_n คือ ค่าเบี่ยงเบนมาตรฐานทั้งหมดใน 1 เฟรม, max_{SD} คือ ค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ LSP-10 ของเฟรมที่มีค่าสูงสุด, SD คือ ค่าเบี่ยงเบนมาตรฐานในแต่ละเฟรม

4. ผลการทดลอง

จากรูปที่ 1 ใช้ตัวอย่างสัญญาณเสียงพูดที่เวลา 1 นาที จัดเก็บในรูปแบบของไฟล์ WAV อัตราการชักตัวอย่าง (Sampling Rate) ที่ 8000 เฮิร์ต ตัวอย่างละ 8 บิตและแบบช่องสัญญาณเดี่ยว (Mono) แล้วแบ่งเฟรม (Frame) เฟรมละ 200 ตัวอย่าง (Sampling) และนำสัญญาณนี้ไปวิเคราะห์กับ Prosodic Technique ซึ่งมีรายละเอียดดังในตารางที่ 1

ตารางที่ 1 ผลการวิเคราะห์แบบ Prosodic Feature Technique

| Number of frame | Voiced (frame) | Unvoiced (frame) |
|-----------------|----------------|------------------|
| 2,400 | 1,411 | 989 |

ส่วนคุณภาพที่ใช้เทคนิคการวิเคราะห์แบบใหม่ทั้ง 4 เทคนิค แสดงดังตาราง 2-5 ในตารางแสดงผลการทดลองในแต่ละวิธี โดยมีการกำหนดค่า Threshold ที่แตกต่างกันเพื่อหาความผิดพลาดของการวิเคราะห์ (Error Estimation) และหาคุณภาพ (quality) เปอร์เซ็นด์วิเคราะห์ที่จำนวนเฟรมเสียง โฆษะ (Voiced Frame) ในแต่ละวิธีเปรียบเทียบกับวิธี Prosodic Feature Technique ในตารางที่ 1

4.1 ผลการทดลองเทคนิคที่ 1

ตารางที่ 2 ผลทดลองของเทคนิคที่ 1

| Threshold (%) | Voiced (frame) | Quality (%) |
|---------------|----------------|-------------|
| 85 | 1,021 | 70.64 |
| 87.5 | 1,322 | 83.05 |
| 90 | 1,926 | 90.74 |
| 95 | 2,283 | 97.1 |
| 100 | 2,326 | 100 |

จากตารางที่ 2 ผลการทดลองเทคนิคที่ 1 Threshold เท่ากับ 87.5% เป็นการวิเคราะห์ที่ดีที่สุดโดยมีเปอร์เซ็นต์คุณภาพที่ 83.05 % มีจำนวนเฟรม โฆษะที่จำนวน 1,322 เฟรม โดยที่ Threshold 90% 95% และ 100 % ถึงจะมีคุณภาพที่มากกว่าแต่จำนวนเฟรม โฆษะมากเกินไป

4.2 ผลการทดลองเทคนิคที่ 2

ตารางที่ 3 ผลทดลองของเทคนิคที่ 2

| Threshold (%) | Voiced (frame) | Quality (%) |
|---------------|----------------|-------------|
| 60 | 2,243 | 55.76 |
| 70 | 2,280 | 98.62 |
| 80 | 2,299 | 100 |
| 90 | 2,299 | 100 |
| 100 | 2,299 | 100 |

จากตารางที่ 3 ผลการทดลองเทคนิคที่ 2 Threshold เท่ากับ 60 % เป็นการวิเคราะห์ที่ดีที่สุดโดยมีเปอร์เซ็นต์คุณภาพที่ 55.76% มีจำนวนเฟรม โฆษะที่จำนวน 2,243 เฟรม โดยที่ Threshold 70% 80% ถึง 100 % ถึงจะมีคุณภาพที่มากกว่าแต่จำนวนเฟรม โฆษะมากเกินไป

การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคล ครั้งที่ 4 วันที่ 3-5 เมษายน 2555

4.3 ผลการทดลองเทคนิคที่ 3

ตารางที่ 4 ผลทดลองของเทคนิคที่ 3

| Threshold (%) | Voiced (frame) | Quality (%) |
|---------------|----------------|-------------|
| 75 | 0 | 0 |
| 85 | 184 | 12.03 |
| 90 | 1,061 | 43.16 |
| 95 | 2,145 | 87.49 |
| 100 | 2,332 | 100 |

จากตารางที่ 4 ผลการทดลองเทคนิคที่ 3 Threshold เท่ากับ 90% เป็นการวิเคราะห์ที่ดีที่สุดโดยมีเปอร์เซ็นต์คุณภาพ 43.16 % มีจำนวนเฟรมโหม่งที่จำนวน 1,061 เฟรม โดยที่ Threshold 90% 95% และ 100 % ถึงแม้จะมีคุณภาพที่มากกว่าแต่จำนวนเฟรมโหม่งมากเกินไป

4.4 ผลการทดลองเทคนิคที่ 4

ตารางที่ 5 ผลทดลองของเทคนิคที่ 4

| Threshold (%) | Voiced (frame) | Quality (%) |
|---------------|----------------|-------------|
| 80 | 2041 | 84.28 |
| 85 | 1375 | 59.61 |
| 90 | 407 | 25.62 |
| 95 | 56 | 3.92 |
| 100 | 2 | 0.08 |

จากตารางที่ 5 ผลการทดลองเทคนิคที่ 4 Threshold เท่ากับ 59.61% เป็นการวิเคราะห์ที่ดีที่สุดโดยมีเปอร์เซ็นต์คุณภาพ 59.61 % มีจำนวนเฟรมโหม่งที่จำนวน 1,375 เฟรม โดยที่ Threshold 90% 95% และ 100 % จะมีคุณภาพน้อยตามลำดับ

5. สรุป

ในการวิจัยได้เสนอเทคนิคใหม่ในการวิเคราะห์แยกแยะเสียงโหม่งและเสียงโหม่งจากคำศัพท์ประสิทธิ์ยูเด็นสเปกตรัม ซึ่งมีด้วยกัน 4 เทคนิคพร้อมเปรียบเทียบกับเทคนิคจันท์ลักษณะ Prosodic Technique โดยมีผลสรุปดังตารางที่ 6

ตารางที่ 6 ผลสรุปทั้ง 4 เทคนิค

| Technique | Threshold (%) | Voiced (Frame) | Quality (%) |
|-----------------|---------------|----------------|-------------|
| 1 st | 87.5 | 1,322 | 83.05 |
| 2 nd | 60 | 2,243 | 55.76 |
| 3 rd | 90 | 1,061 | 43.16 |
| 4 th | 85 | 1,322 | 59.61 |

จากผลสรุปจากตารางที่ 6 แสดงให้เห็นว่าเทคนิคที่ 1 เป็นเทคนิคที่มีประสิทธิภาพที่ดีในการหาคุณภาพการแยกแยะเสียงโหม่งและเสียงโหม่งที่ 83.05 % ที่จำนวนเสียงโหม่งเท่ากับ 1,322 เฟรม โดย

เทคนิคนี้สามารถประยุกต์ใช้กับเทคนิคการเวกเตอร์ควอนไทซ์เซชันสำหรับการบีบอัดสัญญาณเสียงและการเรียนรู้จำของสัญญาณเสียง ซึ่งโดยปกติ LSP Coefficients จะถูกใช้ในส่วนของคุณลักษณะเสียง (Speech Feature)

เอกสารอ้างอิง

- [1] Y. Qi, and B. R. Hunt, "Voice-unvoiced-silence classification of speech using hybrid features and a network classifier, IEEE Trans. Speech Audio Processing, vol 1 No. 2, pp.250-255, 1993.
- [2] J.R. Deller, J.G.Proakis and J.H.Hansen, "Discrete-Time Processing of Speech Signals", 2000.
- [3] J.Srinonchat, "New Technique to Reduce Bit Rate of LPC-10 speech Coder" TENCON 2006, 2006 IEEE Region 10 Conference, pp.1-4, 14-17 Nov, 2006.
- [4] D.Arifanto, "Dual Parameters for Voiced-Unvoiced Speech Signal Determination", International Conference on Acoustics Speech and Signal Processing, 2007, pp. IV-749 - IV-752
- [5] 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
- [6] สุภาวดี กรสิงห์, เกรียงศักดิ์ พัฒนบุรี, เฉลิมเกียรติ สุดาชา และ จักรวีศรี นนทจักร "การวิเคราะห์และเปรียบเทียบเทคนิคการสกัดคุณลักษณะของสัญญาณที่เป็นเสียงและเสียงรบกวน" การประชุมวิชาการเครือข่ายวิศวกรรมไฟฟ้า มหาวิทยาลัยเทคโนโลยีราชมงคลครั้งที่ 4
- [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
- [8] สุรวรรณ รื่นสุคนธ์, จักรวี ศรีนนทจักร "ประสิทธิภาพของ LPC และ LSP ในการรู้จำแนวทางเดินเสียงพูดภาษาไทย", International Joint Conference on Computer Science and Software Engineering, 2008, pp.287-293

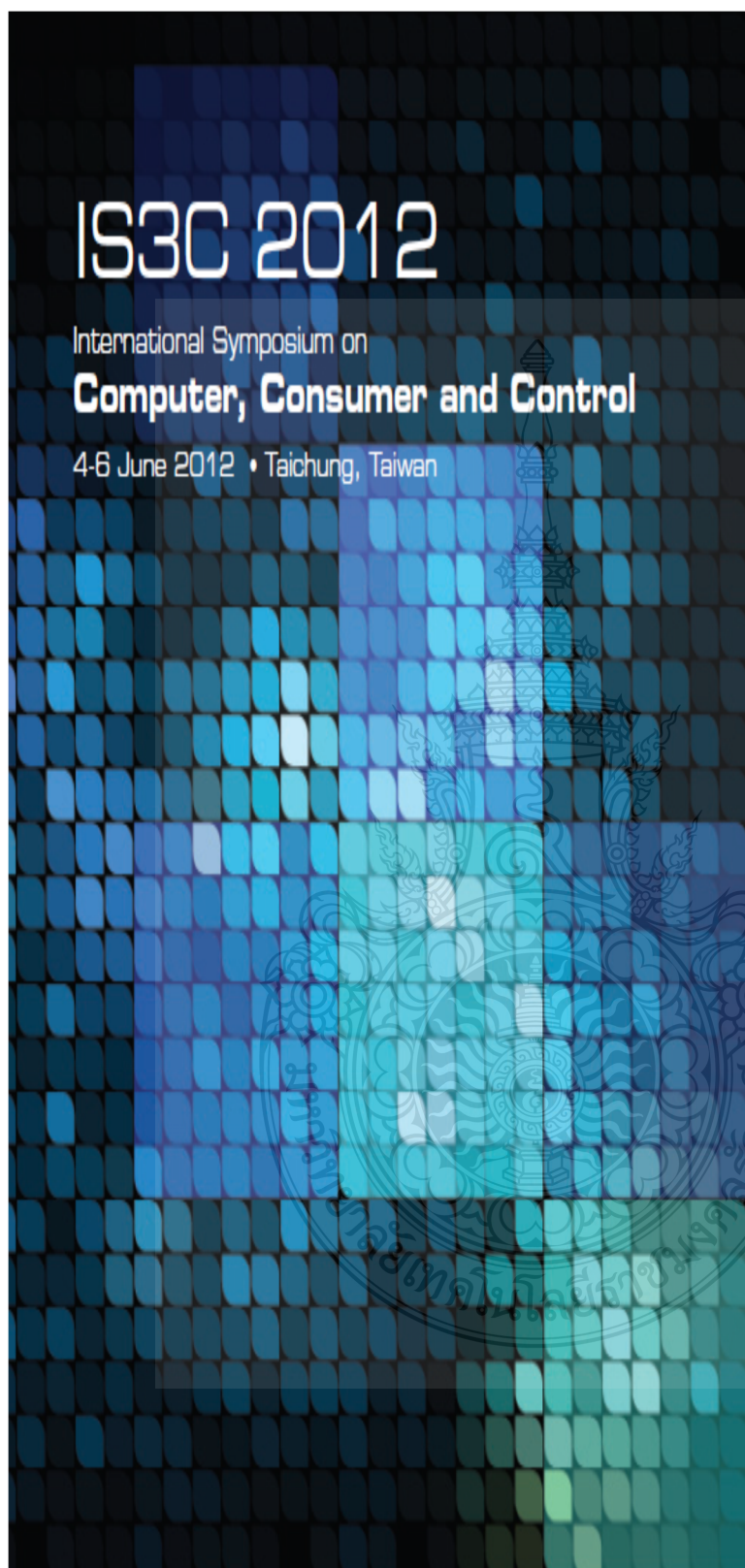


เกรียงศักดิ์ พัฒนบุรี สำเร็จการศึกษาระดับปริญญาตรี ด้านวิศวกรรมอิเล็กทรอนิกส์และ โทรคมนาคม ในปี พ.ศ. 2553 จากมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ปัจจุบันกำลังศึกษาระดับปริญญาโท



ด้านวิศวกรรมอิเล็กทรอนิกส์และ โทรคมนาคม จักรวี ศรีนนทจักร สำเร็จการศึกษาระดับปริญญาเอก จาก Northumbria University, UK. ในปี พ.ศ. 2548 ในสาขาวิศวกรรมไฟฟ้า สาขาเอกวิศวกรรมอิเล็กทรอนิกส์และ โทรคมนาคมปัจจุบันดำรง

ตำแหน่งอาจารย์ผู้สอนที่ภาควิชาอิเล็กทรอนิกส์และ โทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี



The conference cover features a dark blue background with a grid of lighter blue squares. A large, faint watermark of the National Central University (NCU) seal 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-Tsung Yau, *National Chin-Yi University of Technology, Taiwan*
Chuin-Mu Wang, *National Chin-Yi University of Technology, Taiwan*

Publication Chair

Chen-Chung Lin, *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

Pan-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 Yon, *National Chin-Yi University of Technology, Taiwan*

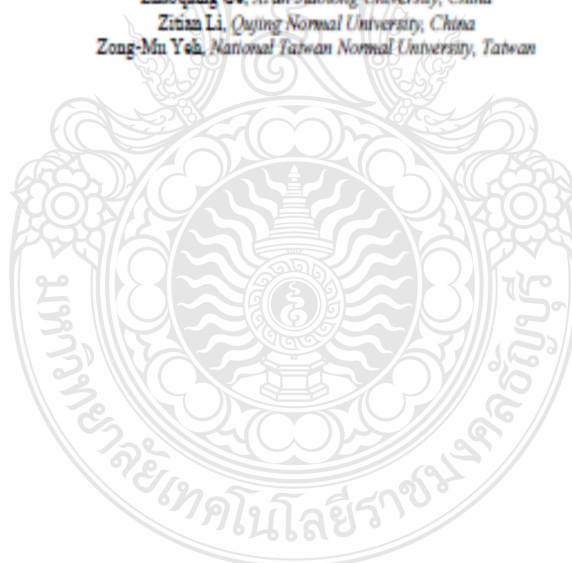
Reviewers

- 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-Tung Yu, *Southern Taiwan University, Taiwan*
 Chao-Wang Young, *National Chiayi University, Taiwan*
 Chao-Yin Hsiao, *Feng Chia University*
 Chen-Chung Lin, *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, *Linghua University of Science and Technology, Taiwan*
 Chien-Chuan Ko, *Chiayi University, Taiwan*
 Chien-Jen Wang, *Kun Shan University, Taiwan*
 Chih-Chyan 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-Jo 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-Pro Huang, *National Chin-Yi University of Technology, Taiwan*
 Chiou-Kou Tung, *National Chin-Yi University of Technology, Taiwan*
 Chin-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*
 Chun-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 Tsang, *National Taipei University of Technology, Taiwan*

Chyi-Wen Huang, *Lan Yang Institute of Technology, Taiwan*
 Cong-Hui Huang, *Far East University, Taiwan*
 Der-Chen Huang, *National Chung-Hsing University, Taiwan*
 Der-Fang Shism, *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*
 Hsai-Yi Lin, *National Chin-Yi University of Technology, Taiwan*
 Hsiang-Yueh Lai, *National Chin-Yi University of Technology, Taiwan*
 Hsin-Chiang Yon, *National Chin-Yi University of Technology, Taiwan*
 Hsing Chia Kuo, *Far East University, Taiwan*
 Hsiuying 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-Chang 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 Sritonchat, *Rajamangala University, Thailand*
 Jeng-Shyang Pan, *National Kaohsiung University of Applied Sciences, Taiwan*
 Jian-Liang Chen, *Kao Yuan University, Taiwan*
 Jian-Der Lee, *Chang Gung University, Taiwan*
 Jian-Shu Lee, *National University of Taiwan, Taiwan*
 Jia-Shang Hu, *National University of Taiwan, Taiwan*
 Jui-Fu Li, *National Central University, Taiwan*
 Jui-Shih Tsai, *National Chung Hsing University, Taiwan*
 Jong Yih Kuo, *National Taipei University of Technology, Taiwan*
 Jui-Nan Chen, *Hsiuping University of Science & Technology, Taiwan*
 Jui-Sheng Lin, *Far East University, Taiwan*
 Jung-Chieh Chen, *National Kaohsiung Normal University, Taiwan*
 Jung-Kuey Tsai, *Asia University, Taiwan*
 Jun-Juh Yan, *Shu-Te University, Taiwan*
 Jyh-Chyang Ruan, *National Yunlin University of Science and Technology, Taiwan*
 Jyun-Jie Wang, *National Chung Hsing University, Taiwan*
 Jzsu-Sheng Lin, *Department of CSE/National Chin-Yi Institute of Technology, Taiwan*
 Jzsu-Sheng Lin, *National Chin-Yi University of Technology, Taiwan*
 Krasangkri 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-Tzung 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*
 Limhong Zhou, *Shanghai University of Finance and Economics, China*
 Lung-Sheng Yang, *Far East University, Taiwan*
 Mao-Jiun J. Wang, *National Yang Hua University, Taiwan*
 Meoi-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-Tyung Tsai, *Southern Taiwan University, Taiwan*
 Ming-Yuan Shieh, *Southern Taiwan University, Taiwan*
 Mirosław Trzupiek, *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*
 Sondrea Xu, *National Taiwan University of Science and Technology, Taiwan*
 Shang-Pin Ma, *National Taiwan Ocean University, Taiwan*
 Shann-Chyi Mon, *Ching Fun University, Taiwan*
 Shao-Ming Yang, *Asia University, Taiwan*
 Sheng-Fui 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*
 Sityam Akshya, *Thapar University, India*
 Shyi-Kao Yang, *Far East University, Taiwan*
 Songdong Xue, *Taiyuan University of Science and Technology, China*
 Tsi-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*
 Trung-Chi Huang, *National Changhua University of Education, Taiwan*
 Trung-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-Chang Pu, *National Chin-Yi University of Technology, Taiwan*
 Wen-Jan Chen, *Da-Yeh University Kaohsiung Normal*
 Wen-Jong 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, *Fusan 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*
 Zhaoxiang Ge, *Nian Jiaotong University, China*
 Zizian Li, *Qiyang 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 Srirongchat</i> | |
| Enhancement Speech Compression Technique Using Modern Wavelet Transforms | 393 |
| <i>Supatinee Komsing and Jakkree Srirongchat</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 Mazinan</i> | |
| An RFID Authentication and Anti-counterfeit Transaction Protocol | 419 |
| <i>Chin-Ling Chen, Yu-Yi Chen, Tzay-Fam Shih, and Tsung-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 Chin-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-Teng Yau</i> | |

Enhancement Pattern Analysis Technique for Voiced/Unvoiced Classification

Kreangsak Pattanaburi
Electronics and Telecommunication Engineering
Department Rajamangala University of Technology
Thanyaburi Pathumthani, Thailand
kreangsak_p@hotmail.com

Jakkrit Onshaunjit and Jakkree Srinonchat
Electronics and Telecommunication Engineering
Department Rajamangala University of Technology
Thanyaburi Pathumthani, Thailand
jakkree_s@hotmail.com

Abstract— Requirement to deciding whether a given frame of a speech waveform should be classified as voiced speech or unvoiced speech arises in many speech analysis systems. Several approaches have been described in the literature for making this decision. In this article presents four enhancement pattern analysis techniques to classify voiced and unvoiced based on the linear predictive coefficients. Those techniques are also compared the performance with the prosodic technique. Ten minutes of speech signal are collected to be input speech. The results show that the 4th technique provides the best performance of the quality of V/UV classification at 89.29% with 19,356 voiced frame. This technique can apply to vector quantization technique for speech compression and speech recognition which usually uses the LP coefficients as the speech feature.

Keywords- voiced/unvoiced classification; pattern analysis; linear predictive coefficients

I. INTRODUCTION

The classification of the short-time speech segments into voiced/unvoiced (V/UV) is critical in many speech analyses-synthesis systems such as speaker identification and speech compression. The essence of classification is to determine whether the speech production system involves vibration of the vocal cords. The problem of V/UV speech determinations is an important one and has been worked on extensively by researchers [1]-[5] during the last three decades. In [1,2] a statistical parametric methods was proposed whereas in [3,4,5] non-parametric method based on linear discrimination functions, multi-layer feedforward and recurrent neural networks were adopted. However, the classification process is the speech feature extraction, which can classify into two groups. Firstly, the prosodic feature such as energy and fundamental frequency. Secondly the spectral-based feature such as the Linear Predictive Coefficient (LPC). This article is focused to use the LPC to be input pattern for V/UV classification.

This article studies the new pattern analysis to classify V/UV speech signal based on the LPC. There are some related research works for V/UV classification such as the application of the notion of instantaneous frequency amplitude spectrum (IFAS) [6] to discriminate V/UV of speech signal. The experimental results show that, the IFAS-based V/UV classifier performs better to the male speaker group than that of the female speaker groups by error rate roughly about 5%. Also [7], it has applies a

wavelet for V/UV classification. The experimental results show that, this technique can classify male speech better than female because the frequency of male speech is lower than female.

All those V/UV classifications can be performed using a single feature, whose behavior could be significantly affected by the presence or absence of voicing activity. The accuracy of such an approach would not go beyond a certain limit, because the range of values of any single parameter generally overlaps between different categories. However, a failure in the detection of periodicity in some voiced regions would results in V/UN error. This article presents and compares the new techniques for errorless V/UV classification based on LPC pattern.

This article is organized as follows. Section II describes the linear predictive coefficients while Section III details the proposed V/UV classification technique. Section IV shows its simulations results. Finally, Section V concludes this work.

II. LINEAR PREDICTIVE COEFFICIENTS

The basic idea of the linear prediction parameters is that the next sample speech signals can be predicted by a linear combination of the past values of the sample signal at time (n). This is shown in the following equation:

$$S_n = e_n + \sum_{k=1}^p a_k S_{(n-k)} \quad (1)$$

Where S_n the value of sample is signal at time (n)

a_k is the predictor parameters

e_n is the prediction error

From the (1), it can be defined that

$$\hat{S}_n = \sum_{k=1}^p a_k S_{(n-k)} \quad (2)$$

Taking z-transforms gives

$$S(z) = E(z) + \left[\sum_{k=1}^p a_k z^{-k} \right] S(z) \quad (3)$$

$$S(z) = \frac{E(z)}{\left(1 - \sum_{k=1}^p a_k z^{-k}\right)} + E(z).H(z) \quad (4)$$

Where $S(z)$ and $E(z)$ are the z transform of $S_{(n)}$ and $e_{(n)}$ respectively. Thus $H(z)$ can be defined as

$$H(z) = \frac{1}{\left(1 - \sum_{k=1}^p a_k z^{-k}\right)} \quad (5)$$

Which $H(z)$ the transfer function of a digital filter is as refers to all-pole system. Thus (5) can be rewrite as

$$H(z) = \frac{1}{A(z)} = \frac{1}{\left(1 - \sum_{k=1}^p a_k z^{-k}\right)} \quad (6)$$

However, a general transfer function of a real vocal has both poles and zeros.

III. V/UV CLASSIFICATION TECHNIQUE

This article presents the new pattern analysis techniques for V/UV classification based on the LPC. The details of those techniques are followed. Then the results of those techniques is compared with the another V/UV classification based on based on the prosodic feature technique.

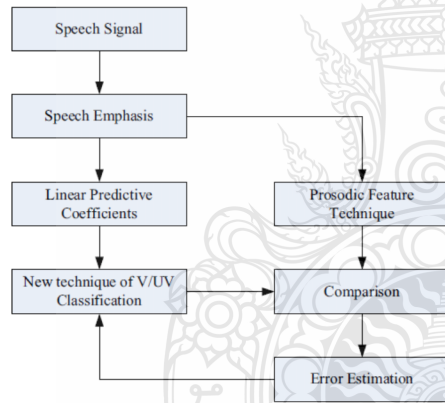


Figure 1 the process diagram

The process diagram is shown in Fig 1. Speech signal is pass through the speech emphasis technique to filter the back ground noise. Then the LPC technique extracts the speech signal to be the speech coefficients as the input pattern for V/UV classification. This article introduces four pattern analysis techniques for errorless V/UV classification, which compares the results with the prosodic

feature technique. The detail of V/UV classification techniques and the prosodic feature technique is followed.

A. The prosodic feature technique

The prosodic feature technique uses the (7) to determine E_i which is used to identify voiced and unvoiced.

$$E_i = \left(\sum_{n=0}^{N-1} |x_i(n)|^2\right) / N \quad (7)$$

B. 1th average technique of V/UV classification

This technique uses the (8) and (9) to determine \bar{X}_n and Th which are used to identify voiced and unvoiced.

$$\bar{X}_n = \frac{1}{w} \sum_{m=1}^w \Psi(n) \quad (8)$$

$$Th = \max_x \times \text{ratio} \quad (9)$$

Where \bar{X}_n is the average value of n frame

Ψ is the LPC coefficients

n is the number of sampling

\bar{X} is the average value in each frame

w is the number of sampling in one frame

\max_x is the maximum value of the LPC coefficients

Th is the threshold

C. 2nd average technique of V/UV classification

This technique uses the (10) and (11) to determine \bar{X}_n and Th which are used to identify voiced and unvoiced.

$$\bar{X}_n = \frac{1}{w} \sum_{m=1}^w \Psi(n) \quad (10)$$

$$Th = \max_m \times \text{ratio} \quad (11)$$

Where \max_m is the maximum value of the LPC coefficients in each frame.

D. Value range technique of V/UV classification

This technique uses the (12) and (13) to determine R_n and Th which are used to identify voiced and unvoiced.

$$R_n = \max \Psi(n) - \min \Psi(n) \quad (12)$$

$$Th = \max_m \times \text{ratio} \quad (13)$$

Where R_n is the value range in each frame

(value range = maximum values - minimum values)

\max_m is the maximum of the LPC coefficients in each frame

E. Standard deviation technique of V/UV classification

This technique uses the (14) and (15) to determine SD_n and Th which are used to identify voiced and unvoiced.

$$SD_n = \sqrt{\frac{\sum_{n=1}^w (\Psi(n) - \bar{X}(n))^2}{w-1}} \quad (14)$$

$$Th = max_{SD} \times ratio \quad (15)$$

Where SD_n is the standard deviation
 max_{SD} is the maximum standard deviation of LPC coefficients
 SD is the standard deviation in each frame

IV. EXPERIMENT RESULTS

10 minutes of speech signal is recorded to be input signal which is used 8000 sampling rate and 8 bits per sampling. This input is divided into frame, which each frame length is 200 sampling. This input is operated as diagram in Fig 1. Firstly, the results of the prosodic feature technique show in Table 1.

TABLE I. THE PROSODIC FEATURE TECHNIQUE

| Number of frame | Voiced (frame) | Unvoiced (frame) |
|-----------------|----------------|------------------|
| 24,000 | 19,752 | 4,248 |

The new pattern analysis techniques are performed as shown in Table 2-5. Each table presents the result of each technique. Also the threshold values are adjusted by the error estimation value of the process. The quality means the percentage of voiced frame number in each technique, which is the same as the prosodic feature technique.

A. 1st average technique results

TABLE II. 1ST AVERAGE TECHNIQUE RESULTS

| Threshold (%) | Voiced | Quality |
|---------------|--------|---------|
| 10 | 22,714 | 71.27 |
| 15 | 20,739 | 72.15 |
| 20 | 16,789 | 76.58 |
| 30 | 9,480 | 47.18 |
| 40 | 4,938 | 25.16 |
| 50 | 2,370 | 11.93 |

Table II shows the performance of the 1st average technique. It can be notice that 20% of threshold provides the maximum quality at 76.58% and also it can detect the 16,789 voiced frames from all speech signal. However at the 10% and 15% of threshold, this technique provides the over

performance of voiced detection. It means it recognizes some of unvoiced frame to be voiced frame.

B. 2nd average technique results

TABLE III. 2ND AVERAGE TECHNIQUE RESULTS

| Threshold (%) | Voiced | Quality |
|---------------|--------|---------|
| 10 | 23,109 | 62.74 |
| 15 | 14,814 | 66.08 |
| 20 | 9,678 | 44.78 |
| 30 | 3,950 | 16.28 |
| 40 | 1,580 | 5.92 |
| 50 | 395 | 1.54 |

Table III shows the performance of the 2nd average technique. It can be notice that 15% of threshold provides the maximum quality at 66.08% and also it can detect the 14,814 voiced frames from all speech signal. However at the 10% of threshold, this technique provides the over performance of voiced detection. It means it recognizes some of unvoiced frame to be voiced frame.

C. Value range technique results

TABLE IV. VALUE RANGE TECHNIQUE RESULTS

| Threshold (%) | Voiced | Quality |
|---------------|--------|---------|
| 10 | 22,517 | 82.12 |
| 15 | 20,739 | 81.54 |
| 20 | 17,776 | 84.73 |
| 30 | 11,258 | 57.34 |
| 40 | 5,530 | 28.22 |
| 50 | 2,370 | 12.46 |

Table IV shows the performance of the value range technique. It can be notice that 20% of threshold provides the maximum quality at 84.73% and also it can detect the 17,776 voiced frames from all speech signal. However at the 10% and 15% of threshold, this technique provides the over performance of voiced detection. It means it recognizes some of unvoiced frame to be voiced frame.

D. Standard deviation technique results

TABLE V. STANDARD DEVIATION TECHNIQUE RESULTS

| Threshold (%) | Voiced | Quality |
|---------------|--------|---------|
| 10 | 21,727 | 80.77 |
| 15 | 19,356 | 89.29 |
| 20 | 15,406 | 76.43 |
| 30 | 8,888 | 44.89 |
| 40 | 3,950 | 20.39 |
| 50 | 1,777 | 8.72 |

Table V shows the performance of the standard deviation technique. It can be notice that 15% of threshold provides the maximum quality at 89.29% and also it can detect the 19,356 voiced frames from all speech signal. However at the 10% of threshold, this technique provides

the over performance of voiced detection. It means it recognizes some of unvoiced frame to be voiced frame.

V. CONCLUSION

This article presents the new pattern analysis technique to classify V/UV speech signal based on LPC. Four techniques are performed and compared with the prosodic technique as summary in Table VI.

TABLE VI. CONCLUSION TABLE

| Technique | Threshold | Voiced | Quality |
|-------------------------|-----------|--------|---------|
| 1 st average | 20% | 16,789 | 76.58 |
| 2 nd average | 15% | 14,814 | 66.08 |
| value range | 20% | 17,776 | 84.73 |
| Standard deviation | 15% | 19,356 | 89.29 |

Table VI shows that provides the best performance of the quality of V/UV classification according to the standard deviation technique classifies voiced frame similar to the prosodic feature technique. It provides 19,356 frame with the quality at 89.92 Furthermore, this technique can apply to vector quantization technique for speech compression and speech recognition which usually uses the LP coefficients as the speech feature. It can also reduce time for process by using the pattern analysis technique of LP coefficients.

ACKNOWLEDGMENT

We would like to express our gratitude to all staff of Signal Processing Research Laboratory, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Thailand, for giving us the speech signal. We have furthermore to thank Prof. Dr. Sean Danaher from the University of Northumbria at Newcastle whose help stimulating suggestions. Also we would like to thank the Office of National Research Council of Thailand for financial support (2012) in this research.

REFERENCES

- [1] B. Atal and L. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Trans on ASSP*, vol. Assp-24, pp.201-212, 1976.
- [2] S. Ahmadi and A.S. Spanias, "Ceptum-based pitch detection using a new statistical V/UV classification algorithm," *IEEE Trans. Speech Audio Processing*, vol.7, No.3, pp. 333-338, 1999.
- [3] Y. Qi, and B. R. Hunt, "Voice-unvoiced-silence classification of speech using hybrid features and a network classifier," *IEEE Trans. Speech Audio Processing*, vol.1 No. 2, pp.250-255, 1993.
- [4] T. L. Burrows, *Speech processing with linear and neural network models*, Ph.D. thesis, Cambridge University Engineering Department, UK, 1996.
- [5] J. Srinonchat, *Investigation and Exploitation of the Repetiveness of Speech Signals in a Speaker Dependent Coding System*, Ph.D Thesis, University of Northumbria at Newcastle, UK, 2004.
- [6] D. Arifianto, "Dual parameters for voiced-unvoiced speech signal determination", *International conference on acoustics speech and signal processing*, pp. IV-749 - IV-752, 2007.

- [7] R. S. Cai, Y. T. Zhu and Y. M. Guo, "Wavelet-based multi-feature voiced/unvoiced speech classification algorithm", *Conference on wireless mobile and sensor networks*, pp. 897 – 900, 2007.

Singapore Institute of Electronics (SIE)

Acceptance Letter

2012 2nd International Conference on Information and Electronics Engineering (ICIEE 2012)
Singapore, October 5 - 7, 2012
<http://www.iciee.org/>



Paper ID: L0026

Paper Title: **Efficiency of Ordered Codebook Learning Vector
Quantization for Speech Compression**

Dear Kreangsak Pattanaburi and Jakkree Srinonchat,

With heartiest congratulations I am pleased to inform you that based on the recommendations of the reviewers and the Technical Program Committees, your paper identified above has been accepted for publication and oral presentation by 2012 2nd International Conference on Information and Electronics Engineering (ICIEE 2012)

ICIEE 2012 conference received over 120 submissions from countries and regions so far, reviewed by international experts; the acceptance ratio is controlled below 38%. Your paper will be published in the conference proceeding after registration.

Herewith, the conference committee sincerely invites you to come to present your paper at ICIEE 2012 to be held on October 5 - 7, 2012, Singapore.

Yours sincerely,



ICIEE 2012 Organizing Committees

<http://www.iciee.org/>

Singapore

Singapore Institute of Electronics (SIE)
E-mail: info@sie-edu.sg, <http://www.sie-edu.sg>

Efficiency of Ordered Codebook Learning Vector Quantization for Speech Compression

Kreangsak Pattanaburi and Jakkree Srinonchat

Abstract—Combined compression and classification problems are becoming increasingly important in many applications with large amounts of data and large sets of classes. This article presents the efficiency of ordered codebook learning vector quantization (OC-LVQ) for speech compression. The algorithm is based on competitive networks. It is developed and analyzed a learning vector quantization based algorithm for combined speech compression and classification. The Peak Signal to Noise Ratio (PSNR), Signal to Noise Ratio (SNR), and Normalized Root Mean Square Error (NEMSE) are used to measure the quality of speech signal. It provides the maximum quality at 28.9432 dB and 15.0333 dB for SNR and PSNR respectively. Also the minimum error of NEMSE is 0.1578.

Index Terms—speech compression, ordered codebook, learning vector quantization.

I. INTRODUCTION

The requirements of a speech compression signal have been sought in mainly speech coding research centers. As a result many different strategies for the suitable speech compression applications have been developed. The exploitation of bit rate speech coders have been standardized in many international and national communication systems [1].

In speech signal processing, the amount of data analyzed require a long time process. To process audio faster, speech coding or speech compression is to reduce size of the speech signal input [2], [3]. It is essentially technique for communication system which obviously uses in many researches such as, presents [4] the Kmean-LBG algorithm and KSOFM algorithm, which were investigated to use in speech coding system. The experimental results show the comparison of the performance of ordered and disordered codebooks which employ to measure and classify the repetition of the speech signal coefficients. Both ordered and disordered codebooks can reduce the number of bit rate transmission around 20% in speech coding system. Also presents [5] a good quality speech data at a low bit rate. In order to accomplish this, the most powerful speech analysis and compression techniques such as Linear Predictive Coding (LPC), Discrete Cosine Transformation (DCT) and Discrete Wavelet Transformation (DWT) are adopted for Tamil speech database. The adopted techniques are

evaluated based on Compression ratio, Peak Signal to Noise Ratio (PSNR),

Normalized Root Mean Square Error (NRMSE) and Mean Opinion Score (MOS). The results show that the DWT achieves greater performance than other two techniques employed in this research.

This article presents an exploitation of Learning Vector Quantization (LVQ), in ordered codebook for speech compression. The adopted techniques are evaluated based on Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE). It is organized as follows. Section II describes the linear predictive coefficients while Section III details the LVQ neural network. Section IV details experiment while V shows its simulations results. Finally, Section VI concludes this work.

II. LINEAR PREDICTIVE COEFFICIENTS

The basic idea of the linear prediction parameters is that the next sample speech signals can be predicted by a linear combination of the past values of the sample signal at time (n). This is shown in the following equation:

$$S_n = e_n + \sum_{k=1}^p a_k S_{(n-k)} \quad (1)$$

Where S_n the value of sample is signal at time (n)

a_k is the predictor parameters

e_n is the prediction error

From the (1), it can be defined that

$$\hat{S}_n = \sum_{k=1}^p a_k S_{(n-k)} \quad (2)$$

Taking z-transforms gives

$$S(z) = E(z) + \left[\sum_{k=1}^p a_k z^{-k} \right] S(z) \quad (3)$$

$$S(z) = \frac{E(z)}{\left(1 - \sum_{k=1}^p a_k z^{-k}\right)} + E(z)H(z) \quad (4)$$

Where $S(z)$ and $E(z)$ are the z transform of $S_{(n)}$ and $e_{(n)}$ respectively. Thus $H(z)$ can be defined as

$$H(z) = \frac{1}{\left(1 - \sum_{k=1}^p a_k z^{-k}\right)} \quad (5)$$

Manuscript received June 29, 2012

Kreangsak Pattanaburi is with the Department of Electronic and Telecommunication, Rajamangala University of Technology Thanyaburi, Thailand.(e-mail: kreangsak_p@hotmail.com).

Jakkree Srinonchat is with the Department of Electronic and Telecommunication, Rajamangala University of Technology Thanyaburi, Thailand.(e-mail: jakkree_s@hotmail.com).

Which $H(z)$ the transfer function of a digital filter is as refers to all-pole system. Thus (5) can be rewrite as

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^M a_k z^{-k}} \quad (6)$$

However, a general transfer function of a real vocal has both poles and zeros.

III. LVQ NEURAL NETWORK

Learning vector quantization (LVQ) was developed by Kohonen network. LVQ network structure is different from unsupervised training structure. LVQ algorithm is a learning algorithm to train the Kohonen layer with teacher's guide. The architecture of the LVQ network used in this paper is shown in Fig. 1. when x is input network. Output neurons are the nearest to x . It is selected as the "winning neuron".

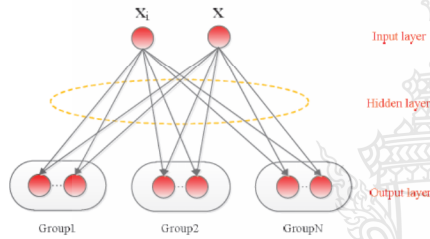


Fig.1. Structure of LVQ network

The proposed learning algorithm of the LVQ networks is as follow:

Prepare the training data. In the input layer, there are C neurons. The continuous are input vectors are

$$X = (x_1, x_2, \dots, x_C) \quad (7)$$

Connection weights vectors between input layer and Kohonen layer are

$$W^1 = (w_1^1, w_2^1, \dots, w_D^1) w_i^1 = (w_{i1}^1, w_{i2}^1, \dots, w_{iC}^1) \quad (8)$$

Where $i = 1, 2, \dots, D$.

Connection weights vectors between Kohonen layer and output layer are

$$W^2 = (w_1^2, w_2^2, \dots, w_k^2) w_k^2 = (w_{k1}^2, w_{k2}^2, \dots, w_{kC}^2) \quad (9)$$

Where $k = 1, 2, \dots, E$.

Every Kohonen neuron is assigned to an output neuron and corresponding connection weights vector is 1 and other connection weights vector is 0. W^2 is fixed during training process.

$$W_{kr}^2 = \begin{cases} 1 & r = k \\ 0 & r \neq k \end{cases} \quad (10)$$

Suppose training mode as follow

$$\{x_1, t_1\}, \{x_2, t_2\}, \dots, \{x_F, t_F\} \quad (11)$$

Where l_j ($j = 1, 2, \dots, Q$) is object output vector. The output of Kohonen layer is calculated as follow

$$V = W^1 X \quad (12)$$

Then the output vector is

$$T = W^2 V \quad (13)$$

W^1 can be determined as follow:

For every input vector, the network will give a classification result. If the result of classification is correct, the connection weights value can be corrected by (13).

$$i^{*w^1}(t+1) = i^{*w^1}(t) + \eta(p(t+1) - i^{*w^1}(t)) \quad (14)$$

If the result of classification is false, the connection weights value can be correct by (14)

$$i^{*w^1}(t+1) = i^{*w^1}(t) - \eta(p(t+1) - i^{*w^1}(t)) \quad (15)$$

Where $\eta \in (0, 1)$, $i^{*w^1}(t)$ is the connection weights value of the i^{th} neurons at t time.

Repeat this step until the achieved classification rate is satisfied or the maximum number of epochs is reached as show in Fig. 2.

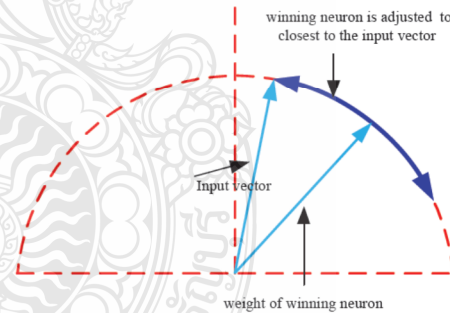


Fig. 2. Learning of LVQ neural network

After learning, the LVQ network can serve to recognize the unknown gas data [6].

IV. EXPERIMENT

There are 20 speech input data which are generated from males and females. The speech signal is sampled at 8 kHz and the frame size is 200 sampling per second. This process research can be classified into three steps as show in Fig. 3.

Firstly speech signal is passing through the speech emphasis technique to filter the back ground noise. Then the

speech signals were calculated the LPC coefficients, where only 10 LPC coefficients represented a speech frame. LPC coefficients were generated in order to compare the effectiveness in the ordered codebook.

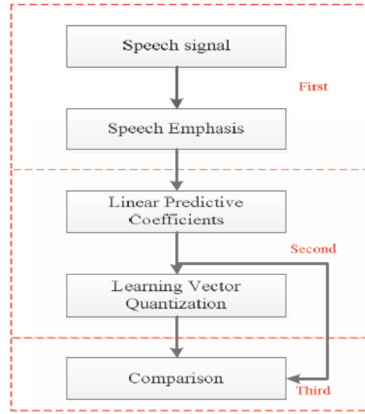


Fig. 3. Experiment processes

Secondly, the LPC coefficients are calculated to become the new LPC coefficients, namely codebooks, using the technique of standard deviation technique of V/U/V classification [7]. The Learning vector quantization (LVQ) is used to classify LPC coefficients into groups for each particular speaker which each address of the codebook contained a set of code vectors, which represents each group. The size of the codebook has an effect on the bit rate which represents the speech coefficients. This experiment uses the sizes of codebooks ranging from 1024 to 64 addressed.

Finally, the comparison on of the performance between the different ordered codebook using PSNR, SNR and NRMSE.

A. Peak Signal to Noise Ratio (PSNR)

$$PSNR = 10 \log_{10} \frac{NP^2}{\|p-r\|^2} \quad (16)$$

Where N is the length of the reconstructed signal.
 P is the maximum absolute square value of the signal p .

$\|p-r\|^2$ is the energy of the difference between the original and reconstructed signals.

B. Signal to Noise Ratio (SNR)

$$SNR = 10 \log_{10} \left| \frac{\sigma_x^2}{\sigma_e^2} \right| \quad (17)$$

Where σ_x^2 is the mean square of speech signal and
 σ_e^2 is mean square difference between the original and reconstructed signal [4].

C. Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \sqrt{\frac{(p(n)-r(n))^2}{(p(n)-\mu p(n))^2}} \quad (18)$$

Where $p(x)$ is the speech signal.

$r(n)$ is reconstructed signal

$\mu p(n)$ is mean of the speech signal.

V. RESULTS

The results show the comparison of different ordered performance codebook based on the frequency domain, PSNR, SNR and NRMSE.

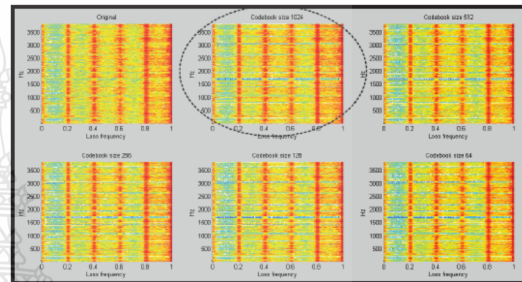


Fig. 4. Comparison of order based on the loss frequency (Male)

In Fig. 4, the 1024 address ordered codebook provides the best performer of male when is compared to all of speech signal in term of the loss frequency domain. It can be notice noise in high of 512,256,128 and 64 address ordered codebook.

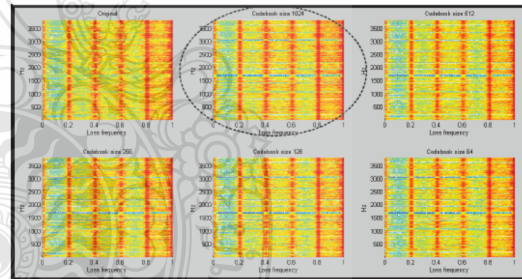


Fig. 5. Comparison of order based on the loss frequency (Female)

In Fig. 5, the 1024 address ordered codebook provides the best performer of female when is compared to all of speech signal in term of the loss frequency domain. It can be notice noise in high of 512,256,128 and 64 address ordered codebook.

TABLE I. COMPARISON OF ORDER BASED ON PSNR

| Speech | Peak signal to noise ratio (dB) | | | | |
|----------|---------------------------------|---------|---------|---------|---------|
| | Codebook size | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| Male 1 | 28.9432 | 28.8432 | 28.5604 | 28.4131 | 28.2671 |
| Male 2 | 24.3278 | 24.3280 | 24.2865 | 24.2376 | 24.1985 |
| Female 1 | 28.1904 | 28.1891 | 28.1448 | 28.1385 | 28.0505 |
| Female 2 | 27.8795 | 27.6994 | 27.7460 | 27.5003 | 27.3852 |

In the Table I, the 1024 address ordered codebook provides the best performer at male1 when is compared to all of speech signal in term of PSNR.

TABLE II: COMPARISON OF ORDER BASED ON PSNR

| Speech | Signal to noise ratio (dB) | | | | |
|----------|----------------------------|---------|---------|---------|---------|
| | Codebook size | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| Male 1 | 14.1268 | 14.1263 | 14.1263 | 14.0962 | 13.9502 |
| Male 2 | 13.0010 | 13.0012 | 12.9365 | 12.8776 | 12.7829 |
| Female 1 | 12.8215 | 12.8189 | 12.8070 | 12.7461 | 12.7093 |
| Female 2 | 15.0333 | 14.8532 | 14.8734 | 14.8195 | 14.6889 |

In the Table II, the 1024 address ordered codebook provides the best performer at female2 when is compared to all of speech signal in term of SNR

TABLE III: COMPARISON OF ORDER BASED ON NRMSE

| Speech | Normalized Root Mean Squared Error | | | | |
|----------|------------------------------------|--------|--------|--------|--------|
| | Codebook size | | | | |
| | 1024 | 512 | 256 | 128 | 64 |
| Male 1 | 0.2000 | 0.2000 | 0.2001 | 0.2007 | 0.2041 |
| Male 2 | 0.1931 | 0.1954 | 0.1945 | 0.1953 | 0.1979 |
| Female 1 | 0.1952 | 0.1954 | 0.1956 | 0.1970 | 0.1979 |
| Female 2 | 0.1578 | 0.1611 | 0.1609 | 0.1616 | 0.1626 |

In the Table III, the 1024 address ordered codebook provides the best performer at female2 when is compared to all of speech signal in term of NRMSE

VI. CONCLUSION

This article presents the efficiency of ordered codebook learning vector quantization for speech compression. In term of the speech quality using PSNR, the large ordered codebook size provides best performer than the small ordered codebook size because of large number of codebooks. It provides the maximum quality at 28.9432 dB and the minimum quality at 24.1985 dB as shown in Table I. In the term of the speech quality using SNR, the large ordered codebook size also provides better performer than the small ordered codebook size. This article provides the maximum quality at 15.0333 dB and the minimum quality at 12.7093 dB as shown in Table II. In the term of the speech quality using NRMSE, the large codebook size was better than the small codebook size because it will have a minimum error value. The article provides the maximum error at 0.2041 and the minimum error at 0.1578 as shown in Table III.

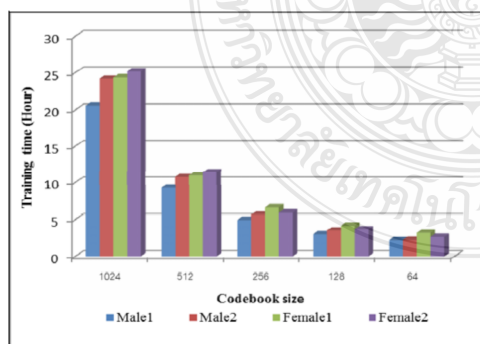


Fig. 6. Comparison of order based on training time

The experiments of using ordered codebook LVQ for low bit rate speech compression show that 1024 address ordered codebook is the best performer but it requires large amount of bit rate to storage and transmission data (10 bits). Also it needs more time to train the winning neuron as in shows in Fig. 6. Therefore this experiment can apply to those speech compression and data storage.

ACKNOWLEDGMENT

We would like to express our gratitude to all staff of Signal Processing Research Laboratory, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Thailand, for giving us the speech signal. We have furthermore to thank Prof. Dr. Sean Danaher from the University of Northumbria at Newcastle whose help stimulating suggestions. Also we would like to thank the Office of National Research Council of Thailand for financial support (2012) in this research.

REFERENCES

- [1] W.T.K.Wong et.al, "Low rate speech coding for telecommunications," BT Tehnology, j, vol14(1), pp.28-43, 1996.
- [2] J.Srinonchat, "Comparison of the efficiency of ordered and disordered codebook techniques in speech coding," IEEE-International Conference on Information and Communication Systems, pp. 195 - 198, 2005.
- [3] J.Srinonchat, "Enhancement artificial neural networks for low-bit rate speech compression system," IEEE- International Symposium on Communications and Information Technologies, pp. 195-198, 2006.
- [4] Weerayuth khunrattanasiri and Jakkree srinonchat, "Comparison efficiency of speech compression using wavelet technique," JICTEE-2010, pp.242-246, 2010
- [5] Dr.V.Radha,Vimala.C and M.Krishnaveni, "Comparison analysis of compression techniques for Tamil speech datasets," IEEE-International conferent on trends in information technology, pp.712-716, 2011.
- [6] Jianye Liu, Yongchun Liang, Xiaoyun Sun, " Application of learning quantization network in faule diagnosis of power transformer," IEEE-International conference on mechatronics and automation, pp.4435-4439, 2009
- [7] Kreangsak Pattanaburi and jakkree Srinonchat, "Enhancement pattern analysis technique for voiced/unvoiced classification," IEEE-International symposium on computer, consumer and control, present june 4-6 2012 , 2012



Kreangsak Pattanaburi received the bachelor's degree in electronic and telecommunication engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 2010 where he is currently studying toward the master's degree in the Department of Electrical Engineering, RMUTT.



Jakkree Srinonchat received bachelor's degree in electronic and telecommunication engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 1995, and his Ph.D. in Electrical Engineering, major signal processing from University of Northumbria at Newcastle, UK, in 2005. He is currently a lecturer of Department of Electronics and Telecommunication Engineering, Faculty of Engineering, RMUTT, Thailand. His research is focus on the signal processing, especially FPGA Design, speech and image processing. He is currently the advisor of the Signal Processing Research Laboratory, which establishes to provide and services the new design and solution to industry.

ประวัติผู้เขียน

| | |
|------------------|--|
| ชื่อ - นามสกุล | นายเกรียงศักดิ์ พัฒนบุรี |
| วัน เดือน ปีเกิด | 28 มีนาคม 2529 |
| ที่อยู่ | 118 หมู่ 2 ตำบล นาคา อำเภอ สุขสำราญ จังหวัด ระนอง 85120 |
| การศึกษา | |
| 2553 - ปัจจุบัน | กำลังศึกษาปริญญาโท คณะวิศวกรรมศาสตร์ สาขาวิชา วิศวกรรมไฟฟ้า แขนง วิศวกรรมอิเล็กทรอนิกส์และโทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคล รัตนบุรี |
| 2550 - 2553 | ปริญญาตรี คณะวิศวกรรมศาสตร์ สาขาวิชา วิศวกรรมอิเล็กทรอนิกส์และ โทรคมนาคม มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนบุรี |
| 2548 - 2550 | ประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาอิเล็กทรอนิกส์ มหาวิทยาลัยเทคโนโลยี ราชมงคลสุวรรณภูมิ วิทยาเขตนนทบุรี |
| 2545 - 2548 | ประกาศนียบัตรวิชาชีพ สาขาวิชาอิเล็กทรอนิกส์ วิทยาลัยเทคนิคระนอง |

