

การแยกประเภทดนตรีอัตโนมัติโดยใช้โครงข่ายประสาทเทียม  
AUTOMATIC MUSIC GENRE CLASSIFICATION USING ARTIFICIAL  
NEURAL NETWORKS

นันทภัสร์ อัมพวานนท์  
NUNTAPUT AMPAWANON

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์  
บัณฑิตวิทยาลัย  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
พ.ศ. 2550

การแยกประเภทดนตรีอัตโนมัติโดยใช้โครงข่ายประสาทเทียม

AUTOMATIC MUSIC GENRE CLASSIFICATION USING ARTIFICIAL  
NEURAL NETWORKS

นันทภัศ อัมพวานนท์

NUNTAPUT AMPAWANON

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิศวกรรมอิเล็กทรอนิกส์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2550

**AUTOMATIC MUSIC GENRE CLASSIFICATION USING ARTIFICIAL  
NEURAL NETWORKS**

**NUNTAPUT AMPAWANON**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF REQUIREMENT FOR THE DEGREE OF MASTER OF  
ENGINEERING IN ELECTRONICS ENGINEERING  
SCHOOL OF GRADUATE STUDIES  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

**2007**

**COPYRIGHT 2007**

**SCHOOL OF GRADUATE STUDIES**

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

หัวข้อวิทยานิพนธ์	การแยกประเภทดนตรีอัตโนมัติโดยใช้โครงข่ายประสาทเทียม
นักศึกษา	นายณัฏฐ์ อัมพวานนท์
รหัสนักศึกษา	47060411
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมอิเล็กทรอนิกส์
พ.ศ.	2550
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.ยุทธนา กิจใจเดียว

### บทคัดย่อ

วิทยานิพนธ์ฉบับนี้เสนอวิธีการแยกประเภทดนตรี ที่ถูกจัดเก็บอยู่ในรูปแบบดิจิทัล ซึ่งเป็นส่วนหนึ่งของการรู้จำรูปแบบ (Pattern Recognition) และการแบ่งแยกข้อมูล (Data Classification) เพื่อเปรียบเทียบประสิทธิภาพในการแยกประเภทดนตรีโดยใช้ลักษณะเด่นทางความถี่ (Spectral Features) สามประเภท คือ ค่าเฉลี่ยกำลังงานทางความถี่ (Spectral Mean), ความตัดกันทางความถี่ (Spectral Contrast) และ ฮิสโทแกรมทางความถี่ (Spectral Histogram) ซึ่งเป็นลักษณะเด่นที่ถูกพัฒนาขึ้นใหม่ในวิทยานิพนธ์นี้ ลักษณะเด่นทั้งหมด ถูกใช้ในการทดลองแยกประเภทด้วย ระบบโครงข่ายประสาทเทียม (Artificial Neural Network) เปรียบเทียบ การดึงลักษณะเด่นมาใช้ นั้น ต้องผ่านการประมวลผลข้อมูลทางดิจิทัลก่อน โดยใช้การปรับสัดส่วนทางความถี่ 2 แบบคือ การปรับสัดส่วนทางความถี่แบบเมล และการปรับสัดส่วนทางความถี่แบบอ็อกเทฟ เพื่อปรับปรุงข้อมูลให้เข้ากับโครงข่ายประสาทเทียม ข้อมูลดนตรีที่ใช้ในการทดลองประกอบด้วยเพลง 5 ประเภทคือ ร็อก, แจ๊ซ, คลาสสิก, ฮิปฮอป และ โพล์ค ทั้งหมด 1,000 เพลง แบ่งเป็นชุดข้อมูลเรียนรู้ 500 เพลง และชุดข้อมูลทดสอบ 500 เพลง ผลที่ได้คือลักษณะเด่นฮิสโทแกรมทางความถี่ที่ใช้ร่วมกับการปรับสัดส่วนทางความถี่แบบเมลนั้น ให้ความถูกต้องในการแยกประเภทสูงถึง 86.77% และเมื่อใช้ร่วมกับการปรับสัดส่วนทางความถี่แบบอ็อกเทฟที่มีจำนวนอินพุทน้อยนั้น ยังคงได้ความถูกต้องในการแยกประเภทดนตรีสูงถึง 77.98%

<b>Thesis Title</b>	Automatic Music Genre Classification using Artificial Neural Networks
<b>Student</b>	Mr. Nuntaput Ampawanon
<b>Student ID.</b>	47060411
<b>Degree</b>	Master of Engineering
<b>Program</b>	Electronics Engineering
<b>Year</b>	2007
<b>Thesis Advisor</b>	Asst. Prof. Dr. Yuttana Kidjaidure

### **ABSTRACT**

This thesis proposes the classification of musical genre of songs recorded in digital files. Our work is in the field of pattern recognition and data classification. To compare the classification efficiency, we used three spectral features (Spectral Mean, Spectral Contrast, and Spectral Histogram developed in this thesis) with the artificial neural network as a classifier. We applied the pre-processor unit to the data before the feature extraction to make the data compatible for the classifier. We provided a thousand of songs in five genres (Rock, Jazz, Classical, Hip-hop, and Folk). All songs were divided into training set data of 500 songs and testing set data of the other 500 songs. The results show that spectral histogram features with Mel-frequency scaling can reach 86.77% of accuracy and with octave-based scaling still reach 77.98% of accuracy.

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความกรุณาจาก ผศ.ดร.ยุทธนา กิจใจเดียว อาจารย์ที่ปรึกษา ที่กรุณาให้ความช่วยเหลือ ให้คำชี้แนะช่วยแก้ปัญหาตลอดจนให้ความรู้แลประสบการณ์ ที่ดี ผู้เขียนรู้สึกซาบซึ้งในความเมตตาของท่านและขอกราบขอบพระคุณเป็นอย่างสูง ขอขอบพระคุณบิดาและมารดา ผู้ซึ่งอบรมสั่งสอน เลี้ยงดู ตลอดจนให้กำลังใจและสนับสนุนในด้านการศึกษากว่าห้าทศวรรษมา

ขอขอบคุณพี่ชาย สำหรับคำแนะนำดีๆเกี่ยวกับการเรียน และคอยส่งเสียมาตลอด

ขอขอบคุณน้องมี ที่ช่วยพิมพ์ตัวหนังสือทุกตัวและสมการทุกสมการในวิทยานิพนธ์เล่มนี้ และขอบคุณสำหรับกำลังใจที่มีให้เสมอมา

ขอขอบคุณน้องโอ น้องเอฟ น้องยัน และเพื่อนหวก เป็นอย่างมาก สำหรับการช่วยเหลือที่ดีเสมอมา

คุณค่าและประโยชน์อันพึงมีจากวิทยานิพนธ์เล่มนี้ ผู้วิจัยขอบอบแด่ผู้มีพระคุณทุกท่าน

นนทภัศ อัมพวานนท์

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	III
สารบัญตาราง.....	VIII
สารบัญรูป .....	XIII
บทที่ 1 บทนำ .....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 สมมติฐานของการศึกษา.....	2
1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย.....	3
1.5 ขอบเขตการวิจัย.....	3
1.6 โครงสร้างวิทยานิพนธ์.....	3
บทที่ 2 การแยกประเภทดนตรีในปัจจุบัน.....	4
2.1 ประวัติการแยกประเภทดนตรี.....	4
2.2 ประเภทของดนตรีจำแนกตามแนวเพลง.....	7
2.2.1 เพลงคลาสสิก (Classical Music).....	7
2.2.2 เพลงอิเล็กทรอนิกส์และเพลงแดนซ์ (Electronic / Dance Music).....	9
2.2.3 เพลงฮิปฮอป (Hip-Hop Music).....	10
2.2.4 เพลงแจ๊สและเพลงบลู (Jazz/Blue Music).....	11
2.2.5 เพลงร็อกและเพลงป๊อป (Rock /Pop Music).....	12
2.2.6 เพลงอาร์แอนด์บี,เพลงโซลและเพลงฟังก์ (R&B/Soul/ Funk Music).....	13
2.2.7 เพลงเวิร์ลด์และเพลงโฟล์ค (World / Folk Music).....	14
2.3 ความสามารถในการแยกประเภทดนตรีของมนุษย์.....	14
บทที่ 3 ลักษณะสัญญาณเสียงดนตรี และการเตรียมข้อมูล เพื่อใช้ในการแยกประเภทดนตรี.....	15
3.1 รูปแบบข้อมูลเสียง.....	15
3.2 การเตรียมข้อมูล.....	16

# สารบัญ(ต่อ)

หน้า

3.2.1 การประมวลผลเบื้องต้น (Pre-processing).....	16
3.2.1.1 การเน้นน้ำหนักเสียงเบื้องต้น (Pre-emphasis).....	16
3.2.1.2 การแปลงฟูริเยร์แบบเต็มหน่วย (DFT).....	18
3.2.1.3 การแปลงฟูริเยร์ในช่วงเวลาสั้นๆ (Short Time Fourier Transform)...	22
3.2.1.4 การปรับสัดส่วนทางความถี่ (Frequency Scaling).....	24
3.2.2 การดึงลักษณะเด่น (Feature Extraction).....	27
3.2.2.1 ค่าเฉลี่ยทางความถี่ (Spectral Mean).....	27
3.2.2.2 ความตัดกันทางความถี่ (Spectral Contrast).....	28
3.2.2.3 ฮิสโทแกรมทางความถี่ (Spectral Histogram).....	31
บทที่ 4 โครงข่ายประสาทเทียม.....	34
4.1 ความเป็นมาของโครงข่ายประสาทเทียม.....	34
4.1.1 ช่วงทศวรรษที่ 40 ยุคเริ่มต้นของโครงข่ายประสาทเทียม.....	34
4.1.2 ช่วงทศวรรษที่ 50-60 ยุคทองของโครงข่ายประสาทเทียม.....	35
4.1.3 ช่วงทศวรรษที่ 70 ยุคเจียบของโครงข่ายประสาทเทียม.....	36
4.1.4 ช่วงทศวรรษที่ 80 ยุคความกระตือรือร้นกลับมาใหม่ของ โครงข่ายประสาทเทียม.....	36
4.1.4.1 โครงข่ายประสาทเทียมชนิด Back-Propagation.....	37
4.1.4.2 โครงข่ายประสาทเทียมชนิด Hopfield.....	37
4.1.4.3 โครงข่ายประสาทเทียมชนิด Neocognitron.....	37
4.1.4.4 โครงข่ายประสาทเทียมชนิด Boltzman.....	38
4.2 โครงข่ายประสาทชีวภาพ.....	38
4.2.1 นิวรอน.....	38
4.2.2 โครงสร้างของนิวรอนและไซแนปส์.....	38
4.2.3 การส่งผ่านกระแสประสาท.....	39
4.3 โครงข่ายประสาทเทียม.....	41
4.3.1 แบบจำลองของนิวรอนที่มี 1 อินพุต.....	41
4.3.2 แบบจำลองของนิวรอนที่มีหลายอินพุต.....	42
4.3.3 ฟังก์ชันกระตุ้น (Activation Function).....	43

## สารบัญ(ต่อ)

หน้า

4.3.3.1 Competitive Transfer Function.....	44
4.3.3.2 Hard limit Transfer Function.....	44
4.3.3.3 Symmetric Hard Limit Transfer Function .....	45
4.3.3.4 Log sigmoid Transfer Function .....	45
4.3.3.5 Positive Linear Transfer Function .....	46
4.3.3.6 Linear Transfer Function .....	46
4.3.3.7 Radial Basis Transfer Function.....	47
4.3.3.8 Saturating Linear Transfer Function .....	47
4.3.3.9 Symmetric Saturating Linear Transfer Function .....	48
4.3.3.10 Soft max Transfer Function.....	48
4.3.3.11 Hyperbolic Tangent Sigmoid Transfer Function .....	49
4.3.3.12 Triangular Basis Transfer Function .....	49
4.3.4 คุณสมบัติและความสามารถของโครงข่ายประสาทเทียม .....	50
4.3.5 ลักษณะของปัญหาที่เหมาะสมกับโครงข่ายประสาทเทียม.....	50
4.4 การเรียนรู้ของโครงข่ายประสาทเทียม .....	51
4.4.1 การเรียนรู้แบบควบคุม (Supervised Learning) .....	51
4.4.2 การเรียนรู้ไม่มีการควบคุม (Unsupervised Learning) .....	51
4.5 สถาปัตยกรรมของโครงข่ายประสาทเทียม (Architectures of Network) .....	52
4.5.1 โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Neural Network) .....	52
4.5.2 โครงข่ายประสาทเทียมแบบหลายชั้น (Multi Layer Neural Network).....	53
4.6 พื้นฐานแนวคิดของโครงข่ายประสาทเทียม (Basis Algorithm of Neural Networks).....	54
4.6.1 โครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ (Back-Propagation Neural Network :BP).....	54
4.6.1.1 โครงสร้าง.....	54
4.6.1.2 แนวคิด.....	55

# สารบัญ(ต่อ)

	หน้า
บทที่ 5 ขั้นตอนการทดลอง.....	64
5.1 การตั้งค่าสำหรับการทดลอง .....	72
5.1.1 สัญญาณเสียง (Audio Signal) .....	72
5.1.2 ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ.....	72
5.1.3 การเน้นน้ำหนักเสียงก่อน (Pre-emphasis) .....	72
5.1.4 การแปลงฟูรีเยร์ในช่วงเวลาสั้นๆ (Short Time Fourier Transfer) .....	72
5.1.5 โครงข่ายประสาทเทียม (Neural Network).....	72
5.2 ตัวแปรต้นและตัวแปรตาม.....	73
5.2.1 ตัวแปรต้น.....	73
5.2.2 ตัวแปรตาม .....	73
บทที่ 6 ผลการทดลอง.....	74
บทที่ 7 สรุปผลการวิจัยและข้อเสนอแนะ.....	106
เอกสารอ้างอิง .....	108
ภาคผนวก ก. โปรแกรมที่ใช้ในการทดลอง .....	110
ภาคผนวก ข. ผลงานวิจัยที่ได้รับการตีพิมพ์ .....	116
ประวัติผู้เขียน.....	122

# สารบัญตาราง

ตารางที่	หน้า
2.1	แสดงผลงานวิจัยต่างๆที่ผ่านมา ..... 7
6.1	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง 24 โหนด ..... 74
6.2	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 18 และชั้นที่สอง 36 โหนด ..... 75
6.3	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง 120 โหนด ..... 75
6.4	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด ..... 76
6.5	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 36 โหนด ..... 76
6.6	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด ..... 77
6.7	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 48 และชั้นที่สอง 96 โหนด ..... 77
6.8	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง 192 โหนด ..... 78
6.9	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384 โหนด ..... 78

## สารบัญตาราง(ต่อ)

ตารางที่	หน้า
6.10	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 96 โหนด ..... 79
6.11	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด ..... 79
6.12	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Meanและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด ..... 80
6.13	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง 24 โหนด .....80
6.14	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 36 และชั้นที่สอง 72 โหนด .....81
6.15	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง 360 โหนด .....81
6.16	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด ..... 82
6.17	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 72 โหนด ..... 82
6.18	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด ..... 83

## สารบัญตาราง(ต่อ)

ตารางที่	หน้า
6.19	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง 192 โหนด .....83
6.20	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384 โหนด .....84
6.21	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 384 และชั้นที่สอง 768 โหนด .....84
6.22	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด ..... 85
6.23	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด ..... 85
6.24	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrastและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 768 โหนด ..... 86
6.25	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง 120 โหนด .....86
6.26	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 180 และชั้นที่สอง 360 โหนด .....87
6.27	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 360 และชั้นที่สอง 720 โหนด .....87

## สารบัญตาราง(ต่อ)

ตารางที่	หน้า
6.28	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด ..... 88
6.29	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด ..... 88
6.30	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 720 โหนด ..... 89
6.31	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง 360 โหนด ..... 89
6.32	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 240 และชั้นที่สอง 480 โหนด ..... 90
6.33	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 480 และชั้นที่สอง 960 โหนด ..... 90
6.34	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด ..... 91
6.35	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 480 โหนด ..... 91
6.36	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 960 โหนด ..... 92

## สารบัญตาราง(ต่อ)

ตารางที่	หน้า
6.37	ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogramและโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 240 และชั้นที่สอง 480 โหนด โดยใช้ข้อมูลฝึกสอนที่ได้จากความยาวเพลงฝึกสอนทั้งเพลง..... 92
6.38	เปรียบเทียบประสิทธิภาพในการแยกประเภทดนตรีโดย ใช้ลักษณะเด่นและการปรับสัดส่วนทางความถี่ต่างๆ ..... 104
6.39	ตารางจัดอันดับความถูกต้องในการแยกประเภทเพลงสูงสุด 3 อันดับแรก ..... 105

# สารบัญรูป

รูปที่		หน้า
2.1	วงออร์เคสตรา (Orchestra) .....	8
2.2	วงแชมเบอร์ (Chamber music) .....	8
2.3	เครื่องดนตรีอิเล็กทรอนิกส์ (Electronics music instruments) ..	9
2.4	ดีเจ (DJ – Disc Jockey) .....	10
2.5	นักร้องเพลงฮิปฮอป (Rapper or MC – Microphone Controller) .....	10
2.6	วงดนตรีแจ๊ส (Jazz music group) .....	11
2.7	วงดนตรีร็อก (Rock music group) .....	12
2.8	วงดนตรีอาร์แอนด์บี (R&B music group) .....	13
2.9	วงดนตรีโฟล์ค (Folk music group) .....	14
3.1	Spectrum ของสัญญาณก่อนการใช้ Pre-emphasis .....	17
3.2	ตัวอย่างของสัญญาณ.....	19
3.3	กราฟของการคำนวณ DFT จำนวน 2 จุด (บัตเตอร์ฟลาย) .....	20
3.4	กราฟของการคำนวณ DFT จำนวน 4 จุด .....	21
3.5	ผลการวิเคราะห์สัญญาณ $y(n)$ ด้วย DFT .....	22
3.6	หน้าต่างของฮาน (Hann Window) .....	23
3.7	หน้าต่างของแฮมมิง (Hamming Window) .....	23
3.8	กราฟความสัมพันธ์ระหว่างความถี่เชิงเส้น (Hertz) กับความถี่เมล (Mel) .....	24
3.9	Mel –Frequency Filter Bank .....	25
3.10	พาวเวอร์สเปกตรัมในหน่วยเฮิรซ์และเมล .....	25
3.12	ตัวอย่างข้อมูลกำลังงานทางความถี่ที่ได้จากการแปลงฟูริเยร์ในช่วงเวลาสั้นๆ .....	27
3.11	ชุดของหน้าต่างสี่เหลี่ยมที่แบ่งช่วงด้วยสเกลลอการิทึมฐาน 2 .....	26
3.13	ตัวอย่างค่าเฉลี่ยทางความถี่ (Spectral Mean) .....	28
3.14	กระบวนการการหาค่าความตัดกันทางความถี่ (Spectral Contrast) .....	29
3.15	ตัวอย่างข้อมูลกำลังงานทางความถี่ที่ถูกเรียงลำดับแล้ว,ค่าสูงสุดและต่ำสุดทางความถี่...30	
3.16	ตัวอย่างลักษณะเด่นความตัดกันทางความถี่ .....	30
3.17	ตัวอย่างกำลังงานทางความถี่ที่ถูกเรียงลำดับแล้ว .....	31
3.18	ตัวอย่างกำลังงานทางความถี่ที่ถูกลดจำนวนตัวอย่าง (Down sampling) แล้ว .....	32
3.19	ตัวอย่างลักษณะเด่นฮิสโทแกรมทางความถี่ .....	32
3.20	เปรียบเทียบขั้นตอนการเตรียมข้อมูลลักษณะเด่น .....	33

## สารบัญรูป(ต่อ)

รูปที่		หน้า
4.1	ตัวอย่างของนิวรอนที่มีสองขั้ว .....	39
4.2	โครงสร้างของประสาท .....	39
4.3	แบบจำลองของนิวรอนที่มี 1 อินพุต .....	42
4.4	แบบจำลองของนิวรอนที่มีหลายอินพุต .....	42
4.5	Competitive Transfer Function .....	44
4.6	Hard Limit Transfer Function .....	44
4.7	Symmetric Hard Limit Transfer Function .....	45
4.8	Log sigmoid Transfer Function .....	45
4.9	Positive Linear Transfer Function ..	46
4.10	Linear Transfer Function .....	46
4.11	Radial Basis Transfer Function.....	47
4.12	Saturating Linear Transfer Function .....	47
4.13	Symmetric Saturating Linear Transfer Function .....	48
4.14	Soft max Transfer Function .....	48
4.15	Hyperbolic Tangent Sigmoid Transfer Function .....	49
4.16	Triangular Basis Transfer Function .....	49
4.17	แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบชั้นเดียว .....	52
4.18	แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบหลายชั้น .....	53
4.19	แสดงแบบจำลองของโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อน .....	54
5.1	ขั้นตอนการแยกประเภทดนตรีด้วยโครงข่ายประสาทเทียม .....	64
5.2	ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Hip-Hop 100 เพลง .....	65
5.3	ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Jazz 100 เพลง .....	65
5.4	ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Classic 100 เพลง .....	65
5.5	ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Rock 100 เพลง .....	66
5.6	ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Folk 100 เพลง .....	66
5.7	ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Hip-Hop 100 เพลง .....	67
5.8	ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Jazz 100 เพลง .....	67
5.9	ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Classic 100 เพลง .....	68
5.10	ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Rock 100 เพลง .....	68

## สารบัญรูป(ต่อ)

รูปที่	หน้า
5.11 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Folk 100 เพลง .....	69
5.12 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Hip-Hop 100 เพลง .....	69
5.13 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Jazz 100 เพลง .....	70
5.16 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Folk 100 เพลง .....	71
5.15 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Rock 100 เพลง .....	71
5.14 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Classic 100 เพลง .....	70
6.1 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Mean ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง .....	93
6.2 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Mean ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง .....	95
6.3 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Contrast ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง .....	97
6.4 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Contrast ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง.....	99
6.5 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Histogram ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง .....	101
6.6 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Histogram ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง .....	103

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน เพลงในท้องตลาดถูกจัดเก็บอยู่ในรูปแบบดิจิทัลเป็นส่วนมาก เครื่องเล่นเพลงในรูปแบบดิจิทัลจึงเป็นที่แพร่หลาย ทั้งที่เป็น ฮาร์ดแวร์และซอฟต์แวร์ ประเภทของคนตรี เป็นคำบรรยายหนึ่ง ที่สามารถบ่งบอกถึงลักษณะต่างๆของคนตรีได้ ทั้งประเภทเครื่องดนตรีที่ใช้เล่น โทนของเสียง สไตส์การเล่น หรือแม้กระทั่งอารมณ์ของเพลง ดังนั้นการทำให้เครื่องเล่นเพลงรู้ว่าเพลงที่เล่นอยู่นั้น เป็นเพลงประเภทอะไร จะเพิ่มขีดความสามารถ และเป็นจุดเด่นใหม่ๆ ของเครื่องเล่นเพลงนั้นๆด้วย ประโยชน์ของการรู้ประเภทเพลงมีหลายอย่าง เช่น เครื่องเล่นเพลงที่เป็นฮาร์ดแวร์สามารถเลือกเล่นเพลงต่อเนื่องกัน ด้วยประเภทเพลงแบบเดียวกัน หรือสามารถปรับ Equalizer ได้ตามประเภทเพลงที่เล่น โดยอัตโนมัติ สำหรับเครื่องเล่นเพลงที่เป็น ซอฟต์แวร์ เช่น โปรแกรม Winamp, Realplayer หรือ Window Media Player มีความสามารถในการปรับ Equalizer ได้เช่นเดียวกับฮาร์ดแวร์ และยังมียอดักประกอบอื่นเช่น Visualization ซึ่งเป็นภาพประกอบที่สามารถแสดงอารมณ์ทางดนตรีได้ ถ้าโปรแกรมรู้ประเภทเพลงที่เล่น ก็จะสามารถเลือก ปรับ Equalizer และ Visualization ที่เหมาะกับเพลงประเภทนั้นๆได้

การศึกษาเรื่องการแยกประเภทดนตรีในปัจจุบัน มีการศึกษาอย่างแพร่หลาย หลังจากปี ค.ศ. 2000 เป็นต้นมา เนื่องจากตลาดเพลงดิจิทัลโตขึ้นอย่างรวดเร็ว มีการขายเพลงผ่านอินเทอร์เน็ต จึงต้องมีการ จัดเก็บข้อมูล และการค้นหาข้อมูลเพลง และการค้นหาเพลงจากประเภทเพลงนั้น สามารถช่วยให้ผู้ค้นหา ได้เลือกสรรเพลงได้ตามความชอบ โดยไม่ต้องรู้จักศิลปินนั้นๆมาก่อน แต่ในงานวิจัยส่วนใหญ่ แยกประเภทดนตรีเป็นหมวดหลัก เพียง 4 ถึง 15 ประเภท ซึ่งประเภทดนตรีที่มนุษย์สามารถแยกย่อยออกมาได้อีกนั้นมีสูงถึง 200 ประเภท เพราะฉะนั้น การที่จะระบุประเภทเพลงให้ละเอียดจริงๆ ตามความต้องการของมนุษย์นั้นเป็นเรื่องที่ยากมาก แต่งานวิจัยส่วนใหญ่ มักจะเสนอประโยชน์ในการแยกประเภทดนตรีเพื่อใช้ในการ จัดเก็บและค้นหาฐานข้อมูลเพลงบนอินเทอร์เน็ต ซึ่งจริงๆแล้ว การแยกประเภทดนตรีเป็นเพียง 4-15 ประเภทนั้น ไม่สามารถใช้ได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ จึงมุ่งเน้นประโยชน์ในการแยกประเภทดนตรีโดยอัตโนมัตินี้ เพื่อประยุกต์ใช้ในเครื่องเล่นเพลง ทั้งที่เป็นฮาร์ดแวร์ และซอฟต์แวร์ ซึ่งในเครื่องเล่นนี้ต้องการข้อมูลประเภทดนตรีเพียงคร่าวๆว่า เป็นเพลงประเภทใด การตั้งจุดมุ่งหมายนี้ เพื่อกำหนดขอบเขตโครงสร้างของระบบทั้งหมด ที่จะทำการวิจัย เช่นการเลือกใช้ตัวแยกประเภท การเลือกใช้ลักษณะเด่น หรือปริมาณข้อมูลที่จะใช้

## 1.2 วัตถุประสงค์ของการศึกษา

วิทยานิพนธ์ฉบับนี้มุ่งหวังเพื่อศึกษาความแตกต่างของเพลงแต่ละประเภท เพื่อหาลักษณะเด่นที่สามารถระบุประเภทของดนตรีในเพลงต่างๆ ให้มีประสิทธิภาพสูงขึ้นจากเดิมที่ใช้ลักษณะเด่นเก่าที่เคยใช้ในงานประเภท รู้จำเสียงพูด (Speech Recognition) เช่น สัมประสิทธิ์เซพตรอลในย่านความถี่แบบเมล (Mel-Frequency Cepstral Coefficient) หรือ สัมประสิทธิ์การทำนายเชิงเส้น (Linear Prediction Coefficient) ในงานวิจัยก่อนหน้านี้ ได้มีการพัฒนาลักษณะเด่นใหม่คือ ความแตกต่างของกำลังงานทางความถี่ (Spectral Contrast) ซึ่งให้ผลการทดลองที่น่าสนใจ โดยได้ทดลองแยกประเภทดนตรีกับ ตัวแยกประเภทเกาเซียนรูปแบบผสม (Gaussian Mixture Model) และได้ผลการแยกประเภทที่ดีกว่า สัมประสิทธิ์เซพตรอลในย่านความถี่แบบเมล ในงานวิจัยนี้ จึงได้ใช้พื้นฐานทฤษฎีดังกล่าว ในการพัฒนาลักษณะเด่นชนิดใหม่ขึ้นมา แล้วเปรียบเทียบประสิทธิภาพในการแยกประเภทดนตรีโดยการนำลักษณะเด่นแบบเดิม

นอกจากเปรียบเทียบลักษณะเด่นแล้ว งานวิจัยนี้ยังมุ่งหวังเพื่อศึกษาความสามารถในการแยกประเภทดนตรี โดยใช้ตัวแยกประเภท แบบโครงข่ายประสาทเทียม โดยทดลองใช้จำนวน โหนดและชั้นซ่อนของประสาทต่างๆ เพื่อสนับสนุนการใช้งาน โครงข่ายประสาทเทียม กับงานแยกประเภทข้อมูลในลักษณะนี้

## 1.3 สมมติฐานของการศึกษา

ลักษณะเด่นทางความถี่ เป็นลักษณะเด่นที่เหมาะสมกับการแยกประเภทเสียงดนตรี ซึ่งมีข้อมูลตามแกนเวลาปริมาณมาก และดนตรีแต่ละประเภทยังมีจังหวะทั้งช้าและเร็วเหมือนกัน ทำให้ข้อมูลทางเวลาไม่มีประสิทธิภาพในการแยกประเภทดนตรี

Spectral Contrast เป็นลักษณะเด่นทางความถี่ที่เพิ่มข้อมูลเชิงปริมาณเข้ามาด้วย โดยการเรียงลำดับ Power Spectrum ในแต่ละความถี่ แล้วนำจุดสูงสุด มาลบกับจุดต่ำสุด เพื่อหาความแตกต่างของ Spectrum การใช้เพียงความแตกต่างระหว่างจุดสูงสุดกับจุดต่ำสุดนั้น ข้อมูลทางปริมาณจะถูกลดทอนออกไปมาก ถ้าเราเลือกเก็บสัญญาณจาก Spectrum ที่เรียงลำดับแล้ว (Spectral Histogram) เหล่านี้ให้ละเอียดขึ้น เราจะสามารถรับรู้ถึง พฤติกรรมทางปริมาณของเพลงนั้นๆด้วย ซึ่งจะช่วยให้เพิ่มประสิทธิภาพในการแยกประเภทดนตรีได้

## 1.4 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

งานวิจัยนี้ ได้สร้างลักษณะเด่นแบบใหม่ ที่พัฒนาต่อยอดมาจากลักษณะเด่นแบบเดิมคือ ค่าความตัดกันทางความถี่ แนวคิดคือ การใช้ลักษณะเด่นทางความถี่ของสัญญาณเพียงอย่างเดียว ไม่สามารถแก้ปัญหาความซับซ้อนของข้อมูลเสียงดนตรีได้ ดังนั้นควรมีวิธีที่จะนำคุณสมบัติอื่นของสัญญาณมาเกี่ยวข้องในการแยกแยะความต่างของเพลงแต่ละแนว คุณสมบัติดังกล่าวคือ คุณสมบัติทางปริมาณ เมื่อผนวกกับข้อมูลทางความถี่แล้ว จะเกิดข้อมูลใน 3 มิติ ซึ่งสามารถบอกได้ว่าสัญญาณที่ความถี่ใดมีปริมาณการใช้ในเพลงๆนั้นเท่าไร ซึ่งสามารถช่วยในการแยกแยะประเภทเพลงได้

## 1.5 ขอบเขตการวิจัย

ในวิทยานิพนธ์ฉบับนี้ได้นำเสนอการแยกประเภทดนตรีด้วยลักษณะเด่นใหม่ คือ Spectral Histogram เปรียบเทียบกับลักษณะเด่น 2 แบบคือ Spectral Contrast และ Power Spectrum โดยใช้โครงข่ายประสาทเทียม เป็นตัวแยกประเภท

การทดลองจะใช้โปรแกรม Matlab ในการประมวลผล และใช้ Neural Network ที่มีให้ใน Matlab ช่วยในการทดลอง เพื่อแยกประเภทเพลง 5 ประเภทได้แก่ เพลงฮิปฮอป (Hip-Hop) เพลงแจ๊ส (Jazz) เพลงคลาสสิก (Classical) เพลงร็อก (Rock) และ เพลงโฟล์ค (Folk)

## 1.6 โครงสร้างวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 7 บท ในแต่ละบทมีเนื้อหาดังต่อไปนี้

บทที่ 1 กล่าวถึงความเป็นมา วัตถุประสงค์ของการศึกษา ความสำคัญของปัญหา ทฤษฎีและขอบเขตที่ใช้ในงานวิจัย

บทที่ 2 กล่าวถึงการแยกประเภทดนตรีในปัจจุบัน วิธีการแยกประเภทต่างๆ และลักษณะเด่นต่างๆ

บทที่ 3 กล่าวถึงลักษณะสัญญาณเสียงดนตรี และการเตรียมข้อมูล เพื่อใช้ในการแยกประเภทดนตรี

บทที่ 4 กล่าวถึงระบบโครงข่ายประสาทเทียมแบบสเกลคอนจูเกตเกรเดียนท์

บทที่ 5 กล่าวถึงขั้นตอนที่ใช้ในการแยกประเภทดนตรีในงานวิจัยนี้

บทที่ 6 กล่าวถึงผลการทดลอง

บทที่ 7 สรุปผลการวิจัยและข้อเสนอแนะ

## บทที่ 2

# การแยกประเภทดนตรีในปัจจุบัน

ในหัวข้อนี้จะกล่าวถึงพื้นฐานต่างๆทางดนตรี การแยกประเภทดนตรีและผลงานต่างๆในด้านการแยกประเภทดนตรีอัตโนมัติที่มีในปัจจุบัน แนะนำตัวแยกประเภทและลักษณะเด่นแบบต่างๆที่ถูกนำมาใช้แยกประเภทดนตรี

### 2.1 ประวัติการแยกประเภทดนตรี

ดนตรีสามารถแบ่งออกเป็นประเภทได้หลายประเภท ตามลักษณะการแบ่งต่างๆ เช่น ความเร็วของเพลง แบ่งเป็น เพลงเร็ว, เพลงช้า ยุคของเพลง แบ่งเป็น ยุค 60s, 70s, 80s และ 90s ความหมายของเพลง แบ่งเป็น เพลงรัก, เพลงเพื่อชีวิต, เพลงปลุกใจ, ฯลฯ ถ้าแบ่งตาม แนวเพลง แบ่งออกได้เป็น Classical, Jazz, Rock, Hip-Hop, Folk, Blue, R&B, Ska, ฯลฯ การแบ่งเพลงเป็นแนวต่าง ๆ นั้น มีประโยชน์อย่างมาก ในการเลือกฟังเพลงของผู้ฟัง เราสามารถค้นหาเพลงในรูปแบบที่ใกล้เคียงกับเพลงที่เราชอบฟังได้ง่าย

แนวเพลงนั้นเป็นการจัดกลุ่มของดนตรีเข้าไว้ด้วยกัน ในหลายๆมุมมอง เช่น เครื่องดนตรีคล้ายกัน Melody คล้ายกัน จังหวะคล้ายกัน อารมณ์เพลงคล้ายกัน ก็จะถูกจัดอยู่ในแนวเพลงเดียวกัน จากการทดลองให้มนุษย์ฟังเพลงหลายๆเพลง แล้วจำไว้ว่าเพลงใดคือแนวเพลงอะไร จากนั้นลองให้ฟังเพลงที่ไม่เคยฟังแล้วลองแยกประเภทดูว่าเป็นเพลงประเภทอะไร บุคคลเหล่านั้นสามารถแยกแนวเพลงได้โดยง่ายในแนวเพลงที่ต่างกัน เช่น Classical และ Rock แต่อย่างไรก็ตามแนวเพลงก็ยังมีขอบเขตที่ไม่ชัดเจนใจการแยก เนื่องจากดนตรีนั้นเป็นศิลปะที่มนุษย์สร้างสรรค์ขึ้นมา การแยกประเภทดนตรีนั้นจึงขึ้นอยู่กับผู้ฟังเองและพื้นฐานทางวัฒนธรรมของผู้ฟังด้วย

อย่างไรก็ตาม การศึกษาการแยกประเภทของดนตรีนี้ ยังไม่แพร่หลายมากนักทั้งๆที่การศึกษาเรื่องการรู้จำเสียงพูดซึ่งมีความใกล้เคียงกัน ถูกให้ความสนใจอย่างสูงมาเป็นทศวรรษ ในช่วงแรกของการศึกษาการแยกประเภทดนตรีจึงอ้างอิงและนำลักษณะเด่นต่างๆมาจากการรู้จำเสียงพูด

เร็วๆนี้ เชียร์เรอร์ (Scheirer) และ สแลนนี่ (Slaney) [1] ได้ศึกษาความแตกต่างระหว่างเสียงพูดกับเสียงดนตรี ซึ่งได้ใช้ลักษณะเด่น 13 อย่าง เช่น อัตราการตัดศูนย์ (Zero crossing rate - ZCR), ศูนย์กลางย่านความถี่ (Spectral centroid), จุดโรลออฟของย่านความถี่ (Spectral roll-off point) และลักษณะเด่นอื่นๆ ได้ทดสอบร่วมกัน โดยใช้ ตัวแยกประเภท 2 ชนิดคือ แบบเกาเซียน

ผสม (Gaussian Mixture Model - GMM) และ (k-Nearest Neighbor - KNN) ที่น่าสนใจคือการเลือกใช้ลักษณะเด่นเพียง 3 ชนิดนั้นได้ผลออกมาดีพอๆกับการใช้ลักษณะเด่นทั้งหมด

ในงานของ ลี (Li) [2] ได้แนะนำลักษณะเด่นที่ชื่อ คอบีซีเวฟเลทโคเอฟฟิเชียนฮิสโทแกรม (Daubechies wavelet coefficient histograms) ซึ่งเป็นลักษณะเด่นแบบใหม่ แล้วเปรียบเทียบกับลักษณะเด่นก่อนหน้านี้ โดยใช้ตัวแยกประเภท 4 ตัว

แลมบรู (Lambrou) [3] ได้ทดสอบ เวฟเลททรานส์ฟอร์ม (Wavelet transforms) แบบอื่นที่ต่างออกไป เพื่อใช้ในการแยกประเภท โดยใช้ตัวแยกประเภทแบบระยะทางต่ำสุด (Minimum distance classifier) และตัวแยกแบบกำลังสองน้อยที่สุดของระยะทางต่ำสุด (Least-square minimum distance classifier) เพื่อแยกประเภทดนตรีให้ออกเป็น ดนตรีร็อก ดนตรีแจ๊ส และดนตรีเปียโน โดยจะมีเปอร์เซ็นต์ความถูกต้อง 60% เมื่อแยกดนตรีออกเป็น 10 ประเภท และมีความถูกต้องสูงถึง 90% เมื่อแยกประเภทของดนตรีออกเป็น 3 ประเภท

แมคคินนี (McKinney) [4] ได้เข้าถึงการแยกเสียงและประเภทดนตรีโดยการเน้นถึงลักษณะเด่น เขาได้แนะนำ ชุดของลักษณะเด่นใหม่ 2 ชุด ซึ่งเป็นแบบเข้าใจง่าย และเปรียบเทียบกับลักษณะเด่นอื่นที่ถูกเสนอก่อนหน้านี้ โดยใช้ การวิเคราะห์ความแตกต่างกำลังสองบนรากฐานของเกาเซียน (Gaussian-based quadratic discriminant analysis) พบว่าลักษณะเด่นที่เข้าใจง่ายนี้ทำงานได้ดีกว่าลักษณะเด่นเก่าๆ โดยการเพิ่ม พฤติกรรมทางเวลาของลักษณะเด่นในช่วงเวลาสั้นๆ (ช่วงละ 23 มิลลิวินาที) และ ผลรวมของค่าเพาเวอร์สเปกตรัม 4 ค่าของแต่ละลักษณะเด่น ที่ถูกพบได้ในช่วงเวลาที่ยาวขึ้น (743 มิลลิวินาที) เข้าไปด้วย ซึ่งวิธีนี้ได้พิสูจน์แล้วว่ามีความแม่นยำทางจังหวะเช่น บีท (beat) รวมอยู่ด้วย

ซานะทาคิส (Tzanetakis) และ กู้ก (Cook) [5] ได้ทดสอบลักษณะเด่นหลายชนิดอย่างเช่น ศูนย์กลางย่านความถี่ (Spectral centroid), MFCCs และที่รวมทั้งของใหม่อย่าง บีทฮิสโทแกรม (beat-histogram) ด้วย โดยใช้ตัวแยกแบบเกาเซียนทั้งสองชนิดคือ GMM และ KNN เพื่อแยกประเภทดนตรีออกเป็นประเภทย่อยๆอีกเช่น ดนตรีคลาสสิก ถูกแบ่งออกเป็น ดนตรีโนโบสท์ (Choir) ดนตรีวงใหญ่ (Orchestra) ดนตรีเปียโน (Piano) และดนตรีเครื่องสายสี่ชิ้น (String quartet)

หนึ่งในหลายๆงานวิจัย จากตารางที่ 2.1 งานวิจัยของ Jiang และคณะ น่าสนใจมาก เพราะว่าจำนวนเพลงที่ใช้ในการทดลองมีถึง 1500 เพลง และใช้ลักษณะเด่นที่เรียกว่า Spectral Contrast เพียงอย่างเดียว แต่ได้ความถูกต้องสูงถึง 82% งานวิจัยนี้จึงมุ่งศึกษาลักษณะเด่นดังกล่าวและพัฒนาเพื่อเพิ่มศักยภาพ ในการแยกประเภทดนตรีต่อไป

ตารางที่ 2.1 แสดงผลงานวิจัยต่างๆที่ผ่านมา

บทความ	จำนวน ประเภท	จำนวน เพลง	ลักษณะเด่น	ตัวแยก ประเภท	ความ ถูกต้อง
Soltau et al., 1998, [6]	4	360	Cepstrum	HMM	79%
				ETM-NN	86%
Pye, 2000, [7]	6	350	MFCC	GMM	92%
				TreeQ	90%
Casey, 2002, [8]	8	หลาย ชั่วโมง	MPEG-7 Low-Level De-scriptor (LLD)	HMM	95%
Jiang et al., 2002, [9]	5	1500	Spectral Contrast	GMM	82%
Tzanetakis et al., 2002, [5]	10	1000	Timbral texture Beat histogram Pitch content	GMM	61%
Burred et al., 2003, [10]	13	850	Timbral Beat histogram MPEG-7 LLD Other	GMM	52%
Li et al., 2003, [2]	10	1000	Daubechies Wavelet Co-efficient Histograms	GMM	64%
				k-NN	62%
				LDA	71%
				SVM	79%
Xu et al., 2003, [11]	4	100	MFCC LPC-derived cepstrum ZCR Spectrum power Beat spectrum	GMM	88%
				HMM	88%
				k-NN	79%
				SVM	93%

## หมายเหตุ

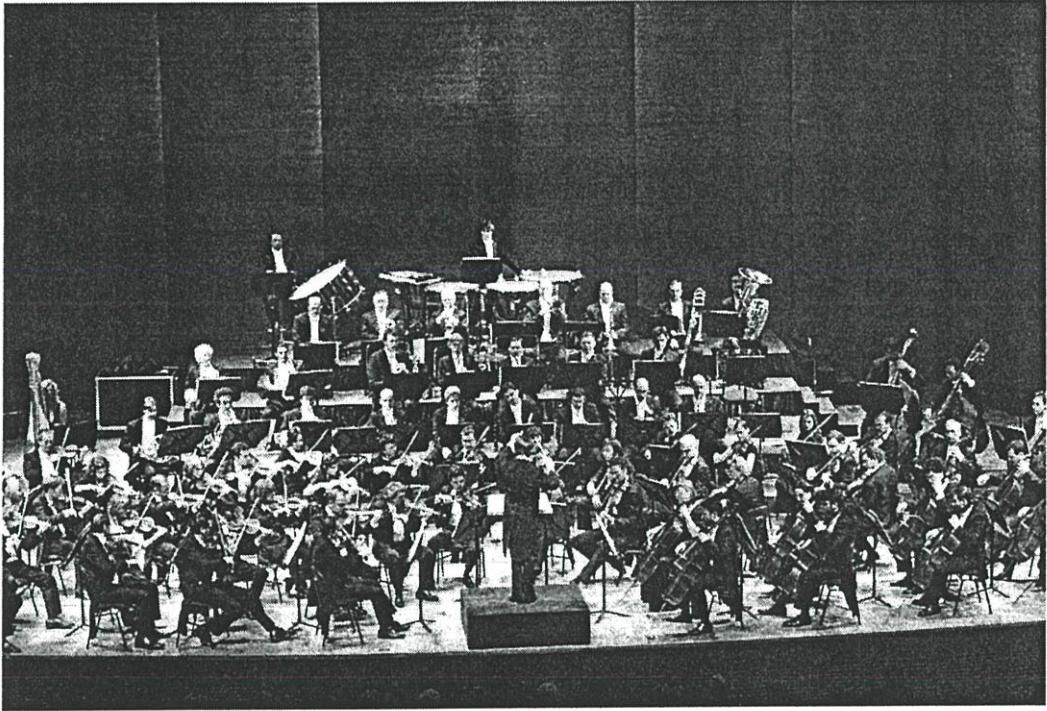
ETM-NN	Explicit Time Modeling with Neural Network
LPC	Linear Prediction Coding
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
k-NN	k-Nearest Neighbor classifier
LDA	Linear Discriminant Analysis
MFCC	Mel Frequency Cepstral Coefficients
SVM	Support Vector Machine
TreeQ	Tree-based Vector Quantization
ZCR	Zero Crossing Rate

## 2.2 ประเภทของคนตรีจำแนกตามแนวเพลง

แนวเพลงเป็นการจำแนกเพลงที่มีลักษณะพื้นฐานต่างๆร่วมกัน โดยทั้งนี้อาจไม่ได้คำนึงถึงเกี่ยวกับด้านดนตรีอย่างเดี่ยว (เช่น ที่มาของเพลง และ เนื้อหาของเพลง เป็นต้น) อาจพูดได้ว่า แนวเพลงนั้นพิจารณาจาก เทคนิค ,สไตส์ ,บริบท ,ที่มา และ เนื้อหาของเพลง เป็นต้น แนวเพลงในโลกนี้มีอยู่มากมายตามแต่จินตนาการของนักดนตรีที่จะสร้างสรรค์ศิลปะด้านนี้ ขึ้นมาได้ ผู้ทำวิจัยจึงขอกกล่าวถึงแนวเพลงหลักๆ เพียง 7 แนวดังต่อไปนี้

### 2.2.1 เพลงคลาสสิก (Classical Music)

บทเพลงที่มีประวัติศาสตร์ยาวนาน เป็นบทเพลงที่ถูกประพันธ์ขึ้นมา ประมาณในช่วงปี ค.ศ. 1770-1830 อย่างไรก็ตาม เรายังรวมไปกับเพลงที่ถูกประพันธ์ขึ้นใหม่ แต่ยังคงใช้แบบฉบับตามวัฒนธรรมทางดนตรีของยุโรปด้วย เพลง Classical ส่วนใหญ่จะถูกประพันธ์แยกเป็นแต่ละเครื่องดนตรี มีเพียงวงที่มีเครื่องดนตรีครบเท่านั้นที่จะสามารถเล่นเพลงๆนั้นได้ เพลง Classical โดยทั่วไปจะถูกเล่น โดยวง Orchestra ซึ่งเป็นวงดนตรีขนาดใหญ่มีเครื่องดนตรีต่างๆอย่างครบครัน โดยมีเครื่องสาย (String instruments) เป็นหลัก เช่น ไวโอลิน วิโอล่า เชลโล และ เบส เล่นร่วมกับเครื่องเป่าโลหะ (brass instruments) เช่น ทรัมเป็ต ทรอมโบน เฟรนชฮอร์น และ ทูบา เครื่องเป่าไม้ (Wood winds) เช่น คลาริเน็ต ฟลูต โอโบ และบาซูน และ เครื่องให้จังหวะ (percussion) เช่น กลอง และ ฉาบ หรืออาจถูกเล่น โดยวง Chamber ที่มีเครื่องดนตรีจำนวนน้อยกว่า



รูปที่ 2.1 วงออร์เคสตรา (Orchestra)



รูปที่ 2.2 วงแชมเบอร์ (Chamber music)

## 2.2.2 เพลงอิเล็กทรอนิกส์ และเพลงแดนซ์ (Electronic / Dance Music)

ในปี ค.ศ. 1970 การทดลองดนตรีอิเล็กทรอนิกส์ Avant-Garde ของวงแบบ Kraft Werk ได้ทำให้การใช้อุปกรณ์อิเล็กทรอนิกส์ ในทางดนตรีแพร่หลายมากขึ้นการพัฒนาของเครื่องสังเคราะห์เสียงทำให้เกิดแนวดนตรีชนิดใหม่ขึ้น การใช้อุปกรณ์กำเนิดเสียงทางอิเล็กทรอนิกส์ในดนตรี จำนวนมากนั้นทำให้ดนตรี มีความเป็นเอกลักษณ์และสามารถเรียกได้ว่า ดนตรีแนวนี้คือ ดนตรีแนวอิเล็กทรอนิกส์ได้ อุปกรณ์อิเล็กทรอนิกส์นี้ได้แก่ เครื่องสังเคราะห์เสียง (Synthesizers), เครื่องเคาะจังหวะ (Drum Machines), เครื่องเก็บตัวอย่างเสียง (Samplers), เครื่องคอมพิวเตอร์ (Computers), เอฟเฟ็ค (Effects), และ เครื่องเล่นเพลง (Record Players) เพลง Electronic ส่วนใหญ่ จะมีจังหวะที่แข็งแรง และมีบีท (Beat) ที่หนักแน่น ถึงแม้จะเป็นแนวเพลงที่เพิ่งกำเนิดขึ้นมาใหม่ก็ แต่ก็มีแนวเพลงย่อยแตกออกมาจนนับไม่ถ้วน



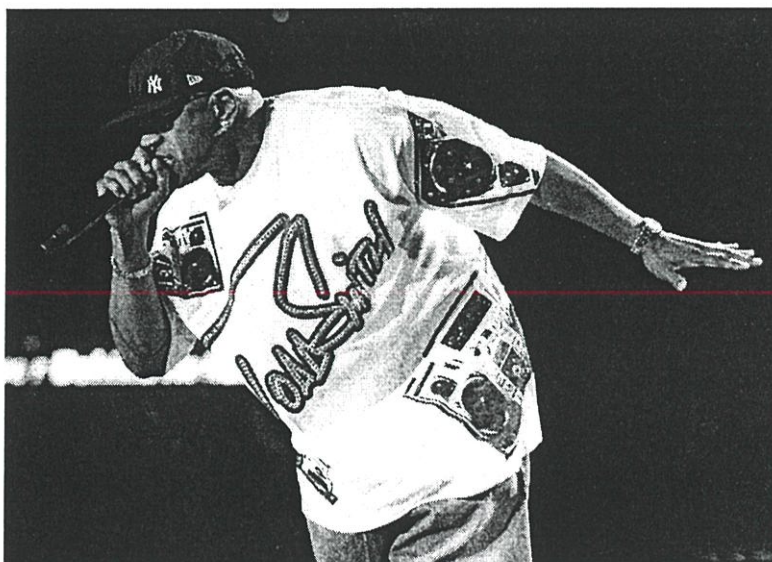
รูปที่ 2.3 เครื่องดนตรีอิเล็กทรอนิกส์ (Electronics music instruments)

### 2.2.3 เพลงฮิปฮอป (Hip-Hop Music)

วัฒนธรรมของชาวฮิปฮอป (Hip-Hop) กำเนิดมากจากสังคมแอฟริกัน-อเมริกัน ที่อยู่ในเมือง ในช่วงปลายปี ค.ศ.1970 และต้นปี ค.ศ. 1980 ส่วนหนึ่งของเพลงแนวนี้คือ คำกลอน และ การแร็ป (Rap) ที่ นักร้อง หรือ แร็ปเปอร์ ร้องไปกับดนตรีที่เป็นบีทแอนด์เบรก (Beat and Break) ซึ่งเล่นจากแผ่นเสียงแนว Disco, Funk, Jazz, Rock หรือ Soul โดยมีดีเจ เป็นคนเล่นแผ่นเสียง หรือจะเป็นคนตรีจากจาก เครื่องเคาะจังหวะ (Drum Machine), เครื่องเก็บตัวอย่างเสียง(Sampler), เครื่องสังเคราะห์เสียง (Synthesizer) ก็ได้ เอกลักษณ์ของเพลงแนวนี้คือ การท่องกลอน หรือคำคล้องจอง ให้เข้ากับจังหวะเคาะในหนึ่งห้องเพลง (Beat) ของเพลงซึ่งอาจจะมาจากหลายๆแนวเพลงดังกล่าว แต่สิ่งที่เหมือนกันของเพลง Hip-Hop ก็คือการท่องกลอนที่คล้องจองและรวดเร็ว



รูปที่ 2.4 ดีเจ (DJ – Disc Jockey)



รูปที่ 2.5 นักร้องเพลงฮิปฮอป (Rapper or MC – Microphone Controller)

#### 2.2.4 เพลงแจ๊สและเพลงบลู (Jazz/Blue Music)

แนวเพลง Blue คือวัฒนธรรมของเพลง Folk ที่กำเนิดโดยคนแอฟริกัน-อเมริกัน ช่วงเริ่มต้นทศวรรษที่ 20 เป็นการรวม วัฒนธรรมทางดนตรีของคน แอฟริกัน กับวัฒนธรรมทางดนตรีของชาวยุโรป เอกลักษณะหนึ่งของแนวเพลง Blue คือการเล่นโน้ตแบบ Blue และใช้เครื่องดนตรีเหมือนกัน ในแต่ละวง โดยมีเครื่องดนตรีหลักๆคือ ดับเบิลเบส ทรัมเป็ต แซ็กโซโฟน เปียโน แนวเพลงJazz เติบโตขึ้นจากเพลง Blue ในช่วงต้นทศวรรษที่ 20 เช่นกัน ซึ่งเพลงJazz มีการเล่นที่เน้นการคิดโน้ตที่จะเล่นขึ้นมาสดๆขณะเล่น (Improvisation) มีการลากจังหวะมีความอิสระใน Melody การคิดโน้ตขึ้นมาสดๆนั้นเป็นเอกลักษณ์ของเพลง Jazz



รูปที่ 2.6 วงดนตรีแจ๊ส (Jazz music group)

### 2.2.5 เพลงร็อกและเพลงป๊อป (Rock /Pop Music)

ปัญหาอย่างหนึ่งที่พบก็คือ ความคลุมเครือระหว่างเพลง Rock และเพลง Pop เนื่องจากเพลงทั้งสองแนวนี้ เป็นเพลงแนวใหม่ที่มีคำอธิบายอย่างกว้าง และค่อนข้างใกล้เคียงกัน จึงแยกออกจากกันได้ยาก การที่จะสามารถแยกเพลง Rock และ Pop ได้จะต้องใช้การแบ่งประเภทขั้นสูง นั่นคือเพลง Rock ที่ใช้กีตาร์ไฟฟ้าและมีการเล่นกลองที่หนักแน่น ปัจจุบันเพลง Rock ได้แตกสาขาออกไปเป็นแนวเพลงย่อยอีกและแนวเพลงย่อยเหล่านี้ได้เติบโตและเป็นแนวเพลงของตัวเอง ส่วนเพลง Pop คือเพลงที่มี Melody สั้นๆง่ายๆต่อการจดจำ ทำให้ติดหูผู้ฟัง



รูปที่ 2.7 วงดนตรีร็อก (Rock music group)

### 2.2.6 เพลงอาร์แอนด์บี,เพลงโซลและเพลงฟังก์ (R&B/Soul/ Funk Music)

Rhythm and Blues (R&B) เป็นแนวเพลงที่ถูกสร้างขึ้นเพื่อเป็นเพลงของคนดำ โดยเฉพาะถูกพัฒนามาจากเพลง Blue และเล่นในแบบของคนในเมือง ซึ่งเพลงR&B เกือบทั้งหมดที่ถูกแต่งขึ้นในปี ค.ศ. 1960 ถึง 1970 เรียกว่าแนวเพลง Soul เพลง Soul นั้นจะมีเอกลักษณ์ คือเสียงร้องที่แสดงอารมณ์อย่างมาก

เพลง Funk เดิมโตมาจากเพลง Soul โดยการใส่ท่วงทำนอง ของเพลง Rock เข้าไป เพลง R&B ในปัจจุบันหรือที่เรียกว่า (RnB) มีลักษณะใกล้เคียงกับเพลง HipHop และเพลง Pop มากขึ้น แต่อย่างไรก็ตามก็ยังเอกลักษณ์ในการร้องที่นุ่มนวลของจังหวะที่หนักแน่น



รูปที่ 2.8 วงดนตรีอาร์แอนด์บี (R&B music group)

### 2.2.7 เพลงเวิร์ลและเพลงโฟล์ค (World / Folk Music)

แนวเพลงทั้งสองนี้เป็นแนวเพลงที่ถูกแบ่งประเภทแบบชั้นสูง เราไม่สามารถแยกประเภทเพลงแนวนี้โดยอัตโนมัติได้ เนื่องจากธรรมชาติหลายๆอย่างของเพลง ซึ่งเพลง World Music คือเพลงอะไรก็ตามที่เล่นด้วยเครื่องดนตรีพื้นบ้านและไม่ได้ถือกำเนิดจากวัฒนธรรมเพลงตะวันตก ส่วนเพลงพื้นบ้านที่เป็นวัฒนธรรมตะวันตกเราจะเรียกว่า เพลง Folk Music



รูปที่ 2.9 วงดนตรีโฟล์ค (Folk music group)

## 2.3 ความสามารถในการแยกประเภทดนตรีของมนุษย์

มีการศึกษาเพียงเล็กน้อยเกี่ยวกับความสามารถของมนุษย์ในการแยกประเภทดนตรี หนึ่งในนั้นมีการศึกษาของเปอร์โรต์ (Perrot) [12] ซึ่งเขาได้ใช้ประเภทดนตรี 10 ประเภทในการศึกษา ได้แก่ Blues, Classical, Country, Dance, Jazz, Latin, Pop, R&B, Rap และ Rock ในประเภทดนตรีแต่ละประเภทประกอบด้วยเพลง 8 เพลง เป็นเพลงที่มีเสียงร้อง 4 เพลง และไม่มีเสียงร้อง 4 เพลง จากการทดลองกับนักศึกษาชั้นปีที่ 1 จำนวน 52 คน โดยให้คนเหล่านี้ฟังเพลงประมาณ 24 ชั่วโมงต่อสัปดาห์ แล้วทดลองฟังท่อนใดท่อนหนึ่งที่ตัดมาจากเพลงที่ให้ทดลองฟัง ผลปรากฏว่า ถ้าให้ฟังเพลงนาน 3 วินาที คนเหล่านี้สามารถแยกประเภทดนตรีได้ถูกต้อง 70 % ถ้าฟังนาน 250 มิลลิวินาที จะได้ความถูกต้อง 40 % ผลการทดลองเป็นที่น่าสนใจอย่างมาก เพราะมันบ่งบอกว่าจริงๆแล้วมนุษย์สามารถแยกแยะประเภทดนตรีได้โดยไม่ต้องใช้ลักษณะของดนตรียุ่งยากอย่างเช่น จังหวะจะโคลนเนื่องจากท่อนของดนตรีที่ยาวเพียง 250 มิลลิวินาที นั้นไม่สามารถแสดงให้เห็นถึงจังหวะจะโคลนได้ สิ่งนี้ทำให้รู้ว่าแค่เพียงลักษณะเด่นทางความถี่ (โทนของดนตรี) ก็ทำให้เราสามารถแยกประเภทดนตรีออกได้แล้ว

## บทที่ 3

# ลักษณะสัญญาณเสียงดนตรี และการเตรียมข้อมูล เพื่อใช้ในการ แยกประเภทดนตรี

ในบทนี้ จะขอลำดับถึงลักษณะการเก็บข้อมูลทางดิจิทัลของเพลง การปรับปรุงข้อมูลที่จะใช้ในกระบวนการแยกประเภทข้อมูล การแปลงข้อมูลทางเวลา ให้อยู่ในโดเมนของความถี่ และการดึงลักษณะเด่น ที่จะนำไปใช้แยกประเภทดนตรีในบทถัดไป

### 3.1 รูปแบบข้อมูลเสียง

ข้อมูลของเพลงต่างๆ ที่ถูกจัดเก็บอยู่ในสื่อประเภท Compact Disc (CD) นั้นจะอยู่ในรูปดิจิทัลและเข้ารหัสแบบ Pulse-Code Modulated (PCM) สัญญาณเสียงจะถูกสุ่มตัวอย่าง (Sampling) ด้วยอัตราสุ่มตัวอย่าง (Sampling Rate) ที่ 44.1 kHz และมีความละเอียด (Resolution) ที่ 16-bit และจำนวนช่องสัญญาณแบบ 2 ช่อง (Stereo)

เนื่องจากเสียงที่มนุษย์ สามารถได้ยินนั้นมีความถี่อยู่ในช่วง 20 HZ-22 kHz อัตราการสุ่มตัวอย่างจึงต้องเป็น 44.1 kHz จะเห็นได้ว่าการเข้ารหัสแบบ PCM นั้นใช้พื้นฐานทางเวลามาเกี่ยวข้อง ซึ่งไม่มีการบีบอัดข้อมูล (Data Compression) เลย ข้อมูลเสียงที่เก็บไว้จึงไม่มีความผิดเพี้ยนจากสัญญาณต้นแบบ

ความละเอียด 16 Bits นั้นหมายถึงเราสามารถบันทึกข้อมูลเสียง ณ การสุ่มตัวอย่างหนึ่งๆ ได้ด้วยระดับที่แตกต่างกันถึง 65,536 ระดับ เพราะฉะนั้นในการเก็บข้อมูลเสียง 1 วินาที เราต้องเก็บข้อมูลมากถึง 705,600 Bits

การบันทึกแบบ Stereo หมายถึงการเก็บข้อมูลของเสียง 2 ช่องสัญญาณสำหรับลำโพงซ้าย และสำหรับลำโพงขวา นั้นหมายถึงข้อมูลที่ต้องการเก็บจะมากขึ้นเป็น 2 เท่า สรุปแล้วการที่เราจะเก็บเพลงหนึ่งเพลงที่มีความยาว 3 นาที เราจะต้องเก็บข้อมูล ปริมาณเท่ากับ 2 ลำโพง x 3 นาที x 44100 Sampling x 16 Bits = 254,016,000 Bits ซึ่งเท่ากับประมาณ 32 MByte

## 3.2 การเตรียมข้อมูล

ในงานด้านการแยกประเภทข้อมูล ( Classification) หรือการรู้จำรูปแบบ (Pattern Recognition) นั้น ขั้นตอนที่สำคัญขั้นตอนหนึ่ง คือการจัดเตรียมข้อมูล เนื่องจากข้อมูลเสียงเพลงที่เรา มีนั้น มีปริมาณมาก และยังไม่พร้อมใช้งาน เราจึงต้องใช้กระบวนการ ประมวลผลเบื้องต้น (Pre-processing) และ การดึงลักษณะเด่น (Feature extraction) เพื่อ ปรับข้อมูลให้เหมาะสมกับการใช้งาน และดึงลักษณะเด่น ที่จะช่วยในการรู้จำประเภทของดนตรีที่เรา กำลังศึกษาอยู่ได้

### 3.2.1 การประมวลผลเบื้องต้น (Pre-processing)

ลักษณะของข้อมูลเสียงเพลงโดยทั่วไป ที่ความถี่ต่ำจะมีขนาดของสัญญาณ (Magnitude) ขนาดใหญ่ และ ที่ความถี่สูงจะมีขนาดของสัญญาณเล็ก ซึ่งขนาดสองขนาดนี้ มีความแตกต่างกันมาก เราจึงต้องใช้ กระบวนการเน้นน้ำหนักเสียงเบื้องต้น ช่วยในการปรับปรุงสัญญาณก่อน เพื่อลดความแตกต่างของข้อมูลลง จากนั้นเราจะใช้ การแปลงฟูรีเยร์ในช่วงเวลาสั้นๆ มาช่วยในการแปลงข้อมูลทางเวลา (Time Domain) ให้กลายเป็นข้อมูลทางความถี่ (Frequency Domain – Spectral) เพื่อนำไปดึงลักษณะเด่นต่อไป

#### 3.2.1.1 การเน้นน้ำหนักเสียงเบื้องต้น (Pre-emphasis)

ถ้ากล่าวถึงกระบวนการเกี่ยวกับเสียงทางดิจิทัลอิเล็กทรอนิกส์ การเน้นน้ำหนักเสียงเบื้องต้นเป็นกระบวนการที่ถูกออกแบบมาเพื่อใช้ในการเพิ่มขนาดของสัญญาณในบางย่านความถี่ (ส่วนใหญ่จะเพิ่มขนาดในย่านความถี่สูง) และลดทอนขนาดของสัญญาณเสียงในบางย่านความถี่ (ส่วนใหญ่จะเป็นย่านความถี่ต่ำ) เพื่อที่จะปรับปรุง อัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal-to-noise ratio) โดยการลดความแตกต่างระหว่างขนาดของสัญญาณย่านความถี่ต่ำ และสัญญาณย่านความถี่สูง ซึ่งเป็นส่วนช่วยในการลดจำนวนข้อมูลที่จะต้องบันทึกลงในสื่อประเภทต่าง แล้วจึงใช้การเน้นน้ำหนักเสียงกลับ (de-emphasis) ช่วยในการกู้คืนสัญญาณเสียงให้เหมือนเดิม ระบบการเน้นน้ำหนักเสียงเบื้องต้น และการเน้นน้ำหนักเสียงกลับ รวมเรียกว่าระบบการเน้นน้ำหนักเสียง (emphasis) เราสามารถหาคุณลักษณะทางความถี่ (frequency curve) ได้โดยใช้ค่าคงที่ทางเวลาเฉพาะ (special time constants) และความถี่คutoffสามารถคำนวณได้จากค่าคงที่ทางเวลาเฉพาะนั้น

การเน้นน้ำหนักเสียงเบื้องต้นถูกใช้อย่างแพร่หลายในงานด้าน การสื่อสาร, การบันทึกเสียงแบบดิจิทัล, การตัดต่อเสียง, การส่งวิทยุกระจายเสียงแบบ FM, และการแสดงสเปกโตรแกรม (spectrograms) ในงานจดจำเสียงพูด

### วิธีการ

เมื่อเราให้  $x_i$  ที่  $i = 1, 2, 3, \dots$  เป็นตัวอย่างข้อมูลในแต่ละการสุ่มตัวอย่าง  $i$  สมการการเน้นน้ำหนักเสียง จะเป็นตามสมการที่ (3.1) โดยที่ ตัวประกอบการเน้นน้ำหนักเสียงก่อน  $\alpha$  คำนวณได้จากสมการที่ (3.2)

$$x_i = x_i - \alpha x_{i-1} \quad (3.1)$$

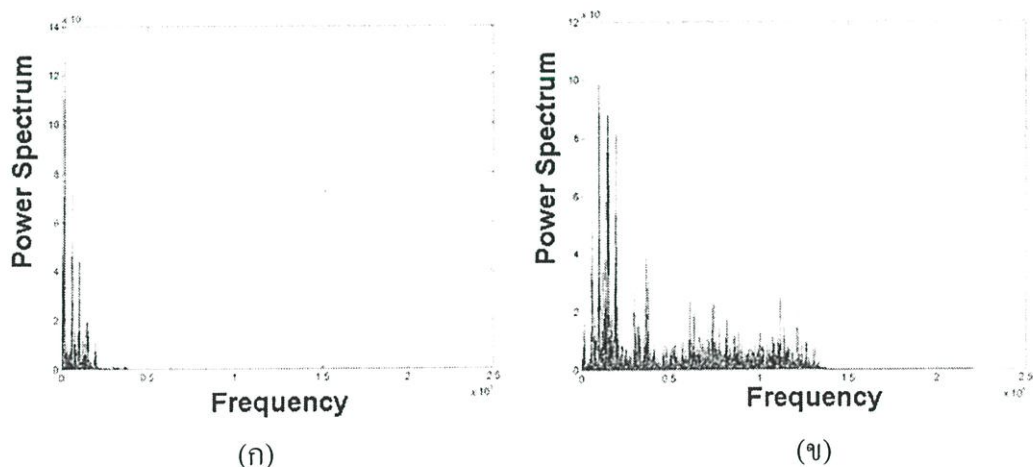
$$\alpha = \exp(-2\pi F \Delta t) \quad (3.2)$$

โดย  $\Delta t$  คือช่วงเวลาเก็บตัวอย่างของเสียงนั้น ในทุกๆตัวอย่าง  $x_i$  ของเสียง(ยกเว้น  $x_1$ ) จะเปลี่ยนไป โดยจะถูกลดทอนด้วยขนาดของตัวอย่างตัวก่อนหน้าคูณกับตัวประกอบ  $\alpha$  และ  $F$  คือความถี่คัทออฟ

ตัวอย่างเช่น ถ้าต้องการเน้นน้ำหนักเสียง กับสัญญาณที่มีความถี่คัทออฟ = 360 Hz และสัญญาณมีอัตราการสุ่มตัวอย่าง 44100 ตัวอย่างต่อวินาที เราสามารถคำนวณหาค่า  $\alpha$  ได้ดังสมการที่ (3.3)

$$\alpha = \exp(-2 \times 3.1415 \times 360 \times (1/44100)) = 0.95 \quad (3.3)$$

หลังจากการเน้นน้ำหนักเสียงแล้ว เราจะได้ตัวอย่างผลดังรูปที่ 3.1



รูปที่ 3.1 (ก) Spectrum ของสัญญาณก่อนการใช้ Pre-emphasis

(ข) Spectrum ของสัญญาณหลังการใช้ Pre-emphasis

### 3.2.1.2 การแปลงฟูรีเยร์แบบเต็มหน่วย (DFT)

ในทางคณิตศาสตร์ การแปลงฟูรีเยร์แบบเต็มหน่วย (Discrete Fourier transform – DFT) หรือบางครั้งเรียกว่า การแปลงฟูรีเยร์แบบจำกัดขอบเขต (Finite Fourier transform) คือการแปลงฟูรีเยร์กับสัญญาณทางเวลาที่ไม่ต่อเนื่อง ซึ่งถูกใช้อย่างแพร่หลายกับงานด้านการประมวลผลสัญญาณ และงานที่เกี่ยวข้องกับการวิเคราะห์แฉกแรงทางความถี่ของสัญญาณที่ได้จากการสุ่มตัวอย่าง เพื่อแก้สมการพาร์เชียลดิฟเฟอเรนเชียล (Partial differential) หรือเพื่อใช้ทำการคอนโวลูชัน (Convolution) เราสามารถแปลงฟูรีเยร์แบบเต็มหน่วยได้อย่างมีประสิทธิภาพ ด้วยการแปลงฟูรีเยร์อย่างรวดเร็ว (Fast Fourier transform - FFT)

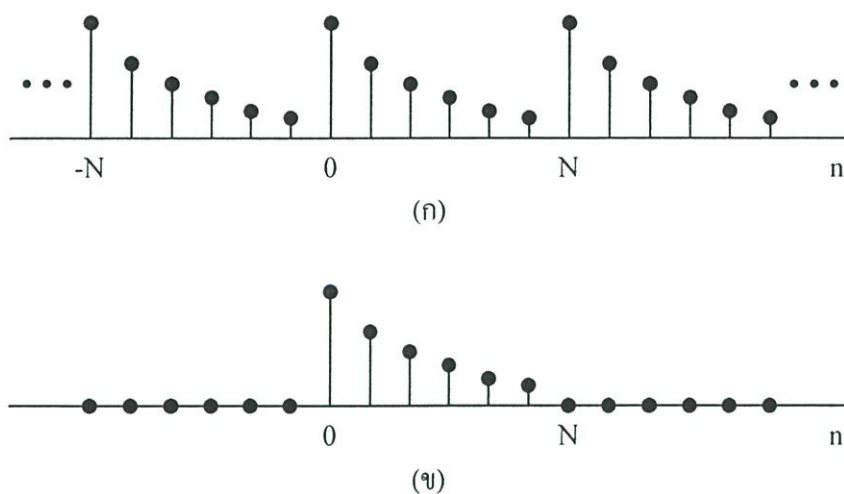
ผลการแปลงฟูรีเยร์แบบเต็มหน่วยของสัญญาณ  $X(n)$  พิจารณาได้จากอนุกรมฟูรีเยร์แบบเต็มหน่วย (discrete Fourier series หรือ DFS) ของสัญญาณเป็นคาบ  $\tilde{X}(n)$  ที่มีคาบเท่ากับ  $N$  แสดงได้ดังสมการ (3.4)

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k) e^{j\left(\frac{2\pi}{N}\right)kn} \quad (3.4)$$

โดยที่  $\tilde{X}(k)$  คือค่าสัมประสิทธิ์ของอนุกรมฟูรีเยร์ ซึ่งคำนวณได้จาก  $\tilde{X}(n)$  ตามความสัมพันธ์ดังสมการ (3.5)

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j\left(\frac{2\pi}{N}\right)kn} \quad (3.5)$$

เมื่อพิจารณา  $X(n)$  ให้เป็นหนึ่งคาบของ  $\tilde{X}(n)$  แสดงตัวอย่างดังรูปที่ 3.2



รูปที่ 3.2 ตัวอย่างของ (ก) สัญญาณ  $\tilde{X}(n)$  และ (ข) สัญญาณ  $X(n)$

ค่าสัมประสิทธิ์ฟูรีเยร์ของ  $X(n)$  จึงมีค่าเท่ากับหนึ่งคาบของ  $\tilde{X}(k)$  แสดงดังสมการ (3.6) ซึ่งเรียกว่า ผลการแปลงฟูรีเยร์เต็มหน่วย หรือ DFT

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\left(\frac{2\pi}{N}\right)kn}, \quad 0 \leq k \leq N-1 \quad (3.6)$$

สเปกตรัมขนาดของ  $X(k)$  จะแสดงองค์ประกอบทางความถี่ของสัญญาณ  $X(n)$  ที่ให้พลังงานสูงสุด หรือเรียกว่า ฟอร์แมนท์ (formant) ในกรณีสัญญาณเสียงดนตรี นำอัลกอริทึมการแปลงฟูรีเยร์อย่างรวดเร็ว (fast Fourier transform หรือ FFT) มาใช้ในการคำนวณ DFT โดยกำหนดให้  $N$  เป็นค่ากำลังของสอง และ  $W_N = e^{-j(2\pi/N)}$  จะได้  $X(k)$  มีค่าตามสมการ (3.7)

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}, \quad 0 \leq k \leq N-1 \quad (3.7)$$

อัลกอริทึม FFT จะทำการแบ่ง  $X(n)$  ออกเป็นสองลำดับเท่าๆกัน คือลำดับของเลขคู่ (แทน  $n = 2r$ ) และลำดับของเลขคี่ (แทน  $n = 2r + 1$ ) โดยที่  $0 \leq r \leq (N/2) - 1$  ดังสมการ (3.8)

$$X(k) = \sum_{r=0}^{(N/2)-1} x(2r) W_N^{2rk} + \sum_{r=0}^{(N/2)-1} x(2r+1) W_N^{(2r+1)k} \quad (3.8)$$

ให้  $W_N^2 = W_{N/2}$  จะได้ DFT จำนวน  $N$  จุด เป็นผลรวมของ DFT จำนวน  $N/2$  จุด สองลำดับ แสดงดังสมการ (3.9)

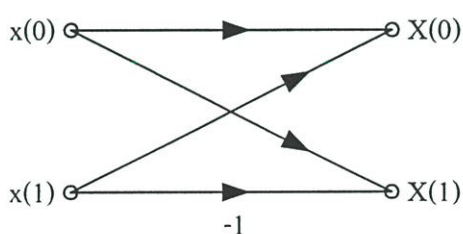
$$X(k) = \sum_{r=0}^{(N/2)-1} x(2r) W_{N/2}^{rk} + W_N^k \sum_{r=0}^{(N/2)-1} x(2r+1) W_{N/2}^{rk} \quad (3.9)$$

ซึ่งแต่ละลำดับสามารถแยกออกเป็นผลรวมของ DFT จำนวน  $N/4$  จุด สองลำดับ จนกระทั่งแยกได้เป็น DFT จำนวน 2 จุด  $N/2$  ลำดับ กระบวนการนี้จะสิ้นสุด การคำนวณ DFT จำนวน 2 จุด แสดงได้ดังสมการ (3.10) และเนื่องจาก  $W_1^{0k} = 1$  และ  $W_2^k = (-1)^k$  จะได้ค่าของ  $X(k)$  แสดงดังสมการ (3.11) และ (3.12) แทนด้วยกราฟที่เรียกว่า บัตเตอร์ฟลาย (butterfly) ได้ดังรูปที่ 3.3 และการคำนวณ DFT จำนวน 4 จุด แสดงดังรูปที่ 3.4

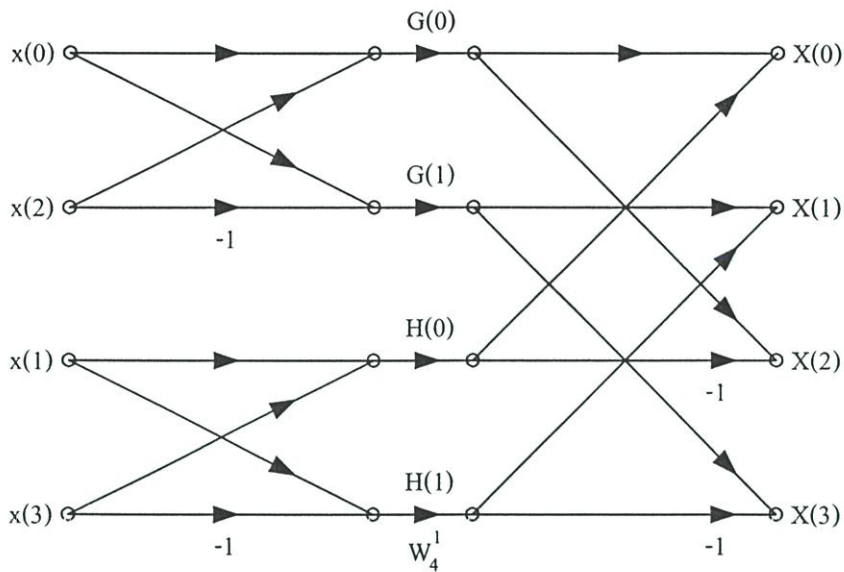
$$X(k) = x(0) W_1^{0k} + W_2^k [x(1) W_1^{0k}] \quad , \quad k = 0, 1 \quad (3.10)$$

$$X(0) = x(0) + x(1) \quad (3.11)$$

$$X(1) = x(0) - x(1) \quad (3.12)$$



รูปที่ 3.3 กราฟของการคำนวณ DFT จำนวน 2 จุด (บัตเตอร์ฟลาย)

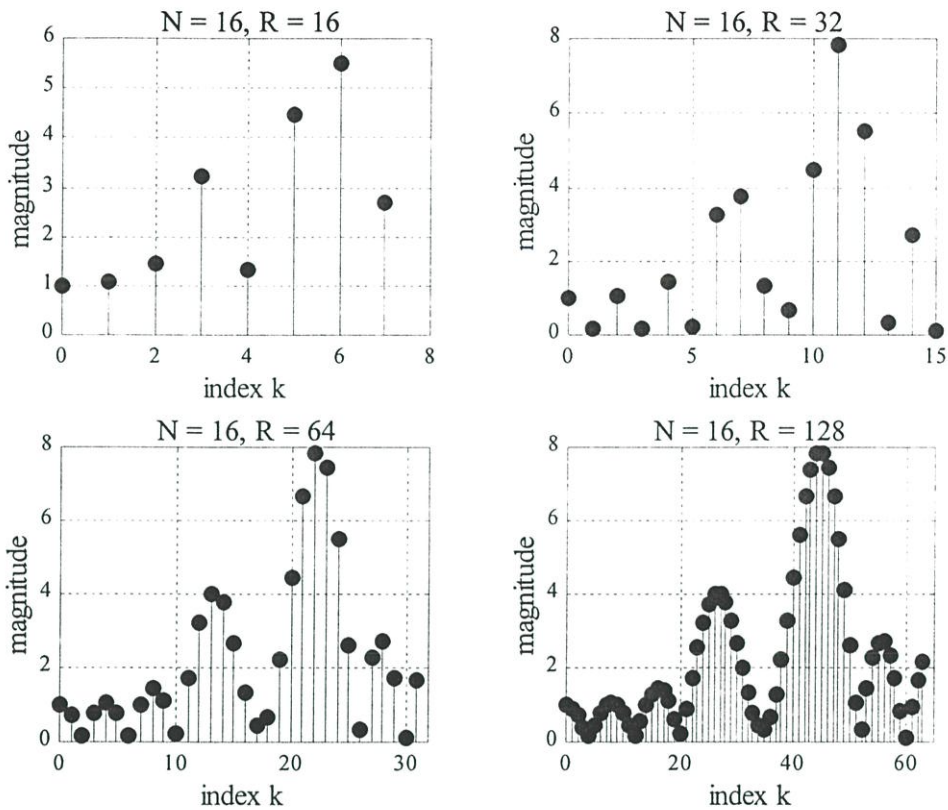


รูปที่ 3.4 กราฟของการคำนวณ DFT จำนวน 4 จุด

สำหรับการวิเคราะห์สัญญาณด้วย DFT การเพิ่มความถูกต้องของตำแหน่งเส้นสเปกตรัม ทำได้โดยเพิ่มขนาดของการคำนวณ DFT ให้เป็น  $R$  จุด ซึ่งจำนวนจุดที่เพิ่มขึ้นจะเป็นค่าศูนย์ ตัวอย่างต่อไปนี้ แสดงการคำนวณ DFT ของสัญญาณไซน์ซอซอด์  $y(n)$  มีค่าตามสมการ (3.13) โดยกำหนดให้  $N = 16$ ,  $f_1 = 0.22$  Hz,  $f_2 = 0.34$  Hz และ  $R$  มีค่าเปลี่ยนแปลงจาก 16 จุด ถึง 128 จุด

$$y(n) = \left(\frac{1}{2}\right) \sin(2\pi f_1 n) + \sin(2\pi f_2 n) , \quad 0 \leq n \leq N-1 \quad (3.13)$$

จากรูปที่ 3.5 จะสังเกตได้ว่า ที่  $R$  เท่ากับ 64 จุด ขนาดสูงสุดปรากฏที่ตำแหน่ง  $k$  เท่ากับ 13 และ 22 ซึ่งตรงกับความถี่ 0.2031 Hz และ 0.3438 Hz ตามลำดับ และที่  $R$  เท่ากับ 128 จุด ขนาดสูงสุดปรากฏที่ตำแหน่ง  $k$  เท่ากับ 27 และ 44 ซึ่งตรงกับความถี่ 0.2109 Hz และ 0.3438 Hz ตามลำดับ นั่นคือเมื่อ  $R$  เพิ่มมากขึ้น สเปกตรัมของ  $y(n)$  ปรากฏขนาดสูงสุดที่สองความถี่อย่างชัดเจน และมีค่าใกล้เคียงกับความถี่จริงของสัญญาณ

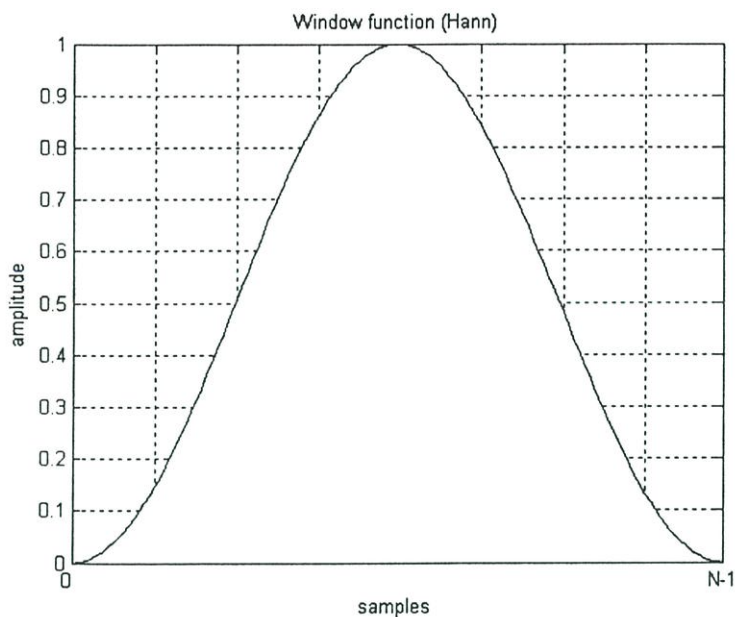


รูปที่ 3.5 ผลการวิเคราะห์สัญญาณ  $y(n)$  ด้วย DFT

### 3.2.1.3 การแปลงฟูรีเยร์ในช่วงเวลาสั้นๆ (Short Time Fourier Transform)

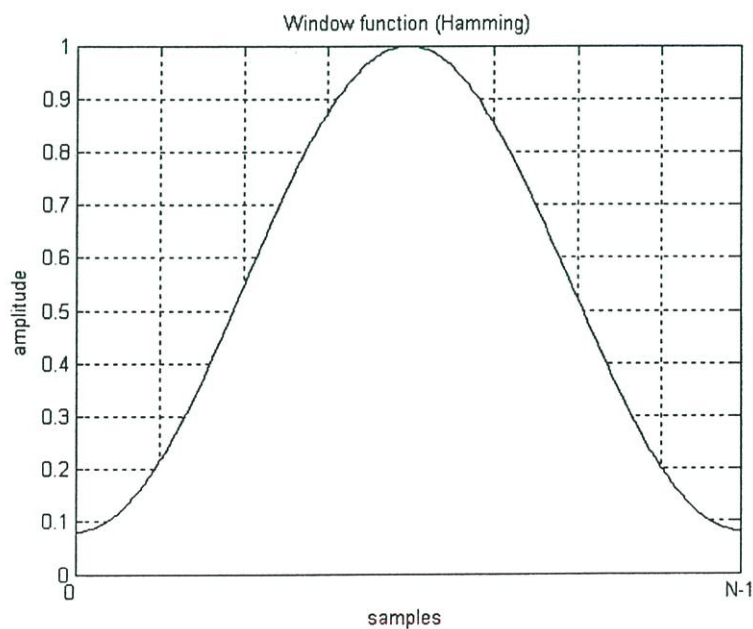
การแปลงฟูรีเยร์ในช่วงเวลาสั้นๆ ก็คือการแปลงฟูรีเยร์กับสัญญาณทางเวลาสั้นๆ ที่ถูกตัดออกมาจากสัญญาณทางเวลาทั้งหมด ซึ่งผ่านการคูณกับหน้าต่างเพื่อปรับข้อมูลเรียบร้อยแล้ว รวมเรียกว่า การทำหน้าต่าง (Windowing) โดยส่วนใหญ่ จะใช้หน้าต่างของฮาน (Hann Window) หรือหน้าต่างแฮมมิง (Hamming Window) ที่มีลักษณะคล้ายกัน

การที่เราตัดสัญญาณออกมาเป็นช่วงๆนั้น เมื่อเราใช้การแปลงฟูรีเยร์ Power Spectrum ที่ได้ จะเกิดความผิดเพี้ยน เนื่องจากปัญหาของขอบของสัญญาณ จะมีการเปลี่ยนแปลงขนาดอย่างรวดเร็ว เหมือนกับการเกิด Impulse ซึ่งให้ผลการตอบสนองความถี่ในทุกย่าน ปัญหานี้เราสามารถแก้ไขได้ โดยการ คูณฟังก์ชันหน้าต่างแบบฮาน หรือแบบแฮมมิง ซึ่งมีฟังก์ชันตามสมการที่ 3.14 และสมการที่ 3.15 ตามลำดับ และมีลักษณะตามรูปที่ 3.6 และรูปที่ 3.7 ตามลำดับ



รูปที่ 3.6 หน้าต่างของฮาน (Hann Window)

$$w[k+1] = 0.5 \left( 1 - \cos \left( 2\pi \frac{k}{n-1} \right) \right), \quad k = 0 \dots n-1 \quad (3.14)$$



รูปที่ 3.7 หน้าต่างของแฮมมิง (Hamming Window)

$$w[k+1] = 0.54 - 0.46 \cos \left( 2\pi \frac{k}{n-1} \right), \quad k = 0 \dots n-1 \quad (3.15)$$

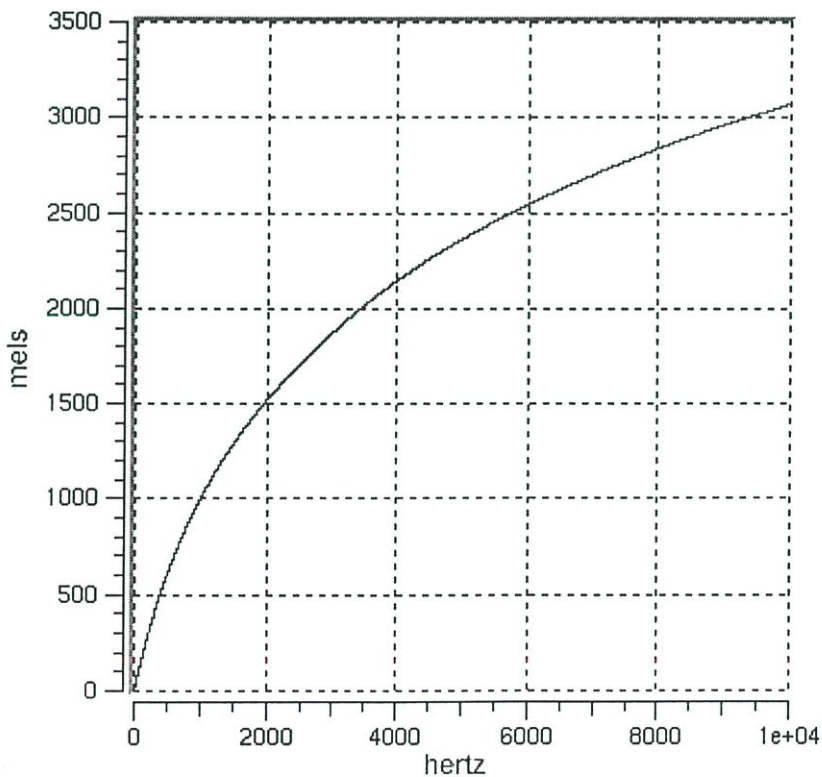
การใช้ฟังก์ชันหน้าต่างแบบนี้ ขอบของข้อมูลจะถูกลดขนาดลง ทำให้ข้อมูลบางส่วนหายไป เราสามารถเช็คข้อมูลเหล่านี้ได้ โดยในการตัดทอนชุดข้อมูลมานั้น ให้ทำร่วมกับข้อมูลชุดเก่า เรียกว่าการซ้อนทับ (Overlapping) ซึ่งจะช่วยให้ข้อมูลยังคงครบถ้วนเหมือนเดิม โดยทั่วไปการซ้อนทับนั้น จะกระทำที่ครึ่งหนึ่งของขนาดชุดข้อมูล หรือที่ 1 ใน 3 ของขนาดชุดข้อมูล

### 3.2.1.4 การปรับสัดส่วนทางความถี่ (Frequency Scaling)

#### 1. การปรับสัดส่วนทางความถี่แบบของเมล (Mel-Frequency Scaling)

เนื่องจากกระบวนการได้ยินของมนุษย์จะมีลักษณะไม่เป็นเชิงเส้น จึงต้องมีการทำ Frequency Scaling เป็นการเปลี่ยนสเกลความถี่เชิงเส้นให้เป็นความถี่ในสเกลที่เหมาะสมกับการได้ยินของมนุษย์ นั่นคือความถี่ในสเกลแบบเมล (Mel Frequency Scale) โดยในช่วงความถี่ต่ำกว่า 1 kHz จะมีการตอบสนองทางความถี่เป็นเชิงเส้น แต่ช่วงความถี่ที่สูงกว่านั้นจะเป็นแบบ logarithm โดยความถี่เมลกับความถี่เชิงเส้นมีความสัมพันธ์กันตามสมการที่ (3.16)

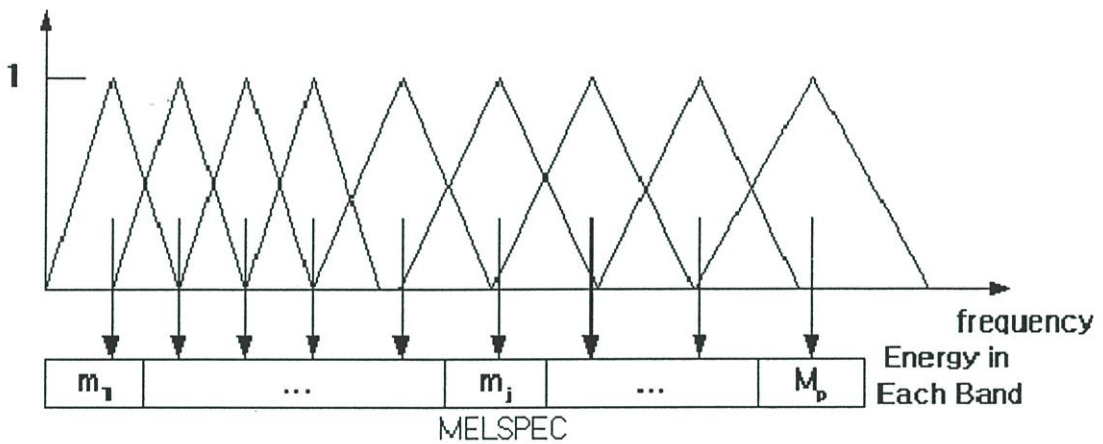
$$\text{Mel} = 2595 \log(1 + \text{freq}/700) \quad (3.16)$$



รูปที่ 3.8 กราฟความสัมพันธ์ระหว่างความถี่เชิงเส้น (Hertz) กับความถี่เมล (Mel)

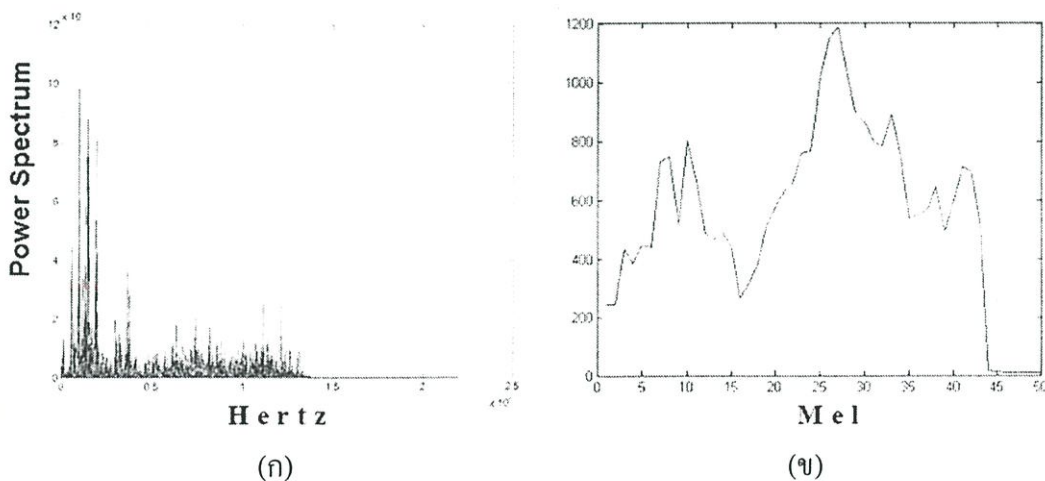
ขั้นตอนการปรับสเกลความถี่แบบเมล คือหลังจากนำข้อมูลเสียงผ่านการแปลง Discrete Fourier Transform (DFT) จะได้ข้อมูลกำลังงานทางความถี่ (Power Spectrum) ของสัญญาณเสียงนั้น แล้วจึงนำมาปรับสเกลทางความถี่โดยการนำ Power Spectrum นั้นผ่านการกรองโดยใช้ชุดการกรองความถี่แบบเมล (Mel-Frequency Filter Bank) ซึ่งเป็นกลุ่มของ Band Pass filter ที่ใช้ช่วงความถี่ที่คำนวณจากสมการที่ (3.16)

การกรองความถี่กับสัญญาณที่อยู่ในโดเมนความถี่อยู่แล้วนั้น ก็คือการนำหน้าต่างของการตอบสนองทางความถี่ ของการกรองนั้นๆ มาคูณเข้ากับ Power Spectrum ของสัญญาณนั้นนั่นเอง ซึ่งการกรองความถี่แบบ Band pass filter นี้เทียบได้กับการใช้ หน้าต่างสามเหลี่ยม (Triangular window) ที่มีลักษณะการสเกลแบบเมลดังรูป 3.9



รูปที่ 3.9 Mel –Frequency Filter Bank

ตัวอย่างผลของการปรับสัดส่วนทางความถี่แบบเมล เป็นดังรูปที่ 3.10



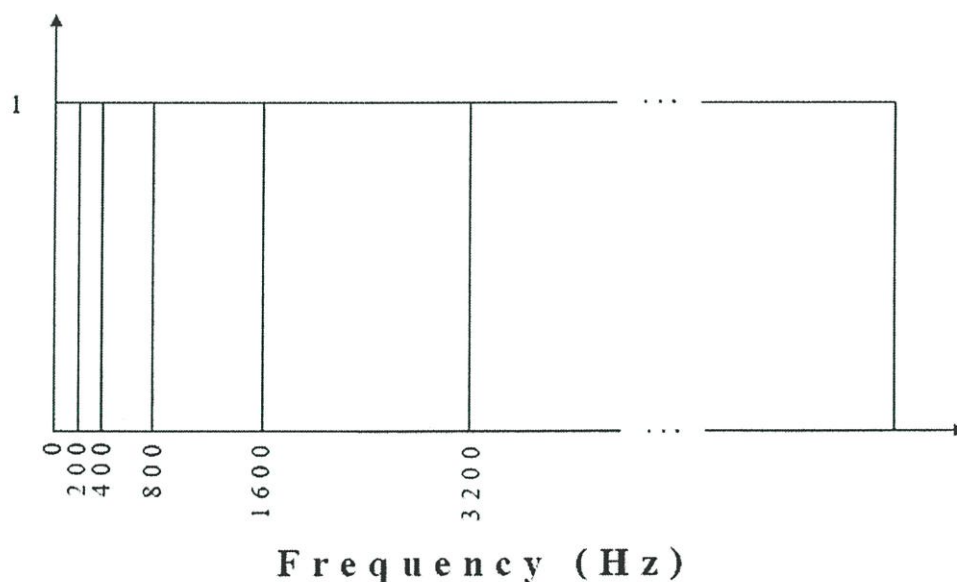
รูปที่ 3.10 (ก) พาวเวอร์สเปกตรัมในหน่วยเฮิรซ์ (ข) พาวเวอร์สเปกตรัมในหน่วยเมล

จะเห็นได้ว่า การใช้ Mel-Frequency Scaling นั้น นอกจากปรับสัดส่วนทางความถี่แล้ว ยังสามารถลดจำนวนข้อมูลลงได้อีกด้วย เนื่องจากการคูณเวกเตอร์ข้อมูลเข้ากับหน้าต่างสามเหลี่ยม นั้น เป็นการหาตัวแทนข้อมูลประเภทหนึ่ง

## 2. การปรับสัดส่วนทางความถี่บนพื้นฐานเสียงคู่แปด (Octave-based Frequency Scaling)

ถึงแม้ว่าการปรับสัดส่วนตามความถี่ของเมล จะเหมาะกับการประมวลผลทางเสียงโดยทั่วไป แต่การปรับสัดส่วนแบบเสียงคู่แปดนั้นเหมาะกับการประมวลผลทางดนตรีมากกว่า โดยมีการแบ่งย่านความถี่ออกเป็น 0Hz-200 Hz, 200 Hz- 400 Hz, 400 Hz-800Hz, 800 Hz -1,600 Hz, 1,600-3,200 Hz, 3,200 Hz -6,400 Hz, 6,400 Hz -12,800 Hz, 12,800 Hz -22,050 Hz

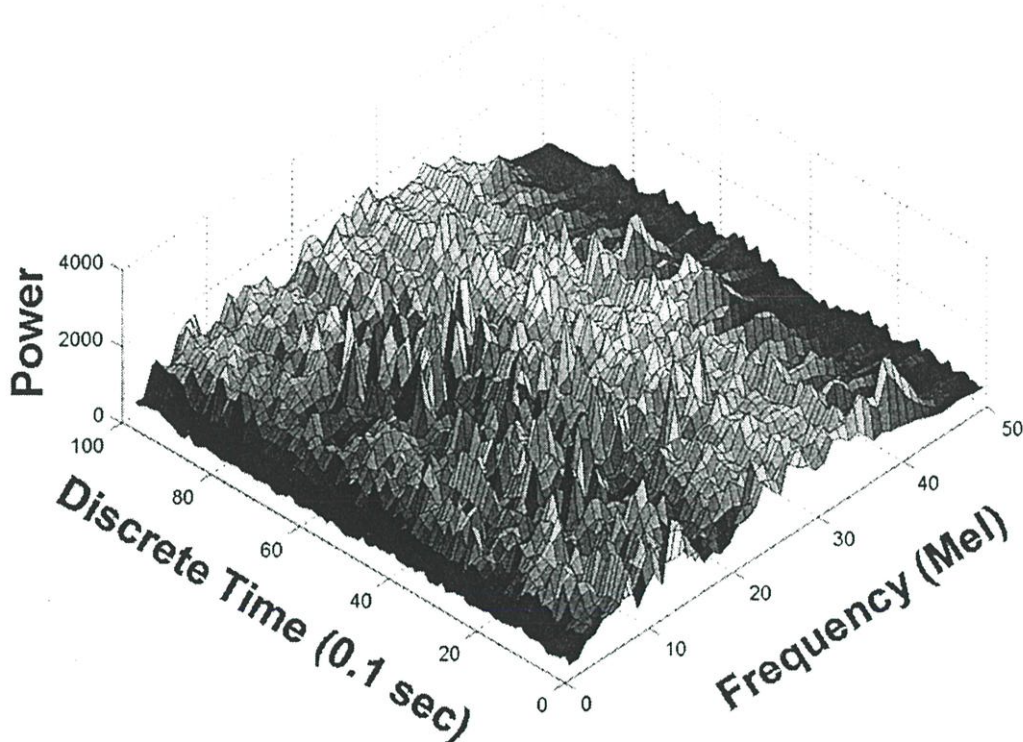
ขั้นตอนการทำงาน คือใช้หน้าต่างสี่เหลี่ยม (Rectangular Window) แบบไม่มีการซ้อนทับ (Overlap) คูณเข้ากับ Power Spectrum ซึ่งก็คือการหาค่าเฉลี่ยของสัญญาณในแต่ละช่วงนั่นเอง การปรับสัดส่วนแบบนี้ ช่วงของการปรับ ตั้งอยู่บนการคำนวณแบบลอการิทึมฐาน 2 ซึ่งต่างจากการปรับสัดส่วนของเมลที่เป็นลอการิทึมฐาน 10 ดังแสดงในรูปที่ 3.11



รูปที่ 3.11 ชุดของหน้าต่างสี่เหลี่ยมที่แบ่งช่วงด้วยสเกลลอการิทึมฐาน 2

### 3.2.2 การดึงลักษณะเด่น (Feature Extraction)

ลักษณะเด่นที่ใช้ในงานวิจัยนี้ เป็นลักษณะเด่นทางความถี่ซึ่งดึงมาจาก ข้อมูลกำลังงานทางความถี่ (Power Spectrum) ที่ได้จากการประมวลผลฟูริเยร์ในช่วงเวลาสั้นๆ มีลักษณะดังแสดงในรูปที่ 3.12 ซึ่งลักษณะเด่นทางความถี่ที่ผู้ทำวิจัยศึกษาและทำการทดลองเปรียบเทียบนั้น มี 3 ลักษณะเด่น ดังต่อไปนี้



รูปที่ 3.12 ตัวอย่างข้อมูลกำลังงานทางความถี่ที่ได้จากการแปลงฟูริเยร์ในช่วงเวลาสั้นๆ

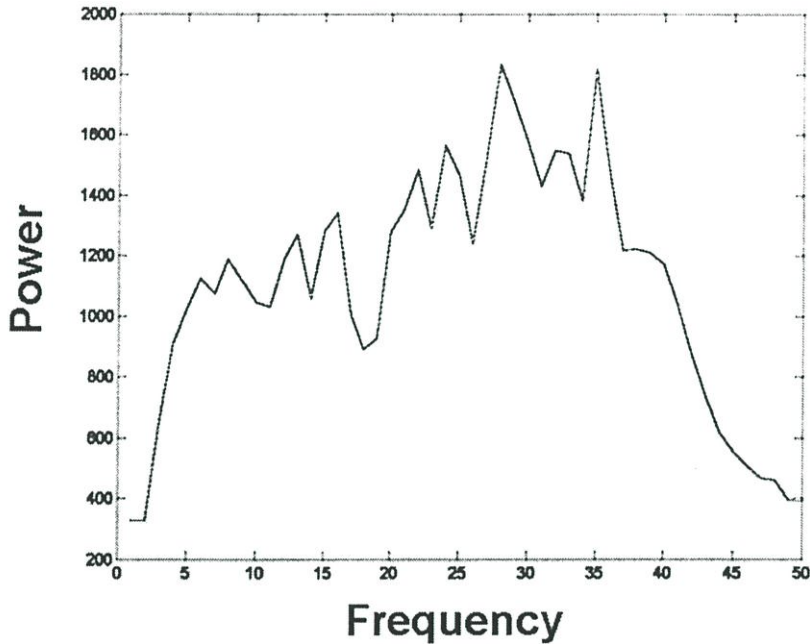
#### 3.2.2.1 ค่าเฉลี่ยทางความถี่ (Spectral Mean)

หลังจากที่ได้ข้อมูลกำลังงานทางความถี่ของเสียงคนตรีในช่วงเวลาต่างๆ จากการทำ Short-Time Fourier Transform แล้ว การดึงลักษณะเด่นที่ทำได้ง่ายมากอย่างหนึ่งก็คือการหาค่าเฉลี่ยของข้อมูล ซึ่งเป็นตัวแทนข้อมูลที่ดีเมื่อข้อมูลทั้งหมดมีค่าใกล้เคียงกัน การคำนวณค่าเฉลี่ยนั้นจะต้องทำแยกกันในแต่ละย่านความถี่ ดังสมการ (3.17)

$$\bar{X}_k = \frac{1}{N} \sum_{i=1}^N X_{k,i} \quad (3.17)$$

โดยที่  $\bar{X}_k$  คือค่าเฉลี่ยของ Power Spectrum ( $X_k$ ) ที่ย่านความถี่ที่  $k$  และ  $N$  คือจำนวน Power Spectrum ทั้งหมดในย่านความถี่ที่  $k$

หลังจากหาค่าเฉลี่ยทางความถี่แล้วเราจะได้ เวกเตอร์ของ Power Spectrum 1 ชุด ที่เป็นตัวแทนของ Power Spectrum ทั้งหมด ดังแสดงในรูปที่ 3.13 ซึ่งสามารถบอกโทนเสียงของเพลงแต่ละแนวเพลงได้

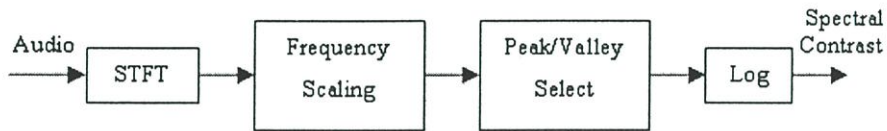


รูปที่ 3.13 ตัวอย่างค่าเฉลี่ยทางความถี่ (Spectral Mean)

### 3.2.2.2 ความตัดกันทางความถี่ (Spectral Contrast)

ในงานวิจัยของ Dan-Ning Jiang และคณะ [9] ได้ทำเรื่องการแยกประเภทดนตรี โดยใช้ลักษณะเด่น Octave -Based Spectral Contrast ซึ่งแสดงถึงความสัมพันธ์ของคุณลักษณะทางความถี่ของดนตรี ในงานวิจัยนี้ได้ใช้ตัวแยกประเภท เกาเขียนแบบผสม (Gaussian Mixture Model) เปรียบเทียบลักษณะเด่น Spectral Contrast กับ Mel -Frequency Cepstral Coefficient (MFCC) โดยใช้ข้อมูลเพลง ทั้งหมด 1500 เพลง แบ่งเป็นชุดข้อมูลฝึกสอน 1250 เพลงและชุดข้อมูลทดสอบ 250 เพลง แยกดนตรีออกเป็น 5 ประเภท ได้แก่ Baroque, Romantic, Pop, Jazz และ Rock ผลการแยกประเภทโดยใช้ลักษณะเด่น Spectral Contrast นั้นได้ ความถูกต้องเฉลี่ย 82.3 % และ MFCC ได้ความถูกต้องเฉลี่ย 78 % จะเห็นได้ว่า ลักษณะเด่น Spectral Contrast ให้ผลการแยกประเภทที่ดีกว่า ลักษณะเด่น MFCC ที่เป็นที่แพร่หลายมาก ข้อดีของลักษณะเด่น Spectral Contrast คือ ได้เพิ่มความสัมพันธ์ระหว่าง Power Spectrum สูงสุดและ Power Spectrum ต่ำสุดเข้าไปด้วยในขณะที่ MFCC สนใจข้อมูลทางความถี่เพียงอย่างเดียว

Spectral Contrast หาได้จากความแตกต่างระหว่างค่าสูงสุด (Spectral Peak) ของ Power Spectrum และค่าต่ำสุดของ Power Spectrum (Spectral Valley) โดยที่ความถี่ที่มี Spectral Peak ค่าสูงๆ ส่วนใหญ่จะเป็นความถี่ของเสียงที่ใช้บ่อยๆ ในเพลง ส่วนความถี่เสียงที่ถูกใช้น้อยและพวก สัญญาณรบกวน จะปรากฏอยู่ที่ค่าต่ำสุดทางความถี่ ลักษณะเด่นของ Spectral Contrast สามารถ แสดงให้เห็นถึงการกระจาย ของ Power Spectrum ในความถี่นั้นๆ ซึ่ง ดีกว่าการเฉลี่ย Power Spectrum ทั้งหมดเข้าด้วยกัน เพราะจะทำให้ข้อมูลการกระจายนี้หายไป



รูปที่ 3.14 กระบวนการหาค่าความตัดกันทางความถี่ (Spectral Contrast)

ขั้นแรกใช้ Short-Time Fourier Transform เพื่อแปลงสัญญาณทางเวลาให้เป็นสัญญาณทางความถี่ ก่อนโดยให้  $\{X_{k,1}, X_{k,2}, \dots, X_{k,N}\}$  เป็นเวกเตอร์ของ Power Spectrum ที่ย่านความถี่  $k$  จากนั้น เรียงลำดับเวกเตอร์นี้ใหม่ โดยเรียงจากมากไปหาน้อย จะได้เวกเตอร์ใหม่ เป็น  $\{X'_{k,1}, X'_{k,2}, \dots, X'_{k,N}\}$  โดยที่  $X'_{k,1} > X'_{k,2} > \dots > X'_{k,N}$  จากนั้นหาค่าความแข็งแรงของ Spectrum สูงสุดและ Spectrum ต่ำสุดได้ โดย

$$\text{Peak}_k = \log \left\{ \frac{1}{\alpha N} \sum_{i=1}^{\alpha N} X'_{k,i} \right\} \quad (3.18)$$

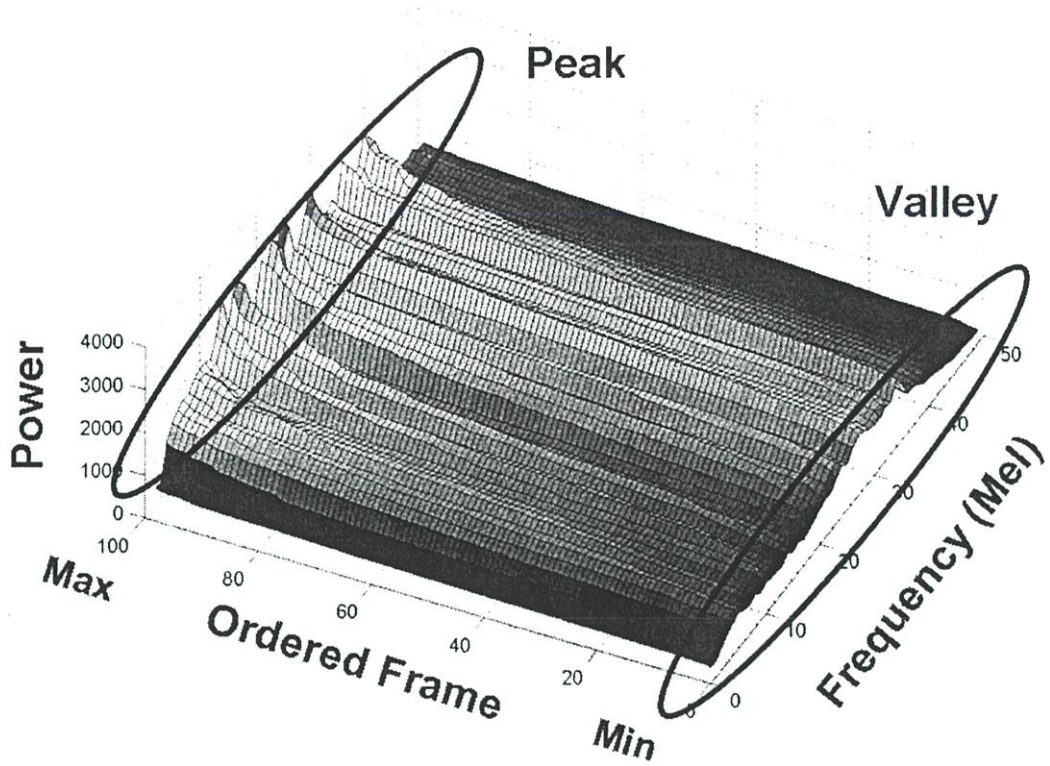
$$\text{Valley}_k = \log \left\{ \frac{1}{\alpha N} \sum_{i=1}^{\alpha N} X'_{k,N-i+1} \right\} \quad (3.19)$$

และค่าต่าง คือ

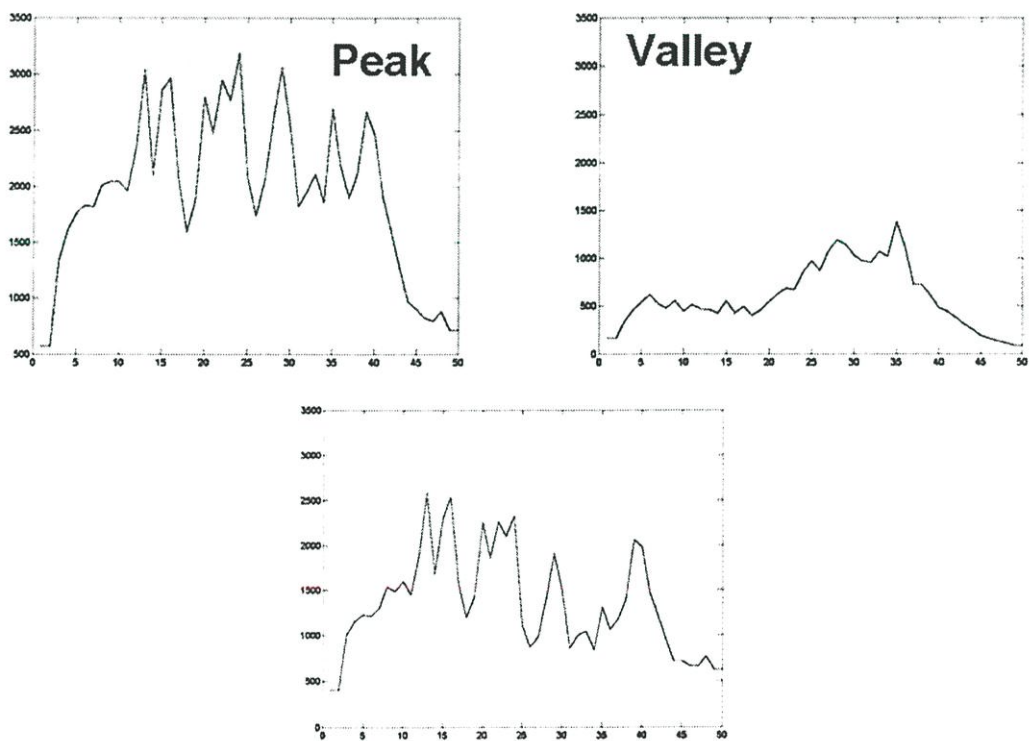
$$\text{SC}_k = \text{Peak}_k - \text{Valley}_k \quad (3.20)$$

เมื่อ  $N$  คือจำนวน Spectrum ทั้งหมดที่ย่านความถี่ย่อย  $k$  และ  $\alpha$  มีค่าเท่ากับ 0.02

ลักษณะของค่าสูงสุดทางความถี่และค่าต่ำสุดทางความถี่หลังจากเรียงลำดับ Power Spectrum แล้ว แสดงดังรูปที่ 3.15 และลักษณะของลักษณะเด่นความตัดกันทางความถี่ แสดงในรูปที่ 3.16



รูปที่ 3.15 ตัวอย่างข้อมูลกำลังงานทางความถี่ที่ถูกเรียงลำดับแล้ว, ค่าสูงสุดและต่ำสุดทางความถี่



**Spectral Contrast = Peak - Valley**

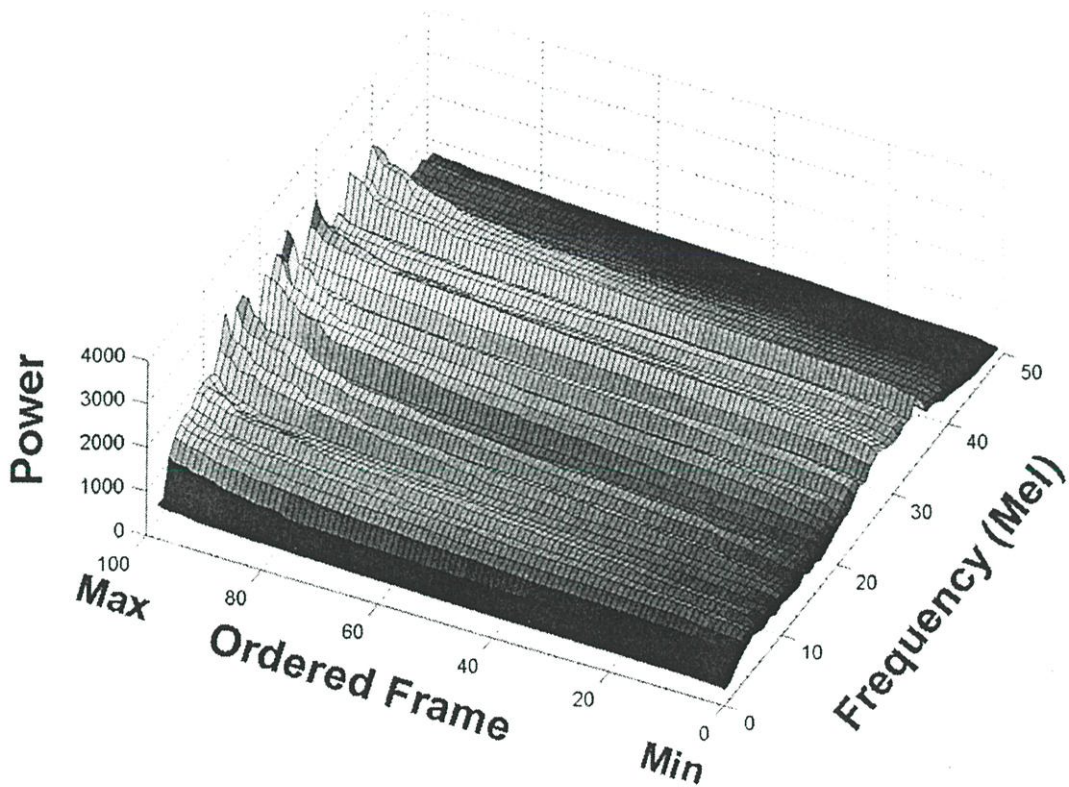
รูปที่ 3.16 ตัวอย่างลักษณะเด่นความตักันทางความถี่

### 3.2.2.3 ฮิสโทแกรมทางความถี่ (Spectral Histogram)

จากลักษณะเด่นของ Spectral Contrast แสดงให้เห็นถึงความสัมพันธ์ระหว่าง Power Spectrum ที่มีค่ามากที่สุดและ Power Spectrum ที่มีค่าน้อยที่สุด โดยการเรียงลำดับ Power Spectrum ทั้งหมดจากมากไปน้อย ในแต่ละย่านความถี่อย่างอิสระต่อกัน

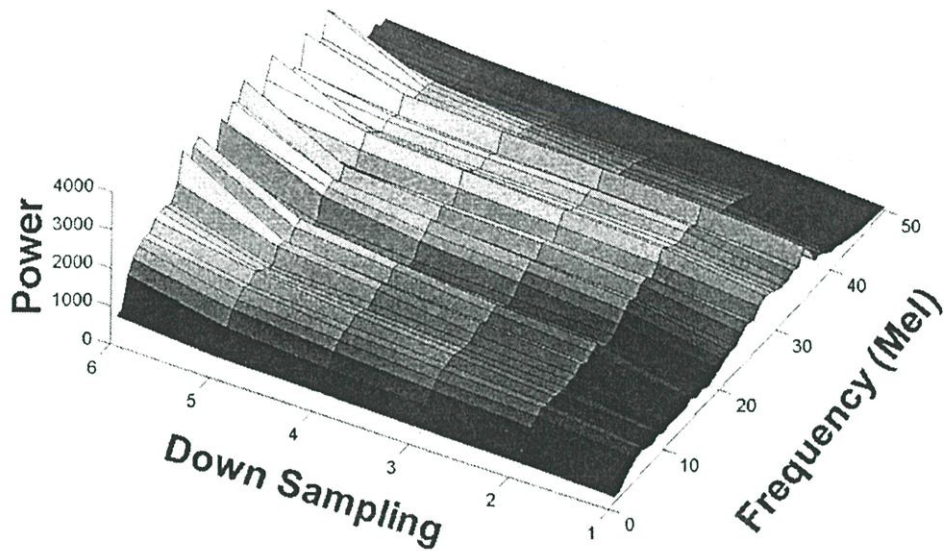
การเรียงลำดับ Power Spectrum นี้ทำให้เห็นถึงลักษณะการกระจายและการใช้เสียงในแต่ละความถี่ของเพลงแต่ละประเภท ผู้ทำวิจัย จึงสนใจในลักษณะเด่นนี้และพยายามพัฒนาต่อยอดจนได้ลักษณะเด่นใหม่คือ Spectral Histogram

ขั้นตอนการทำ Spectral Histogram นั้นเหมือนกับ Spectral Contrast ตรงที่ต้องนำ Power Spectrum ทั้งหมดมาเรียงลำดับจากมากไปน้อย แยกกันในแต่ละย่านความถี่ ให้  $\{X_{k,1}, X_{k,2}, \dots, X_{k,N}\}$  เป็นค่า Power spectrum ที่เรียงลำดับแล้วในย่านความถี่ที่  $k$  ดังรูปที่ 3.17



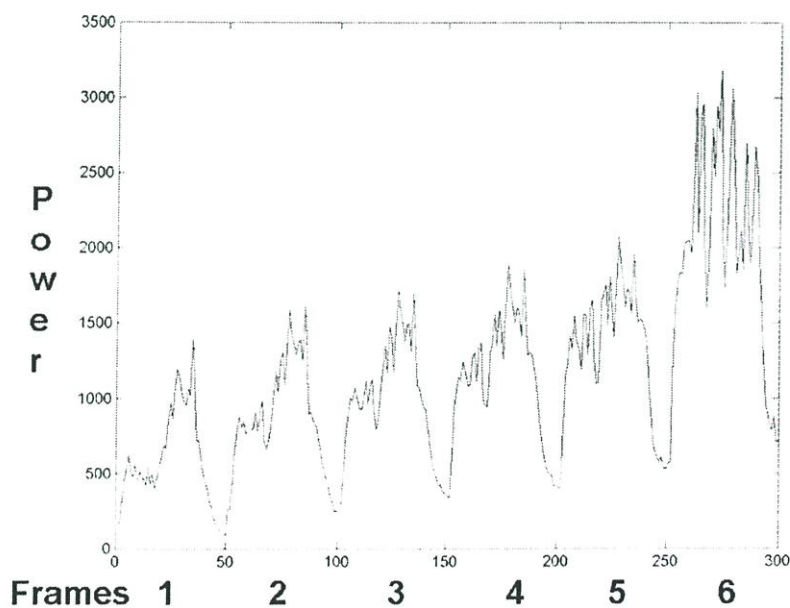
รูปที่ 3.17 ตัวอย่างกำลังงานทางความถี่ที่ถูกเรียงลำดับแล้ว

จากนั้นลดจำนวนข้อมูลทางแกนปริมาณลงด้วยการเก็บตัวอย่าง(Down Sampling) จะเหลือข้อมูลอย่างหายาดังรูปที่ 3.18

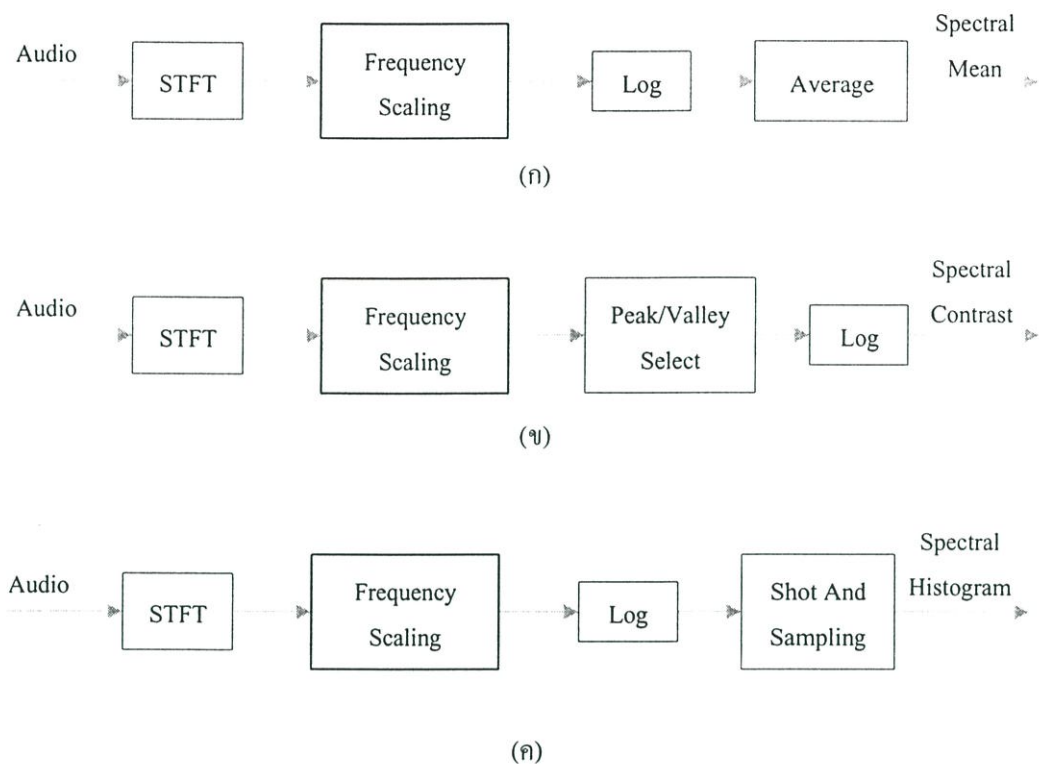


รูปที่ 3.18 ตัวอย่างกำลังงานทางความถี่ที่ถูกลดจำนวนตัวอย่าง (Down sampling) แล้ว

เมื่อได้ข้อมูลจากการเก็บตัวอย่างมาแล้วแต่ข้อมูลที่ได้ยังเป็น 3 มิติอยู่ ดังนั้นเพื่อให้สามารถนำข้อมูลมาใช้กับโครงข่ายประสาทเทียมได้ เราจึงต้องเรียงข้อมูลใหม่ให้อยู่ในรูป 2 มิติ โดยการนำชุดของ Spectrum ที่ได้จากการเก็บตัวอย่างแล้ว มาเรียงต่อกันจะได้ลักษณะเด่น ซีสโทแกรมทางความถี่ ดังรูปที่ 3.19



รูปที่ 3.19 ตัวอย่างลักษณะเด่นซีสโทแกรมทางความถี่



รูปที่ 3.20 เปรียบเทียบขั้นตอนการเตรียมข้อมูลลักษณะเด่น (ก) ค่าเฉลี่ยทางความถี่ (ข) ค่าความตักกันทางความถี่ (ค) ฮิสโทแกรมทางความถี่

จากรูปที่ 3.20 เป็นขั้นตอนการเตรียมข้อมูลลักษณะเด่นทั้ง 3 แบบ ที่ใช้ในการทดลองนี้

## บทที่ 4

# โครงข่ายประสาทเทียม

ในปัจจุบันคอมพิวเตอร์ได้ถูกนำมาใช้งานอย่างกว้างขวางเกือบจะทุกด้านก็ว่าได้ แต่ยังคงมีขีดความสามารถจำกัดอยู่เมื่อเทียบกับสมองมนุษย์ ซึ่งสมองมนุษย์สามารถจดจำและเรียนรู้จากประสบการณ์ในอดีตและสามารถปรับใช้กับสถานการณ์ปัจจุบันได้ ดังนั้นจึงได้มีการศึกษาการทำงานของสมอง เพื่อนำมาเป็นแบบจำลองของเซลล์ประสาท (Artificial Neural) โครงข่ายประสาทเทียม (Artificial Neural Network) และศึกษากระบวนการการเรียนรู้ (Learning Algorithm) เพื่อนำมาประยุกต์ใช้กับคอมพิวเตอร์ ซึ่งการนำทฤษฎีต่างๆของโครงข่ายประสาทเทียมที่ผ่านการพัฒนามาหลายสิบปี มาประยุกต์ใช้กับงานจริงมีอย่างมากมายหลายด้าน เช่น โทรมนาคมนำมาใช้ในการบีบข้อมูลภาพ (Image and Data Compression) ระบบนักบินอัตโนมัติ (Auto Pilot Aircraft) การแปลภาษาพูด (Real Time Translation of spoken language) การจดจำเสียง (Voice recognition) การจดจำภาพ (Image recognition) การจดจำรูปแบบ (Pattern recognition) ซึ่งจะจดจำแบบเดิมและจำรูปภาพนั้นเมื่อพบอีกครั้ง นอกจากนี้ยังมีการใช้งานแบบอื่นๆ เช่น การค้นหาเป้าหมาย (Trajectory Control) การวิเคราะห์การตลาด (Market Analysis) ฯลฯ

### 4.1 ความเป็นมาของโครงข่ายประสาทเทียม

เป็นความพยายามที่จะทำความเข้าใจการทำงานของสมองมนุษย์ได้มีมานานนับศตวรรษแล้ว เริ่มต้นใจปี ค.ศ. 1890 William James นักสรีรศาสตร์ชาวอเมริกันได้แต่งหนังสือชื่อ Psychology (Briefer Course) ซึ่งอธิบายถึงการทำงานของสมอง และยังคงกล่าวถึงทฤษฎีที่มีอยู่ในปัจจุบันไว้ล่วงหน้าอีกด้วย

#### 4.1.1 ช่วงทศวรรษที่ 40 ยุคเริ่มต้นของโครงข่ายประสาทเทียม

ในปี ค.ศ. 1936 Alan Turing ได้นำสมองไปใช้เป็นแบบอย่างของการคำนวณใน A way of looking at the world of computing ในปี ค.ศ. 1943 Warren McCulloch นักประสาทวิทยา และ Walter Pitts นักคณิตศาสตร์ ที่มีอายุเพียง 18 ปี ในขณะนั้นได้ออกแบบและเป็นที่สนใจโดยทั่วไปว่าเป็นโครงข่ายประสาทเทียมชนิดแรก นักวิจัยเหล่านั้นยอมรับว่า การรวมหลายๆ นิวรอนไม่ซับซ้อนเข้าสู่ระบบโครงข่ายให้เป็นการเพิ่มกำลังในการคำนวณค่าถ่วงน้ำหนักของนิวรอน McCulloch-Pitts ได้กำหนดให้มีค่าคงที่ ดังนั้นนิวรอน McCulloch-Pitts ประกอบด้วยชั้นของโครงข่ายจำนวน 2 ชั้น คือ ชั้นอินพุตและเอาต์พุต โครงข่ายจะมีการเชื่อมโยงจากชั้นอินพุตไปยังชั้นเอาต์พุตโดยผ่านน้ำหนัก (weight) ความคิดของค่าขีดจำกัด (threshold) กล่าวคือ ถ้าโครงข่าย

อินพุตถึงนิวรอนมีค่าใหญ่กว่าค่าขีดจำกัดแล้ว โหนดกระตุ้นเป็นลักษณะอย่างหนึ่งของนิวรอน McCulloch-Pitts ต่อมาในปี ค.ศ. 1949 Donald Habb นักจิตวิทยาของมหาวิทยาลัย McGill ได้ออกแบบกฎการเรียนรู้ชนิดแรกสำหรับโครงข่ายประสาทเทียม กฎการเรียนรู้นี้ ก็คือ ถ้าสองนิวรอนถูกกระตุ้นพร้อมกันแล้วความแข็งแรงของการเชื่อมโยงระหว่างพวกมันควรจะเพิ่มขึ้น และได้แต่งหนังสือชื่อ The organization of behavior ซึ่งกล่าวถึงการเชื่อมโยงกันของจิตวิทยาและสรีรศาสตร์

#### 4.1.2 ช่วงทศวรรษที่ 50-60 ยุคทองของโครงข่ายประสาทเทียม

การพัฒนาด้านฮาร์ดแวร์และซอฟต์แวร์ ในช่วงทศวรรษที่ 1950 ทำให้การวิจัยด้านต่างๆ เข้าสู่ยุคของการสร้างแบบจำลองด้วยคอมพิวเตอร์ เป็นผลให้การทำสอนทฤษฎีที่เกี่ยวข้องกับโครงข่ายประสาทเทียมมีความเป็นไปได้มากยิ่งขึ้น ในปีค.ศ. 1956 Dartmouth Summer ได้เริ่มทำงานวิจัยด้านปัญญาประดิษฐ์ (Artificial intelligence :AI) ซึ่งเป็นแรงผลักดันให้เกิดการศึกษาทั้งในด้านปัญญาประดิษฐ์และโครงข่ายประสาทมากขึ้น ในปีค.ศ. 1958 John Von Neumann เขียนหนังสือเรื่อง The Computer and the Brain ซึ่งได้นำเสนอฟังก์ชันการทำงานของเซลล์ประสาทโดยใช้รีเลย์ที่ใช้สำหรับส่งโทรเลขและหลอดสุญญากาศ (Vacuum Tube) และ

ในปี ค.ศ. 1957 Frank Rosenblatt ได้เริ่มศึกษาและพัฒนาประเภทใหญ่ของโครงข่ายประสาทเทียมที่เรียกว่า Perceptron ซึ่งเป็นรูปแบบที่สำคัญรูปแบบหนึ่งของโครงข่ายประสาทเทียม โดยส่วนมากสถาปัตยกรรมของโครงข่ายประสาทเทียมชนิดนี้จะประกอบด้วยชั้นอินพุตเชื่อมโยงผ่านด้วยค่าถ่วงน้ำหนัก เพื่อเป็นตัวเชื่อมกับนิวรอนที่ชั้นเอาต์พุต ค่าถ่วงน้ำหนักบนทางผ่านเชื่อมโยงถูกปรับเปลี่ยนกฎการเรียนรู้ Perceptron ปรับเปลี่ยนค่าถ่วงน้ำหนักซ้ำๆ กัน ซึ่งมีประสิทธิภาพดีกว่ากฎของ Habb การเรียนรู้ของ Perceptron สามารถแสดงให้เห็นว่าการลู่เข้าสู่ค่าถ่วงน้ำหนักที่ถูกต้อง ค่าถ่วงน้ำหนักเหล่านั้นจะแก้ปัญหาก็เตรียมพร้อม และต่อมานางของ Frank Rosenblatt ได้อธิบาย Perceptron หลายๆ ชนิดว่า เหมือนนิวรอนได้พัฒนาจาก McCulloch และ Pitts รวมถึง Habb ด้วย โครงข่ายประสาทเทียมชนิด Perceptron ใช้ค่าเริ่มต้นฟังก์ชันเอาต์พุต

ในปี ค.ศ. 1959 Bernard Widrow และนักศึกษาของเขา Marcian Hoff ได้พัฒนาการเรียนรู้ซึ่งใกล้เคียงกับกฎการเรียนรู้ของ Perceptron โดยการเรียนรู้ของ Perceptron ปรับค่าถ่วงน้ำหนักที่เชื่อมโยงที่โหนด เมื่อผลตอบสนองของโหนดไม่ถูกต้อง กฎ Delta ปรับค่าถ่วงน้ำหนักเพื่อลดความแตกต่างระหว่างโครงข่ายอินพุตถึงโหนดเอาต์พุตและค่าเอาต์พุตที่ต้องการ สิ่งเหล่านี้จะสรุปในรูปของค่าผิดพลาดยกกำลังสอง เหล็กที่น้อยที่สุด (Smallest Mean squared Error: SMS) และนำสิ่งต่างๆ เหล่านี้ไปใช้สร้างแบบจำลองของ MADALINE (Multiple ADaptive LINear Element) ซึ่งพัฒนามาจาก ADLINE และเป็นการนำเอาโครงข่ายประสาทมาใช้งานในทางปฏิบัติครั้งแรก โดยนำมาใช้ในการลดเสียงก้องในสายโทรศัพท์

#### 4.1.3 ช่วงทศวรรษที่ 70 ยุคเงียบของโครงข่ายประสาทเทียม

ในปี ค.ศ. 1969 Marvin Minsky และ Seymour Papert ได้ตีหนังสือ Perceptron ซึ่งกล่าวถึงข้อจำกัดของ Perceptron ว่าไม่สามารถนำไปใช้ในการแก้ปัญหาหลายๆปัญหาที่มีความน่าสนใจได้ ซึ่งจุดนี้เองทำให้นักวิจัยหลายคนกลับไปให้ความสนใจด้านปัญญาประดิษฐ์ (AI) มากขึ้น แต่ยังมีนักวิจัยบางกลุ่มที่ไม่ย่อท้อและยังคงทำงานวิจัยด้านโครงข่ายประสาทต่อไป John Anderson ของมหาวิทยาลัย Brown เริ่มวิจัยงานของเขาในปี ค.ศ. 1968 โดยทำการวิจัยโครงข่ายประสาทเทียมด้วยโครงข่ายรวมกลุ่มความสามารถในการจดจำ (Associative Memory Nets) เขาได้พัฒนาแบบจำลองของโครงข่ายประสาทขึ้นมา และให้ชื่อว่า Brain-State-in-A-Box (BSB) งานประยุกต์ที่ใช้โครงข่ายประสาทเทียมชนิดนี้คือ การวินิจฉัยทางการแพทย์ (Medical Diagnosis)

ในปี ค.ศ. 1972 งานเริ่มต้นของ Teuvo Kohonen ของมหาวิทยาลัยเทคโนโลยี Helsinki วิศวกรไฟฟ้าชาวฟินแลนด์ได้จัดการกับโครงข่ายประสาทรวมกลุ่มความสามารถในการจดจำ และต่อมาได้พัฒนากลายเป็น Self Organizing Map ขึ้นมา สถาปัตยกรรมของโครงข่ายประสาทเทียมชนิด Self Organizing Map นี้ประกอบด้วยจำนวน 2 ชั้น คือ ชั้นอินพุต และ เอาต์พุต การเชื่อมโยงที่ชั้นอินพุตไปยังนิวรอนที่ชั้นเอาต์พุตนั้นจะต่อผ่านน้ำหนัก ข้อแตกต่างจากโครงข่ายอื่น คือ กระบวนการเรียนรู้ที่สามารถจัดกลุ่มข้อมูลอินพุตได้ด้วยตัวของมันเองและเรียกการเรียนรู้ว่าการเรียนรู้แบบคอมแพติทิฟ (Competitive Learning) โครงข่ายประสาทเทียมชนิดนี้ถูกนำไปใช้เกี่ยวกับการจดจำคำพูด (Speech Recognition), การแก้ปัญหาของ “Traveling Salesman Problem” เป็นต้น

#### 4.1.4 ช่วงทศวรรษที่ 80 ยุคความกระตือรือร้นกลับมาใหม่ของโครงข่ายประสาทเทียม

ในปี ค.ศ. 1982 สหรัฐอเมริกาและญี่ปุ่นได้ร่วมมือกันจัด Conference on Cooperative /Competitive Neural Network ขึ้นที่ญี่ปุ่น ซึ่งเป็นจุดเริ่มต้นของการประชุมสัมมนาทางด้านโครงข่ายประสาท ในปี ค.ศ. 1985 American Institute of Physics ได้จัดงาน Neural Network for Computing Meeting ในปี ค.ศ. 1987 Institute of Electrical and Electronic Engineers (IEEE) ได้จัดการประชุมนานาชาติเกี่ยวกับโครงข่ายประสาทเทียมขึ้นเป็นครั้งแรก โดยมีผู้ร่วมงานมากกว่า 1,800 คน ในปีเดียวกัน International Neural Network Society (INNS) ถูกก่อตั้งขึ้น โดยมี Grossberg จากสหรัฐอเมริกา Kohonen จากฟินแลนด์ และ Amari จากญี่ปุ่นเป็นผู้นำในการก่อตั้ง โดยหลังจากที่ INNS ก่อตั้งได้ 2 ปี ก็มีสมาชิกเข้าร่วมมากกว่า 3,000 คน จากทั่วโลก

#### 4.1.4.1 โครงข่ายประสาทเทียมชนิด Back-Propagation

สองเหตุผลหลักที่ทำให้เกิดยุคเงียบของโครงข่ายประสาทเทียมคือ ความผิดพลาดของ Perceptron ชั้นเดียวที่ไม่สามารถแก้ปัญหาที่ไม่ซับซ้อนได้ เช่น ฟังก์ชัน XOR และอีกเหตุผลก็คือไม่มีวิธีฝึกสอนโครงข่ายประสาทเทียมหลายชั้น วิธีสำหรับแพร่ข้อมูลค่าผิดพลาดที่โหนดเอาต์พุตย้อนกลับไปที่ชั้นซ่อนได้ถูกค้นพบโดย Paul Werbos ในปี ค.ศ. 1974 แต่ไม่ได้กระจายใช้อย่างกว้างขวาง ซึ่งวิธีนี้ถูกค้นพบอีกครั้งในปี ค.ศ. 1985 โดย David Parker และ LeCun ก่อนใช้อย่างกว้างขวาง เปรียบเสมือนเป็นวิธีคิดแรกเริ่มในทฤษฎีควบคุมอย่างเหมาะสม (Optimal Control Theory) ในปี ค.ศ. 1986 งานของ Parker กลายเป็นที่สนใจของกลุ่ม Parallel Distributed Processing ซึ่งนำทีมโดยนักจิตวิทยา David Rumelhart ของมหาวิทยาลัย California ที่ Sandiego และ James McClelland ของมหาวิทยาลัย Carnegie Mellon ผู้ซึ่งนำวิธีดังกล่าวมาปรับปรุงให้ดีขึ้นและถูกใช้อย่างกว้างขวางในเวลาต่อมา หลักการของโครงข่ายประสาทเทียมชนิดนี้สามารถนำมาใช้แก้ปัญหาในลักษณะเชิงเส้น และปัญหาที่ไม่เชิงเส้นได้ ทำให้ประยุกต์เข้ากับโครงการต่างๆ ได้อย่างหลากหลาย

#### 4.1.4.2 โครงข่ายประสาทเทียมชนิด Hopfield

ในปี ค.ศ. 1982 John Hopfield ทำงานที่สถาบันเทคโนโลยี California พร้อมกับ David Tank นักวิจัยที่ AT & T Hopfield ได้พัฒนาโครงข่ายประสาทเทียมที่ประสบความสำเร็จเป็นอย่างดีในนามโครงข่ายประสาทเทียมชนิด Hopfield ส่งผลให้ John Hopfield ได้รับรางวัลโนเบลสาขาฟิสิกส์ และได้พัฒนาจำนวนของโครงข่ายประสาทเทียมพื้นฐานค่าถ่วงน้ำหนักคงที่ และ Adaptive Activation โครงข่ายประสาทเทียมเหล่านั้นสามารถมีประโยชน์เหมือนกับโครงข่ายประสาทเทียมรวมกลุ่มความสามารถในการจดจำและสามารถนำไปใช้เพื่อแก้ปัญหาการบีบบังคับความพอใจ เช่น “Traveling Salesman Problem” และเขาได้นำเสนอบทความต่อ National Academy of Sciences ซึ่งมีเนื้อหาเกี่ยวกับการใช้โครงข่ายประสาทเทียมในการทำนายส่วนที่ขาดหายไปของข้อมูลหรือรูปภาพ

#### 4.1.4.3 โครงข่ายประสาทเทียมชนิด Neocognitron

ในปี ค.ศ. 1975-1988 Kumihiko Fukushima และเพื่อนร่วมงานที่ห้องทดลอง NHK ในกรุงโตเกียว ได้พัฒนาตระกูลโครงข่ายประสาทเทียมที่ใช้งานเฉพาะสำหรับการจดจำตัวอักษรหนึ่งในตัวอย่างของโครงข่ายประสาทเทียมกลุ่มนี้คือ Neocognitron โครงข่ายประสาทเทียมชนิด Self-Organizing ในช่วงแรกถูกเรียกว่า Cognitron แต่เนื่องจากมีความผิดพลาดเกี่ยวกับสภาวะการจดจำข้อบกพร่องเหล่านี้ได้ถูกแก้ไขด้วยโครงข่ายประสาทเทียมชนิด Neocognitron

#### 4.1.4.4 โครงข่ายประสาทเทียมชนิด Boltzman

ในช่วงปี ค.ศ. 1983-1987 นักวิจัยหลายท่านได้พัฒนาการหาค่าเชื่อมโยงและฟังก์ชันกระตุ้น โดยใช้ฟังก์ชันความหนาแน่นของความจะเป็น (Probability Density Function) ต่างจากเดิมที่มักจะใช้แบบ Nondeterministic โดยพัฒนาโครงข่ายประสาทเทียมในนาม Boltzman ซึ่งอาศัยแนวคิดจากการจำลองการอ่อนตัวลงสู่สภาวะเสถียรภาพ (Simulated Annealing) และทฤษฎีการตัดสินใจของ บายเอเซียน (Bayesian Decision Theory)

## 4.2 โครงข่ายประสาทชีวภาพ

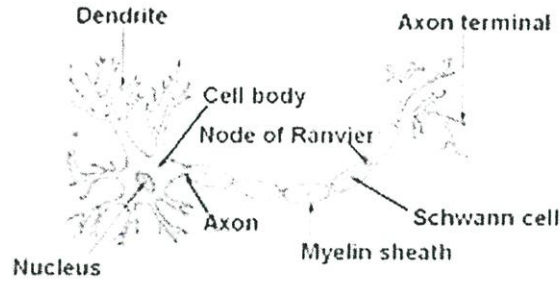
สมองมนุษย์ประกอบไปด้วยเซลล์ประสาท (Neuron) จำนวนมาก (ประมาณ  $10^{11}$  ยูนิต) ที่มีการเชื่อมโยงกันอย่างหนาแน่น (การเชื่อมโยงแต่ละเซลล์ประสาทประมาณ  $10^4$  ยูนิต) แต่ละเซลล์ประสาทประกอบด้วยเดนไดรต์ (Dendrites) ทำหน้าที่รับรู้อารมณ์จากเส้นใยประสาทโดยส่งข้อมูลเป็นสัญญาณไฟฟ้า (Electrical Signals) มาที่เดนไดรต์ และที่เดนไดรต์แต่ละกิ่งจะรับได้ด้วยค่าถ่วงน้ำหนัก (Weight) ที่ต่างกันและการแทนค่าถ่วงน้ำหนักด้วยความแข็งแรง (Strength) ของแต่ละไซแนปส์ (Synapse) จากนั้นจึงส่งข้อมูลตัวเซลล์ (Cell Body or Soma) ซึ่งทำหน้าที่รวบรวมสิ่งที่ได้รับรู้แล้วส่งให้แอกซอน (Axon) แอกซอนจะส่งสัญญาณออกไป โดยสัญญาณที่ส่งออกมาจะเป็นฟังก์ชันของผลรวมของสิ่งที่ได้รับรู้จากตัวเซลล์

### 4.2.1 นิวรอน

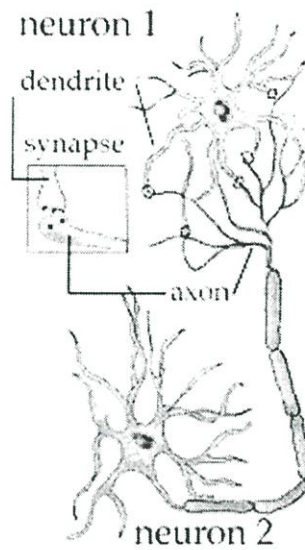
นิวรอน หมายความว่า เซลล์ประสาทและกระบวนการทั้งหมดของเซลล์ประสาท ในรูปที่ 4.1 แสดงตัวอย่างของนิวรอน ที่สามารถพบได้ในเรตินา เป็นนิวรอนที่มีสองขั้ว (Bipolar Neural) ในตัวเซลล์ (Cell Body) ของนิวรอนจะมีนิวเคลียส (Nucleus) อยู่ตรงกลางเซลล์ และมีแขนงอย่างน้อย 1 เส้นเชื่อมต่อกับเซลล์ แขนงที่ทำหน้าที่นำกระแสประสาท ซึ่งมีลักษณะเป็นสัญญาณอิมพัลส์ (Impulse Signal) เข้ามายังตัวเซลล์นั้นเรียกว่า เดนไดรต์ (Dendrite) ส่วนแขนงที่ทำหน้าที่นำกระแสประสาทออกจากตัวเซลล์เรียกว่า แอกซอน (Axon)

### 4.2.2 โครงสร้างของนิวรอนและไซแนปส์

เมื่อนิวรอนมากกว่า 1 นิวรอนอยู่รวมกันจะมีการจัดเรียงเป็นโครงสร้างของประสาทดังแสดงในรูปที่ 4.2 จะมีการส่งกระแสประสาทจากนิวรอนอื่น โดยผ่านจุดเชื่อมต่อระหว่างปลายของแอกซอนของนิวรอนหนึ่งกับตัวเซลล์หรือเดนไดรต์ของอีกนิวรอนหนึ่ง หรือเรียกว่า ไซแนปส์ (Synapse) ซึ่งเป็นช่องว่างที่เล็กมากระหว่างนิวรอนสองนิวรอนที่อยู่ติดกันแต่ไม่ได้สัมผัสกันโดยตรง ไซแนปส์ จะเป็นบริเวณที่ควบคุมให้มีการส่งผ่านกระแสประสาทต่อไปหรือยับยั้งการส่งกระแสประสาท รวมถึงเปลี่ยนแปลงลักษณะของกระแสประสาท อีกด้วย



รูปที่ 4.1 ตัวอย่างของนิวรอนที่มีสองขั้ว



รูปที่ 4.2 โครงสร้างของประสาท

#### 4.2.3 การส่งผ่านกระแสประสาท

ในสภาวะปกติเมื่อนิวรอนกระแสประสาทที่ได้รับจากอวัยวะที่ทำหน้าที่รับรู้ (อาทิเช่น ตา หรือ หู) ไปยังอวัยวะที่ทำหน้าที่ตอบสนอง (อาทิเช่น กล้ามเนื้อ หรือ ต่อมต่างๆภายในร่างกาย) กระแสประสาทจะถูกส่งผ่านจากนิวรอนหนึ่งไปยังอีกนิวรอนหนึ่ง โดยผ่านไซแนปส์ เมื่อกระแสประสาทผ่านเข้าไปในไซแนปส์ (เสมือนเป็นอินพุตของไซแนปส์) สารเคมีในไซแนปส์จะมีผลต่อกระแสประสาทได้สองลักษณะ คือการกระตุ้นให้กระแสประสาทมีขนาดเพิ่มขึ้นหรือยับยั้งให้กระแสประสาทมีขนาดลดลง ซึ่งทั้งสองลักษณะดังกล่าวเปรียบได้กับการที่สัญญาณอินพุตถูกถ่วงน้ำหนักให้มีความมากขึ้นหรือลดลง เมื่อผ่านออกจากไซแนปส์กระแสประสาทที่ถูกกระตุ้นและกระแสประสาทที่ถูกยับยั้งจะถูกรวมเข้าด้วยกัน หากผลรวมของกระแสประสาทดังกล่าวมีค่ามากกว่าหรือเท่ากับศักยค่าที่ผิวของนิวรอนถัดไป หรือเรซโซลด์ (Threshold) กระแสประสาท

ดังกล่าวก็จะสามารถไหลผ่านเข้าไปในนิวรอนได้ หากผลรวมของกระแสประสาทนั้นมีค่าน้อยกว่าค่าเทรชโฮลด์ กระแสประสาทก็จะถูกยับยั้งไม่ให้ผ่านเข้าไปในนิวรอนนั้น การไหลของกระแสประสาทจะเป็นไปในทิศทางเดียว (One Direction) และมีความเร็วอยู่ระหว่าง 10 ถึง 120 เมตรต่อวินาที เมื่อการส่งกระแสประสาทครั้งหนึ่งสิ้นสุด ส่วนต่างๆของนิวรอนจะกลับเข้าสู่สภาวะเริ่มต้นอีกครั้งหนึ่งเพื่อเตรียมพร้อมสำหรับกระแสประสาทครั้งต่อไป

อาจสรุปคุณสมบัติของโครงข่ายประสาทชีวภาพได้ดังนี้

- 1) แต่ละนิวรอนจะรับข้อมูลอินพุตเป็นจำนวนมาก
- 2) ข้อมูลอินพุตอาจถูกปรับให้เหมาะสมด้วยค่าเชื่อมโยง ( Connection Value of Weight Value of Strength) ที่ไซแนปส์
- 3) ข้อมูลอินพุตที่ผ่านการปรับด้วยค่าการเชื่อมโยงจะถูกรวมและทำการประมวลผล
- 4) ภายใต้อาหารที่เหมาะสม เช่น ข้อมูลอินพุตจำนวนเพียงพอ นิวรอนจะส่งข้อมูลออกมาหนึ่งค่า
- 5) ข้อมูลเอาต์พุตจากนิวรอนหนึ่งอาจจะถูกส่งไปที่นิวรอนอื่นๆ หลายนิวรอนผ่านทางแอกซอน
- 6) นิวรอนแต่ละนิวรอนมีค่าเป็นเฉพาะตำแหน่งหรือเฉพาะที่ (Local)
- 7) ความจำแบ่งออกเป็น 2 ส่วน คือ
  - ความจำระยะยาว จะเก็บที่ไซแนปส์ของนิวรอนหรือเก็บในค่าการเชื่อมโยง
  - ความจำระยะสั้น สัมพันธ์กับข้อมูลเอาต์พุตจากนิวรอน
- 8) ความถูกต้องของการเชื่อมโยงขึ้นอยู่กับประสบการณ์ในการเรียนรู้ที่ผ่านมา
- 9) การควบคุมการถ่ายโอนของไซแนปส์อาจจะเป็นการกระตุ้นแบบเสริมหรือแบบหักล้าง

นอกจากที่กล่าวมา โครงข่ายประสาทชีวภาพยังยึดหยุ่นต่อความผิดพลาด

(Fault Tolerance) 2 ประเด็นด้วยกัน ประการแรกคือ ความสามารถในการจดจำและจำแนกข้อมูลอินพุตที่แตกต่างจากข้อมูลเอาต์พุตที่เคยได้รับ ไม่มากนัก ตัวอย่างเช่น ความสามารถในการจำแนกภาพบุคคลที่ไม่เคยเห็นมาก่อน หรือสามารถจดจำบุคคลที่ไม่ได้พบเห็นกันเป็นเวลานาน ประการที่สองความสามารถในการซ่อมแซมหรือชดเชยความเสียหายของตัวนิวรอนเอง ในสมองมนุษย์เมื่อนิวรอนหนึ่งชำรุดเสียหายก็จะมีการสร้างนิวรอนใหม่ขึ้นมาแทนที่ นั่นคือยังคงมีการเรียนรู้อย่างต่อเนื่องถึงแม้ว่าจะสูญเสียนิวรอนบางส่วน ในบางครั้งนิวรอนที่เหลือจะได้รับการฝึกสอนเพื่อทำหน้าที่แทนนิวรอนที่ชำรุดเสียหาย

จากจุดเด่นหลายประการของโครงข่ายประสาทชีวภาพผนวกกับเทคโนโลยี ด้านคอมพิวเตอร์การทำงานแต่ละครั้งของอุปกรณ์สารกึ่งตัวนำในคอมพิวเตอร์จะใช้เวลาอยู่ในช่วงนาโนวินาที ( $10^{-9}$ s) ซึ่งมีความรวดเร็วมากเมื่อเปรียบเทียบกับการทำงานของสมองมนุษย์ทำให้มีการคิดค้นแนวทางแก้ปัญหา โดยมีโครงสร้างองค์ประกอบและกลไกการแก้ปัญหาลักษณะเดียวกับโครงข่ายประสาทชีวภาพ โดยเรียกว่า โครงข่ายประสาทเทียม (Artificial Neural Network)

### 4.3 โครงข่ายประสาทเทียม

เมื่อทราบหลักการการทำงานของเซลล์ประสาทแล้ว จะสามารถสร้างแบบจำลอง ของเซลล์ประสาทได้ ซึ่งเรียกว่า เซลล์ประสาทเทียม (Artificial Neuron) หรือยูนิต (Unit) เซลล์ประสาทเทียมมีความเร็วในการประมวลผลสูงกว่าเซลล์ประสาท แต่สมองมนุษย์สามารถทำงานหลายๆอย่าง งานที่ซับซ้อนพร้อมๆกันได้เร็วกว่าคอมพิวเตอร์ เพราะโครงสร้างของเซลล์ประสาทมีโครงสร้างที่ขนานกันอย่างมาก จึงสรุปได้ว่าโครงข่ายประสาทเทียมเป็นอัลกอริทึมที่เลียนแบบการเรียนรู้ของสมองมนุษย์ โดยมีองค์ประกอบของโครงสร้างพื้นฐานที่สำคัญดังนี้ คือ ยูนิต (Unit), ตัวแปรอินพุท (Input Layer), ตัวแปรเอาต์พุท (Output Layer) และค่าถ่วงน้ำหนัก (Weighted Value) ซึ่งสามารถสรุปความสัมพันธ์ระหว่างเซลล์ประสาทกับเซลล์ประสาทเทียมได้ดังนี้

#### เซลล์ประสาท

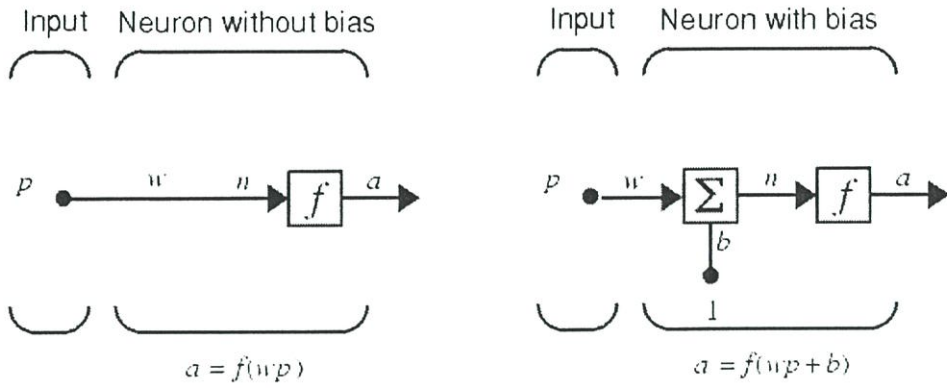
- ตัวเซลล์ (Cell body)
- เดนไดรต์ (Dendrite)
- แอกซอน (Axon)
- ไซแนปส์ (Synapse)

#### เซลล์ประสาทเทียม

- ยูนิต (Unit)
- ตัวแปรอินพุท (Input Layer)
- ตัวแปรเอาต์พุท (Output Layer)
- ค่าถ่วงน้ำหนัก (Weight Value)

#### 4.3.1 แบบจำลองของนิวรอนที่มี 1 อินพุท

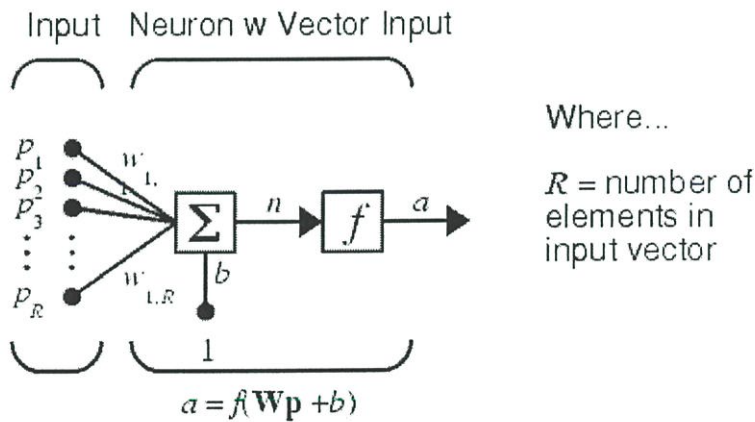
นิวรอนที่มี 1 อินพุท ที่ไม่มีค่าไบแอสแสดงในรูปที่ 4.3 ด้านซ้ายมือ และนิวรอนด้านขวามือจะมีค่าไบแอส “b” โดยที่  $f$  คือฟังก์ชันถ่ายโอน (Transfer Function) อาทิเช่น Linear Function หรือ Sigmoid Function ที่ถูกป้อนด้วยค่า “n” และให้ผลลัพธ์เป็นค่า “a” โดยส่วนสำคัญที่ทำให้โครงข่ายประสาทเทียมแสดงพฤติกรรมต่างๆ ได้ก็คือ ค่าถ่วงน้ำหนัก “w” และค่าไบแอส “b” ซึ่งเป็นสเกลลาร์ที่สามารถปรับเปลี่ยนค่าได้



รูปที่ 4.3 แบบจำลองของนิวรอนที่มี 1 อินพุต

4.3.2 แบบจำลองของนิวรอนที่มีหลายอินพุต

นิวรอนที่มีอินพุต “R” ชุด แสดงในรูปที่ 4.4 ในรูปนี้ อินพุต  $P_1, P_2, \dots, P_R$  ถูกเชื่อมต่อเข้ากับนิวรอนด้วยค่าถ่วงน้ำหนักจะถูกรวมเข้ากันเป็นอินพุตของฟังก์ชันถ่ายโอน “n” ดังสมการที่ (4.1)



รูปที่ 4.4 แบบจำลองของนิวรอนที่มีหลายอินพุต

$$n = P_1 * W_{1,1} + P_2 * W_{1,2} + \dots + P_R * W_{1,R} + b \tag{4.1}$$

### 4.3.3 ฟังก์ชันกระตุ้น (Activation Function)

ฟังก์ชันกระตุ้นหรือหนังสือบางเล่มเรียกว่า ฟังก์ชันการถ่ายโอน(Transfer Function) เป็นตัวกำหนดค่าเอาต์พุต ฟังก์ชันกระตุ้นที่ใช้ในการฝึกสอนโครงข่ายประสาทเทียมมีอยู่หลายชนิด เราสามารถแบ่งฟังก์ชันกระตุ้นพื้นฐานออกเป็น 3 ชนิด คือ

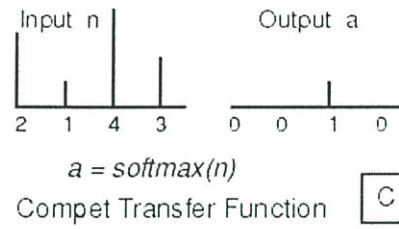
1. ฟังก์ชันเทรชโฮลด์ (Threshold Function หรือ Hard Limit Function หรือ Step Function )
2. ฟังก์ชันเชิงเส้นอิ่มตัว (Saturating Linear Function หรือ Piecewise – Linear Function)
3. ฟังก์ชันซิกมอยด์ (Sigmoid Function)

จากฟังก์ชันพื้นฐานทั้ง 3 ชนิด ฟังก์ชันซิกมอยด์เป็นฟังก์ชันที่ได้รับความนิยมนำมาใช้เป็นฟังก์ชันกระตุ้นของโครงข่ายประสาทเทียมกันมาก เนื่องจากเป็นฟังก์ชันที่เพิ่มขึ้นอย่างสม่ำเสมอ (smoothness) นอกเหนือจากฟังก์ชันกระตุ้นพื้นฐานทั้งสามแล้วยังสามารถนำความสัมพันธ์ระหว่างชุดข้อมูลในรูปอื่นๆ มาเป็นฟังก์ชันกระตุ้นได้ ตัวอย่างเช่น ฟังก์ชันกระตุ้นแบบคอมแพทิทิฟ (Competitive Function) โดยทั่วไปฟังก์ชันกระตุ้นจะจำกัดช่วงเอาต์พุตอยู่ตั้งแต่ 0 ถึง 1 ในการประยุกต์ใช้งานบางกรณีอาจมีความจำเป็นต้องใช้ฟังก์ชันที่มีค่าออกมานอกเหนือช่วงดังกล่าว ตัวอย่างเช่นต้องการเอาต์พุตอยู่ในช่วงตั้งแต่ -1 ถึง +1 สำหรับการพิจารณาเลือกใช้ฟังก์ชันการกระตุ้นแบบใดนั้นนับถึงปัจจุบันยังไม่มีข้อกำหนดที่แน่นอน แต่อาจจะพิจารณาองค์ประกอบดังต่อไปนี้

1. ความเป็นเชิงเส้นหรือไม่เป็นเชิงเส้น
2. ลักษณะของข้อมูลอินพุต
3. ขอบเขตที่ต้องการและลักษณะของข้อมูลเอาต์พุต
4. ปัจจัยอื่นที่มีนัยสำคัญต่อผลตอบสนองของนิวรอนหรือ โครงข่ายประสาทเทียมโดยรวม

ในบางครั้งอาจเลือกใช้ฟังก์ชันสโตกาสติก (Stochastic Function) เป็น ฟังก์ชันกระตุ้นในกรณีค่าความน่าจะเป็นของข้อมูลอินพุตมีนัยสำคัญ หรือในบางกรณีอาจเลือกใช้ฟังก์ชันแบบเชิงเส้นและไม่เป็นเชิงเส้นในโครงข่ายประสาทเทียมเดียวกัน ตัวอย่างเช่น โครงข่ายประสาทเทียมชนิดฟังก์ชันฐานรัศมี (RBF) ซึ่งใช้ฟังก์ชันกระตุ้นแบบไม่เป็นเชิงเส้นในชั้นซ่อน (Hidden Layer) ของโครงข่ายประสาท ขณะที่ชั้นเอาต์พุต (Output Layer) ใช้ฟังก์ชันกระตุ้นแบบเป็นเชิงเส้น ซึ่งสามารถแสดงฟังก์ชันกระตุ้นที่ใช้ใน การฝึกสอนโครงข่ายประสาทเทียมดังนี้

### 4.3.3.1 Competitive Transfer Function

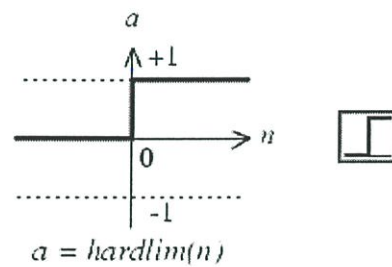


รูปที่ 4.5 Competitive Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 1; & i = i^* \\ 0; & i \neq i^* \end{cases} \quad (4.2)$$

### 4.3.3.2 Hard limit Transfer Function



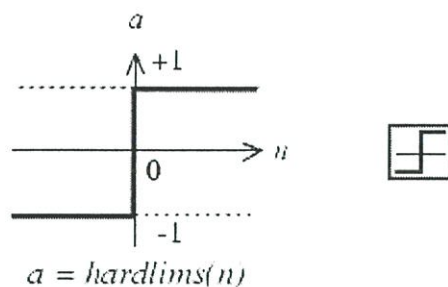
Hard-Limit Transfer Function

รูปที่ 4.6 Hard Limit Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 1 & \text{if } n \geq 0 \\ 0 & \text{if } n < 0 \end{cases} \quad (4.3)$$

### 4.3.3.3 Symmetric Hard Limit Transfer Function



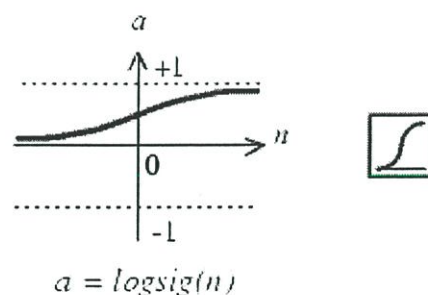
Symmetric Hard-Limit Trans. Funct.

รูปที่ 4.7 Symmetric Hard Limit Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 1 & \text{if } n \geq 0 \\ -1 & \text{if } n < 0 \end{cases} \quad (4.4)$$

### 4.3.3.4 Log sigmoid Transfer Function



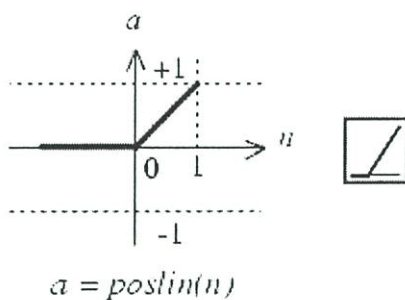
Log-Sigmoid Transfer Function

รูปที่ 4.8 Log sigmoid Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n) = \frac{1}{1 + e^{-kn}} \quad (4.5)$$

### 4.3.3.5 Positive Linear Transfer Function



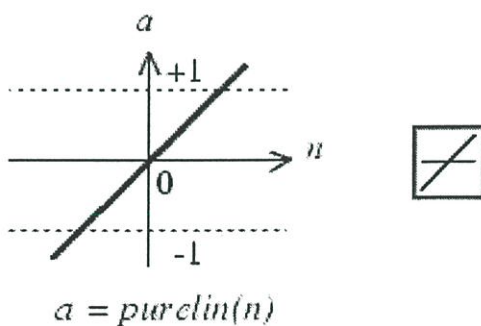
Positive Linear Transfer Funct.

รูปที่ 4.9 Positive Linear Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} n & \text{if } n \geq 0 \\ 0 & \text{if } n \leq 0 \end{cases} \quad (4.6)$$

### 4.3.3.6 Linear Transfer Function



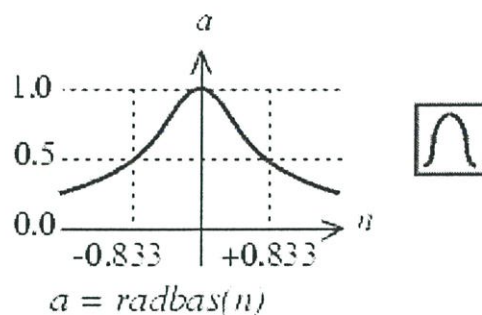
Linear Transfer Function

รูปที่ 4.10 Linear Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n) = n \quad (4.7)$$

### 4.3.3.7 Radial Basis Transfer Function



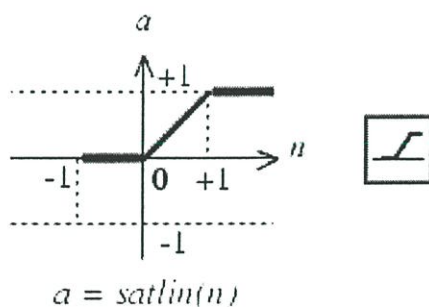
### Radial Basis Function

รูปที่ 4.11 Radial Basis Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n) = e^{-n^2} \quad (4.8)$$

### 4.3.3.8 Saturating Linear Transfer Function



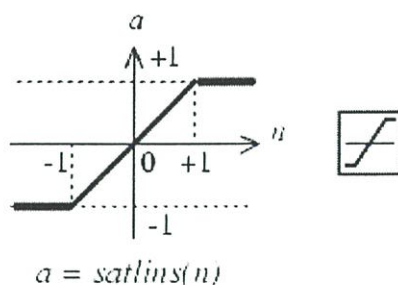
### Satlin Transfer Function

รูปที่ 4.12 Saturating Linear Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 0 & n \leq 0 \\ n & \text{if } 0 \leq n \leq 1 \\ 1 & 1 \leq n \end{cases} \quad (4.9)$$

### 4.3.3.9 Symmetric Saturating Linear Transfer Function



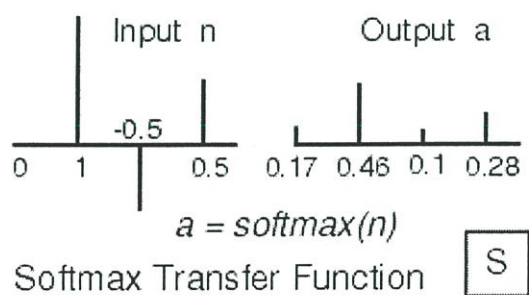
Satlins Transfer Function

รูปที่ 4.13 Symmetric Saturating Linear Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 0 & n \leq 0 \\ n & \text{if } 0 \leq n \leq 1 \\ 0 & 1 \leq n \end{cases} \quad (4.10)$$

### 4.3.3.10 Soft max Transfer Function

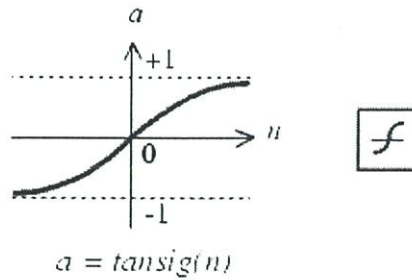


รูปที่ 4.14 Soft max Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n) = \frac{e^n}{\sum e^n} \quad (4.11)$$

### 4.3.3.11 Hyperbolic Tangent Sigmoid Transfer Function



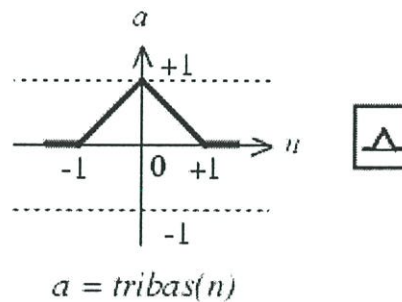
Tan-Sigmoid Transfer Function

### รูปที่ 4.15 Hyperbolic Tangent Sigmoid Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n) = \frac{1 - e^{(-2n)}}{1 + e^{(-2n)}} \quad (4.12)$$

### 4.3.3.12 Triangular Basis Transfer Function



Triangular Basis Function

### รูปที่ 4.16 Triangular Basis Transfer Function

สมการฟังก์ชันกระตุ้น คือ

$$a(n_i) = \begin{cases} 1 - |n| & -1 \leq n \leq 1 \\ \text{if} & \\ 0 & \text{otherwise} \end{cases} \quad (4.13)$$

#### 4.3.4 คุณสมบัติและความสามารถของโครงข่ายประสาทเทียม

คุณสมบัติและความสามารถของโครงข่ายประสาทเทียม สามารถสรุปได้ดังนี้

1. โครงข่ายประสาทเทียมไม่ใช่โปรแกรมคอมพิวเตอร์แต่โครงข่ายจะเรียนรู้จากตัวอย่าง
2. มีความยืดหยุ่นสูงจนสามารถจำลองกระบวนการของปัญหาใดๆได้
3. มีความสามารถในการจำชุดของค่าอินพุต เอาท์พุทที่มีความซับซ้อนมากจนไม่สามารถจำลองแบบในเชิงความน่าจะเป็นได้
4. มีความสามารถในการปรับตัวเข้ากับการเปลี่ยนแปลงของสิ่งแวดล้อม
5. มีความสามารถในการตอบสนองต่อข้อมูลที่ไม่เคยเห็น
6. ความรู้กระจายอยู่ทั่วโครงสร้างของโครงข่ายประสาทเทียม

#### 4.3.5 ลักษณะของปัญหาที่เหมาะสมกับโครงข่ายประสาทเทียม

ลักษณะของปัญหาที่เหมาะสมกับโครงข่ายประสาทเทียมนั้นมีดังนี้

1. ตัวอย่างของข้อมูลสำหรับการฝึกสอนประกอบด้วยองค์ประกอบหลายๆค่า
2. เอาท์พุทที่ต้องการสามารถเป็นค่าจำนวนเต็ม หรือจำนวนจริงหรือเป็นเวกเตอร์ของจำนวนเต็มหรือจำนวนจริงก็ได้
3. ตัวอย่างของข้อมูลสามารถมีความผิดพลาดปะปนอยู่ได้
4. ไม่มีปัญหากับขั้นตอนการฝึกหัดซึ่งอาจใช้เวลานานมาก
5. ต้องการประเมินผลเอาท์พุทอย่างรวดเร็ว
6. ไม่ต้องการเข้าใจวิธีการแก้ปัญหา ต้องการให้ปัญหาได้รับการแก้ไขเท่านั้น

โครงข่ายประสาทเทียมอาจพิจารณาให้เป็นขั้นตอนวิธีทางคณิตศาสตร์ (Mathematical Algorithm) ของโครงข่ายประสาทชีวภาพโดยตั้งบนพื้นฐานดังนี้

1. หน่วยประมวลผลเรียกว่า นิวรอน การส่งผ่านข้อมูลระหว่างนิวรอนทำได้โดยการเชื่อมโยงระหว่างนิวรอน
2. การเชื่อมโยงแต่ละส่วนนี้จะขึ้นอยู่กับค่าเชื่อมโยงหรือค่าถ่วงน้ำหนัก
3. ข้อมูลเอาท์พุทจะถูกกำหนดด้วยฟังก์ชันกระตุ้น (Activation Function) ซึ่งปกติความสัมพันธ์ระหว่างผลรวมของข้อมูลอินพุทที่คูณด้วยค่าการเชื่อมโยงกับข้อมูลเอาท์พุทเป็นความสัมพันธ์แบบไม่เป็นเชิงเส้น (Non-Linear)

## 4.4 การเรียนรู้ของโครงข่ายประสาทเทียม

การเรียนรู้ของโครงข่ายประสาทเทียมจะมีประสิทธิภาพเพียงใดนั้น ขึ้นอยู่กับค่าถ่วงน้ำหนักของโครงข่ายประสาทเทียม ซึ่งการสอนโครงข่ายประสาทเทียมก็คือการหาค่าถ่วงน้ำหนักที่เหมาะสมให้แก่โครงข่ายนั้นๆ ซึ่งโครงข่ายประสาทเทียมสามารถแบ่งออกเป็น 2 ประเภทตามลักษณะการเรียนรู้ คือ การเรียนรู้แบบควบคุม (Supervised Learning) และการเรียนรู้ไม่มีการควบคุม (Unsupervised Learning)

### 4.4.1 การเรียนรู้แบบควบคุม (Supervised Learning)

การเรียนรู้วิธีนี้จะมีการกำหนดคู่ของการฝึกสอนระหว่างอินพุตกับเอาต์พุตที่ต้องการให้โครงข่ายประสาทเทียม ที่เรียกว่า เทรนนิ่งแพร์ (Training Pairs) โครงข่ายประสาทเทียมจะถูกฝึกสอนไปตามจำนวนของคู่ที่ป้อน (จำนวนคู่ของอินพุตกับเอาต์พุตที่ต้องการให้โครงข่ายประสาทเทียมรู้จัก) เอาต์พุตที่คำนวณได้จากโครงข่ายประสาทเทียมจะถูกเปรียบเทียบให้สอดคล้องกับเป้าหมายค่าผิดพลาดที่เกิดขึ้นจะถูกป้อนกลับไปยังโครงข่ายประสาทเทียมและเปลี่ยนแปลงค่าถ่วงน้ำหนักให้สอดคล้องกับอัลกอริทึม ที่ทำให้แนวโน้มของข้อผิดพลาดที่เกิดขึ้นระหว่างเอาต์พุตกับเป้าหมายโดยเฉลี่ยมีค่าต่ำลง หลังจากการฝึกสอน โครงข่ายประสาทเทียมจะถูกทดสอบโดยให้เฉพาะค่าอินพุต แต่ไม่ให้ค่าเอาต์พุตที่ต้องการและพิจารณาเอาต์พุตที่ได้จากโครงข่ายประสาทเทียมใกล้เคียงกับค่าเอาต์พุตที่ถูกต้องหรือไม่ โครงข่ายประสาทเทียมประเภทนี้จะแบ่งออกเป็น Feed-Forward Neural Networks และ Radial Basis Function Networks ซึ่งสองชนิดนี้นิยมใช้การแก้ปัญหาลักษณะการประมาณค่าฟังก์ชัน (Function Approximation)

### 4.4.2 การเรียนรู้ไม่มีการควบคุม (Unsupervised Learning)

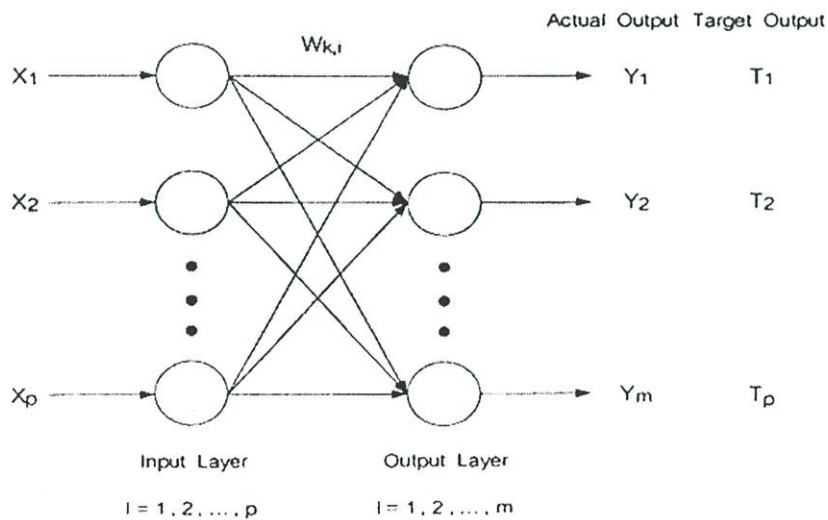
การเรียนรู้โดยวิธีนี้จะป้อนข้อมูลอินพุตเข้าสู่โครงข่ายและภายใน โครงข่ายจะมีโหนดเอาต์พุตจำนวนหลายโหนดด้วยกัน โดยแต่ละโหนดจะแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนข้อมูลอินพุตเข้าสู่โครงข่าย โครงข่ายจะคำนวณค่าความสัมพันธ์ที่มีอยู่ภายในเซตของอินพุตโดยอาศัยค่าน้ำหนักเป็นตัวแยกความแตกต่างของข้อมูลอินพุตไปเก็บไว้ในโหนดเอาต์พุตของโครงข่าย การสอนโดยวิธีนี้จะไม่สามารถระบุได้ว่าข้อมูลเอาต์พุตโหนดใดเป็นข้อมูลของกลุ่มไหน ผู้ใช้จะต้องกำหนดเอง ซึ่งต่างจากการสอนแบบชี้หน้าที่โครงข่ายสามารถระบุกลุ่มของข้อมูลเอาต์พุตได้อย่างแน่นอน โครงข่ายประสาทเทียมประเภทนี้จะแบ่งออกเป็น 2 ประเภท ได้แก่ Kohonen Networks ซึ่งนิยมใช้ในการแก้ปัญหาลักษณะการแยกแยะ (Classification Problem) และ Hopfield Networks ซึ่งนิยมใช้ในการแก้ปัญหาลักษณะการหาค่าที่เหมาะสม (Optimization Problem)

#### 4.5 สถาปัตยกรรมของโครงข่ายประสาทเทียม (Architectures of Network)

ในการสร้างโครงข่ายประสาทเทียมเพียงหนึ่งเซลล์ โดยใช้แนวความคิดจากเซลล์ประสาทชีวภาพ การจะนำเซลล์ประสาทเทียมมาใช้งานได้นั้น ต้องใช้เซลล์ประสาทเทียมที่มีคุณลักษณะต่างกัน (ค่าถ่วงน้ำหนักจะทำให้คุณสมบัติของเซลล์ประสาทเทียมแต่ละเซลล์มีคุณลักษณะแตกต่างกันไป) มาเชื่อมต่อเซลล์ประสาทหลายๆ ตัวเข้าด้วยกันให้เกิดลักษณะของโครงข่ายเป็นชั้นๆ หรือที่เรียกว่าเลเยอร์ (Layer) ซึ่งประสาทเซลล์แต่ละตัวที่อยู่ในชั้นเดียวกันจะไม่มี การเชื่อมต่อถึงกัน ประเภทของโครงสร้างการเชื่อมต่อโครงข่ายประสาทเทียม พอจะแบ่งออกเป็น 2 แบบ คือ โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Neural Network), โครงข่ายประสาทเทียมแบบหลายชั้น (Multi Layer Neural Network) ซึ่งแต่ละชนิดมีคุณลักษณะเด่นแตกต่างกันไป

##### 4.5.1 โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Neural Network)

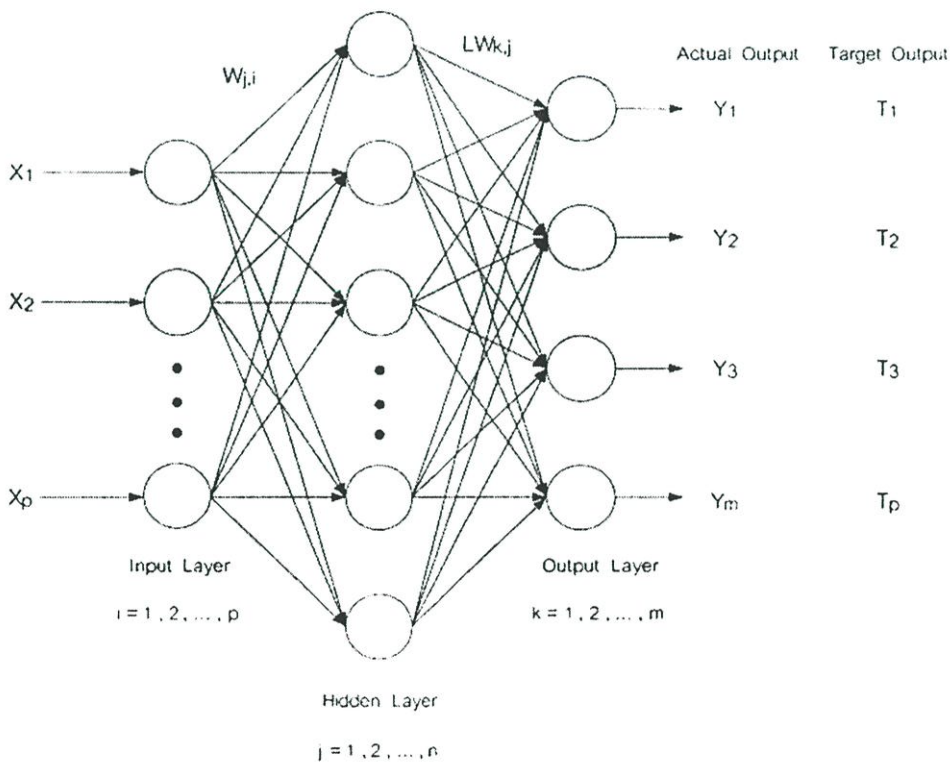
จะประกอบไปด้วยเซลล์ประสาทที่จัดเรียงตัวอยู่ในชั้นต่างกัน ได้แก่ ชั้นอินพุตและชั้นเอาต์พุต สาเหตุที่จัดโครงสร้างแบบนี้เป็นแบบชั้นเดียว เนื่องจากเรานับชั้นเอาต์พุตเพียงชั้นเดียวเท่านั้นว่าเป็นชั้นของเซลล์ประสาท สำหรับชั้นอินพุตเราไม่พิจารณาว่าเป็นชั้นของเซลล์ประสาทก็เพราะเป็นชั้นที่ไม่มีการประมวลใดๆ จะทำหน้าที่เพียงแครับข้อมูลอินพุตเข้ามาและส่งต่อไปกับชั้นถัดไปเท่านั้น รูปที่ 4.17 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบชั้นเดียว



รูปที่ 4.17 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบชั้นเดียว

#### 4.5.2 โครงข่ายประสาทเทียมแบบหลายชั้น (Multi Layer Neural Network)

โครงข่ายประสาทเทียมแบบชั้นเดียวมีข้อจำกัด ไม่สามารถนำมาใช้แก้ปัญหาต่างๆในบางกรณีได้ นักวิจัยจึงได้นำเสนอรูปแบบของโครงข่ายประสาทเทียมแบบหลายชั้น ซึ่งมีโครงสร้างประกอบขึ้นจากการจัดเรียงของตัวเซลล์ประสาทตั้งแต่ 2 ชั้นขึ้นไป (ไม่นับรวมชั้นอินพุท) โดยปกติโครงข่ายประสาทเทียมแบบหลายชั้นจะประกอบด้วยโครงข่ายประสาทเทียมแบบชั้นเดียวหลายๆตัวต่ออยู่ในโครงสร้างที่มีลำดับชั้นลดหลั่นกันลงไป โดยอยู่ในรูปของการทำงานที่ป้อนไปข้างหน้า (Feed -Forward) ซึ่งชั้นที่อยู่ระหว่างชั้นอินพุทและชั้นเอาต์พุทเรียกว่า ชั้นซ่อน (Hidden Layer) ซึ่งชั้นซ่อนนี้เองที่เป็นตัวเพิ่มความสามารถให้แก่โครงข่าย ทั้งนี้จำนวนของชั้นซ่อนสามารถมีได้มากกว่า 1 ชั้น จำนวนของชั้นซ่อนและจำนวนของนิวรอนในแต่ละชั้นจะมีค่าไม่คงที่ ซึ่งแต่ละชั้นอาจจะมีจำนวนนิวรอนที่แตกต่างกันขึ้นอยู่กับการประยุกต์ใช้งาน จากรูปที่ 4.18 แสดงชุดข้อมูลเอาต์พุทที่ได้จากชั้นที่หนึ่งจะถูกส่งให้เป็นข้อมูลอินพุทสำหรับชั้นต่อไป โดยค่าเอาต์พุทชั้นสุดท้ายซึ่งก็คือ เอาต์พุทของโครงข่ายประสาทเทียมจะถูกนำไปเปรียบเทียบกับค่าความผิดพลาดกับค่าเอาต์พุทที่ต้องการ (Target Output) ค่าความผิดพลาดที่คำนวณได้จะนำไปใช้ในการปรับค่าถ่วงน้ำหนักของโครงข่ายประสาทเทียมให้เปลี่ยนไปในทางที่ถูกต้องมากขึ้น



รูปที่ 4.18 แสดง โครงสร้างของ โครงข่ายประสาทเทียมแบบหลายชั้น

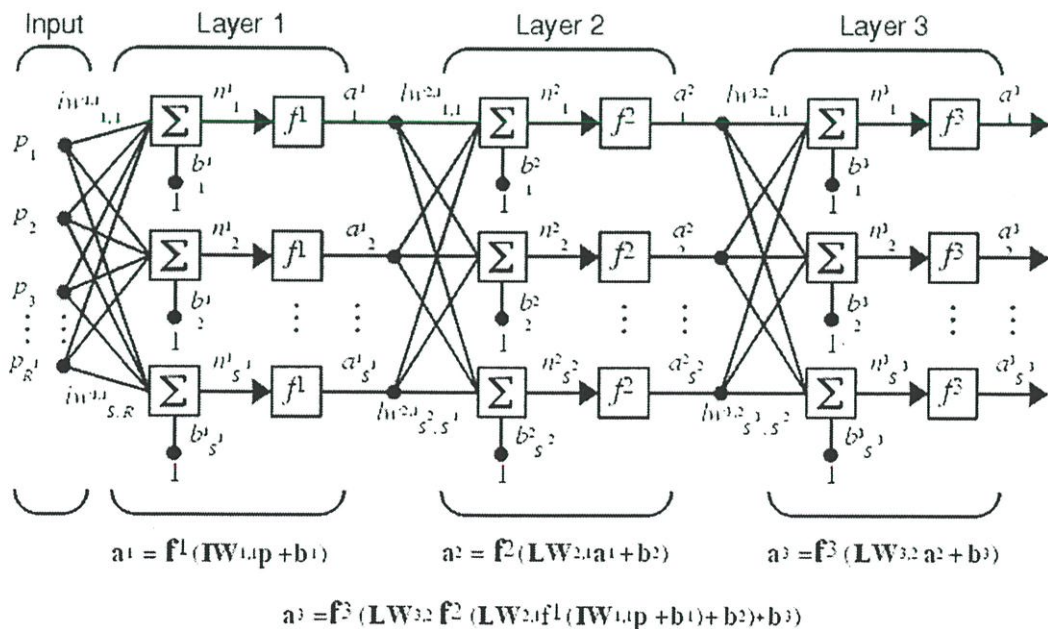
## 4.6 พื้นฐานแนวคิดของโครงข่ายประสาทเทียม (Basis Algorithm of Neural Networks)

โครงข่ายประสาทเทียมในปัจจุบันมีหลากหลายชนิด ซึ่งนำไปใช้ในงานลักษณะที่ต่างกัันจึงจะขอกกล่าวถึงรายละเอียดเฉพาะ โครงข่ายประสาทเทียมที่ในวิทยานิพนธ์ฉบับนี้เท่านั้นซึ่งเป็นโครงข่ายประสาทเทียมที่มีอยู่ในโปรแกรม MATLAB/Neural Network Toolbox

### 4.6.1 โครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ (Back-Propagation Neural Network :BP)

#### 4.6.1.1 โครงสร้าง

โครงสร้างพื้นฐานของโครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ (Back-Propagation Neural Network :BP) ประกอบด้วย นิวรอนเซลล์ที่เรียงกันอยู่มากกว่า 2 ชั้นกล่าวคือ มีได้ตั้งแต่ 3 ชั้นขึ้นไป ซึ่งประกอบด้วยชั้นอินพุท (Input Layer) ,ชั้นซ่อน (Hidden Layer) และชั้น เอาท์พุท (Output Layer) ลักษณะของโครงข่ายชนิดนี้เป็นโครงข่ายที่ป้อนไปข้างหน้า โดยมีการเชื่อมโยงในแต่ละชั้นแบบต่อกันหมด นั่นคือ ทุกๆนิวรอนในชั้นอินพุทจะส่งสัญญาณไปยังทุกๆนิวรอนในชั้นซ่อนชั้นแรกและทุกๆนิวรอนในชั้นซ่อนชั้นแรกจะส่งสัญญาณไปยังทุกๆนิวรอนในชั้นถัดไป จนในที่สุดทุกๆนิวรอนในชั้นสุดท้ายจะส่งสัญญาณไปยังทุกๆนิวรอนในชั้นเอาท์พุท จากรูปที่ 4.19 แสดงโครงสร้างพื้นฐานของโครงข่าย 3 ชั้น



รูปที่ 4.19 แสดงแบบจำลองของโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อน

จากรูปที่ 4.19 แสดงโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อน มีอินพุต  $R$  ชุด, นิวรอนในชั้นที่ 1 และ 2 เท่ากับ  $S^1$  และ  $S^2$  และมีเอาต์พุต  $S^3$  ชุด สมาชิกของอินพุตเวกเตอร์  $P$  แต่ละตัวจะถูกเชื่อมเข้ากับนิวรอนแต่ละนิวรอนในชั้นซ่อนที่ 1 ด้วยค่าถ่วงน้ำหนัก  $W^1$  (เรียกว่า Input Weight,  $IW^{1,1}$ ) ซึ่งเป็นเมตริกซ์ขนาด  $S^1 \times R$  และเวกเตอร์  $a^1$  ซึ่งผลลัพธ์ของชั้นซ่อนที่ 1 และขนาด  $S^1$  จะเป็นอินพุตของชั้นซ่อนที่ 2 ด้วยค่าถ่วงน้ำหนัก  $W^2$  (เรียกว่า Layer Weight,  $LW^{2,1}$ ) ซึ่งเป็นเมตริกซ์ขนาด  $S^2 \times S^1$  เช่นเดียวกับชั้นซ่อนที่ 1 เวกเตอร์  $a^2$  ซึ่งผลลัพธ์ของชั้นซ่อนที่ 2 และขนาด  $S^2$  จะเป็นอินพุตของชั้นข้อมูลเอาต์พุต สมาชิกแต่ละตัวของเวกเตอร์  $a^2$  จะถูกเชื่อมเข้ากับนิวรอนแต่ละตัวชั้นข้อมูลเอาต์พุตด้วยค่าถ่วงน้ำหนัก  $W^3$  (เรียกว่า  $LW^{3,2}$ ) ซึ่งเป็นเมตริกซ์ขนาด  $S^3 \times S^2$  โดยที่เวกเตอร์  $a^3$  ซึ่งเป็นเมตริกซ์ขนาด  $S^3$  จะเป็นผลลัพธ์ของโครงข่ายประสาทเทียม

#### 4.6.1.2 แนวคิด

ขั้นตอนการฝึกสอนของโครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ จะสามารถแบ่งออกเป็น 3 ขั้นตอนด้วยกัน คือ ป้อนข้อมูลอินพุตไปข้างหน้า, แพร่ค่าย้อนกลับของค่าผิดพลาดที่เกี่ยวข้องและปรับค่าถ่วงน้ำหนัก ระหว่างป้อนไปข้างหน้า จะเริ่มต้นด้วยการนำค่าอินพุตที่ได้รับจากชั้นอินพุตมาคำนวณหาผลรวมที่ได้รับ ในแต่ละนิวรอน แล้วจึงส่งผลลัพธ์ที่ได้ส่งต่อไปยังชั้นซ่อนที่ 1 แต่ละนิวรอนในชั้นซ่อนที่ 1 จะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นนั้น แล้วจึงส่งผลลัพธ์ที่ได้ส่งไปยังชั้นซ่อนที่อยู่ถัดไป ในชั้นซ่อนที่ 2 แต่ละนิวรอนในชั้นซ่อนที่ 2 จะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นนั้น จากนั้นจะส่งผลลัพธ์ที่ได้ส่งต่อไปยังชั้นถัดไป สุดท้ายแต่ละนิวรอนในชั้นเอาต์พุตจะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นเอาต์พุต เพื่อหาค่าตอบหรือผลตอบสนองของโครงข่ายสำหรับแต่ละรูปแบบข้อมูลที่ป้อนให้ ทิศทางการไหลของข้อมูลเหล่านี้จะเป็นลักษณะเคลื่อนไปข้างหน้าจากชั้นอินพุตผ่านชั้นของนิวรอนเซลล์ภายในไปสู่ชั้นเอาต์พุต วิธีการคำนวณของโครงข่ายลักษณะนี้ จึงมีชื่อตามลักษณะดังกล่าวว่า Forward – Propagation ขั้นตอนต่อไปจะเป็นการคำนวณแบบ Backward – Propagation สำหรับขั้นตอนนี้จะเป็นการนำผลลัพธ์ที่ได้จากเอาต์พุตของโครงข่ายมาเปรียบเทียบกับเอาต์พุตเป้าหมายเพื่อคำนวณค่าผิดพลาดที่เกี่ยวข้อง แล้วค่าผลต่างระหว่างเอาต์พุตของโครงข่ายกับเอาต์พุตเป้าหมายที่ได้จากโครงข่ายจะถูกส่งถอยหลังกลับมายังชั้นซ่อนที่ 2 และส่งต่อไปยังนิวรอนต่างๆ ภายในชั้นซ่อนที่ 2 ซึ่งแต่ละโหนดจะได้รับค่าผิดพลาดเพียงบางส่วนขึ้นอยู่กับว่านิวรอนเซลล์นั้นเป็นตัวส่งผลมากหรือน้อยไปสู่เอาต์พุตนั้น เพื่อปรับค่าถ่วงน้ำหนักระหว่างชั้นเอาต์พุตกับชั้นซ่อนที่ 2 กระบวนการของการส่งค่าความผิดพลาดกลับมาจะซ้ำกับชั้นถัดลงมาอีกจนกระทั่งทุกโหนดในโครงข่ายได้รับส่วนแบ่งค่าความผิดพลาดนั้น

ขั้นตอนสุดท้ายจะนำค่าผิดพลาดที่ได้มาใช้ในการปรับค่าถ่วงน้ำหนัก ซึ่งค่าถ่วงน้ำหนักจะเปลี่ยนไปมากหรือน้อยขึ้นอยู่กับปริมาณของค่าความผิดพลาดที่ได้รับ กระบวนการจะกลับไป

ทำซ้ำในขั้นตอนแรกจนกระทั่งค่าความผิดพลาดที่ได้มีค่าน้อยกว่าค่าผิดพลาดต่ำสุดที่กำหนดไว้ต่อไปเป็นรายละเอียดของขั้นตอนการทำงานในแต่ละขั้นตอน

### 1. แพร่ค่าไปข้างหน้า (Forward-Propagation)

เมื่อข้อมูลอินพุตปรากฏบนชั้นอินพุต ( $P_R ; R = 1, \dots, R$ ) สัญญาณจากชั้นอินพุตทั้งหมดถูกส่งมาตามสายเชื่อมโยงไปยังนิวรอนเซลล์ในชั้นถัดไป แต่ละนิวรอนในชั้นช่อนที่ 1 ถูกเชื่อมโยงด้วยค่าถ่วงน้ำหนัก ( $iw^{1,1}_{S^1,R} = S^1 = 1, \dots, S^1$ ) จะทำการคำนวณสัญญาณอินพุตของนิวรอนเซลล์ในชั้นช่อนที่ 1 ได้สมการ

$$\text{net}_{S^1} = b_{S^1}^1 + \sum_{R=1}^R iw_{S^1,R}^{1,1} P_R \quad (4.14)$$

ใช้ฟังก์ชันกระตุ้นเพื่อคำนวณสัญญาณเอาต์พุต ( $a^1$ ) ในชั้นช่อนที่ 1

$$a^1 = f(\text{net}_{S^1}) \quad (4.15)$$

เมื่อสัญญาณเอาต์พุต ( $a^1$ ) ในชั้นช่อนที่ 1 ได้ส่งต่อไปยังชั้นช่อนที่ 2 โดยแต่ละนิวรอนในชั้นช่อนที่ 2 ถูกเชื่อมโยงด้วยค่าถ่วงน้ำหนัก ( $lw^{2,1}_{S^2,S^1} = S^2 = 1, \dots, S^2$ ) จะทำการคำนวณสัญญาณอินพุตของนิวรอนเซลล์ในชั้นช่อนที่ 2 ได้จากสมการ

$$\text{net}_{S^2} = b_{S^2}^2 + \sum_{S^1=1}^{S^1} lw_{S^2,S^1}^{2,1} a^1 \quad (4.16)$$

ใช้ฟังก์ชันกระตุ้นเพื่อคำนวณสัญญาณเอาต์พุต ( $a^2$ ) ในชั้นช่อนที่ 2

$$a^2 = f(\text{net}_{S^2}) \quad (4.17)$$

เมื่อสัญญาณเอาต์พุต ( $a^2$ ) ในชั้นช่อนที่ 2 ได้ส่งต่อไปยังแต่ละนิวรอนในชั้นเอาต์พุตถูกเชื่อมโยงด้วยค่าถ่วงน้ำหนัก ( $lw^{3,1}_{S^3,S^2} = S^3 = 1, \dots, S^3$ ) จะทำการคำนวณสัญญาณอินพุตของนิวรอนเซลล์ในชั้นเอาต์พุต ได้จากสมการ

$$\text{net}_{s^3} = b^3 s^3 + \sum_{s^2=1}^{s^2} lw_{s^3, s^2}^{3,2} a^2 \quad (4.18)$$

ใช้ฟังก์ชันกระตุ้นเพื่อคำนวณสัญญาณเอาต์พุต ( $a^3$ ) ในชั้นเอาต์พุต

$$a^3 = f(\text{net}_{s^3}) \quad (4.19)$$

## 2. แพร่ค่าย้อนกลับของค่าผิดพลาด (Back-Propagation of Error)

ในแต่ละนิวรอนของชั้นซ่อนเอาต์พุต ค่าผิดพลาดของโครงข่ายประสาทเทียมจะเป็นผลต่างระหว่างเอาต์พุตเป้าหมายกับเอาต์พุตของโครงข่ายประสาทเทียม ( $t^3 - a^3$ ) ค่าผิดพลาดที่ได้จะถูกส่งกลับเข้ามายังโครงข่ายประสาทเทียม ในรูปสัญญาณค่าผิดพลาด (Error Signal) การหาสัญญาณค่าผิดพลาดที่ชั้นเอาต์พุต หาได้จากสมการ

$$\delta^3 = (t^3 - a^3) f'(\text{net}_{s^3}) \quad (4.20)$$

ค่าถ่วงน้ำหนักที่เกี่ยวข้อง (ใช้เพื่อปรับปรุ้ค่าถ่วงน้ำหนัก  $lw_{s^3, s^2}^{3,2}$  หลังจากนี้) โดยที่  $\alpha$  คือ อัตราการเรียนรู้ (Learning Rate)

$$\Delta lw_{s^3, s^2}^{3,2} = \alpha \delta^3 a^2 \quad (4.21)$$

คำนวณค่าไบแอส (Bias) ที่เกี่ยวข้อง (ใช้เพื่อปรับปรุ้  $b^3_{s^3}$  หลังจากนี้)

$$\Delta b^3_{s^3} = \alpha \delta^3 \quad (4.22)$$

และส่ง  $\delta^3$  ไปที่ชั้นซ่อนที่ 2 โดยแต่ละนิวรอนของชั้นซ่อนที่ 2 การหาสัญญาณค่าผิดพลาดที่ชั้นซ่อนที่ 2 หาได้จากสมการ

$$\delta^2 = \sum_{s^3=1}^{s^3} \delta^3 lw_{s^3, s^2}^{3,2} f'(\text{net}_{s^2}) \quad (4.23)$$

คำนวณค่าถ่วงน้ำหนักที่เกี่ยวข้อง (ใช้เพื่อปรับปรุงค่าถ่วงน้ำหนัก  $lw_{s^2,s^1}^{2,1}$  หลังจากนั้น )

$$\Delta lw_{s^2,s^1}^{2,1} = \alpha \delta^2 a^1 \quad (4.24)$$

คำนวณค่าไบแอส (Bias) ที่เกี่ยวข้อง (ใช้เพื่อปรับปรุง  $b_{s^2}^2$  หลังจากนั้น)

$$\Delta b_{s^2}^2 = \alpha \delta^2 \quad (4.25)$$

และส่ง  $\delta^2$  ไปที่ชั้นซ่อนที่ 1 โดยแต่ละนิวรอนของชั้นซ่อนที่ 2 การหาสัญญาณค่าผิดพลาดที่ชั้นซ่อนที่ 1 หาได้จากสมการ

$$\delta^1 = \sum_{s^2=1}^{s^2} \delta^2 lw_{s^2,s^1}^{2,1} f'(net_{s^1}) \quad (4.26)$$

คำนวณค่าถ่วงน้ำหนักที่เกี่ยวข้อง (ใช้เพื่อปรับปรุงค่าถ่วงน้ำหนัก  $lw_{s^1,s^R}^{1,1}$  หลังจากนั้น )

$$\Delta lw_{s^1,s^R}^{1,1} = \alpha \delta^1 P_R \quad (4.27)$$

คำนวณค่าไบแอส (Bias) ที่เกี่ยวข้อง (ใช้เพื่อปรับปรุง  $b_{s^1}^1$  หลังจากนั้น)

$$\Delta b_{s^1}^1 = \alpha \delta^1 \quad (4.28)$$

### 3. ขั้นตอนการปรับค่าถ่วงน้ำหนัก (Weight) และค่าไบแอส (Bias)

ถ้าหากมีข้อผิดพลาดเกิดขึ้น ค่าผิดพลาดที่คำนวณได้จะถูกนำมาใช้ในการปรับน้ำหนัก เพื่อให้โครงข่ายประสาทเทียมได้ผลลัพธ์ที่ถูกต้อง การปรับน้ำหนักสำหรับแต่ละนิวรอนในชั้นเอาต์พุท คำนวณได้จาก

$$lw_{s^3,s^2}^{3,2} = lw_{s^3,s^2}^{3,2} + \Delta lw_{s^3,s^2}^{3,2} \quad (4.29)$$

ส่วนการปรับค่าถ่วงน้ำหนักของชั้นซ่อนที่ 2 การทำงานจะเป็นเช่นเดียวกันกับสมการที่ (4.29) แต่ตัวแปรต่างๆจะเป็นของชั้นซ่อนอยู่ สำหรับแต่ละนิวรอนในชั้นซ่อนที่ 2 ซึ่งสามารถหาได้จาก

$$lw_{S^2, S^1(\text{new})}^{2,1} = lw_{S^2, S^1(\text{old})}^{2,1} + \Delta lw_{S^2, S^1}^{2,1} \quad (4.30)$$

สำหรับแต่ละนิวรอนในชั้นที่ซ่อนที่ 1

$$iw_{S^1, R(\text{new})}^{1,1} = iw_{S^1, R(\text{old})}^{1,1} + \Delta iw_{S^1, R}^{1,1} \quad (4.31)$$

การปรับค่าถ่วงน้ำหนักจะปรับที่ละชั้นจากชั้นเอาต์พุตไปยังชั้นอินพุต เมื่อเสร็จสิ้นก็จะกลับสู่ส่วนของการป้อนไปข้างหน้าอีกครั้ง เพื่อหาค่าเอาต์พุตของโครงข่ายประสาทเทียมใหม่ เปรียบเทียบกับค่าเอาต์พุตเป้าหมาย และดำเนินการปรับค่าถ่วงน้ำหนักในส่วนของการแพร่ค่าย้อนกลับเช่นนี้ สลับกันไปโดยวัฏจักรของการคำนวณหาค่าความผิดพลาดและปรับน้ำหนักจะคงดำเนินต่อไปจนกระทั่งค่าความผิดพลาดต่ำกว่าที่กำหนดไว้ค่าหนึ่งแล้วจึงหยุดกระบวนการ ค่าถ่วงน้ำหนักที่ได้จะเป็นค่าที่เหมาะสมสำหรับทุกคู่ของอินพุตและค่าเป้าหมาย ในการฝึกสอนโครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ การคำนวณหาความผิดพลาดรวมของตัวอย่าง หาได้จากค่าเฉลี่ยของค่าผิดพลาดยกกำลังสอง (Mean Square Error ,MSE) ระหว่างคำตอบที่แท้จริง ( $t_i^3$ ) กับผลลัพธ์ของโครงข่ายประสาทเทียม ( $a_i^3$ ) ดังสมการ (4.32)

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i^3 - a_i^3)^2 \quad (4.32)$$

โดยที่ N = จำนวนชุดข้อมูลฝึกสอน

ในขั้นตอนการปรับน้ำหนักที่กล่าวมาข้างต้น เป็นหลักการพื้นฐานแต่ละลำดับขั้นตอนการปรับค่าถ่วงน้ำหนักและค่าไบแอสของการฝึกสอนโครงข่ายประสาทเทียมชนิด Feed – Forward Neural Network โดยใช้ Neural Network Toolbox นั้นมีหลายวิธี อาทิเช่น Levenberg- Marquard Algorithm, Resilient Back-Propagation, Conjugate Gradient เป็นต้น แต่ละวิธีจะมีประสิทธิภาพและความรวดเร็วในการฝึกสอนแตกต่างกันตามระบุใน Neural Network Toolbox: Backpropagation: Summary นั้น Scaled Conjugate Gradient ซึ่งเป็นประเภทหนึ่งของการปรับน้ำหนักแบบ Conjugate Gradient มีความเหมาะสมที่สุด กับงานประเภท Pattern Recognition ที่มี

จำนวนค่าน้ำหนักและโหนดในแต่ละชั้น จำนวนมาก โดยสามารถทำงานได้เร็วพอๆกับ Resilient Back - Propagation และบางทีก็เร็วกว่า Levenberg - Marquard Algorithm (ในโครงข่ายขนาดใหญ่) ในขณะที่ Scaled Conjugate Gradient ใช้หน่วยความจำน้อยกว่า และได้รับความถูกต้องมากกว่า ในงานวิทยานิพนธ์เล่มนี้จึงนำลำดับขั้นตอนการปรับถ่วงน้ำหนักและค่าไบแอสวิธีนี้มาใช้

#### 4. การปรับค่าถ่วงน้ำหนักและค่าไบแอสแบบคอนจูเกตเกรเดียนท์ (Conjugate Gradient)

วิธีการปรับค่าถ่วงน้ำหนักและค่าไบแอสแบบคอนจูเกตเกรเดียนท์ ถูกพัฒนามาเพื่อแก้ปัญหาการคำนวณที่ซับซ้อน จากวิธีของนิวตัน Newton มีคุณสมบัติการลดลงแบบกำลังสอง (Quadratic termination) ซึ่งต้องการการคำนวณอนุพันธ์อันดับที่ 2 (second order derivative) เพื่อหา เมทริกของเฮสเซียน (Hessian Matrix) เมื่อมีตัวแปรจำนวนมาก มันแทบจะเป็นไปไม่ได้เลยที่จะคำนวณอนุพันธ์อันดับสองนี้ได้หมด เพราะในการใช้งานจริงอาจจะต้องใช้ตัวถ่วงน้ำหนักเป็นหลายร้อยถึงหลายพันเลขก็ได้ วิธีของคอนจูเกตเกรเดียนท์นี้ จึงใช้การประมาณค่าอนุพันธ์อันดับ 2 ด้วย การคำนวณที่ใช้เพียงแค่ อนุพันธ์อันดับ 1 (Gradient)

การปรับค่าถ่วงน้ำหนักและค่าไบแอสแบบคอนจูเกตเกรเดียนท์ เป็นดังสมการ (4.33)

$$x_{k+1} = x_k + \alpha_k p_k \quad (4.33)$$

ทิศทางการค้นหาขั้นนั้นจะมี  $P_0$  เป็นจุดเริ่มต้นและเราจะหา  $P_1$  อันไหนก็ได้เพื่อที่จะทำให้ทิศทางการค้นหาตั้งฉาก (Orthogonal) กับ  $\Delta g_0$  เพราะฉะนั้นจะสามารถมี เซ็ทของ คอนจูเกตเวกเตอร์ (Conjugate Vector) อยู่มากมายนับไม่ถ้วน โดยทั่วไปทิศทางการค้นหาจะเริ่มต้นด้วยการใช้ จุดที่ ลึกที่สุด (Steepest Descent) ก่อน

$$P_0 = -g_0 \quad (4.34)$$

จากนั้นในการคำนวณแต่ละครั้งเราจะสร้างเวกเตอร์  $P_k$  ที่ตั้งฉากกับ  $\{\Delta g_0, \Delta g_1, \dots, \Delta g_{k-1}\}$  เขียนเป็นสมการอย่างง่ายได้เป็น

$$P_k = -g_k + \beta_k P_{k-1} \quad (4.35)$$

ค่า  $\beta_k$  เราสามารถหาได้จากหลายๆวิธีที่จะให้ผลลัพธ์เทียบได้กับฟังก์ชันกำลังสอง (Quadratic Function) เช่น

$$\beta_k = \frac{\Delta \mathbf{g}_{k-1}^T \mathbf{g}_k}{\Delta \mathbf{g}_{k-1}^T \mathbf{P}_{k-1}} \quad (4.36)$$

สมการของ Hestenes and Steifel

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \quad (4.37)$$

สมการของ Fletcher and Reeves

$$\beta_k = \frac{\Delta \mathbf{g}_{k-1}^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \quad (4.38)$$

สมการของ Polak and Ribiere

โดยขั้นตอนการปรับค่าถ่วงน้ำหนักสามารถทำได้ดังนี้

1. เลือกทิศทางการค้นหาทิศทางแรกจากค่าลบของ Gradient ดังสมการ (4.34)
2. เลือกอัตราการเรียนรู้ (Learning Rate) ตามสมการ (4.39) ที่จะทำให้ฟังก์ชันการหามีค่าน้อยที่สุดตามสมการ (4.33)

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{P}_k}{\mathbf{P}_k^T \mathbf{A}_k \mathbf{P}_k} \quad (4.39)$$

3. เลือกทิศทางการค้นหาต่อไปโดยคำนวณ  $\beta_k$  จากสมการ (4.36) , (4.37) หรือ(4.38)
4. ถ้าทำตามขั้นตอนดังที่กล่าวมา แล้วไม่สามารถทำให้ค่าตอบลู่เข้า (Converge) ได้ ให้กลับไปทำขั้นตอนที่ 2 ใหม่

5. การปรับค่าถ่วงน้ำหนักและค่าไบแอสวิธี สเกลคอนจูเกตเกรเดียนท์ (Scaled-Conjugate Gradient) [13]

ใน Conjugate Gradient การประมาณค่า Hessian Matrix  $\mathbf{A}_k$  นั้นทำโดยให้

$$\Delta \mathbf{P}_k = \nabla^2 F(\mathbf{x}_k) \mathbf{P}_k \approx \frac{\nabla F(\mathbf{X}_k + \sigma_k \mathbf{P}_k) - \nabla F(\mathbf{X}_k)}{\sigma_k}, 0 < \sigma_k \ll 1 \quad (4.40)$$

ความซับซ้อนของการคำนวณจาก  $O(N^2)$  จะลดลงเหลือ  $O(N)$  นั่นคือเราไม่ต้องคำนวณ Derivative อันดับที่ 2 เพียงแค่คำนวณอันดับที่ 1 ก็พอ การทำเช่นนี้เป็นการลดปริมาณหน่วยความจำที่ต้องใช้คำนวณและคิดขั้นตอนการคำนวณลงอย่างมาก

วิธีการคำนวณนี้ได้ถูกประยุกต์ใช้โดยตรงกับ Feed-Forward Neural Network แล้ว ผลการทดสอบปรากฏว่า ล้มเหลว แทบจะทุกกรณีและไม่สามารถหาค่าจุดคงที่ได้ เหตุผลก็คือ วิธีการนี้สามารถใช้ได้กับ Hessian Matrix ที่เป็นบวกเท่านั้น และการประมาณค่ากำลังสอง (Quadratic Approximation) นั้น แทบจะใช้ไม่ได้เลยกับ จุดเริ่มต้นที่อยู่ไกลจากจุดต่ำสุดที่ต้องการมาๆ

Lavenberg-Marquardt ได้แก้ปัญหานี้โดยการเพิ่มค่า  $\lambda_k$  ลงไปในสมการ (4.41) ได้เป็น

$$S_k = AP_k \approx \frac{\nabla F(X_k + \sigma_k P_k) - \nabla F(X_k)}{\sigma_k} + \lambda_k P_k \quad (4.41)$$

สำหรับการปรับค่า  $\lambda_k$  ทำได้โดยการดูที่สัญลักษณ์ของ  $\delta_k$  จากสมการ (4.42)

$$\delta_k = P_k^T \Delta P_k = P_k^T S_k \quad (4.42)$$

ถ้า  $\delta_k \leq 0$  ให้เพิ่ม  $\lambda_k$  แล้วคำนวณ  $S_k$  ใหม่อีกครั้ง แล้วเรียกว่า  $\bar{S}_k$  และ  $\lambda_k$  ที่เพิ่มขึ้นเรียกว่า  $\bar{\lambda}_k$  ดังนั้น  $\bar{S}_k$  จะเป็นไปตามสมการ

$$\bar{S}_k = S_k + (\bar{\lambda}_k - \lambda_k) P_k \quad (4.43)$$

ทฤษฎีของ Lavenberg – Marquardt ให้เพิ่ม  $\lambda_k$  แบบคงที่เมื่อไหร่ที่ Hessian ไม่เป็นค่าบวก วิธีการนี้ไม่ได้คำนวณว่าจะต้องเพิ่ม  $\lambda_k$  เท่าไหร่จึงจะไม่ทำให้  $\delta_k > 0$  แต่จริงแล้วเราสามารถหาค่า  $\lambda_k$  ที่น่าจะเป็นได้ โดยสมการ (4.44)

$$\bar{\delta}_k = P_k^T \bar{S}_k = P_k (S_k + (\bar{\lambda}_k - \lambda_k) P_k)^T > 0 = \bar{\lambda}_k > \lambda_k - \frac{\delta_k}{|P_k|^2} \quad (4.44)$$

จากสมการ (4.44) ควรจะเป็นเท่าไรถึงจะเป็นค่าที่เหมาะสม คำถามนี้ยังไม่สามารถหาคำตอบได้ แต่มั่นใจจนกว่า  $\lambda_k$  น่าจะขึ้นอยู่กับ  $\lambda_k, \delta_k$  และ  $|P_k|^2$  ตัวเลือกหนึ่งที่ดีมีเหตุผลก็คือ

$$\bar{\lambda}_k = 2\left(\lambda_k - \frac{\delta_k}{|P_k|^2}\right) \quad (4.45)$$

สมการ (4.45) นี้จะทำให้เกิด

$$\bar{\delta}_k = S_k + (\bar{\lambda}_k - \lambda_k)|P_k|^2 = S_k + (2\lambda_k - 2\frac{\delta_k}{|P_k|^2} - \lambda_k)|P_k|^2 = -\delta_k + \lambda_k|P_k|^2 > 0 \quad (4.46)$$

เพราะฉะนั้น อัตราการเรียนรู้จะได้เป็น

$$\alpha_k = \frac{-g_k^T P_k}{P_k^T S_k + \lambda_k |P_k|^2} \quad (4.47)$$

โดยค่า  $\lambda_k$  จะปรับขนาดของอัตราการเรียนรู้โดยตรง ยิ่ง  $\lambda_k$  มากอัตราการเรียนรู้จะน้อยลง ซึ่งตรงกับความต้องการในขั้นต้นที่ใส่  $\lambda_k$  ลงไปในสมการ เนื่องจาก  $\lambda_k$  ปรับขนาดของ Hessian Matrix โดยอ้อมเพราะฉะนั้นการประมาณ Quadratic Function นั้นอาจจะใช้ได้ไม่ดีมากในบางจุด ถ้าต้องการให้การประมาณนั้นใช้งานได้ดีเราจะต้องเพิ่ม  $\lambda_k$  ขึ้นเมื่อ Hessian Matrix มีค่าเป็นบวก

$$\Delta_k = \frac{\nabla F(X_k) - \nabla F(X_k + \alpha_k P_k)}{\nabla F(X_k) - \nabla F_{g^n}(\alpha_k P_k)} = \frac{2\delta_k [\nabla F(X_k + \alpha_k P_k)]}{-P_k^T g_k} \quad (4.48)$$

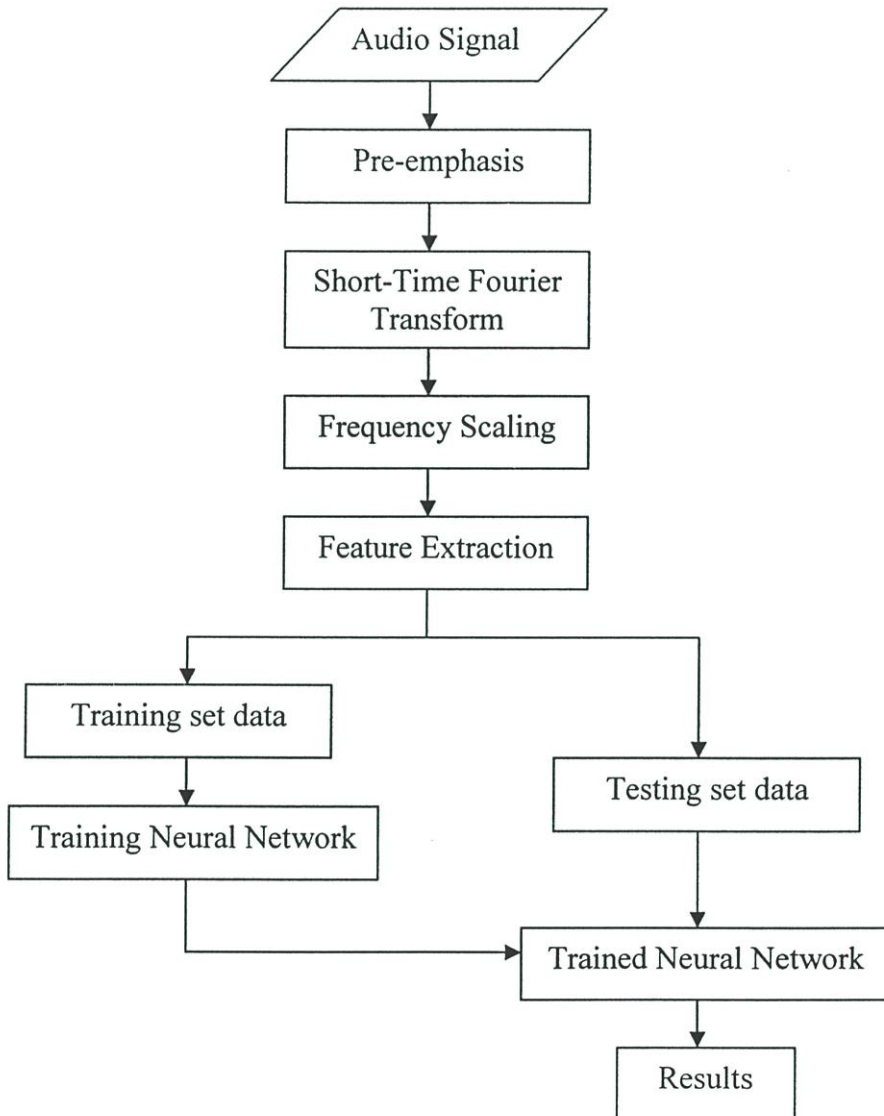
$\Delta_k$  คือตัววัดว่า  $\nabla F_{g^n}(\alpha_k P_k)$  ประมาณค่าแทน  $\nabla F(X_k + \alpha_k P_k)$  ได้ดีเพียงใด ยิ่ง  $\Delta_k$  เข้าใกล้ 1 เท่าใดก็แปลว่าการประมาณทำได้ใกล้เคียงเท่านั้น เพราะฉะนั้นเราจะได้เงื่อนไขการปรับค่า  $\lambda_k$  เพิ่มอีกโดย

$$\begin{aligned} \text{if } \Delta_k > 0.75 \text{ then } \lambda_k &= \frac{1}{2}\lambda_k \\ \text{if } \Delta_k < 0.25 \text{ then } \lambda_k &= 4\lambda_k \end{aligned} \quad (4.49)$$

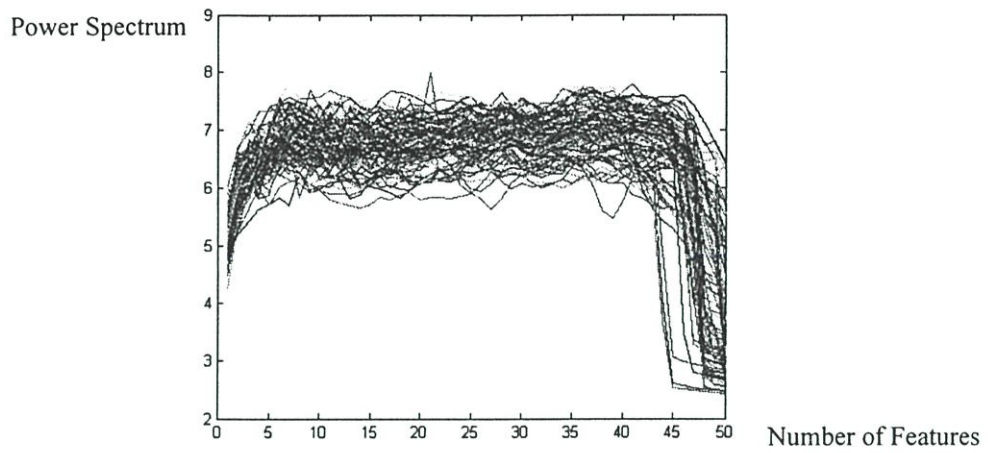
## บทที่ 5

### ขั้นตอนการทดลอง

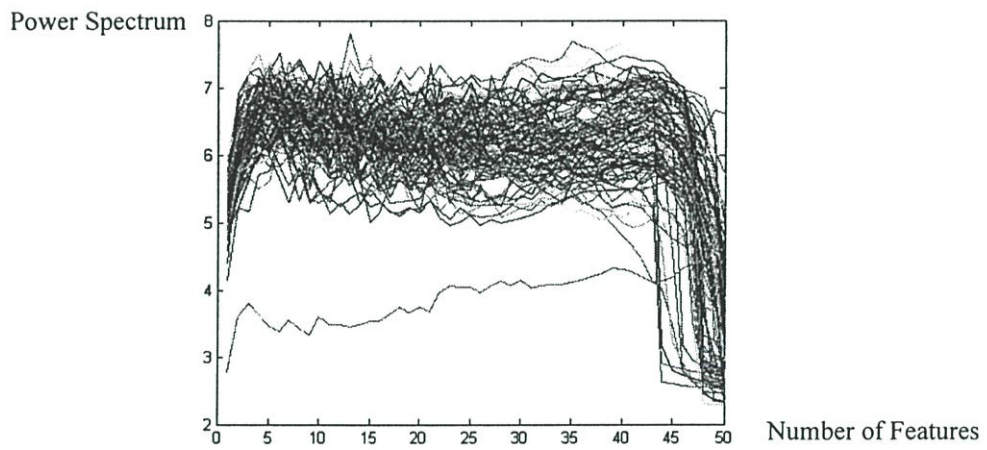
การแยกประเภทดนตรีอัตโนมัติเป็นงานวิจัยประเภทการเรียนรู้จำรูปแบบ (Pattern Recognition) ซึ่งมีขั้นตอนการทำงานดังรูป 5.1



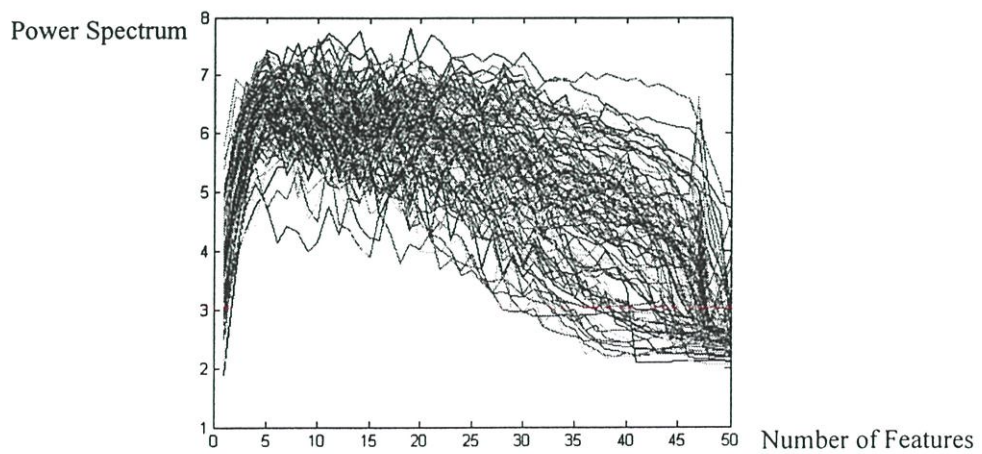
รูปที่ 5.1 ขั้นตอนการแยกประเภทดนตรีด้วยโครงข่ายประสาทเทียม



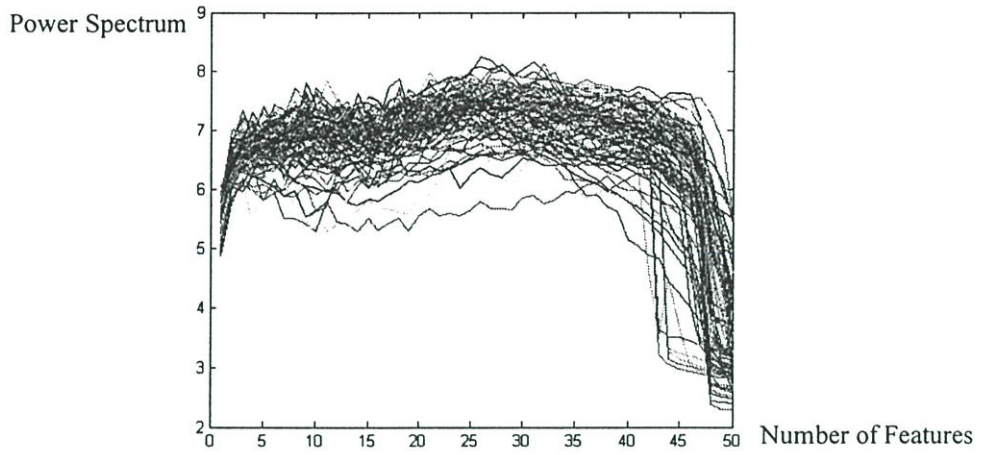
รูปที่ 5.2 ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Hip-Hop 100 เพลง



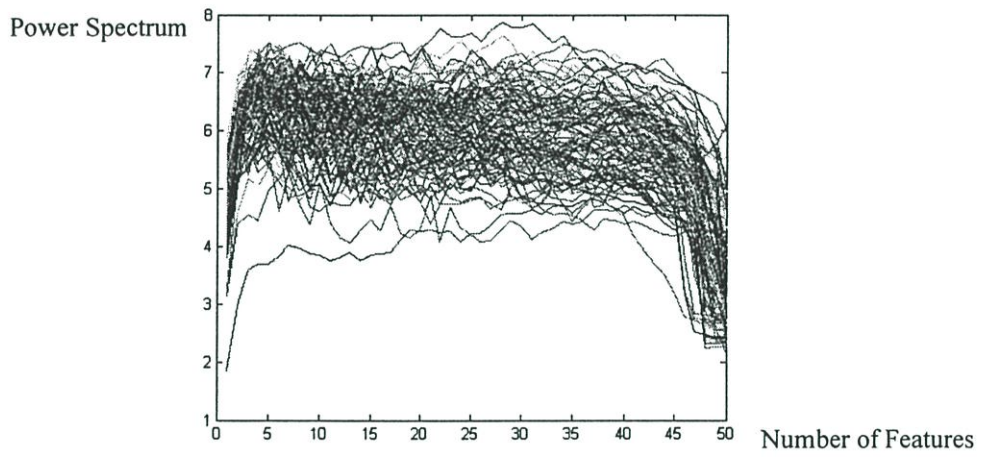
รูปที่ 5.3 ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Jazz 100 เพลง



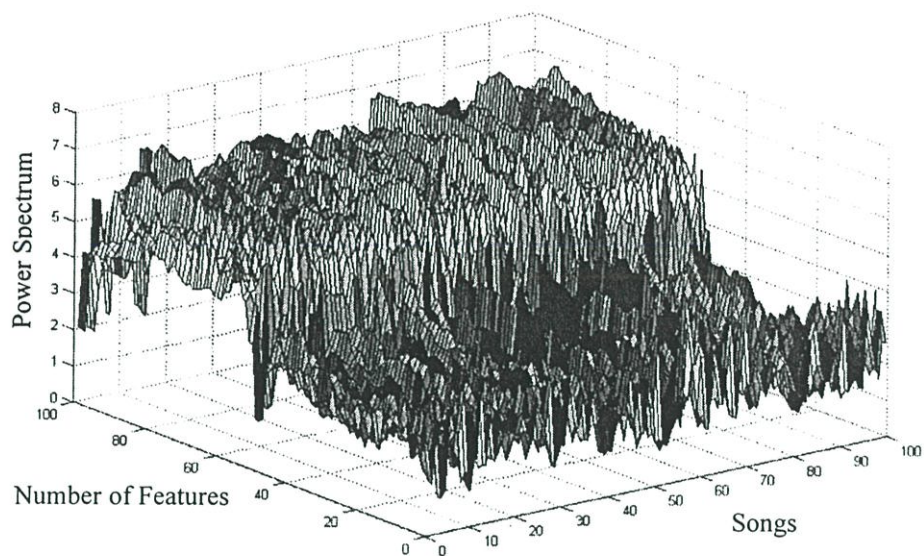
รูปที่ 5.4 ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Classic 100 เพลง



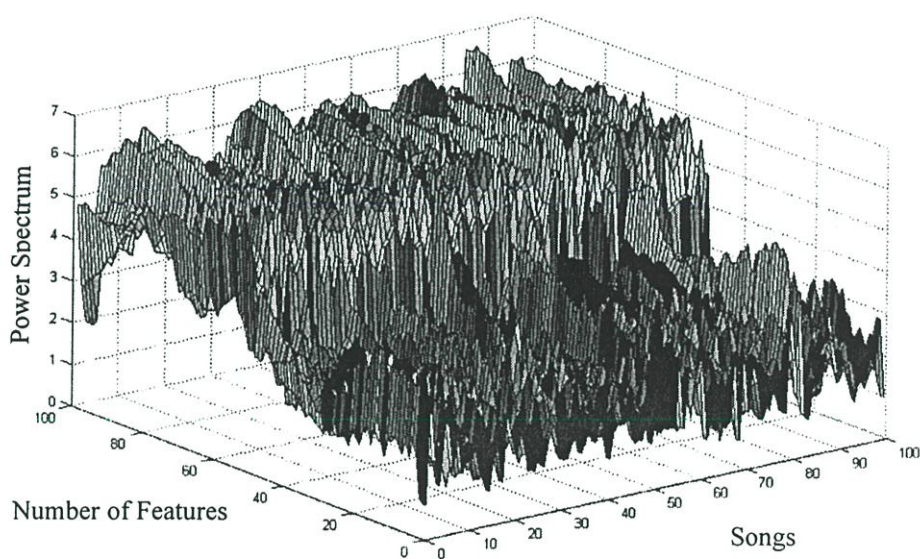
รูปที่ 5.5 ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Rock 100 เพลง



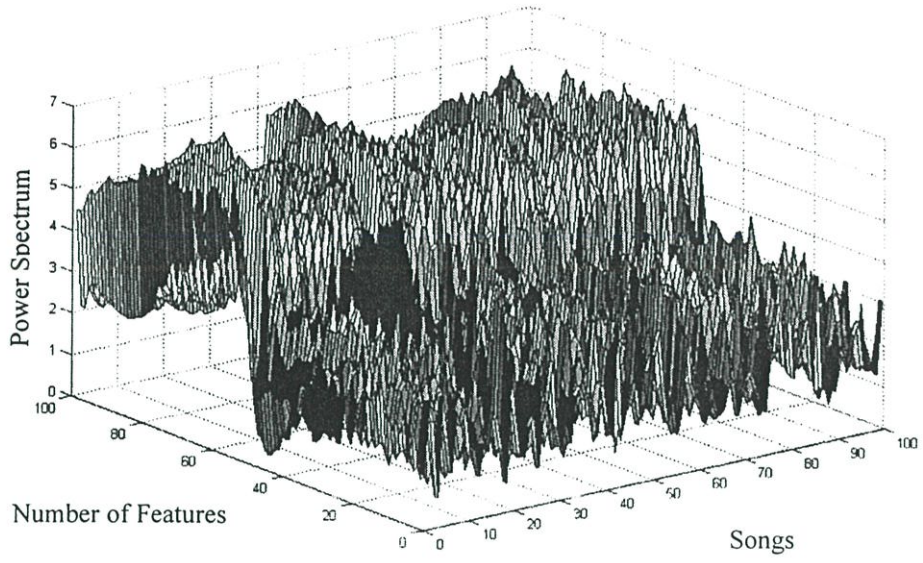
รูปที่ 5.6 ตัวอย่าง Mel-Frequency Spectral Mean ของเพลง Folk 100 เพลง



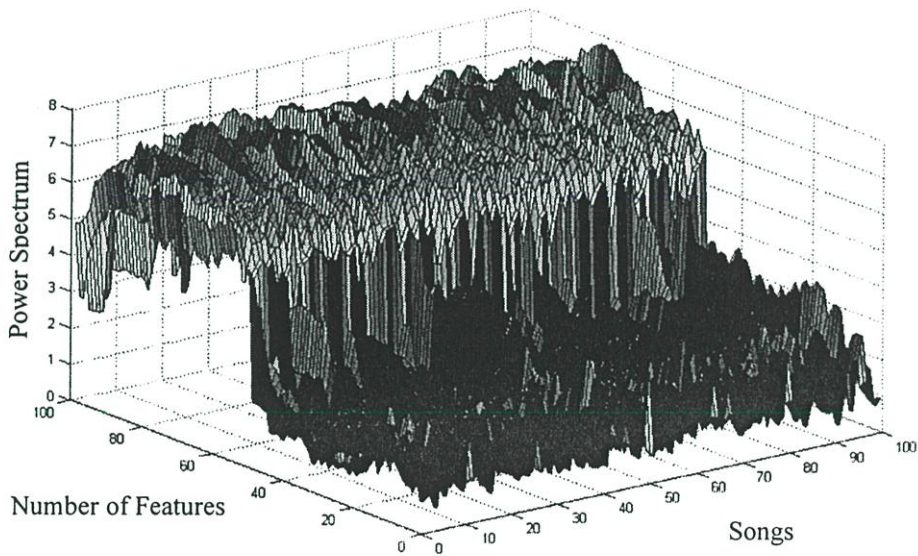
รูปที่ 5.7 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Hip-Hop 100 เพลง



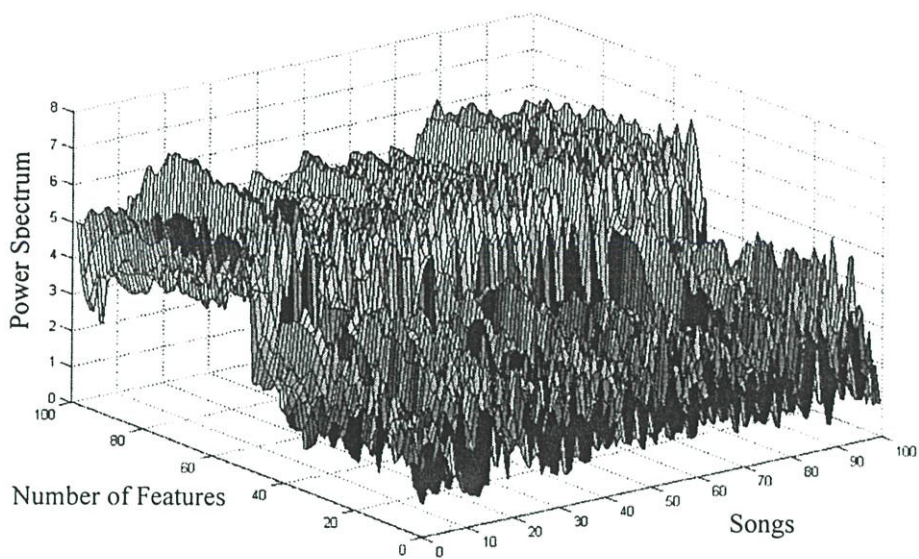
รูปที่ 5.8 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Jazz 100 เพลง



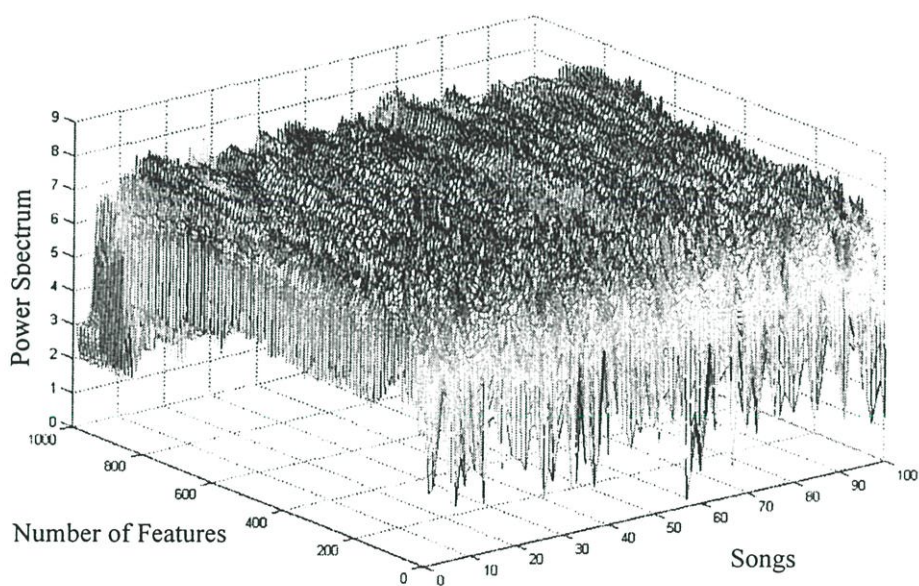
รูปที่ 5.9 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Classic 100 เพลง



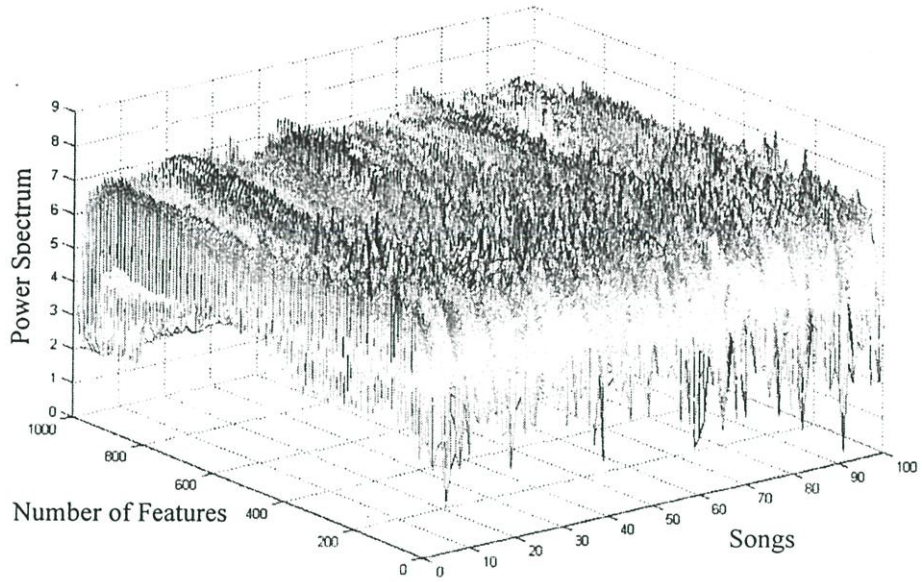
รูปที่ 5.10 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Rock 100 เพลง



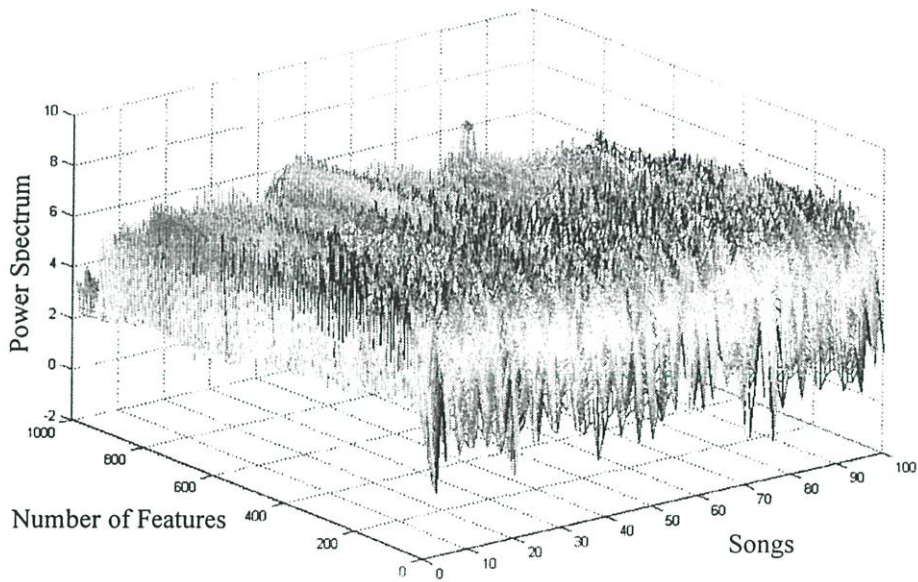
รูปที่ 5.11 ตัวอย่าง Mel-Frequency Spectral Contrast ของเพลง Folk 100 เพลง



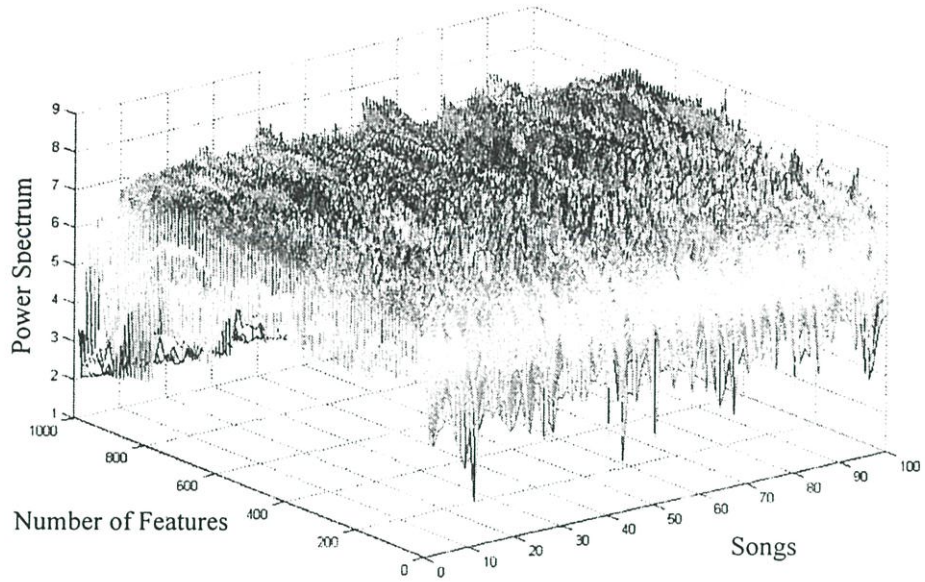
รูปที่ 5.12 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Hip-Hop 100 เพลง



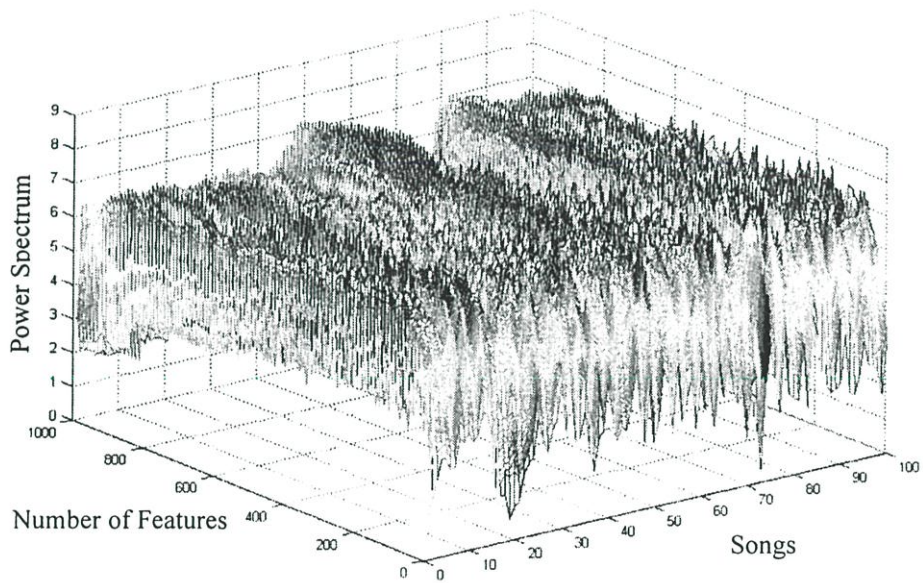
รูปที่ 5.13 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Jazz 100 เพลง



รูปที่ 5.14 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Classic 100 เพลง



รูปที่ 5.15 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Rock 100 เพลง



รูปที่ 5.16 ตัวอย่าง Mel-Frequency Spectral Histogram ของเพลง Folk 100 เพลง

## 5.1 การตั้งค่าสำหรับการทดลอง

### 5.1.1 สัญญาณเสียง (Audio Signal)

ในการทดลองจะใช้ Audio Signal จากไฟล์นามสกุล WAV.

ซึ่งมีองค์ประกอบต่างๆดังนี้

อัตราการสุ่มตัวอย่าง Sampling Rate = 44100 Hz

ช่องสัญญาณ Channel = Mono

ความละเอียด Resolution = 16 Bits

### 5.1.2 ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ

เพลงที่ใช้ในการทดลองทั้งหมด 1,000 เพลง แบ่งเป็น 5 ประเภทแนวดนตรี ได้แก่ Classical, Hip-Hop, Jazz, Rock และ Folk ประเภทละ 200 เพลง โดยใช้เป็นข้อมูลฝึกสอน 100 เพลง (50%) และชุดข้อมูลทดสอบอีก 100 เพลง (50%) โดยเพลงทั้งหมดนี้ได้มาจากเว็บไซต์ [www.download.com](http://www.download.com) และ [www.amazon.com](http://www.amazon.com) ตามประเภทดนตรีหลักที่ต้องการ ซึ่งประกอบด้วยประเภทดนตรีย่อยหลากหลายประเภท

โดยตัดช่วงข้อมูลเพลงออกมา 5 ท่อน ท่อนละ 10 วินาที ที่ 15%, 30%, 45%, 60%, และ 75% ของความยาวเพลง เพื่อใช้ในการหาลักษณะเด่นของเพลง 1 เพลง และทดลองใช้ข้อมูลเพลงทั้งเพลงกับตัวแยกประเภทที่มีประสิทธิภาพสูงสุด

### 5.1.3 การเน้นน้ำหนักเสียงก่อน (Pre-emphasis)

โดยใช้ค่า Pre-Emphasis Factor = 0.7 เพื่อให้ได้ Cut-Off Frequency ที่ 2500 Hz

### 5.1.4 การแปลงฟูริเยร์ในช่วงเวลาสั้นๆ (Short Time Fourier Transfer)

ใช้ขนาดของหน้าต่าง (Window Size) = 0.2 Sec

ขนาดการเลื่อนหน้าต่าง (Hop Size) = 0.1 Sec.

ชนิดการทำหน้าต่าง (Windowing) = Hamming Window

### 5.1.5 โครงข่ายประสาทเทียม (Neural Network)

ชนิดของโครงข่ายประสาทเทียม = Feed Forward Back-Propagation

ชนิดของการเรียนรู้ = Scaled Conjugate Gradient

จำนวน Node ที่ชั้น Input ขึ้นอยู่กับลักษณะเด่นแต่ละชนิด

จำนวน Node ที่ชั้น Output = 5 Nodes

Function ของชั้นต่างๆ = Tan Sigmoid

โดยให้ Output เป็นประเภทคนตรี 5 ประเภทและแต่ละ Node จะให้ค่าเป้าหมายเป็น 1 เมื่อระบบถูกฝึกสอนด้วยประเภทคนตรีนั้นๆ และให้ค่าเป้าหมายเป็น 0 เมื่อระบบถูกฝึกสอนด้วยประเภทคนตรีที่ไม่ตรงกับ Output นั้น

## 5.2 ตัวแปรต้นและตัวแปรตาม

การทดลองนี้ มุ่งศึกษาลักษณะเด่นที่จะนำมาใช้เพื่อแยกประเภทคนตรีให้มีประสิทธิภาพ โดยใช้ตัวแปรต้นดังนี้

### 5.2.1 ตัวแปรต้น

#### 1. Frequency Scaling

- Mel-Frequency Scaling (50 Mels)
- Octave –Based Scaling (6 Octaves)

#### 2. Feature Extraction

- Spectral Mean
- Spectral Contrast
- Spectral Histogram (10 Frames of sampling)

3. จำนวนชั้นซ่อนของโครงข่ายประสาทเทียม
4. จำนวน Node ของโครงข่ายประสาทเทียมในแต่ละชั้น
5. จำนวนรอบของการฝึกสอน โครงข่ายประสาทเทียม

### 5.2.2 ตัวแปรตาม

การวัดผลการทดลองใช้ ตัวแปรตามดังนี้

1. ความถูกต้องในการแยกประเภทคนตรีจากชุดข้อมูลฝึกสอน
2. ความถูกต้องในการแยกประเภทคนตรีจากชุดข้อมูลทดสอบ

## บทที่ 6

### ผลการทดลอง

ผลการทดลองทั้งหมด เป็นผลการทดลองจากการเปลี่ยนตัวแปรต้นคือ การปรับสัดส่วนทางความถี่ ลักษณะเด่นต่างๆ จำนวนชั้นของโครงข่ายประสาทเทียม จำนวนโหนดของโครงข่ายประสาทเทียม และจำนวนครั้งในการฝึกสอน โครงข่ายประสาทเทียม โดยค่าต่างๆอยู่ในช่วงที่ให้ประสิทธิภาพในการแยกประเภทดนตรีที่น่าสนใจ

ในตารางที่ 6.1 ถึง 6.37 เป็นข้อมูลดิบซึ่งได้จากการทดลองซ้ำ 4 ครั้ง โดยตารางที่ 6.37 เป็นผลจากการทดลองใช้ข้อมูลฝึกสอนที่ได้จากเพลงชุดฝึกสอนทั้งเพลง ในตารางที่ 6.38 คือตารางเปรียบเทียบการแยกประเภทโดยอิงจากการปรับสัดส่วนทางความถี่และลักษณะเด่นประเภทต่างๆ ในตารางที่ 6.39 คือการจัดอันดับประสิทธิภาพในการแยกประเภทดนตรีสูงสุด 3 อันดับแรก โดยใช้ตัวแปรต้นต่างๆตามตารางที่ 6.1 ถึง 6.36

ในรูปที่ 6.1 ถึง 6.6 เป็นกราฟแสดงการเปรียบเทียบความสามารถในการแยกประเภทดนตรีในแนวเพลงต่างๆ และเปรียบเทียบผลจากการเพิ่มจำนวนชั้นของโครงข่ายประสาทเทียม จำนวนโหนดของโครงข่ายประสาทเทียม และจำนวนครั้งในการฝึกสอน

ตารางที่ 6.1 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง 24 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 12 จำนวนโหนดในชั้นที่ 2 = 24																
จำนวนรอบในการฝึกสอน		50					100					300				
		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	82.2	0.93	0.93	13.1	2.8	66.1	8.26	1.83	22	1.83	72.1	5.41	1.8	17.1	3.6
	Jazz	52.6	7.76	25.9	3.45	10.3	4.46	42	26.8	17.9	8.93	16.7	19.3	13.2	3.51	47.4
	Classic	3.7	4.63	82.4	2.78	6.48	0	3.85	93.3	0.96	1.92	0	2.8	86.9	1.87	8.41
	Rock	55.7	0	0	40	4.35	69.2	8.41	0	21.5	0.93	17.8	4.67	0	72	5.61
	Folk	30.5	1.9	36.2	5.71	25.7	0.87	22.6	48.7	10.4	17.4	2.97	7.92	11.9	1.98	75.2
	Average	47.62					48.04					65.10				
Testing Set Data	HipHop	83.7	1.92	0	14.4	0	71.8	3.64	1.82	22.7	0	70.4	3.7	0	22.2	3.7
	Jazz	46.5	6.14	32.5	3.51	11.4	4.27	35.9	37.6	17.9	4.27	18.5	21	13.4	2.52	44.5
	Classic	5.45	5.45	78.2	1.82	9.09	1.94	1.94	90.3	1.94	3.88	2.8	0.93	85	0.93	10.3
	Rock	63.9	0	0	34.3	1.85	73.7	3.39	0	22.9	0	13.9	4.63	0.93	77.8	2.78
	Folk	33.6	0.93	32.7	1.87	30.8	0.89	28.6	50.9	6.25	13.4	1.96	9.8	11.8	0	76.5
	Average	46.62					46.86					66.13				

ตารางที่ 6.2 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 18 และชั้นที่สอง 36 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 18 จำนวนโหนดในชั้นที่ 2 = 36																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	74.6	4.39	0.88	19.3	0.88	73.6	7.27	0.91	9.09	9.09	69.4	8.33	0.93	20.4	0.93
	Jazz	10.6	15.4	34.6	24	15.4	16.8	22.1	23	7.96	30.1	13.5	35.1	9.91	10.8	30.6
	Classic	0	0	91.3	2.88	5.77	0.95	1.9	85.7	0.95	10.5	0	0.93	90.7	2.78	5.56
	Rock	7.84	2.94	0	83.3	5.88	11	7.34	0	79.8	1.83	13.5	3.85	0	78.8	3.85
	Folk	0.93	5.56	44.4	23.1	25.9	7.55	15.1	28.3	5.66	43.4	1.87	18.7	7.48	3.74	68.2
	Average	58.11					60.94					68.48				
Testing Set Data	HipHop	75.9	2.68	0	18.8	2.68	72.3	3.57	1.79	16.1	6.25	76.6	5.41	1.8	14.4	1.8
	Jazz	12.6	16.8	38.7	18.5	13.4	13	23.5	27.8	6.09	29.6	14	27.2	13.2	5.26	40.4
	Classic	0.93	0	92.5	1.87	4.67	0.94	0.94	82.1	1.89	14.2	2.88	0.96	83.7	1.92	10.6
	Rock	9	2	0	87	2	11.2	0.93	0	86	1.87	8.26	3.67	0.92	83.5	3.67
	Folk	1.72	3.45	48.3	18.1	28.4	9.62	16.3	26.9	1.92	45.2	2.83	16	9.43	3.77	67.9
	Average	60.13					61.81					67.77				

ตารางที่ 6.3 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean  
และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง  
120 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 60 จำนวนโหนดในชั้นที่ 2 = 120																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	77.1	7.34	1.83	10.1	3.67	73.9	9.01	0	16.2	0.9	79.4	5.61	0	14	0.93
	Jazz	10.3	51.7	19	12.1	6.9	14.8	33.3	13	8.33	30.6	13.5	37.8	9.01	13.5	26.1
	Classic	0	6.73	88.5	3.85	0.96	0	0.96	89.4	2.88	6.73	0	1.9	89.5	1.9	6.67
	Rock	20.6	8.41	0	71	0	18.3	6.73	0	72.1	2.88	14.2	1.89	0	80.2	3.77
	Folk	4.72	43.4	17.9	10.4	23.6	4.85	17.5	10.7	9.71	57.3	2.94	19.6	7.84	5.88	63.7
	Average	62.37					65.21					70.14				
Testing Set Data	HipHop	77.9	7.08	0.88	13.3	0.88	76.6	4.67	0.93	15.9	1.87	78.5	5.61	1.87	13.1	0.93
	Jazz	13.1	53.3	16.8	8.41	8.41	15.3	28.8	13.5	5.41	36.9	17.9	29.5	13.4	8.04	31.3
	Classic	1.89	7.55	86.8	0.94	2.83	2.86	0	85.7	2.86	8.57	2.8	2.8	83.2	2.8	8.41
	Rock	19.6	3.74	0	74.8	1.87	14.6	3.88	0	78.6	2.91	9.62	2.88	0	84.6	2.88
	Folk	3.64	42.7	18.2	5.45	30	2.75	11.9	11	4.59	69.7	1.9	15.2	7.62	3.81	71.4
	Average	64.54					67.91					69.44				

ตารางที่ 6.4 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 24																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	48.6	1.8	0	49.5	0	79	4.76	1.9	13.3	0.95	75.9	1.85	0.93	15.7	5.56
	Jazz	22.9	17.4	38.5	15.6	5.5	12.4	14.9	14.9	14	43.8	21.7	9.57	9.57	5.22	53.9
	Classic	0	8.41	85	2.8	3.74	0	0	92.6	1.85	5.56	0	0	89.4	0	10.6
	Rock	9.35	2.8	0.93	85	1.87	17.9	4.72	0	75.5	1.89	18.9	1.89	0	72.6	6.6
	Folk	6.03	16.4	49.1	6.9	21.6	4.59	6.42	33.9	6.42	48.6	4.85	5.83	10.7	2.91	75.7
	Average	51.55					62.12					64.66				
Testing Set Data	HipHop	48.6	0.9	4.5	45.9	0	78.7	3.7	0.93	14.8	1.85	77.5	3.6	0.9	17.1	0.9
	Jazz	21.1	21.1	43	12.3	2.63	14	12.3	16.7	7.89	49.1	21.7	10.8	13.3	1.67	52.5
	Classic	1.89	8.49	85.8	2.83	0.94	1.87	0.93	87.9	2.8	6.54	2.8	0	83.2	0.93	13.1
	Rock	5.88	0.98	0.98	92.2	0	13.9	6.48	0.93	77.8	0.93	15.4	1.92	0	77.9	4.81
	Folk	5.26	12.3	59.6	7.02	15.8	2.78	4.63	26.9	6.48	59.3	3.81	0.95	8.57	2.86	83.8
	Average	52.70					63.17					66.64				

ตารางที่ 6.5 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 36 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 36																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	37.5	3.57	3.57	48.2	7.14	65.2	4.46	2.68	22.3	5.36	79.6	2.78	0.93	13.9	2.78
	Jazz	6.31	36	22.5	12.6	22.5	15.8	36	20.2	10.5	17.5	16.8	22.7	8.4	10.9	41.2
	Classic	0	6.73	88.5	1.92	2.88	0	1.89	92.5	2.83	2.83	0	1.85	88	0.93	9.26
	Rock	2.91	0.97	0.97	89.3	5.83	19.2	0.96	0	76	3.85	14.4	0.96	0	80.8	3.85
	Folk	4.42	38.9	23	3.54	30.1	3.67	22	22	6.42	45.9	2.88	7.69	7.69	2.88	78.8
	Average	56.28					63.09					69.98				
Testing Set Data	HipHop	50	4.46	1.79	40.2	3.57	76.4	1.89	0.94	18.9	1.89	76.8	1.79	0	16.1	5.36
	Jazz	10.6	45.1	16.8	7.08	20.4	12.3	42.1	17.5	5.26	22.8	20	18.3	13	4.35	44.3
	Classic	0.94	10.4	82.1	2.83	3.77	1.87	5.61	88.8	2.8	0.93	2.86	2.86	84.8	0	9.52
	Rock	3.74	1.87	0.93	88.8	4.67	14.6	3.88	0	79.6	1.94	9.35	1.87	0	82.2	6.54
	Folk	3.54	41.6	18.6	4.42	31.9	4.85	24.3	17.5	4.85	48.5	2.91	6.8	9.71	2.91	77.7
	Average	59.57					67.09					67.94				

ตารางที่ 6.6 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 120																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	65.8	5.41	0.9	24.3	3.6	67.6	8.11	0	21.6	2.7	81.5	6.48	0	11.1	0.93
	Jazz	14.4	29.7	18.9	7.21	29.7	5.61	41.1	10.3	10.3	32.7	9.91	44.1	7.21	10.8	27.9
	Classic	0.94	1.89	93.4	1.89	1.89	0	2.94	91.2	1.96	3.92	0	2.91	89.3	0.97	6.8
	Rock	15.1	0.94	1.89	75.5	6.6	6.86	1.96	0	87.3	3.92	11.7	1.94	0	83.5	2.91
	Folk	11.2	12.1	22.4	3.74	50.5	2.86	16.2	15.2	3.81	61.9	6.42	15.6	9.17	3.67	65.1
	Average	62.97					69.81					72.72				
Testing Set Data	HipHop	71.6	2.75	0	23.9	1.83	75	4.63	0	16.7	3.7	79	1.9	1.9	13.3	3.81
	Jazz	23.7	28.1	14.9	3.51	29.8	12.1	31	15.5	5.17	36.2	18	41.4	12.6	3.6	24.3
	Classic	0.93	4.63	89.8	1.85	2.78	0.93	3.74	86	1.87	7.48	1.92	4.81	86.5	0.96	5.77
	Rock	11.4	1.9	0.95	80	5.71	8.41	1.87	0	85	4.67	10.6	0.96	0	84.6	3.85
	Folk	11.2	11.2	17.8	1.87	57.9	2.88	17.3	12.5	4.81	62.5	2.94	18.6	10.8	2.94	64.7
	Average	65.48					67.91					71.27				

ตารางที่ 6.7 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean  
และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 48 และชั้นที่สอง  
96 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 48 จำนวนโหนดในชั้นที่ 2 = 96																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	56	11.2	2.59	26.7	3.45	60.9	14.5	0.91	21.8	1.82	76.1	12.4	2.65	6.19	2.65
	Jazz	17	56.3	10.7	6.25	9.82	10.7	66.1	5.36	5.36	12.5	12.8	71.6	3.67	2.75	9.17
	Classic	0	2.88	90.4	1.92	4.81	0	0.99	89.1	1.98	7.92	0	0.97	91.3	0.97	6.8
	Rock	17.3	4.81	0	75	2.88	5.66	1.89	0	88.7	3.77	5.88	3.92	0	89.2	0.98
	Folk	6.84	36.8	12.8	9.4	34.2	5.66	10.4	4.72	3.77	75.5	4.95	6.93	3.96	1.98	82.2
	Average	62.37					76.05					82.06				
Testing Set Data	HipHop	55.9	9.91	0.9	31.5	1.8	58.2	13.6	0	25.5	2.73	75.7	9.91	0.9	9.91	3.6
	Jazz	17.6	51.9	13.9	5.56	11.1	12.1	56.1	5.61	5.61	20.6	6.31	68.5	5.41	5.41	14.4
	Classic	2.94	0.98	90.2	2.94	2.94	3.85	0.96	86.5	2.88	5.77	1.85	2.78	89.8	1.85	3.7
	Rock	11.7	3.88	0	81.6	2.91	7.77	0.97	0.97	86.4	3.88	7.69	1.92	1.92	87.5	0.96
	Folk	3.88	38.8	16.5	5.83	35	0	11.5	5.77	2.88	79.8	2.91	10.7	4.85	0.97	80.6
	Average	62.88					73.40					80.41				

ตารางที่ 6.8 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง 192  
โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 96 จำนวนโหนดในชั้นที่ 2 = 192																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	48.3	16.4	1.72	31	2.59	74.8	9.01	0	12.6	3.6	82.1	11.3	0	4.72	1.89
	Jazz	12	63.9	7.41	7.41	9.26	12.7	62.7	5.45	8.18	10.9	13.1	72.9	3.74	0.93	9.35
	Classic	0.97	0.97	95.1	0.97	1.94	1.87	0	89.7	2.8	5.61	0	0.98	92.2	0.98	5.88
	Rock	6.93	1.98	0.99	85.1	4.95	6.73	1.92	0	87.5	3.85	5.66	2.83	0	88.7	2.83
	Folk	4.67	35.5	11.2	8.41	40.2	8.41	11.2	4.67	2.8	72.9	4.81	4.81	3.85	1.92	84.6
	Average	66.53					77.52					84.08				
Testing Set Data	HipHop	51.4	14.7	0.92	30.3	2.75	65.5	11.5	2.65	18.6	1.77	78	7.34	0.92	12.8	0.92
	Jazz	8.41	62.6	9.35	6.54	13.1	9.43	55.7	5.66	7.55	21.7	7.41	63.9	6.48	5.56	16.7
	Classic	2.78	2.78	88.9	2.78	2.78	1.89	1.89	87.7	0.94	7.55	1.92	1.92	87.5	1.92	6.73
	Rock	4.81	1.92	3.85	87.5	1.92	9.62	0	0.96	85.6	3.85	7.92	0	0.99	88.1	2.97
	Folk	2.73	30.9	14.5	1.82	50	0.93	15.7	8.33	2.78	72.2	2.91	7.77	4.85	0.97	83.5
	Average	68.08					73.34					80.20				

ตารางที่ 6.9 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384  
โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 192 จำนวนโหนดในชั้นที่ 2 = 384																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	69.6	14.3	1.79	13.4	0.89	76.5	9.8	1.96	8.82	2.94	81.7	8.26	0.92	6.42	2.75
	Jazz	13	60.2	4.63	8.33	13.9	16.1	69.6	3.57	2.68	8.04	10.2	78.7	2.78	2.78	5.56
	Classic	0.95	0.95	89.5	1.9	6.67	0.95	0.95	91.4	1.9	4.76	0	0.96	94.2	0.96	3.85
	Rock	7.62	1.9	0	86.7	3.81	4.81	3.85	0	90.4	0.96	5	3	0	91	1
	Folk	12.6	18.4	8.74	3.88	56.3	9.43	12.3	3.77	1.89	72.6	5.94	1.98	2.97	1.98	87.1
	Average	72.47					80.11					86.54				
Testing Set Data	HipHop	67.9	9.82	1.79	18.8	1.79	69.5	9.52	0.95	20	0	72.6	10.4	0.94	12.3	3.77
	Jazz	12.4	57.5	4.42	7.08	18.6	8.33	63.9	6.48	6.48	14.8	5.61	65.4	5.61	6.54	16.8
	Classic	2.83	0	89.6	0	7.55	1.9	1.9	88.6	0.95	6.67	0.94	2.83	89.6	0.94	5.66
	Rock	7.62	1.9	1.9	87.6	0.95	6.6	0	0.94	88.7	3.77	7.84	0.98	0.98	88.2	1.96
	Folk	1.89	13.2	10.4	0.94	73.6	2.88	12.5	6.73	0.96	76.9	2.94	9.8	6.86	0	80.4
	Average	75.24					77.52					79.26				

ตารางที่ 6.10 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 96 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 96																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	48.2	17.5	2.63	30.7	0.88	54.2	15	1.87	27.1	1.87	78.1	12.4	0	6.67	2.86
	Jazz	6.31	62.2	9.01	7.21	15.3	7.41	69.4	7.41	5.56	10.2	12.6	73	2.7	2.7	9.01
	Classic	1.96	2.94	89.2	0.98	4.9	2.78	4.63	88	0.93	3.7	0	0.93	91.7	1.85	5.56
	Rock	3.96	1.98	0.99	90.1	2.97	2.8	3.74	0	91.6	1.87	7.62	3.81	0	86.7	1.9
	Folk	1.94	13.6	6.8	8.74	68.9	9.52	19	9.52	3.81	58.1	4.76	8.57	2.86	1.9	81.9
	Average	71.73					72.26					82.26				
Testing Set Data	HipHop	48.7	19.1	0.87	29.6	1.74	52.7	16.1	0.89	28.6	1.79	75	5.56	0.93	14.8	3.7
	Jazz	10.7	50.5	5.83	5.83	27.2	4.67	65.4	7.48	6.54	15.9	8.04	65.2	6.25	7.14	13.4
	Classic	1.96	1.96	84.3	3.92	7.84	1.89	3.77	90.6	0	3.77	0.96	2.88	88.5	0	7.69
	Rock	3.88	0.97	0.97	89.3	4.85	3.85	0.96	0.96	90.4	3.85	9.71	0	1.94	84.5	3.88
	Folk	0.95	6.67	6.67	2.86	82.9	2.88	20.2	8.65	0.96	67.3	2.83	12.3	6.6	0	78.3
	Average	71.13					73.27					78.28				

ตารางที่ 6.11 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 192																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	60.4	10.4	2.83	23.6	2.83	63.2	15.8	1.75	15.8	3.51	85	8.41	0.93	4.67	0.93
	Jazz	13.8	59.6	12.8	7.34	6.42	16.3	64.4	5.77	3.85	9.62	13.6	74.8	2.91	1.94	6.8
	Classic	2.8	0.93	91.6	1.87	2.8	0	0	90.6	1.89	7.55	0	1.96	89.2	0.98	7.84
	Rock	4.81	2.88	0	90.4	1.92	5	3	0	91	1	5.71	4.76	0	88.6	0.95
	Folk	7.34	20.2	9.17	7.34	56	6.8	10.7	4.85	1.94	75.7	6.67	1.9	2.86	1.9	86.7
	Average	71.59					76.98					84.85				
Testing Set Data	HipHop	53.9	13	0.87	29.6	2.61	63.5	9.62	1.92	24	0.96	77.1	7.34	0	13.8	1.83
	Jazz	9.17	57.8	6.42	8.26	18.3	9.09	68.2	5.45	6.36	10.9	7.48	65.4	4.67	7.48	15
	Classic	2.86	0.95	90.5	0.95	4.76	0.93	3.74	88.8	1.87	4.67	2.75	1.83	88.1	0.92	6.42
	Rock	6.93	0.99	0.99	90.1	0.99	4.72	0.94	0	91.5	2.83	7.77	0	1.94	87.4	2.91
	Folk	3.88	25.2	9.71	1.94	59.2	3.96	15.8	3.96	0	76.2	2.86	11.4	2.86	0.95	81.9
	Average	70.30					77.64					79.97				

ตารางที่ 6.12 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Mean และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 384																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	67	17.4	1.83	12.8	0.92	73.3	9.52	1.9	12.4	2.86	80.9	9.09	1.82	4.55	3.64
	Jazz	7.62	66.7	8.57	8.57	8.57	12	71.3	3.7	3.7	9.26	11.3	75.5	1.89	1.89	9.43
	Classic	0	1.9	91.4	1.9	4.76	0	0.93	90.7	2.78	5.56	0	0.94	91.5	2.83	4.72
	Rock	1.9	3.81	0	92.4	1.9	3.92	2.94	0	91.2	1.96	4.72	2.83	0	89.6	2.83
	Folk	9.17	7.34	7.34	5.5	70.6	6.8	9.71	4.85	1.94	76.7	3.96	3.96	2.97	1.98	87.1
	Average	77.62					80.65					84.93				
Testing Set Data	HipHop	55.9	16.2	0	27.9	0	70.3	8.11	0	18.9	2.7	75.9	8.33	0.93	13.9	0.93
	Jazz	5.77	61.5	6.73	8.65	17.3	8.57	56.2	6.67	6.67	21.9	5.66	66	3.77	7.55	17
	Classic	1.9	0.95	89.5	1.9	5.71	1.85	0.93	90.7	1.85	4.63	0.94	1.89	89.6	1.89	5.66
	Rock	6.54	0	0.93	91.6	0.93	5.61	1.87	0.93	89.7	1.87	8.74	0	1.94	87.4	1.94
	Folk	1.94	14.6	7.77	1.94	73.8	0.98	12.7	7.84	0.98	77.5	2.97	9.9	3.96	1.98	81.2
	Average	74.46					76.87					80.03				

ตารางที่ 6.13 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง  
24 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 12 จำนวนโหนดในชั้นที่ 2 = 24																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	53.3	27.1	1.87	13.1	4.67	81.7	4.59	0.92	4.59	8.26	82.3	8.85	1.77	0.88	6.19
	Jazz	22.9	42.4	4.24	16.9	13.6	32.4	13.5	10.8	3.6	39.6	3.7	88	0.93	2.78	4.63
	Classic	1.94	1.94	86.4	3.88	5.83	1.94	0.97	90.3	1.94	4.85	0	5.71	91.4	0.95	1.9
	Rock	7.48	3.74	2.8	84.1	1.87	6.54	0	0.93	85	7.48	2.91	1.94	0.97	90.3	3.88
	Folk	6.36	27.3	10	10	46.4	0.96	2.88	8.65	6.73	80.8	2.88	10.6	0	2.88	83.7
	Average	62.51					70.25					87.13				
Testing Set Data	HipHop	67.3	21.8	0.91	9.09	0.91	77.1	4.59	0.92	8.26	9.17	69.4	14.8	0	10.2	5.56
	Jazz	32.1	39.6	6.6	7.55	14.2	24.4	13.4	13.4	10.1	38.7	2.88	70.2	6.73	3.85	16.3
	Classic	4.63	4.63	79.6	6.48	4.63	3.7	2.78	84.3	2.78	6.48	1.9	3.81	90.5	1.9	1.9
	Rock	8.41	7.48	0	81.3	2.8	9.8	0	0.98	82.4	6.86	8.91	5.94	0	82.2	2.97
	Folk	4.63	34.3	16.7	10.2	34.3	2.78	7.41	8.33	3.7	77.8	3.7	22.2	4.63	2.78	66.7
	Average	60.42					66.98					75.79				

ตารางที่ 6.14 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 36 และชั้นที่สอง  
72 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 36 จำนวนโหนดในชั้นที่ 2 = 72																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	70.3	18	0	4.5	7.21	80.4	5.36	2.68	4.46	7.14	88.9	3.7	0.93	2.78	3.7
	Jazz	19.4	51.9	11.1	9.26	8.33	12.3	62.3	6.6	1.89	17	3.77	87.7	1.89	0.94	5.66
	Classic	7.77	1.94	81.6	6.8	1.94	0	6.42	89	1.83	2.75	0	4.67	91.6	0.93	2.8
	Rock	3.88	2.91	0.97	85.4	6.8	7.62	2.86	0.95	84.8	3.81	1.9	0.95	0.95	92.4	3.81
	Folk	4.63	16.7	13.9	10.2	54.6	1.85	8.33	8.33	4.63	76.9	1.96	8.82	0	2.94	86.3
	Average	68.75					78.65					89.37				
Testing Set Data	HipHop	67.9	11.6	1.79	10.7	8.04	78	8.26	0	5.5	8.26	76.5	7.84	0	9.8	5.88
	Jazz	11.6	54.5	6.25	9.82	17.9	10	54.5	7.27	5.45	22.7	3	74	4	4	15
	Classic	5.61	3.74	82.2	5.61	2.8	1.89	3.77	84.9	5.66	3.77	1.92	3.85	88.5	2.88	2.88
	Rock	7.62	4.76	0.95	81	5.71	8.49	4.72	0.94	78.3	7.55	7.69	6.73	0	82.7	2.88
	Folk	8.49	20.8	17.9	4.72	48.1	5.83	11.7	15.5	2.91	64.1	6.67	19	4.76	1.9	67.6
	Average	66.73					71.96					77.85				

ตารางที่ 6.15 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง  
360 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 120 จำนวนโหนดในชั้นที่ 2 = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	81.8	2.73	0.91	3.64	10.9	83.5	4.59	0.92	3.67	7.34	89.8	4.63	0	2.78	2.78
	Jazz	16.4	56.4	5.45	3.64	18.2	9.35	76.6	4.67	0.93	8.41	2.75	91.7	2.75	0.92	1.83
	Classic	0.96	3.85	90.4	1.92	2.88	0.95	2.86	92.4	1.9	1.9	0.96	1.92	93.3	0.96	2.88
	Rock	6.93	0.99	0.99	86.1	4.95	3.74	1.87	0.93	89.7	3.74	1.94	0.97	0.97	93.2	2.91
	Folk	2.91	11.7	7.77	3.88	73.8	1.92	12.5	2.88	4.81	77.9	0.98	2.94	0	1.96	94.1
	Average	77.70					84.02					92.43				
Testing Set Data	HipHop	78	7.34	0.92	7.34	6.42	75.2	7.96	1.77	6.19	8.85	72.6	9.43	0.94	11.3	5.66
	Jazz	10.5	54.3	7.62	4.76	22.9	9.65	64	7.02	3.51	15.8	6.54	66.4	6.54	2.8	17.8
	Classic	2.83	6.6	82.1	1.89	6.6	2.83	4.72	84.9	3.77	3.77	2.83	3.77	86.8	1.89	4.72
	Rock	7.69	4.81	1.92	78.8	6.73	10.5	3.81	0.95	80	4.76	9.35	4.67	0.93	81.3	3.74
	Folk	4.63	13	6.48	2.78	73.1	1.94	15.5	8.74	3.88	69.9	4.9	12.7	4.9	1.96	75.5
	Average	73.27					74.81					76.52				

ตารางที่ 6.16 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด

		จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 24																			
		จำนวนรอบในการฝึกสอน					50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk					
Training Set Data	HipHop	67.9	16	0.94	3.77	11.3	78.8	7.96	0.88	3.54	8.85	84.7	5.41	1.8	1.8	6.31					
	Jazz	26.3	43	9.65	4.39	16.7	17.9	52.7	7.14	5.36	17	11.1	65.7	4.63	1.85	16.7					
	Classic	5.71	3.81	78.1	6.67	5.71	0.94	2.83	91.5	0.94	3.77	1.85	6.48	88	1.85	1.85					
	Rock	7.84	1.96	0.98	82.4	6.86	4.76	0.95	0.95	89.5	3.81	7.77	0.97	0.97	88.3	1.94					
	Folk	9.26	25.9	15.7	3.7	45.4	2.86	8.57	8.57	6.67	73.3	3	5	2	4	86					
	Average	63.35					77.16					82.55									
Testing Set Data	HipHop	71.6	17.4	0.92	7.34	2.75	77.1	7.34	0.92	8.26	6.42	78.2	9.09	0.91	5.45	6.36					
	Jazz	21.6	43.2	7.21	7.21	20.7	14.7	52.3	6.42	5.5	21.1	6.48	58.3	9.26	5.56	20.4					
	Classic	6.42	3.67	74.3	7.34	8.26	2.91	1.94	88.3	3.88	2.91	2.83	3.77	84.9	2.83	5.66					
	Rock	10.6	9.62	0	76.9	2.88	8.74	2.91	0.97	81.6	5.83	13	4.63	0.93	77.8	3.7					
	Folk	6.36	22.7	17.3	4.55	49.1	2.88	17.3	13.5	5.77	60.6	8.41	14	4.67	2.8	70.1					
	Average	63.03					71.97					73.86									

ตารางที่ 6.17 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 72 โหนด

		จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 72																			
		จำนวนรอบในการฝึกสอน					50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk					
Training Set Data	HipHop	73.6	7.27	2.73	6.36	10	76.9	7.41	0.93	4.63	10.2	90.7	4.67	1.87	0.93	1.87					
	Jazz	17.4	56.9	6.42	5.5	13.8	11.6	63.4	7.14	1.79	16.1	4.63	86.1	1.85	1.85	5.56					
	Classic	1.9	5.71	78.1	5.71	8.57	0	2.91	91.3	1.94	3.88	0.98	2.94	94.1	0.98	0.98					
	Rock	6.73	2.88	0.96	82.7	6.73	3.88	2.91	0.97	89.3	2.91	2.88	0.96	0.96	92.3	2.88					
	Folk	9.62	9.62	3.85	6.73	70.2	1.96	9.8	2.94	4.9	80.4	1.9	5.71	0	1.9	90.5					
	Average	72.30					80.24					90.73									
Testing Set Data	HipHop	77.1	4.59	0.92	11	6.42	77.6	5.61	1.87	7.48	7.48	72.1	9.91	1.8	9.01	7.21					
	Jazz	19.6	49.1	5.36	6.25	19.6	11.7	52.3	9.01	5.41	21.6	6.31	65.8	5.41	3.6	18.9					
	Classic	2.91	2.91	80.6	4.85	8.74	1.96	3.92	86.3	3.92	3.92	1.85	4.63	87	2.78	3.7					
	Rock	8.41	1.87	0	83.2	6.54	8.49	2.83	2.83	82.1	3.77	9.9	4.95	0.99	79.2	4.95					
	Folk	9.8	19.6	6.86	8.82	54.9	0.93	16.7	5.56	4.63	72.2	3.88	20.4	5.83	1.94	68					
	Average	68.97					74.08					74.41									

ตารางที่ 6.18 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Contrast และ  
โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	85	4.67	0	5.61	4.67	86	5.61	0	3.74	4.67	93.3	0.96	0.96	2.88	1.92
	Jazz	10.4	68.9	3.77	4.72	12.3	10.2	77.8	2.78	1.85	7.41	2.8	91.6	2.8	0	2.8
	Classic	0	4.63	86.1	4.63	4.63	0	3.77	91.5	1.89	2.83	0	1.96	98	0	0
	Rock	4.72	1.89	0.94	88.7	3.77	3.88	0.97	0.97	91.3	2.91	2.88	0.96	0.96	92.3	2.88
	Folk	3.74	21.5	4.67	6.54	63.6	1.98	9.9	3.96	2.97	81.2	0.97	2.91	0	1.94	94.2
	Average	78.45					85.54					93.88				
Testing Set Data	HipHop	72.7	9.09	0.91	10.9	6.36	74.5	8.49	0.94	10.4	5.66	73.5	11.5	0	8.85	6.19
	Jazz	5.77	63.5	8.65	5.77	16.3	7.55	62.3	5.66	2.83	21.7	4.9	70.6	5.88	2.94	15.7
	Classic	0.93	6.54	85	3.74	3.74	1.98	5.94	86.1	3.96	1.98	1.85	4.63	84.3	4.63	4.63
	Rock	4.81	4.81	1.92	82.7	5.77	12.7	4.9	0.98	77.5	3.92	6.86	5.88	0.98	83.3	2.94
	Folk	3.88	18.4	8.74	4.85	64.1	3.67	17.4	8.26	0.92	69.7	3.81	15.2	6.67	1.9	72.4
	Average	73.60					74.02					76.80				

ตารางที่ 6.19 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast  
และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง  
192 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 96 จำนวนโหนดในชั้นที่ 2 = 192																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	69.2	10.3	3.74	12.1	4.67	78.2	8.18	3.64	7.27	2.73	86.7	3.81	1.9	4.76	2.86
	Jazz	10.2	59.3	5.56	7.41	17.6	8.77	64	5.26	7.02	14.9	2.88	86.5	3.85	0.96	5.77
	Classic	1.87	3.74	86.9	4.67	2.8	0.94	2.83	89.6	2.83	3.77	0	3.77	92.5	1.89	1.89
	Rock	12.1	3.74	1.87	73.8	8.41	17.3	0	1.92	76	4.81	3.74	2.8	0	91.6	1.87
	Folk	1.92	9.62	8.65	11.5	68.3	3.88	7.77	2.91	8.74	76.7	0.96	1.92	0.96	1.92	94.2
	Average	71.49					76.90					90.30				
Testing Set Data	HipHop	65.7	13	0	13.9	7.41	73	9.91	0.9	9.01	7.21	63.3	20.2	0.92	7.34	8.26
	Jazz	13.6	50	11.8	8.18	16.4	10.6	51.3	15	4.42	18.6	9.09	53.6	12.7	3.64	20.9
	Classic	2.83	2.83	84	5.66	4.72	2.86	4.76	86.7	1.9	3.81	2.86	8.57	82.9	1.9	3.81
	Rock	11.8	12.7	0.91	70	4.55	17.1	6.67	1.9	70.5	3.81	11.3	5.66	2.83	74.5	5.66
	Folk	2.78	11.1	9.26	11.1	65.7	2.91	8.74	4.85	5.83	77.7	4.85	5.83	3.88	3.88	81.6
	Average	67.09					71.82					71.18				

ตารางที่ 6.20 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 192 จำนวนโหนดในชั้นที่2 = 384																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	78.4	6.31	2.7	9.91	2.7	79.6	6.8	1.94	10.7	0.97	89.7	2.8	0.93	3.74	2.8
	Jazz	12.3	51.8	7.89	7.02	21.1	6.67	64.8	5.71	7.62	15.2	1.83	91.7	1.83	1.83	2.75
	Classic	1.83	3.67	87.2	3.67	3.67	1.89	3.77	86.8	2.83	4.72	0	1.92	93.3	0.96	3.85
	Rock	13.1	0	1.87	79.4	5.61	15.1	1.89	1.89	75.5	5.66	3.92	0.98	0	94.1	0.98
	Folk	3.7	6.48	3.7	10.2	75.9	0.93	5.61	2.8	6.54	84.1	0.98	1.96	0.98	1.96	94.1
	Average	74.53					78.15					92.59				
Testing Set Data	HipHop	74.1	7.14	1.79	8.93	8.04	72.1	12.6	0	9.01	6.31	68.2	15.9	1.87	10.3	3.74
	Jazz	15	45.8	12.1	8.41	18.7	11.7	52.4	15.5	3.88	16.5	5.26	60.5	12.3	4.39	17.5
	Classic	3.88	1.94	83.5	5.83	4.85	0	5.71	85.7	5.71	2.86	1.83	7.34	84.4	1.83	4.59
	Rock	15.2	6.67	1.9	69.5	6.67	14.4	7.69	1.92	72.1	3.85	13.2	6.6	0.94	73.6	5.66
	Folk	1.85	7.41	5.56	12	73.1	2.83	10.4	3.77	4.72	78.3	5.66	6.6	5.66	6.6	75.5
	Average	69.21					72.13					72.44				

ตารางที่ 6.21 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 384 และชั้นที่สอง 768 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 384 จำนวนโหนดในชั้นที่2 = 768																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	73.7	8.77	3.51	11.4	2.63	80.7	8.26	0.92	9.17	0.92	91.6	1.87	0.93	2.8	2.8
	Jazz	11.7	57.7	7.21	9.91	13.5	11.2	69.2	2.8	3.74	13.1	1.85	90.7	0.93	3.7	2.78
	Classic	1.87	1.87	92.5	1.87	1.87	0	2.94	91.2	1.96	3.92	0	1.9	94.3	0.95	2.86
	Rock	10.2	0	1.85	83.3	4.63	14.6	0.97	0.97	80.6	2.91	2.83	0.94	0	95.3	0.94
	Folk	2.91	1.94	3.88	7.77	83.5	0.97	4.85	1.94	3.88	88.3	1	0	1	2	96
	Average	78.14					82.00					93.58				
Testing Set Data	HipHop	71.9	10.5	0	9.65	7.89	71.4	12.5	0.89	9.82	5.36	72.7	11.8	1.82	10.9	2.73
	Jazz	13.8	44	13.8	7.34	21.1	11.8	52.7	14.5	4.55	16.4	6.36	59.1	12.7	4.55	17.3
	Classic	4.72	4.72	84	2.83	3.77	2.8	4.67	86.9	2.8	2.8	2.88	7.69	84.6	1.92	2.88
	Rock	13	4.63	0	75.9	6.48	14.5	9.09	0.91	70.9	4.55	8.82	7.84	1.96	75.5	5.88
	Folk	3.88	10.7	3.88	3.88	77.7	1.85	13.9	7.41	2.78	74.1	3.77	2.83	6.6	3.77	83
	Average	70.70					71.21					74.99				

ตารางที่ 6.22 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 192																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	77.8	5.56	2.78	10.2	3.7	77.9	7.69	0.96	11.5	1.92	87.3	4.55	0.91	3.64	3.64
	Jazz	11.4	57.9	6.14	5.26	19.3	9.35	67.3	3.74	4.67	15	4.59	80.7	4.59	2.75	7.34
	Classic	0.96	2.88	88.5	3.85	3.85	0.96	2.88	88.5	2.88	4.81	0.97	2.91	92.2	0.97	2.91
	Rock	12.7	1.96	1.96	76.5	6.86	14.8	1.85	0.93	77.8	4.63	6.67	0.95	0	89.5	2.86
	Folk	2.97	8.91	4.95	8.91	74.3	1.9	6.67	3.81	4.76	82.9	0.94	2.83	0.94	1.89	93.4
	Average	74.97					78.85					88.63				
Testing Set Data	HipHop	73.6	7.27	0	11.8	7.27	77.6	9.35	0.93	7.48	4.67	69.6	14.3	0.89	8.04	7.14
	Jazz	13.9	50	12	6.48	17.6	14.2	49.1	16	3.77	17	6.84	58.1	12.8	2.56	19.7
	Classic	0.92	7.34	83.5	5.5	2.75	1.9	4.76	83.8	5.71	3.81	2.86	9.52	81.9	0.95	4.76
	Rock	14.8	8.33	0	72.2	4.63	14.7	9.17	0.92	70.6	4.59	10.6	10.6	0	73.1	5.77
	Folk	1.9	6.67	4.76	4.76	81.9	1.96	11.8	3.92	2.94	79.4	1.94	7.77	3.88	1.94	84.5
	Average	72.25					72.10					73.44				

ตารางที่ 6.23 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 384																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	73.5	8.85	3.54	12.4	1.77	82.1	5.66	1.89	8.49	1.89	90.6	3.77	0.94	2.83	1.89
	Jazz	7.62	59	10.5	10.5	12.4	2.75	76.1	5.5	3.67	11.9	2.78	87	2.78	2.78	4.63
	Classic	2	2	90	3	3	0.97	2.91	89.3	1.94	4.85	0	1.85	93.5	0.93	3.7
	Rock	10.7	0.97	1.94	79.6	6.8	11.9	0.99	0.99	84.2	1.98	6	1	0	93	0
	Folk	0.93	7.41	4.63	13	74.1	0.97	2.91	0.97	3.88	91.3	1	2	1	2	94
	Average	75.24					84.59					91.62				
Testing Set Data	HipHop	74.1	9.26	0.93	8.33	7.41	70.1	15	0.93	7.48	6.54	70.9	13.6	0	9.09	6.36
	Jazz	12.3	49.1	16	4.72	17.9	7.48	57	13.1	4.67	17.8	12.3	56.1	11.4	3.51	16.7
	Classic	0.93	5.61	86	3.74	3.74	1.87	5.61	86	3.74	2.8	1.89	7.55	85.8	0.94	3.77
	Rock	13.1	5.61	0.93	72.9	7.48	13.8	6.42	1.83	73.4	4.59	14.7	4.9	0	73.5	6.86
	Folk	3.74	9.35	6.54	6.54	73.8	2.86	9.52	5.71	3.81	78.1	3	5	6	3	83
	Average	71.17					72.91					73.89				

ตารางที่ 6.24 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Contrast และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 768 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 768																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	77.1	6.42	3.67	10.1	2.75	82.3	5.31	1.77	7.96	2.65	87.7	3.77	2.83	3.77	1.89
	Jazz	11.4	54.4	10.5	9.65	14	3.6	74.8	5.41	6.31	9.91	4.76	82.9	2.86	3.81	5.71
	Classic	2.78	7.41	87	1.85	0.93	0	2.83	91.5	2.83	2.83	0	3.81	93.3	0.95	1.9
	Rock	11.8	3.92	1.96	78.4	3.92	8.65	1.92	0.96	83.7	4.81	7.77	1.94	0	88.3	1.94
	Folk	1.98	13.9	4.95	12.9	66.3	0.94	4.72	1.89	4.72	87.7	0.98	0.98	0.98	1.96	95.1
	Average	72.65					83.99					89.47				
Testing Set Data	HipHop	76.7	6.8	0.97	9.71	5.83	73.1	12	0	8.33	6.48	72.9	12.1	1.87	8.41	4.67
	Jazz	15.7	50.4	14.8	5.22	13.9	10.8	55.9	9.91	4.5	18.9	11.2	57.9	11.2	3.74	15.9
	Classic	2.78	8.33	80.6	4.63	3.7	3.96	6.93	85.1	0.99	2.97	0.96	7.69	86.5	0	4.81
	Rock	10.6	8.65	0.96	74	5.77	15.5	9.71	0	70.9	3.88	11	5.5	0.92	73.4	9.17
	Folk	6.6	13.2	4.72	11.3	64.2	3.77	10.4	4.72	2.83	78.3	3.81	4.76	2.86	3.81	84.8
	Average	69.18					72.67					75.11				

ตารางที่ 6.25 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง 120 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 60 จำนวนโหนดในชั้นที่ 2 = 120																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	86.4	3.88	0.97	6.8	1.94	84.5	7.27	2.73	3.64	1.82	94.3	2.86	0	0.95	1.9
	Jazz	30.3	37	10.9	4.2	17.6	14.8	55.6	11.1	3.7	14.8	8.41	66.4	4.67	2.8	17.8
	Classic	2.88	0.96	83.7	8.65	3.85	0	0	96.1	2.94	0.98	0	0.97	94.2	1.94	2.91
	Rock	5.5	2.75	0.92	82.6	8.26	1.87	0.93	0.93	85	11.2	1.92	0.96	0.96	87.5	8.65
	Folk	8.85	18.6	23	13.3	36.3	3.74	9.35	12.1	13.1	61.7	0.95	12.4	0.95	8.57	77.1
	Average	65.18					76.58					83.89				
Testing Set Data	HipHop	73.2	6.25	2.68	12.5	5.36	72.5	14.7	2.75	6.42	3.67	83	5.36	1.79	8.04	1.79
	Jazz	19.3	38.5	12.8	9.17	20.2	8.11	57.7	8.11	0.9	25.2	8.33	67.6	3.7	2.78	17.6
	Classic	1.92	3.85	84.6	5.77	3.85	1.83	0.92	81.7	6.42	9.17	0.92	0.92	81.7	5.5	11
	Rock	6.8	0.97	2.91	83.5	5.83	6.67	1.9	2.86	81.9	6.67	8.41	1.87	1.87	77.6	10.3
	Folk	16.3	18.3	20.2	9.62	35.6	6.48	13.9	16.7	4.63	58.3	7.55	17.9	5.66	3.77	65.1
	Average	63.09					70.40					74.99				

ตารางที่ 6.26 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 180 และชั้นที่สอง 360 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 180 จำนวนโหนดในชั้นที่2 = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	88.8	2.8	0.93	6.54	0.93	92.6	2.78	0.93	1.85	1.85	94.5	1.83	0.92	1.83	0.92
	Jazz	20.6	53.3	7.48	3.74	15	15.3	60.4	4.5	4.5	15.3	5.41	73	4.5	2.7	14.4
	Classic	0	0	89.8	4.63	5.56	0	0	92.5	3.77	3.77	0	2.97	94.1	0	2.97
	Rock	2.8	1.87	0.93	85	9.35	2.83	1.89	0.94	84	10.4	0	1.89	0.94	87.7	9.43
	Folk	2.83	16	2.83	16	62.3	1.9	12.4	1.9	12.4	71.4	0	8.57	0	2.86	88.6
	Average	75.84					80.16					87.57				
Testing Set Data	HipHop	80	9.09	1.82	7.27	1.82	75	12	0.93	8.33	3.7	82.2	6.54	0.93	7.48	2.8
	Jazz	7.62	58.1	7.62	2.86	23.8	13.8	54.1	6.42	1.83	23.9	10.6	70.2	2.88	0.96	15.4
	Classic	2.83	0	79.2	6.6	11.3	3.88	2.91	78.6	5.83	8.74	2.91	0.97	82.5	4.85	8.74
	Rock	6.8	2.91	0.97	85.4	3.88	5.71	1.9	1.9	81.9	8.57	7.55	2.83	1.89	77.4	10.4
	Folk	6.93	17.8	6.93	6.93	61.4	8.74	14.6	3.88	5.83	67	10.1	17.4	3.67	1.83	67
	Average	72.83					71.33					75.86				

ตารางที่ 6.27 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 360 และชั้นที่สอง 720 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 360 จำนวนโหนดในชั้นที่2 = 720																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	89.8	4.63	0.93	1.85	2.78	93.4	1.89	1.89	1.89	0.94	94.3	2.86	0	1.9	0.95
	Jazz	16.8	57.5	7.96	2.65	15	15.9	60.2	5.31	4.42	14.2	5.61	77.6	1.87	3.74	11.2
	Classic	0	0.97	94.2	1.94	2.91	0	0	95.1	0.98	3.92	0	0.96	92.3	1.92	4.81
	Rock	3.92	1.96	0.98	82.4	10.8	2.83	1.89	0.94	84	10.4	0.95	0.95	0	88.6	9.52
	Folk	2.8	13.1	4.67	13.1	66.4	1.92	15.4	0.96	11.5	70.2	0	9.43	0	3.77	86.8
	Average	78.04					80.57					87.91				
Testing Set Data	HipHop	75.2	10.1	1.83	7.34	5.5	81.9	7.62	1.9	4.76	3.81	75.7	11.7	2.7	6.31	3.6
	Jazz	10.8	56.8	8.11	1.8	22.5	11.6	64.3	2.68	0.89	20.5	10.9	71.8	3.64	0.91	12.7
	Classic	0.93	2.8	81.3	4.67	10.3	7.34	0.92	77.1	5.5	9.17	2.91	1.94	83.5	4.85	6.8
	Rock	4	3	4	83	6	5.66	2.83	1.89	83	6.6	4.72	2.83	2.83	78.3	11.3
	Folk	10.1	13.8	7.34	8.26	60.6	9.43	12.3	2.83	4.72	70.8	4.72	16	3.77	0.94	74.5
	Average	71.37					75.41					76.76				

ตารางที่ 6.28 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 120																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	84.6	3.85	2.88	5.77	2.88	81.3	10.7	1.79	4.46	1.79	90.8	2.75	0.92	3.67	1.83
	Jazz	20.4	54	6.19	4.42	15	14	55.3	5.26	3.51	21.9	11.2	63.6	5.61	3.74	15.9
	Classic	0	1.92	86.5	10.6	0.96	0	0.96	92.3	5.77	0.96	0	1.85	88.9	3.7	5.56
	Rock	2.86	1.9	0.95	84.8	9.52	2.86	1.9	0.95	83.8	10.5	2.86	1.9	0.95	85.7	8.57
	Folk	2.83	8.49	12.3	20.8	55.7	2.83	15.1	4.72	19.8	57.5	0	8.82	2.94	10.8	77.5
	Average	73.11					74.04					81.29				
Testing Set Data	HipHop	75.5	6.6	1.89	12.3	3.77	70.1	16.8	1.87	10.3	0.93	82.4	5.56	1.85	6.48	3.7
	Jazz	8.26	46.8	6.42	7.34	31.2	7.55	53.8	8.49	3.77	26.4	9.43	68.9	2.83	0	18.9
	Classic	0.96	3.85	83.7	4.81	6.73	5.61	1.87	78.5	6.54	7.48	2.88	0.96	83.7	3.85	8.65
	Rock	5.56	2.78	2.78	82.4	6.48	4.85	2.91	0.97	87.4	3.88	5.94	1.98	1.98	81.2	8.91
	Folk	5.45	11.8	11.8	14.5	56.4	5.83	14.6	13.6	11.7	54.4	3.74	15.9	3.74	2.8	73.8
	Average	68.94					68.82					77.99				

ตารางที่ 6.29 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	84.8	10.5	0.95	3.81	0	89.5	4.76	0.95	3.81	0.95	94.3	0.94	0.94	2.83	0.94
	Jazz	12.7	60	5.45	4.55	17.3	13.5	63.5	3.85	4.81	14.4	10.1	71.6	3.67	2.75	11.9
	Classic	0	0	92.2	2.94	4.9	0	0.96	92.3	1.92	4.81	0	1.9	96.2	0.95	0.95
	Rock	1.92	0.96	0.96	85.6	10.6	2.8	1.87	0.93	85	9.35	0.94	1.89	0.94	86.8	9.43
	Folk	0.94	13.2	7.55	15.1	63.2	0.98	13.7	0.98	13.7	70.6	0	7.62	0	8.57	83.8
	Average	77.14					80.19					86.54				
Testing Set Data	HipHop	71.6	16.5	1.83	6.42	3.67	76.1	12.8	1.83	6.42	2.75	81.8	6.36	1.82	5.45	4.55
	Jazz	11.1	60.2	5.56	0.93	22.2	9.26	61.1	4.63	0.93	24.1	9.91	68.5	2.7	0.9	18
	Classic	1.85	2.78	80.6	5.56	9.26	4.67	3.74	80.4	4.67	6.54	2.86	1.9	81.9	6.67	6.67
	Rock	4.67	2.8	0.93	86	5.61	4.72	3.77	1.89	80.2	9.43	6.67	1.9	2.86	79	9.52
	Folk	8.18	15.5	10	8.18	58.2	5.61	15	4.67	3.74	71	3.67	14.7	4.59	3.67	73.4
	Average	71.29					73.77					76.93				

ตารางที่ 6.30 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Octave-based Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 720 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 720																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	85.5	3.64	3.64	4.55	2.73	87.2	6.42	1.83	2.75	1.83	92.6	2.78	0.93	1.85	1.85
	Jazz	14	57.9	5.26	3.51	19.3	13.6	68.2	4.55	3.64	10	6.67	77.1	1.9	2.86	11.4
	Classic	0	0	89.6	3.77	6.6	0	0.93	92.5	2.8	3.74	0	0	94.3	1.89	3.77
	Rock	2.88	0.96	0.96	83.7	11.5	2.78	1.85	0.93	84.3	10.2	0.94	0.94	0.94	88.7	8.49
	Folk	1.89	12.3	2.83	13.2	69.8	0	14.6	0.97	11.7	72.8	0	7.77	0	4.85	87.4
	Average	77.29					80.99					88.03				
Testing Set Data	HipHop	82.6	6.42	1.83	7.34	1.83	76.8	8.93	2.68	8.04	3.57	80.2	9.01	1.8	7.21	1.8
	Jazz	10.4	54.8	7.83	2.61	24.3	9.01	64	4.5	2.7	19.8	9.82	69.6	2.68	2.68	15.2
	Classic	3.7	0.93	78.7	5.56	11.1	0.99	1.98	83.2	6.93	6.93	1.82	1.82	80.9	6.36	9.09
	Rock	3.74	2.8	2.8	80.4	10.3	5.61	2.8	1.87	82.2	7.48	4.9	2.94	2.94	80.4	8.82
	Folk	5.71	16.2	7.62	2.86	67.6	5.71	16.2	4.76	5.71	67.6	5.83	15.5	2.91	3.88	71.8
	Average	72.81					74.76					76.59				

ตารางที่ 6.31 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง 360 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 120 จำนวนโหนดในชั้นที่2 = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	81.5	8.33	0.93	3.7	5.56	94.2	1.94	0	1.94	1.94	98.1	0.97	0	0	0.97
	Jazz	8.33	74.1	7.41	1.85	8.33	0	88.9	3.7	0.93	6.48	0	96.3	0.93	0.93	1.87
	Classic	0.95	0.95	93.3	0.95	3.81	0	0	98.1	0.97	0.97	0	0	98.1	0.95	0.95
	Rock	0.96	0	0	98.1	0.96	0.98	0	0	97.1	1.96	0	0	0	99	0.96
	Folk	1.98	6.93	3.96	2.97	84.2	0.97	6.8	1.94	0	90.3	0	0	0	0	100
	Average	86.22					93.69					98.29				
Testing Set Data	HipHop	77.3	12.7	1.82	5.45	2.73	86	5.61	0.93	4.67	2.8	80.4	8.04	1.79	6.25	3.57
	Jazz	9.26	64.8	10.2	3.7	12	6.67	76.2	7.62	2.86	6.67	9.17	69.7	8.26	4.59	8.26
	Classic	2.88	0.96	88.5	0.96	6.73	1.85	1.85	88	2.78	5.56	2.86	1.9	88.6	0.95	5.71
	Rock	5.61	1.87	0	90.7	1.87	4.85	1.94	0	90.3	2.91	2.88	1.92	0	93.3	1.92
	Folk	1.92	9.62	5.77	0.96	81.7	3.96	8.91	1.98	0	85.1	0.97	5.83	2.91	0.97	89.3
	Average	80.59					85.11					84.25				

ตารางที่ 6.32 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 240 และชั้นที่สอง 480 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 240 จำนวนโหนดในชั้นที่2 = 480																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	84.6	8.65	0.96	2.88	2.88	94.5	0.92	0.92	1.83	1.83	95.3	0.93	0.93	0.93	1.87
	Jazz	10.2	74.1	7.41	2.78	5.56	0	87.6	5.71	1.9	4.76	0	96.2	0.96	0.96	1.92
	Classic	2.86	0	94.3	0.95	1.9	0	0	97.1	0.97	1.94	0	0	100	0	0
	Rock	3.81	0.95	0.95	93.3	0.95	0	0	0	98.1	1.9	0	0	0	99	0.95
	Folk	0.94	4.72	6.6	3.77	84	0.98	0.98	0.98	0	97.1	0	0	0	0	100
	Average	86.05					94.87					98.11				
Testing Set Data	HipHop	82.4	8.33	0.93	5.56	2.78	83.8	6.31	1.8	7.21	0.9	87.7	6.6	0	5.66	0
	Jazz	6.19	68.1	8.85	6.19	10.6	7.34	70.6	9.17	2.75	10.1	8.04	72.3	8.04	2.68	8.93
	Classic	2.91	0.97	91.3	1.94	2.91	3.81	0.95	86.7	0.95	7.62	0.96	0.96	90.4	1.92	5.77
	Rock	4.9	0.98	0	93.1	0.98	2.83	3.77	0	90.6	2.83	2.83	1.89	0	94.3	0.94
	Folk	3.88	8.74	5.83	1.94	79.6	2.94	8.82	1.96	0.98	85.3	1.98	4.95	2.97	0.99	89.1
	Average	82.91					83.39					86.78				

ตารางที่ 6.33 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 480 และชั้นที่สอง 960 โหนด

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 480 จำนวนโหนดในชั้นที่2 = 960																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	86.8	6.6	0.94	0.94	4.72	96.1	1.94	0	0.97	0.97	94.3	0.94	0.94	1.89	1.89
	Jazz	11.1	74.1	6.48	1.85	6.48	1.89	86.8	4.72	2.83	3.77	0	96.3	1.87	0	1.87
	Classic	1.87	0	94.4	0.93	2.8	0	0	99	0	0.96	0	0	99	0	0.98
	Rock	4.85	0.97	0	93.2	0.97	0	0	0	99	0.98	0	0	0	98.1	1.89
	Folk	1.87	9.35	3.74	1.87	83.2	0	2	3	1	94	0	0	0	0	100
	Average	86.33					94.99					97.55				
Testing Set Data	HipHop	81.4	9.73	1.77	5.31	1.77	85.5	7.27	0.91	4.55	1.82	81.1	8.49	1.89	6.6	1.89
	Jazz	7.96	67.3	10.6	3.54	10.6	3.7	71.3	10.2	5.56	9.26	2.78	78.7	7.41	3.7	7.41
	Classic	2.75	0.92	89.9	1.83	4.59	1.89	2.83	90.6	1.89	2.83	0.94	0.94	93.4	0.94	3.77
	Rock	2.97	1.98	0	94.1	0.99	2.94	1.96	0	94.1	0.98	0.97	3.88	0.97	92.2	1.94
	Folk	2.97	8.91	1.98	1.98	84.2	0.95	10.5	2.86	0	85.7	0	8.91	5.94	0.99	84.2
	Average	83.36					85.43					85.92				

ตารางที่ 6.34 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 360																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	85.3	6.42	1.83	3.67	2.75	88.7	5.66	0.94	0.94	3.77	97.1	0.96	0.96	0	0.96
	Jazz	11.4	75.2	4.76	0	8.57	2.78	86.1	5.56	0	5.56	0	96.2	0.96	0.96	1.92
	Classic	1.9	0	93.3	0.95	3.81	0.99	0	97	0.99	0.99	0	0	97.2	0.94	1.89
	Rock	6.8	2.91	0.97	87.4	1.94	0.96	0.96	0	97.1	0.96	0	0	0	99	0.99
	Folk	1.94	12.6	8.74	1.94	74.8	0	7.55	1.89	0.94	89.6	0	0	0	0	100
	Average	83.21					91.71					97.89				
Testing Set Data	HipHop	81.3	11.2	1.87	3.74	1.87	82.9	9.01	1.8	5.41	0.9	81.5	8.33	0.93	6.48	2.78
	Jazz	8.04	63.4	13.4	2.68	12.5	6.31	71.2	9.91	4.5	8.11	3.7	77.8	7.41	5.56	5.56
	Classic	1.96	1.96	92.2	0	3.92	0.94	3.77	91.5	0	3.77	2.78	0	89.8	2.78	4.63
	Rock	4.9	2.94	0	91.2	0.98	3.74	3.74	0	90.7	1.87	2.94	0.98	0	95.1	0.98
	Folk	0.93	8.41	6.54	1.87	82.2	1.89	12.3	2.83	0	83	0	9.52	3.81	1.9	84.8
	Average	82.06					83.85					85.79				

ตารางที่ 6.35 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 480 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 480																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	82.4	9.26	0.93	3.7	3.7	86.4	8.18	0.91	1.82	2.73	94.5	1.82	1.82	1.82	0
	Jazz	7.08	72.6	7.08	0.88	12.4	3.67	80.7	9.17	0.92	5.5	0	95.2	1.92	1.92	0.96
	Classic	0.95	0.95	90.5	2.86	4.76	0	0	96.3	0.93	2.78	0	0	98.1	0.97	0.97
	Rock	3.81	0	0.95	91.4	3.81	1.94	0	0	95.1	2.91	0	0	0	99.1	0.94
	Folk	0.95	13.3	13.3	2.86	69.5	0.99	7.92	3.96	0.99	86.1	0	0	0	0	100
	Average	81.28					88.94					97.37				
Testing Set Data	HipHop	81.5	11.1	0	3.7	3.7	83.7	9.62	1.92	3.85	0.96	85	6.54	0.93	5.61	1.87
	Jazz	8.18	62.7	11.8	2.73	14.5	7.77	70.9	11.7	3.88	5.83	3.7	77.8	7.41	4.63	6.48
	Classic	1.96	1.96	92.2	0	3.92	0.93	2.8	91.6	0.93	3.74	0.96	1.92	89.4	1.92	5.77
	Rock	2.83	1.89	0	91.5	3.77	2.91	3.88	0	92.2	0.97	2.86	0.95	0.95	93.3	1.9
	Folk	0.93	13.1	6.54	3.74	75.7	1.96	10.8	7.84	0.98	78.4	0.93	9.35	1.87	0	87.9
	Average	80.72					83.36					86.69				

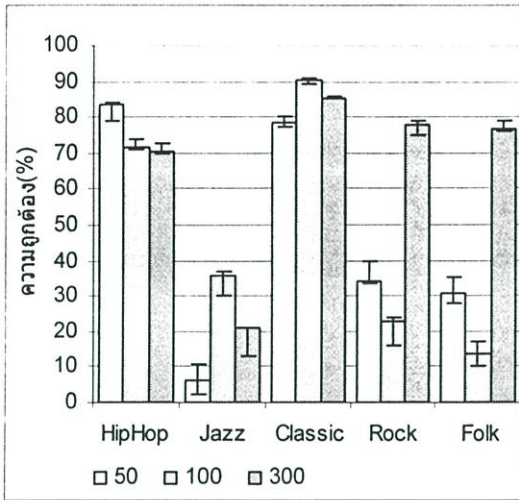
ตารางที่ 6.36 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 960 โหนด

จำนวนชั้นซ่อน = 1 จำนวนโหนดในชั้นแรก = 960																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	86	7.48	0.93	1.87	3.74	90.6	4.72	0.94	1.89	1.89	94.4	1.87	0	1.87	1.87
	Jazz	6.42	79.8	6.42	0.92	6.42	1.96	88.2	4.9	0.98	3.92	0	94.5	2.75	0.92	1.83
	Classic	0.97	0	94.2	1.94	2.91	0	0	98.1	0	1.89	0	0	97.2	0.93	1.87
	Rock	3.81	0.95	0	93.3	1.9	2.83	0	0	95.3	1.89	0	0	0	99	0.95
	Folk	0.97	5.83	4.85	1.94	86.4	0	4.95	5.94	0.99	88.1	0	0	0	0	100
	Average	87.94					92.06					97.03				
Testing Set Data	HipHop	84.3	9.26	0.93	4.63	0.93	87.5	6.73	0	3.85	1.92	81.4	6.19	0.88	7.96	3.54
	Jazz	7.77	66	10.7	2.91	12.6	5.56	70.4	10.2	4.63	9.26	5.56	72.2	10.2	4.63	7.41
	Classic	2.83	0.94	88.7	1.89	5.66	2.86	2.86	90.5	0.95	2.86	1.89	0.94	91.5	2.83	2.83
	Rock	4.85	3.88	0	90.3	0.97	3.88	3.88	0	90.3	1.94	2.91	1.94	0.97	93.2	0.97
	Folk	2.88	5.77	4.81	2.88	83.7	1.94	7.77	2.91	1.94	85.4	0.93	9.35	3.74	0	86
	Average	82.58					84.81					84.87				

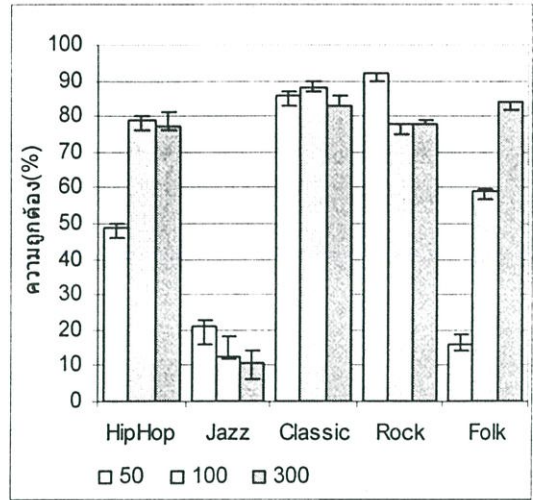
ตารางที่ 6.37 ตารางผลการแยกประเภทเพลงด้วยลักษณะเด่น Mel-Frequency Spectral Histogram และโครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 240 และชั้นที่สอง 480 โหนด โดยใช้ข้อมูลฝึกสอนที่ได้จากความยาวเพลงฝึกสอนทั้งเพลง

จำนวนชั้นซ่อน = 2 จำนวนโหนดในชั้นแรก = 240 จำนวนโหนดในชั้นที่ 2 = 480																
จำนวนรอบในการฝึกสอน		50					100					300				
แนวเพลง		HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk	HipHop	Jazz	Classic	Rock	Folk
Training Set Data	HipHop	86.8	8.49	0.94	0.94	2.83	88.7	5.66	1.89	1.89	1.89	95.1	2.91	0.97	0.97	0
	Jazz	10.1	74.3	7.34	0.92	7.34	4.72	81.1	6.6	0.94	6.6	0	89.8	3.7	0.93	5.56
	Classic	2.86	0	93.3	0.95	2.86	1.87	0	93.5	0.93	3.74	0	0	97.1	0.98	1.96
	Rock	3.85	0.96	0.96	91.3	2.88	1.89	0.94	0.94	94.3	1.89	0	0	0	100	0
	Folk	0.96	4.81	5.77	3.85	84.6	0.97	2.91	3.88	1.94	90.3	0.98	0.98	0.98	0	97.1
	Average	86.08					89.58					95.82				
Testing Set Data	HipHop	83.2	8.41	1.87	3.74	2.8	83.8	5.71	0.95	7.62	1.9	86.8	5.66	0	6.6	0.94
	Jazz	6.19	69.9	7.96	6.19	9.73	8.18	72.7	7.27	3.64	8.18	8.57	72.4	6.67	3.81	8.57
	Classic	2.88	0.96	92.3	0	3.85	3.85	0.96	86.5	0.96	7.69	0.98	0.98	91.2	1.96	4.9
	Rock	4.85	0.97	0	92.2	1.94	2.88	3.85	0	90.4	2.88	2.86	1.9	0	93.3	1.9
	Folk	3.96	8.91	5.94	1.98	79.2	2.86	7.62	1.9	0.95	86.7	1.92	4.81	2.88	0.96	89.4
	Average	83.37					84.03					86.62				

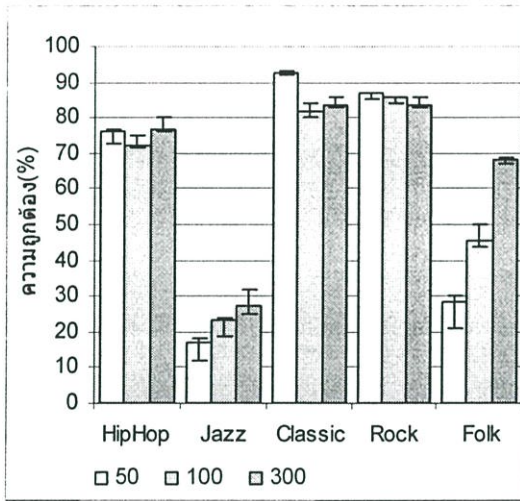
จากตารางที่ 6.37 จะเห็นว่าเมื่อเพิ่มจำนวนข้อมูลฝึกสอนจาก 50 วินาที เป็นความยาวเพลงฝึกสอนทั้งเพลงนั้น ไม่สามารถเพิ่มความถูกต้องของการแยกประเภทดนตรีได้อีก แต่ความถูกต้องของการทดสอบกับข้อมูลฝึกสอน จะมีค่าลดลงเนื่องจากข้อมูลฝึกสอนมีการกระจายตัวของข้อมูลมากขึ้น



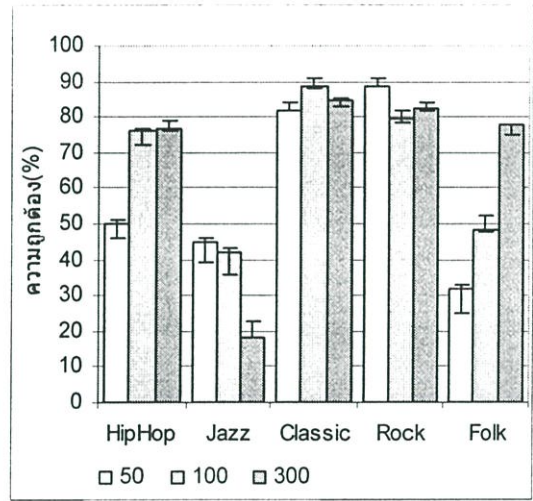
(ก)



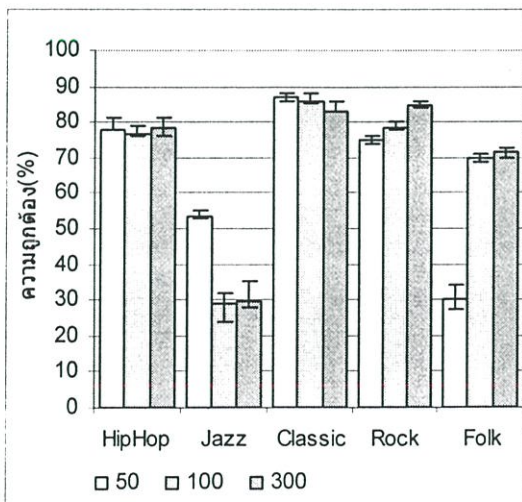
(ข)



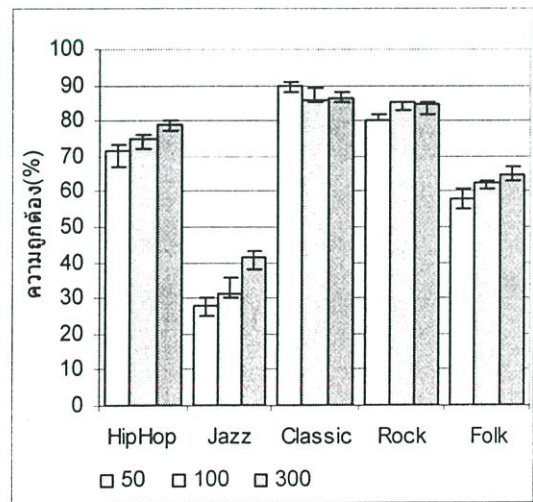
(ค)



(ง)



(จ)



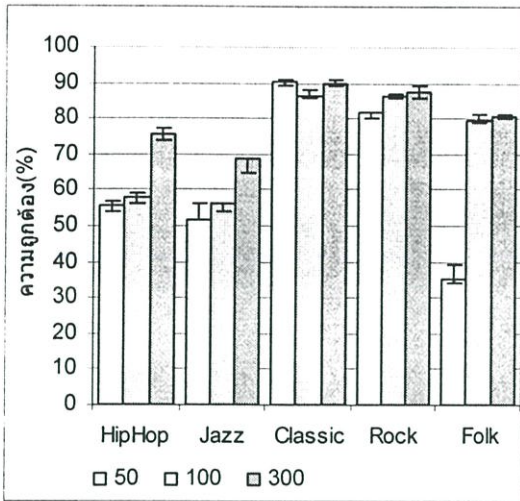
(ฉ)

รูปที่ 6.1 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Mean ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

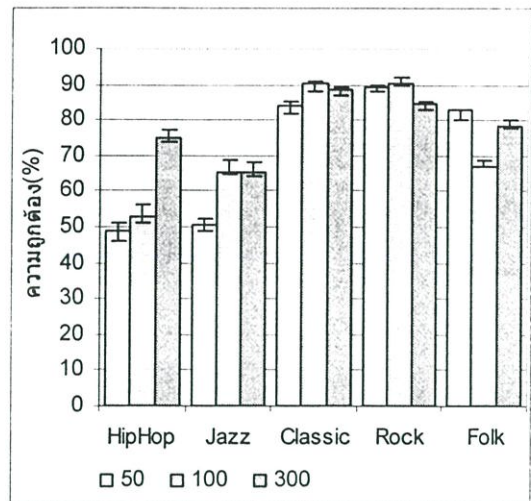
- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง 24 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 18 และชั้นที่สอง 36 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 36 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง 120 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด

จากรูปที่ 6.1 เมื่อเปรียบเทียบความถูกต้องระหว่างประเภทของเพลง จะเห็นได้ว่า ลักษณะเด่น Octave-based Spectral Mean ที่มีเพียง 6 ค่าต่อ 1 เพลง สามารถแยกประเภทดนตรี Classic, Rock และ Hip-Hop ได้ดี แต่ล้มเหลวมากเมื่อใช้แยกประเภทดนตรี Jazz และ Folk ถึงแม้จะเพิ่มจำนวนครั้งที่ใช้ในการฝึกสอน หรือเพิ่มจำนวนโหนด ในโครงข่ายประสาทเทียมแล้วก็ตาม แต่ประสิทธิภาพในการแยกแยะ เพลง 2 ประเภทนี้ ก็ไม่ดีขึ้นเท่าที่ควร เนื่องจากเพลง 2 ประเภทนี้มีความซับซ้อนเกินกว่าที่จะถูกแยกประเภทด้วยตัวแปรเพียงแค่ 6 ค่า

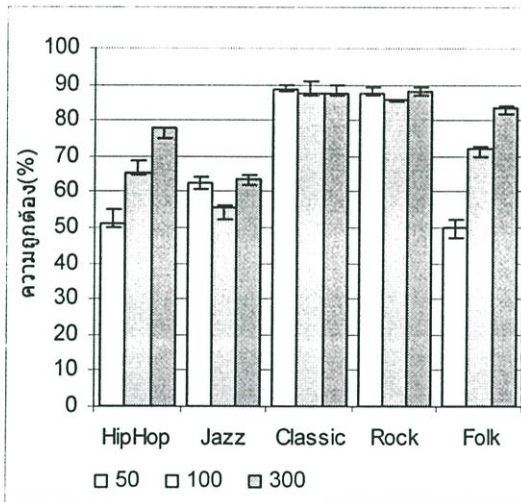
เมื่อกลับไปดูที่ตาราง 6.1 - 6.6 จะเห็นได้ว่า เพลง Jazz นั้น ถูกแยกได้เป็น เพลง Hip-Hop, Classic และ Folk กระจายไปพอกๆกัน และเพลง Folk ก็ถูกแยกได้เป็น Hip-Hop, Jazz และ Classic เช่นกัน เพราะฉะนั้น การใช้ลักษณะเด่น Octave-based Spectral Mean นั้น ไม่เพียงมีความสับสนระหว่างเพลง Jazz กับเพลง Folk เท่านั้น แต่มีความคลุมเครืออยู่ในหลายแนวเพลงด้วยกัน



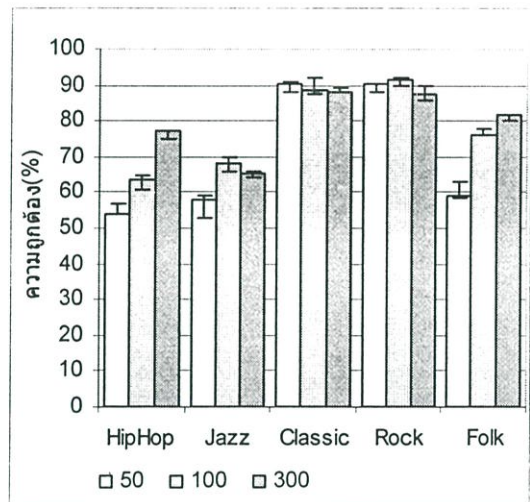
(ก)



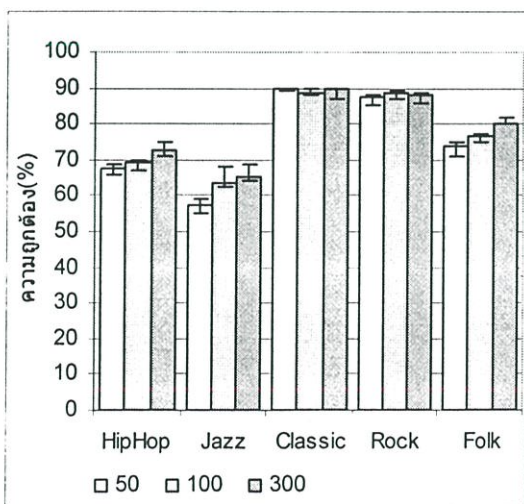
(ข)



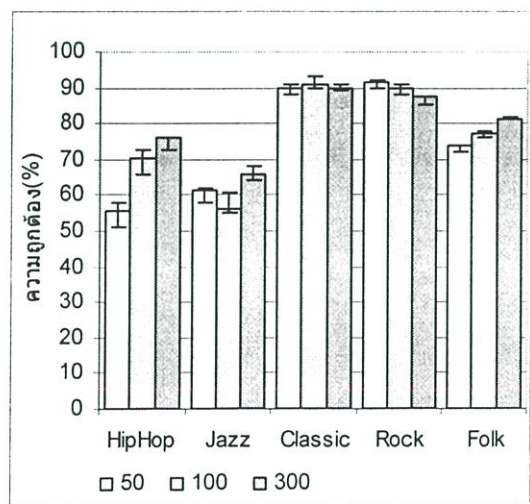
(ค)



(ง)



(จ)



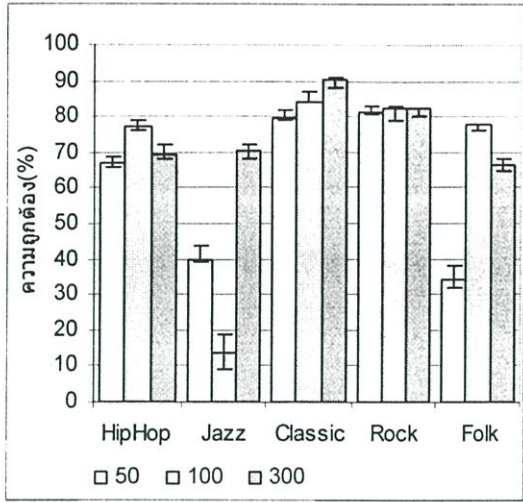
(ฉ)

รูปที่ 6.2 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Mean ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

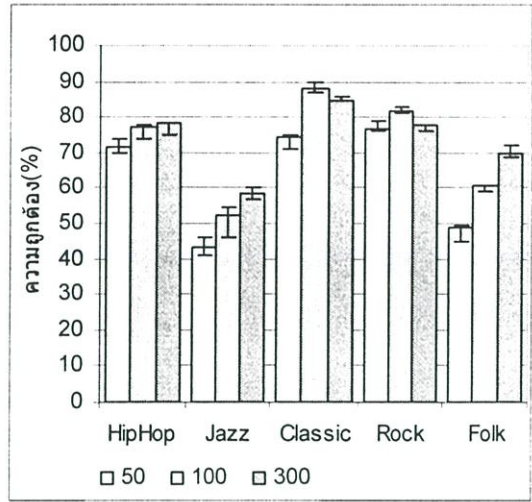
- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 48 และชั้นที่สอง 96 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 96 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง 192 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด

จากรูปที่ 6.2 เมื่อดูกราฟโดยรวมแล้ว ลักษณะเด่น Mel-Frequency Spectral Mean นี้สามารถใช้แยกประเภทได้ดีพอสมควร เมื่อเปรียบเทียบความสามารถในการแยกประเภทเพลงแต่ละแนว ลักษณะเด่นนี้สามารถแยกประเภทเพลง Classic และ Rock ได้เป็นอย่างดี และสามารถเพิ่มความถูกต้องในการแยกประเภทเพลง Hip-Hop ได้เมื่อเพิ่มจำนวนโหนดของโครงข่ายประสาทเทียม

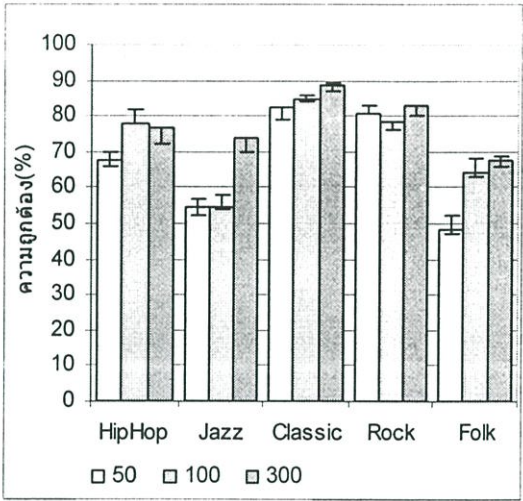
การเพิ่มจำนวนครั้งในการฝึกสอนมีผลทำให้ความถูกต้องในการแยกประเภทเพลง Hip-Hop Jazz และ Folk เพิ่มขึ้น



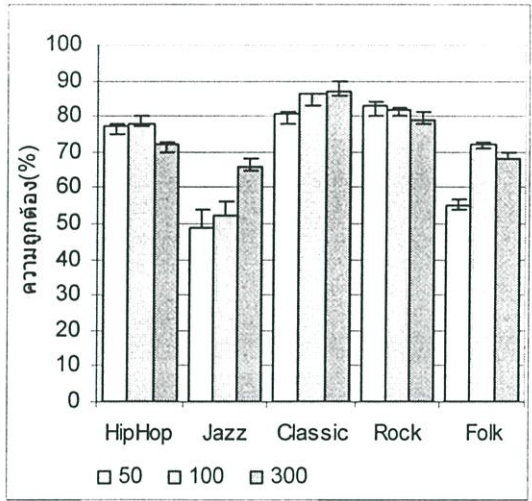
(ก)



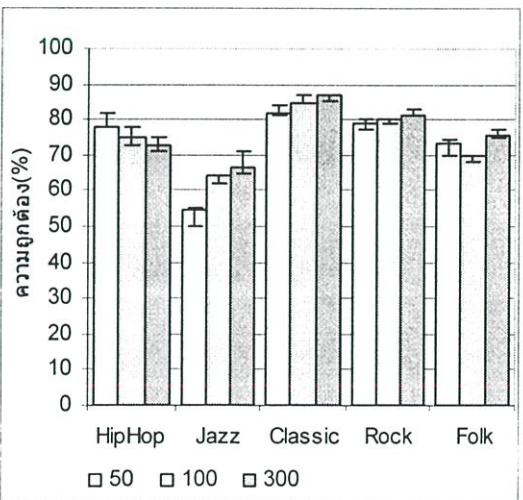
(ข)



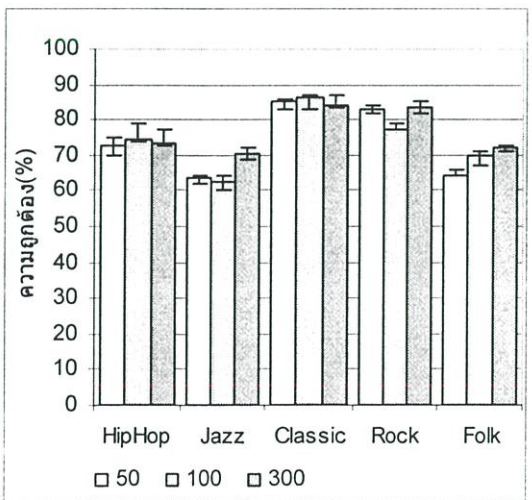
(ค)



(ง)



(จ)

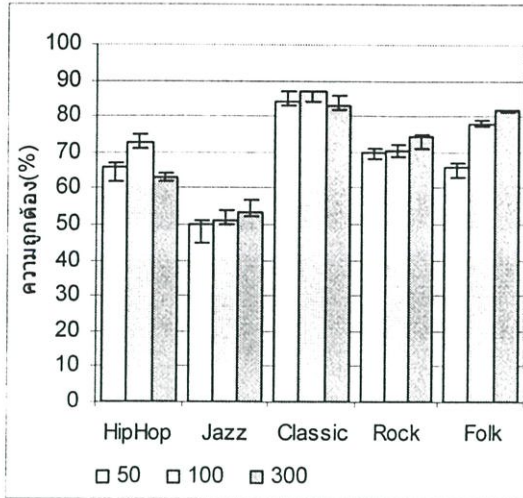


(ฉ)

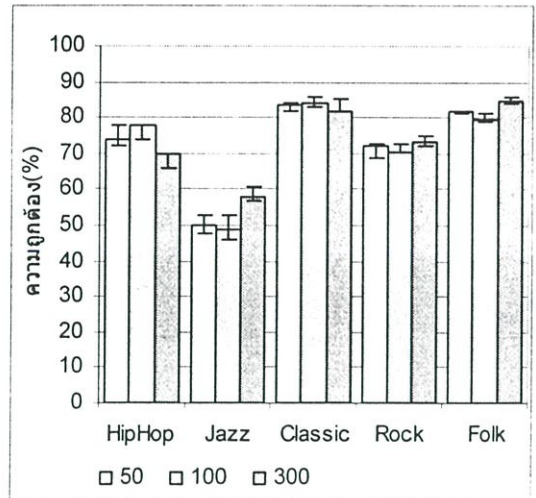
รูปที่ 6.3 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Contrast ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 12 และชั้นที่สอง 24 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 24 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 36 และชั้นที่สอง 72 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 72 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง 360 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด

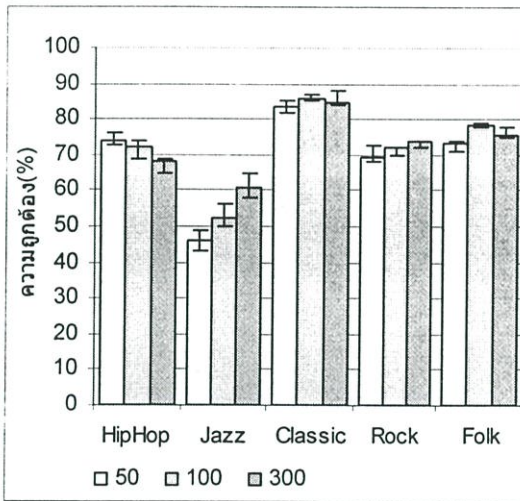
จากรูปที่ 6.3 ในลักษณะเด่นแบบ Octave-based Spectral Contrast นี้ การเพิ่มจำนวนโหนดของโครงข่ายประสาทเทียมและจำนวนครั้งในการฝึกสอน มีผลอย่างมากกับประสิทธิภาพในการแยกประเภทเพลง Jazz และเพลง Folk และมีผลเพียงเล็กน้อยกับเพลง Hip-Hop Classic และ Rock



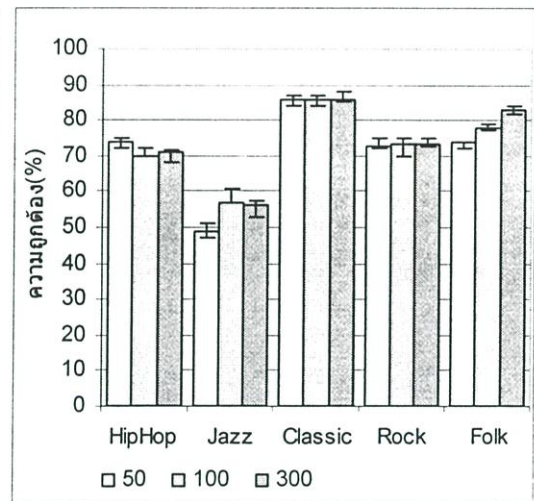
(ก)



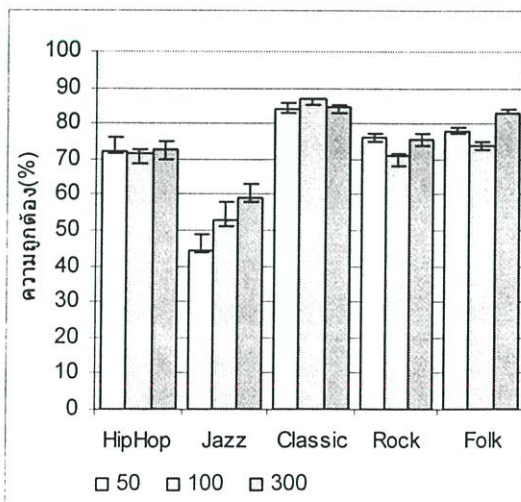
(ข)



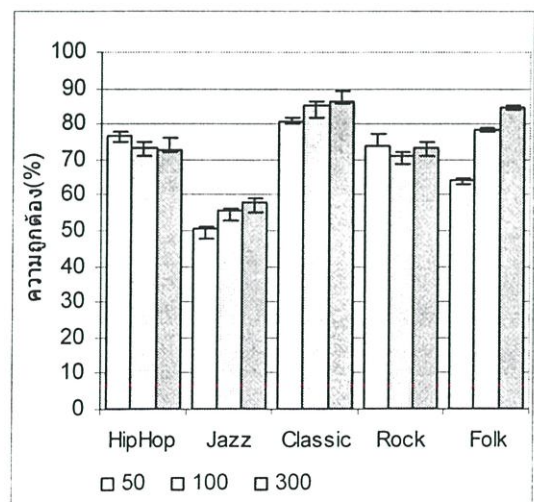
(ค)



(ง)



(จ)

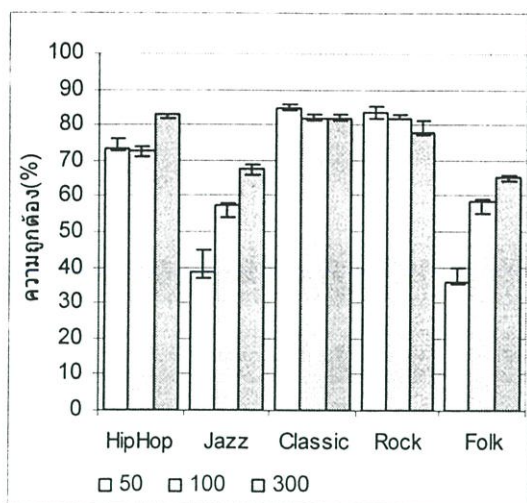


(ฉ)

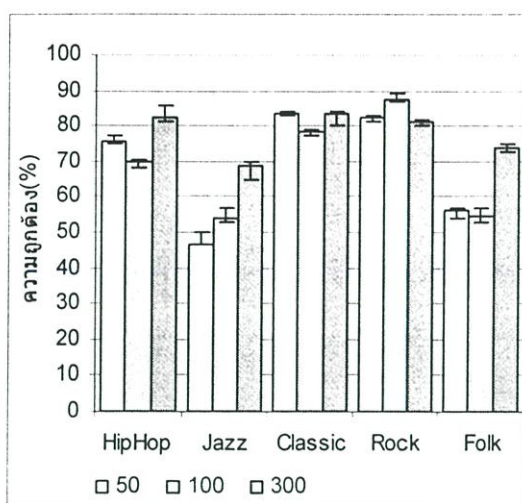
รูปที่ 6.4 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Contrast ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 96 และชั้นที่สอง 192 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 192 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 192 และชั้นที่สอง 384 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 384 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 384 และชั้นที่สอง 768 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 768 โหนด

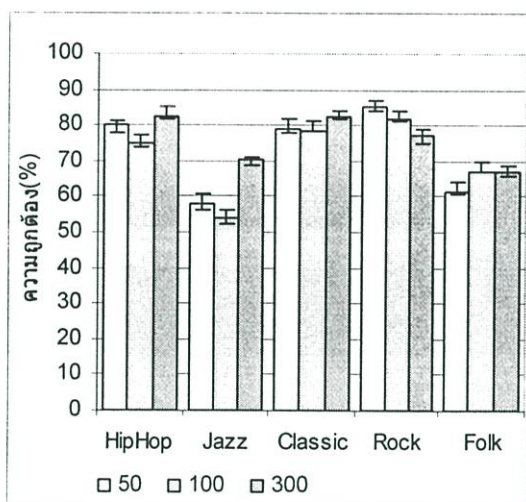
จากรูปที่ 6.4 ในลักษณะเด่น Mel-Frequency Spectral Contrast นี้ การเพิ่มจำนวนครั้งในการฝึกสอนมีผลทำให้ความถูกต้องในการแยกประเภทดนตรีของ เพลง Jazz และ Folk เพิ่มขึ้น แต่การเพิ่มจำนวนโหนดของ โครงข่ายประสาทเทียมมีผลเพียงเล็กน้อยกับความถูกต้อง



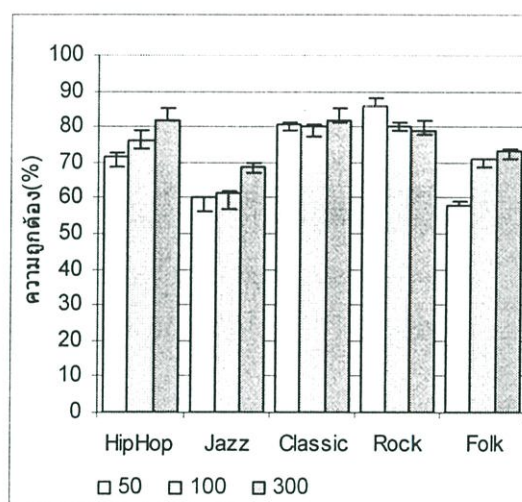
(ก)



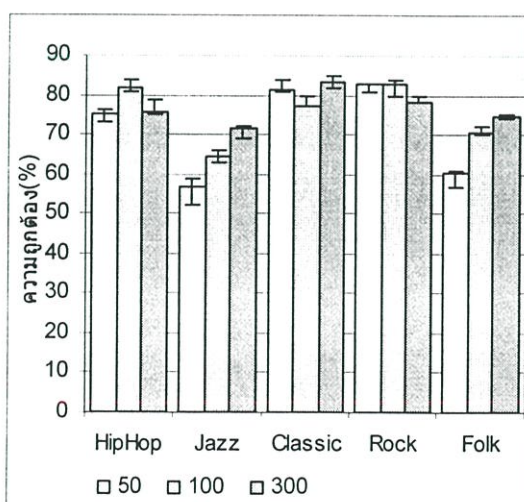
(ข)



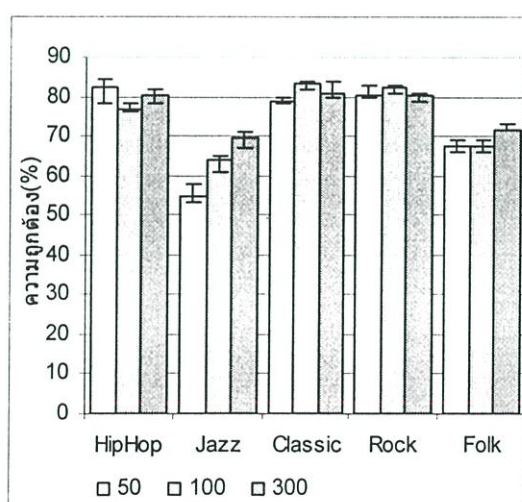
(ค)



(ง)



(จ)



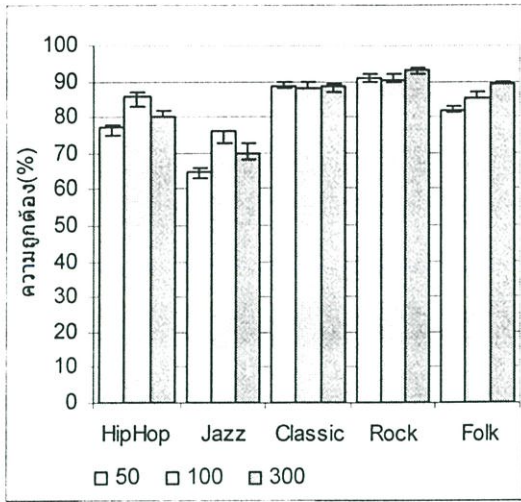
(ฉ)

รูปที่ 6.5 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Octave-based Spectral Histogram ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

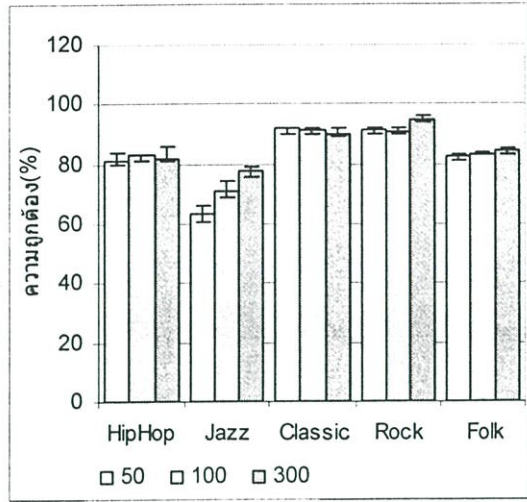
- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 60 และชั้นที่สอง 120 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 120 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 180 และชั้นที่สอง 360 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 360 และชั้นที่สอง 720 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 720 โหนด

จากรูปที่ 6.5 ในลักษณะเด่น Octave-based Spectral Histogram นี้คล้ายกับของลักษณะเด่น Mel-Frequency Spectral Contrast โดยการเพิ่มจำนวนครั้งในการฝึกสอนมีผลทำให้ความถูกต้องในการแยกประเภทดนตรีของ เพลง Jazz และ Folk เพิ่มขึ้น แต่การเพิ่มจำนวน โหนดของโครงข่ายประสาทเทียมมีผลเพียงเล็กน้อยกับความถูกต้อง

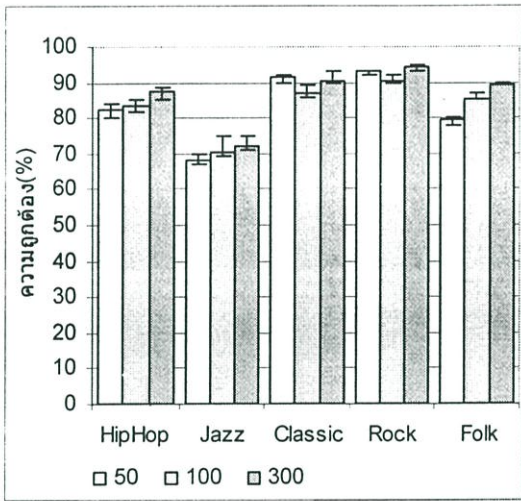
แต่การเพิ่มจำนวนครั้งในการสอนนั้น ทำให้ความถูกต้องในการแยกประเภทเพลง Hip-Hop เพิ่มขึ้นหรือลดลงไม่เป็นไปในทางเดียวกัน



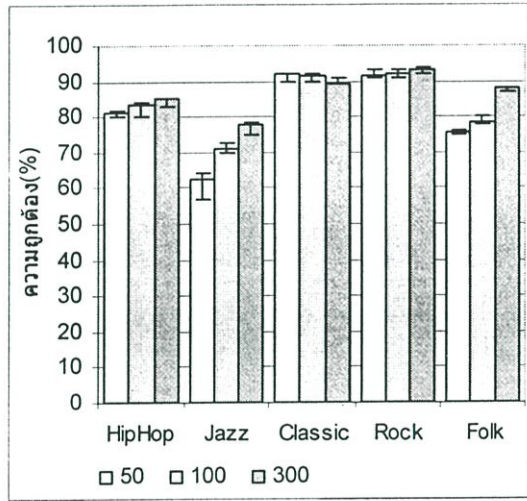
(ก)



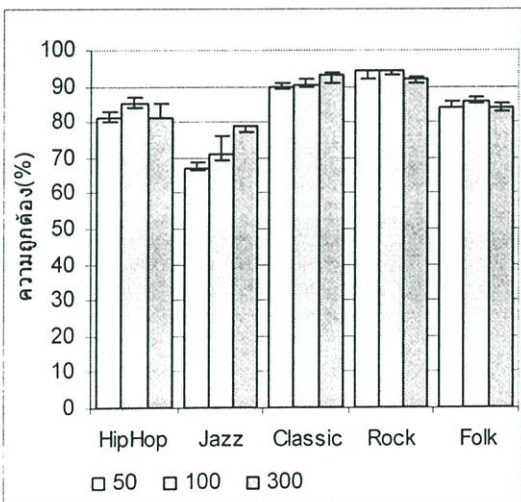
(ข)



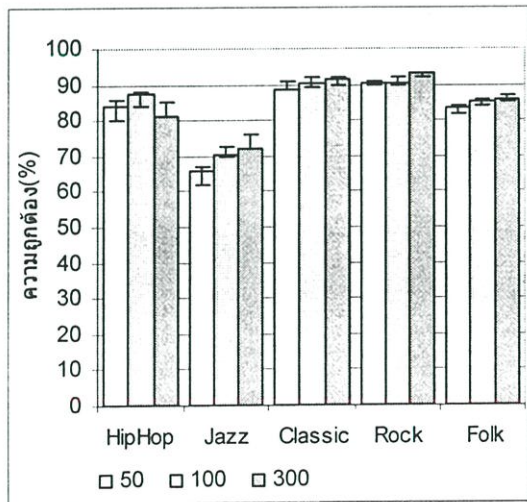
(ค)



(ง)



(จ)



(ฉ)

รูปที่ 6.6 กราฟเปรียบเทียบความถูกต้องโดยใช้ลักษณะเด่น Mel-Frequency Spectral Histogram ที่จำนวนครั้งฝึกสอน 50, 100, 300 ครั้ง

- ก) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 120 และชั้นที่สอง 360 โหนด
- ข) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 360 โหนด
- ค) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 240 และชั้นที่สอง 480 โหนด
- ง) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 480 โหนด
- จ) ใช้โครงข่ายประสาทเทียม 2 ชั้น โดยมีจำนวนโหนดในชั้นแรก 480 และชั้นที่สอง 960 โหนด
- ฉ) ใช้โครงข่ายประสาทเทียม 1 ชั้น โดยมีจำนวนโหนด 960 โหนด

จากรูป 6.6 ลักษณะเด่น Mel-Frequency Spectral Histogram นี้ ความถูกต้องใกล้เคียงกันเมื่อเพิ่มจำนวนโหนด หรือเพิ่มจำนวนครั้งในการฝึกสอน มีเพียงเพลง Jazz และ Folk ที่ความถูกต้องเพิ่มขึ้น เมื่อเพิ่มจำนวนครั้งในการฝึกสอน

ความถูกต้องโดยรวม ลดลงอย่างเห็นได้ชัด เมื่อลดจำนวนชั้นของโครงข่ายประสาทเทียมเหลือ 1 ชั้น ที่มีจำนวนโหนด 360 โหนด

ตารางที่ 6.38 เปรียบเทียบประสิทธิภาพในการแยกประเภทดนตรีโดยใช้ลักษณะเด่นและการปรับสัดส่วนทางความถี่ต่างๆ

	Spectral Mean	Spectral Contrast	Spectral Histogram
Octave-based Scaling	71.26%	77.84%	77.98%
Mel-Frequency Scaling	80.40%	75.10%	86.77%

จากตาราง 6.38 จะเห็นว่าในการใช้การปรับสัดส่วนทางความถี่แบบ Octave-based กับลักษณะเด่นแบบ Spectral Mean, Spectral Contrast และ Spectral Histogram มีประสิทธิภาพสูงขึ้นตามลำดับ แต่ความแตกต่างระหว่าง Spectral Contrast และ Spectral Histogram นั้นมีเพียงเล็กน้อยเท่านั้น

ส่วนในการใช้การปรับสัดส่วนทางความถี่แบบ Mel-Frequency นั้น ในลักษณะเด่นแบบ Spectral Mean ซึ่งใช้จำนวนตัวแปรน้อยกว่า Spectral Contrast ครั้งหนึ่ง แต่กลับให้ความถูกต้องมากกว่าถึง 5% นั้นหมายความว่าลักษณะเด่น Spectral Contrast นั้น จะช่วยเพิ่มประสิทธิภาพได้เมื่อใช้กับ ชุดข้อมูลที่ Spectrum ที่มีจำนวนน้อยเช่น Octave-based Spectrum ที่มีเพียง 6 ค่า ต่างกับลักษณะเด่น Spectral Histogram ซึ่งสามารถเพิ่มประสิทธิภาพได้ถึง 6% โดยต้องเพิ่มจำนวนข้อมูลทางความถี่ไปอีก 10 เท่า

ตารางที่ 6.39 ตารางจัดอันดับความถูกต้องในการแยกประเภทเพลงสูงสุด 3 อันดับแรก

Genre	อันดับ	ค่าความถูกต้อง (%)	Feature	จำนวน Input Node	จำนวน โหนดใน Layer 1	จำนวน โหนดใน Layer 2	จำนวน ครั้งฝึกสอน
Hip-Hop	1	87.7	Mel-Frequency Spectral Histogram	480	240	480	300
	2	87.5	Mel-Frequency Spectral Histogram	480	960	-	100
	3	86	Mel-Frequency Spectral Histogram	480	120	360	100
Jazz	1	78.7	Mel-Frequency Spectral Histogram	480	480	960	300
	2	77.7	Mel-Frequency Spectral Histogram	480	480	-	300
	3	77.7	Mel-Frequency Spectral Histogram	480	360	-	300
Classic	1	93.3	Mel-Frequency Spectral Histogram	480	960	-	300
	2	92.5	Octave-based Spectral Mean	6	18	36	50
	3	92.2	Mel-Frequency Spectral Histogram	480	360	-	50
Rock	1	95.1	Mel-Frequency Spectral Histogram	480	360	-	300
	2	94.33	Mel-Frequency Spectral Histogram	480	240	480	300
	3	94.11	Mel-Frequency Spectral Histogram	480	480	960	50
Folk	1	89.3	Mel-Frequency Spectral Histogram	480	120	360	300
	2	89.1	Mel-Frequency Spectral Histogram	480	240	480	300
	3	87.9	Mel-Frequency Spectral Histogram	480	480	-	300
Average	1	86.78	Mel-Frequency Spectral Histogram	480	240	480	300
	2	86.69	Mel-Frequency Spectral Histogram	480	480	-	300
	3	85.92	Mel-Frequency Spectral Histogram	480	480	960	300

ตาราง 6.39 คือการจัดอันดับความสามารถในการแยกประเภทดนตรีที่ได้จากการทดลองทั้งหมด จะเห็นว่า ลักษณะเด่นแบบ Mel-Frequency Spectral Histogram ให้ประสิทธิภาพในการแยกประเภทดนตรีที่ดีที่สุดในทุกแนวดนตรี

## บทที่ 7

### สรุปผลการวิจัยและข้อเสนอแนะ

วิทยานิพนธ์นี้ได้ศึกษาการแยกประเภทดนตรีออกเป็น 5 ประเภทโดยมีจุดประสงค์เพื่อเพิ่มความสามารถในการแยกประเภทดนตรีให้มีประสิทธิภาพสูงขึ้น ด้วยการทดลองใช้การปรับสัดส่วนทางความถี่ 2 แบบ คือการปรับสัดส่วนทางความถี่แบบเมล และการปรับสัดส่วนทางความถี่แบบฮ็อกเพฟ ซึ่งเป็นขั้นตอนหนึ่งในการปรับปรุงข้อมูลขั้นต้น ประกอบกับการทดลองเปรียบเทียบการดึงลักษณะเด่น 3 แบบ คือ ค่าเฉลี่ยทางความถี่ซึ่งเป็นการดึงลักษณะเด่นแบบพื้นฐาน, การตัดกันทางความถี่ซึ่งเป็นลักษณะเด่นที่ได้ศึกษาจากงานวิจัยก่อนหน้า ที่มีประสิทธิภาพในการแยกประเภทดนตรีสูงและใช้เพลงในการวิจัยจำนวนมาก เมื่อเทียบกับงานวิจัยอื่นๆ และลักษณะเด่นฮิสโทแกรมทางความถี่ ที่ถูกพัฒนาขึ้นในงานวิทยานิพนธ์เล่มนี้ ซึ่งต่อยอดมาจากลักษณะเด่นการตัดกันทางความถี่ เพื่อเพิ่มประสิทธิภาพในการแยกประเภทดนตรีให้สูงขึ้น โดยลักษณะเด่นฮิสโทแกรมทางความถี่ เป็นลักษณะเด่น ที่จะช่วยเพิ่มข้อมูล ให้มีมิติของปริมาณเพิ่มขึ้น เพื่อแยกแยะข้อมูลที่มีความซับซ้อนมากๆ ได้ วิทยานิพนธ์นี้ใช้ตัวแยกประเภทแบบ โครงข่ายประสาทเทียม ที่มีความสามารถในการแยกประเภทข้อมูลที่มีความซับซ้อนมากๆ ได้ โดยวิทยานิพนธ์นี้ได้ทดลองเปลี่ยน โครงสร้างของโครงข่ายประสาทเทียมเพื่อให้ได้ประสิทธิภาพในการแยกประเภทดนตรีดีขึ้น

จากผลการทดลอง พบว่า การแยกประเภทดนตรีโดยใช้โครงข่ายประสาทเทียม ร่วมกับลักษณะเด่นฮิสโทแกรมทางความถี่ นั้นให้ผลเป็นที่น่าพอใจ โดยการใช้การปรับสัดส่วนทางความถี่แบบเมล สามารถให้ความถูกต้องในการแยกประเภทดนตรีเฉลี่ยสูงถึง 86.77% และสามารถให้ความถูกต้องสูงในทุกๆแนวเพลง เมื่อเปรียบเทียบกับลักษณะเด่นอื่นๆ ข้อดีของลักษณะเด่นฮิสโทแกรมทางความถี่นี้ คือสามารถแยกความแตกต่างของเพลงฮิพฮอป แจ๊ส และ โฟล์ค ซึ่งมีความคล้ายกันได้

ส่วนการใช้ลักษณะเด่นการตัดกันทางความถี่นั้น มีประโยชน์มากเมื่อใช้กับการปรับสัดส่วนทางความถี่แบบฮ็อกเพฟ ซึ่งจะ ได้ข้อมูลของลักษณะเด่นจำนวนน้อยแต่ให้ประสิทธิภาพในการแยกประเภทดนตรีสูง และมีประโยชน์อย่างมาก หากนำไปประยุกต์ใช้ร่วมกับอุปกรณ์อิเล็กทรอนิกส์

หากเปรียบเทียบจำนวน โหนดของโครงข่ายประสาทเทียม ที่ใช้กับข้อมูลลักษณะเด่นแบบต่างๆแล้ว จะเห็นว่า การใช้ลักษณะเด่นฮิสโทแกรมทางความถี่ ที่ได้จากการปรับสัดส่วนทางความถี่แบบเมลนั้น ใช้จำนวนโหนดปริมาณมากที่สุด ซึ่งหมายถึงการใช้เวลาในการคำนวณปริมาณมากด้วย เพื่อที่จะ ได้ประสิทธิภาพในการแยกข้อมูลที่สูง แต่โครงข่ายประสาทเทียมที่มีจำนวนโหนด

มาก ไม่ได้ให้ประสิทธิภาพในการแยกประเภทสูงเสมอไป จำนวนโหนดนั้นต้องสัมพันธ์กับความซับซ้อนของข้อมูล และสัมพันธ์กับจำนวนอินพุตด้วย ในขณะที่การใช้การปรับสัดส่วนความถี่แบบอ็อกเทฟนั้น จะใช้จำนวนอินพุตที่น้อยกว่า แต่ให้ประสิทธิภาพสูงพอที่จะนำไปใช้ได้

จำนวนครั้งในการฝึกสอนโครงข่ายประสาทเทียมนั้น มีผลกับการรู้จำข้อมูลที่มีความซับซ้อนมาก เช่น เพลงแจ๊สและเพลงโฟล์ค และขึ้นอยู่กับจำนวนโหนดของโครงข่ายประสาทเทียมด้วย โดยถ้าโครงข่ายประสาทเทียมมีจำนวนโหนดมาก เวลาที่ใช้ในการฝึกสอนยิ่งต้องมากตาม โดยในผลการทดลองแต่ละครั้ง การใช้จำนวนครั้งในการฝึกสอนสูง จะได้ความถูกต้องสูงด้วย ดังนั้นไม่เพียงแต่จำนวนโหนดที่เหมาะสมเท่านั้นที่จะเพิ่มประสิทธิภาพของการแยกประเภทได้ แต่เรายังเพิ่มความสามารถในการรู้จำของโครงข่ายประสาทเทียมด้วยการปรับจำนวนครั้งในการฝึกสอนได้อีกด้วย

การประยุกต์ใช้งานโครงข่ายประสาทเทียมนี้ จะใช้กระบวนการฝึกสอนหรือปรับน้ำหนักในครั้งแรกเพียงครั้งเดียว แล้วเราสามารถนำโครงข่ายที่ถูกฝึกสอนแล้วนี้ ไปใช้งานแยกประเภทดนตรีได้ต่อไป เพราะฉะนั้นระยะเวลาในการฝึกสอนจึงไม่ใช่ปัญหาในการใช้งาน เราจึงทดลองเพิ่มปริมาณข้อมูลฝึกสอน โดยเพิ่มจากท่อนของเพลง 10 วินาที จำนวน 5 ท่อน เป็นการใช้ข้อมูลจากเพลงทั้งเพลง มาเปรียบเทียบกัน เพื่อหาความเป็นไปได้ที่จะเพิ่มความถูกต้องในการแยกประเภทดนตรี จากการทดลองพบว่า ความถูกต้องเมื่อทำการทดสอบกับชุดข้อมูลทดสอบนั้น มีค่าใกล้เคียงกัน แต่ความถูกต้องเมื่อทำการทดสอบกับชุดข้อมูลฝึกสอนนั้นกลับมีค่าลดลง เนื่องจากการกระจายของข้อมูล มีมากเกินไป โครงข่ายประสาทเทียมจึงไม่สามารถแยกความแตกต่างของข้อมูลมากมายเหล่านั้นได้

## เอกสารอ้างอิง

- [1] E. Scheirer and M. Slaney. **Construction and evaluation of a ro-bust multifeature speech/music discriminator.** vol. 2, pages 1331-1334. IEEE International Acoustics, Speech, and Signal Processing (ICASSP), April 1997.
- [2] Li, Toa and Tzanetakis, George “**Factors in Automatic Musical Genre Classification of Audio Signals**” *In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY October 2003.
- [3] T. Lambrou *et al.*, “**Classification of Audio Signals using Statistical Features on Time and Wavelet Transform Domains,**” in *Proc. ICASSP '98* , Seattle, USA, May 1998, pp. 3621–3624.
- [4] M.F. McKinney and J. Breebaart, “**Features for Audio and Music Classification,**” in *4th International Conference on Music Information*, [Online].Available : <http://ismir2003.ismir.net/papers/McKinney.PDF>
- [5] G. Tzanetakis and P. Cook, “**Musical Genre Classification of Audio Signal,**” *IEEE Transections on Speech and Audio Processing*, vol. 10, pp. 293-302, July 2002.
- [6] H. Soltau, T. Schultz, M. Westphal and A. Waibel. **Recognition of Music Types.** Seattle, WA, 1998. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [7] D. Pye, **Content-based methods for the management of digital music.** vol. 4, pages 2437-2440. IEEE International Conference on, Acoustics, Speech, and Signal Processing, 2000.
- [8] M. Casey, **General sound classification and similarity in MPEG-7.** Organized Sound, vol. 6, no. 2, pages 153-164, 2002.
- [9] D.-N. Jiang, L. Lu and H.-J. Zhang. **Music Type Classification by Spectral Contrast Features.** vol. 1, pages 113-116, Lausanne Switzerland, August 2002. IEEE International Conference on Multimedia and Expo (ICME).

- [10] J. J. Burred and A. Lerch. **A Hierarchical approach to automatic musical genre classification.** pages 308-311, London, UK, September 2003. International Conference on Digital Audio Effects (DAFx-03).
- [11] C. Xu, N. C. Maddage, X. Shao, F. Cao and Q. Tian. **Musical Genre Classification Using Support Vector Machines.** vol. 5, pages 429-432. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2003.
- [12] D. Perrot and R. Gjerdigen. **Scanning the dial: An exploration of factors in the identification of musical style.** page 88(abstract). Society for Music Perception and Cognition, 1999.
- [13] Moller, M. F., "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525-533, 1993.

ภาคผนวก ก.

โปรแกรมที่ใช้ในการทดลอง

โปรแกรมทั้งหมดถูกเขียนโดยใช้โปรแกรม MATLAB รุ่น 6.5.0.180913a Release 13

## 1. การอ่านข้อมูลจากไฟล์ .wav

```
MusicData = wavread(['F:\Music\HIPHOP\HipHop1.wav'],[1 44100]);
```

คือการเปิดข้อมูลตัวที่ 1 ถึงตัวที่ 44100 จากไฟล์ HipHop1.wav

```
SongSize = wavread(['F:\Music\HIPHOP\HipHop1.wav'],'size');
```

คือการหาขนาดข้อมูลของเพลง HipHop1.wav

## 2. การทำ Pre-emphasis

```
Preemph = 0.7;
```

```
MusicData = filter([1 -Preemph], 1, MusicData);
```

คือการทำ Frequency Filter ด้วยค่า Pre-emphasis Factor 0.7

## 3. ฟังก์ชันการหา Power Spectrum ด้วยการทำให้ Short-Time Fourier Transform

```
PowerSpectrum = powspec(MusicData, 44100, 0.2, 0.1)
```

คือการหา Power Spectrum ด้วยการทำให้ STFT จากข้อมูล MusicData ที่มี Sampling Rate 44100 และมี window size = 0.2 วินาที และ hop size = 0.1 วินาที โดยที่ Function "powspec" คือ

```
function y = powspec(x, sr, wintime, steptime)
```

% compute the powerspectrum of the input signal.

% basically outputs a power spectrogram

% each column represents a power spectrum for a given frame

% each row represents a frequency

```
winpts = round(wintime*sr);
```

```
steppts = round(steptime*sr);
```

```

%to find the match number of FFT and the Sampling Rate
NFFT = 2^(ceil(log(winpts)/log(2)));
WINDOW = hamming(winpts);
NOVERLAP = winpts - steppts;

SAMPRATE = sr;
% Values coming out of rasta treat samples as integers,
% not range -1..1, hence scale up here to match (approx)
y = abs(specgram(x*32768,NFFT,SAMPRATE,WINDOW,NOVERLAP)).^2;

```

#### 4. การทำ Mel-Frequency Scaling

```

sr = 44100;
nbands = 50;
minfreq = 0;
maxfreq = 22050;
bwidth = 1;
MelSpectrum = audspec(PowerSpectrum, sr, nbands, minfreq, maxfreq, bwidth);
MelSpectrum = postaud(MelSpectrum, maxfreq);

```

คือการแปลง PowerSpectrum ให้อยู่ใน Scale ของ Mel ซึ่งตัวแปร MelSpectrum โดยมี sr คือ sampling rate = 44100, nbands คือจำนวน Mel ที่ต้องการ = 50 Mel, ความถี่ต่ำสุดที่ต้องการแปลง 0 Hz, ความถี่สูงสุดที่ต้องการแปลง 22050 Hz

โดยที่ Function "**audspec**" คือ

```

function [aspectrum,wts] = audspec(pspectrum, sr, nfilts, minfreq, maxfreq, bwidth)
% perform critical band analysis (see PLP)
% takes power spectrogram as input

[nfreqs,nframes] = size(pspectrum);
nfft = (nfreqs-1)*2;
wts = fft2melmx(nfft, sr, nfilts, bwidth, minfreq, maxfreq);

```

```

wts = wts(:, 1:nfreqs);

% Integrate FFT bins into Mel bins, in abs domains:
aspectrum = wts * pspectrum;

%% Function "postaud"
function [y,eql] = postaud(x,fmax)

% do loudness equalization and cube root compression
% x = critical band filters
% rows = critical bands
% cols = frames

[nbands,nframes] = size(x);

% Include frequency points at extremes, discard later
nfpts = nbands;

bandcfhz = mel2hz(linspace(0, hz2mel(fmax), nfpts));

% Remove extremal bands (the ones that will be duplicated)
bandcfhz = bandcfhz(1:nfpts);

% Hynes's magic equal-loudness-curve formula
fsq = bandcfhz.^2;
ftmp = fsq + 1.6e5;
eql = ((fsq./ftmp).^2) .* ((fsq + 1.44e6)./(fsq + 9.61e6));

% weight the critical bands
z = repmat(eql',1,nframes).*x;

```

```
% cube root compress
z = z .^ (.33);

% replicate first and last band (because they are unreliable as calculated)
y = z([2,2:(nbands-1),nbands-1],:);
```

## 5. การหา Spectral Contrast

```
for j = 1:50;
    SortSpectrum = (sort(MelSpectrum(j,(1:100))));
    (เพื่อเรียงลำดับ Mel Spectrum 100 Frames ที่ความถี่เมล ที่ j)
    Valley = log(mean(SortSpectrum (1:2)));
    Peak = log(mean(SortSpectrum (99:100)));
    (หา Peak และ Valley จากค่าเฉลี่ยของ Spectrum 2 จุดสุดท้ายและ 2 จุดแรก)
    SpectContrast(j,i) = Peak - Valley;
    SpectCpnttrast(j+50,i) = Valley;
end;
```

## 6. Neural Network

```
InputData =[SpectConHipHop SpectConJazz SpectConClassic SpectConRock SpectConFolk];
net = newff(minmax(InputData),[360 720 5],{'tansig' 'tansig' 'tansig'},'trainscg');
(สร้าง Neural Network ใหม่ชนิด Feed Forward โดยมีชั้นซ่อน 2 ชั้น ชั้นแรกมี 360 Node ชั้นที่
2 มี 720 Node และแต่ละชั้นใช้ Function Tan Sigmoid การ Train ใช้ขั้นตอนของ Scaled Conjugate
Gradient - trainscg)

net = init(net);
(สุ่มค่าเริ่มต้นของน้ำหนักและไบแอส)
```

```
net.trainParam.epochs = 100;
```

```
net.trainParam.show = 5;
```

(ตั้งจำนวนรอบในการ Train = 100 ครั้ง และแสดงผลทุกๆ 5 รอบที่ทำการ Train)

```
OutputData (1,1:100)=1;
```

```
OutputData (2,101:200)=1;
```

```
OutputData (3,201:300)=1;
```

```
OutputData (4,301:400)=1;
```

```
OutputData (5,401:500)=1;
```

(ตั้ง Output สำหรับ Output Node ให้เป็น 1 เมื่อถูกเทรนด้วย Input ที่เป็นเพลงแนวนั้นๆ)

```
net = train(net,InputData,OutputData);
```

(สั่งให้ Train Neural Network)

```
Result = sim(net,InputData);
```

(ทดลองแยกประเภทด้วย Neural Network ที่ Train แล้ว ด้วย Input Data ชุดที่ใช้ Train)

ภาคผนวก ข.

## ผลงานวิจัยที่ได้รับการตีพิมพ์

1. **N. Ampawanon** and Y. Kidjaidure. "**Automatic Music Genre Classification using Neural Network**," Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology International Conference (ECTI'2006), pp. 769-772, Ubon Ratchathani, Thailand, May 10-13, 2006.

# ECTI-CON 2006

THE 2006 ECTI INTERNATIONAL CONFERENCE



**ECTI**  
Association



## Enter to the Proceeding

Proceedings of the 2006 Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology (ECTI) International Conference

May 10-13, 2006

Ubonburi Hotel, Ubon Ratchathani, THAILAND

ECTI Association



HELEUP  
Thailand Chapter

ECTI THAILAND

NECTEC

THAIPOST

ECTI CAS Society

ECTI CAS Society

ECTI CAS Society

# Automatic Music Genre Classification Using Artificial Neural Network

Nuntaput Anpawanon and Yuttana Kidjaidure

Faculty of Engineering King Mongkut's Institute of Technology Ladkrabang,  
Ladkrabang, Bangkok 10520 Thailand.  
Email: amman213@hotmail.com and kkyuttan@kmitl.ac.th

## ABSTRACT

This research is about classification of music genre into hip-hop, jazz, classic, rock and folk that is always seen in the pre-set equalizer of stereo. Difficulty is the variety of music. Spectral feature seems to be the most appropriate feature so we propose a new model that uses spectral based called Spectral Histogram. Using Mel-Frequency scale and Octave-Based scale, we can reduce the number and variation of spectral data. Three kinds of spectral feature - Spectral Mean, Spectral Contrast and Spectral Histogram- are tested with Neural Network to compare the performance of each feature. In this paper we use Back-propagation Neural Network with scaled conjugate gradient training method for data classifying. There are 1000 songs in the database. Five-hundred songs are training-set data and other 500 songs are test-set data.

**Keywords:** Music Genre Classification, Neural Network, Scaled Conjugate Gradient, Spectral Histogram, Spectral Contrast, Mel-Frequency, Octave-Based

## 1. Introduction

Musical Genre is an important description for music classifying which can represent the type of instrument being played in the song and its configuration and also represent playing styles, notes and scales of music.

Nowadays, music in digital media is very popular. Digital songs are distributed throughout the internet. Database management is important for a large library of songs. When people search the songs, they would talk about artist name, album name and genre. Automatic music genre classification is helpful. Music genre identification can also be the new feature of digital media player both in software and hardware.

Automatic Music Genre Classification is attracting interest from the researchers. Since 2000, there are many researches in this field. Many kinds of feature are tested with various classifiers. In early work [1], MFCC (Mel-Frequency Cepstral Coefficient) is used as feature with GMM (Gaussian Mixture Model) and TreeQ to classify 175 songs into 6 genres. The accuracy is 92% in GMM and 90% in TreeQ. In [2] Tzanetakis compared 3 feature sets including Trimbral Texture, Beat Histogram and Pitch Content with GMM classifier, classifies 1000 songs into 10 genres and get 61% accuracy. Dan-Ning Jiang [3] can reach 82% accuracy with the new feature Spectral Contrast in classifying 1500 songs to 5 genres. This new

feature is very interesting because it combines spectral data and temporal data. Li [4] used Daubechies Wavelet Coefficient Histogram with 4 classifiers GMM, k-NN, LDA and SVM. The accuracy of 79% in SVM classifier is the best performance. In [5] they tried to classify 1,699 songs into 4 genres consisting of Pop, Rock, Jazz and Classic with the high-level feature that they call Instrumental Feature. This feature retrieves the kind of instrument using in the songs. This kind of feature is good for classifying Jazz and Rock music. In the recent work [6-8], they used MFCC as features and tried to develop the classifier model for the higher accuracy.

Principal Component Analysis is often used as the pre-processor before the classification. This way, the data would be destroyed. Also, MFCC produced from Mel-Frequency Spectrum via DCT destroys some of information. This makes limitation of the classification method. Thus we selected the Artificial Neural Network that can be used with many kinds of feature. It is a very flexible classifier. Furthermore, the trained Neural Network is very easy to implement, both in software and hardware. Three features used in this paper are Spectral Mean, Spectral Contrast and Spectral Histogram.

In section 2, we will describe the features extraction. Artificial Neural Network that is used is described in section 3. Experimental configuration is explained in section 4. The results are shown in section 5 and the conclusion provided in section 6.

## 2. FEATURES EXTRACTION

In this paper, we will focus on spectral features because high variation of music tempo would not affect the classification. Mel-Frequency Scaling and Octave-Based Scaling are used to reduce the number of spectrum. Log scale is used to reduce the variance of spectral amplitude. All power spectrums are retrieved from audio signal by Short-Time Fourier Transform (STFT). Three features -Spectral Mean, Spectral Contrast and Spectral Histogram- are compared based on both Mel-Frequency Scale and Octave-Based scale. Spectral Histogram is a new feature that is the improved version of Spectral Contrast.

### 2.1 Spectral Mean

First, the audio signal would be equalized with pre-emphasis. Then we perform the STFT on the signal to get the power spectrum with the window size of 0.2 sec and

hop size of 0.1 sec and all frames are smoothed by Hamming's Window. We scale the spectrums with Mel-Frequency scale to reduce a large number of spectrums to 50 Mels, and also scale these spectrums with Octave-Based scale to reduce number of them to 6 Octave-Based sub-bands. We then take log to equalize amplitude of the spectrums. Finally, we average 99 frames of spectrum to one frame of spectral mean. This kind of feature can represent the overall tone of the song.

## 2.2 Spectral Contrast

In [3] we have seen that spectral contrast can make a high accuracy for classification. The spectral contrast is the difference between the strength of spectral peak and the strength of spectral valley. First the audio signal is equalized with pre-emphasis. And then the STFT is performed on the signal with window size of 0.2 sec and hop size of 0.1 sec like the Spectral Mean to get power spectrums. The spectrums would be scaled with both Mel-Frequency scaling and Octave-Based scaling. Then the scaled spectrums are sorted in descending order in each band. We can get the spectral peak and the spectral valley by taking log to the average value in the small neighborhood around the maximum and minimum value. In [3], they use 0.02 of the total number of frames in each sub-band and also used in this paper. To keep more spectral information, the strength of spectral valley is kept in the spectral contrast feature.

The strength of peak and valley can be represented in these equations:

$$Peak_k = \log\left\{\frac{1}{\alpha N} \sum_{i=0}^{\alpha N} x_{k,i}\right\} \quad (1)$$

$$Valley_k = \log\left\{\frac{1}{\alpha N} \sum_{i=0}^{\alpha N} x_{k,N-i}\right\} \quad (2)$$

And their difference is:

$$SC_k = Peak_k - Valley_k \quad (3)$$

where  $N$  is the total number in  $k$ -th sub-band,  $k \in [1,6]$  in Octave-based scale and  $k \in [1,50]$  in Mel-Frequency scale, and  $x$  is the sorted power spectrum in  $k$ -th sub-band.

$\{SC_k, Valley_k\}$  is used as the Spectral Contrast feature. The strength of spectral valleys is also contained in the feature to keep more spectral information.

## 2.3 Spectral Histogram

From the Spectral Contrast feature, sorted spectrums in each sub-band contain some information. They represent the distribution of power in each band of frequencies. Therefore, we call it Spectral Histogram. In Spectral Contrast, only the difference between spectral peak and spectral valley is used. However, the shape of spectral distribution is significant.

To get the Spectral Histogram, the scaled spectrums must be sorted. In this paper, we have 99 frames of spectrums. Thus, the sorted spectrums need to be sampled to reduce the number of feature. After the sampling, we will get 10 values in each sub-band, which are totally 500 values on Mel-Frequency scaling input and 60 values on Octave-Based scaling input.

Comparison of calculation procedure of these three feature extractions is shown in figure 1.

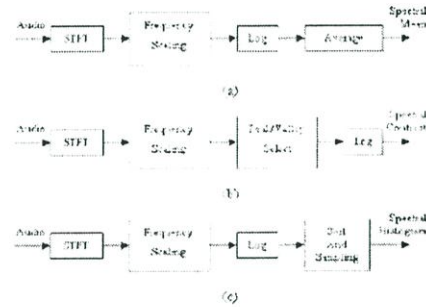


Fig.1: Procedure of (a) Spectral Mean (b) Spectral Contrast (c) Spectral Histogram

## 3. NEURAL NETWORK

The neural network that is used in this paper is Multi Layers Back-Propagation Neural Network. There are 400 nodes in the first layer, 800 nodes in the second layer and 5 Output nodes. The number of Input Neurons depends on the number of dimensions of each feature. We used 6 and 50 dimensions with Octave-Based Spectral Mean and Mel-Frequency Spectral Mean, 12 and 100 dimensions in Octave-Based Spectral Contrast and Mel-Frequency Spectral Contrast, 60 and 500 dimensions in Octave-Based Spectral Histogram and Mel-Frequency Spectral Histogram respectively. The Output classes are Hip-Hop, Jazz, Classical, Rock, and Folk. Tan Sigmoid function is used in all layers. The Neural Network is trained using Scaled Conjugate Gradient method ("traincg" procedure in MATLAB 6.5) in [9]. This training algorithm was developed from Levenberg-Marquardt algorithm and Conjugate Gradient Algorithm to decrease time consumed by Conjugate Gradient algorithm and also use less memory than Levenberg-Marquardt algorithm. Structure of the Neural Network is shown in figure 2.

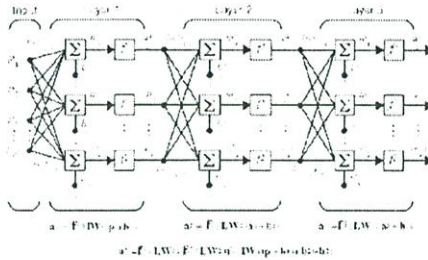


Fig. 2: Structure of The Neural Network

4. EXPERIMENTS

All songs in our experiments are collected from www.amazon.com [10] and www.download.com [11] in the free download section. They are full MP3 files including 1,000 songs in 5 genres. In each genre, the sets of 200 songs are divided into 100 songs of training-set data and 100 songs of testing-set data. Afterward we convert MP3 files to WAV files with 44100-KHz sampling rate, 16 Bits of depth in mono type. Then we take a 10-second clip from the middle of each song. All clips are equalized with 0.7 pre-emphasis. STFT is used to get the power spectrums of audio signal with 200-ms window size and 100-ms hop size. Thus, 99 frames are obtained from a clip. A set of feature is extracted from these 99 frames of spectrum. All sets of feature in all genres are used to train the Neural Network for 100 times in batch mode. The target values are set to 1 for right genre class and 0 for the wrong ones.

5. CLASSIFICATION RESULTS

Results would be represented in confusion matrix. These are the results of classification in testing-set data. Using Spectral Mean Features, accuracy of the classification is shown in table 1 and 4. Table 2 and 5 is the results of using Spectral Contrast features. Table 3 and 6 shows the results of using Spectral Histogram features.

Table 1: Confusion Matrix of Using 6 Octave-Based Spectral Mean Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	82	1	0	17	0
Jazz	8	56	7	2	27
Classic	1	4	86	2	7
Rock	12	8	1	78	1
Folk	3	40	3	6	48

Table 2: Confusion Matrix of Using 6 Octave-Based Spectral Contrast Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	83	9	1	6	1
Jazz	10	65	3	0	22
Classic	2	1	84	5	8

Rock	6	3	3	81	7
Folk	7	16	3	4	70

Table 3: Confusion Matrix of Using 6 Octave-Based Spectral Histogram Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	80	9	0	11	0
Jazz	3	76	3	0	18
Classic	7	1	86	3	3
Rock	4	1	1	89	5
Folk	3	12	3	2	80

Table 4: Confusion Matrix of Using 50 Mel-Frequency Spectral Mean Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	80	8	0	10	2
Jazz	10	59	7	4	20
Classic	2	1	94	1	2
Rock	7	0	3	88	2
Folk	1	17	7	0	75

Table 5: Confusion Matrix of Using 50 Mel-Frequency Spectral Contrast Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	78	9	0	9	4
Jazz	8	65	7	2	18
Classic	2	5	89	1	3
Rock	10	4	1	83	2
Folk	9	15	9	1	66

Table 6: Confusion Matrix of Using 50 Mel-Frequency Spectral Histogram Features

	HipHop	Jazz	Classic	Rock	Folk
HipHop	80	9	0	6	5
Jazz	10	74	6	3	7
Classic	1	5	93	1	0
Rock	3	2	1	90	4
Folk	1	6	5	3	85

Table 7: Comparison of Overall Classification Accuracy

	SM	SC	SH
Octave-Scale	70%	76.6%	82.2%
Mel-Scale	79.2%	76.2%	84.4%

Table 7 is the comparison of overall classification accuracy of each feature where SM is Spectral Mean Feature, SC is Spectral Contrast Feature and SH is Spectral Histogram Feature. Using Spectral Mean Feature, we obtain a noticeable difference between Octave-Scale and Mel-Scale because there are too few dimensions of Octave-Scale features. The best

performance feature is Spectral Histogram with both the Octave-Scale and Mel-Scale.

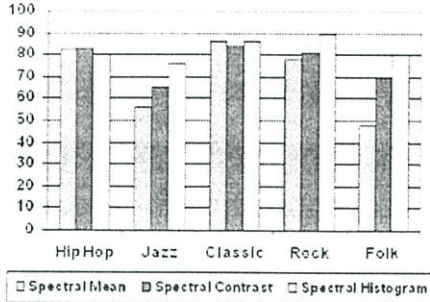


Fig. 2: Comparison of Octave-Based Features

In Fig. 2, we compare each feature in Octave-Based Scale. There is a little difference between the features in HipHop and Classic type while in Jazz and Folk, the classification accuracy is higher in Spectral Contrast and highest in Spectral Histogram. In Rock type, the classification is the best when we used Spectral Histogram.

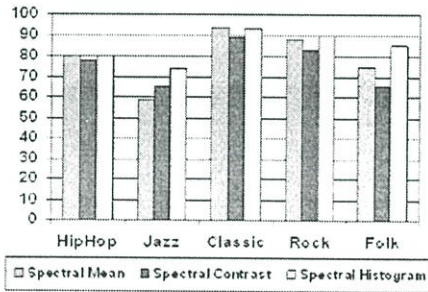


Fig. 3: Comparison of Mel-Frequency Features

Figure 3 is the comparison of the features in Mel-Frequency Scale. In Jazz and Folk type, we can increase the performance with the Spectral Histogram feature. The classification accuracy of Jazz is always low in many researches because of its variety.

## 6. CONCLUSION

The proposed feature, Spectral Histogram can perform the best classification with Neural Network. It can reach 84.4% of accuracy more than Spectral Mean and Spectral Contrast on the 10-second music clips. With this kind of feature, there is a little difference between scaling with Mel-Scale and Octave-Scale. As a result, if we would like to avoid the time and memory consumed, we can use Octave-Based Spectral Histogram to get

higher accuracy. Classification accuracy of Jazz and Folk can be increased because it can represent the shape of power distribution of each frequency band. This shape of power distribution also contains the temporal behavior of music signal that could play an important role in identifying the genre of musics.

## 7. REFERENCES

- [1] D. Pye, "Content-based methods for the management of digital music," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pages 2437-2440, 2000.
- [2] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 293-302, July 2002.
- [3] Dan-ning Jiang, Lie Lu, Hong-Jiang Zhang, Lian-Hong Cai, Jian-Hua Tao, "Music Type Classification By Spectral Contrast Features", *Proc. of IEEE International Conference on Multimedia and Expo (ICME'02)*, Lansanne Switzerland, Aug 26-29, 2002.
- [4] Li, Tao and Tzanetakis, George "Factors in Automatic Musical Genre Classification of Audio Signals" *In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY October 2003
- [5] Jiajun Zhu, Xiangyang Xue and Hong Lu, "Musical genre classification by instrumental features," *the International Computer Music Conference (ICMC2004)*, New Jersey, USA, 2004.
- [6] Tao Li and Mitsunori Ogihara, "Music genre classification with taxonomy," *In International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, USA, March 2005.
- [7] P. Ahrendt, A. Meng, and J. Larsen, "Decision time horizon for music genre classification using shorttime features," *In Proc. of IUSPACO*, pages 1293-1296, 2004.
- [8] A. Meng, P. Ahrendt, and J. Larsen, "Improving music genre classification by short-time feature integration," *In Proc. of ICASSP*, pages 1293-1296, 2005.
- [9] Moller, M. F., "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525-533, 1993.
- [10] [www.amazon.com](http://www.amazon.com), "Free Music Download section," 2005
- [11] [www.download.com](http://www.download.com), "Free Music Download section," 2005

## ประวัติผู้เขียน

นายันทภัส อัมพวานนท์ เกิดเมื่อวันที่ 6 กุมภาพันธ์ พ.ศ.2525 ที่จังหวัดอุดรดิตถ์ สำเร็จ การศึกษาระดับประกาศนียบัตรวิชาชีพ (ปว.ช.) สาขาไฟฟ้าและอิเล็กทรอนิกส์ จากคณะวิทยาลัย เทคโนโลยีอุตสาหกรรม สถาบันพระจอมเกล้าพระนครเหนือ สำเร็จการศึกษาระดับปริญญาตรี วิศวกรรมศาสตรบัณฑิต (วศ.บ.) สาขาวิศวกรรมอิเล็กทรอนิกส์ จากภาควิชาวิศวกรรมไฟฟ้า คณะ วิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2546 และเข้า ศึกษาต่อในระดับปริญญาโท หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต (วศ.ม.) สาขาวิศวกรรม อิเล็กทรอนิกส์ ภาควิชาวิศวกรรมอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอม ก่อเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2547