

บทที่ 3

วิธีการดำเนินงานวิจัย

ในส่วนนี้จะกล่าวถึงรายละเอียดขั้นตอนการดำเนินการวิจัยการจำแนกเพลงไทยและฝรั่ง ด้วยโครงข่ายประสาทเทียมโดยใช้คุณลักษณะเชิงคุณภาพเสียง รายละเอียดในขั้นตอนนี้ประกอบด้วย ส่วนแรกเป็นการจำแนกประเภทเพลง 8 ประเภทใน 1 ขั้นตอน ส่วนที่สองเป็นการจำแนกเพลง 8 ประเภทใน 2 ขั้นตอน ส่วนที่สามเป็นผลการจำแนกเพลงที่มีทำนองเดียวกันแต่ใช้ภาษาแตกต่างกัน และส่วนที่สี่ส่วนสุดท้ายที่จะสรุปผลการทดลองโดยรวมทั้งหมด

3.1 การจำแนกเพลง 8 ประเภทใน 1 ขั้นตอน

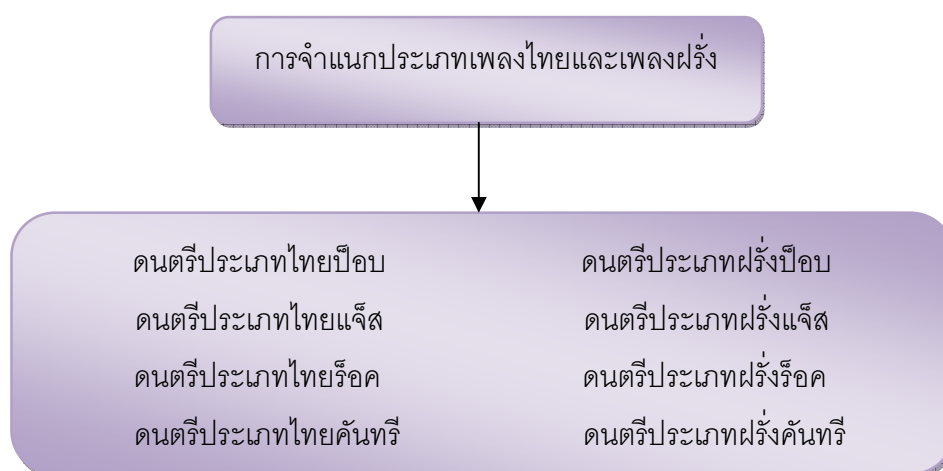
การทดลองในส่วนนี้เป็นการใช้ค่าคุณลักษณะเชิงคุณภาพเสียงทั้ง 7 ค่าประกอบด้วย MFCC LPC ZC SC SF SR และ SSB เพื่อทำการจำแนกเพลงป๊อป ร็อค คันทรี และ แจ๊ส ของทั้งสองภาษารวมเป็น 8 ประเภทโดยการทดลองในส่วนนี้ประกอบด้วย 7 ส่วน ส่วนแรกเป็นขั้นตอนเตรียมการทดลอง ส่วนที่สองเป็นการปรับค่าพารามิเตอร์ที่ให้การจำแนกที่ดีที่สุด ส่วนที่สามเป็นการวิเคราะห์การหาค่ากลาง ส่วนที่สี่เป็นการเปรียบเทียบค่าคุณลักษณะเชิงคุณภาพเสียงที่กึ่งกลางเพลง ส่วนที่ห้าเป็นการเปรียบเทียบค่าคุณลักษณะที่ช่วงเวลา 0 – 10 0 – 20 และ 0 - 30 วินาทีและในส่วนที่หกเป็นการจำแนกโดยใช้ขนาดเฟรม ในส่วนสุดท้ายส่วนที่เจ็ด การจำแนกประเภทเพลงที่ดีที่สุดในการค่าคุณลักษณะ MFCC 0 – 4

3.1.1 ขั้นตอนเตรียมการทดลอง

ในขั้นตอนการเตรียมการศึกษาคือจะรวบรวมบทความ วิจารณ์ วิชานิพนธ์ รวมถึง การศึกษาในส่วนอื่นๆที่เกี่ยวข้องกับการจำแนกประเภทดนตรีทั้งภายในและภายนอกประเทศเพื่อทำ ความเข้าใจทั้งในด้านทฤษฎีและปฏิบัติโดยขั้นตอนในส่วนแรกการจำแนกประเภทเพลง 8 ประเภท 1 ขั้นตอน เป็นไปตามภาพ 3.1

ภาพที่ 3.1

ภาพแสดงการจำแนกเพลง 8 ประเภทใน 1 ขั้นตอน



จากภาพแสดงการจำแนกเพลง 8 ประเภทใน 1 ขั้นตอน เมื่อได้ดึงค่าคุณลักษณะเชิง คุณภาพจากเพลงแล้วจะจำแนกเพลงออกเป็น 8 ประเภท เพื่อหาค่าความถูกต้องที่ดีที่สุดจากกลุ่มค่า คุณลักษณะเชิงคุณภาพเสียง

หลังจากนั้นจะทำการรวบรวมเพลงทั้งเพลงไทยและเพลงฝรั่งจากทีมงาน 116 Zstudio Record มีการระบุหมวดหมู่ไว้แล้ว โดยคัดเลือกประเภทเพลง ป๊อป ร็อค แจ๊ส และคันทรีของทั้งสอง ภาษา เนื่องจากข้อมูลเพลงที่ได้เป็นประเภทที่แตกต่างหลายประเภท ดังนั้นเราจะต้องแปลงสัญญาณ เพลงที่ได้ให้มีลักษณะที่เป็นประเภทเดียวกันก่อน ซึ่งขั้นตอนในการแปลงจะต้องแปลงให้เป็นนามสกุล .wav ด้วยโปรแกรมแปลงสัญญาณเสียง Goldwave โดยในการแปลงค่าสุ่มหน้าคลื่นสัญญาณเพื่อ

แทนด้วยสัญญาณดิจิทัลที่ 44.1 KHz ซึ่งค่านี้จะใกล้เคียงกับต้นฉบับ ช่องสัญญาณโมโนประเภทละ 50 เพลง 8 ประเภทรวมทั้งหมด 400 เพลง

ตารางที่ 3.1
แสดงจำนวนเพลงตามประเภทดนตรีทั้ง 8 ประเภท

ชนิดดนตรี	ภาษา	ข้อมูลที่ใช้ในการสอนระบบ(Training Set)	ข้อมูลที่ใช้ในการทดสอบระบบ(Testing)
1.ป๊อบ	ฝรั่ง	35	15
	ไทย	35	15
2.ร็อค	ฝรั่ง	35	15
	ไทย	35	15
3.แจ๊ส	ฝรั่ง	35	15
	ไทย	35	15
4.คันทรี่	ฝรั่ง	35	15
	ไทย	35	15
รวม	ฝรั่ง	140	60
	ไทย	140	60

จากตารางแสดงจำนวนเพลงทั้ง 400 เพลง เพลงประเภทป๊อบ 100 เพลง เป็นเพลงฝรั่ง 50 เพลง เพลงไทย 50 เพลง เพลงประเภทร็อค 100 เพลงเป็นเพลงฝรั่ง 50 เพลง เพลงไทย 50 เพลง เพลงประเภทแจ๊ส 100 เพลงเป็นเพลงฝรั่ง 50 เพลง เพลงไทย 50 เพลง เพลงประเภทคันทรี่ 100 เพลง เป็นเพลงฝรั่ง 50 เพลง เพลงไทย 50 เพลง โดยทุกประเภทจะแยก ข้อมูลที่ใช้ในการสอนระบบ (Training Set) 35 เพลง และใช้ในการทดสอบ 15 เพลง โดยคิดเป็นอัตราส่วน 70 % เพื่อใช้ในการสอนระบบ และ ใช้ในการทดสอบ 30 %

3.1.2 การปรับค่าพารามิเตอร์ของวิธี Artificial Neural Network (ANN)

Artificial Neural Network (ANN) มีโครงสร้างที่มีกระบวนการฝึกฝนเป็นแบบ Supervise และใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) สำหรับการฝึกฝนกระบวนการส่งค่าย้อนกลับประกอบด้วย 2 ส่วนย่อยคือ การส่งผ่านไปข้างหน้า (Forward Pass) การส่งผ่านย้อนกลับ (Backward Pass) ANN แบบ Back – Propagation ซึ่งเป็นวิธีที่มีประสิทธิภาพและได้รับความนิยมใน Multilayer Perceptron ใน Weka โดยประกอบด้วย หน่วยย่อย 3 หน่วย คือ หน่วย Input หน่วย Hidden Unit และ หน่วย Output โดยแต่ละหน่วยจะเชื่อมต่อกัน การที่ให้ค่าในการจำแนกประเภทเพลงให้ผลในการจำแนกที่ดีก็คือส่วนของ Output ใน ANN นั้นเองซึ่งจะเกี่ยวข้องกับการปรับค่าพารามิเตอร์ใน ANN ซึ่งมีค่าที่สำคัญ 5 ค่าดังนั้นการหาค่าที่ดีที่สุดเพื่อนำมาใช้ในการทดลองก็จะเป็นการปรับค่าพารามิเตอร์ในหัวข้อนี้ โดยการทดลองแรกเป็นการปรับค่าพารามิเตอร์เป็นค่า Learning Rate ที่ 0.2 0.5 0.75 และ 0.9 ส่วนค่าอื่นๆจะคงที่ไว้โดย Hidden Unit อยู่ที่ 20 โหนด Momentum เท่ากับ 0.5 Threshold เท่ากับ 20 Training Time เท่ากับ 3000 การทดลองต่อมาเป็นการปรับค่าพารามิเตอร์เป็นค่า Momentum ที่ 0.2 0.5 0.75 และ 0.9 ส่วนค่าอื่นๆจะคงที่ไว้โดย Hidden Unit อยู่ที่ 20 โหนด Learning Rate เท่ากับ 0.5 Threshold เท่ากับ 20 Training Time เท่ากับ 3000 ในการทดลองที่เหลือ เช่น Hidden Unit จะปรับค่าตั้งแต่ 10 20 30 และ 40 ส่วนค่าอื่นๆจะคงที่ไว้โดย Hidden Unit อยู่ที่ 20 โหนด Learning Rate เท่ากับ 0.5 Threshold เท่ากับ 20 Training Time ปรับค่าตั้งแต่ 1000 2000 3000 และ 100000 รอบ ส่วนค่าอื่นๆจะคงที่ไว้โดย Hidden Unit อยู่ที่ 20 โหนด Learning Rate เท่ากับ 0.5 Threshold อยู่ที่ 20 – 40 ส่วนค่าอื่นๆจะคงที่ไว้โดย Hidden Unit อยู่ที่ 20 โหนด Learning Rate เท่ากับ 0.5

3.1.3 การวิเคราะห์ค่ากลาง

เมื่อได้ค่าพารามิเตอร์ที่ดีที่สุดแล้ว ในขั้นตอนต่อไปเป็นการหาค่ากลุ่มคุณลักษณะที่ดีที่สุดในแต่ละกลุ่มเพื่อเป็นค่ากลางในการจำแนกต่อไป ค่าคุณลักษณะเชิงคุณภาพเสียงทั้ง 7 ค่า MFCC LPC ZC SC SF SR และ SSB

ตารางที่ 3.2

แสดงกลุ่มค่าคุณลักษณะที่ใช้ค่าทางสถิติหาค่าในการจำแนก

Feature	Mean	Median	Mode	Mean-Median	Mean-Mode	Median-Mode	Mean-median-Mode
MFCC							
SC							
SR							
SF							
ZC							
SSB							

จากตารางที่ 3.2 เป็นการหาค่าคุณลักษณะที่ดีที่สุดที่ 7 กลุ่มและค่าทางสถิติที่ดีที่สุด 7 กลุ่มคือ Mean Median Mode Mean-Median Mean-Mode Median-Mode และ Mean-Median-Mode นำค่าที่ดีที่สุดเป็นค่ากลางเพื่อนำมาใช้ในการจำแนกต่อไป

3.1.4 การเปรียบเทียบค่าคุณลักษณะเชิงคุณภาพเสียง

ในงานวิจัยนี้จะทำการนำค่า MFCC ZC และ LPC เป็นค่าคุณลักษณะที่มีบทบาทในการจำแนกประเภทเพลงที่ดีโดยจะนำสามค่านี้เป็นค่าเริ่มต้นของการทดลองซึ่งแต่ละชุดการทดลองซึ่งจะเพิ่มค่าคุณลักษณะ 1 ค่าคุณลักษณะที่ช่วยให้ค่าในการจำแนกประเภทเพลงเพิ่มขึ้นมาใช้เป็นค่าคุณลักษณะต่อไปจนครบทั้ง 7 ค่า การดึงค่าคุณลักษณะเพลงไทยและเพลงฝรั่งเพื่อจำแนกเป็นประเภทต่าง ๆ นั้น เครื่องมือที่นำมาใช้ในการดึงค่าคุณลักษณะในครั้งนี้ใช้ JAudio (McEnnis et al. 2005) ซึ่งเป็นส่วนหนึ่งของชุดโปรแกรม JMIR ที่พัฒนาขึ้นเพื่องานวิจัยด้านค้นคืนดนตรี (Music Information Retrieval) ค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ที่ใช้ประกอบไปด้วย MFCC LPC ZC SC SF SR และ SSB ซึ่งจะขออธิบายรายละเอียดของค่าคุณลักษณะทั้ง 7 ค่าดังนี้

3.4.1.1 Mel-frequency cepstral coefficients (MFCC) เป็นค่าที่ปรับปรุงโดยปรับหรือกรองส่วนที่มนุษย์ไม่ได้ยินออก ซึ่งจะได้ความถี่ที่เหมาะสมสำหรับการรับฟังของมนุษย์ ซึ่งสามารถเก็บรายละเอียดของสัญญาณความถี่ต่ำได้ดีโดยค่า MFCC ที่ได้จากโปรแกรมมีทั้งหมด 12 ค่าคือ MFCC 0 – MFCC12 แต่เนื่องจากผลการทดลองเบื้องต้นการจำแนกดนตรีคันทรี่ไทยและฝรั่งโดยใช้แบบโครงข่ายประสาทเทียม อธิบายค่าของ MFCC 0 – MFCC 4 ก็เพียงพอสำหรับการจำแนกประเภทดนตรีคันทรี่ไทยและฝรั่งได้ (ลักษณะันท์ และ ณัฐธนนท์, 2010) ในวิทยานิพนธ์ชุดนี้จึงนำค่า MFCC 0 – MFCC 4 นำมาใช้ในการทดลองครั้งนี้ โดยจากโปรแกรม JAudio อ้างอิงการใช้งานข้อมูล MFCC จาก (Pfeiffer, Parker, Vincent, and Maate, 2005) [Accessed April 14, 2005] URL <http://www.cmis.csiro.au/maate/>

3.4.1.2 Linear Predictive Coefficients (LPC) เป็นค่าคุณลักษณะที่ได้จากการสุ่มตัวอย่างของสัญญาณเสียง ได้จากผลรวมเชิงเส้น (Linear Combination) ของการสุ่มจากสัญญาณเสียงหลายๆส่วนจากตัวอย่างที่ผ่านมาจากการเปรียบเทียบกับค่าสัญญาณจริง โดยค่าจริงที่ได้จากโปรแกรม JAudio มีด้วยกันทั้งหมด 13 ค่าคือ LPC

0 -12 นี้ โดยจากโปรแกรม JAudio (McEnnis et al. 2005) [Accessed May 17, 2005].
www.musicdsp.org/archive.php?classid=2#137

3.4.1.3 Zero Crossing Rate (ZC) เป็นค่าคุณลักษณะที่ได้จากการนับจำนวนครั้งของสัญญาณที่ผ่านแกนศูนย์

3.4.1.4 Spectral Centroid (SC) เป็นค่าคุณลักษณะที่แสดงถึงค่ากึ่งกลางของ Power Spectrum โดยคำนวณจากค่าเฉลี่ยของค่า Power Spectrum

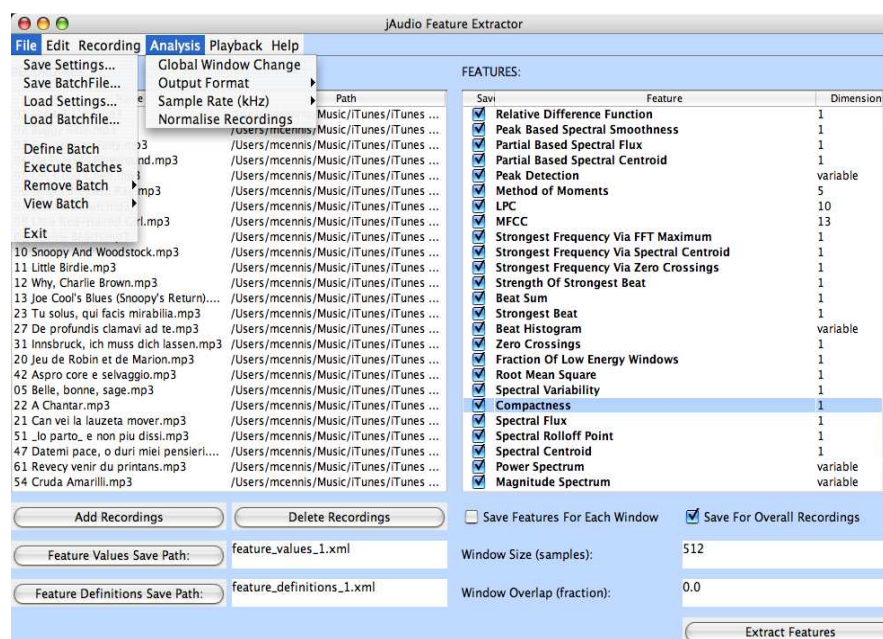
3.4.1.5 Spectral Flux (SF) เป็นค่าที่คำนวณความแตกต่างระหว่างค่าปัจจุบันของแต่ละพลังงานเสียงกับค่าพลังงานก่อนหน้าโดยแต่ละค่าที่แตกต่างจะถอดด้วยรากที่สองของผลรวมจากการยกกำลังสอง

3.4.1.6 Strength of Strongest Beat (SSB) เป็นค่าที่ได้ค่า Beat Histogram แสดงกำลังคาบจังหวะที่แตกต่างของสัญญาณโดยคำนวณจากการใช้ RMS ของ 256 ค่า จากนั้นใช้ Fast Fourier Transform ซึ่งได้เป็นค่า Beat Histogram ส่วน Strength of Strongest Beat นำค่าสัญญาณจาก Beat Histogram โดยใช้ค่าที่สอดคล้องกับ Strongest Beat ค่าของผลรวมที่ถูกแบ่งจากสัญญาณ Beat Histogram ที่ผ่านเข้ามา

3.4.1.7 Spectral Roll Off (SR) เป็นค่าของความถี่ภายใต้ 85 % ของการกระจายค่าความถี่ใช้วัดค่ารูปร่างที่เปลี่ยนแปลง

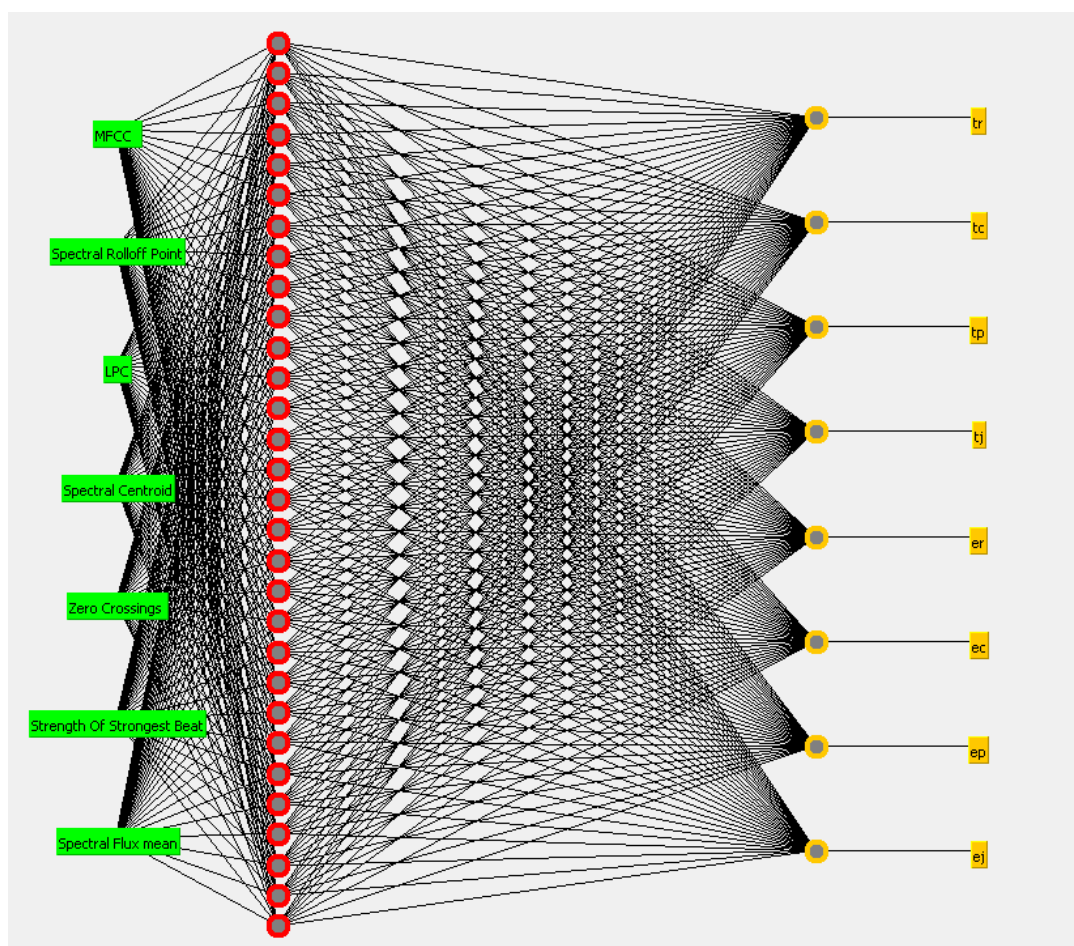
ภาพที่ 3.2

แสดงการใช้งานโปรแกรม JAudio



จากรูปโปรแกรม JAudio ซึ่งใช้ในการดึงค่าคุณลักษณะจากเพลงที่ใช้ระยะเวลา 4 ช่วงเวลา คือวินาทีที่ 0 -10 วินาทีที่ 0-20 วินาทีที่ 0-30 และสุดท้ายตัดแบ่งที่กึ่งกลางของเพลงมีความยาวประมาณ 10 วินาที จากเพลงทั้ง 4 ประเภท ร็อค ป๊อป คันทรี และแจ๊สที่ใช้ภาษาไทยและภาษาฝรั่งโดยเลือกขนาดของการสุ่มที่ 44.1 โดยใช้รูปแบบ .arff เพื่อเตรียมส่งให้โปรแกรม weka ทำการจำแนกประเภทตามค่าคุณลักษณะที่ได้ การดึงค่าคุณลักษณะจากเพลงช่วงระยะเวลาต่างๆ เช่น ช่วงระยะ 0 – 10 วินาทีจะได้ขนาดตัวเลข 300 กว่าบรรทัดจากการดึงค่าคุณลักษณะ ช่วงระยะ 0 – 20 วินาทีจะได้ขนาดตัวเลข 600 กว่าบรรทัด ยิ่งช่วงระยะเวลา 0 – 30 วินาที จะให้ค่าตัวเลขถึง 900 บรรทัด การนำค่าทั้งหมดมาคำนวณจะใช้เวลาและทรัพยากรในการประมวลผลอย่างสิ้นเปลืองโดยในหัวข้อต่อไปจะเป็นการนำค่าทางสถิติมาใช้เพื่อความสะดวกและประหยัดเวลาในการประมวลผล

ภาพที่ 3.3
แสดงค่าคุณลักษณะทั้งหมดโดยใช้ Neural Network



จากรูป 3.4 แสดงตัวอย่างการทดลองจำแนกประเภทเพลงทั้ง 8 ประเภทโดยในภาพแสดงค่าคุณลักษณะ 7 ค่าโดย input เป็นค่าคุณลักษณะ 7 ค่า ชั้น Hidden Unit ใช้ 30 unit Output เป็น 8 ประเภท โดย er แทนเพลงร็อคฝรั่ง ep แทนเพลงป๊อปฝรั่ง ej แทนเพลงแจ๊สฝรั่ง ec แทนเพลงคันทรี่ฝรั่ง tr แทนเพลงร็อคไทย tc แทนเพลงคันทรี่ไทย tp แทนเพลงป๊อปไทย และ tj แทนเพลงแจ๊สไทย

3.1.5 การเปรียบเทียบค่าคุณลักษณะในช่วงเวลาที่ 0 - 10 0 - 20 และ 0 - 30 วินาที

ในส่วนนี้เป็นการทดลองเหมือนกับการทดลองที่ 3.1.4 คือ นำค่า MFCC ZC และ LPC เป็นค่าคุณลักษณะที่มีบทบาทในการจำแนกประเภทเพลงที่ดีโดยจะนำสามค่านี้เป็นค่าเริ่มต้นของการทดลองซึ่งแต่ละชุดการทดลองซึ่งจะเพิ่มค่าคุณลักษณะ 1 ค่าคุณลักษณะที่ช่วยให้ค่าในการจำแนกประเภทเพลงเพิ่มขึ้นมาใช้เป็นค่าคุณลักษณะต่อไปจนครบทั้ง 7 ค่า เพียงแต่จะใช้ช่วงเวลาที่แตกต่างกันโดยจะเป็นการทดลองในช่วงเวลาที่แตกต่าง 3 ช่วงคือ ในช่วงระยะเวลาที่ 0 – 10 วินาที ส่วนที่สองจะเป็นช่วงเวลา 0 – 20 วินาที และ ในช่วงระยะเวลาที่ 0 – 30 วินาที

3.1.6 การจำแนกโดยใช้ขนาดเฟรม

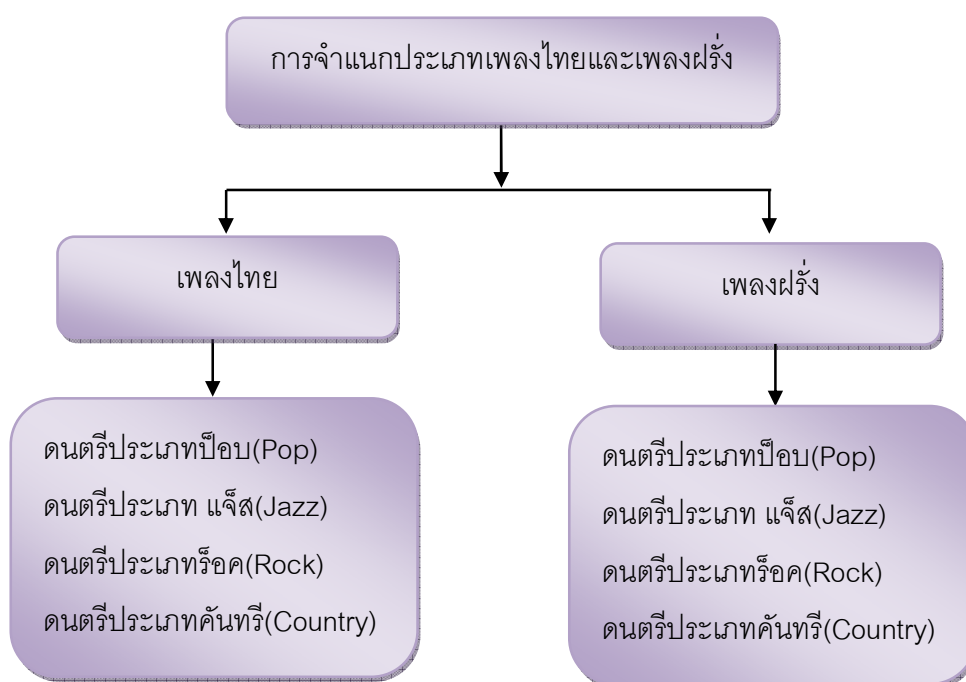
ส่วนการแบ่งขนาดของเฟรมที่แบ่งขนาดแตกต่างกัน มีผลต่อค่าความถูกต้องต่อการจำแนกหรือไม่ โดยทำการแบ่งขนาดเฟรมของเพลงเป็น 128 256 512 1024 และ 2048 เพื่อนำหาค่าคุณลักษณะโอบาส แก้วต่าย (2552) ทำการแบ่งเฟรมขนาด 32 64 128 256 512 1024 และ 2048 แต่จากการทดลองขนาดของเฟรมที่ให้ค่าความถูกต้องตั้งแต่ 80 % ขึ้นไปก็คือขนาดที่ 128 256 512 1024 และ 2048 ดังนั้นจึงขอเลือกขนาดเฟรมดังกล่าวเพื่อนำมาใช้ในการทดลอง

3.1.7 การจำแนกประเภทเพลงที่ดีที่สุดในค่าคุณลักษณะ MFCC 0 – 4

การทดลองในขั้นตอนนี้เป็นการหาค่าที่ดีที่สุด ใน MFCC 0 – 4 แล้วนำค่าที่ดีที่สุด 1 ค่าใน MFCC เปรียบเทียบกับค่าคุณลักษณะอื่นๆ ที่มีค่าเดียวเพื่อทำการเปรียบเทียบความสำคัญของค่าคุณลักษณะที่มีค่าเดียวทั้งหมด เนื่องจากการทดลองในขั้นตอนนี้พบว่ากลุ่มค่าคุณลักษณะเชิงคุณภาพที่มีค่าเป็นกลุ่มคือ MFCC และ LPC มีค่าในความสำคัญในการจำแนกที่ดีมากกว่ากลุ่มที่มีค่าเดียว ดังนั้นในการทดลองในส่วนนี้นำค่า MFCC ที่มีค่าเดียวไปเปรียบเทียบกับค่าคุณลักษณะที่มีค่าเดียว

3.2 การจำแนกประเภทดนตรีแบบ 2 ขั้นตอน

ภาพที่ 3.4
แสดงขั้นตอนการจำแนกประเภทเพลงแบบ 2 ขั้นตอน



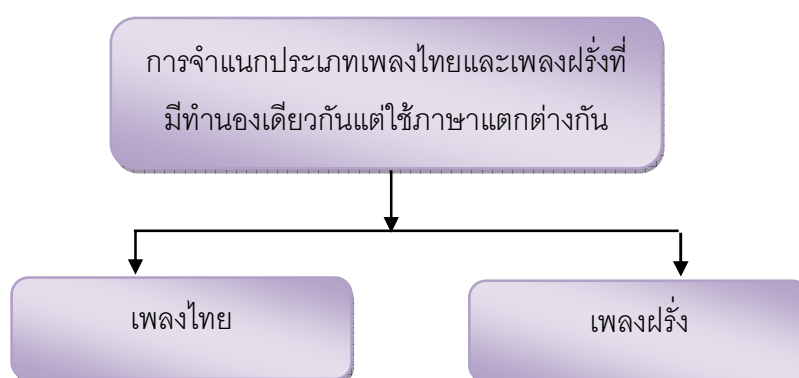
ในขั้นตอนนี้จะเป็นการทดลองซึ่งมีลักษณะคล้ายกับขั้นตอนแรกโดยเลือกค่ากลางทางสถิติที่ดีที่สุด 7 กลุ่มเป็นค่ากลางจากการดึงค่าคุณลักษณะและแบ่งกลุ่มทดลองค่าคุณลักษณะออกเป็น 7 กลุ่มเลือกค่าที่ดีที่สุดในแต่ละกลุ่มเพียงแต่ในขั้นตอนนี้จะแบ่งขั้นตอนในการจำแนกออกเป็น 2 ขั้นตอน โดยขั้นตอนแรกจะเป็นการจำแนกระหว่างเพลงไทยและเพลงฝรั่ง ขั้นตอนต่อมาเป็นการจำแนกประเภทเพลงในแต่ละภาษา

3.3 การจำแนกเพลงที่มีทำนองเดียวกันแต่ใช้ภาษาแตกต่างกัน

ในขั้นตอนนี้ต้องการศึกษาถึงค่าคุณลักษณะเชิงคุณภาพเสียงมีผลต่อการจำแนกเพลงที่ใช้ทำนองเดียวกันแต่ใช้ภาษาที่แตกต่างกันได้หรือไม่โดยการทดลองนี้ใช้ค่าคุณลักษณะแยกประเภทเพลงที่ใช้ภาษาไทยและฝรั่งเท่านั้นไม่ทำการแยกประเภทเพลง บ๊อบ ร็อค คันทรี และ แจ๊ส

ภาพที่ 3.5

แสดงขั้นตอนการทดสอบเพลงที่มีทำนองเดียวกันแต่ภาษาแตกต่างกัน



จากภาพการจำแนกประเภทเพลงไทยและเพลงฝรั่งที่มีทำนองเดียวกันแต่ใช้ภาษาแตกต่างกันจะทำการจำแนกเฉพาะเพลงที่ใช้ภาษาไทยและฝรั่งเท่านั้นซึ่งเพลงที่ได้ทั้งหมด 26 เพลง เป็นเพลงไทย 13 เพลง และเป็นเพลงฝรั่ง 13 เพลง

3.4 สรุปผลการทดลองโดยรวมทั้งหมด

เมื่อทำการทดลองตามหัวข้อก็จะนำค่ามาสรุปและเปรียบเทียบค่าเพื่อนำมาใช้ในการวิเคราะห์ต่อไป