บทที่ 5

สรุปและอภิปรายผลงานวิจัย

บทที่ 5 เป็นบทสุดท้ายของวิทยานิพนธ์ฉบับนี้ โดยจะสรุปผลการวิจัยซึ่งจะนำไปสู่การอภิปรายในส่วนถัดไปและส่วนสุดท้ายด้วยข้อเสนอแนะงานวิจัยในอนาคต

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้เป็นการศึกษาและพัฒนาให้คอมพิวเตอร์สามารถจำแนกประเภทเพลงที่ขับร้องด้วยภาษาไทยและเพลงฝรั่งโดยประเภทของเพลงที่ใช้ในการจำแนกนั้นจะใช้ ร็อค ป็อบ แจ็ส และ คันทรีของแต่ละภาษาเพื่อเปรียบเทียบความถูกต้องในการจำแนกโดยใช้ตัวจำแนกโครงข่ายประสาทเทียมแบบหลายชั้น จากการใช้ค่าคุณลักษณะ 7 ค่า MFCC, LPC, SC, SF, ZC, SR และ SSB จากเพลงในช่วงเวลา 0-10 วินาที 0-20 วินาที และ 0-30 วินาทีแรกของเพลงและจาก 10 วินาทีตรงกึ่งกลางของเพลง และคำนวณวิเคราะห์หาค่ากลางของคุณลักษณะเพื่อที่จะนำมาใช้ในการจำแนกประเภทเพลงได้แก่ค่ากลาง Mean Median Mode และค่ากลาง 3 ค่าร่วมกัน Mean - Median Mean - Mode Median - Mode และ Mean - Median - Mode การศึกษาในวิทยานิพนธ์ฉบับนี้แบ่งออกเป็น 4 ส่วน ส่วนที่ 1 เป็นการจำแนกประเภทเพลงโดยแบ่งเป็น 8 ประเภท 1 ขั้นตอน ไทยร็อค ไทยป็อบ ไทยคันทรี ไทยแจ็ส ฝรั่งร็อค ฝรั่งป็อบ ฝรั่งคันทรี และฝรั่งแจ็ส ส่วนแรกขั้นตอนเตรียมการทดลองโดยการจำแนกประเภทเพลงโดยการทดลองแบบทวนสอบ 10-Fold Cross Validation ส่วนที่สองเป็นการปรับค่าพารามิเตอร์ที่ให้การจำแนกดีที่สุด ส่วนที่สามเป็นการวิเคราะห์การหาค่ากลาง ส่วนที่สี่เป็นการเปรียบเทียบค่าคุณลักษณะเชิงคุณภาพเสียงที่กึ่งกลางเพลง ส่วนที่ห้าเป็นการเปรียบเทียบค่าคุณลักษณะที่ช่วงเวลาที่ 0 – 10 0 – 20 และ 0 - 30 วินาทีและในส่วนที่หกเป็นการจำแนกโดยใช้ขนาดเฟรม ในส่วนสุดท้ายส่วนที่เจ็ดเป็นการจำแนกประเภทเพลงที่ดีที่สุดในค่าคุณลักษณะ MFCC 0 – 4

ในขั้นตอนเป็นการปรับค่าพารามิเตอร์ (Parameter Tuning) ของโครงข่ายประสาทเทียมแบบหลายชั้นโดยจะทำการปรับค่าพารามิเตอร์เพื่อที่จะหาค่าที่ดีที่สุด ในการทดลองนี้ค่าพารามิเตอร์ที่ดีที่สุดคือ Learning Rate เท่ากับ 0.75 Momentum เท่ากับ 0.25 Hidden Unit เท่ากับ 30 Training Rate เท่ากับ 3000 และ Threshold เท่ากับ 40 ผลการวิเคราะห์ค่ากลางพบว่าจากการทดลองใช้ค่าทางสถิติ 7 กลุ่มเพื่อเป็นค่ากลางพบว่าค่าที่นำมาใช้ในการจำแนกแล้วได้ค่าในการจำแนกดีที่สุดคือ ค่า Mean - Median ผลการเปรียบเทียบกลุ่มค่าคุณลักษณะที่ดีที่สุดในช่วงกึ่งกลางเพลงที่ 10 วินาที ZC+SR+MFCC+LPC+SC+SF ที่ 74.45% ค่ากลุ่มคุณลักษณะการจำแนก 10 วินาทีแรกของเพลง MFCC+SR+LPC ที่ 67.32% ค่ากลุ่มคุณลักษณะการจำแนก 20 วินาทีแรกของเพลง ZC+MFCC+SSB+LPC+SC+SF ที่ 58.72% ค่ากลุ่มคุณลักษณะการจำแนก 30 วินาทีแรกของเพลง ZC+LPC+MFCC+SF+SC+SSB ที่ 58.72% ผลการจำแนกโดยใช้ขนาดเฟรมจากการทดลองครั้งนี้เลือกใช้ขนาดเฟรม 5 ค่า 128 256 512 1024 และ 2048 พบว่าค่าที่ขนาดเฟรมที่ใช้ในการจำแนกได้ค่าดีที่สุด 512 หลังจากนั้นหาค่าที่ดีที่สุดในกลุ่มของ MFCC 0 - 4 พบว่าค่า MFCC 1 ให้ค่าดีที่สุด หลังจากนั้นนำค่า MFCC 1 ไปเปรียบเทียบค่าคุณลักษณะที่มีค่าเดียวพบว่า MFCC 1 เป็นรองค่า SR ค่าเดียว ในส่วนที่ 2 ส่วนการจำแนกเพลงแบบ 8 ประเภท 2 ขั้นตอนพบว่าค่าในขั้นตอนแรกการจำแนกเพลงที่ใช้ภาษาไทยและฝรั่งตั้งแต่กลุ่มที่มี 2 ค่าคุณลักษณะขึ้นไปสามารถจำแนกประเภทเพลงได้ดีที่ 99.75% ขั้นตอนที่สองการจำแนกประเภทเพลงที่ใช้ภาษาไทยกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC+SC+SF ให้ค่าความถูกต้องที่ 72.47% ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาฝรั่งกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC ให้ค่าความถูกต้องที่ 73.43% ในส่วนสุดท้ายเพลงที่ใช้ทำนองเดียวกันแต่ใช้ภาษาแตกต่างกันพบว่าค่า ZC ให้ค่าในการจำแนกสูงสุดที่ 61.54%

5.2 อภิปรายผลการวิจัย

ส่วนที่แรกเป็นการอภิปรายผลการปรับค่าพารามิเตอร์ที่ให้การจำแนกดีที่สุด ส่วนที่สองเป็นการการอภิปรายวิเคราะห์การหาค่ากลาง ส่วนที่สามเป็นการการอภิปรายการเปรียบเทียบค่าคุณลักษณะเชิงคุณภาพเสียงที่กึ่งกลางเพลง ส่วนที่สี่เป็นการการอภิปรายผลการเปรียบเทียบค่าคุณลักษณะที่ช่วงเวลาที่ 0 – 10 0 – 20 และ 0 - 30 วินาทีและในส่วนที่ห้าเป็นการการอภิปรายผล

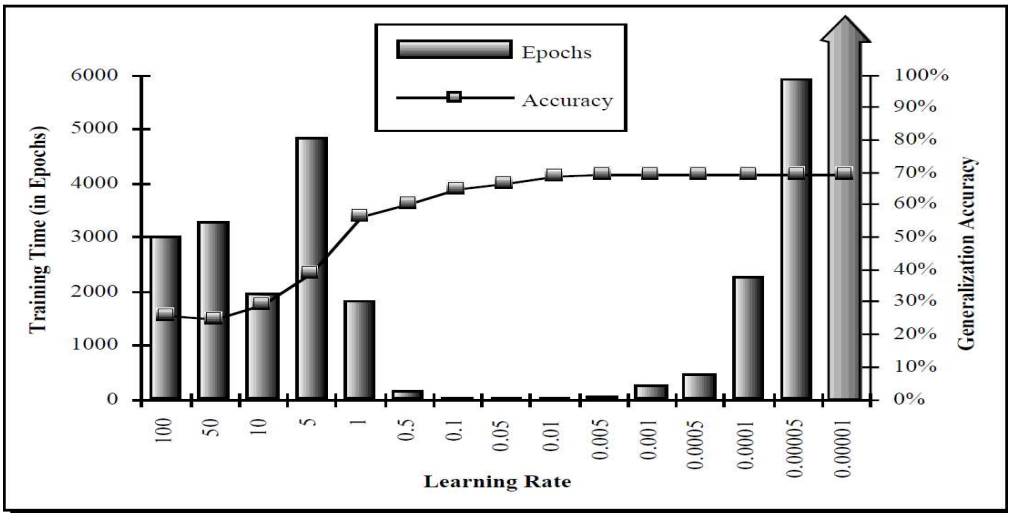
จำแนกโดยใช้ขนาดเฟรมและในส่วนสุดท้ายเป็นการการอภิปรายผลการจำแนกประเภทเพลงที่ดีที่สุดในค่าคุณลักษณะ MFCC 0 – 4

5.2.1 การอภิปรายสรุปผลจากการปรับค่าพารามิเตอร์

ในส่วนนี้เป็นการปรับค่าพารามิเตอร์ของโครงข่ายประสาทเทียมแบบหลายชั้นเพื่อให้ผลในการจำแนกดีที่สุดซึ่งมีพารามิเตอร์ที่สำคัญอยู่ 5 ค่าคือ Learning Rate, Momentum,Hidden Unit Training Rate และ Threshold จากผลการทดลองพบว่าโครงข่ายประสาทเทียมแบบหลายชั้นใช้ค่า Learning Rate เท่ากับ 0.75 Momentum เท่ากับ 0.25 Hidden Unit เท่ากับ 30 Training Rate เท่ากับ 3000 และ Threshold เท่ากับ 40 ให้ผลการทดลองดีที่สุดแต่ใช้เวลาประมวลค่อนข้างเยอะเมื่อเทียบกับงานวิจัยอื่นๆจะใช้ Learning Rate ค่อนข้างต่ำ รมิดา ธนกุลภารัชต์ (2552) ใช้ Learning Rate เท่ากับ 0.3 และใช้ Training Rate เท่ากับ 500 โดยใช้โครงข่ายประสาทเทียมแบบหลายชั้น ในการจำแนกประเภทดนตรีไทยเดิมตามอารมณ์หลัก 3 ประเภท คือ สนุกสนาน อารมณ์ผ่อนคลาย และอารมณ์เศร้า ส่วนในงานวิจัยของต่างประเทศ Wilson and Martinez (2001) อธิบายถึงการทดลองเรื่อง The Need for Small Learning Rate on Large Problems แยกประเภทเสียงจากการอ่านออกเสียงคำศัพท์ซึ่งจัดหมวดหมู่ไว้ 23 กลุ่มทดสอบ ผลของการใช้ Learning Rate และ Trainining ที่มีผลต่อเปอร์เซ็นความถูกต้องต่อการจำแนกเพลงดังภาพ 5.1

ภาพที่ 5.1

แสดงอัตราส่วนการเรียนรู้และค่าความถูกต้องในการจำแนก



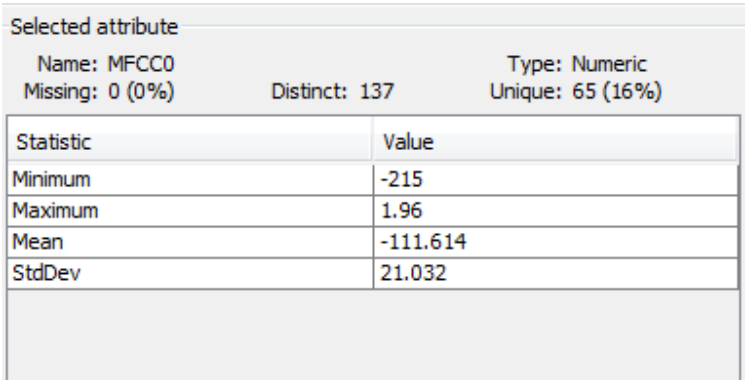
จากภาพแสดงให้เห็นว่าอัตราระหว่างการจำแนกเปอร์เซ็นค่าความถูกต้องกับค่า Leaning Rate แสดงให้เห็นว่าค่า Learning Rate เริ่มมีค่าลดลงเมื่อค่าเริ่มเข้าใกล้ 1 เปอร์เซ็นความถูกต้องจะใกล้กับ 60 เปอร์เซ็นต์ และลดลงต่ำลงไปถึงประมาณ 40 % เมื่อค่า Learning Rate เท่ากับ 5 ซึ่งในงานวิจัยชุดนี้ใช้ค่า Learning Rate เท่ากับ 0.75 ค่าความถูกต้องอยู่ที่ประมาณใกล้ ๆ 70% Wilson and Martinez ได้อธิบายถึงหลักการเลือกค่า Learning Rate สูงจะทำให้ในการจำแนกลดลงและความเร็วในการทำงานของเครื่องลดลง ขณะที่ลดค่า Learning Rate ต่ำลงจะพบว่าค่าในการจำแนกและประสิทธิภาพในการทำงานจะดีขึ้น

5.2.2 การอภิปรายผลการเปรียบเทียบค่ากลาง

การหาค่าตัวแทนเพื่อเป็นค่ากลางนำไปใช้ในการจำแนกจากค่าทางสถิติ 7 กลุ่ม คือ Mean, Median, Mode, Mean – Median, Mean – Mode, Median – Mode และ Mean – Median – Mode พบว่าค่า Mean-Median ให้ในการจำแนกดีที่สุด ซึ่งค่ากลางทั้ง 3 ค่านี้เป็นค่าเพียงค่าเดียวเพื่อ

ใช้อธิบายข้อมูลทั้งชุดที่เป็นตัวแทนของข้อมูลทั้งชุด จากการใช้โปรแกรม Weka วิเคราะห์ค่าของกลุ่มทางสถิติของค่า MFCC ที่ใช้ค่ากลาง Mean – Median

ภาพที่ 5.2
แสดงค่าทางสถิติของค่าคุณลักษณะที่ได้จากการดึงค่าคุณลักษณะ


Selected attribute

Name: MFCC0 Type: Numeric
Missing: 0 (0%) Distinct: 137 Unique: 65 (16%)

Statistic	Value
Minimum	-215
Maximum	1.96
Mean	-111.614
StdDev	21.032

จากภาพอธิบายการจำแนกดีที่สุดใน 7 กลุ่มพบว่าช่วงตัวเลขจากการดึงค่าคุณลักษณะค่าน้อยที่สุดที่ -215 และ ค่ามากที่สุดที่ 1.96 ค่า Standard Deviation ที่ 21.032 และค่า Mean ที่ -111.614 จากค่า Standard Deviation จะทำให้ทราบถึงการกระจายของข้อมูลที่ทำให้ทราบถึงการกระจายข้อมูลที่ค่าแตกต่างจากค่าของค่าเฉลี่ยมากน้อยเพียงใดซึ่งพบว่าค่า Standard Deviation อยู่ที่ 21.032 ซึ่งถือว่ามีค่าไม่มากเท่าไรนักดังนั้นเราจะถือว่าการกระจายข้อมูลไม่แตกต่างกับค่าเฉลี่ยเท่าไรและการกระจายข้อมูลยังเป็นลักษณะค่าปกติซึ่งเหมาะสมกับค่า Mean และ Median ให้ผลในการจำแนกค่อนข้างดีซึ่งเหมาะกับการกระจายข้อมูลที่ปกติเมื่อนำค่า Mean และ Median มาใช้ร่วมกันทำให้สามารถจำแนกเพลงได้ดีขึ้น

5.2.3 การอภิปรายผลการเปรียบเทียบค่าคุณลักษณะเชิงคุณภาพเสียงที่กึ่งกลางเพลง

จากการทดลองส่วนแรกการแยกประเภทเพลง 8 ประเภทใช้ค่าคุณลักษณะทั้ง 7 ค่าในการทดลองจากผลการวิจัยจะเห็นได้ว่าอัตราความถูกต้องที่ดีที่สุดให้ค่าความถูกต้องอยู่ที่ 74.45 % ซึ่งให้ผลเป็นที่น่าพอใจในระดับหนึ่ง ค่าคุณลักษณะที่มีผลต่อการจำแนกประเภทของเพลงที่สำคัญคือ

ZC+SR+MFCC+LPC+SC+SF ซึ่งเป็นค่าคุณลักษณะที่สำคัญใน Timbral Feature ซึ่งมีบางส่วนที่สอดคล้องกับงานวิจัยของ Xu,Maddage,and Shao (2005) อธิบายถึงค่าคุณลักษณะที่สำคัญ ZC MFCC และLPC ซึ่งมีความสำคัญในการจำแนกประเภทเพลงอย่างมาก Xu,Wang ,and Yan (2008) ได้ทำการจำแนกประเภทเพลงพื้นบ้านจีน 10 ประเภทใช้ค่าคุณลักษณะ ZC MFCC และLPC ดึงค่าคุณลักษณะจะเห็นได้ว่าค่าคุณทั้ง 3 ค่ามีอยู่ในกลุ่มค่าที่ดีที่สุดของการทดลองนี้ด้วย

5.2.4 การอภิปรายผลการเปรียบเทียบค่าคุณลักษณะที่ช่วงเวลาที่ 0 - 10 0 - 20 และ 0 - 30 วินาที

ในงานวิจัยนี้ใช้ช่วงระยะเวลาที่ต่างกัน 4 แบบ ความยาวประมาณ 10 วินาที 20 วินาที 30 วินาทีและสุดท้ายตัดเพลงจากกึ่งกลางของเพลงความยาวประมาณ 10 วินาที ช่วงระยะเวลาของเพลงที่ใช้ในการจำแนกจากการทบทวนวรรรกรรมที่ผ่านมารายงานถึง ระยะเพลงที่ใช้ 10 วินาที สามารถจำแนกประเภทเพลงให้ค่าความถูกต้องสูงสุดโดยใช้วินาทีที่ 0 –10 ส่วนงานวิจัยชุดนี้ใช้ระยะเพลงที่10 วินาทีที่กึ่งกลางเพลงให้ผลดีที่สุดเนื่องรายละเอียดของเพลงส่วนใหญ่จะอยู่ที่กึ่งกลางเพลง ซึ่งส่วนใหญ่ช่วงระยะเวลานี้จะอยู่ในช่วงท่อนสร้อยของเพลง ส่วนช่วงเพลงที่ 0 – 10 ให้ผลดีรองลงมาเนื่องจากท่อนสร้อยของเพลงจะอยู่ช่วงเริ่มต้นเพลงด้วยเหมือนกัน ส่วนวินาทีที่ 0-20 จะเป็นส่วนใหญ่จะเป็นช่วงเนื้อร้องซึ่งรายละเอียดของเพลงจะน้อยลงและส่วนของเนื้อเพลงก็จะเริ่มขึ้น ส่วนวินาทีที่ 0-30 จะคล้ายกับส่วนของวินาทีที่ 0-20 แต่ส่วนนี้จะเป็นส่วนที่ใกล้กับท่อนแยกซึ่งเป็นช่วงเริ่มจะมีรายละเอียดของดนตรีบางหรือน้อยลงเพื่อรอเข้าท่อนแยกของดนตรีบ้างช่วงในผลการทดลองมีผลการจำแนกที่ดีกว่าช่วง 0-20

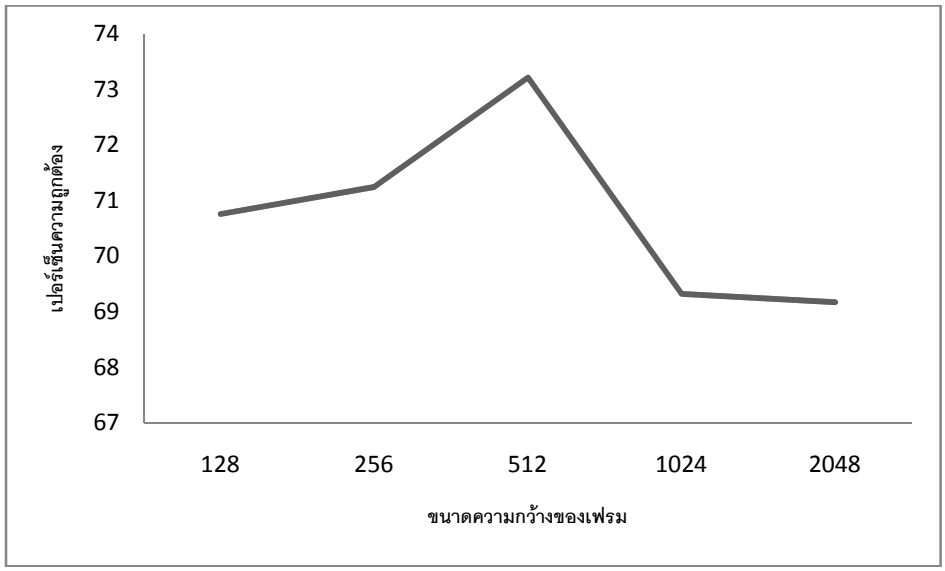
5.2.5 การอภิปรายผลการเปรียบเทียบขนาดของเฟรมที่ใช้ในการจำแนก

จากการใช้ค่าคุณลักษณะ ZC+SR+MFCC+LPC+SC+SF พบว่าเมื่อทำการจำแนกด้วยขนาดเฟรมเท่ากับ 512 sample จะให้ค่าความถูกต้องในการจำแนกที่ดีที่สุดด้วยวิธีโครงข่ายประสาทเทียมแบบหลายชั้น ค่าความถูกต้องจะลดลงเมื่อกำหนดขนาดเฟรมมากกว่า 512 sample เนื่องจากขนาดของเฟรมมีผลต่อการดึงค่าคุณลักษณะที่ได้ดังนั้น เมื่อทำการแบ่งเฟรมตามขนาดต่าง ๆ แล้ว

เฟรมที่มีขนาดใหญ่หมายถึงจำนวนเฟรมมีน้อยลงต่อเพลง ทำให้ค่าคุณลักษณะที่ได้ไม่สามารถบ่งบอกเอกลักษณ์ของเพลงได้ จึงให้ค่าความถูกต้องในการจำแนกที่ไม่ดีเท่าที่ควร

ภาพที่ 5.3

แสดงค่าความถูกต้องในการจำแนกประเภทเพลงกับขนาดความกว้างของเฟรม



จากภาพผลแสดงการทดลองในการจำแนกประเภทเพลงพบว่าขนาดเฟรมที่มีความกว้าง 512 Sample จะให้ผลในการจำแนกดีที่สุด หลังจากนั้นจะให้ผลการจำแนกลดลงที่ 69.32% และ 69.17%จากขนาดเฟรมที่เพิ่มขึ้นที่ 1024 และ 2048

ตารางที่ 5.1
แสดงจำนวนเฟรมเมื่อแบ่งตามความกว้างของเฟรมขนาดต่างๆที่ช่วงความยาวของเพลงที่แตกต่างกันต่อหนึ่งเพลง

ช่วงความยางเพลง	จำนวนของเพลงตามขนาดความกว้างต่าง ๆ ต่อหนึ่งเพลง						
	32Sample	64Sample	128Sample	256Sample	512Sample	1024Sample	2024Sample
0 -10 วินาที	13,782	6,891	3,446	1,723	862	431	216
0 -20 วินาที	27,563	13,782	6,891	3,446	1,723	862	431
0 -30 วินาที	41,334	20,672	10,336	5,168	2,584	1,292	646

ที่มา: “การจำแนกประเภทกลุ่มเพลงไทยเดิม”โดย โอภาส แก้วต่าย, 2552, เทคโนโลยีสารสนเทศ, มหาวิทยาลัยศิลปากร

จากตารางการทดลองของ โอภาส แก้วต่าย (2552) พบว่าเมื่อใช้ขนาดเฟรมที่ 512 Sample ที่ 0 -10 วินาทีจะใช้จำนวนเฟรมทั้งหมด 862 ต่อเพลงและเมื่อเพิ่มขนาดเป็น 1024 Sample และ 2024 Sample ใช้จำนวนเฟรมทั้งหมด 431ต่อเพลง และ 216ต่อเพลงตามลำดับ ซึ่งแสดงให้เห็นว่าการเพิ่มขนาดเฟรมทำให้จำนวนเฟรมต่อความยาวเพลงจะลดลงเรื่อยๆ ทำให้การจำแนกประเภทเพลงลดลงซึ่งจะสอดคล้องกับผลการทดลองในงานวิจัยครั้งนี้

5.2.6 การอภิปรายผลการจำแนกประเภทเพลงที่ดีที่สุดในค่าคุณลักษณะ MFCC 0 – 4

ค่าจากโปรแกรม Jaudio มีค่า MFCC ทั้งหมด 13 ค่าผลจากการทดลองของ ลักษนันท์ แกล้วทนงค์ และ ณัฐธนนท์ หงส์วริทธิ์ธร,(2553) อธิบายถึงใช้ค่า MFCC 0 -4 ทั้งหมด 5 ค่าก็เพียงพอต่อการจำแนกประเภทเพลง ดังนั้นในการทดลองนี้ได้นำค่า MFCC ค่าเดียวที่ดีที่สุดมาทำการทดลองโดยการจำแนกประเภทเพลง 8 ประเภท 1 ในขั้นตอนซึ่งได้ผลการทดลอง

จากการเปรียบเทียบค่าคุณลักษณะ พบว่าเมื่อใช้ค่า MFCC1 ค่าเดียวค่าความถูกต้องที่ได้ 35.38% จากตารางที่ 4.27 ซึ่งลดลงไปเกือบ 20% เปรียบเทียบจากค่าเดิมที่ 50.86 % จากตารางที่ 4.3 เมื่อเปรียบเทียบกับค่าคุณลักษณะที่มีค่าเดียวจะมีค่ารองจากค่า SR ทำให้สามารถสรุปได้ว่าค่าคุณลักษณะแบบกลุ่มสามารถจำแนกได้ดีกว่าค่าที่มีลักษณะในการจำแนกแบบค่าเดียวซึ่งจากการการทดลองเมื่อนำค่า MFCC 1 ซึ่งเป็นค่าที่ดีที่สุดจากค่าทั้งหมด 5 ค่า นำมาเปรียบเทียบค่ากลุ่มคุณลักษณะค่าเดียวพบว่า ค่า MFCC 1 เป็นรองค่า SR ซึ่งมีเปอร์เซ็นต์ความถูกต้องที่ 37.10 % จึงสามารถสรุปได้ว่า เมื่อนำค่า MFCC 1 เพียงค่าเดียวเปรียบเทียบกับกลุ่มค่าคุณลักษณะที่มีค่าเดียวก็ยังมีเปอร์เซ็นต์การจำแนกประเภทเพลงที่ค่อนข้างดี ซึ่งจะสอดคล้องกับงานของ Ahrendt , Meng and Larsen (2008) อธิบายถึงการจำแนกประเภทเพลงโดยใช้กรอบเวลาในการตัดสินใจข้อมูลในการจำแนกปนะเภทดนตรีร่วมกับค่าคุณลักษณะ Short Time (Decision Time Horizon for Music Genre Classification Using Short Time Feature)

ตารางที่ 5.2

แสดงค่าคุณลักษณะที่ดีที่สุดจากงานวิจัยของ Ahrendt , Meng and Larsen (2008)

L=0 (1 to 5)	LPC2	LPC1	MFCC2	LPC3	MFCC4
L=50 (1 to5)	**MFCC1**	MFCC4	MFCC6	MFCC2	LPC2
L=100(1to5)	**MFCC1**	MFCC4	MFCC6	MFCC7	ASE19
L=0 (6 to10)	LPC4	LPC5	GAIN	MFCC1	MFCC3
L=50(6to10)	MFCC7	ASE19	LPC1	ASS	MFCC10
L=100(to10)	MFCC2	LPC2	MFCC13	ASS	ZCR

ที่มา: “Decision Time Horizon for Music Genre Classification Using Short Time Feature,”โดย Ahrendt , Meng and Larsen (2008), Informatics and Mathematical Modelling, University of Denmark

จากภาพที่ 5.3 แสดงค่าคุณลักษณะที่ดีที่สุด ค่าส่วนใหญ่ที่ดีที่สุดจากการใช้โครงข่ายประสาทเทียมแบบเชิงเส้นจำแนกประเภทเพลง และใช้ร่วมกับ Function Time Stack ซึ่งเป็นฟังก์ชั่นที่ใช้คำนวณหาค่าคุณลักษณะที่ดีที่สุดในช่วงระยะเวลาสั้น พบว่าเมื่อค่า L = 50 และ L = 100 ค่าจะทำให้ MFCC 1 เป็นค่าคุณลักษณะที่ดีที่สุด 1 ใน 10 ค่าคุณลักษณะ โดยที่ค่า L เป็น Lag Parameter ที่ใช้คำนวณใน Function Time Stack และพบว่าสามารถจำแนกประเภทเพลงแจ๊สได้ดีภายในระยะเวลา 5 วินาที

5.3 การอภิปรายผลการเปรียบเทียบการจำแนกเพลงไทยและฝรั่ง

ในส่วนการจำแนกเพลงแบบ 8 ประเภท 2 ขั้นตอน ขั้นตอนแรกจะทำการแยกประเภทเพลงระหว่างเพลงที่ขับร้องด้วยภาษาไทยและภาษาฝรั่งพบว่าค่าในการจำแนกเพลงที่ใช้ภาษาไทยและฝรั่งตั้งแต่กลุ่มที่มี 2 ค่าคุณลักษณะขึ้นไปสามารถจำแนกประเภทเพลงได้ดีที่ 99.75% เกิดจากจำนวนกลุ่มในการจำแนกลดลงเหลือแค่เพลงไทยและฝรั่ง 2 กลุ่มเท่านั้นทำให้ความถูกต้องในการจำแนกจะสูงขึ้นอีก 20%

ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาไทยกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC+SC+SF ให้ค่าความถูกต้องที่ 72.47% ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาฝรั่งกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC ให้ค่าความถูกต้องที่ 73.43% เมื่อเปรียบเทียบค่าในการจำแนกประเภทเพลงในแต่ละภาษาพบว่าให้ผลในการจำแนกใกล้เคียงกันเนื่องจากเพลงไทยในปัจจุบันได้รับอิทธิพลจากเพลงฝรั่งมากขึ้นอีกทั้งความต้องการของตลาดของคนฟังเพลงไทยในปัจจุบันนั้นต้องการประเภทเพลงที่เป็นสากลมากขึ้นทำให้การผลิตเพลงของค่ายเพลงในปัจจุบันต้องการที่จะสร้างเพลงเพื่อให้ได้เพลงที่ทัดเทียมกับเพลงฝรั่งและเพื่อต้องการที่จะส่งเพลงไทยออกไปขายยังต่างประเทศ

5.4 การอภิปรายการทดสอบเพลงที่ใช้ทำนองเดียวกันแต่ภาษาแตกต่างกัน

ในขั้นตอนนี้จะทำการแยกประเภทเพลงระหว่างเพลงที่ขับร้องด้วยภาษาไทยและภาษาฝรั่งเท่านั้น เนื่องจากประเภทเพลงที่ได้เป็นเพลงอคูสติกและแนวประสานเสียงซึ่งไม่ตรงตามประเภทเพลงที่ต้องการ จากการทดลองพบว่าค่า ZC ให้ค่าในการจำแนกสูงสุดที่ 61.54 % ซึ่ง Xu,Maddage and Shao (2005) อธิบายถึงค่าคุณลักษณะของ LPC และ ZC จะมีความไวต่อการรับค่าในการจำแนกค่าเพลงที่เป็นเสียงเพลงและร้องได้ดี

5.5 สรุปผลที่ได้จากการวิจัย

ในส่วนนี้ประกอบด้วยกัน 2 ส่วนโดยส่วนแรกจะอภิปรายผลที่ได้การวิจัยเชิงทฤษฎี (Theoretical Implication) ซึ่งจะรวมประโยชน์ที่ได้ในทางทฤษฎี ส่วนที่ 2 จะอภิปรายผลที่ได้จากการวิจัยเชิงปฏิบัติ (Practical Implication) ซึ่งมีรายละเอียดดังนี้

5.5.1 ผลที่ได้จากการวิจัยเชิงทฤษฎี (Theoretical Implication)

จากการปรับค่าพารามิเตอร์ของโครงข่ายประสาทเทียมแบบหลายชั้นของงานวิจัยฉบับนี้พบว่าส่วนของ Learning rate และ Training Time ค่อนข้างสูงโดยโครงสร้าง (Topology) จะเป็น 34 30 8 อธิบายโครงสร้างได้ดังนี้ จำนวนโหนด Input เท่ากับ 34 จำนวนโหนดของ Hidden Unit เท่ากับ 30 และ จำนวนโหนดของ Output เท่ากับ 8 ซึ่งจะเป็นข้อมูลพื้นฐานในการพัฒนาต่อไป

ค่าคุณลักษณะที่มีผลต่อการจำแนกประเภทของเพลงที่สำคัญคือ MFCC, SR, LPC และ SC ซึ่งเป็นค่าคุณลักษณะที่สำคัญใน Timbral Feature จากการใช้ค่า Mean และ Median เป็นค่าตัวแทนทางสถิติเพื่อนำไปทดสอบ ค่าคุณลักษณะที่ใช้ในการจำแนกเสียงร้องที่สำคัญยังคงเป็น MFCC LPC และ ZC ซึ่งเป็นค่าคุณลักษณะที่สำคัญในการแยกเสียงร้อง เสียงพูด จากการทดลองทำให้เห็นค่าคุณลักษณะ 3 ค่านี้มีผลต่อการจำแนกประเภทเพลงอย่างเห็นได้ชัด

ผลการเปรียบเทียบระหว่างการจำแนกประเภทเพลง 8 ประเภท 1 ขั้นตอน และ การจำแนกประเภทเพลง 8 ประเภท 2 ขั้นตอนพบว่าในกลุ่มการจำแนกประเภทเพลง 8 ประเภท 1 ขั้นตอน กลุ่มค่าคุณลักษณะที่ให้ค่าในการจำแนกประเภทเพลงดีที่สุดคือ กลุ่ม ZC+SR+MFCC+

LPC+SC+SF ที่ 74.45% ส่วนการจำแนกประเภทเพลง 8 ประเภท 2 ขั้นตอนพบว่าค่าในการจำแนกเพลงที่ใช้ภาษาไทยและฝรั่งตั้งแต่กลุ่มที่มี 2 ค่าคุณลักษณะขึ้นไปสามารถจำแนกประเภทเพลงได้ดีที่ 99.75% ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาไทยกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC+SC+SF ให้ค่าความถูกต้องที่ 72.47% ส่วนการจำแนกประเภทเพลงที่ใช้ภาษาฝรั่งกลุ่มค่าคุณลักษณะที่ได้ค่าในการจำแนกดีที่สุดคือ ZC+SR+MFCC+LPC ให้ค่าความถูกต้องที่ 73.43% จากการเปรียบเทียบทั้ง 2 ขั้นตอนค่าความถูกต้องที่ได้จากการจำแนกประเภทเพลงต่างกันไม่เกิน 2% ทั้งที่กลุ่มในการจำแนกประเภทเพลงแบบ 2 ขั้นตอนประเภทกลุ่มการจำแนกประเภทลดลงเนื่องจากเพลงไทยสากลใส่เนื้อร้องภาษาไทยโดยใช้มาตรฐานของโน้ตเพลงแบบสากลอาจจะมีความใกล้เคียงกับโน้ตที่เป็นสากลที่เป็นเพลงฝรั่งทำให้การจำแนกประเภทเพลงไม่ต่างจากเดิมมากนัก

จากผลการทดลองการแยกประเภทเพลงที่ใช้ภาษาแตกต่างกันระหว่างภาษาไทยและฝรั่งพบว่าค่าคุณลักษณะ ZC,SR,MFCC,LPC,SC และ SF ตัดจากกึ่งกลางของเพลงความยาว 10 วินาที ให้ค่าความถูกต้องในการจำแนกสูงสุด เมื่อเปรียบเทียบกับช่วงวินาทีที่ 0 -10 ช่วงวินาทีที่ 0 -20 และช่วงวินาทีที่ 0 - 30 ใช้ความกว้างของเฟรมที่ 512 ซึ่งผลการทดลองดังกล่าวเหมาะสมกับการนำไปประยุกต์ใช้จริง ส่วนการจำแนกประเภทเพลงจากการจำแนกดนตรีคันทรีไทยและฝรั่งโดยใช้แบบโครงข่ายประสาทเทียม ลักษนันท์ แกล้วทนงค์ และ ณัฐธนนท์ หงส์วริทธิ์ธร (2553) พบว่าใช้ค่า MFCC 0 - 4 ก็พอเพียงต่อการจำแนกประเภทเพลง ซึ่งในส่วนการทดลองหาค่าที่ดีที่สุดระหว่าง MFCC 0 - 4 เพื่อนำค่าที่ได้เปรียบเทียบกับกลุ่มค่าคุณลักษณะเดียวพบว่าค่า MFCC 1 ให้ค่าดีที่สุดในการจำแนกเพลง 8 ประเภทและเมื่อนำค่า MFCC 1 เปรียบเทียบกับกลุ่มค่าคุณลักษณะเดียวพบว่าค่า MFCC 1 มีค่าความสามารถในการจำแนกประเภทเพลงเป็นลำดับที่สองรองจากค่า Spectral Roll Off ดังนั้นจากการทดลองนี้พบว่าการนำค่า MFCC 1 ค่าเดียวไปใช้ก็สามารถนำไปใช้ในการจำแนกประเภทเพลงได้ดีแต่ถ้าต้องการการจำแนกประเภทเพลงให้ได้ผลดีขึ้นนั้นในทางปฏิบัติถ้าสามารถนำค่า MFCC มาใช้ได้ทั้ง 5 ค่าจะให้ผลในการจำแนกที่ดีกว่าการนำค่า MFCC ค่าเดียวมาใช้งาน

ส่วนหนึ่งที่ทำให้เกิดการจำแนกยังมีค่าความถูกต้องค่อนข้างต่ำเป็นเพราะการพัฒนาของเพลงในปัจจุบันมีการผลมผสานของประเภทเพลงที่หลากหลายขึ้น อาทิเช่น เพลงประเภทร็อคในปัจจุบันมีการนำเครื่องดนตรีแบบอิเล็คทรอนิคที่มีความถี่หลายย่านความถี่มาผสมผสานทำให้ย่านความถี่ของเพลงร็อคแบบเดิมผิดแปลกไป ค่า Strength of Strongest Beat ที่ใช้ในการวิเคราะห์

จังหวะเพลงมีความถูกต้องในการจำแนกต่ำเนื่องจากในการทดลองไม่ได้แยกเพลงช้าและเร็วในแต่ละประเภทจึงส่งผลให้ไม่เกิดความแตกต่างของเพลงที่มีจังหวะใกล้เคียงกันเช่น จังหวะเพลงเศร้าในเพลงประเภทร็อคอาจจะมีจังหวะใกล้เคียงกับเพลงเร็วในเพลงประเภทป็อบ ส่วนสุดท้ายคือจำนวนค่าคุณลักษณะในช่วงแรกให้ค่าความถูกต้องเพิ่มขึ้นและเพิ่มขึ้นสูงสุดหลังจากนั้นได้ค่าความถูกต้องลดลงเมื่อจำนวนค่าคุณลักษณะมากขึ้นในระดับหนึ่งนั้นอาจจะหมายถึง ความไม่สัมพันธ์หรือซ้ำซ้อนของค่าคุณลักษณะที่เพิ่มเข้ามาไม่ได้มีส่วนเพิ่มทำให้ประสิทธิภาพในการจำแนกอีกต่อไป

5.5.2 **ผลที่ได้จากการวิจัยเชิงปฏิบัติ** (Practical Implication)

จำนวนเพลงที่เพิ่มขึ้นอย่างมากในแต่ละปีทำให้การจัดจำแนกประเภทเพลงเป็นไปได้ยากขึ้นอีกทั้งบุคลากรที่มีอยู่นั้นจำเป็นต้องประสบการณ์ อาศัยความชำนาญจากการฟังและความเข้าใจทางดนตรี ยิ่งในปัจจุบันเพลงในแต่ละประเภทแต่ละภาษาเพิ่มขึ้นอย่างรวดเร็ว การหาเครื่องทุนแรงเพื่ออำนวยความสะดวกและช่วยลดความผิดพลาดจึงเป็นสิ่งที่จำเป็น จากภาพ 4.2 Confusion Matrix พบว่าเปอร์เซ็นต์ความถูกต้องในการจำแนกประเภทเพลงพบว่าเพลงแจ็สฝรั่งสามารถทำนายเพลงถูก 43 เพลงจากทั้งหมด 49 เพลง เพลงประเภทร็อคฝรั่งสามารถทำนายเพลงถูก 40 จากทั้งหมด 49 เพลง เพลงร็อคไทยสามารถทำนายเพลงถูก 40 จากทั้งหมด 49 เพลง และ เพลงคันทรีไทยสามารถทำนายเพลงถูก 40 จากทั้งหมด 49 เพลงเช่นกัน ส่วนประเภทเพลงอื่น ๆ ที่เหลือไม่สามารถทำนายเพลงถูกเกิน 40 เพลง ดังนั้นการทำระบบการจำแนกเพลงไทยและฝรั่งด้วยโครงข่ายประสาทเทียมแบบหลายชั้นโดยใช้คุณลักษณะเชิงคุณภาพเสียงไปประยุกต์ใช้จริงนั้น ควรนำประเภทเพลงทั้ง สี่ประเภทไปประยุกต์ใช้ ซึ่งเมื่อนำจำนวนเพลง 40 เพลงจากทั้งหมด 49 เพลง คิดเป็นเปอร์เซ็นต์จะได้ประมาณ 81%ขึ้นไป จากกลุ่มคุณลักษณะ ZC+SR+MFCC+LPC+SC+SF

5.6 **ข้อเสนอแนะผลงานวิจัย**

5.6.1 เนื่องจากการทดลองครั้งนี้ต้องใช้เวลาในการเตรียมข้อมูล (Dataset) ค่อนข้างนาน หากต้องใช้ข้อมูลขนาดใหญ่อย่างจำนวนข้อมูลประมาณ 1,000 เพลงขึ้นไปจำเป็นต้องใช้เวลาในเตรียมข้อมูล ดังนั้นผู้ที่มีความสนใจในงานวิจัยด้านนี้ควรเผื่อเวลาในการเตรียมข้อมูลเป็นอย่างยิ่ง

ยกตัวอย่าง ข้อมูล (Dataset) ในแต่ละเพลงที่ได้จากการดึงค่าคุณลักษณะมีความยาวหลายร้อยบรรทัดเช่น ข้อมูลที่ดึงจากการเพลงช่วงระยะเวลา 0 – 30 มีความยาวประมาณ 900 กว่าบรรทัด การนำข้อมูลที่ได้จากงานวิจัยนี้ไปใช้จะสามารถประหยัดเวลาในการทำงานได้พอสมควร

5.6.2 ปัญหาจากการปรับค่าพารามิเตอร์ยังคงเป็นปัญหาที่ต้องใช้ระยะเวลาในการปรับหรือกลุ่มค่าคุณลักษณะที่มีผลต่อการจำแนกประเภทเพลงก็มีส่วนสำคัญเช่นกัน ข้อมูลในงานวิจัยนี้จะเป็นข้อมูลอย่างดีในการพัฒนาครั้งต่อไป

5.6.3 การจำแนกโดยใช้เสียงร้องภาษาไทยและฝรั่งอย่างเดียว (Vocal Track) ซึ่งในการทดลองครั้งนี้ยังไม่ได้ทดลองใช้อาจจะเป็นอีกหนึ่งการทดลองที่น่าสนใจ

5.6.4 ข้อมูล (Dataset) ในงานวิจัยฉบับนี้อาจจะน้อยเกินไป ในงานวิจัยครั้งต่อไปควรเตรียมข้อมูลในการจำแนกเพิ่มขึ้นซึ่งอาจจะมีผลในการจำแนกเพิ่มขึ้น

5.6.5 จากการทบทวนวรรณกรรมในการจำแนกประเภทเพลงพบว่าเทคนิคที่ให้ค่าจำแนกประเภทเพลงได้ดีนั้นยังมีเทคนิค Support Vector Machine และ Hidden Markov Model หากต้องการพัฒนาระบบให้สามารถจำแนกประเภทเพลงให้ดีขึ้นนั้นการนำเทคนิคนี้มาประยุกต์ใช้จะสามารถพัฒนาระบบการจำแนกเพลงได้ดีขึ้น