

บทที่ 1

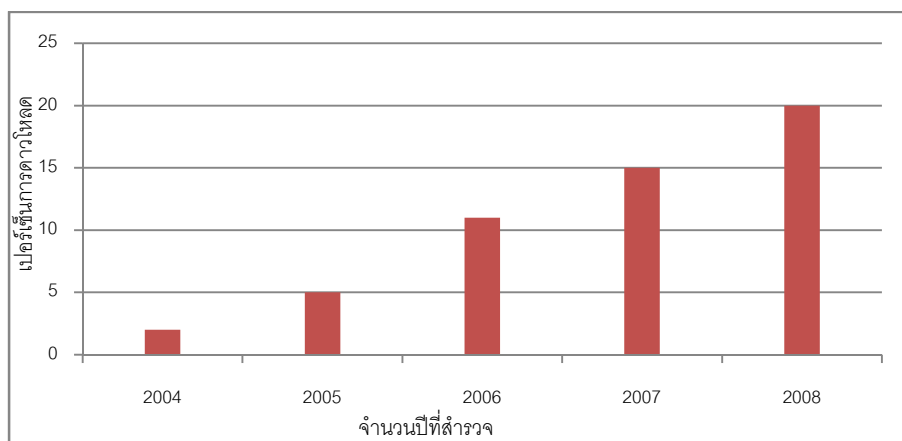
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การค้นคืนข้อมูลเสียงเพลง (Music Information Retrieval) ได้รับความสนใจอย่างต่อเนื่องจนถึงปัจจุบันเพื่อปรับปรุงเพื่อให้ได้ข้อมูลที่ดีที่สุด จากอดีตการจัดเก็บข้อมูลเสียงเจ้าหน้าที่จัดการข้อมูลเสียงจะต้องมีความชำนาญในด้านดนตรีและมีความเข้าใจในทฤษฎีดนตรีเพื่อทำหน้าที่ในการจัดเก็บประเภทดนตรีได้อย่างถูกต้อง ดังนั้นหากเจ้าหน้าที่ขาดความรู้ความชำนาญหรือขาดประสบการณ์การจำแนกประเภทของดนตรีจะส่งผลทำให้เกิดการผิดพลาดในการจำแนกประเภทดนตรีได้ ในปัจจุบันฐานข้อมูลเพลงในสถานีวิทยและช่องโทรทัศน์แนวโน้มของฐานข้อมูลเพลงเพิ่มขึ้นอย่างต่อเนื่อง ฐานข้อมูลเพลงประเภทดิจิทัลเป็นประเภทของข้อมูลที่มีความสำคัญและได้รับความนิยมอย่างมากในปัจจุบัน

ภาพที่ 1.1

แสดงยอดการดาวน์โหลดเพลงระหว่างปี 2004-2008



ที่มา Adrian Strain or Alex Jacob at IFPI Communications London,(16 th January 2009).”IFPI publishes Digital Music Report 2009”.Retrieved 13 ,September,2011,from <http://www.ifpi.org/content /section-resources/dmr2009.htm/>

จากรายงานของ IFPI (International Federation of the Phonographic Industry) ซึ่งนำเสนออุตสาหกรรมทางด้านแผ่นเสียงดนตรีทั่วโลกจาก 1400 สมาชิกใน 66 ประเทศและบริษัทร่วมใน 45 ประเทศที่มีการดาวโหลดสูงขึ้นจากปี 2004 – 2008 เพิ่มขึ้นเกือบ 18% จากจำนวนคนกว่า 3 พันล้านคนแสดงให้เห็นค่านิยมในการฟังเพลงและดาวโหลดเพลงเพิ่มขึ้นอย่างต่อเนื่องใน 4 ปีที่ผ่านมาความนิยมในการฟังเพลงตามคลื่นวิทยุในประเทศไทยจะสังเกตได้ว่าจะมีเพลงที่ได้รับความนิยมไม่ใช่เฉพาะแต่เพลงในประเทศไทยเพียงอย่างเดียว ความนิยมในการฟังเพลงของคนไทยนิยมฟังเพลงสากลมานานตั้งแต่อดีตในช่วงสงครามโลกครั้งที่ 2 (พ.ศ.2481) ดนตรีทางตะวันตกจะเข้ามามีบทบาทในประเทศไทย หลังจากนั้นค่านิยมในการฟังเพลงสากล ก็เริ่มเพิ่มขึ้นและมีบทบาทในประเทศไทยอย่างต่อเนื่องและมีอิทธิพลต่อคนไทยเรื่อยมา สังเกตการจัดอันดับเพลงในคลื่นวิทยุปัจจุบันที่มีการจัดเพลงสากลรวมกับเพลงไทยแสดงให้เห็นว่ามีเพลงสากลได้รับความนิยมอย่างมาก บางสัปดาห์ได้รับความนิยมจนได้รับการจัดอันดับให้อยู่ลำดับที่ 1 ดังนั้นเมื่ออิทธิพลของเพลงที่มีภาษาแตกต่างกันได้รับความนิยมจากคนไทยและมีการเพิ่มขึ้นอย่างต่อเนื่องของฐานข้อมูล การหาเครื่องมือช่วยในการแบ่งเบาภาระจึงเป็นสิ่งสำคัญในการจำแนกประเภทดนตรี

ในการศึกษางานวิจัยที่ผ่านมาทั้งภายในและภายนอกประเทศให้ความสนใจในการศึกษาการจำแนกประเภทดนตรีสากลหลายประเภททั้ง Classical Jazz Dance Lullaby Country Bossa Piano Blues และ Hip-hop เช่น Tzanetakis and Cook (2002) ค้นคว้าการดึงลักษณะของเสียงร้องหรือเสียงเครื่องดนตรีที่ผิดแผกกัน จังหวะของเพลง และระดับของเสียงดนตรี Xu, Maddage, Shao, Cao, and Tian (2003) ได้แยกแยะดนตรี ร็อค ป๊อป แจ๊ส และ คลาสสิก โดยใช้ Support Vector Machine ให้ความถูกต้องในระดับที่น่าพอใจ ในส่วนของภาษาอื่นเช่น Shayamala Doraisamy et al. (2008) ได้จำแนกประเภทดนตรีของมาเลเซียออกเป็น 10 หมวดหมู่โดยผลที่จำแนกความถูกต้องดีที่สุดคือ MLP Multi-layer neural network with back propagation โดยให้ค่า 88.6 % งานวิจัยภายในประเทศได้มีการศึกษาการจำแนกประเภทเพลงไทยเดิม เพลงไทยลูกทุ่ง ส่วนใหญ่มุ่งเน้นไปในส่วนของการหาภาษาใดภาษาหนึ่ง ยังไม่มีงานวิจัยใดที่ทำการศึกษาจำแนกหมวดหมู่ดนตรีที่ใช้ภาษาที่แตกต่างกันมาทำการจำแนกหมวดหมู่ดนตรี รวมถึงตัวคุณลักษณะที่ใช้วิเคราะห์จังหวะ (Rhythm Based Feature) ที่ใช้ในการวิเคราะห์จังหวะดนตรี ดังนั้นงานวิจัยชิ้นนี้ได้มีการนำเสนอ คุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ร่วมกับ Strength of Strongest Beat ซึ่งเป็นค่าคุณลักษณะในการวิเคราะห์จังหวะดนตรีซึ่งเป็นส่วนประกอบที่สำคัญของเพลงและใช้ตัวจำแนก Artificial Neural Network (ANN)

โดยใช้คลิปดนตรี 4 แบบ 10 วินาที 20 วินาที 30 วินาที และสุดท้ายคลิปที่กึ่งกลางเพลงความยาว 10 วินาที

1.2 วัตถุประสงค์

1.2.1 เพื่อศึกษาเปรียบเทียบความถูกต้องในการจำแนกภาษาที่ใช้ดนตรีที่แตกต่างกันด้วยวิธี Artificial Neural Network (ANN)

1.2.2 เพื่อศึกษาเปรียบเทียบขนาดของเฟรมและขนาดของระยะเวลาของเพลงที่ใช้ในการจำแนกประเภทเพลงและค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ที่เหมาะสมในการจำแนก

1.2.3 เพื่อศึกษาค่าคุณลักษณะค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ร่วมกับ Strength of Strongest Beat ที่เหมาะสมในการจำแนก

1.2.4 เพื่อศึกษาการหาพารามิเตอร์ที่เหมาะสมของ Artificial Neural Network (ANN)

1.3 ขอบเขตการศึกษา

1.3.1 ศึกษาค่าคุณลักษณะ (Feature) ของเพลงที่สามารถนำมาใช้ในการจำแนกเพลงไทยและเพลงฝรั่งแยกเป็น เพลงร็อค เพลงป๊อบ เพลงคันทรี่ และเพลงแจ๊ส ในงานวิจัยจะศึกษาเฉพาะเพลงไทยและเพลงภาษาฝรั่งเท่านั้นซึ่งในวิทยานิพนธ์ฉบับนี้จะขอใช้การเรียกเพลงที่ซับซ้อนด้วยภาษาฝรั่งเป็นเพลงฝรั่งเพื่อความเข้าใจง่าย

1.3.1.1 ศึกษาทฤษฎีและวิธีการทำงานของค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) ได้แก่ Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZC), Linear Predictive Coefficients (LPC), Spectral Centroid (SC), Spectral Flux (SF), Spectral Roll- Off (SR) และ Strength of Strongest Beat (SSB)

1.3.1.2 ศึกษาการจำแนกประเภทของเพลงร็อค เพลงป๊อบ เพลงคันทรี่ และเพลงแจ๊สที่ใช้ภาษาในที่แตกต่างกัน ภาษาฝรั่งและภาษาไทยด้วยค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) และ Strength of Strongest Beat

1.3.1.3 เทคนิคที่ใช้ในการดึงค่าคุณลักษณะเด่นด้วยค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) โดย Timbre หรือ Tone Color คือลักษณะที่แตกต่างกันของเสียงแต่ละชนิดโดยคลื่นเสียงที่ออกมาจากแหล่งกำเนิดที่ต่างชนิดกันก็จะมีลักษณะรูปร่างที่แตกต่างกันรูปคลื่นที่เล่นจากเครื่องดนตรีต่างชนิดกันก็จะมีรูปคลื่นไม่เหมือนกัน ทำให้เราสามารถแยกเสียงของเครื่องดนตรีได้ดังนั้นค่าคุณลักษณะเชิงคุณภาพเสียง (Timbral Feature) เป็นค่าคุณลักษณะที่บ่งบอกเอกลักษณ์ของเสียงร้องหรือเสียงเครื่องดนตรี Tzanetakis and Cook (2002) โดยค่าคุณลักษณะจะขึ้นกับการคำนวณ Short Time Fourier transform (STFT) และทุก Short Time Frame โดยค่าคุณลักษณะเด่นที่ใช้ 1) Mel Frequency Cepstral Coefficients 2) Spectral Centroid 3) Spectral Flux 4) Spectral Roll Off 5) Zero Crossing Rate ในส่วนของวิทยานิพนธ์นี้ Rhythm Feature จังหวะเป็นส่วนประกอบของบทเพลง ในเพลงหนึ่งจะต้องประกอบด้วยจังหวะ (Tempo) ควบคุมความช้าเร็วของเพลงดังนั้นการเลือกค่าที่จะดึงคุณลักษณะจึงเป็นส่วนสำคัญที่ทำให้คุณภาพการจำแนกดีขึ้นในงานวิจัยชิ้นนี้จึงเลือก Strength of Strongest Beat โดยค่าคุณลักษณะตัวนี้จะคำนวณจากผลรวมของค่า Beat Histogram

1.3.2 ศึกษาทฤษฎีและวิธีการทำงานของ Artificial Neural Network (ANN ซึ่งเป็นเทคนิคที่ใช้ในการจำแนกเป็นแบบจำลองจากระบบเซลล์ประสาทมนุษย์เพื่อสร้างเครื่องมือช่วยในการจดจำหรือจำแนกโดยปกติในสมองซึ่งประกอบด้วย เซลล์ประสาท หรือ “นิวรอน” (Neurons) และ จุดประสานประสาท (Synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า “เดนไดรต์” (Dendrite) ซึ่งเป็น input และปลายในการส่งกระแสประสาทเรียกว่า “แอกซอน” (Axon) ซึ่งเป็นเหมือน Output ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมีเมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่น ๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางแอกซอนของมัน Back - Propagation เป็นอัลกอริทึมที่ใช้ในการเรียนรู้ของเครือข่ายประสาทวิธีหนึ่งที่นิยมใช้ใน Multilayer Perceptron เพื่อปรับค่าน้ำหนักในเส้นเชื่อมต่อระหว่างโหนดให้เหมาะสม โดยการปรับค่านี้อาจจะขึ้นกับความแตกต่างของค่าเอาต์พุตที่คำนวณได้กับค่าเอาต์พุตที่ต้องการ ในเครือข่ายประสาทเทียมมีอัลกอริทึมที่นิยมใช้ในการเรียนรู้เช่น อัลกอริทึมแบบ RBFNetwork อัลกอริทึมแบบ แพร่กระจายงานนี้ได้เลือกใช้ Back- Propagation Algorithm ซึ่งเป็นเครือข่ายประสาทเทียมแบบ Multiple Layer Perceptron

1.4 ขั้นตอนการศึกษา

1.4.1 ทำการรวบรวมข้อมูลและเอกสารจากแหล่งข้อมูลที่เกี่ยวข้องในการจำแนกภาษาและหมวดหมู่เพลง

1.4.2 ทำการรวบรวมข้อมูลและเอกสารจากแหล่งข้อมูลที่เกี่ยวข้องกับการจำแนกประเภทโดยวิธี Artificial Neural Network (ANN)

1.4.3 ทำการรวบรวมข้อมูลและเอกสารจากแหล่งข้อมูลที่เกี่ยวข้องในการค่าดึงคุณลักษณะ (Feature Extraction) Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZC), Linear Predictive Coefficients (LPC), Spectral Centroid (SC), Spectral Flux (SF), Spectral Roll- Off (SR) และ Strength of Strongest Beat (SSB)

1.4.4 ทำการรวบรวมบทเพลงที่จะนำมาใช้ในการจำแนกทั้งประเภทของเพลงร็อค เพลงป๊อป เพลงคันทรี่ และเพลงแจ๊ส ที่ใช้ภาษาในการร้องเป็นภาษาฝรั่งเศสและภาษาไทย

1.4.5 ทำการทดลองและทดสอบความถูกต้องในการจำแนกดนตรีที่ใช้ภาษาที่แตกต่างกัน

1.4.6 วิเคราะห์ค่าผลการทดลองเปรียบเทียบความถูกต้องโดยวิธีของ Artificial Neural Network (ANN)

1.4.7 สรุปผลการทดลองและรวบรวมข้อเสนอนะ

1.5 คำจำกัดความในการวิจัย

1.5.1 การจำแนก (Classification) หมายถึง การจำแนกออกเป็นประเภท

1.5.2 การดึงค่าคุณลักษณะ (Feature Extraction) หมายถึง กระบวนการแยกหาค่าเอกลักษณ์ของเสียงของเสียงดนตรีที่ใช้ภาษาที่แตกต่างกัน

1.5.3 Mel Frequency Cepstral Coefficients (MFCC) หมายถึง เป็นค่าที่ปรับปรุงโดยปรับหรือกรองส่วนที่มนุษย์ไม่ได้ยินออก จะได้ความถี่ที่เหมาะสมสำหรับการรับฟังของมนุษย์ ซึ่งสามารถเก็บรายละเอียดของสัญญาณความถี่ต่ำได้ดี ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า MFCC แทนในครั้งต่อไป

1.5.4 Linear Predictive Coefficients (LPC) หมายถึง ค่าคุณลักษณะที่ได้จากการสุ่มอย่างของสัญญาณเสียง ได้จากผลรวมเชิงเส้น (Linear Combination) ของการสุ่มจาก

สัญญาณเสียงหลาย ๆ ส่วนจากตัวอย่างที่ผ่านมา ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า LPC แทนในครั้งต่อไป

1.5.5 Zero Crossing Rate (ZC) หมายถึง ค่าคุณลักษณะที่ได้จากการนับจำนวนครั้งของสัญญาณที่ผ่านแกนศูนย์ ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า ZC แทนในครั้งต่อไป

1.5.6 Spectral Centroid (SC) หมายถึง ค่าคุณลักษณะที่แสดงถึงค่ากึ่งกลางของ Power Spectrum โดยคำนวณจากค่าเฉลี่ยของค่า Power Spectrum ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า SC แทนในครั้งต่อไป

1.5.7 Spectral Flux (SF) หมายถึง ค่าที่คำนวณความแตกต่างระหว่างค่าปัจจุบันของแต่ละ Spectrum กับค่าที่สอดคล้องกับ Spectrum ที่สอดคล้องกับค่าก่อนโดยแต่ละค่าที่แตกต่างกันจะถอดด้วยรากที่สองของผลรวมจากการยกกำลังสอง ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า SF แทนในครั้งต่อไป

1.5.8 Spectral Roll Off (SR) หมายถึง ค่าของความถี่ภายใต้ 85 % ของการกระจายค่าความถี่ใช้วัดค่ารูปร่างที่เปลี่ยนแปลง ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า SR แทนในครั้งต่อไป

1.5.9 Strength of Strongest Beat (SSB) หมายถึง ค่าที่ได้ค่า Beat Histogram แสดงกำลังคาบจังหวะที่แตกต่างของสัญญาณโดยคำนวณจากการใช้ RMS ของ 256 ค่าจากนั้นใช้ Fast Fourier Transforms ซึ่งได้เป็นค่า Beat Histogram ส่วน Strength of Strongest Beat นำค่าสัญญาณ จาก Beat Histogram โดยใช้ค่าที่สอดคล้องกับ Strongest beat ค่าของผลรวมที่ถูกแบ่งจากสัญญาณ Beat Histogram ที่ผ่านเข้ามา ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า SSB แทนในครั้งต่อไป

1.5.10 Artificial Neural Network (ANN) หมายถึง วิธีการที่ใช้ในการเรียนรู้และทำนายหรือจำแนกค่าต่างโดยเรียนแบบลักษณะของเครือข่ายประสาทของสมองมนุษย์ในวิทยานิพนธ์ฉบับนี้ขอใช้คำว่า โครงข่ายประสาทเทียมแบบหลายชั้น แทนในครั้งต่อไป

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 ได้ค่าพารามิเตอร์ต่างๆของ Artificial Neural Network สำหรับใช้ในการจำแนกประเภทดนตรี

1.6.2 ทำให้ทราบค่าคุณลักษณะที่เหมาะสมในการจำแนกประเภทเพลงภาษาไทย และภาษาฝรั่งเศส แจ๊ส ป๊อป ร็อค และคันทรี่เพื่อเป็นการเตรียมข้อมูลในการพัฒนาต่อไป

1.6.3 ทำให้ทราบถึงขนาดของเฟรมและช่วงของคลิปเพลงในการจำแนกประเภทดนตรีที่สามารถจำแนกประเภทได้ดี