การจำแนกกลุ่มข้อมูลทางดนตรี

สุนันท์ ธาติ

วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การจำแนกกลุ่มข้อมูลทางดนตรี

สุนันท์ ธาติ

วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง "การจำแนกกลุ่มข้อมูลทางดนตรี"

ของ สุนันท์ ธาติ

ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปรัชญาดุษฎีบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการสอบวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.มงคล เอกปัญญาพงศ์)

..... ประธานที่ปรึกษาวิทยานิพนธ์

(ดร.วรลักษณ์ คงเด่นฟ้า)

..... กรรมการที่ปรึกษาวิทยานิพนธ์

(ผู้ช่วยศาสตราจารย์ ดร.พงศ์พันธ์ กิจสนาโยธิน)

..... กรรมการผู้ทรงคุณวุฒิภายใน

(ผู้ช่วยศาสตราจารย์ ดร.พนมขวัญ ริยะมงคล)

อนุมัติ

.....

(ศาสตราจารย์ ดร.ไพศาล มุณีสว่าง)

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง การจำแนกกลุ่มข้อมูลทางดนตรี
ผู้วิจัย สุนันท์ ธาติ
ประธานที่ปรึกษา ดร. วรลักษณ์ คงเด่นฟ้า
กรรมการที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.พงศ์พันธ์ กิจสนาโยธิน
ประเภทสารนิพนธ์ วิทยานิพนธ์ ปร.ด. สาขาวิชาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยนเรศวร, 2561
คำสำคัญ การทำเหมืองข้อมูล, การจัดกลุ่มข้อมูล, การรู้จำข้อมูลดนตรี, สเปคโตแกรม, ลายนิ้วมือทางเสียง

บทคัดย่อ

วิธีการทั่วไปในการตรวจสอบการละเมิดลิขสิทธิ์หรือการระบุข้อมูลเพลงคือการฟังโดยคน แต่การฟังมีข้อจำกัดในกรณีที่ต้องวิเคราะห์ข้อมูลเพลงจำนวนมากๆ อีกทั้งความถูกต้องแม่นยำยังขึ้นอยู่กับความเชี่ยวชาญของผู้ฟังแต่ละคน ด้วยข้อจำกัดดังกล่าวจึงมีการประยุกต์ใช้หลักการการรู้จำดนตรี (Music recognition) ในการแก้ไขปัญหานี้แทนและคุณลักษณะของข้อมูลที่ถูกนำมาใช้ในการวิเคราะห์อย่างกว้างขวางคือลายนิ้วมือทางเสียง (Audio fingerprint) การวิเคราะห์ลายนิ้วมือทางเสียงมีประสิทธิภาพมากในการตรวจสอบว่าข้อมูลเสียงใดเป็นเพลงเดียวกันในกรณีที่มีเนื้อหาตรงกัน (Exactly match) แต่ยังไม่สามารถตรวจสอบได้ในกรณีที่เนื้อหาคล้ายคลึงกัน (Similar) งานวิจัยฉบับนี้นำเสนอวิธีการหาความคล้ายคลึงกันของเพลงโดยใช้ฟังก์ชันความสัมพันธ์ (Relation function) สำหรับการเปรียบเทียบลายนิ้วมือทางเสียงแทนการเปรียบเทียบข้อมูลเสียงขนาดใหญ่ทีละคู่ และทดลองใช้วิธีการที่นำเสนอเพื่อระบุเพลงต้นฉบับจากเพลงคัฟเวอร์ ผลการทดลองพบว่าวิธีการดังกล่าวสามารถระบุเพลงต้นฉบับได้อย่างมีประสิทธิภาพครอบคลุมกลุ่มเพลงทุกประเภท (Genre) โดยมีพื้นที่ใต้กราฟเฉลี่ย (Average AUC) เป็น 0.790

Title MUSIC DATA CLUSTERING

Author SUNUN TATI

Advisor Woralak Kongdenfha, Ph.D.

Co-Advisor Phongphun Kijsanayothin, Ph.D.

Academic Paper Thesis Ph.D. in Computer Engineering, Naresuan University, 2018

Keywords Data mining, Data clustering, Music recognition, Audio fingerprint, Spectrogram

ABSTRACT

Listening is the most common way to detect copyright infringement or identify unknown music data, but it is difficult to analyze a large amount of music data. The accuracy also depends on the listener's level of expertise. As mentioned earlier, music recognition is applied to solve this problem and the audio fingerprint is widely used as data feature. Audio fingerprint analysis is effective at finding audio tracks which are duplicate content (exactly match) however it cannot detect in the case of similar content. This research proposes a method for finding the similarity between two songs using relation functions for comparing audio fingerprints instead of comparing bigger music content. For a case study, we try to find the original song from the cover song to assess the efficiency of our approach. The findings of this study indicate that proposed approach can be use effectively to identify the original song covered with many genres. The overall average area under curve (Average AUC) is 0.790.

ประกาศคุณูปการ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาเป็นอย่างดีจากอาจารย์ที่ปรึกษา ดร.วรลักษณ์ คงเด่นฟ้าที่เสียสละเวลาให้คำแนะนำและสร้างแรงบันดาลใจให้แก่ข้าพเจ้า อีกทั้งการให้ความช่วยเหลือในการให้คำแนะนำ และดูแลความก้าวหน้าอย่างสม่ำเสมอจาก ผศ.ดร.พงศ์พันธ์ กิจสนาโยธิน อาจารย์ที่ปรึกษาร่วม อีกทั้ง ผศ.ดร. มงคล เอกปัญญาพงศ์ และผศ.ดร. พนมขวัญ ริยะมงคล ที่ให้ข้อเสนอแนะในการปรับปรุงงานวิจัย ข้าพเจ้าขอขอบคุณท่านทั้งสี่เป็นอย่างยิ่ง

ขอขอบคุณอาจารย์ผู้สอนประจำหลักสูตรปริญญาเอก สาขาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ได้ให้ความรู้ซึ่งเป็นพื้นฐานในการดำเนินการวิจัย ขอขอบคุณ ผศ.ดร.อัครพันธ์ วงศ์กังแห ซึ่งให้คำแนะนำในการใช้ชีวิตในช่วงเวลาที่ขาดความหนักแน่นต่อจุดมุ่งหมายในการเรียนต่อ และ Mr. Roy I. Morien สำหรับการให้คำแนะนำในการเขียนเอกสารวิชาการเป็นภาษาอังกฤษ (Writing Academic Paper in English) ภายใต้โครงการ Graduate School 1-Stop English Shop ของบัณฑิตวิทยาลัย

และขอขอบคุณคณะวิศวกรรมศาสตร์ มหาวิทยาลัยนเรศวรที่มอบทุนในการศึกษา ขอขอบคุณเจ้าหน้าที่ประจำภาควิชาและคณะที่ช่วยอำนวยความสะดวกในระหว่างที่ศึกษา เจ้าหน้าที่บัณฑิตวิทยาลัยทุกท่านที่ช่วยดูแลและประสานงานระหว่างขั้นตอนการทำวิทยานิพนธ์ ตลอดจนน้องๆ ในทีมหน่วยวิจัยด้านวิศวกรรมระบบและวิศวกรรมซอฟต์แวร์ขั้นสูง (Advance Systems and Software Engineering Research Team : ASSERT) ที่เป็นเพื่อนร่วมงานที่ช่วยสนับสนุนให้เกิดการเรียนรู้และพัฒนาทักษะอยู่เสมอ

ท้ายนี้ผู้เขียนขอขอบบิดา มารดา ที่ให้การอุปการะอบรมเลี้ยงดู ตลอดจนส่งเสริมการศึกษาและให้กำลังใจเป็นอย่างดี ขอบคุณทุกคนในครอบครัวที่คอยเป็นแรงสนับสนุนที่สำคัญ

สุนันท์ ธาติ

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... ง

ประกาศคุณูปการ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ญ

บทที่ 1 1

บทนำ 1

1.1 ความเป็นมาของปัญหา 1

1.2 จุดมุ่งหมายของการศึกษา 2

1.3 ขอบเขตของงานวิจัย 2

1.4 ข้อตกลงเบื้องต้น 2

1.5 นิยามศัพท์เฉพาะ 2

บทที่ 2 1

เอกสารและงานวิจัยที่เกี่ยวข้อง 1

2.1 การสืบค้นข้อมูลดนตรี (Music Information Retrieval) 1

2.2 คุณลักษณะของข้อมูลดนตรี (Music feature) 1

2.3 การรู้จำข้อมูลดนตรี (Music recognition) 3

2.4 ลายนิ้วมือทางเสียง (Audio fingerprint) 4

2.5 แอปพลิเคชันรู้จำข้อมูลดนตรี (Music recognition application) 5
2.6 ความหลายหลากของชนิดของเพลง (Various of song) 5
2.7 จัดการลิขสิทธิ์ของข้อมูลทางดนตรี 6
บทที่ 3 8
วิธีดำเนินงานวิจัย 8
3.1 ประชากรและกลุ่มตัวอย่าง 8
3.2 เครื่องมือที่ใช้ในการวิจัย 8
3.3 กระบวนการเตรียมข้อมูล (Data pre-processing) 8
3.4 วิธีการที่นำเสนอ (Proposed Approach) 11
3.5 การประเมินประสิทธิภาพของวิธีการที่นำเสนอ 14
บทที่ 4 17
ผลการวิจัย 17
4.1 ผลการวิจัย 17
บทที่ 5 23
บทสรุป 23
5.1 สรุปผลการวิจัย 23
5.1.1 กลุ่มเพลงที่มีความแม่นยำในการทำนายสูง 23
5.1.2 กลุ่มเพลงที่มีสัญญานรบกวน (Noise) 23
5.1.3 กลุ่มเพลงที่ถูกคัฟเวอร์โดยผู้ร้องต่างเพศกับเพลงต้นฉบับ 23
5.1.4 กลุ่มเพลงที่ความเร็วของจังหวะในการเล่นเปลี่ยนแปลง แต่เนื้อร้องคงเดิม 24
5.1.5 กลุ่มเพลงที่มีการเปลี่ยนเนื้อร้อง เพิ่มเนื้อร้องหรือท่อนอิมโพรไวส์ (improvise) 24

5.2 ข้อเสนอแนะ 24

บรรณานุกรม 25

ประวัติผู้วิจัย 27

สารบัญตาราง

หน้า

ตาราง 1 แสดงรายละเอียดของเพลงที่เป็นกลุ่มข้อมูลตัวอย่าง 17

ตาราง 2 แสดงค่าเฉลี่ยของพื้นที่ใต้กราฟของ 20 เพลงในกลุ่มตัวอย่าง..... 18

สารบัญภาพ

หน้า

ภาพ 1 การแสดงข้อมูลดนตรี (Music representation) 2
ภาพ 2 ลักษณะของข้อมูลดนตรีและตัวอย่างวิธีการวิเคราะห์ข้อมูล 2
ภาพ 3 ขั้นตอนการใช้ลายนิ้วมือทางเสียงในการตรวจสอบ การละเมิดลิขสิทธิ์และการระบุข้อมูลเพลง 4
ภาพ 4 Music recognition application: Shazam and SoundHound 5
ภาพ 5 เซตของเพลงแยกตามรูปแบบการสร้าง (Various of song) 6
ภาพ 6 ภาพรวมของกระบวนการเตรียมข้อมูล (Data pre-processing) 9
ภาพ 7 หน้าต่างลายนิ้วมือทางเสียงของเพลงเต็มและลำดับค่าในแต่ละหน้าต่าง 11
ภาพ 8 ตัวอย่างการเปรียบเทียบเพื่อหาความคล้ายคลึงกันของเพลงเต็ม 11
ภาพ 9 ตัวอย่างการเปรียบเทียบหน้าต่างลายนิ้วมือทางเสียง A และ B 12
ภาพ 10 การวัดความคล้ายคลึงกันของเนื้อหาทางเสียง 13
ภาพ 11 ตัวอย่างในการคำนวณค่านัยสำคัญทางสถิติที่เกี่ยวข้อง 14
ภาพ 12 การเปรียบเทียบข้อมูลเสียงเพื่อระบุเพลงต้นฉบับจากเพลงคัฟเวอร์ 15
ภาพ 13 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0.5 15
ภาพ 14 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0.66 16
ภาพ 15 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 1 16
ภาพ 16 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0 16
ภาพ 17 กราฟแจกแจงความถี่ของค่าเฉลี่ยพื้นที่ใต้กราฟของทั้ง 20 เพลง 19
ภาพ 18 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงป๊อป 20

ภาพ 19 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงร็อค 20
ภาพ 20 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงโซลหรือแจ๊ส 21
ภาพ 21 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของ กลุ่มเพลงริทึมแอนด์บลูส์ ฮิปฮอป หรืออีดีเอ็ม 21
ภาพ 22 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของเพลงที่ไม่ดัดแปลงดนตรี 22
ภาพ 23 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของเพลงที่ดัดแปลงดนตรี 22

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

ในยุคดิจิทัลการละเมิดลิขสิทธิ์ทำได้ง่ายและก่อให้เกิดความเสียหายแก่ผู้เป็นเจ้าของผลงานเป็นมูลค่ามหาศาล ซึ่งอุตสาหกรรมที่เกี่ยวข้องจำนวนมากได้รับผลกระทบนี้ อาทิ สิ่งประดิษฐ์ สื่อสิ่งพิมพ์ ภาพยนตร์ ดนตรี เป็นต้น สำหรับอุตสาหกรรมดนตรีนั้น การเผยแพร่ผลงานเพลงโดยผู้ที่ไม่ใช่เจ้าของผลงานถือเป็นปัญหาที่ส่งผลกระทบในวงกว้างอย่างมาก การตรวจสอบการละเมิดลิขสิทธิ์ทำได้หลายวิธี ซึ่งวิธีการทั่วไปใช้การสุ่มตรวจโดยการฟังของเจ้าหน้าที่ลิขสิทธิ์ แต่เนื่องด้วยข้อจำกัดหลายประการ อาทิ ความยากในการวิเคราะห์ข้อมูลจำนวนมากๆโดยคนฟัง ประสบการณ์ของผู้ฟังที่ส่งผลต่อความถูกต้องของการวิเคราะห์ข้อมูลโดยตรง เป็นต้น ดังนั้นจึงมีการนำการสืบค้นข้อมูลดนตรี (Music information retrieval) มาช่วยในการตรวจสอบการละเมิดลิขสิทธิ์แทนการฟัง

การสืบค้นข้อมูลดนตรีเป็นหนึ่งในกลุ่มการวิจัยที่บูรณาการความรู้จากหลายสาขาวิชา โดยอาศัยความรู้ในด้านต่างๆ อาทิ การประมวลผลสัญญาณดิจิทัล (Digital signal processing) การจดจำรูปแบบ (Pattern recognition) ทฤษฎีดนตรี (Music theory) เป็นต้น เทคนิคการรู้จำดนตรี (Music recognition technique) [1] ถูกนำมาประยุกต์ใช้เพื่อการดึงคุณลักษณะเฉพาะที่เป็นตัวแทนของเนื้อหาเสียงและการเปรียบเทียบเพื่อการตรวจสอบการทำซ้ำ คุณลักษณะเฉพาะที่ถูกเลือกเป็นตัวแทนของเนื้อหาเสียงมีหลากหลายรูปแบบทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ของระบบ ซึ่งหนึ่งในคุณลักษณะเฉพาะที่ถูกนำมาใช้อย่างกว้างขวางในการตรวจสอบการละเมิดลิขสิทธิ์ คือ ลายนิ้วมือทางเสียง (Audio fingerprint) [2]

ลายนิ้วมือทางเสียงคือคุณลักษณะของข้อมูล (Data feature) ที่สามารถระบุถึงเนื้อหาของเสียงเช่นเดียวกับลายนิ้วมือที่แตกต่างกันของแต่ละบุคคลที่สามารถช่วยในการระบุตัวตนของมนุษย์ได้ สัญญาณเสียงที่เป็นข้อมูลเข้าของระบบจะถูกนำมาสร้างเป็นลายนิ้วมือทางเสียงเพื่อนำไปเปรียบเทียบกับลายนิ้วมือทางเสียงที่มีอยู่ในฐานข้อมูลอ้างอิง หากพบว่าเนื้อหาตรงกันถือว่าตรวจพบการละเมิดลิขสิทธิ์หรือสามารถระบุได้ว่าเป็นเพลงเดียวกันที่ถูกทำซ้ำนั้นเอง การประยุกต์ใช้ลายนิ้วมือทางเสียงที่มีอยู่ในปัจจุบันสามารถตรวจสอบข้อมูลเสียงว่าเป็นเพลงเดียวกันได้ในกรณีที่เนื้อหาตรงกัน (Exactly match) แต่ยังไม่สามารถตรวจสอบได้ในกรณีที่เนื้อหาทางเสียงคล้ายคลึงกัน (Similar) ทั้งที่การละเมิดลิขสิทธิ์ไม่ใช่เพียงการทำซ้ำโดยอ้างอิงตรงตามเนื้อหาเดิมเท่านั้น แต่การดัดแปลงรูปแบบของเพลงและนำมาเผยแพร่เพื่อแสวงหาผลกำไรโดยไม่จ่ายค่าลิขสิทธิ์แก่ผู้เป็นเจ้าของผลงานก็ถือเป็นการละเมิดลิขสิทธิ์รูปแบบหนึ่งเช่นกัน เนื่องด้วยกลุ่มเพลงดังกล่าวมีเนื้อหาคล้ายกันเท่านั้นจึงทำให้ไม่สามารถตรวจสอบได้เมื่อใช้วิธีการค้นหาเนื้อหาที่ซ้ำกัน เพลงคัฟเวอร์ (Cover song) ซึ่งเป็นเพลงที่ถูก

นำมาร้องใหม่และบันทึกเสียงอีกครั้งโดยผู้อื่นที่ไม่ใช่ศิลปินหรือนักแต่งที่เป็นเจ้าของถือเป็นตัวอย่างหนึ่งของการดัดแปลงรูปแบบของเพลงที่ไม่สามารถระบุข้อมูลเพลงที่เป็นต้นฉบับได้

1.2 จุดมุ่งหมายของการศึกษา

งานวิจัยนี้เสนอวิธีการในการหาความคล้ายกันของเพลงโดยหาระยะห่างระหว่างเพลงสองเพลงโดยการเปรียบเทียบลายนิ้วมือทางเสียง และนำวิธีการที่นำเสนอมาทดลองเพื่อค้นหาเพลงต้นฉบับจากเพลงคัฟเวอร์ โดยมีสมมุติฐานตั้งต้นว่าระยะห่างระหว่างเพลงต้นฉบับและเพลงคัฟเวอร์ของเพลงเดียวกันจะมีค่าน้อยกว่าระยะห่างระหว่างเพลงคัฟเวอร์นั้นและเพลงต้นฉบับอื่นที่ไม่ใช่เพลงๆนั้น นอกจากนี้ยังออกแบบการทดลองโดยคำนึงถึงปัจจัยที่เกี่ยวข้องต่างๆ อาทิ ช่วงความถี่ เวลาในการประมวลผล และความถูกต้องแม่นยำในการระบุเพลงต้นฉบับ เป็นต้น

1.3 ขอบเขตของงานวิจัย

เราใช้วิธีการที่นำเสนอในการหาความคล้ายคลึงกันของเพลงต้นฉบับกับเพลงคัฟเวอร์ ซึ่งเพลงคัฟเวอร์เกิดจากการนำเพลงที่ถูกปล่อยขายในเชิงพาณิชย์โดยผู้ซึ่งเป็นศิลปินหรือนักแต่งเพลงต้นฉบับมาบันทึกเสียงหรือแสดงใหม่ สำหรับการทดลองในการวิจัยนี้ เพลงคัฟเวอร์จะไม่รวมถึงการคัฟเวอร์เฉพาะดนตรี (Instrument cover) อาทิ กีตาร์, เบส, เปียโน, กลองหรือ การคัฟเวอร์เฉพาะดนตรีของดนตรีประเภทอื่นๆ

1.4 ข้อตกลงเบื้องต้น

เพลงคัฟเวอร์ที่ถูกสร้างมาจากเพลงต้นฉบับมากกว่าหนึ่งเพลงจะไม่ถูกนำมาพิจารณาในการวิจัยครั้งนี้

1.5 นิยามศัพท์เฉพาะ

การทำเหมืองข้อมูล, การจัดกลุ่มข้อมูล, การรู้จำข้อมูลดนตรี, สเปคโตแกรม และลายนิ้วมือทางเสียง

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

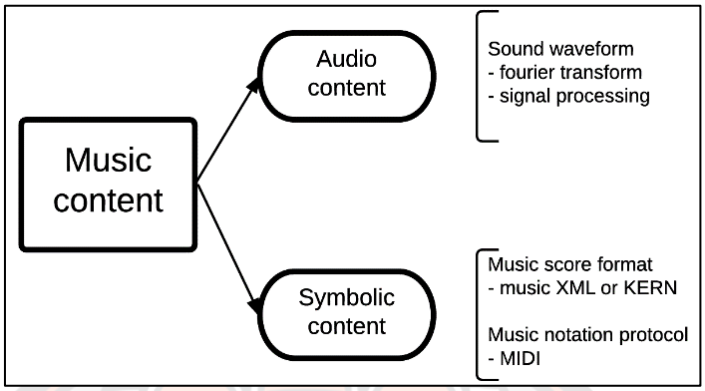
2.1 การสืบค้นข้อมูลดนตรี (Music Information Retrieval)

การสืบค้นข้อมูลดนตรี (Music Information Retrieval) เป็นหนึ่งในกลุ่มการวิจัยที่บูรณาการความรู้จากหลายสาขาวิชา โดยต้องอาศัยความรู้ในด้านต่างๆ อาทิ การประมวลผลสัญญาณดิจิตอล (Digital signal processing), การจดจำรูปแบบ (Pattern recognition) ทฤษฎีดนตรี (Music Theory) เป็นต้น การสืบค้นข้อมูลดนตรี (MIR) เป็นวิธีการค้นหาข้อมูลเพลงด้วยเทคนิคทางคอมพิวเตอร์ แอปพลิเคชันการสืบค้นข้อมูลดนตรีที่ผู้ใช้ทั่วไปรู้จักและใช้งานอย่างแพร่หลาย อาทิ iTunes, Spotify หรือ YouTube music ซึ่งเป็นแอปพลิเคชันสำหรับค้นหา เล่น และสร้างเพลย์ลิสเพลง อย่างไรก็ตามไม่ใช่เพียงผู้ฟังทั่วไปเท่านั้นแต่กลุ่มผู้เชี่ยวชาญด้านดนตรีก็สามารถใช้ประโยชน์จากเทคนิคการสืบค้นข้อมูลดนตรี (MIR technique) เช่นเดียวกัน ตัวอย่างเช่น นักดนตรีใช้แอปพลิเคชันสำหรับฟังเสียงและหาตำแหน่งของตัวโน๊ต (Score following) ในเพลงโดยอัตโนมัติ นอกจากนี้เทคนิคทางการรู้จำดนตรี (Music recognition technique) ยังสามารถช่วยจัดกลุ่ม แยกกลุ่ม และระบุลักษณะเฉพาะของงานดนตรีได้ โดยที่กระบวนการรู้จำดนตรีจะดึงข้อมูลจากเพลงเพื่อเป็นข้อมูลเข้าของระบบและหารูปแบบ (Pattern) ที่เหมือนหรือคล้ายกันของกลุ่มข้อมูลนั้นๆ เทคนิคการรู้จำดนตรี (Music recognition technique) ถูกนำเข้ามาประยุกต์ใช้เพื่อแก้ปัญหา อาทิ การแยกประเภทของเพลง การรู้จำเสียงเครื่องดนตรี การตรวจหาการละเมิดลิขสิทธิ์ หรือการระบุชื่อเพลง เป็นต้น

2.2 คุณลักษณะของข้อมูลดนตรี (Music feature)

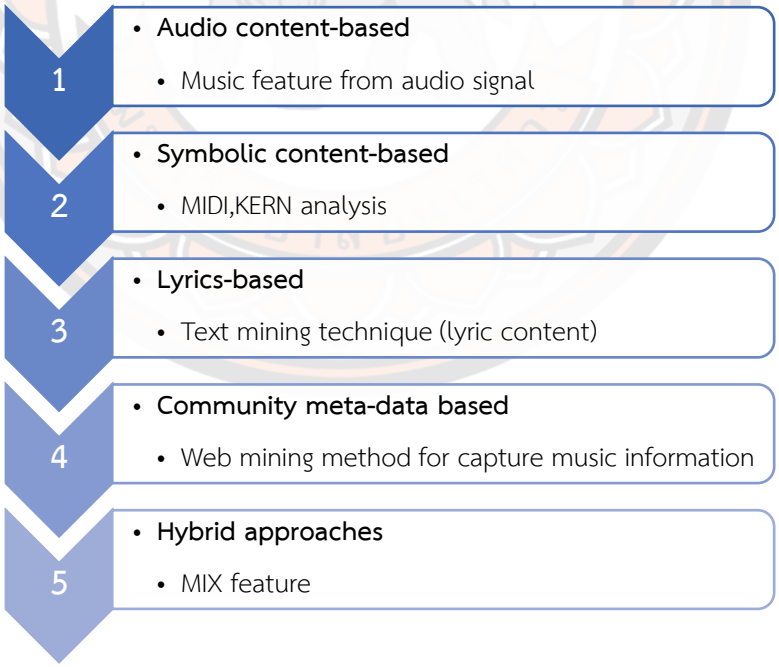
ไฟล์เสียง (Audio file) ถูกนำมาใช้เป็นข้อมูลตั้งต้นในกระบวนการดึงข้อมูล (Data extraction) คุณลักษณะของข้อมูล (Data feature) ถูกสร้างขึ้นหลังจากข้อมูลตั้งต้นถูกนำไปผ่านกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้ ข้อมูลดนตรีถูกแสดงในรูปแบบที่แตกต่างกันขึ้นอยู่กับวัตถุประสงค์ของการเก็บข้อมูลนั้นๆ ด้วยเหตุผลดังกล่าว การเลือกข้อมูล (Data feature extraction) จึงขึ้นอยู่กับวัตถุประสงค์ของการนำข้อมูลนั้นไปใช้เป็นหลัก โดยการแสดงข้อมูลดนตรีมีสองประเภท ประการแรกการแสดงข้อมูลดนตรีด้วยสัญลักษณ์ (Symbolic representation) [3, 4] ใช้สำหรับการอธิบายเนื้อหาทางดนตรีเพื่อให้นักดนตรีสามารถสื่อสารกันเพื่อบรรเลงเพลงได้ ตัวอย่างเช่น โน้ต สกอร์เพลง และการเข้ารหัสคอมพิวเตอร์ (Music notation protocol) เป็นต้น ประการที่สองการแสดงข้อมูลดนตรีด้วยเสียง (Audio representation) [1] ใช้สำหรับการบันทึกเสียง ซึ่งก็คือข้อมูลสัญญาณเสียงที่แสดงในรูปแบบอนาล็อกหรือดิจิทัล ตัวอย่างเช่น

แผ่นซีดี (CD) เทป แผ่นเสียงในรูปแบบอนาล็อก และไฟล์เสียงรูปแบบดิจิทัล เป็นต้น ประเภทของการแสดงข้อมูลดนตรีพร้อมด้วยตัวอย่างดังแสดงในภาพ 1



ภาพ 1 การแสดงข้อมูลดนตรี (Music representation)

วิธีการวิเคราะห์ข้อมูลดนตรีแยกตามลักษณะของข้อมูลดนตรีแบ่งออกเป็น 5 ลักษณะ โดยสรุปพร้อมระบุตัวอย่างวิธีการวิเคราะห์ของข้อมูลแต่ละแบบ ดังสรุปในภาพ 2



ภาพ 2 ลักษณะของข้อมูลดนตรีและตัวอย่างวิธีการวิเคราะห์ข้อมูล

2.3 การรู้จำข้อมูลดนตรี (Music recognition)

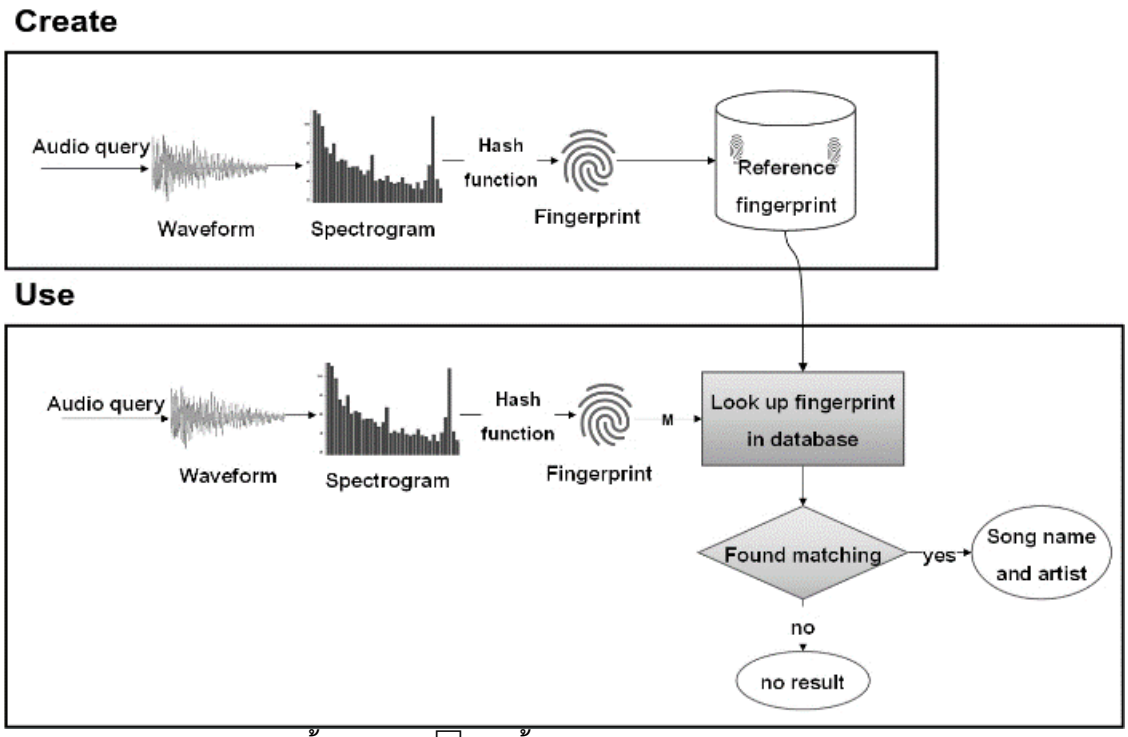
การฟังโดยมนุษย์เป็นวิธีการดั้งเดิมในการวิเคราะห์ข้อมูลดนตรี ซึ่งวิธีการดังกล่าวไม่สามารถทำได้ในกรณีที่มีข้อมูลจำนวนมาก อีกทั้งประสบการณ์ของผู้ฟังส่งผลต่อความถูกต้องของการวิเคราะห์ข้อมูลโดยตรง เนื่องด้วยข้อจำกัดที่กล่าวมาแล้ว วิธีการค้นหาข้อมูลเพลงด้วยเทคนิคทางคอมพิวเตอร์จึงถูกนำมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลดนตรีแทนการฟัง เทคนิคทางการรู้จำดนตรีสามารถช่วยจัดกลุ่ม แยกกลุ่ม และระบุลักษณะเฉพาะของงานดนตรีได้ โดยที่กระบวนการรู้จำดนตรีจะดึงข้อมูลจากเพลงเพื่อเป็นข้อมูลเข้าและหารูปแบบ (Pattern) ที่เหมือนหรือคล้ายกันของกลุ่มข้อมูลนั้นๆเพื่อนำมาใช้ในการวิเคราะห์ ตัวอย่างของการนำความรู้เรื่องการรู้จำดนตรีมาประยุกต์ใช้ เช่น การตรวจหาการละเมิดลิขสิทธิ์ การระบุข้อมูลเพลงในกรณีที่ไม่ทราบแหล่งที่มา การระบุประเภทเครื่องดนตรีจากเสียง และการแยกประเภท เป็นต้น

งานวิจัยก่อนหน้าที่ได้ศึกษาผลการวัดประสิทธิภาพวิธีการฟังโดยคนเทียบกับการประยุกต์ใช้เทคนิคการรู้จำดนตรี จากการทดลองของ Perrot และ Gjerdingen [5] พบว่ากรณีการจำแนกประเภทดนตรีด้วยวิธีการฟังโดยคน หลังจากฟังสามวินาที ผู้ที่ไม่เชี่ยวชาญด้านดนตรีสามารถระบุประเภทดนตรีได้ถูกต้อง 72% และการฟังมากกว่า สามวินาทีไม่ส่งผลให้ความถูกต้องเพิ่มขึ้น ข้อจำกัดในการฟังโดยคน ประการแรกในกรณีที่ต้องแยกประเภทเพลงจำนวนมาก เวลาในการฟังจะเพิ่มขึ้นตามไปด้วย ประการที่สองความถูกต้องในการแยกประเภทขึ้นอยู่กับความเชี่ยวชาญและประสบการณ์ในการฟังเป็นหลัก ในขณะที่ Tzanetakis และ Cook [6] ได้ทดลองจำแนกประเภทดนตรีสิบประเภท โดยประยุกต์ใช้เทคนิคการรู้จำดนตรีเพื่อแยกประเภทเพลงจำนวนมาก ความถูกต้องที่ได้เป็น 61% และบ่งชี้ว่าปัจจัยสำคัญที่ส่งผลต่อความถูกต้องมากที่สุดคือการเลือกข้อมูลที่เหมาะสมเพื่อเป็นข้อมูลเข้าของระบบ

การสืบค้นข้อมูลโดยใช้ตัวอย่าง (Query-by-example) เป็นเทคนิคที่ถูกนำมาประยุกต์ใช้ในการสืบค้นข้อมูลดนตรี ซึ่งเทคนิคดังกล่าวเป็นการเปรียบเทียบส่วนของข้อมูลเพลงเพื่อเลือกเพลงมีเนื้อหาทางเสียง (Audio content) ที่เหมือนกันมากที่สุดจากฐานข้อมูลอ้างอิง Chandrasekhar และทีมวิจัยของเขาได้ประเมินประสิทธิภาพการสืบค้นข้อมูลโดยใช้ตัวอย่าง (Query-by-example) ผลการวิจัยชี้ให้เห็นว่าวิธีนี้สามารถตรวจจับการละเมิดลิขสิทธิ์ได้ในเวลา 30-60 วินาที แต่ไม่สามารถทำงานได้ในแอปพลิเคชันบนอุปกรณ์เคลื่อนที่ ในขณะที่การระบุข้อมูลเพลง (Song identification) ต้องใช้เวลาอย่างน้อย 10 วินาทีแต่สามารถทำงานได้บนอุปกรณ์เคลื่อนที่ แอปพลิเคชันระบุข้อมูลเพลง (Song identification application) บนอุปกรณ์เคลื่อนที่จำนวนมากประยุกต์ใช้การสืบค้นข้อมูลโดยใช้ตัวอย่าง (Query-by-example) กับการวิเคราะห์ลายนิ้วมือทางเสียง (Audio fingerprint)

2.4 ลายนิ้วมือทางเสียง (Audio fingerprint)

งานวิจัยนี้เน้นการศึกษาการใช้ข้อมูลดนตรีที่แสดงด้วยเสียงเพื่อการวิเคราะห์ข้อมูล ซึ่งกระบวนการดึงข้อมูล เริ่มต้นด้วยการนำเข้าข้อมูลตั้งต้นซึ่งก็คือไฟล์เสียง (Audio file) จากนั้นคุณลักษณะของข้อมูลจะถูกสร้างขึ้นหลังจากไฟล์เสียงถูกนำไปผ่านกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้โดยพิจารณาลำดับเวลา ตัวอย่างของคุณลักษณะของข้อมูล อาทิ รูปแบบของคลื่น (Waveform) สเปคโตแกรม (Spectrogram) ภาพไบนารี (Binary Image) ลายนิ้วมือทางเสียง เป็นต้น ลายนิ้วมือทางเสียงเกิดจากการแปลงรูปแบบของคลื่น โดยใช้หลักการทางคณิตศาสตร์ในการคิดคำนวณเพื่อใช้เป็นตัวแทนข้อมูลเสียงและเป็นหนึ่งในคุณลักษณะของข้อมูลที่ถูกนำไปใช้อย่างกว้างขวางในการตรวจสอบการละเมิดลิขสิทธิ์และการระบุข้อมูลเพลง (Song Identification) สำหรับขั้นตอนการใช้ลายนิ้วมือทางเสียงที่กล่าวมาแล้วข้างต้นแสดงดังในภาพ 3



ภาพ 3 ขั้นตอนการใช้ลายนิ้วมือทางเสียงในการตรวจสอบการละเมิดลิขสิทธิ์และการระบุข้อมูลเพลง

ทั้งนี้การใช้ลายนิ้วมือทางเสียงในการระบุข้อมูลเพลงที่ไม่ทราบแหล่งที่มา โดยใช้การค้นหาข้อมูลจากเนื้อหาบางส่วน (Query-by-Example) [7] และการตรวจสอบการทำซ้ำของเพลง [4] มีประสิทธิภาพสูงในการตรวจสอบแทร็กเสียงที่ผลิตจากแหล่งเดียวหรือเนื้อหาซ้ำกัน (Duplicate

content) การใช้ลายนิ้วมือทางเสียงที่สร้างจากภาพไบนารีมีประสิทธิภาพดีกว่าการใช้ภาพไบนารีที่สร้างจากสเปคโตแกรมของข้อมูลเพลงโดยตรง [8] แต่กลับไม่สามารถตรวจสอบกรณีที่ข้อมูลเสียงมีเนื้อหาคล้ายคลึงกันแต่ไม่ตรงกันได้

2.5 แอปพลิเคชันรู้จำข้อมูลดนตรี (Music recognition application)

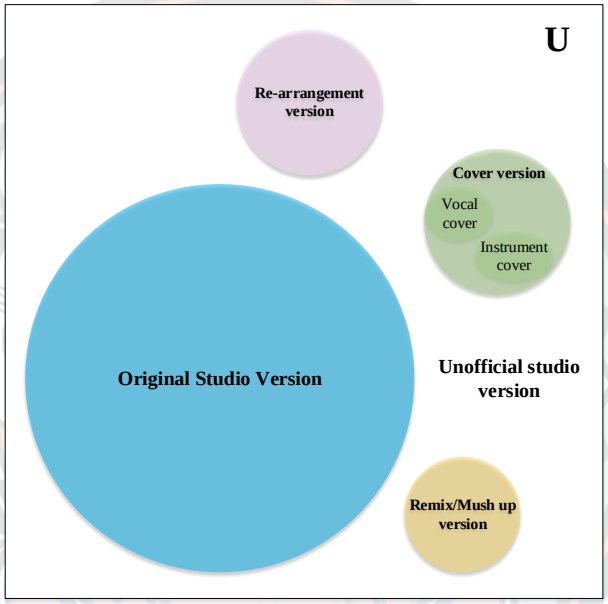
แอปพลิเคชันที่ใช้ในการระบุข้อมูลเพลงที่ถูกใช้งานอย่างกว้างขวางและเป็นที่รู้จัก ได้แก่ Shazam [9] และ Soundhound [10] ซึ่งแอปพลิเคชันดังกล่าวนี้สามารถช่วยระบุชื่อศิลปิน ชื่อเพลง และรายละเอียดอื่นๆของเพลงที่เรากำลังฟังได้โดยใช้สัญญาณเสียงเป็นข้อมูลเข้าเพื่อเปรียบเทียบกับลายนิ้วมือทางเสียง (Audio fingerprint) ที่มีอยู่ในฐานข้อมูลอ้างอิงเพื่อหาเพลงที่เหมือนกัน แอปพลิเคชันเหล่านี้สามารถตรวจสอบข้อมูลเสียงว่าเป็นเพลงเดียวกันได้ในกรณีที่เนื้อหาตรงกัน แต่ยังไม่สามารถทำได้ในกรณีที่ข้อมูลเสียงเหมือนกันเพียงบางส่วนหรือคล้ายกันเท่านั้น เพลงคัฟเวอร์ (Cover song) ซึ่งถูกนำมาร้องใหม่และบันทึกเสียงอีกครั้งโดยผู้อื่นที่ไม่ใช่ศิลปินหรือนักแต่งเพลงต้นฉบับเป็นตัวอย่างหนึ่งที่ไม่สามารถตรวจสอบได้โดยแอปพลิเคชันเหล่านี้

ภาพ 4 Music recognition application: Shazam and SoundHound

2.6 ความหลายหลากของชนิดของเพลง (Various of song)

เพลง Version Official คือเพลงที่เผยแพร่โดยศิลปินหรือผู้แต่งที่เป็นเจ้าของเพลง ในทางกลับกัน unofficial studio version คือเพลงที่เผยแพร่โดยผู้อื่นที่ไม่ใช่เจ้าของหรือพูดแต่งเพลงต้นฉบับ ตัวอย่างของ unofficial Version คือเพลงคัฟเวอร์ (Cover version) เพลงที่ถูกเรียบเรียงใหม่ (Re-arrangement version) หรือเพลงที่เกิดจากการรวมกันของเพลงมากกว่าหนึ่งเพลง (mush-up version) ลำดับแรกเพลงคัฟเวอร์เกิดจากการนำเพลงที่ถูกปล่อยขายในเชิงพาณิชย์โดยผู้ซึ่งเป็นศิลปินหรือนักแต่งเพลงต้นฉบับมาบันทึกเสียงหรือแสดงใหม่ ซึ่งแบ่งเป็น 2 ประเภทคือการคัฟเวอร์เฉพาะดนตรี (Instrumental cover) และการคัฟเวอร์ทั้งเพลงด้วยการร้อง (Vocal

cover) การคัฟเวอร์เฉพาะดนตรี (Instrumental cover) เป็นการบรรเลงดนตรีโดยไม่ร้องเนื้อร้อง ยกตัวอย่างเช่น กีตาร์, เบส, เปียโน, กลองหรือ การคัฟเวอร์เฉพาะดนตรีของดนตรีประเภทอื่นๆ และการคัฟเวอร์ด้วยการร้อง (Vocal cover) เป็นการร้องเนื้อเพลงและบันทึกเสียงใหม่ด้วยท่วงทำนองดนตรีที่คงเค้าโครงเดิมไว้ ลำดับถัดมาเพลงที่ถูกเรียบเรียงใหม่ คือเพลงที่ถูกนำไปบันทึกเสียงใหม่โดยที่การเรียบเรียงดนตรีเปลี่ยนแปลงไป สุดท้ายเพลงที่เกิดจากการรวมกันของเพลงมากกว่าหนึ่งเพลง (mush-up version) คือเพลงที่สร้างขึ้นจากการดัดแปลงโครงสร้างดนตรีเพื่อรวมเพลงมากกว่าหนึ่งเพลงขึ้นไปเข้าไว้ด้วยกัน ภาพแสดงเซตของเพลงแยกตามรูปแบบการสร้างดังแสดงในภาพ 5



ภาพ 5 เซตของเพลงแยกตามรูปแบบการสร้าง (Various of song)

2.7 จัดการลิขสิทธิ์ของข้อมูลทางดนตรี

การละเมิดลิขสิทธิ์ พระราชบัญญัติของไทยแบ่งออกเป็น 2 ประเภทคือการละเมิดลิขสิทธิ์โดยตรงและโดยอ้อม การละเมิดลิขสิทธิ์โดยตรงคือการทำซ้ำโดยไม่ได้รับความยินยอมจากเจ้าของลิขสิทธิ์ ในขณะที่การละเมิดลิขสิทธิ์โดยอ้อมคือการค้าหรือการมีส่วนสนับสนุนถึงการค้าสำหรับงานที่มีเจ้าของลิขสิทธิ์

สำหรับการคัฟเวอร์ที่กำลังเป็นที่นิยมอย่างมากสำหรับนักดนตรีสมัครเล่น การคัฟเวอร์เป็นการดัดแปลงเพลงของผู้ร้องเดิมโดยนำเพลงนั้นมาร้องใหม่ทั้งนี้หากทำขึ้นเพื่อความเพลิดเพลินส่วนบุคคล พระราชบัญญัติลิขสิทธิ์พ.ศ. 2537 มีการบัญญัติข้อยกเว้นที่ไม่ถือว่าเป็นการละเมิดลิขสิทธิ์ไว้ หากการคัฟเวอร์ดังกล่าวทำเพื่อผลประโยชน์ของตนเองหรือตนเองและบุคคลในครอบครัว หากมีการ

นำไปขายหรือหาผลประโยชน์จากการคัฟเวอร์ ดังกล่าวถือเป็นการละเมิดลิขสิทธิ์ เนื่องจากหากการค้าและการกระทำดังกล่าวมีผลต่อการแสวงหาผลประโยชน์ของเจ้าของลิขสิทธิ์ที่แท้จริง

การละเมิดลิขสิทธิ์โดยการทำซ้ำ ดัดแปลง เผยแพร่สู่สาธารณชน โดยไม่ได้รับการยินยอมจากเจ้าของลิขสิทธิ์ สำหรับในกรณีการตรวจจับการละเมิดลิขสิทธิ์ในประเทศไทย เจ้าของลิขสิทธิ์หรือตัวแทนที่ได้รับมอบอำนาจจะต้องไปร้องทุกข์กับเจ้าหน้าที่ตำรวจ เพราะตัวแทนหรือเจ้าของลิขสิทธิ์เองไม่มีอำนาจในการดำเนินคดีโดยลำพัง

บทที่ 3
วิธีดำเนินงานวิจัย

3.1 ประชากรและกลุ่มตัวอย่าง

การทดลองประยุกต์ใช้วิธีการหาความคล้ายกันของเพลงเพื่อระบุเพลงต้นฉบับจากเพลงคัฟเวอร์ ข้อมูลที่ใช้สำหรับการทดลองเป็นไฟล์เสียงจำนวน 120 ข้อมูลตัวอย่าง ซึ่งประกอบด้วย 1) เพลงเวอร์ชันต้นฉบับ 20 ข้อมูลตัวอย่าง โดยเพลงที่ถูกเลือกเป็นเพลงที่มีจำนวนการถูกคัฟเวอร์จำนวนมากซึ่งประกอบด้วยเพลงสากลจำนวน 15 เพลงและเพลงเกาหลีจำนวนห้าเพลง และ 2) เพลงเวอร์ชันคัฟเวอร์จำนวนห้าข้อมูลตัวอย่างต่อหนึ่งเพลงต้นฉบับ รวมเป็น 100 ข้อมูลตัวอย่าง

3.2 เครื่องมือที่ใช้ในการวิจัย

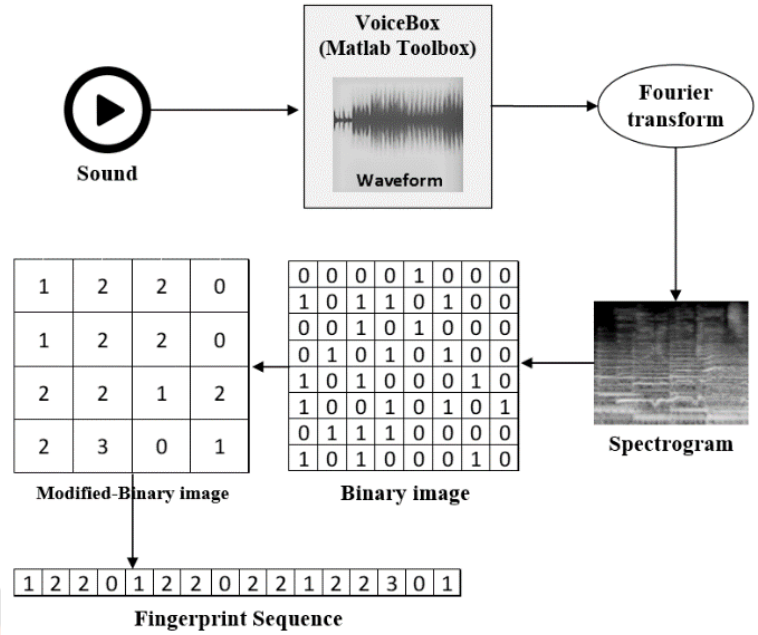
แมตแล็บ (MATLAB) เป็นโปรแกรมสำหรับการคำนวณและเขียนโปรแกรมเพื่อสร้างแบบจำลองทางคณิตศาสตร์ รวมถึงการจำลองการทำงานของระบบ โดยสำหรับวิจัยนี้ แมตแล็บ (MATLAB) ถูกนำมาใช้ในการเขียนโปรแกรมเพื่อสร้างลายนิ้วมือทางเสียงและการคำนวณหาระยะห่างระหว่างเพลงสองเพลงด้วยวิธีการที่นำเสนอเพื่อระบุความคล้ายกันของเพลง

VOICEBOX เป็นส่วนขยายของโปรแกรมแมตแล็บ (MATLAB toolbox) ที่ใช้สำหรับการประมวลผลข้อมูลสัญญาณเสียง โดยสำหรับงานวิจัยนี้ใช้ส่วนขยายนี้ในการเปลี่ยนเนื้อหาของเพลงที่อยู่ในรูปแบบของไฟล์ดิจิตอลให้เป็นรูปแบบของคลื่น (Waveform) และการเปลี่ยนรูปแบบของคลื่นจากที่อยู่ในโดเมนของเวลาไปอยู่ในโดเมนของความถี่โดยการแปลงฟูเรียรอย่างเร็ว (Fast Fourier transform-FFT)

3.3 กระบวนการเตรียมข้อมูล (Data pre-processing)

สำหรับกระบวนการเตรียมข้อมูลนั้น ข้อมูลตัวอย่างที่เป็นไฟล์เสียงถูกนำมาสร้างเป็นเฟรมของรูปคลื่นสัญญาณความยาว 0.3 วินาที จากนั้นนำไปสร้างสเปคโตแกรมที่ช่วงความถี่ 40 ถึง 1200 Hz ซึ่งเป็นช่วงความถี่ของเสียงร้องต่ำที่สุดของผู้ชายถึงเสียงร้องสูงสุดของผู้หญิงในเพลง (Vocal frequency range) [11] ขั้นตอนถัดมานำเฟรมของสเปคโตแกรมทีละแปดเฟรมมาสร้างเป็นภาพไบนารี และขั้นตอนสุดท้ายสร้างหน้าต่างของลายนิ้วมือทางเสียงจากภาพไบนารีดังกล่าว

สำหรับงานวิจัยนี้การเตรียมข้อมูล (Data pre-processing) ประยุกต์แนวคิดการสร้างลายนิ้วมือทางเสียง [3] ซึ่งมีสามขั้นตอนประกอบด้วย (1) การสร้างสเปคโตแกรม (2) การสร้างภาพไบนารี และ (3) การสร้างลายนิ้วมือทางเสียงจากภาพไบนารี ภาพรวมของกระบวนการเตรียมข้อมูลดังแสดงในภาพ 6



ภาพ 6 ภาพรวมของกระบวนการเตรียมข้อมูล (Data pre-processing)

เพลงที่มีความยาวเต็มทั้งเพลงในรูปแบบไฟล์เสียงเป็นข้อมูลตั้งต้นที่ถูกนำไปสร้างรูปคลื่นของสัญญาณเสียง ซึ่งแสดงลักษณะสัญญาณที่เปลี่ยนไปในช่วงเวลาต่างๆด้วยอัตราการสุ่มตัวอย่าง 44,100 ครั้งต่อวินาที โดยใช้ VOICEBOX [12] ที่เป็นกล่องเครื่องมือการประมวลผลเสียง (Speech processing toolbox) ในโปรแกรม MATLAB [13] หลังจากนั้นแบ่งรูปคลื่นของสัญญาณเสียงที่มีความยาวทั้งเพลงเป็นเฟรมเสียง (Spectrogram frame) และแปลงรูปคลื่นสัญญาณของเฟรมเสียงบนโดเมนเวลา (Time domain) ให้อยู่บนโดเมนความถี่ (Frequency domain) โดยการแปลงฟูเรียร์แบบเร็ว (FFT) รูปคลื่นสัญญาณเสียงที่ได้ของแต่ละเฟรมจะถูกนำไปสร้างสเปคโตแกรมดังขั้นตอนที่อธิบายในอัลกอริทึมที่ 1

Algorithm 1: Generate spectrogram

```
Data: x : matrix represent audio file
Result: y : matrix of spectrogram
Set initial value ;
Set frequency band i=1,2,3,...8;
Slice full length waveform to frame;
fftdata=fourier transform of each frame;
for j=1 to length(fftdata) do
    currentSample=fftdata(j);
    if currentSample is in the i band then
        y(j,i)=y(j,i)+currentSample;
    else
        do nothing;
    end
end
```

ในขั้นตอนต่อมาการสร้างภาพไบนารี เริ่มจากการสร้างหน้าต่างของสเปคโตแกรม (Spectrogram window) โดยการนำแปดเฟรมของสเปคโตแกรมที่ได้จากขั้นตอนก่อนหน้ามารวมเข้าด้วยกัน จากนั้นนำมาสร้างเป็นภาพไบนารีโดยการเปรียบเทียบกับค่าเฉลี่ยเลขคณิต (Arithmetic mean) ซึ่งถูกคำนวณมาจากค่าของสมาชิกทุกตัวในหน้าต่างนั้นๆ [4] ภาพไบนารีจะถูกสร้างโดยกำหนดค่าศูนย์หรือหนึ่งแทนที่ค่าเดิมของสมาชิกแต่ละตัวในหน้าต่างตามเงื่อนไงดังอธิบายในอัลกอรึทึมที่ 2

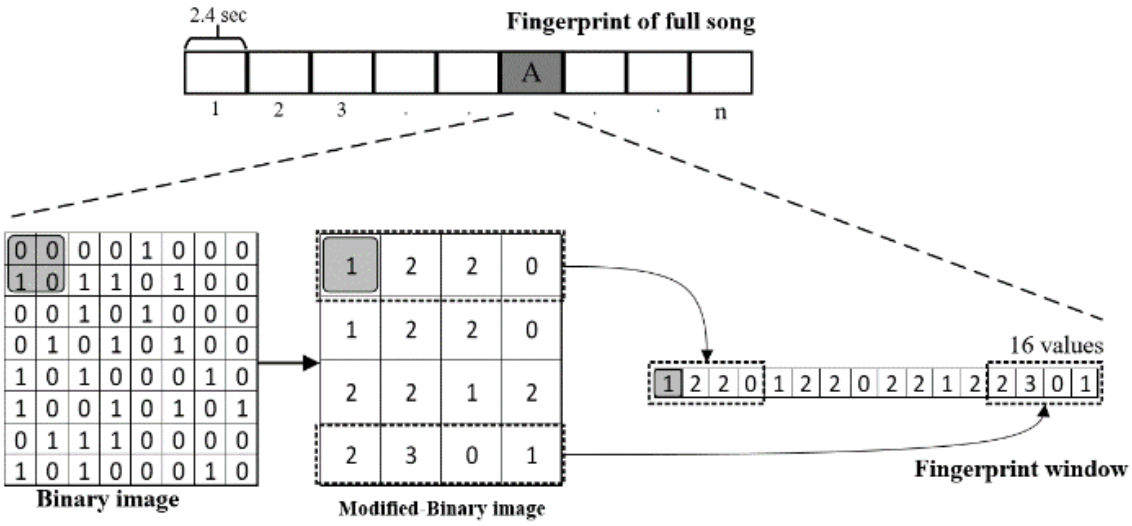
Algorithm 2: Generate binary image

```
Data: x : matrix of spectrogram
Result: y : matrix of binary image
split spectrogram (x) to eight frames window (y);
for i=1 to length(y) do
    if y(i) <mean of y window then
        replace y(i) by 0;
    else
        replace y(i) by 1;
    end
end
```

ในขั้นตอนสุดท้ายเป็นการสร้างลายนิ้วมือทางเสียงจากภาพไบนารี ซึ่งทำโดยการบวกรวมค่าของภาพไบนารีขนาดสองคูณสองเป็นค่าเดียว ดังนั้นหนึ่งหน้าต่างจะได้ค่าทั้งหมด 16 ค่าตามลำดับเรียกลำดับดังกล่าวว่า ลายนิ้วมือทางเสียง

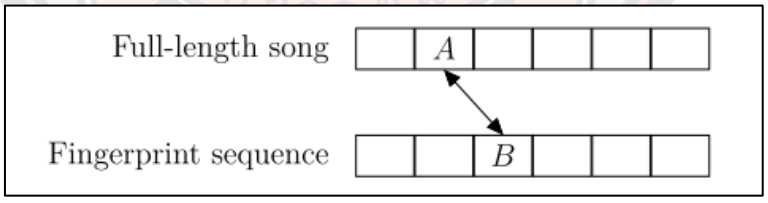
3.4 วิธีการที่นำเสนอ (Proposed Approach)

หลังจากที่ข้อมูลเพลงเต็มผ่านกระบวนการเตรียมข้อมูลแล้วจะถูกแบ่งเป็นหน้าต่างลายนิ้วมือทางเสียงขนาดเล็กที่ประกอบไปด้วย 16 ค่าข้อมูลต่อหนึ่งหน้าต่าง ดังแสดงรายละเอียดในภาพ 7



ภาพ 7 หน้าต่างลายนิ้วมือทางเสียงของเพลงเต็มและลำดับค่าในแต่ละหน้าต่าง

การเปรียบเทียบเพื่อหาความคล้ายคลึงกันของแต่ละหน้าต่าง เมื่อกำหนดให้ A และ B คือหน้าต่างลายนิ้วมือทางเสียงของเพลงสองเพลงที่แตกต่างกัน ดังแสดงในภาพ 8



ภาพ 8 ตัวอย่างการเปรียบเทียบเพื่อหาความคล้ายคลึงกันของเพลงเต็ม

ความคล้ายคลึงกันของสองหน้าต่างลายนิ้วมือทางเสียงจะถูกคำนวณโดยการวัดคะแนนที่ตรงกัน (Matching score) เมื่อหน้าต่างลายนิ้วมือทางเสียง A และ B มีค่าดังสมการที่ (1) และ (2) จะสามารถนิยามค่า (A-B) ได้ดังสมการที่ (3) โดยที่ w_i เป็นค่าที่สอดคล้องกับเงื่อนไขในสมการที่ (4)

$$A = < a_1, a_2, \ldots, a_n > \quad (1)$$

$$B = < b_1, b_2, \ldots, b_n > \quad (2)$$

$$A - B = < w_1, w_2, \ldots, w_n > \quad (3)$$

$$w_i = \begin{cases} 1, & if\ (a_i - b_i) = 0 \\ 0, & otherwise \end{cases} \quad (4)$$

สำหรับคะแนนที่ตรงกันของหน้าต่างลายนิ้วมือทางเสียง A และ B เป็นผลรวมค่าของสมาชิกทุกตัวใน (A-B) ซึ่งคะแนนที่ตรงกันจะเพิ่มขึ้นทีละหนึ่งเมื่อค่าในลำดับเดียวกันของทั้งสองหน้าต่างเท่ากัน ดังสอดคล้องกับสมการที่ (5)

$$\mu(A - B) = \sum_{i=1}^{n} w_i \mid w_i = (A - B)_{i^{th}} \quad (5)$$

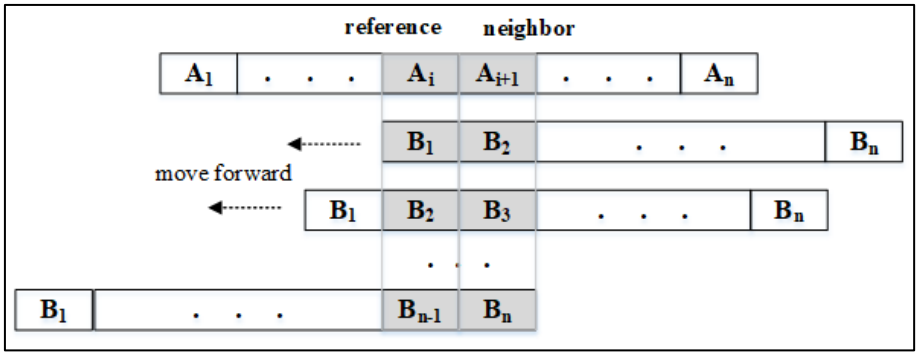
ตัวอย่างในการคำนวณคะแนนที่ตรงกันของหน้าต่างลายนิ้วมือทางเสียง A และ B แสดงดังภาพ 9

A =	2	2	1	1	1	2	1	4	2	2	1	0	1	2	0	2
B =	1	2	1	1	1	3	4	4	1	3	1	0	2	1	0	1
(A-B) =	0	1	1	1	1	0	0	1	0	0	1	1	0	0	1	0

$\mu(A, B) = 8$

ภาพ 9 ตัวอย่างการเปรียบเทียบหน้าต่างลายนิ้วมือทางเสียง A และ B

เนื่องด้วยเพลงหนึ่งประกอบไปด้วยหลายหน้าต่างลายนิ้วมือทางเสียง การวัดความคล้ายคลึงกันระหว่างเพลงสองเพลงนั้นจึงทำได้โดยการประเมินค่าคะแนนที่ตรงกันของทุกๆหน้าต่างลายนิ้วมือทางเสียง แต่งานวิจัยฉบับนี้ได้นำเสนอฟังก์ชันความสัมพันธ์ (Relation function) สำหรับการเปรียบเทียบลายนิ้วมือทางเสียงแทนการเปรียบเทียบข้อมูลขนาดใหญ่ทีละคู่ ดังแสดงในภาพ 10



ภาพ 10 การวัดความคล้ายคลึงกันของเนื้อหาทางเสียง

ฟังก์ชันความสัมพันธ์ (Relation function) ที่นำเสนอประกอบด้วยค่านัยสำคัญทางสถิติที่ถูกคำนวณคือค่าความคล้ายคลึงกัน (Sum of similarity: σ) และค่าความแตกต่าง (Degree of difference: δ) สำหรับค่าความคล้ายคลึงกันคือผลรวมของคะแนนที่ตรงกันของหน้าต่างอ้างอิง $(A_i - B_i)$ และหน้าต่างเพื่อนบ้าน $(A_{i+1} - B_{i+1})$ ในขณะที่ค่าความแตกต่างคือสัมบูรณ์ของความแตกต่างระหว่างคะแนนที่ตรงกันของหน้าต่างอ้างอิงและหน้าต่างเพื่อนบ้านดังสมการที่ (6) และ (7)

$$\sigma_i = \mu(A_i - B_i) + \mu(A_{i+1} - B_{i+1}) \qquad (6)$$

$$\delta_i = \left|\mu(A_i - B_i) - \mu(A_{i+1} - B_{i+1})\right| \qquad (7)$$

จากสมการข้างต้น σ_i และ δ_i หมายถึงค่าความคล้ายคลึงกันและความแตกต่างของหน้าต่างลายนิ้วมือทางเสียงที่ i เมื่อคำนวณค่าความคล้ายคลึงกันและค่าความแตกต่างแล้ว ทั้งนี้เพื่อวิเคราะห์ความคล้ายกันของทั้งเพลงจึงนิยามค่าสหสัมพันธ์ระหว่างสองเพลง (Correlation: c) ขึ้น โดยค่าสหสัมพันธ์ระหว่างเพลงสองเพลงถูกกำหนดให้เป็นเซตของข้อมูลที่มีสมาชิกเป็นคู่ลำดับของค่าความคล้ายคลึงกันและค่าความแตกต่าง ดังสมการที่ (8)

$$C_{A,B} = \left\{(\sigma_{11}, \delta_{11}), (\sigma_{12}, \delta_{12}), \ldots, (\sigma_{(n-1)(n-1)}, \delta_{(n-1)(n-1)})\right\} \qquad (8)$$

โดยที่ $C_{A,B}$ มีจำนวนสมาชิกเป็น $(n-1)^2$ เมื่อเปรียบเทียบเพลงที่ประกอบด้วยหน้าต่างลายนิ้วมือทางเสียงจำนวน n หน้าต่าง สมาชิก หนึ่งตัวที่มีค่าความคล้ายคลึงกันมากที่สุดโดยที่มีค่า

ความแตกต่างน้อยที่สุดด้วยจะถูกเลือกให้เป็นตัวแทนของคะแนนความคล้ายกันของเพลง A และ B รายละเอียดดังแสดงในสมการที่ (9) (10) และ (11)

$$max(C_{A,B}) = \left\{(\sigma,\delta) \mid (\sigma,\delta),(\sigma',\delta') \in C_{A,B} \text{ and } \sigma \geq \sigma'\right\} \quad (9)$$

$$min(C_{A,B}) = \left\{(\sigma,\delta) \mid (\sigma,\delta),(\sigma',\delta') \in C_{A,B} \text{ and } \delta \leq \delta'\right\} \quad (10)$$

$$similarity(A - B) \in min(max(C_{A,B})) \quad (11)$$

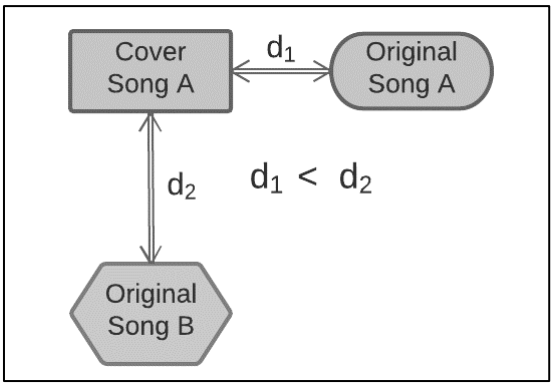
ตัวอย่างในการคำนวณค่านัยสำคัญทางสถิติที่เกี่ยวข้องกับฟังก์ชันความสัมพันธ์ (Relation function) เพื่อหาค่าความคล้ายกันของเพลงดังแสดงในภาพ 11

$$C_{A,B} = \left\{(20,5),(28,1),(28,5),(28,1)\right\}$$
$$max(C_{A,B}) = \left\{(28,1),(28,5),(28,1)\right\}$$
$$min(max(C_{A,B})) = \left\{(28,1),(28,1)\right\}$$
$$similarity(A - B) = (28,1)$$

ภาพ 11 ตัวอย่างในการคำนวณค่านัยสำคัญทางสถิติที่เกี่ยวของ

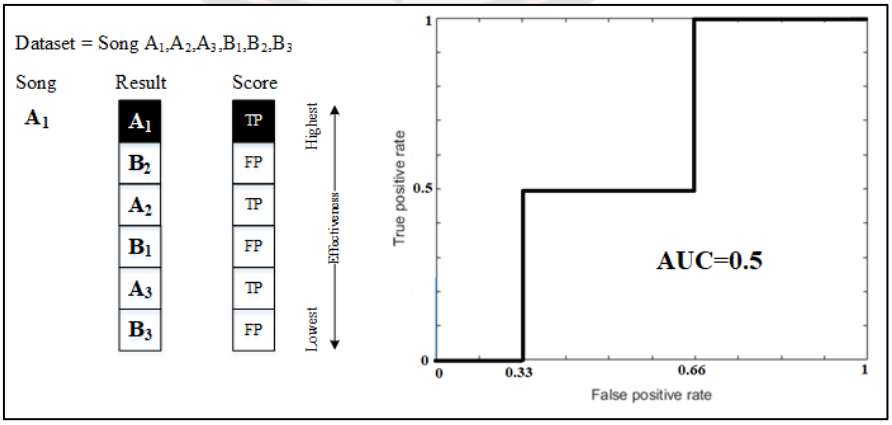
3.5 การประเมินประสิทธิภาพของวิธีการที่นำเสนอ

คะแนนความคล้ายกันระหว่างเพลงต้นฉบับและเพลงคัฟเวอร์ของเพลงเดียวกันจะมีค่ามากกว่าคะแนนความคล้ายกันระหว่างเพลงคัฟเวอร์นั้นและเพลงต้นฉบับอื่นที่ไม่ใช่เพลงๆนั้น ซึ่งนั้นก็หมายความว่าในกรณีเพลงเดียวกันแต่คนละเวอร์ชันจะมีระยะห่างระหว่างกันน้อยกว่ากรณีที่เป็นคนละเพลง ซึ่งสอดคล้องกับตัวอย่างในภาพ 12

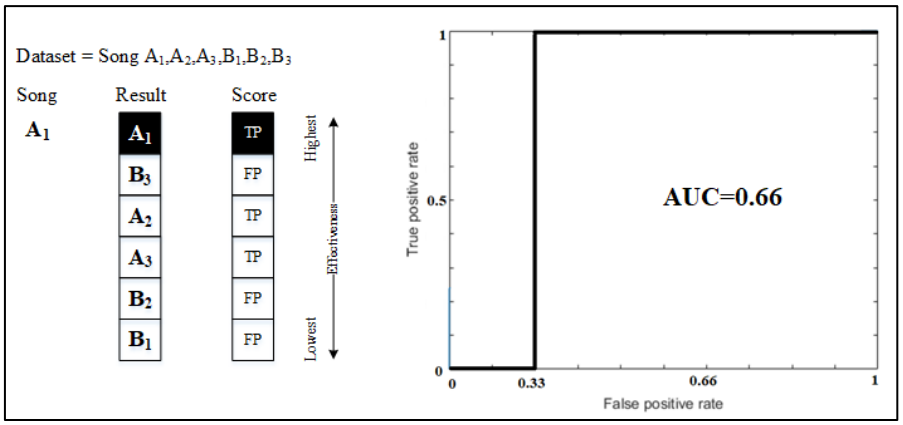


ภาพ 12 การเปรียบเทียบข้อมูลเสียงเพื่อระบุเพลงต้นฉบับจากเพลงคัฟเวอร์

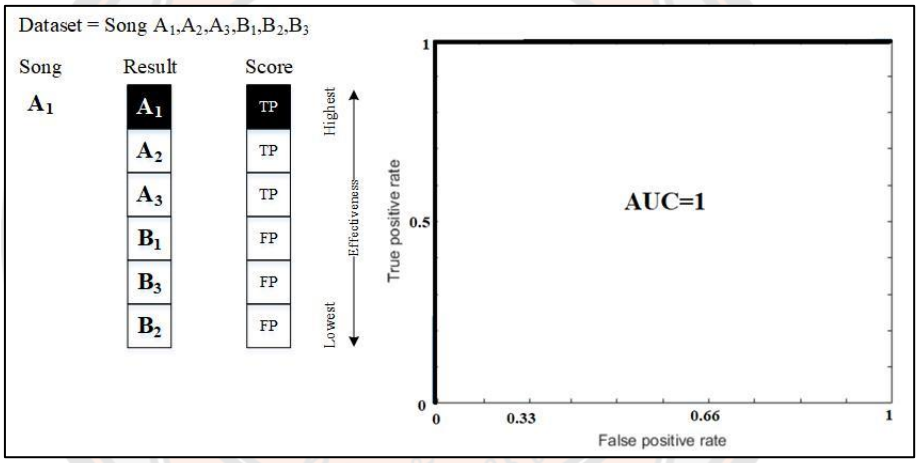
การประเมินประสิทธิภาพของวิธีการที่นำเสนอทำได้โดยการเปรียบเทียบข้อมูลขาเข้า (Input) กับข้อมูลทุกตัวในกลุ่มตัวอย่าง ซึ่งการเปรียบเทียบแต่ละครั้งจะมีการคำนวณค่าคะแนนความคล้ายกันของเพลงเพื่อจัดอันดับเพลงที่มีคะแนนจากมากไปหาน้อย ดังนั้นเพลงที่มีความคล้ายกับเพลงที่เป็นข้อมูลขาเข้ามากกว่าจะถูกจัดให้อยู่ในลำดับที่ดีกว่า อันดับความคล้ายของเพลง (Similarity rank) ที่ได้มาจะถูกนำไปสร้างกราฟเปรียบเทียบระหว่างอัตราผลบวกจริง (True positive rate) และอัตราผลบวกเท็จ (False positive rate) โดยที่เส้นกราฟจะขยับไปตามแนวแกน X เมื่อเพลงที่อยู่ในลำดับดังกล่าวเป็นเพลงต้นฉบับเพลงอื่นที่ไม่ถูกต้อง (False positive) และกราฟจะขยับไปตามแนวแกน Y เมื่อเพลงที่อยู่ในลำดับดังกล่าวเป็นเพลงต้นฉบับที่ถูกต้อง (True positive) ค่าของทั้งสองแกนอยู่ระหว่างศูนย์ถึงหนึ่งและประสิทธิภาพของการการจัดกลุ่มแปรผันตรงกับพื้นที่ใต้กราฟที่ได้ การจัดกลุ่มที่ถูกต้องสมบูรณ์จะมีพื้นที่ใต้กราฟเท่ากับหนึ่ง ตัวอย่างอันดับความคล้ายกันของเพลงและกราฟที่ถูกสร้างขึ้น รวมทั้งพื้นที่ใต้กราฟที่คำนวณได้ดังแสดงในภาพ 13 ถึง 16 ตามลำดับ



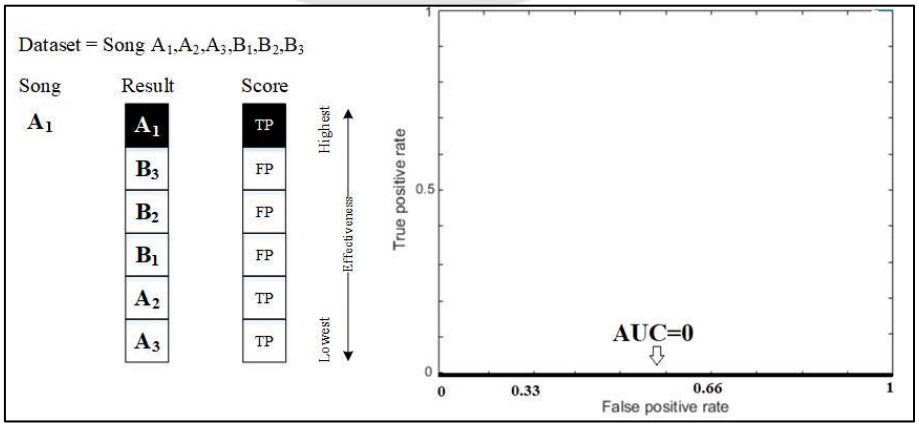
ภาพ 13 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0.5



ภาพ 14 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0.66



ภาพ 15 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 1



ภาพ 16 อันดับความคล้ายกันและกราฟที่ถูกสร้างขึ้นของเพลงที่มี AUC เป็น 0

บทที่ 4
ผลการวิจัย

4.1 ผลการวิจัย

การทดลองประยุกต์ใช้วิธีการหาความคล้ายกันของเพลงเพื่อระบุเพลงต้นฉบับจากเพลงคัฟเวอร์ ข้อมูลที่ใช้สำหรับการทดลองเป็นไฟล์เสียงจำนวน 120 ข้อมูลตัวอย่าง ซึ่งประกอบด้วย 1) เพลงเวอร์ชันต้นฉบับ 20 ข้อมูลตัวอย่าง โดยเพลงที่ถูกเลือกเป็นเพลงที่มีจำนวนการถูกคัฟเวอร์จำนวนมากซึ่งประกอบด้วยเพลงสากลจำนวน 15 เพลงและเพลงเกาหลีจำนวนห้าเพลง และ 2) เพลงเวอร์ชันคัฟเวอร์จำนวนห้าข้อมูลตัวอย่างต่อหนึ่งเพลงต้นฉบับ รวมเป็น 100 ข้อมูลตัวอย่าง ดังแสดงในตารางที่ 1

ตาราง 1 แสดงรายละเอียดของเพลงที่เป็นกลุ่มข้อมูลตัวอย่าง

Song No	Song Tittle	Artist	Genre
1	All For You	Jung Eun Ji & Seo In Guk	Pop
2	(They Long to Be) Close to You	Carpenters	
3	Counting Stars	OneRepublic	
4	Love Yourself	Justin Bieber	
5	Good day	IU	
6	I (feat. Verbal Jint)	Taeyeon	
7	Dream	Suzy &BAEKHYUN	Soul / Jazz
8	If I Ain't Got You	Alicia Keys	
9	Just The Two Of Us	Bill Withers	
10	Thinking Out Loud	Ed Sheeran	
11	Ordinary People	John Legend	
12	Don't Know Why	Norah Jones	
13	With or Without You	U2	Rock
14	Tears In Heaven	Eric Clapton	
15	Zombie	The Cranberries	
16	Still Feel Like Your Man	John Mayer	
17	Loser	Bigbang	R&B / Hip-Hop / EDM
18	That's What I Like	Bruno Mars	
19	Habits (Stay High)	Tove Lo	
20	Same Old Love	Selena Gomez	

สำหรับกระบวนการเตรียมข้อมูลนั้น ข้อมูลตัวอย่างที่เป็นไฟล์เสียงถูกนำมาสร้างเป็นเฟรมของรูปคลื่นสัญญาณความยาว 0.3 วินาที จากนั้นนำไปสร้างสเปคโตแกรมที่ช่วงความถี่ 40 ถึง 1200 Hz ซึ่งเป็นช่วงความถี่ของเสียงร้องต่ำที่สุดของผู้ชายถึงเสียงร้องสูงสุดของผู้หญิงในเพลง (Vocal frequency range) [11] ขั้นตอนถัดมานำเฟรมของสเปคโตแกรมทีละแปดเฟรมมาสร้างเป็นภาพไบนารี และขั้นตอนสุดท้ายสร้างหน้าต่างของลายนิ้วมือทางเสียงจากภาพไบนารีดังกล่าว

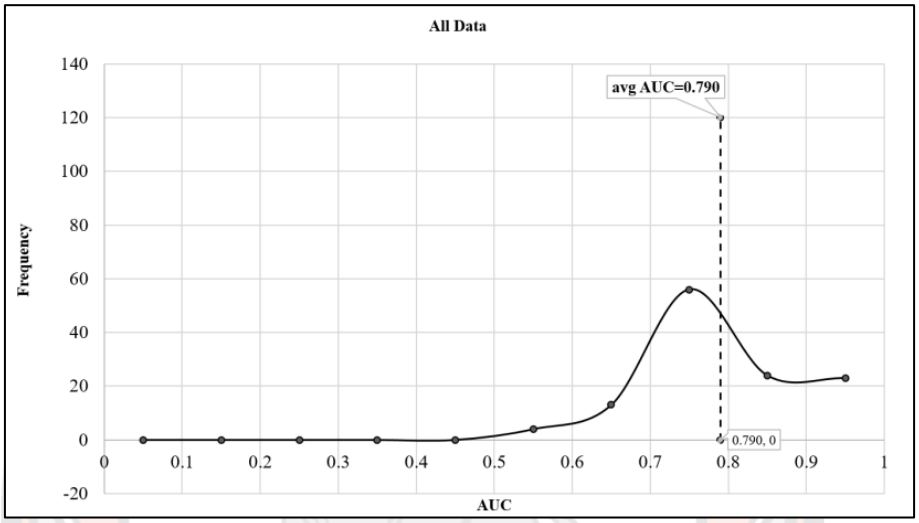
พื้นที่ใต้กราฟเป็นตัวบ่งชี้สำคัญต่อประสิทธิภาพของวิธีการที่นำเสนอ ค่าเฉลี่ยของพื้นที่ใต้กราฟ (Average AUC) โดยรวมของทั้ง 20 เพลงในกลุ่มตัวอย่างอยู่ที่ 0.790 และค่าเฉลี่ยของพื้นที่ใต้กราฟของแต่ละเพลงดังแสดงไว้ในตารางที่ 2

ตาราง 2 แสดงค่าเฉลี่ยของพื้นที่ใต้กราฟของ 20 เพลงในกลุ่มตัวอย่าง

#Song	V1	V2	V3	V4	V5	V6	Avg AUC
1	0.779	0.737	0.818	0.626	0.804	0.812	0.763
2	0.756	0.725	0.796	0.651	0.753	0.743	0.737
3	0.746	0.723	0.675	0.751	0.791	0.732	0.736
4	0.975	0.963	0.904	0.967	0.940	0.988	0.956
5	0.967	0.981	0.784	0.923	0.916	0.930	0.917
6	0.991	0.995	0.993	0.974	0.995	0.993	0.990
7	0.789	0.807	0.654	0.819	0.737	0.547	0.726
8	0.707	0.739	0.747	0.701	0.793	0.725	0.735
9	0.893	0.695	0.896	0.781	0.696	0.935	0.816
10	0.746	0.746	0.807	0.795	0.754	0.888	0.789
11	0.911	0.826	0.904	0.814	0.791	0.704	0.825
12	0.791	0.626	0.715	0.701	0.774	0.702	0.718
13	0.717	0.678	0.690	0.718	0.802	0.774	0.730
14	0.870	0.830	0.754	0.721	0.589	0.812	0.763
15	0.826	0.893	0.805	0.828	0.821	0.637	0.802
16	0.733	0.720	0.589	0.696	0.703	0.547	0.665
17	0.781	0.833	0.774	0.746	0.720	0.699	0.759
18	0.739	0.781	0.809	0.788	0.795	0.763	0.779
19	0.754	0.746	0.772	0.668	0.712	0.782	0.739
20	0.925	0.904	0.809	0.916	0.868	0.707	0.855

จากตารางที่ 1 แต่ละแถวของตารางแสดงพื้นที่ใต้กราฟของเพลงหนึ่งเพลง ซึ่งประกอบไปด้วยหกเวอร์ชัน โดยที่ V1 เป็นเวอร์ชันต้นฉบับ(Original version) ของเพลงดังกล่าว V2-V6 เป็น

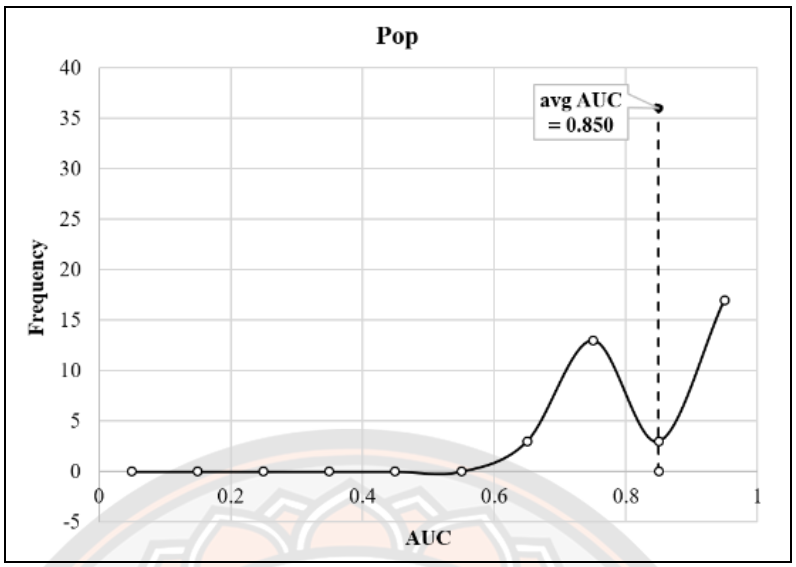
เวอร์ชันคัฟเวอร์ (Cover version) ของเพลงดังกล่าวจำนวนห้าเวอร์ชันตามลำดับ และคอลัมน์สุดท้ายของตารางเป็นค่าเฉลี่ยของพื้นที่ใต้กราฟของเพลงนั้นๆ เมื่อนำค่าเฉลี่ยของพื้นที่ใต้กราฟของแต่ละเพลงมานับความถี่ที่เกิดขึ้นโดยการแบ่งช่วงกว้าง (Range) ของค่าเฉลี่ยของพื้นที่ใต้กราฟออกเป็นช่วงๆ ดังนี้ 0-0.1, 0.1-0.2, 0.3-0.4 ,..., 0.9-1.0 ตามลำดับจะได้กราฟแสดงการแจกแจงความถี่ของค่าเฉลี่ยของพื้นที่ใต้กราฟของทุกเพลงในกลุ่มตัวอย่างดังแสดงในภาพ 17



ภาพ 17 กราฟแจกแจงความถี่ของค่าเฉลี่ยพื้นที่ใต้กราฟของทั้ง 20 เพลง

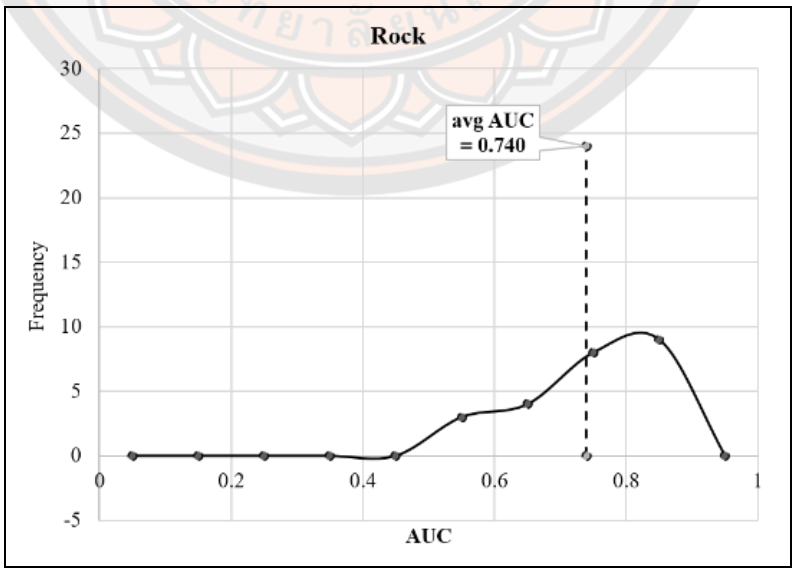
กลุ่มข้อมูลเพลงตัวอย่างถูกแบ่งเป็นสี่กลุ่มเมื่อตามประเภทของเพลง (Genre) ดังนี้ กลุ่มที่หนึ่งเพลงป๊อป (Pop) จำนวน 36 ข้อมูลตัวอย่าง กลุ่มที่สองเพลงร็อค (Rock) จำนวน 36 ข้อมูลตัวอย่าง กลุ่มที่สามเพลงโซลหรือแจ๊ส (Soul/Jazz) จำนวน 24 ข้อมูลตัวอย่าง และสุดท้ายกลุ่มที่สี่เพลงริทึมแอนด์บลูส์ (R&B) ฮิปฮอป (Hip-Hop) หรืออีดีเอ็ม (EDM) จำนวน 24 ข้อมูลตัวอย่าง เมื่อพิจารณาพื้นที่ใต้กราฟแยกตามกลุ่มจะได้กราฟแจกแจงความถี่ ดังแสดงในภาพ 18 ถึง 21

กลุ่มเพลงป๊อปมีค่าเฉลี่ยพื้นที่ใต้กราฟมากที่สุดเมื่อเทียบกับเพลงในกลุ่มอื่นเป็น 0.85 และเพลงที่มีพื้นที่ใต้กราฟสูงที่สุดมีพื้นที่ใต้กราฟถึง 0.995 ซึ่งเกือบเป็นการจัดกลุ่มแบบสมบูรณ์แบบ ดังกราฟที่แสดงในภาพ 18

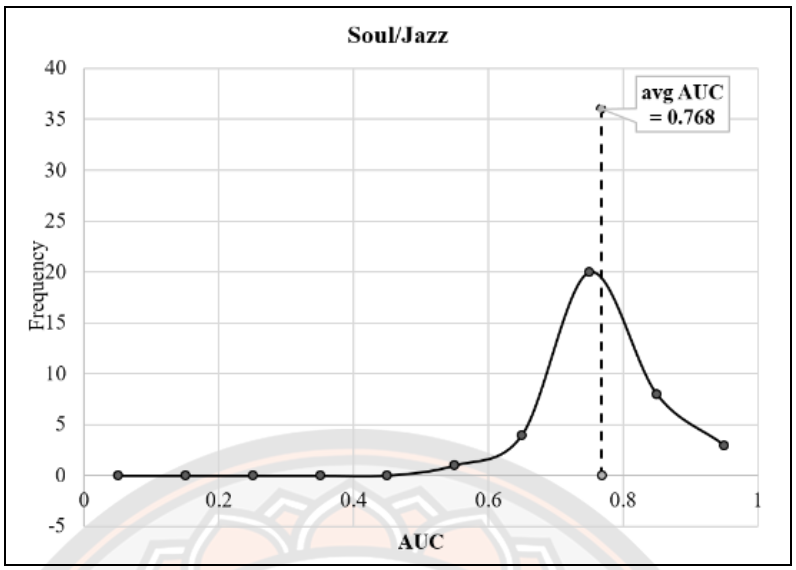


ภาพ 18 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงป๊อป

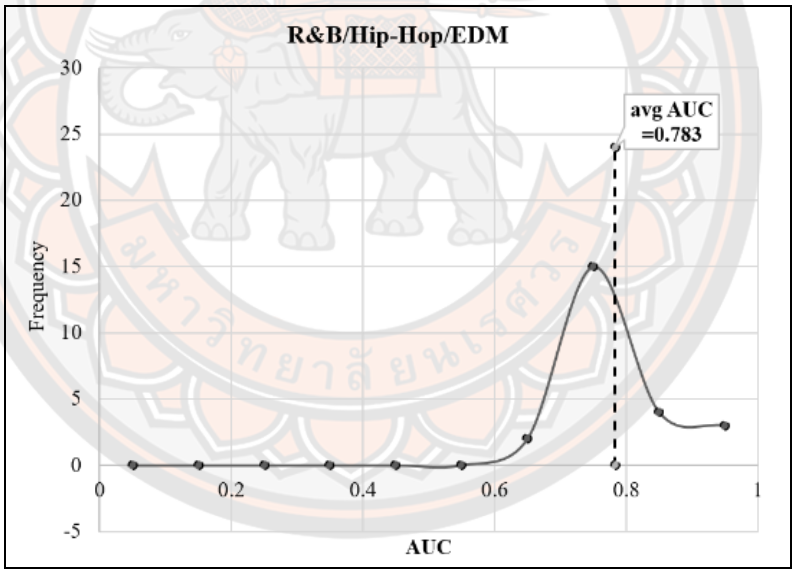
ในขณะที่อีกสามกลุ่ม กลุ่มที่สองเพลงร็อค (Rock) กลุ่มที่สามเพลงโซลหรือแจ๊ส (Soul/Jazz) และกลุ่มที่สี่เพลงริทึมแอนด์บลูส์ (R&B) ฮิปฮอป (Hip-Hop) หรืออีดีเอ็ม (EDM) มีค่าเฉลี่ยพื้นที่ใต้กราฟใกล้เคียงกันอยู่ระหว่าง 0.74-0.78 ดังแสดงในภาพ 19 ถึง 21 จากผลการทดลองดังกล่าวบ่งชี้ว่า วิธีการที่นำเสนอสามารถจัดกลุ่มเพลงเดียวกันแต่คนละเวอร์ชันให้อยู่ในกลุ่มเดียวกันได้โดยไม่ขึ้นอยู่กับประเภทของเพลง



ภาพ 19 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงร็อค

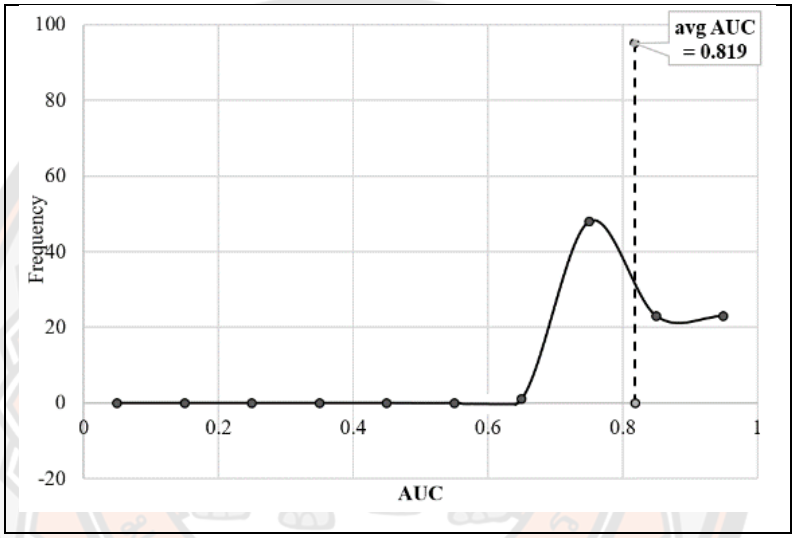


ภาพ 20 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของกลุ่มเพลงโซลหรือแจส

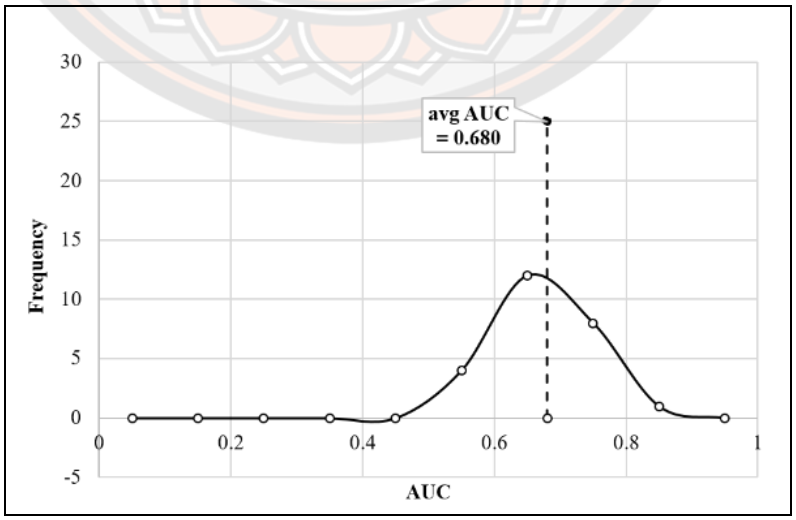


ภาพ 21 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของ
กลุ่มเพลงริทึมแอนด์บลูส ฮิปฮอป หรืออีดีเอ็ม

นอกจากการพิจารณาพื้นที่ใต้กราฟตามประเภทของเพลงแล้ว ยังมีการวิเคราะห์ผลการจัดกลุ่มโดยพิจารณาลักษณะของการบันทึกเสียงใหม่จากต้นฉบับเดิม ซึ่งแบ่งออกเป็นสองกลุ่มตามลักษณะการบรรเลงดนตรี ลักษณะแรกเป็นเวอร์ชันคัฟเวอร์ที่เพียงแค่บรรเลงซ้ำเดิมและร้องใหม่เท่านั้น ไม่ดัดแปลงการบรรเลงดนตรี ลักษณะที่สองมีการดัดแปลงการบรรเลงดนตรีเมื่อบันทึกเสียงอีกครั้งเพื่อสร้างเวอร์ชันคัฟเวอร์ ตัวอย่างเช่น การนำเพลงต้นฉบับที่เป็นเพลงป๊อปมาบรรเลงใหม่เป็นเพลงร็อคแทน ผลการทดลองพบว่า ลักษณะแรกที่ไม่ดัดแปลงการบรรเลงดนตรีมีพื้นที่ใต้กราฟมากกว่าลักษณะที่สองอย่างเห็นได้ชัด ดังแสดงในภาพ 22 และ 23 ตามลำดับ



ภาพ 22 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของเพลงที่ไม่ดัดแปลงดนตรี



ภาพ 23 กราฟแจกแจงความถี่ของพื้นที่ใต้กราฟของเพลงที่ดัดแปลงดนตรี

บทที่ 5
บทสรุป

5.1 สรุปผลการวิจัย

ข้อสรุปที่ได้จากการวิจัยบ่งชี้ว่า การหาความคล้ายกันของเพลงโดยใช้ฟังก์ชันความสัมพันธ์ (Relation function) สามารถระบุเพลงต้นฉบับจากเพลงคัฟเวอร์ได้อย่างมีประสิทธิภาพครอบคลุมเพลงทุกประเภท โดยมีพื้นที่ใต้กราฟอยู่ระหว่าง 0.74-0.85 ทั้งนี้เพลงในกลุ่มตัวอย่างบางส่วนมีผลการทดลองที่มีนัยสำคัญผู้วิจัยจึงนำมาวิเคราะห์เป็นรายกรณีดังต่อไปนี้

5.1.1 กลุ่มเพลงที่มีความแม่นยำในการทำนายสูง

เมื่อนำเพลงกลุ่มตัวอย่างมีความแม่นยำในการทำนายสูงมาวิเคราะห์เพื่อเป็นกรณีศึกษาพบว่ากลุ่มเพลง **คัฟเวอร์** ของเพลงดังกล่าวมีลักษณะการเรียบเรียงดนตรีและเนื้อร้องยังคงเดิมเพียงแต่เปลี่ยนผู้ร้องเท่านั้น เช่น เพลงลำดับที่ 4 Love Yourself โดย Justin Bieber ที่อยู่ในกลุ่มตัวอย่างมีความแม่นยำในการทำนายสูงที่สุดเฉลี่ยถึง 0.956 เป็นต้น

5.1.2 กลุ่มเพลงที่มีสัญญานรบกวน (Noise)

เมื่อพิจารณาเพลง**คัฟเวอร์**ที่มีสัญญาณรบกวนในลักษณะต่างๆ กรณีแรกเมื่อมีเสียงพูดแทรกระหว่างเพลง ตัวอย่างเช่น เพลงลำดับที่ 9 Just The Two Of Us เวอร์ชัน 3 มีค่า AUC เฉลี่ย 0.896 และกรณีที่ 2 เมื่อมีสัญญาณรบกวนในกระบวนการการบันทึกเสียง เช่น เสียงลม เสียงแก้วน้ำ เป็นต้น เพลงกลุ่มดังกล่าวมีค่า AUC เฉลี่ยอยู่ระหว่าง 0.75-0.91 ซึ่งบ่งชี้ว่าสัญญาณรบกวนไม่ส่งผลต่อความแม่นยำในการทำนายโดยวิธีการที่นำเสนอ

5.1.3 กลุ่มเพลงที่ถูก**คัฟเวอร์**โดยผู้ร้องต่างเพศกับเพลงต้นฉบับ

เมื่อพิจารณาเพลงที่ถูก**คัฟเวอร์**โดยผู้ร้องต่างเพศกับเพลงต้นฉบับ ตัวอย่างเช่น เพลงลำดับที่ 3 Counting the star ซึ่งเพลงต้นฉบับมีผู้ร้องเป็นผู้ชาย เมื่อถูกนำมา**คัฟเวอร์**โดยผู้ร้องที่เป็นเพศหญิงมีค่า AUC เฉลี่ยที่ 0.67 ซึ่งต่ำกว่าเพลงที่ถูกนำมา**คัฟเวอร์**โดยผู้ชายเหมือนกับต้นฉบับ หรือเพลงลำดับที่ 9 Just The Two Of Us ที่ต้นฉบับร้องโดยเพศชาย เมื่อพิจารณาเวอร์ชัน 2 และ 4 ซึ่งถูกร้องโดยผู้ร้องเพศหญิงมี AOC เฉลี่ยอยู่ที่ประมาณ 0.69-0.78 ในขณะที่เพลงนี้หากถูกร้องโดยเพศชายเหมือนกับต้นฉบับจะมีค่า AOC เฉลี่ยตั้งแต่ 0.78-0.93 หลังจากนำมาวิเคราะห์และทบทวนกระบวนการทดลองพบว่า เนื่องด้วยงานวิจัยในขั้นตอนการสร้างสเปคโตแกรมเลือกช่วงความถี่ 40 ถึง 1200 Hz ซึ่งเป็นช่วงความถี่เสียงร้องต่ำที่สุดถึงเสียงร้องสูงสุด (Vocal frequency range) [11] ของมนุษย์โดยภาพรวมทั้งเพศชายและเพศหญิง แต่หากแบ่งเสียงร้องตามเพศ ช่วงความถี่เสียงร้องผู้ชายและผู้หญิงมีความแตกต่างกัน โดยในเสียงร้องของเพศหญิงมีความถี่สูงกว่าอยู่ระหว่าง 160-

1200 Hz และเพศชายมีความถี่ต่ำกว่าอยู่ระหว่าง 40-500 Hz ด้วยเหตุนี้จึงส่งผลให้การแยกแยะเพลงเดียวกันที่ถูก**คัฟเวอร์**โดยผู้ร้องต่างเพศด้วยวิธีการที่นำเสนอมีความแม่นยำลดลง

5.1.4 กลุ่มเพลงที่ความเร็วของจังหวะในการเล่นเปลี่ยนแปลง แต่เนื้อร้องคงเดิม

จากการวิเคราะห์ผลการทดลองในกลุ่มเพลงที่มีความเร็วของจังหวะเปลี่ยนแปลงไปจากต้นฉบับ ผู้วิจัยพบข้อบ่งชี้ที่ว่า ความเร็วของจังหวะในการเล่นมีผลต่อความแม่นยำในการทำนาย เพราะถึงแม้เนื้อร้องจะคงเดิมแต่ความเร็วของจังหวะในการเล่นที่เปลี่ยนแปลงไปนั้นส่งผลต่อความเร็วในการออกเสียงเนื้อร้อง เช่นกรณีของเพลงลำดับที่ 16 Still Feel Like Your Man โดย John Mayer มีค่า AOC เฉลี่ยทุกเวอร์ชันเป็น 0.665 โดยเวอร์ชัน 3 และ 6 มีค่า AOC ต่ำอยู่ในช่วง 0.5 เท่านั้น ซึ่งเมื่อนำมาวิเคราะห์แยกแต่ละเวอร์ชันด้วยการฟังพบว่า เพลงลำดับที่ 16 เวอร์ชัน 3 เป็นเพลง**คัฟเวอร์** ที่ถูกสร้างขึ้นจากการบรรเลงดนตรีใหม่โดยจังหวะในการเล่นถูกทำให้เร็วขึ้น ในขณะที่เวอร์ชัน 6 จังหวะในการเล่นถูกทำให้ช้าลง

5.1.5 กลุ่มเพลงที่มีการเปลี่ยนเนื้อร้อง เพิ่มเนื้อร้องหรือท่อนอิมโพรไวส์ (improvise)

กรณีดังกล่าวส่วนใหญ่เกิดขึ้นในประเภทกลุ่มเพลงโซลหรือแจ๊ส ซึ่งความแม่นยำของการทำนายขึ้นอยู่กับความมากน้อยในการเพิ่มหรือเปลี่ยนเนื้อร้อง และความยาวของท่อนอิมโพรไวส์ (Improvise) ที่ถูกใส่เพิ่มลงไปในเพลง เช่น กรณีตัวอย่างเพลงลำดับที่ 12 Don’t know why เวอร์ชัน 2 ซึ่งมีการเพิ่มท่อนอิมโพรไวส์ (improvise) และเพลงลำดับที่ 9 Just two of us เวอร์ชัน 5 ซึ่งมีการเพิ่มเนื้อร้องและท่อนอิมโพรไวส์ (improvise) เป็นต้น

5.2 ข้อเสนอแนะ

การจัดกลุ่มเพลงมีประสิทธิภาพในกรณีที่มีการบรรเลงดนตรีใหม่ในรูปแบบคล้ายกับต้นฉบับเดิม ในขณะที่การบรรเลงโดยเปลี่ยนรูปแบบดนตรีที่ส่งผลต่อความเร็วและลักษณะการออกเสียงของการร้องส่งผลให้ความถูกต้องของการจัดกลุ่มน้อยลง ซึ่งข้อจำกัดดังกล่าวผู้วิจัยจะนำมาศึกษาและปรับปรุงวิธีการต่อไป รวมทั้งการศึกษา Deep learning ที่เกี่ยวข้องกับการรู้จำข้อมูลทางดนตรีเพื่อนำมาประยุกต์ใช้กับงานวิจัยในอนาคตต่อไป

บรรณานุกรม

1. Downie, J.S., *Music information retrieval.* Annual Review of Information Science and Technology, 2003. **37**(1): p. 295--340.
2. Tao, L. and M. Ogihara, *Toward intelligent music information retrieval.* Multimedia, IEEE Transactions on, 2006. **8**(3): p. 564-574.
3. Corrêa, D.C. and F.A. Rodrigues, *A survey on symbolic data-based music genre classification.* Expert Systems with Applications, 2016. **60**: p. 190-210.
4. Ouali, C., P. Dumouchel, and V. Gupta, *A spectrogram-based audio fingerprinting system for content-based copy detection.* Multimedia Tools Appl., 2016. **75**(15): p. 9145-9165.
5. Perrot, D.a.G., R. O, *Scanning the dial: An exploration of factors in the identification of musical style.* Proc. Int. Conf. on Music Perception and Cognition, 1999.
6. Tzanetakis, G. and P. Cook, *Musical genre classification of audio signals.* Speech and Audio Processing, IEEE Transactions on, 2002. **10**(5): p. 293-302.
7. Chandrasekhar, V.a.S., Matt and Ross, David, *Survey and Evaluation of Audio Fingerprinting Schemes for Mobile Query-by-Example Applications.* 2011.
8. Ouali, C., P. Dumouchel, and V. Gupta. *Efficient spectrogram-based binary image feature for audio copy detection*. in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015.
9. Limited, S.E. *Music Discovery, Charts & Song Lyrics*. 2017 [cited 2017 19 July]; Available from: https://www.shazam.com/.
10. Inc, S. *SoundHound*. [cited 2017 19 July]; Available from: https://soundhound.com/.
11. Snow, W.B., *Audible Frequency Ranges of Music, Speech and Noise.* The Journal of the Acoustical Society of America, 1931. **3**(1A): p. 10-10.
12. Brookes, M. *VOICEBOX: Speech Processing Toolbox for MATLAB*. Available from: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.
13. The MathWorks, I. *MATLAB*. Available from: https://www.mathworks.com/products/matlab.html.