โครงการ

การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

Multimodal Retrieval of Thai Music from Singing and Humming

ชื่อนิสิต นางสาวขวัญชนก ศิลปพิบูลย์ เลขประจำตัว 5933606023

นางสาวณีญาภรณ์ เพชรปาน เลขประจำตัว 5933624323

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์

สาขาวิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2562

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

นางสาวขวัญชนก ศิลปพิบูลย์
นางสาวณีญาภรณ์ เพชรปาน

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Multimodal Retrieval of Thai Music from Singing and Humming

Khwanchanok Sillapapibul
Neeyaporn Petchapan

A Project Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science Program in Computer Science Department of Mathematics and Computer Science Faculty of Science
Chulalongkorn University
Academic Year 2019

Copyright of Chulalongkorn University

หัวข้อโครงงาน	การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม
โดย	นางสาวขวัญชนก ศิลปพิบูลย์
	นางสาวณีญาภรณ์ เพชรปาน
สาขาวิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาโครงงานหลัก	ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร
อาจารย์ที่ปรึกษาโครงงานร่วม	ผู้ช่วยศาสตราจารย์ ดร.ศุภกานต์ พิมลธเรศ

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาบัณฑิตในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

.....

(ศาสตราจารย์ ดร.กฤษณะ เนียมมณี) หัวหน้าภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงงาน

.....

(ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร) อาจารย์ที่ปรึกษาโครงงานหลัก

.....

(ผู้ช่วยศาสตราจารย์ ดร.ศุภกานต์ พิมลธเรศ) อาจารย์ที่ปรึกษาโครงงานร่วม

.....

(รองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์) กรรมการ

.....

(ผู้ช่วยศาสตราจารย์ ดร. กรุง สินอภิรมย์สราญ) กรรมการ

นางสาวขวัญชนก ศิลปพิบูลย์, นางสาวณีญาภรณ์ เพชรปาน: การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม. (Multimodal Retrieval of Thai Music from Singing and Humming) อ.ที่ปรึกษาโครงงานหลัก : ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร, อ.ที่ปรึกษาโครงงงานร่วม : ผู้ช่วยศาสตราจารย์ ดร.ศุภกานต์ พิมลธเรศ, 57 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัมโดยพิจารณาความเหมือนของลักษณะของเสียงที่สกัดได้จากข้อมูลนำเข้า ได้แก่ เสียงร้องหรือเสียงฮัมขณะที่ผลลัพธ์เป็นชุดเพลงต้นฉบับที่ค้นคืนและเปอร์เซนต์ระดับความมั่นใจ เพลงไทยห้าประเภท ได้แก่ พอป ร็อค แจ๊ส ฮิปฮอป และลูกทุ่งประเภทละ 50 เพลง ถูกเลือกมาวิเคราะห์ในการศึกษานี้ ลักษณะของเสียงเจ็ดลักษณะ ได้แก่ เซนทรอยด์ของสเปกตรัม โรลออฟของสเปกตรัม อัตราการข้ามศูนย์ ความกว้างแถบของสเปกตรัม สัมประสิทธิ์เซปสตรัมความถี่เมล การแปลงฟูเรียร์ช่วงสั้นโครมา และรากกำลังสองเฉลี่ย ได้ถูกนำมาใช้วัดความเหมือนระหว่างเสียงร้องหรือเสียงฮัมเทียบกับเพลงต้นฉบับ นอกจากนี้เทคนิคไดนามิกไทม์วอร์ปปิง ได้ถูกนำมาใช้เพื่อศึกษาผลกระทบของการประสานจังหวะเพลงในกระบวนการการเปรียบเทียบ ในท้ายที่สุดชุดเพลงต้นฉบับถูกค้นคืนภายใต้การพิจารณาค่าความเหมือนร่วมกับเกณฑ์การยอมรับและเกณฑ์การปฏิเสธที่ให้เปอร์เซนต์ของระดับความมั่นใจ

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต ขวัญชนก ศิลปพิบูลย์

ลายมือชื่อนิสิต ณีญาภรณ์ เพชรปาน

สาขาวิชา วิทยาการคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาโครงงานหลัก ศศิภา พันธุวดีธร

ปีการศึกษา 2562 ลายมือชื่อ อ.ที่ปรึกษาโครงงานร่วม ศุภกานต์ พิมลธเรศ

5933606023, 5933624323: MAJOR COMPUTER SCIENCE
KEYWORDS: THAI MUSIC RETRIEVAL/ SPECTRAL CENTROID (SPC)/ SPECTRAL ROLL-OFF (SRO)/ ZERO CROSSING RATE (ZCR)/ SPECTRAL BANDWIDTH (SBW)/ CHROMA SHORT-TIME FOURIER TRANSFORM (CHM)/ ROOT MEAN SQUARE (RMS)/ EUCLIDEAN DISTANCE/ DYNAMIC TIME WARPING (DTW)

KHWANCHANOK SILLAPAPIBUL, NEEYAPORN PETCHAPAN: MULTIMODAL RETRIEVAL OF THAI MUSIC FROM SINGING AND HUMMING. ADVISOR : ASST PROF. SASIPA PANTHUWADEETHORN, CO-ADVISOR : ASST PROF. SUPHAKANT PHIMOLTARES, Ph.D., 57 pp.

The purpose of this research is to develop multimodal retrieval of Thai music from singing and humming by considering the similarity of audio features that were extracted from input data: singing or humming voice, whereas the output is a set of retrieved original songs with percentage of confidence level. Five types of Thai songs, namely, pop, rock, jazz, hip-hop and folk songs, each of which contains 50 songs, were selected to be analyzed in this study. Seven audio features, i.e., spectral centroid (SPC), spectral roll-off (SRO), zero crossing rate (ZCR), spectral bandwidth (SBW), chroma short-time fourier transform (CHM) and root mean square (RMS), were applied to measure similarity between singing or humming voice and original song. Additionally, the dynamic time warping technique was employed to study the impact of music synchronization in the matching process. Finally, the set of original song was retrieved under the consideration of similarity value together with acceptance and rejection criteria providing the percentage of confidence level.

Department : Mathematics and Computer Science	Student's Signature ขวัญชนก ศิลปพิบูลย์
	Student's Signature นีญาภรณ์ เพชรปาน
Field of Study : Computer Science	Advisor's Signature Sasipa Panthuwadeethorn
Academic Year : 2019	Co-advisor's Signature Suphakant Phimoltares

กิตติกรรมประกาศ

โครงการวิจัยเรื่อง การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม ฉบับนี้ สำเร็จลุล่วงด้วยดีโดยได้รับความอนุเคราะห์อย่างดียิ่งจาก ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร อาจารย์ที่ปรึกษาโครงงานหลัก และผู้ช่วยศาสตราจารย์ ดร.ศุภกานต์ พิมลธเรศ อาจารย์ที่ปรึกษาโครงงานร่วม ที่สละเวลาให้คำปรึกษา แนะนำแนวทางอันเป็นประโยชน์ในโครงการวิจัย ตรวจทานข้อผิดพลาดด้วยความเอาใจใส่เป็นอย่างยิ่ง และให้กำลังใจตลอดการดำเนินโครงการวิจัยฉบับนี้

ขอขอบพระคุณรองศาสตราจารย์ ดร. รัชลิดา ลิปิกรณ์ และผู้ช่วยศาสตราจารย์ ดร. กรุง สินอภิรมย์สราญ กรรมการสอบงานวิจัย ซึ่งช่วยให้คำแนะนำให้โครงการวิจัยฉบับนี้มีความสมบูรณ์มากขึ้น

ขอขอบพระคุณภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ที่ได้อำนวยความสะดวกในด้านสถานที่สำหรับค้นคว้าและจัดทำโครงการวิจัย และให้ความอนุเคราะห์งบประมาณค่าใช้จ่ายในการดำเนินโครงการวิจัยนี้ รวมถึงเพื่อน ๆ ทุกท่านที่คอยช่วยเหลือและมอบกำลังใจให้เสมอตลอดการดำเนินโครงการวิจัยฉบับนี้

สุดท้ายนี้ผู้วิจัยขอขอบพระคุณบิดา มารดา และครอบครัว ที่สนับสนุนและให้กำลังใจผู้วิจัยเสมอมาตลอดการจัดทำโครงการวิจัยนี้จนสำเร็จลุล่วง

สารบัญ

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและเหตุผลการวิจัย 1
1.2 วัตถุประสงค์ของงานวิจัย 1
1.3 ขอบเขตการวิจัย 1
1.4 ขั้นตอนการวิจัย 2
1.5 ประโยชน์ที่คาดว่าจะได้รับ 4
1.6 โครงสร้างของรายงาน 4
บทที่ 2 หลักการและงานวิจัยที่เกี่ยวข้อง 5
2.1 หลักการและทฤษฎีที่เกี่ยวข้อง 5
2.2 งานวิจัยที่เกี่ยวข้อง 12
บทที่ 3 วิธีการวิจัย 14
3.1 วิธีการเก็บข้อมูล 14
3.2 วิธีการเตรียมข้อมูลก่อนนำมาใช้ 14
3.3 การค้นคืนเพลง 15
3.4 การวัดผลการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม 18
บทที่ 4 ผลการวิจัย 22
4.1 การตั้งค่าการทดลอง 22
4.2 ผลการวิจัย 23
4.3 การอภิปรายผล 32

บทที่ 5 สรุปผลการดำเนินงานและข้อเสนอแนะ 35
5.1 สรุปผลการดำเนินงาน 35
5.2 ข้อเสนอแนะ 35
รายการอ้างอิง 37
ภาคผนวก ก แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal ปีการศึกษา 2562 42
ประวัติผู้เขียน 48

สารบัญตาราง

ตารางที่ 3.1 ค่าพารามิเตอร์ของการสกัดลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมล 16
ตารางที่ 4.1 การเปรียบเทียบความถูกต้องของการค้นคืน Top 1 และ Top 5 ของลักษณะ 7 ชนิด 23
ตารางที่ 4.2 การเปรียบเทียบความถูกต้องของการค้นคืน Top 1 และ Top 5 ของลักษณะ 7 ชนิดเมื่อมีการทำให้เป็นบรรทัดฐาน (Normalization) และตัดเพลงก่อนและหลังสกัดลักษณะ 24
ตารางที่ 4.3 การเปรียบเทียบความถูกต้องของการค้นคืน (%) เมื่อมีการตัดเสียงดนตรี ร่วมกับการทำให้เป็นบรรทัดฐาน 25
ตารางที่ 4.4 การเปรียบเทียบความถูกต้องของการค้นคืน n อันดับแรก เมื่อมีการใช้ เทคนิคไดนามิกไทม์วอร์ปปิง 26
ตารางที่ 4.5 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงจากเสียงร้อง เมื่อมีการผสานลักษณะ 27
ตารางที่ 4.6 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงจากเสียงฮัม เมื่อมีการผสานลักษณะ 27
ตารางที่ 4.7 การเปรียบเทียบผลของการระบุประเภทของเพลงในการค้นคืนเพลง 28
ตารางที่ 4.8 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงด้วยเสียงฮัมเมื่อเทียบกับเพลงทั้งหมด และเมื่อเทียบกับเพลงประเภทเดียวกัน 28
ตารางที่ 4.9 ค่าเฉลี่ยอันดับคำตอบของการค้นคืนเพลงด้วยเสียงฮัม แยกตามประเภทของเพลงและเพลงทั้งหมด 29
ตารางที่ 4.10 เกณฑ์การยอมรับและเกณฑ์การปฏิเสธของเพลงไทยประเภทต่าง ๆ 32

สารบัญภาพ

ภาพที่ 2.1 ตัวอย่างค่าลักษณะอัตราการข้ามศูนย์ของสัญญาณเสียง เพลงเจ็บจนพอ – แว่นใหญ่ 6
ภาพที่ 2.2 ตัวอย่างค่าลักษณะเซนทรอยด์ของสเปกตรัมของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่ 7
ภาพที่ 2.3 ตัวอย่างค่าลักษณะโรลออฟของสเปกตรัมของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่..... 7
ภาพที่ 2.4 ตัวอย่างค่าลักษณะความกว้างแถบของสเปกตรัมของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่... 8
ภาพที่ 2.5 ตัวอย่างสเปคโตแกรมแสดงลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่ 8
ภาพที่ 2.6 โครมาแกรมของการไล่สเกล C เมเจอร์ 9
ภาพที่ 2.7 ตัวอย่างค่าลักษณะการแปลงฟูเรียร์ช่วงสั้นโครมาของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่ .. 9
ภาพที่ 2.8 ตัวอย่างค่าลักษณะรากกำลังสองเฉลี่ยของสัญญาณเสียง เพลงเจ็บจนพอ - แว่นใหญ่..... 9
ภาพที่ 2.9 การเปรียบเทียบลำดับข้อมูลสองชุดโดยใช้เทคนิคไดนามิกไทม์วอร์ปปิงบนอนุกรมเวลา 10
ภาพที่ 2.10 การเปรียบเทียบลำดับข้อมูลสองชุดโดยใช้การวัดระยะห่างแบบยุคลิดบนอนุกรมเวลา..... 11
ภาพที่ 3.1 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากเสียงร้องและเสียงฮัม..... 16
ภาพที่ 3.2 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากการฮัมด้วยการผสานลักษณะ..... 17
ภาพที่ 3.3 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากเสียงร้องและเสียงฮัม โดยคำนึงถึงประเภทเพลง 18
ภาพที่ 4.1 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทพอป..... 30
ภาพที่ 4.2 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทร็อค..... 30
ภาพที่ 4.3 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทแจ๊ส 31
ภาพที่ 4.4 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทฮิปฮอป 31
ภาพที่ 4.5 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทลูกทุ่ง 31
ภาพที่ 4.6 ตัวอย่างคลื่นเสียงของเพลงรู้บ้างไหม - เครสเชนโด้ (Crescendo)..... 34

บทที่ 1

บทนำ

1.1 ความเป็นมาและเหตุผลการวิจัย

เสียงเพลงเป็นสิ่งใกล้ตัวที่อยู่ในชีวิตประจำวันของเรา อาจเป็นเพลงที่รู้จัก หรือเป็นเพลงที่ไม่รู้จักและบังเอิญได้ยิน ซึ่งอาจทำให้เกิดความสงสัย และต้องการค้นหาว่าเพลงนี้คือเพลงอะไร ร้องโดยศิลปินคนใด ในปัจจุบันมีแอปพลิเคชันค้นหาเพลงจากเสียง เช่น Shazam [1] , Soundhound [2] เป็นต้น โดยแอปพลิเคชันเหล่านี้ค้นหาเพลงด้วยวิธีแปลงเสียงข้อมูลรับเข้าเป็นลายพิมพ์เสียง (audio fingerprint) แล้วนำไปค้นหาลายพิมพ์เสียงที่ตรงกันในฐานข้อมูลของแอปพลิเคชัน แล้วเล่นเพลงนั้น แต่แอปพลิเคชันเหล่านี้ยังไม่สามารถค้นหาเพลงไทยได้อย่างมีประสิทธิภาพและยังไม่สามารถค้นหาเพลงจากเสียงร้องและเสียงฮัมของผู้ใช้ได้ ดังนั้นผู้พัฒนาจึงต้องการพัฒนาการค้นหาเพลงจากการร้องและฮัมที่สามารถใช้กับเพลงไทยได้

จากที่กล่าวมาข้างต้นทำให้ผู้พัฒนามีความสนใจที่จะศึกษาเกี่ยวกับการค้นคืนเพลงไทยจากการร้องและการฮัม เพื่อช่วยให้ผู้ใช้มีเครื่องมือช่วยค้นหาเพลงไทยได้สะดวกยิ่งขึ้น รวมไปถึงการแนะนำเพลงไทยที่มีความใกล้เคียงกับเพลงที่ผู้ใช้ค้นหา โดยระบบที่พัฒนาจะรับข้อมูลนำเข้าเป็นเสียงร้องหรือเสียงฮัม โดยถ้าเป็นเสียงจากการฮัมจะมีการรับข้อมูลประเภทของเพลงที่สอบถามมาช่วยในการค้นคืนเพลงไทย และนำข้อมูลเสียงร้องหรือเสียงฮัมมาสกัดลักษณะเสียงซึ่งจะได้เป็นทำนองของการร้องหรือการฮัม เพื่อใช้ค้นหาเพลงที่มีความคล้ายกันกับข้อมูลเพลงต้นฉบับในชุดข้อมูลที่เตรียมไว้แสดงเป็นผลลัพธ์

1.2 วัตถุประสงค์ของงานวิจัย

เพื่อพัฒนาเครื่องมือช่วยค้นหาเพลงไทยและแนะนำเพลงที่ใกล้เคียงกับเพลงที่ค้นหาซึ่งผู้ใช้อาจสนใจได้

1.3 ขอบเขตการวิจัย

1. เพลงไทย 5 ประเภทได้แก่ เพลงพอป เพลงร็อค เพลงลูกทุ่ง เพลงแจ๊สและเพลงฮิปฮอป ประเภทละ 50 เพลง รวม 250 เพลง โดยเพลงแต่ละเพลงจะเป็นเพลงประเภทใดประเภทหนึ่งเท่านั้น
2. ท่อนฮุค คือ ท่อนของเนื้อร้องที่ปรากฏซ้ำกันในเพลง ซึ่งทำให้ผู้ฟังจำเพลงนั้นได้ง่าย ในโครงงานนี้ความยาวของท่อนฮุคตั้งแต่ 7 ถึง 15 วินาที สำหรับท่อนฮุคที่มีความยาวมากกว่า 15 วินาทีจะตัดเฉพาะส่วนต้นของท่อนฮุคเพียงแค่ 15 วินาทีเท่านั้น

3. ข้อมูลนำเข้าเป็นเสียงร้องหรือเสียงฮัมที่เริ่มจากส่วนต้นของท่อนฮุคของเพลง โดยมีความยาวเท่ากับเพลงต้นฉบับในชุดข้อมูล และการร้องและฮัมจะทำด้วยเสียงของบุคคลเดียว โดยไม่มีเสียงดนตรีหรือเสียงอื่นประกอบ

1.4 ขั้นตอนการวิจัย

การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม มีขั้นตอนการดำเนินการดังต่อไปนี้

ก.แผนการศึกษา

1. ศึกษาการค้นคืนเพลงจากการร้องและการฮัม
2. ศึกษางานวิจัยที่เกี่ยวข้องกับการการค้นคืนเพลงจากการร้องและการฮัม
3. ศึกษาประสิทธิภาพของขั้นตอนวิธีกับลักษณะต่าง ๆ ของเพลงไทย
4. ออกแบบระบบและพัฒนาโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม
5. ทดสอบประสิทธิภาพโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม
6. ประเมินและอภิปรายผล
7. จัดทำเอกสาร

ข.ระยะเวลาที่ศึกษา

ขั้นตอน	ปี 2562					ปี 2563			
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาการค้นคืนเพลงจากการร้องและการฮัม									
2. ศึกษางานวิจัยที่เกี่ยวข้องกับการการค้นคืนเพลงจากการร้องและการฮัม									
3. ศึกษาประสิทธิภาพของขั้นตอนวิธีกับลักษณะต่าง ๆ ของเพลงไทย									
4. ออกแบบระบบและพัฒนาโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม									
5. ทดสอบประสิทธิภาพโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม									
6. ประเมินและอภิปรายผล									
7. จัดทำเอกสาร									

1.5 ประโยชน์ที่ได้รับ

ประโยชน์ที่ได้รับจากการวิจัยในครั้งนี้มีดังนี้

ก. ประโยชน์ด้านความรู้และประสบการณ์ต่อนิสิต

1. ได้ความรู้ ความเข้าใจเกี่ยวกับการพัฒนาระบบ
2. ได้ฝึกการคิดวิเคราะห์ การทำงานอย่างมีระบบ และการทำงานเป็นกลุ่ม
3. ได้รับความรู้เรื่องการทำงานที่เกี่ยวข้องกับเสียงเพลงมากขึ้น
4. ได้รับความรู้เกี่ยวกับการเขียนโปรแกรมและการพัฒนาระบบค้นคืนเพลงไทย

ข. ประโยชน์ที่ได้จากโครงงานที่พัฒนาขึ้น

1. ได้เครื่องมือที่ช่วยในการค้นคืนเพลงไทยได้อย่างมีประสิทธิภาพมากขึ้น
2. ได้เครื่องมือที่สามารถแนะนำเพลงที่ใกล้เคียงกับเพลงที่ต้องการค้นหาได้

1.6 โครงสร้างของรายงาน

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

บทที่ 3 กล่าวถึงวิธีและขั้นตอนในงานวิจัยเรื่องการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

บทที่ 4 กล่าวถึงผลการวิจัยที่ได้จากการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

บทที่ 5 กล่าวถึงการสรุปผลการดำเนินงาน ข้อเสนอแนะ ที่เกี่ยวกับการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

บทที่ 2

หลักการและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงหลักการและทฤษฎีที่เกี่ยวข้องกับการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม รวมถึงงานวิจัยที่เกี่ยวข้องกับการสกัดลักษณะของเพลง

2.1 หลักการและทฤษฎีที่เกี่ยวข้อง

2.1.1 การเตรียมข้อมูล (Data preprocessing)

ในส่วนนี้จะกล่าวถึงหลักการต่าง ๆ ที่นำมาใช้ในขั้นตอนการเตรียมข้อมูลของเสียงร้อง เสียงฮัม และเพลงต้นฉบับ

- **Open-Unmix [3]**

  เป็นตัวแบบที่พัฒนาขึ้นโดยใช้โครงข่ายประสาทเชิงลึกเพื่อแยก แหล่งที่มาต่าง ๆ ของเพลงจากข้อมูลเพลงหนึ่งไฟล์จะสามารถแยกได้สี่ไฟล์ ได้แก่ เสียงร้อง (vocals) เสียงกลอง (drums) เสียงเบส (bass) และเสียงอื่น ๆ (other)

- **การทำให้เป็นบรรทัดฐาน (Normalization)**

  การทำให้เป็นบรรทัดฐานด้วยวิธีค่าสูงสุดต่ำสุด (Min-Max Normalization)

  คือการแปลงข้อมูลให้อยู่ในช่วงใหม่ที่กำหนด โดยทั่วไปนิยมกำหนดให้ข้อมูลอยู่ในช่วง (0,1) โดยสามารถแปลงข้อมูลโดยใช้สมการ

$$z_i = \frac{x_i - min(X)}{max(X) - min(X)}$$

โดยที่ z_i คือ ค่าบรรทัดฐานของข้อมูลตัวที่ i

x_i คือ ค่าข้อมูลตัวที่ i ในชุดข้อมูล X ที่ต้องการทำให้เป็นบรรทัดฐาน

$min(X)$ คือ ค่าต่ำสุดในชุดข้อมูล X

$max(X)$ คือ ค่าสูงสุดในชุดข้อมูล X

การทำให้เป็นบรรทัดฐานด้วยวิธีคะแนนมาตรฐาน (Z-Score Normalization)

คือการแปลงข้อมูลเพื่อให้ชุดข้อมูลนั้นมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยสามารถแปลงข้อมูลโดยใช้สมการ

$$z_i = \frac{x_i - \mu}{\sigma}$$

โดยที่ z_i คือ ค่าบรรทัดฐานของข้อมูลตัวที่ i

x_i คือ ค่าข้อมูลที่ต้องการแปลงเป็นค่าบรรทัดฐานของข้อมูลตัวที่ i

μ คือ ค่าเฉลี่ยเลขคณิตจากชุดข้อมูลที่มี x เป็นสมาชิก

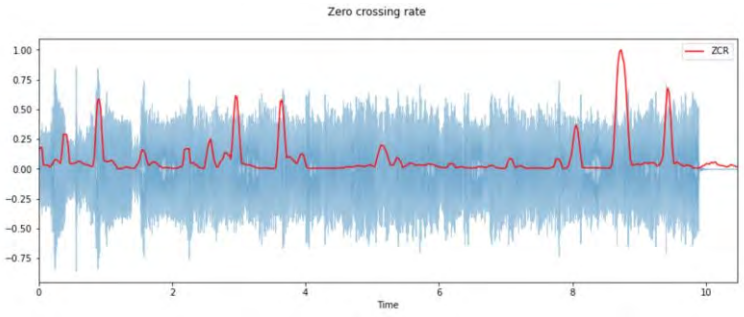
σ คือ ส่วนเบี่ยงเบนมาตรฐานชุดข้อมูลที่มี x เป็นสมาชิก

2.1.2 ลักษณะของเสียง

ลักษณะของเสียง คือข้อมูลที่แสดงลักษณะของเสียงในด้านใดด้านหนึ่ง ในงานวิจัยชิ้นนี้ใช้ลักษณะเสียงทั้งหมดเจ็ดลักษณะในการทดลองค้นคืนเพลง และใช้ลักษณะเสียงแทนลักษณะของข้อมูลเสียงทั้งสามแบบได้แก่ เสียงร้อง เสียงฮัม และเพลงต้นฉบับ เพื่อนำไปเปรียบเทียบในการทดลองค้นคืนเพลงจากการร้องและการฮัม

1. **อัตราการข้ามศูนย์ (Zero crossing rate : ZCR)**

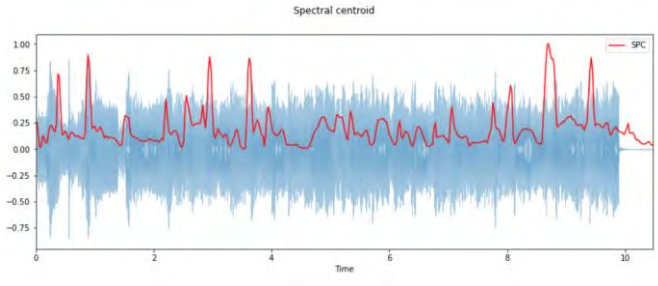
อัตราการข้ามศูนย์ คืออัตราการเปลี่ยนเครื่องหมายของแอมพลิจูดจากบวกเปลี่ยนเป็นศูนย์ เปลี่ยนเป็นลบ หรือจากลบเปลี่ยนเป็นศูนย์เปลี่ยนเป็นบวกของสัญญาณเสียงในแต่ละช่วง ดังภาพที่ 2.1



ภาพที่ 2.1 ตัวอย่างค่าลักษณะอัตราการข้ามศูนย์ของสัญญาณเสียง เพลงเจ็บจนพอ – แว่นใหญ่

2. เซนทรอยด์ของสเปกตรัม (Spectral centroid : SPC)

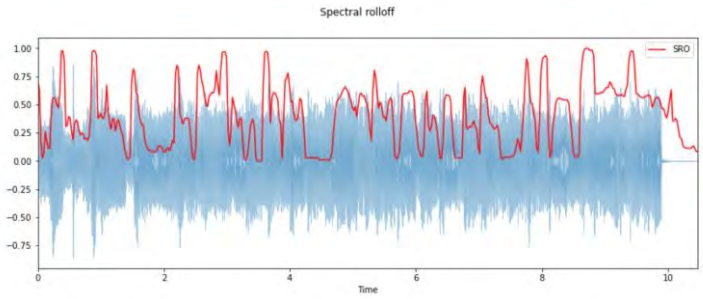
เซนทรอยด์ของสเปกตรัม คือจุดศูนย์กลางมวลของสเปกตรัมของสัญญาณเสียงในแต่ละช่วง ดังภาพที่ 2.2



ภาพที่ 2.2 ตัวอย่างค่าลักษณะเซนทรอยด์ของสเปกตรัมของสัญญาณเสียงเพลงเจ็บจนพอ - แว่นใหญ่

3. โรลออฟของสเปกตรัม (Spectral roll-off : SRO)

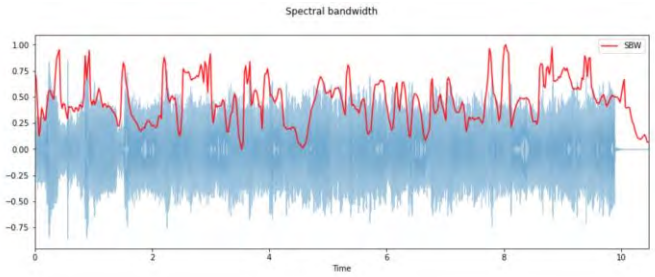
โรลออฟของสเปกตรัม คือสเปกตรัมที่มีพลังงานต่ำกว่าเกณฑ์โรลออฟที่กำหนดของสัญญาณเสียงในแต่ละช่วง ซึ่งกำหนดให้เกณฑ์โรลออฟมีค่า 0.85 สามารถวิเคราะห์ลักษณะการแผ่วลงของสัญญาณเสียง ณ ตำแหน่งต่าง ๆ ดังภาพที่ 2.3



ภาพที่ 2.3 ตัวอย่างค่าลักษณะโรลออฟของสเปกตรัมของสัญญาณเสียงเพลงเจ็บจนพอ - แว่นใหญ่

4. ความกว้างแถบของสเปกตรัม (Spectral bandwidth : SBW)

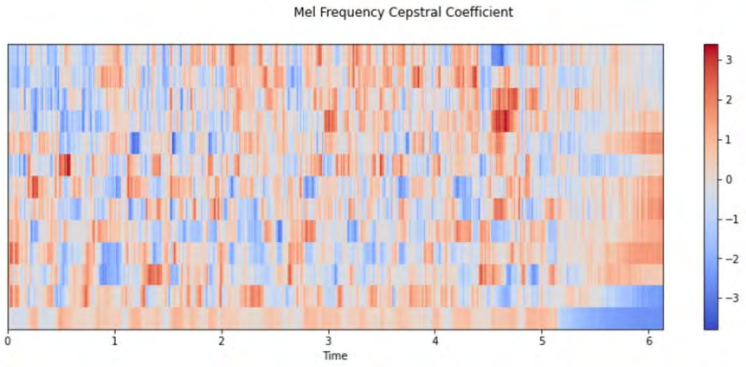
ความกว้างแถบของสเปกตรัม คือความกว้างของสเปกตรัมของสัญญาณเสียงในแต่ละช่วง ดังภาพที่ 2.4



ภาพที่ 2.4 ตัวอย่างค่าลักษณะความกว้างแถบของสเปกตรัมของสัญญาณเสียงเพลงเจ็บจนพอ - แว่นใหญ่

5. **สัมประสิทธิ์เซปสตรัมความถี่เมล (Mel-Frequency Cepstral Coefficients : MFCC)**

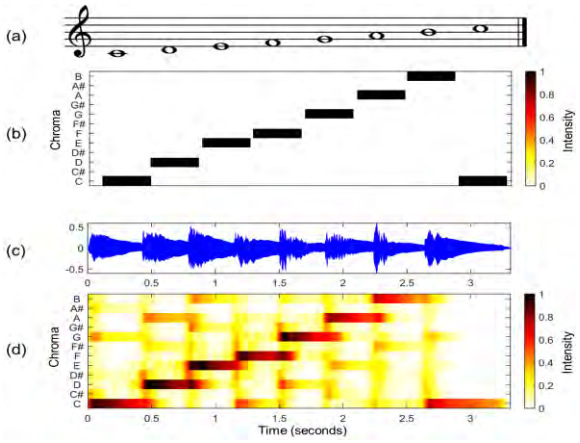
สัมประสิทธิ์เซปสตรัมความถี่เมล เป็นการนำสัญญาณเสียงมาผ่านตัวกรองที่จำลองตามการได้ยินของมนุษย์บนสเกลความถี่เมล เพื่อวิเคราะห์สัญญาณเสียงในรูปแบบเสียงมนุษย์ โดยปกติใช้ค่าสัมประสิทธิ์ 13 ค่า ทำให้ได้เวกเตอร์ลักษณะ 13 มิติในแต่ละช่วง และสามารถแสดงค่าลักษณะที่ได้ด้วยสเปคโตแกรม (spectogram) ดังภาพที่ 2.5



ภาพที่ 2.5 ตัวอย่างสเปคโตแกรมแสดงลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลของสัญญาณเสียงเพลงเจ็บจนพอ - แว่นใหญ่

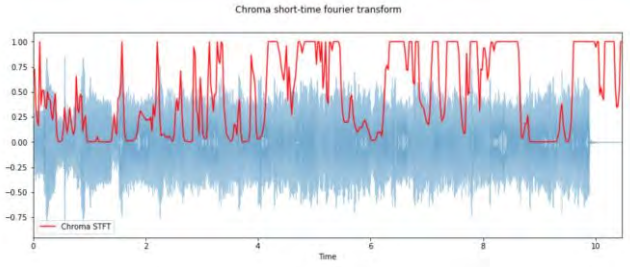
6. **การแปลงฟูเรียร์ช่วงสั้นโครมา (Chroma short-time fourier transform : CHM)**

การแปลงฟูเรียร์ช่วงสั้นโครมา คือการแยกระดับเสียงโน้ตดนตรีตามสเกลโครมาจากค่าฟูเรียร์ช่วงสั้นของสัญญาณเสียงในแต่ละช่วง ซึ่งระดับเสียงของสเกลโครมาจะแบ่งเป็น 12 ระดับ ได้แก่ C C# D D# E F F# G G# A A# B และสามารถแสดงได้ด้วยโครมาแกรม (chromagram) ดังภาพที่ 2.6 - 2.7



ภาพที่ 2.6 โครมาแกรมของการไล่สเกล C เมเจอร์

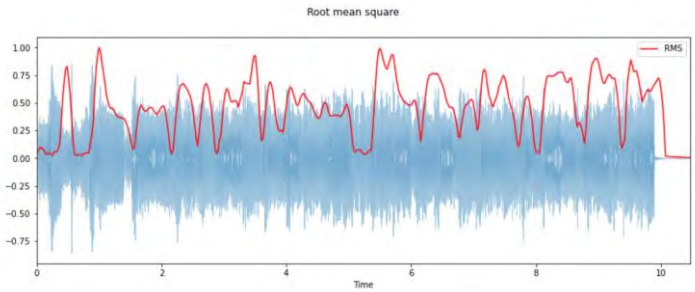
(ที่มา: https://en.wikipedia.org/wiki/Chroma_feature)



ภาพที่ 2.7 ตัวอย่างค่าลักษณะการแปลงฟูเรียร์ช่วงสั้นโครมาของสัญญาณเสียง
เพลงเจ็บจนพอ - แว่นใหญ่

7. รากกำลังสองเฉลี่ย (Root mean square : RMS)

รากกำลังสองเฉลี่ย คือพลังงานโดยรวมของสัญญาณเสียง ได้จากรากกำลังสองเฉลี่ย ความถี่ในแต่ละช่วงของสัญญาณเสียง ดังภาพที่ 2.8

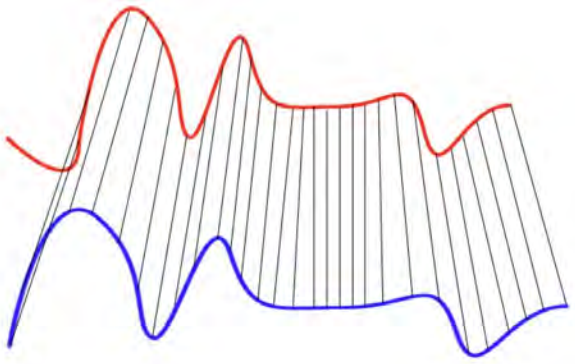


ภาพที่ 2.8 ตัวอย่างค่าลักษณะรากกำลังสองเฉลี่ยของสัญญาณเสียง
เพลงเจ็บจนพอ - แว่นใหญ่

2.1.3 การวัดค่าความเหมือนของข้อมูล

ไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping) [4]

ไดนามิกไทม์วอร์ปปิงเป็นขั้นตอนวิธีที่ใช้ในการเปรียบเทียบความเหมือนของลำดับข้อมูลสองชุดในอนุกรมเวลาที่มีความต่างในด้านความเร็วหรือด้านเวลา โดยในการเปรียบเทียบจะมีการยืดหรือหดตัวข้อมูลเพื่อเทียบกับตัวข้อมูลอีกชุดหนึ่ง ทำให้ข้อมูลมีความยืดหยุ่น เพราะรูปแบบข้อมูลการเคลื่อนที่วัตถุชนิดหนึ่งจะมองว่ามีความเหมือนกันไม่ว่าวัตถุชิ้นนั้นจะเคลื่อนที่ด้วยความเร็วที่ต่างกันขนาดใดก็ตาม ดังภาพที่ 2.9



ภาพที่ 2.9 การเปรียบเทียบลำดับข้อมูลสองชุดโดยใช้เทคนิคไดนามิกไทม์วอร์ปปิงบนอนุกรมเวลา

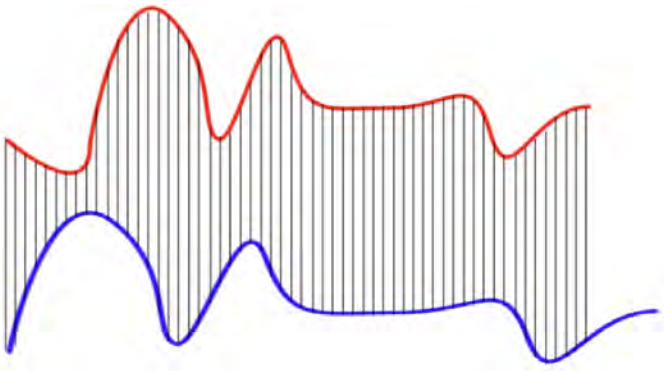
(ที่มา : https://th.wikipedia.org/wiki/Dynamic_time_warping)

จากภาพที่ 2.9 จะเห็นได้ว่าลำดับข้อมูลทั้งสองชุดเป็นข้อมูลที่คล้ายกันแม้ว่ามีการเคลื่อนของอนุกรมเวลาที่แตกต่างกัน การใช้เทคนิคไดนามิกไทม์วอร์ปปิงก็ยังสามารถหาความเหมือนของลำดับข้อมูลทั้งสองชุดได้ซึ่งจะได้ผลลัพธ์ออกมาเป็นระยะห่างสะสม

ไดนามิกไทม์วอร์ปปิงนิยมนำมาใช้ในงานด้านการแยกประเภทและการจัดกลุ่ม ในงานวิจัยฉบับนี้ได้ใช้เทคนิคไดนามิกไทม์วอร์ปปิงเพื่อช่วยในการแก้ปัญหาของเสียงร้องเสียงฮัมที่มีความเร็วหรือการประสานจังหวะไม่ตรงกับเพลงต้นฉบับ

ระยะห่างแบบยุคลิด Euclidean distance

เป็นหนึ่งในวิธีที่นิยมใช้ในการเปรียบเทียบความเหมือนของลำดับข้อมูลสองชุดในอนุกรมเวลาเดียวกัน ดังภาพที่ 2.10



ภาพที่ 2.10 การเปรียบเทียบลำดับข้อมูลสองชุดโดยใช้การวัดระยะห่างแบบยุคลิดบนอนุกรมเวลา

(ที่มา : https://th.wikipedia.org/wiki/Dynamic_time_warping)

เนื่องจากการคำนวณระยะห่างแบบยูคลิดคล้ายกับพีทาโกรัส โดยสามารถหาค่าความเหมือน ได้ตามสมการ

$$d = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

โดยที่ d คือ ค่าระยะห่าง

p_i คือ ข้อมูลตำแหน่งที่ i ในชุดข้อมูล P ในอนุกรมเวลา

q_i คือ ข้อมูลตำแหน่งที่ i ในชุดข้อมูล Q ในอนุกรมเวลา

n คือ จำนวนลำดับข้อมูลในชุดข้อมูล

จากสมการจะได้ว่า ถ้าค่าระยะห่างมีค่ามากแสดงว่าข้อมูลสองชุดนั้นมีความเหมือนกันน้อย ในทางกลับกัน ถ้าค่าระยะห่างมีค่าน้อยแสดงว่าข้อมูลสองชุดนั้นมีความเหมือนกันมาก และถ้าได้ค่าระยะห่างเท่ากับศูนย์ นั่นคือ ข้อมูลสองชุดนี้เหมือนกันหรือเป็นชุดเดียวกัน

2.1.4 การเปรียบเทียบความเหมือนและการวัดผลการค้นคืนเพลง

ในงานวิจัยนี้ได้กำหนดนิยามที่เกี่ยวข้องดังต่อไปนี้

นิยาม Top n

Top n หมายถึงชุดเพลงต้นฉบับที่ค้นคืนได้โดยมีความเหมือนมากที่สุดเมื่อเทียบกับเสียงร้องหรือเสียงฮัมจำนวน n ลำดับ

นิยาม ความถูกต้อง

ความถูกต้อง (accuracy) หมายถึงเปอร์เซนต์สัดส่วนของจำนวนเพลงที่ถูกค้นคืนได้อย่างถูกต้องบนชุด n อันดับสูงสุด (Top n) ต่อจำนวนเพลงทั้งหมดที่ได้นำมาใช้ในการค้นคืนเพลง สามารถหาได้จากสมการ

$$accuracy\ (\%)\ =\ \frac{N_R}{N}\ \times\ 100\%$$

โดยที่ $accuracy$ คือ เปอร์เซนต์ความถูกต้องที่ได้จากการทดลองนี้
N_R คือ จำนวนเพลงที่ถูกค้นคืนได้อย่างถูกต้องบนชุด n อันดับสูงสุด (Top n)
N คือ จำนวนเพลงทั้งหมดที่นำมาทดลองในการทดลองนี้

นิยาม ความเหมือนและความแตกต่าง

ความเหมือนหมายถึงความใกล้เคียงกันระหว่างข้อมูลสองชุด เนื่องจากในงานวิจัยฉบับนี้ ได้ใช้การวัดระยะห่างแบบยุคลิดซึ่งเป็นการเปรียบเทียบความแตกต่าง แต่ทางผู้วิจัยต้องการ ที่จะหาความเหมือนระหว่างเสียงร้อง เสียงฮัมกับเพลงต้นฉบับ เนื่องจากความเหมือนผกผันกับความแตกต่าง ดังนั้นในบทถัดไป ความแตกต่างที่น้อยที่สุดมีความหมายเช่นเดียวกับความเหมือนที่มากที่สุด

2.2 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้อง ทางผู้วิจัยพบว่าการศึกษาการสอบถามด้วยการร้องและการฮัม (Query by Singing/Humming : QBSH) จากงานวิจัยของ Kichul Kim และคณะ [5] งานวิจัยของ Rifki Afina Putri และ Dessi Puji Lestari [6] และงานวิจัยของ Hung-Ming Yu และคณะ [7] เสนอระบบสกัดระดับเสียง (pitch extraction) โดยนำเอาเวกเตอร์ของระดับเสียงจากข้อมูลต้นฉบับมาเปรียบเทียบความเหมือนด้วยวิธี Dynamic Time Warping matching ซึ่ง [5] จะให้ผลลัพธ์เป็นเพลงต้นฉบับที่ตรงกับข้อมูลนำเข้ามากที่สุด แต่ [6] เปรียบเทียบความเหมือนของข้อมูลเสียงในอนุกรมเวลาที่มีความคลาดเคลื่อนไปจากต้นฉบับ ซึ่งหลังจากทำ DTW แล้วจะได้ผลลัพธ์ออกมาเป็นความแตกต่างที่น้อยที่สุดและนำมาตั้งเกณฑ์ เพื่อป้องกันไม่ให้ระบบ QBSH ตอบเพลงที่ไม่ใช่คำตอบออกมาให้กับผู้ใช้ และ [7] ได้ใช้การเปลี่ยนรูปแบบ (transposition) ของข้อมูลนำเข้าของผู้ใช้งานที่เป็นลำดับโน้ต ร่วมกับการใช้ DTW ในการหาความแตกต่าง

ที่น้อยที่สุดเหมือนกัน เพื่อแก้ไขกรณีที่ผู้ใช้ร้องเพลงไม่ตรงกับคีย์ของเพลงต้นฉบับ แล้วให้ผลลัพธ์เป็นรายชื่อเพลงที่มีความคล้ายเพลงต้นฉบับมากที่สุด n อันดับ (Top n)

ในการศึกษาเกี่ยวกับการสกัดลักษณะของเสียง (audio feature extraction) จากงานวิจัยของ Marta Jaczyńska และคณะ [8] ศึกษาการรู้จำเพลง (music recognition) ด้วยวิธีการสกัดระดับเสียง [9] แบบต่าง ๆ พบว่าตัวประมาณค่าความถี่พื้นฐานของเสียงพูดและระดับเสียงที่ชื่อว่า YIN สามารถสกัดระดับเสียงได้ดีที่สุด รวมถึงขั้นตอนวิธี FFT YIN ที่นำการแปลงฟูเรียร์แบบเร็ว (fast fourier transform : FFT) มาประยุกต์ ทำให้ใช้เวลาสกัดระดับเสียงน้อยลงถึงสามเท่า โดยความถูกต้องของระดับเสียงที่สกัดได้ยังอยู่ในระดับเดิม และงานวิจัยของ Tomasz Maka และ Piotr Dziurzansk [10] ศึกษาการสกัดลักษณะของเสียงหลากหลายแบบ โดยจำแนกเป็นกลุ่มต่าง ๆ เช่น คือ สัมประสิทธิ์เซปสตรัมความถี่เมล (mel-frequency cepstral coefficients : MFCC), การแปลงฟูเรียร์แบบเร็ว (fast fourier transform : FFT), กลุ่มที่มีเวลาเป็นฐาน (time-based features) เช่น อัตราการข้ามศูนย์ (zero crossing rate : ZCR) [11] และพลังงานช่วงเวลาสั้น (short-time energy : STE) และกลุ่มที่มีความถี่เป็นฐาน (frequency-based features) เช่น spectral centroid, spectral roll-off และ bandwidth

ลักษณะของเสียงแต่ละชนิดสามารถบ่งบอกลักษณะของเสียงในด้านต่าง ๆ กัน โดยลักษณะของเสียงที่เป็นที่แพร่หลายในงานวิจัยที่เกี่ยวข้องกับเสียงของมนุษย์คือ สัมประสิทธิ์เซปสตรัมความถี่เมล (mel-frequency cepstral coefficients : MFCC) [12] ซึ่งเป็นการนำสัญญาณเสียงมาวิเคราะห์บนสเกลความถี่เมล โดยสัญญาณเสียงจะถูกกรองด้วยตัวกรองที่เป็นการจำลองการสร้างเสียงของมนุษย์

ในการศึกษาเกี่ยวกับการวัดค่าความคล้าย (Similarity measure) ของลักษณะต่าง ๆ ของเสียง งานวิจัยของ Nastaran Borjian [9] ศึกษาการวัดความคล้ายของลายพิมพ์เสียงจากข้อมูลนำเข้า และข้อมูลต้นฉบับในฐานข้อมูลซึ่งใช้เป็นเกณฑ์สำหรับการค้นคืนเพลงที่ใกล้เคียงกับข้อมูลนำเข้าที่สุด

นอกจากนี้งานวิจัยของ Chung-Che wang และคณะ [13] ศึกษาการรู้จำทำนองโดยนำ Linear Scaling มาช่วยในการค้นคืนเพลงจากข้อมูลนำเข้าที่เป็นเสียงจากการร้องและการฮัม แต่วิธีนี้ใช้เวลาในการคำนวณสูง ผู้วิจัยจึงเพิ่มการวิเคราะห์ความคล้ายของคำร้อง (lyrics) จากข้อมูลนำเข้าที่เป็นเสียงจากการร้อง ด้วยเทคนิคแปลงเสียงพูดเป็นตัวอักษร (speech recognition technique) แล้วเปรียบเทียบกับคำร้องในฐานข้อมูล เมื่อนำวิธีทั้งสองข้างต้นมาผนวกกันแล้วสามารถเพิ่มอัตราการรู้จำของระบบโดยรวม

จากงานวิจัยที่ได้กล่าวมาข้างต้น ผู้วิจัยจึงนำลักษณะของเสียงเจ็ดลักษณะ การเปรียบเทียบความเหมือนด้วยเทคนิคไดนามิกไทม์วอร์ปปิง การเปรียบเทียบระยะห่างแบบยุคลิด และการกำหนดเกณฑ์การยอมรับ ปฏิเสธ มาใช้ในงานวิจัยครั้งนี้ โดยจะกล่าวรายละเอียดในบทที่ 3 ต่อไป

บทที่ 3

วิธีการวิจัย

ในบทนี้จะกล่าวถึงวิธีการวิจัยของการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม ซึ่งในงานวิจัยนี้จะแบ่งกระบวนการหลักออกเป็นสี่ขั้นตอน ได้แก่

1. วิธีการเก็บข้อมูล
2. วิธีการเตรียมข้อมูลก่อนนำมาใช้
3. การค้นคืนเพลง
    a. การสกัดลักษณะเสียง
    b. การเปรียบเทียบความเหมือนของลักษณะเสียง
    c. วิธีการผสานลักษณะ
    d. การพิจารณาผลของประเภทเพลงต่อความถูกต้อง
4. การวัดผลการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม
    a. การจัดอันดับคำตอบ
    b. การวัดค่าความถูกต้องของการค้นคืนเพลงไทย
    c. การหาค่าอันดับเฉลี่ย
    d. การคำนวณเกณฑ์การยอมรับและปฏิเสธ

3.1 วิธีการเก็บข้อมูล

ผู้วิจัยได้ทำการบันทึกเพลงเป็นไฟล์นามสกุลเอ็มพีสาม (.mp3) โดยประเภทของเพลงมีทั้งหมดห้าประเภท ได้แก่ พอป ร็อค แจ๊ส ฮิปฮอป และลูกทุ่ง ประเภทละ 50 เพลง บันทึกชื่อไฟล์เป็น 1-250 และบันทึกชื่อไฟล์ลงใน excel จากนั้นนำเพลงที่ได้มาตัดด้วยโปรแกรมการาจแบนด์ (GarageBand) [14] โดยเลือกเฉพาะท่อนฮุคและใช้ความยาว 7-15 วินาทีมาบันทึกเป็นเพลงต้นฉบับ และผู้วิจัยเก็บเสียงร้องและเสียงฮัมของเพลงต้นฉบับด้วยโปรแกรมการาจแบนด์ (GarageBand) บันทึกเป็นไฟล์นามสกุลเอ็มพีสาม (.mp3) เช่นเดียวกัน ซึ่งจะได้ไฟล์เพลงต้นฉบับ ไฟล์เสียงร้อง และไฟล์เสียงฮัม รวมทั้งหมด 750 ไฟล์

3.2 วิธีการเตรียมข้อมูลก่อนนำมาใช้

ผู้วิจัยนำตัวแบบ Open-Unmix [3] มาใช้เพื่อตัดเสียงดนตรีหรือเสียงอื่น ๆ ออกจากไฟล์เพลงต้นฉบับ ไฟล์เสียงร้อง และไฟล์เสียงฮัม ที่ได้จากข้อ 3.1 ให้เหลือแต่เสียงร้องหรือเสียงฮัมที่ชัดเจนขึ้น จากนั้นแปลงสัญญาณเสียงของเพลงต้นฉบับ เสียงร้อง และเสียงฮัม ที่ตัดเสียงอื่นออกแล้วให้เป็นแถวลำดับของเลขจำนวน

จริงแล้วแปลงข้อมูลโดยการทำให้เป็นบรรทัดฐานด้วยวิธีค่าสูงสุดต่ำสุด (min-max normalization) ซึ่งจะทำให้ได้แถวลำดับของจำนวนจริงที่มีค่าในช่วง -1 ถึง 1 ออกมาสามชุดที่แทนสัญญาณเสียงของเพลงต้นฉบับ เสียงร้อง และเสียงฮัม ของเพลงหนึ่งเพลง จากนั้นนำแถวลำดับทั้งสามที่ได้ไปบันทึกในไฟล์รูปแบบ .xlsx ภายในไฟล์ประกอบด้วยสามคอลัมน์ ได้แก่ แถวลำดับของเพลงต้นฉบับ แถวลำดับของเสียงร้อง และแถวลำดับของเสียงฮัม ตามลำดับ ซึ่งจะทำให้ได้ไฟล์ .xlsx จำนวนรวมทั้งสิ้น 250 ไฟล์ จากนั้นจะสกัดลักษณะเสียงจากข้อมูลในแถวลำดับ เพื่อนำไปใช้ในการเปรียบเทียบความเหมือนของลักษณะเสียงต่อไป เนื่องจากความยาวของเพลงแต่ละเพลงไม่เท่ากันซึ่งทำให้ได้แถวลำดับที่มีความยาวไม่เท่ากัน จึงมีการตัดขนาดของแถวลำดับในแต่ละไฟล์ให้มีช่วงความยาวที่เท่ากันเพื่อให้สามารถนำมาเปรียบเทียบกันได้

3.3 การค้นคืนเพลง

ในการทดลองการค้นคืนเพลง จะเลือกใช้ลักษณะเสียงและเทคนิคการเปรียบเทียบความเหมือนของลักษณะเสียงที่ให้คำตอบในการค้นคืนที่ดีที่สุด

3.3.1 การสกัดลักษณะเสียง

เมื่อรับข้อมูลนำเข้าเป็นเสียงร้องหรือเสียงฮัมและเตรียมข้อมูลแล้ว จะได้ข้อมูลนำเข้าเก็บเป็นแถวลำดับ และนำมาสกัดลักษณะเสียง เพื่อนำไปเปรียบเทียบกับลักษณะเสียงกับเพลงต้นฉบับอื่น ๆ ลักษณะเสียงที่ใช้ในการทดลองที่นำมาจากไลบรารี librosa [15] มีทั้งหมดหกชนิดดังนี้

1. เซนทรอยด์ของสเปกตรัม (Spectral centroid : SPC)
2. โรลออฟของสเปกตรัม (Spectral roll-off : SRO)
3. อัตราการข้ามศูนย์ (Zero crossing rate : ZCR)
4. ความกว้างแถบของสเปกตรัม (Spectral bandwidth : SBW)
5. การแปลงฟูเรียร์ช่วงสั้นโครมา (Chroma short-time fourier transform : CHM)
6. รากกำลังสองเฉลี่ย (Root mean square : RMS)

เมื่อนำข้อมูลแถวลำดับหนึ่งมิติในไฟล์ที่ได้จากหัวข้อ 3.2 มาสกัดเพื่อหาลักษณะเสียงโดยกำหนดค่าพารามิเตอร์ตามค่าโดยปริยายที่ไลบรารีได้กำหนดไว้ให้ จะได้ผลลัพธ์เป็นค่าของลักษณะเสียงซึ่งเก็บเป็นแถวลำดับของเลขจำนวนจริงขนาดหนึ่งมิติ จากนั้นปรับค่าให้เป็นบรรทัดฐานด้วยวิธีคะแนนมาตรฐาน (z-score normalization)

สำหรับลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมล (Mel-Frequency Cepstral Coefficients : MFCC) นำมาจากไลบรารี essentia [16] โดยกำหนดค่าพารามิเตอร์ดังตารางที่ 3.1

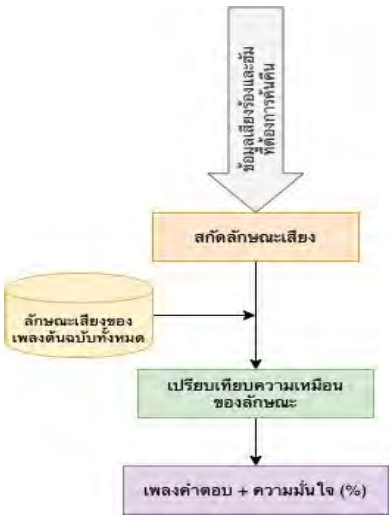
ตารางที่ 3.1 ค่าพารามิเตอร์ของการสกัดลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมล

ตัวแปร	การกำหนดค่า
ประเภทของกรอบหน้าต่าง (Windowing type)	หน้าต่างแบบแฮมมิง (hamming)
จำนวนของสัมประสิทธิ์เมล (numberCoefficients)	13
ขนาดของกรอบ (frameSize)	1024
ขนาดของการกระโดด (hopSize)	500

ซึ่งจะได้ผลลัพธ์เป็นค่าของลักษณะเก็บเป็นแถวลำดับสองมิติ จากนั้นปรับค่าให้เป็นบรรทัดฐานด้วยวิธีคะแนนมาตรฐาน (z-score normalization)

3.3.2 การเปรียบเทียบความเหมือนของลักษณะเสียง

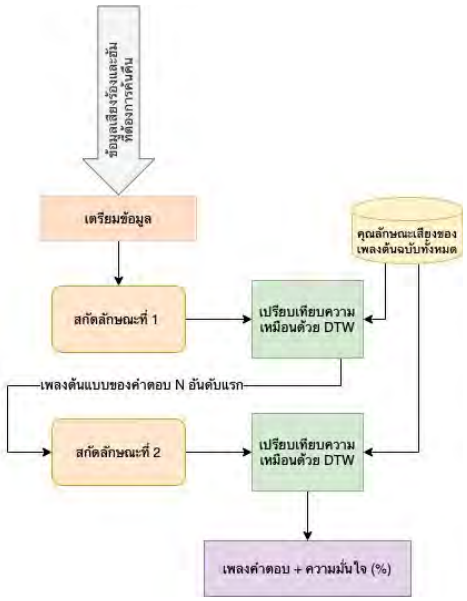
เมื่อได้ลักษณะเสียงของข้อมูลนำเข้า จะเปรียบเทียบความเหมือนของลักษณะเสียงกับเพลงต้นฉบับโดยใช้เทคนิคไดนามิกไทม์วอร์ปปิงและการวัดระยะห่างแบบยุคลิด ได้คำตอบเป็นค่าความเหมือนเมื่อเปรียบเทียบกับเพลงต้นฉบับที่กำหนด จากนั้นเรียงลำดับตามความเหมือนมากที่สุด และนำเพลงต้นฉบับที่มีความเหมือนมากที่สุดอันดับ 1 (Top 1) มาพิจารณาค่าระดับความมั่นใจของคำตอบ ซึ่งคำนวณจากเกณฑ์การยอมรับและปฏิเสธของค่าความเหมือนของเสียงร้องหรือฮัม แสดงดังภาพที่ 3.1



ภาพที่ 3.1 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากเสียงร้องและเสียงฮัม

3.3.3 วิธีการผสานลักษณะ

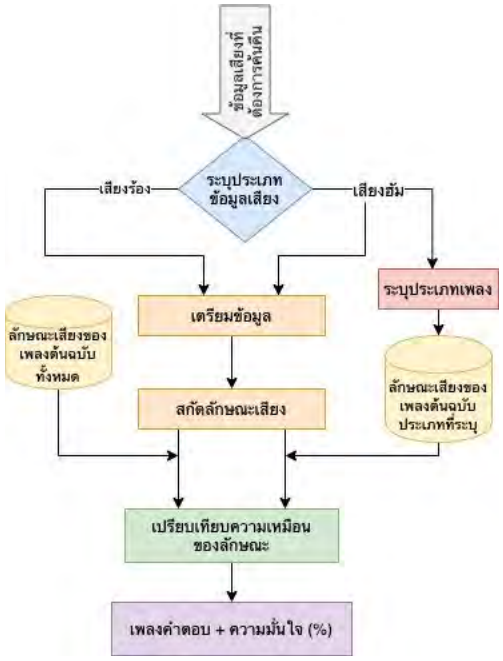
เป็นการค้นคืนเพลงโดยใช้ลักษณะสองลักษณะทำงานผสานกันเพื่อพัฒนาการค้นคืนเพลงให้มีความถูกต้องมากขึ้น ลักษณะที่นำมาทดลองผสานลักษณะ ได้แก่ สัมประสิทธิ์เซปสตรัมความถี่เมล เซนทรอยด์ของสเปกตรัม และโรลออฟของสเปกตรัม โดยการผสานลักษณะจะนำคำตอบ n อันดับแรก (Top n) ที่ได้จากการเปรียบเทียบความเหมือนของลักษณะแรกด้วยเทคนิคไดนามิกไทม์วอร์ปปิง มาเป็นข้อมูลนำเข้าของการสกัด ลักษณะที่สองและนำคำตอบที่ได้เป็นคำตอบของการค้นคืน แสดงดังภาพที่ 3.2



ภาพที่ 3.2 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากการฮัมด้วยการผสานลักษณะ

3.3.4 การพิจารณาผลของประเภทเพลงต่อความถูกต้อง

ผู้วิจัยมีสมมติฐานว่าประเภทของเพลงคำถามน่าจะมีผลต่อความถูกต้องของการค้นคืน จึงเพิ่มขั้นตอนการค้นคืนเพลงโดยคำนึงถึงประเภทของเพลงคำถาม และลักษณะการค้นคืนว่าเป็นข้อมูลเสียงที่รับเข้าเป็นเสียงร้องหรือเสียงฮัม กรณีการค้นคืนจากเสียงร้องจะเปรียบเทียบความเหมือนของลักษณะกับเพลงต้นฉบับทั้งหมด และในกรณีการค้นคืนจากเสียงฮัมจะเปรียบเทียบความเหมือนของลักษณะกับเพลงต้นฉบับตามประเภทที่ผู้ใช้ระบุ โดยใช้เทคนิคไดนามิกไทม์วอร์ปปิง และได้คำตอบเป็นค่าความเหมือน เมื่อเปรียบเทียบกับเพลงต้นฉบับที่กำหนด จากนั้นนำค่าความเหมือนของเพลงคำตอบมาพิจารณาค่าระดับความมั่นใจของคำตอบ ซึ่งคำนวณจากเกณฑ์การยอมรับและปฏิเสธ ของค่าความเหมือนของเสียงร้องหรือฮัม แสดงดังภาพที่ 3.3



ภาพที่ 3.3 แผนภาพแสดงการทำงานของการค้นคืนเพลงจากเสียงร้องและเสียงฮัม โดยคำนึงถึงประเภทเพลง

3.4 การวัดผลการค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

การวัดผลการค้นคืนเพลงไทยจากการร้องและการฮัมประกอบด้วยสี่ขั้นตอนย่อย ได้แก่ การจัดอันดับคำตอบ การวัดค่าความถูกต้องของการค้นคืนเพลงไทย การหาค่าอันดับเฉลี่ย และการคำนวณเกณฑ์การยอมรับและปฏิเสธ

3.4.1 การจัดอันดับคำตอบ

ในงานวิจัยนี้การวัดค่าความแตกต่างจะใช้เทคนิคไดนามิกไทม์วอร์ปปิง จากนั้นนำค่าความแตกต่างที่ได้มาเรียงลำดับจากค่าความแตกต่างน้อยที่สุดไปยังค่าความแตกต่างที่มากที่สุด ในกรณีที่เพลงต้นฉบับลำดับที่หนึ่งซึ่งมีค่าความแตกต่างที่น้อยที่สุด ตรงกับเสียงร้องหรือเสียงฮัมที่นำมาสอบถาม จะได้ว่าเพลงนี้เป็นเพลงคำตอบและอยู่ในลำดับที่หนึ่ง (Rank 1) ในทางกลับกันถ้าเพลงต้นฉบับลำดับที่หนึ่งไม่ตรงกับเสียงร้อง หรือเสียงฮัมที่นำมาสอบถาม ผู้วิจัยจะนำเพลงลำดับที่สองมาตรวจสอบว่าตรงกับเสียงร้อง หรือเสียงฮัมที่นำมาสอบถามหรือไม่ จนกระทั่งพบเพลงคำตอบที่ตรงกับเสียงร้องหรือเสียงฮัม ที่นำมาสอบถาม และกำหนดให้เป็นเพลงคำตอบที่อยู่ในลำดับที่ n (Rank n) เช่น เสียงฮัมของเพลงที่สี่ เมื่อนำมาสอบถาม กับระบบได้คำตอบว่า เพลงต้นฉบับเพลงที่สี่

เรียงลำดับตามความแตกต่างแล้วอยู่ในลำดับที่สาม (Rank 3) และจะสรุปได้ว่าการค้นคืนเพลงด้วยเสียงฮัมของเพลงที่สี่ได้คำตอบอยู่ในสามอันดับแรก (Top 3)

นอกจากนี้ผู้วิจัยยังมีการนับผลรวมของอันดับคำตอบที่ได้จากการค้นคืนในแต่ละลำดับเก็บเป็นความถี่สะสมของอันดับคำตอบที่ได้จากการค้นคืนในรูปแบบ n อันดับแรก (Top n) โดย n แทนเลขอันดับที่ต้องการ และค่า Top n จะแสดงผลรวมของลำดับคำตอบที่ได้จากการค้นคืนตั้งแต่ 1 จนถึง n เช่น เมื่อผู้วิจัยกล่าวถึง Top 10 ของ 250 เพลง จะหมายถึงผลรวมของเพลงคำตอบที่พบในลำดับที่ 1 จนถึงลำดับที่ 10 (Rank 1-10)

3.4.2 การวัดค่าความถูกต้องของการค้นคืนเพลงไทย

วัดค่าความถูกต้องโดยนำจำนวนเพลงที่ถูกค้นคืนได้อย่างถูกต้องบนชุด n อันดับสูงสุด (Top n) หารด้วยจำนวนเพลงต้นฉบับทั้งหมดที่ได้นำไปเปรียบเทียบกับเสียงร้องหรือเสียงฮัมในครั้งนั้น ๆ คูณด้วย 100 ดังสมการในหัวข้อ 2.1.4 ในงานวิจัยฉบับนี้ผู้วิจัยได้วัดค่าความถูกต้องด้วย Top 1, Top 5, Top 10 และ Top 20 เท่านั้น

3.4.3 การหาค่าอันดับเฉลี่ย

ผู้วิจัยหาค่าอันดับเฉลี่ยโดยนำผลรวมของผลคูณระหว่างเลขลำดับที่ m (rank m) กับจำนวนคำตอบของเลขลำดับที่ m ตั้งแต่ 1 จนถึง n หารด้วย จำนวนของ Top n โดยในงานวิจัยฉบับนี้ ได้หาค่าอันดับเฉลี่ยโดยให้ n มีค่า 10 และ 20 เท่านั้น

3.4.4 การคำนวณเกณฑ์การยอมรับและปฏิเสธ

ผู้วิจัยนำค่าความเหมือนระหว่างเสียงฮัมกับเพลงต้นฉบับที่เหมือนกันมากที่สุด และพิจารณาว่าเพลงต้นฉบับนั้น ตรงกับเสียงฮัมของเพลงแต่ละประเภทที่สอบถามหรือไม่ ถ้าตรงกันกำหนดให้สถานะของคำตอบเป็น Top 1 ถ้าไม่ตรงกันกำหนดให้สถานะของคำตอบเป็น not Top 1 จากนั้นนำคำตอบทั้งหมดมาสร้างเป็นฮิสโตแกรมสองชุดบนระนาบเดียวกัน ได้แก่ ชุดเพลงที่เป็นคำตอบ (Top 1) และชุดเพลงที่ไม่ใช่คำตอบ (not Top 1) โดยแกนระนาบแทนค่าความต่างจากน้อยที่สุดไปยังมากที่สุด สำหรับฮิสโตแกรมที่ได้ค่าความต่างที่น้อยที่สุด (min) และค่าความต่างที่มากที่สุด (max) จะนำมาใช้สร้างตาราง เช่น min = 274 และ max = 320 จะได้ตารางที่แสดงค่าความถี่ของเพลงที่เป็นคำตอบและเพลงที่ไม่ใช่คำตอบที่วัดค่าความแตกต่างเป็นจำนวนเต็มได้ตั้งแต่ 274 จนถึง 320 จากนั้นกำหนดเกณฑ์ยอมรับและเกณฑ์ปฏิเสธดังนี้

การคำนวณเกณฑ์การยอมรับ กำหนดให้ i เป็นค่าขีดแบ่งที่เป็นไปได้ และ $H_{T1}(i)$ และ $H_{\overline{T1}}(i)$ เป็นความถี่ของบิน (bin) ที่ i ในฮิสโตแกรมของชุดเพลงที่เป็นคำตอบและไม่ใช่คำตอบตามลำดับ ค่าขีดแบ่ง (threshold) ที่ใช้เป็นเกณฑ์การยอมรับ t_{acc} คำนวณได้จากสมการ

$$t_{acc} = \underset{j}{\operatorname{argmax}} \frac{\sum_{i=min}^{j} H_{T1}(i)}{\sum_{i=min}^{j} H_{T1}(i) + \sum_{i=min}^{j} H_{\overline{T1}}(i)}$$

และระดับความมั่นใจ $conf_{acc}$ ของเกณฑ์การยอมรับคำนวณได้จากสมการ

$$conf_{acc} = \max_{j} \frac{\sum_{i=min}^{j} H_{T1}(i)}{\sum_{i=min}^{j} H_{T1}(i) + \sum_{i=min}^{j} H_{\overline{T1}}(i)}$$

ภายใต้เงื่อนไข

$$\frac{\sum_{i=min}^{t_{acc}} H_{T1}(i)}{\sum_{i=min}^{t_{acc}} H_{T1}(i) + \sum_{i=min}^{t_{acc}} H_{\overline{T1}}(i)} > \frac{\sum_{i=min}^{t_{acc}+1} H_{T1}(i)}{\sum_{i=min}^{t_{acc}+1} H_{T1}(i) + \sum_{i=min}^{t_{acc}+1} H_{\overline{T1}}(i)}$$

โดยเราจะหาตำแหน่ง *j* ที่มากที่สุดที่ค่าระดับความมั่นใจ $conf_{acc}$ ของเกณฑ์การยอมรับมากกว่าที่ตำแหน่งถัดไปและใช้ค่าขีดแบ่ง t_{acc} นั้น เป็นเกณฑ์การยอมรับ

ในกรณีที่ต้องการค่าระดับความมั่นใจ 100% หรือ $conf_{acc} = 1$ นั้น $\sum_{i=min}^{t_{acc}} H_{\overline{T1}}(i)$ ต้องเท่ากับศูนย์ด้วย

อาจกล่าวได้ว่าถ้ากำหนดเกณฑ์การยอมรับโดยใช้ค่าขีดแบ่งที่มากกว่า t_{acc} จะทำให้ค่าระดับความมั่นใจในคำตอบที่ได้น้อยลง และในกรณีที่ต้องการค่าระดับความมั่นใจ 100% หรือ $conf_{acc} = 1$ นั้น $\sum_{i=min}^{t_{acc}} H_{\overline{T1}}(i)$ ต้องเท่ากับศูนย์ด้วย

การคำนวณเกณฑ์การปฏิเสธ กำหนดให้ i เป็นค่าขีดแบ่งที่เป็นไปได้ ค่าขีดแบ่ง (threshold) ที่ใช้เป็นเกณฑ์การปฏิเสธ t_{rej} คำนวณได้จากสมการ

$$t_{rej} = \underset{j}{\operatorname{argmax}} \frac{\sum_{i=j}^{max} H_{\overline{T1}}(i)}{\sum_{i=j}^{max} H_{T1}(i) + \sum_{i=j}^{max} H_{\overline{T1}}(i)}$$

และระดับความมั่นใจ $conf_{rej}$ ของเกณฑ์การยอมรับคำนวณได้จากสมการ

$$conf_{rej} = \max_{j} \frac{\sum_{i=j}^{max} H_{\overline{T1}}(i)}{\sum_{i=j}^{max} H_{T1}(i) + \sum_{i=j}^{max} H_{\overline{T1}}(i)}$$

ภายใต้เงื่อนไข

$$\frac{\sum_{i=t_{rej}}^{max} H_{\overline{T1}}(i)}{\sum_{i=t_{rej}}^{max} H_{T1}(i) + \sum_{i=t_{rej}}^{max} H_{\overline{T1}}(i)} > \frac{\sum_{i=t_{rej}-1}^{max} H_{\overline{T1}}(i)}{\sum_{i=t_{rej}-1}^{max} H_{T1}(i) + \sum_{i=t_{rej}-1}^{max} H_{\overline{T1}}(i)}$$

โดยเราจะหาตำแหน่ง *j* ที่มากที่สุดที่ค่าระดับความมั่นใจ $conf_{rej}$ ของเกณฑ์การปฏิเสธมากกว่าที่ตำแหน่งก่อนหน้าและใช้ค่าขีดแบ่ง t_{rej} นั้น เป็นเกณฑ์การปฏิเสธ

อาจกล่าวได้ว่าถ้ากำหนดเกณฑ์การปฏิเสธโดยใช้ค่าขีดแบ่งที่น้อยกว่า t_{rej} จะทำให้ค่าระดับความมั่นใจในคำตอบที่ได้น้อยลง และในกรณีที่ต้องการค่าระดับความมั่นใจ 100% หรือ $conf_{acc} = 1$ นั้น $\sum_{i=t_{rej}}^{max} H_{T1}(i)$ ต้องเท่ากับศูนย์ด้วย

บทที่ 4
ผลการวิจัย

ในบทนี้จะกล่าวถึง การตั้งค่าการทดลอง ผลการวิจัย และการอภิปรายผลการทดลอง การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม โดยพิจารณาเปรียบเทียบความเหมือนลักษณะของเสียง

4.1 การตั้งค่าการทดลอง

เพลงต้นฉบับที่นำมาใช้ในงานวิจัยนี้ เป็นท่อนฮุคของเพลงไทยจำนวน 250 เพลงประกอบด้วยเพลงพอป ร็อค แจ๊ส ฮิปฮอปและลูกทุ่ง ชนิดละ 50 เพลง แต่ละเพลงมีความยาว 7 - 15 วินาที และบันทึกเสียงร้องและเสียงฮัมของเพลงต้นฉบับให้มีความยาวเท่ากับความยาวของเพลงต้นฉบับ

การทดลองทั้งหมดทำในภาษาไพธอน โดยใช้ไลบรารีของ librosa และ essentia ในการสกัดลักษณะของเสียง ไลบรารี dtw สำหรับเทคนิคไดนามิกไทม์วอร์ปปิงในการวัดความเหมือนของลักษณะของเสียง และตัวแบบ Open-Unmix ในการแยกเสียงดนตรีออกจากเพลงต้นฉบับ และแยกเสียงรบกวนอื่น ๆ ออกจากเสียงร้องและเสียงฮัม ทำให้ได้เสียงร้องที่ชัดเจนขึ้นสำหรับเพลงต้นฉบับ เสียงร้องและเสียงฮัม ตามลำดับ

การทดลองทั้งหมดเป็นการทดลองต่อยอดจากการทดลองขั้นก่อนหน้าเพื่อต้องการให้ได้ผลลัพธ์ของการค้นคืนเพลงที่ดียิ่งขึ้น ดังนี้

1. การเปรียบเทียบความเหมือนของเพลงต้นฉบับกับเสียงร้องและเสียงฮัม โดยนำข้อมูลเสียงมาสกัดหาลักษณะเสียงและนำไปเปรียบเทียบหาความเหมือนเพื่อหาผลลัพธ์ของการค้นคืนเพลง
2. มีการเตรียมข้อมูลเสียงนำเข้าด้วยการปรับข้อมูลให้เป็นบรรทัดฐาน และมีการตัดเพลงทุกเพลงที่จะทำการเปรียบเทียบความเหมือนให้มีขนาดความยาวเท่ากันทุกเพลง
3. มีการเพิ่มขั้นตอนการเตรียมข้อมูลเสียงนำเข้าด้วยการตัดเสียงดนตรีหรือเสียงอื่นออกด้วยตัวแบบ Open-Unmix
4. มีการใช้เทคนิคไดนามิกไทม์วอร์ปปิงมาใช้ในการหาค่าความเหมือน
5. มีการผสานลักษณะเสียงเพื่อใช้ในการหาความเหมือน
6. มีการพิจารณาประเภทของเพลงในการค้นคืนเพลง

4.2 ผลการวิจัย

4.2.1 การเปรียบเทียบความเหมือนของเพลงจากเสียงร้องและเสียงฮัม

ข้อมูลนำเข้าได้แก่ เพลงต้นฉบับ เสียงร้อง และเสียงฮัม เก็บข้อมูลเป็นรายการ จากนั้นนำมาสกัดลักษณะโดยใช้ไลบรารี librosa ดังนี้

1. เซนทรอยด์ของสเปกตรัม (Spectral centroid : SPC)
2. โรลออฟของสเปกตรัม (Spectral roll-off : SRO)
3. อัตราการข้ามศูนย์ (Zero crossing rate : ZCR)
4. ความกว้างแถบของสเปกตรัม (Spectral bandwidth : SBW)
5. การแปลงฟูเรียร์ช่วงสั้นโครมา (Chroma short-time fourier transform : CHM)
6. รากกำลังสองเฉลี่ย (Root mean square : RMS)

และสกัดลักษณะ สัมประสิทธิ์เซปสตรัมความถี่เมล (Mel-Frequency Cepstral Coefficients : MFCC) โดยใช้ไลบรารี essentia

จากนั้นวัดค่าความเหมือนของลักษณะ ระหว่างเสียงร้อง-เพลงต้นฉบับ และเสียงฮัม-เพลงต้นฉบับ และให้ผลลัพธ์เป็นความถูกต้องของการค้นคืน ดังตารางที่ 4.1

ตารางที่ 4.1 การเปรียบเทียบความถูกต้องของการค้นคืน Top 1 และ Top 5 ของลักษณะ 7 ชนิด

ลักษณะ	เสียงร้อง - เพลงต้นฉบับ		เสียงฮัม - เพลงต้นฉบับ	
	Top 1	Top 5	Top 1	Top 5
SPC	33.87	53.23	3.23	9.68
SRO	25.81	50.00	4.84	16.13
ZCR	19.35	29.03	4.84	11.29
SBW	24.19	41.94	3.23	12.90
MFCC	83.87	90.32	48.39	70.97
CHM	11.29	30.65	16.13	35.48
RMS	12.90	24.19	14.52	29.03

จากตารางที่ 4.1 พบว่าลักษณะที่ให้คำตอบที่ดีที่สุด สำหรับเสียงร้องคือ สัมประสิทธิ์เซปสตรัมความถี่เมล ตามด้วยเซนทรอยด์ของสเปกตรัม และสำหรับเสียงฮัมสัมประสิทธิ์เซปสตรัมความถี่เมล ให้ค่าความถูกต้องสูงที่สุด

4.2.2 การเปรียบเทียบความเหมือนของเพลง เมื่อมีการทำให้เป็นบรรทัดฐาน (Normalization) และตัดเพลงก่อน และหลังสกัดลักษณะ

การทดลองนี้เป็นการทดลองเพื่อเพิ่มประสิทธิภาพการค้นคืน โดยการทำให้เป็นบรรทัดฐานด้วยวิธีค่าสูงสุดต่ำสุด (min-max normalization) โดยเปรียบเทียบการทำข้อมูลให้เป็นบรรทัดฐานและตัดเพลงก่อน และหลังสกัดลักษณะ แสดงผลความถูกต้องของการค้นคืน ดังตารางที่ 4.2

ตารางที่ 4.2 การเปรียบเทียบความถูกต้องของการค้นคืน Top 1 และ Top 5 ของลักษณะ 7 ชนิดเมื่อมีการทำให้เป็นบรรทัดฐาน (Normalization) และตัดเพลงก่อนและหลังสกัดลักษณะ

ตัดก่อน	เพลงต้นฉบับ - เสียงร้อง		เพลงต้นฉบับ - เสียงฮัม		ตัดหลัง	เพลงต้นฉบับ - เสียงร้อง		เพลงต้นฉบับ - เสียงฮัม	
	Top 1	Top 5	Top 1	Top 5		Top 1	Top 5	Top 1	Top 5
SPC	40.32	56.45	14.84	34.52	SPC	33.87	53.23	3.23	9.68
SRO	27.42	53.23	1.61	17.74	SRO	25.81	50.00	4.84	16.13
ZCR	25.81	41.94	1.61	18.06	ZCR	19.35	29.03	4.84	11.29
SBW	20.97	33.87	4.84	18.06	SBW	24.19	41.94	3.23	12.90
MFCC	87.10	88.71	51.61	70.97	MFCC	83.87	90.32	48.39	70.97
CHM	8.06	43.55	9.68	19.35	CHM	11.29	30.65	16.13	15.48
RMS	6.45	40.32	12.90	27.42	RMS	12.90	24.19	14.52	19.03

จากตารางที่ 4.2 พบว่าการทำให้เป็นบรรทัดฐานด้วยวิธีค่าสูงสุดต่ำสุด (min-max normalization) ร่วมกับตัดเพลงก่อนสกัดลักษณะ ทำให้การค้นคืนมีความถูกต้องมากขึ้น โดยลักษณะที่ให้คำตอบที่ดีที่สุดสำหรับเสียงร้อง คือ สัมประสิทธิ์เซปสตรัมความถี่เมลและเซนทรอยด์

ของสเปกตรัม สำหรับเสียงฮัมลักษณะที่ให้คำตอบที่ดีที่สุดคือ คือ สัมประสิทธิ์เซปสตรัมความถี่เมล และเซนทรอยด์ของสเปกตรัม

4.2.3 การเปรียบเทียบความเหมือนของลักษณะ เมื่อมีการตัดเสียงดนตรี ร่วมกับการทำให้เป็นบรรทัดฐาน (Normalization)

จากการทดลองที่ 4.2.1 และ 4.2.2 ทางผู้วิจัยจึงเลือกลักษณะที่ให้ความถูกต้องในการค้นคืนมากที่สุด ได้แก่ เซนทรอยด์ของสเปกตรัม โรลออฟของสเปกตรัม และสัมประสิทธิ์เซปสตรัมความถี่เมล มาใช้ในการทดลองนี้ ซึ่งเป็นการทดลองเพื่อพัฒนาการค้นคืนเพลง โดยมีการทำให้เป็นบรรทัดฐานร่วมกับการตัดเสียงดนตรีด้วยตัวแบบ Open-Unmix ซึ่งเป็นการตัดเสียงดนตรีออกสำหรับเพลงต้นฉบับ และตัดเสียงรบกวนสำหรับเสียงร้อง และเสียงฮัมแสดงผลลัพธ์เป็นความถูกต้องของการค้นคืน (%) ดังตารางที่ 4.3

ตารางที่ 4.3 การเปรียบเทียบความถูกต้องของการค้นคืน (%) เมื่อมีการตัดเสียงดนตรี ร่วมกับการทำให้เป็นบรรทัดฐาน

ไม่ตัดดนตรี	เพลงต้นฉบับ - เสียงร้อง		เพลงต้นฉบับ - เสียงฮัม		ตัดดนตรี	เพลงต้นฉบับ - เสียงร้อง		เพลงต้นฉบับ - เสียงฮัม	
	Top 1	Top 5	Top 1	Top 5		Top 1	Top 5	Top 1	Top 5
SPC	40.32	56.45	4.84	14.52	SPC	50.4	67.2	5.6	18.8
SRO	27.42	53.23	1.61	17.74	SRO	42	58.8	6.4	20.8
MFCC	88.4	94	31.6	52	MFCC	88.8	94.8	36	62.4

จากตารางที่ 4.3 พบว่าเมื่อมีการทำให้เป็นบรรทัดฐานร่วมกับการตัดเสียงดนตรีและเสียงรบกวน ทำให้การค้นคืนเพลงจากเสียงร้อง และเสียงฮัมมีความถูกต้องมากขึ้น โดยสำหรับเสียงร้องลักษณะที่มีความถูกต้องมากที่สุดคือ สัมประสิทธิ์เซปสตรัมความถี่เมล ตามด้วยเซนทรอยด์ของสเปกตรัม และโรลออฟของสเปกตรัมตามลำดับ สำหรับเสียงฮัมลักษณะที่มีความถูกต้องมากที่สุดคือ สัมประสิทธิ์เซปสตรัมความถี่เมล ตามด้วยโรลออฟของสเปกตรัม และ เซนทรอยด์ของสเปกตรัมตามลำดับ

4.2.4 การใช้เทคนิคไดนามิกไทม์วอร์ปปิงเพื่อเพิ่มประสิทธิภาพการค้นคืนเพลง

จากการทดลองที่ 4.2.3 ทางผู้วิจัยใช้เทคนิคไดนามิกไทม์วอร์ปปิงมาใช้ในการหาค่าความเหมือนจากการสกัดลักษณะด้วยสัมประสิทธิ์เซปสตรัมความถี่เมล เพื่อเพิ่มความถูกต้องของการค้นคืนเพลง n อันดับแรก โดยแสดงผลลัพธ์ ดังตารางที่ 4.4

ตารางที่ 4.4 การเปรียบเทียบความถูกต้องของการค้นคืน n อันดับแรก เมื่อมีการใช้เทคนิคไดนามิกไทม์วอร์ปปิง

Top n	ความถูกต้องของการค้นคืนด้วยเสียงร้อง		ความถูกต้องของการค้นคืนด้วยเสียงฮัม	
	ใช้ DTW	ไม่ใช้ DTW	ใช้ DTW	ไม่ใช้ DTW
1	90	88.8	39.6	36
10	96.8	96.4	72.8	72
20	98.8	98.4	83.6	84.8
30	98.8	98.8	90.8	89.2
40	98.8	98.8	94	91.6
50	98.8	98.8	94	93.6
60	98.8	98.8	95.2	94.4
70	98.8	98.8	95.2	94.4
80	98.8	98.8	95.2	94.4
90	99.6	98.8	96.4	96.4
100	99.6	99.2	97.2	96.8

จากตารางที่ 4.4 พบว่าค่าความถูกต้องของการค้นคืนด้วยเสียงร้องมีค่าความถูกต้องที่ซ้ำกันบางส่วน เนื่องจากในการค้นคืนนั้นระบบสามารถค้นคืนโดยได้ผลลัพธ์ส่วนใหญ่อยู่ภายในคำตอบของ Top 20 และ Top 30 และสามารถพบผลลัพธ์ของการค้นคืนอีกครั้งเมื่อเป็น Top 80 และ Top 90 จากการค้นคืนเพลงไทยด้วยเสียงร้องโดยใช้เทคนิคไดนามิกไทม์วอร์ปปิงและไม่ใช้เทคนิคไดนามิกไทม์วอร์ปปิงตามลำดับ ดังนั้นสามารถสรุปได้ว่าเมื่อนำเอาเทคนิคไดนามิกไทม์วอร์ปปิงนั้นมาใช้หาค่าความเหมือนของสัมประสิทธิ์เซปสตรัมความถี่เมลของเสียงร้อง เสียงฮัม และเพลงต้นฉบับสามารถช่วยให้การค้นคืนเพลง มีความถูกต้องมากขึ้น

4.2.5 การเปรียบเทียบการค้นคืนเพลงด้วยการผสานลักษณะ

ผู้วิจัยนำลักษณะที่มีความถูกต้องที่สูงที่สุดจากการทดลองที่ 4.2.1 ได้แก่ สัมประสิทธิ์เซปสตรัมความถี่เมล เซนทรอยด์ของสเปกตรัม และโรลออฟของสเปกตรัม มาทำงานผสานกันในการค้นคืนเพลง แสดงผลลัพธ์ดังตารางที่ 4.5 และตารางที่ 4.6

ตารางที่ 4.5 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงจากเสียงร้อง เมื่อมีการผสานลักษณะ

Top n	SPC + MFCC	MFCC + SPC	SRO + MFCC	MFCC + SRO	SPC + SRO	SRO + SPC
1	82.4	58.4	79.6	49.2	42	49.2
10	85.6	85.6	82.8	83.6	65.6	82.8
20	85.6	94.8	83.2	94.8	74.8	92

ตารางที่ 4.6 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงจากเสียงฮัม เมื่อมีการผสานลักษณะ

Top n	SPC + MFCC	MFCC + SPC	SRO + MFCC	MFCC + SRO	SPC + SRO	SRO + SPC
1	32.8	14.4	32.8	14	6.4	12
10	51.6	51.6	49.6	49.6	27.6	50
20	52.4	69.6	50.4	69.6	37.2	68.8

จากตารางที่ 4.5 และ ตารางที่ 4.6 พบว่าการผสานลักษณะระหว่าง สัมประสิทธิ์เซปสตรัมความถี่เมลและเซนทรอยด์ของสเปกตรัม ได้ค่าความถูกต้องของการค้นคืนเพลงมากที่สุด ทั้งในการค้นคืนเพลงจากเสียงร้องและเสียงฮัม เมื่อพิจารณาคำตอบ Top 1 ของการค้นคืนเพลง

4.2.6 การทดลองผลของการค้นคืนเพลงด้วยเสียงฮัม ในเพลงแต่ละประเภท

การทดลองนี้เป็นการวิเคราะห์ผลของการค้นคืนเพลงด้วยเสียงฮัมว่าให้คำตอบอันดับแรก (Top 1) เป็นเพลงในประเภทเดียวกับเพลงคำถามหรือไม่ เพื่อวิเคราะห์ว่าเพลงประเภทเดียวกันมีความเหมือนกันมากกว่าเพลงประเภทอื่น ๆ หรือไม่ ได้ผลดังตารางที่ 4.7

ตารางที่ 4.7 การเปรียบเทียบผลของการระบุประเภทของเพลงในการค้นคืนเพลง

ประเภทของเพลงคำถาม	ผลของการระบุประเภทของเพลงในการค้นคืนเพลงด้วยเสียงฮัมในเพลงแต่ละประเภท (%)				
	พอป	ร็อค	แจ๊ส	ฮิปฮอป	ลูกทุ่ง
พอป	88	2	4	2	4
ร็อค	16	44	10	0	30
แจ๊ส	6	16	64	0	14
ฮิปฮอป	20	22	14	34	10
ลูกทุ่ง	20	18	14	4	44

จากตารางที่ 4.7 พบว่า การค้นคืนเพลงด้วยเสียงฮัมให้คำตอบเป็นเพลงประเภทเดียวกับเพลงคำถามมากกว่าประเภทอื่น ๆ ทำให้ทราบว่าเพลงในประเภทเดียวกันมีความเหมือนกันมากกว่าเพลงต่างประเภทกัน

4.2.7 การเปรียบเทียบการค้นคืนเพลงด้วยเสียงฮัม เมื่อพิจารณาแยกตามประเภทของเพลง

จากการทดลองที่ 4.2.6 ผู้วิจัยทำการทดลองนี้เพื่อวิเคราะห์ความถูกต้องของการค้นคืนเพลงด้วยเสียงฮัมเมื่อมีการค้นคืนกับเพลงแต่ละประเภท เปรียบเทียบกับการค้นคืนกับเพลงทั้งหมด แสดงผลการเปรียบเทียบความถูกต้องของการค้นคืนเพลงดังตารางที่ 4.8

ตารางที่ 4.8 การเปรียบเทียบความถูกต้องของการค้นคืนเพลงด้วยเสียงฮัมเมื่อเทียบกับเพลงทั้งหมดและเมื่อเทียบกับเพลงประเภทเดียวกัน

ประเภทเพลง	ความถูกต้องของการค้นคืนเมื่อเทียบกับเพลงทั้งหมด (%)	ความถูกต้องของการค้นคืนเมื่อเทียบกับเพลงประเภทเดียวกัน (%)
พอป	42	68
ร็อค	28	52
แจ๊ส	50	64
ฮิปฮอป	28	60
ลูกทุ่ง	32	48

จากตารางที่ 4.8 พบว่าการค้นคืนเพลงด้วยเสียงฮัมแยกตามประเภทของเพลงได้ความถูกต้องเพิ่มขึ้น เมื่อเปรียบเทียบกับการค้นคืนเพลงทั้งหมดโดยไม่แยกประเภท

4.2.8 การหาอันดับเฉลี่ยคำตอบของการค้นคืนเพลงด้วยเสียงฮัมแยกตามประเภทของเพลง

จากการทดลองที่ 4.2.6 ผู้วิจัยทำการทดลองนี้เพื่อเป็นการวิเคราะห์ผลการค้นคืนเพลงด้วยเสียงฮัมเมื่อแยกตามประเภทของเพลง ว่าประเภทของเพลงส่งผลต่ออันดับของคำตอบของการค้นคืนหรือไม่ โดยหาอันดับเฉลี่ยคำตอบของการค้นคืนเพลงด้วยเสียงฮัม 10, 20 อันดับแรก (Top 10, 20) ของเพลงแต่ละประเภทและเพลงทั้งหมด แสดงผลดังตารางที่ 4.9

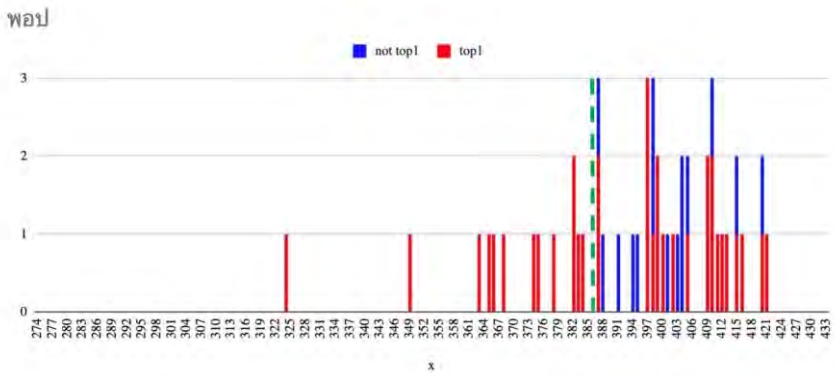
ตารางที่ 4.9 ค่าเฉลี่ยอันดับคำตอบของการค้นคืนเพลงด้วยเสียงฮัม แยกตามประเภทของเพลงและเพลงทั้งหมด

ประเภทของเพลง	การค้นคืนจากเพลงทั้งหมด		การค้นคืนจากเพลงประเภทเดียวกัน	
	อันดับเฉลี่ยของคำตอบ 10 อันดับแรก	อันดับเฉลี่ยของคำตอบ 20 อันดับแรก	อันดับเฉลี่ยของคำตอบ 10 อันดับแรก	อันดับเฉลี่ยของคำตอบ 20 อันดับแรก
พอป	2.52	3.80	1.88	1.88
ร็อค	2.88	5.86	2.27	2.68
แจ๊ส	2.05	3.07	2.08	2.29
ฮิปฮอป	3.09	4.87	1.73	2.67
ลูกทุ่ง	2.88	4.87	2.50	3.52

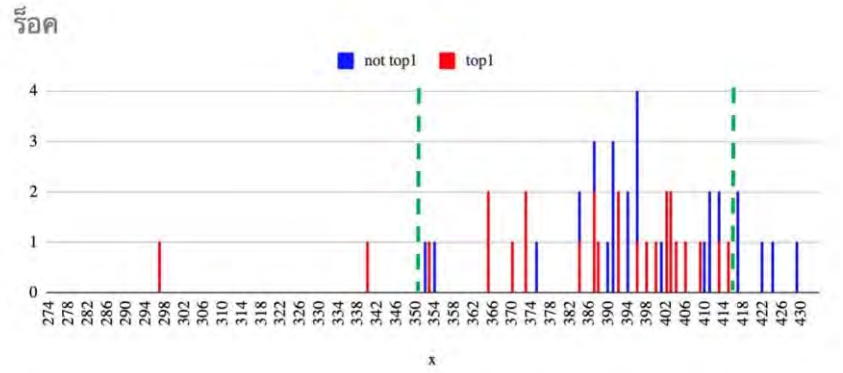
จากตารางที่ 4.9 พบว่าอันดับเฉลี่ยคำตอบของการค้นคืนเพลงในประเภทเดียวกับเพลงที่สอบถามได้อันดับดีกว่าการค้นคืนจากเพลงทั้งหมดที่รวมเพลงทั้งห้าประเภท ทำให้ทราบว่าการค้นคืนเพลง ในประเภทเดียวกับเพลงที่สอบถามจะได้คำตอบได้ดีกว่าการค้นคืนเพลงโดยไม่แยกประเภท

4.2.9 การหาเกณฑ์การยอมรับและเกณฑ์การปฏิเสธคำตอบจากการค้นคืนเพลงแต่ละประเภทด้วยเสียงฮัม

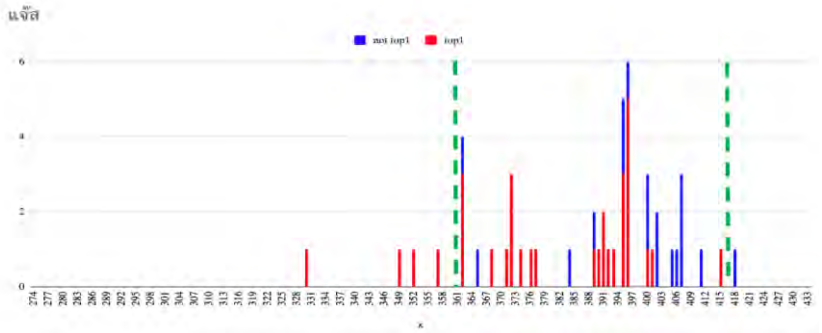
การทดลองนี้เป็นการวิเคราะห์ความเหมือนของเสียงร้องและเสียงฮัมของเพลงทั้งห้าประเภท ได้แก่ พอป ร็อค แจ๊ส ฮิปฮอป และลูกทุ่ง โดยใช้ฮิสโตแกรมของค่าความถี่ของความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับในแต่ละประเภท เพื่อหาเกณฑ์การยอมรับและเกณฑ์การปฏิเสธคำตอบ ซึ่งเป็นเกณฑ์ของความเหมือนที่แยกเพลงที่เป็นคำตอบ (Top 1) และเพลงที่ไม่ใช่คำตอบ (not Top 1) ได้โดยมีค่าระดับความมั่นใจ 100% เมื่อเทียบกับเพลงในประเภทเดียวกัน แสดงผลดังภาพที่ 4.1 - 4.5



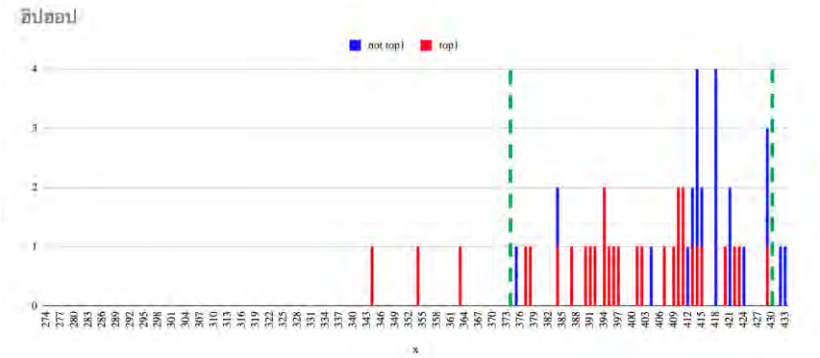
ภาพที่ 4.1 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทพอป



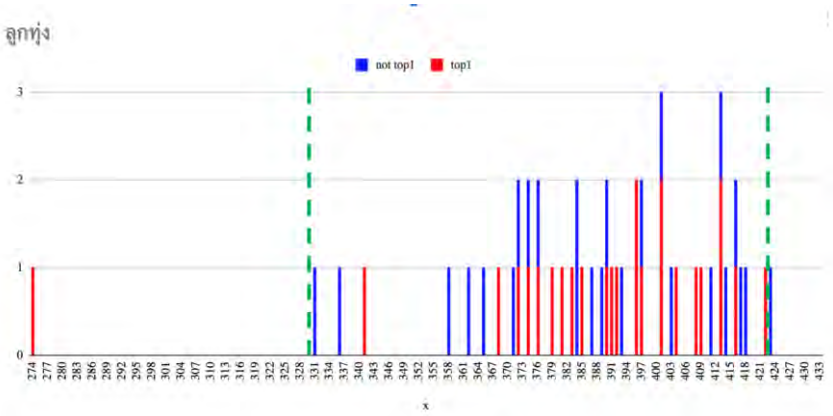
ภาพที่ 4.2 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทร็อค



ภาพที่ 4.3 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทแจ๊ส



ภาพที่ 4.4 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทฮิปฮอป



ภาพที่ 4.5 ฮิสโตแกรมของความถี่ของเพลงที่เป็นคำตอบ (สีแดง) และเพลงที่ไม่ใช่คำตอบ (สีน้ำเงิน) ที่ได้จากค่าความแตกต่างระหว่างเสียงฮัมและเพลงต้นฉบับของเพลงไทยประเภทลูกทุ่ง

จากฮิสโตแกรมของภาพที่ 4.1 – 4.5 แสดงเกณฑ์การยอมรับและเกณฑ์การปฏิเสธแสดงได้ด้วยค่าขีดแบ่ง (เส้นประ สีเขียว) เพื่อยืนยันเพลงคำตอบที่ค้นคืนได้ว่ายอมรับหรือปฏิเสธด้วยค่าความมั่นใจ 100% หรือไม่พบคำตอบที่ถูกต้องจากเสียงร้องหรือเสียงฮัมที่รับเข้ามา และหากค่าความแตกต่างอยู่ระหว่างเกณฑ์การยอมรับและเกณฑ์การปฏิเสธจะคำนวณค่าระดับความมั่นใจของคำตอบที่ได้จากการค้นคืนให้กับผู้ใช้ เกณฑ์การยอมรับและเกณฑ์การปฏิเสธจากการทดลองกับเพลงไทยประเภทต่าง ๆ แสดงได้ดังตารางที่ 4.10

ตารางที่ 4.10 เกณฑ์การยอมรับและเกณฑ์การปฏิเสธของเพลงไทยประเภทต่าง ๆ

ประเภทของเพลง	เกณฑ์การยอมรับ	เกณฑ์การปฏิเสธ
พอป	386	-
ร็อค	351	416
แจ๊ส	361	416
ฮิปฮอป	374	470
ลูกทุ่ง	330	423

จากตารางที่ 4.10 สามารถนำเกณฑ์ดังกล่าวมาประกอบการวิเคราะห์ความน่าเชื่อถือของคำตอบ หากคำตอบที่ได้ของเพลงแต่ละประเภทมีค่าความแตกต่างต่ำกว่าเกณฑ์การยอมรับจะสามารถมั่นใจได้ว่าคำตอบที่ได้ถูกต้อง 100% และหากคำตอบที่ได้มีค่าความแตกต่างสูงกว่าเกณฑ์การปฏิเสธจะสามารถมั่นใจได้ว่าการค้นคืนไม่พบคำตอบที่ถูกต้อง

4.3 การอภิปรายผล

4.3.1 การค้นคืนเพลงไทยด้วยเสียงร้อง

จากการค้นคืนเพลงไทยด้วยเสียงร้องพบว่าเมื่อนำเสียงร้องและเพลงต้นฉบับมาตัดเสียงดนตรีออก แล้วใช้ลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมล ในการเปรียบเทียบความเหมือน โดยใช้เทคนิคไดนามิกไทม์วอร์ปปิงให้ค่าความถูกต้องของการค้นคืนเพลงด้วยเสียงร้องมากที่สุดถึง 90%

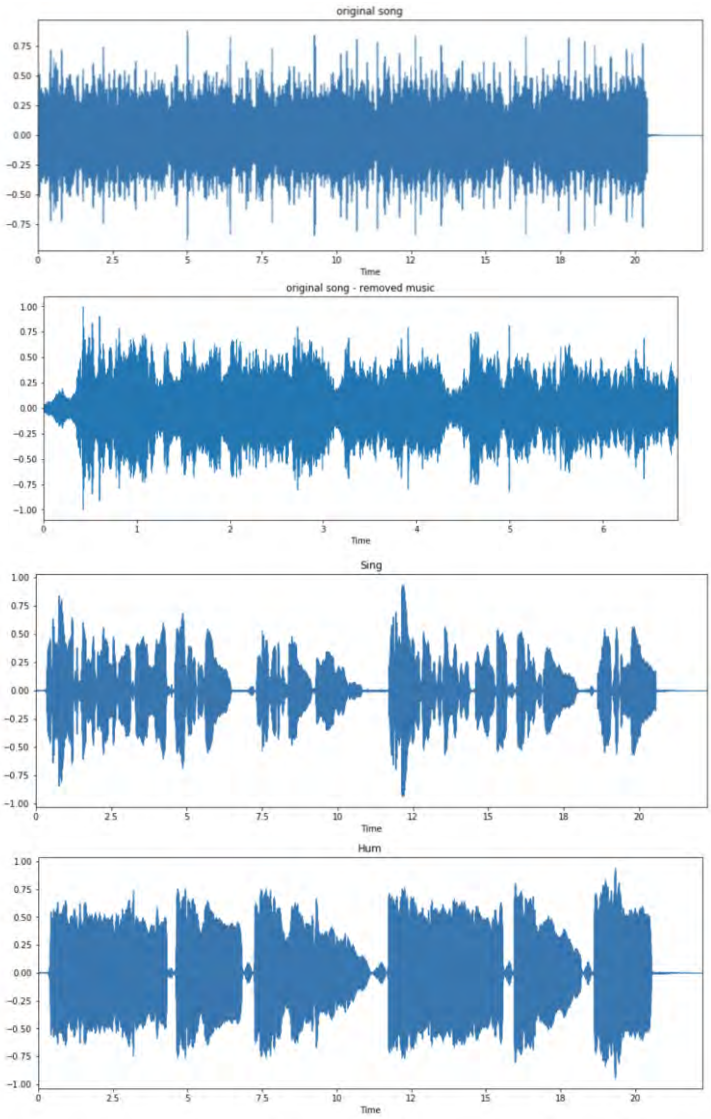
4.3.2 การค้นคืนเพลงไทยด้วยเสียงฮัม

การค้นคืนเพลงไทยด้วยเสียงฮัมพบว่าเมื่อนำเสียงฮัมและเพลงต้นฉบับมาตัดเสียงดนตรีออก แล้วใช้ลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมล ในการเปรียบเทียบความเหมือน โดยใช้เทคนิคไดนามิกไทม์วอร์ปปิง ให้ค่าความถูกต้องของการค้นคืนเพลงด้วยเสียงฮัม 39.6% และเมื่อเปรียบเทียบเสียงฮัมกับเพลงต้นฉบับในประเภทเดียวกันให้ค่าความถูกต้องเพิ่มเป็น 59.2%

เมื่อวิเคราะห์เกณฑ์การยอมรับและปฏิเสธร่วมด้วยจะสามารถรับรองผลการค้นคืนเพลงไทยด้วยเสียงฮัมในเพลงประเภทเดียวกันได้ โดยกรณีที่ค่าความเหมือนที่ได้จากการเทียบระหว่างเสียงฮัมกับเพลงต้นฉบับมีค่าน้อยกว่าหรือเท่ากับเกณฑ์การยอมรับจะได้ว่าการค้นคืนเพลงมีความถูกต้อง 100% และกรณีที่ค่าความเหมือนที่ได้จากการเทียบระหว่างเสียงฮัมกับเพลงต้นฉบับที่มีค่ามากกว่าหรือเท่ากับเกณฑ์การปฏิเสธ จะได้ว่าการค้นคืนเพลงไม่สามารถค้นคืนเพลงที่สอบถามได้

4.3.3 การเปรียบเทียบความแตกต่างระหว่างเสียงร้องและเสียงฮัม

จากการทดลองพบว่าผลการค้นคืนเพลงของเสียงร้องมีความถูกต้องมากกว่าเสียงฮัม เนื่องจากเสียงร้องมีคุณลักษณะที่มีรายละเอียดมากกว่าเสียงฮัม โดยคุณลักษณะของเสียงร้องเป็นการร้อง เป็นคำตามเนื้อร้องคล้ายกับเพลงต้นฉบับมากกว่าเสียงฮัมที่เป็นการทำระดับเสียงตามเพลงต้นฉบับเพียงอย่างเดียว แต่ไม่มีการแยกคำตามเนื้อร้อง ดังนั้นเสียงฮัมจึงมีคุณลักษณะที่ไม่ละเอียดเท่ากับเสียงร้อง แสดงดังภาพที่ 4.6 ซึ่งจะเห็นว่าเพลงต้นฉบับหลังตัดเสียงดนตรีออก มีลักษณะคลื่นเสียงคล้ายกับเสียงร้องมากกว่าเสียงฮัม มีการเว้นจังหวะและเปลี่ยนแปลงความเข้มเสียงชัดเจนกว่าเสียงฮัม ทำให้การค้นคืนเพลงด้วยเสียงร้องจึงมีความถูกต้องมากกว่าเสียงฮัม



ภาพที่ 4.6 ตัวอย่างคลื่นเสียงของเพลงรู้บ้างไหม - เครสเชนโด้ (Crescendo)

(บนลงล่าง) เพลงต้นฉบับ เพลงต้นฉบับที่ตัดเสียงดนตรี เสียงร้อง และเสียงฮัม

บทที่ 5

สรุปผลการดำเนินงานและข้อเสนอแนะ

ในบทนี้จะกล่าวถึง สรุปผลการค้นคืนเพลงไทยจากเสียงร้องและเสียงฮัม และข้อเสนอแนะโดยมีรายละเอียดดังต่อไปนี้

5.1 สรุปผลการดำเนินงาน

ในงานวิจัยการค้นคืนเพลงไทยจากเสียงร้องและเสียงฮัมฉบับนี้ ได้พบว่าการสกัดลักษณะของเสียงโดยใช้วิธีสัมประสิทธิ์เซปสตรัมความถี่เมล (Mel-Frequency Cepstral Coefficients : MFCC) กับเทคนิคไดนามิกไทม์วอร์ปปิง ได้ค่าความถูกต้องถึง 90% ซึ่งเป็นความถูกต้องที่สูงสุดในการค้นคืนเพลงด้วยเสียงร้อง และในกรณีของการค้นคืนเพลงด้วยเสียงฮัมต้องมีการพิจารณาถึงประเภทเพลงของเสียงฮัมด้วยจึงจะทำให้ได้ค่าความถูกต้องเฉลี่ยสูงถึง 58.4% ซึ่งเป็นผลมาจากค่าความถูกต้องของเสียงฮัมประเภทพอป ร็อค แจ๊ส ฮิปฮอป และลูกทุ่ง เท่ากับ 68%, 52%, 64%, 60% และ 48% ตามลำดับ ทั้งนี้ถ้าเพิ่มเกณฑ์การยอมรับและเกณฑ์การปฏิเสธเพลงคำตอบที่ค้นคืนได้ก็ช่วยเสริมความน่าเชื่อถือให้กับคำตอบที่ได้

5.2 ข้อเสนอแนะ

จากการวิจัยการค้นคืนเพลงไทยจากเสียงร้องและเสียงฮัม ผู้วิจัยเห็นว่าควรมีการทดลองเพิ่มเติมตามข้อเสนอแนะต่อไปนี้

1. การทดลองปรับค่าพารามิเตอร์ในการสกัดลักษณะแบบต่าง ๆ และการเลือกใช้ลักษณะอื่น ๆ เพื่อเพิ่มประสิทธิภาพของลักษณะเสียง และทำให้การค้นคืนเพลงมีความถูกต้องมากยิ่งขึ้น
2. การทดลองค้นคืนเพลงประเภทอื่น ๆ นอกจากเพลงห้าประเภทที่ทดลองในงานวิจัยชิ้นนี้ เพื่อเพิ่มประสิทธิภาพของการค้นคืนเพลงให้สามารถค้นคืนเพลงในขอบเขตประเภทเพลงที่หลากหลายมากขึ้นได้
3. การทดลองทำนายประเภทของเพลงจากเสียงร้องและเสียงฮัม และจำแนกเสียงร้องและเสียงฮัมเพื่อลดขั้นตอนในการค้นคืนเพลงของผู้ใช้ และทำให้การค้นคืนเพลงจากการร้องและการฮัมมีความสะดวกมากขึ้น
4. การพัฒนาวิธีการค้นคืนเพลงจากการร้องและการฮัมที่ทนทานต่อสภาพแวดล้อมที่มีเสียงรบกวนได้มากขึ้น เพื่อเพิ่มประสิทธิภาพของการค้นคืนเพลงให้มีความถูกต้องมากขึ้น

5. งานวิจัยนี้ยังสามารถนำไปต่อยอดพัฒนาเป็นโปรแกรมค้นคืนเพลงจากการร้องและการฮัม ที่สามารถรับข้อมูลเข้าผ่านการร้องและการฮัมโดยตรงได้ เพื่อให้ผู้ใช้สามารถค้นคืนเพลงได้สะดวกมากขึ้น

รายการอ้างอิง

[1] Shazam. [https://www.shazam.com]. เข้าถึงเมื่อ 7 พฤศจิกายน พ.ศ. 2562

[2] Soundhound. [https://www.soundhound.com/soundhound]. เข้าถึงเมื่อ 7 พฤศจิกายน พ.ศ. 2562

[3] Open Unmix. [https://sigsep.github.io/open-unmix]. เข้าถึงเมื่อ 12 กุมภาพันธ์ พ.ศ. 2563

[4] Dynamic time warping. [https://pypi.org/project/dtw-python]. เข้าถึงเมื่อ 12 กุมภาพันธ์ พ.ศ. 2563

[5] Kichul Kim, Kang Ryoung Park, Sung-Joo Park, Soek-Pil Lee, and Moo Young Kim, “Robust Query-by-Singing/Humming System against Background Noise Environments”,IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 720-725, 2011

[6] Rifki Afina Putri and Dessi Puji Lestari, “Music Information Retrieval Using Query-by-Humming Based on the Dynamic Time Warping”, in Proceedings of the 5th International Conference on Electrical Engineering and Informatics 2015, Denpasar, Indonesia, pp. 65-70, 2015

[7] Hung-Ming Yu, Wei-Ho Tsai, and Hsin-Min Wang, “A Query-by-Singing Technique for Retrieving Polyphonic Objects of Popular Music”, in Proceedings of the Second Asia Information Retrieval Symposium, Jeju Island, Korea,
pp. 439-453, 2005

[8] Marta Jaczyńska, Piotr Bobiński and Agnieszka Pietrzak, “Music Recognition Algorithms Using Queries by Example”, in Proceedings of the 2018 Joint Conference - Acoustics, Ustka, Poland, pp. 108-111, 2018

[9] Nastaran Borjian, “Query-by-Example Music Information Retrieval by Score-Based Genre Prediction and Similarity Measure”, International Journal of Multimedia Information Retrieval, Vol. 6, No. 2, pp. 155–166, 2017

[10] Tomasz Maka and Piotr Dziurzanski, “Parallel Audio Features Extraction for Sound Indexing and Retrieval Systems”, in Proceedings of the Electronics in Marine-2013, Zadar, Croatia, pp. 185-189, 2013

[11] Yu Song, Wen-Hong Wang, and Feng-Juan Guo, “Feature Extraction and Classification for audio information in news video”, in Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition, Baoding, China, pp. 43-46, 2009

[12] Priyanka Bansal, Dr. Syed Akhtar Imam and Roma Bharti, “Speaker Recognition using MFCC, shifted MFCC with Vector Quantization and Fuzzy”, International Conference on Soft Computing Techniques and Implementations- (ICSCTI), Faridabad, India, pp. 41-44, 2015

[13] Chung-Che Wang and Jyh-Shing Roger Jang, “Improving Query-by-Singing/Humming by Combining Melody and Lyric Information”, IEEE/ACM Transaction on audio, speech and language processing , Vol. 23, No. 4, pp. 798-806, 2015

[15] Librosa. [https://librosa.github.io/librosa/feature.html]. เข้าถึงเมื่อ 7 เมษายน พ.ศ. 2563

[16] Essentia. [https://essentia.upf.edu/reference/streaming_MFCC.html]. เข้าถึงเมื่อ 7 เมษายน พ.ศ. 2563

ภาคผนวก

ภาคผนวก ก

แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal

ปีการศึกษา 2562

ชื่อโครงงาน (ภาษาไทย) การค้นคืนหลายรูปแบบของเพลงไทยจากการร้องและการฮัม

ชื่อโครงงาน (ภาษาอังกฤษ) Multimodal Retrieval of Thai Music from Singing and Humming

อาจารย์ที่ปรึกษา 1. ผู้ช่วยศาสตราจารย์ ศศิภา พันธุวดีธร

2. ผู้ช่วยศาสตราจารย์ ดร.ศุภกานต์ พิมลธเรศ

ผู้ดำเนินการ 1. นางสาวขวัญชนก ศิลปพิบูลย์ เลขประจำตัวนิสิต 5933606023

2. นางสาวณีญาภรณ์ เพชรปาน เลขประจำตัวนิสิต 5933624323

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

เสียงเพลงเป็นสิ่งใกล้ตัวที่อยู่ในชีวิตประจำวันของเรา อาจเป็นเพลงที่รู้จัก หรือเป็นเพลงที่ไม่รู้จักและบังเอิญได้ยิน ซึ่งอาจทำให้เกิดความสงสัย และต้องการค้นหาว่าเพลงนี้คือเพลงอะไร ร้องโดยศิลปินคนใด ในปัจจุบันมีแอปพลิเคชันค้นหาเพลงจากเสียง เช่น Shazam[1] , Soundhound[2] เป็นต้น โดยแอปพลิเคชันเหล่านี้ค้นหาเพลงด้วยวิธีแปลงเสียงข้อมูลรับเข้าเป็นลายพิมพ์เสียง (audio fingerprint) แล้วนำไปค้นหาลายพิมพ์เสียงที่ตรงกันในฐานข้อมูลของแอปพลิเคชัน แล้วเล่นเพลงนั้น แต่แอปพลิเคชันเหล่านี้ยังไม่สามารถค้นหาเพลงไทยได้อย่างมีประสิทธิภาพ และยังไม่สามารถค้นหาเพลงจากเสียงร้องและเสียงฮัมของผู้ใช้ได้ ดังนั้นผู้พัฒนาจึงอยากพัฒนาการค้นหาเพลงจากการร้องและฮัมที่สามารถใช้กับเพลงไทยได้

การศึกษาการสอบถามด้วยการร้องหรือการฮัม (Query by Singing/Humming : QBSH) จากงานวิจัยของ Kichul Kim และคณะ [3] ,งานวิจัยของ Rifki Afina Putri และ Dessi Puji Lestari [4] และงานวิจัยของ Hung-Ming Yu และคณะ [5] เสนอระบบสกัดระดับเสียง (pitch extraction) โดยนำเอาเวกเตอร์ของระดับเสียง จากข้อมูลต้นแบบที่มีนามสกุลเป็นมิดิ (.MIDI) มาหาค่าความคล้ายด้วยวิธี Dynamic Time Warping matching ซึ่ง [3] จะให้ผลลัพธ์เป็นเพลงต้นแบบที่ตรงกับข้อมูลนำเข้ามากที่สุดแต่ [4] เปรียบเทียบความคล้ายของข้อมูลเสียงในอนุกรมเวลาที่มีความคลาดเคลื่อนไปจากต้นแบบซึ่งหลังจากทำ DTW แล้วจะได้ผลลัพธ์ออกมาเป็นความแตกต่างที่น้อยที่สุดและนำมาตั้งเกณฑ์ เพื่อป้องกันไม่ให้ระบบ QBSH ตอบเพลงที่ไม่ใช่คำตอบออกมาให้กับผู้ใช้และ [5] ได้ใช้การเปลี่ยนรูปแบบ (transposition) ของข้อมูลนำเข้าของผู้ใช้งานที่เป็นลำดับโน้ต เข้ามาร่วมกับการใช้ DTW ในการหาความแตกต่าง

ที่น้อยที่สุดเหมือนกันเพื่อแก้ไขกรณีที่ผู้ใช้ร้องเพลงไม่ตรงกับคีย์ของต้นแบบ แล้วให้ผลลัพธ์เป็นรายชื่อเพลงที่มีความคล้ายเพลงต้นแบบมากที่สุด N อันดับ (Top-N)

ในการศึกษาเกี่ยวกับการสกัดคุณลักษณะของเสียง (audio feature extraction) จากงานวิจัยของ Marta Jaczyńska และคณะ [6] ศึกษาการรู้จำเพลง (music recognition) ด้วยวิธีการสกัดระดับเสียง [7] แบบต่าง ๆ พบว่าตัวประมาณค่าความถี่พื้นฐานของเสียงพูดและระดับเสียงที่ชื่อว่า YIN สามารถสกัดระดับเสียงได้ดีที่สุด รวมถึงขั้นตอนวิธี FFT YIN ที่นำ Fast Fourier transform (FFT) มาประยุกต์ ทำให้ใช้เวลาสกัดระดับเสียงน้อยลงถึง 3 เท่า โดยความถูกต้องของระดับเสียงที่สกัดได้ยังอยู่ในระดับเดิม และงานวิจัยของ Tomasz Maka และ Piotr Dziurzansk [8] ศึกษาการสกัดคุณลักษณะของเสียงหลากหลายแบบ โดยจำแนกเป็นกลุ่มต่างๆ เช่น Mel-Frequency Cepstral Coefficients (MFCC), Fast Fourier Transform (FFT), กลุ่มที่มีเวลาเป็นฐาน(time-based features) เช่น Zero Crossing Rate (ZCR) [9] และ Short-Time Energy (STE) และกลุ่มที่มีความถี่เป็นฐาน(frequency-based features) เช่น spectral centroid, spectral roll-off และ bandwidth

ในการศึกษาเกี่ยวกับการวัดค่าความคล้าย (Similarity measure) ของคุณลักษณะต่าง ๆ ของเสียง งานวิจัยของ Nastaran Borjian [7] ศึกษาการวัดความคล้ายของลายพิมพ์เสียงจากข้อมูลนำเข้า และข้อมูลต้นแบบในฐานข้อมูล ซึ่งใช้เป็นเกณฑ์สำหรับการค้นคืนเพลงที่ใกล้เคียงกับข้อมูลนำเข้าที่สุด

นอกจากนี้งานวิจัยของ Chung-Che wang และคณะ [10] ศึกษาการรู้จำทำนองโดยนำ Linear Scaling มาช่วยในการค้นคืนเพลงจากข้อมูลนำเข้าที่เป็นเสียงจากการร้องและการฮัม แต่วิธีนี้ใช้เวลาในการคำนวณสูง ผู้วิจัยจึงเพิ่มการวิเคราะห์ความคล้ายของคำร้อง (lyrics) จากข้อมูลนำเข้าที่เป็นเสียงจากการร้อง ด้วยเทคนิคแปลงเสียงพูดเป็นตัวอักษร (speech recognition techniques) แล้วเปรียบเทียบกับคำร้องในฐานข้อมูล เมื่อนำ 2 วิธีข้างต้นมาผนวกกันแล้ว สามารถเพิ่มอัตราการรู้จำของระบบโดยรวม

จากที่กล่าวมาข้างต้น ทำให้ผู้พัฒนามีความสนใจที่จะศึกษาเกี่ยวกับการค้นคืนเพลงไทยจากการร้องและการฮัม เพื่อช่วยให้ผู้ใช้มีเครื่องมือช่วยค้นหาเพลงไทยได้สะดวกยิ่งขึ้น รวมไปถึงการแนะนำเพลงไทยที่มีความใกล้เคียงกับเพลงที่ผู้ใช้ค้นหา โดยระบบที่พัฒนาจะรับข้อมูลนำเข้าเป็นเสียงจากการร้องหรือการฮัม และนำมาสกัดคุณลักษณะเสียงซึ่งจะได้เป็นทำนองของการร้องและการฮัม เพื่อใช้ค้นหาเพลงที่มีความคล้ายกันกับข้อมูลเพลงต้นแบบในชุดข้อมูลที่เตรียมไว้แสดงเป็นผลลัพธ์ นอกจากนี้ระบบยังรับข้อมูลนำเข้าที่เป็นเสียงร้อง เพื่อนำมาแปลงเป็นคำร้อง สำหรับการเทียบความคล้ายของคำร้องกับข้อมูลคำร้องของเพลงต้นแบบในชุดข้อมูลที่เตรียมไว้ แล้วนำค่าความคล้ายทั้งหมดมาวิเคราะห์ ซึ่งจะช่วยให้ได้ผลลัพธ์ที่แม่นยำยิ่งขึ้น

วัตถุประสงค์

เพื่อพัฒนาเครื่องมือช่วยค้นหาเพลงไทยและแนะนำเพลงที่ใกล้เคียงกับเพลงที่ค้นหาซึ่งผู้ใช้อาจสนใจได้

ขอบเขตของโครงงาน

1. เพลงไทย 5 ประเภทได้แก่ เพลงพอป เพลงร็อค เพลงลูกทุ่ง เพลงแจ๊สและเพลงฮิปฮอป ประเภทละ 50 เพลง รวม 250 เพลง โดยเพลงแต่ละเพลงจะเป็นเพลงประเภทใดประเภทหนึ่งเท่านั้น
2. ชุดข้อมูลเก็บท่อนฮุคของเพลงต้นแบบซึ่งประกอบด้วยข้อมูลเนื้อร้อง และข้อมูลเสียงชนิดมิดิ (.MIDI)
3. ท่อนฮุค คือ ท่อนของเนื้อร้องที่ปรากฏซ้ำกันในเพลง ซึ่งทำให้ผู้ฟังจำเพลงนั้นได้ง่าย ในโครงงานนี้ความยาวของท่อนฮุคตั้งแต่ 10 ถึง 15 วินาที สำหรับท่อนฮุคที่มีความยาวมากกว่า 15 วินาทีจะตัดเฉพาะส่วนต้นของท่อนฮุคเพียงแค่ 15 วินาทีเท่านั้น
4. ข้อมูลนำเข้าเป็นเสียงร้องหรือเสียงฮัมที่เริ่มจากส่วนต้นของท่อนฮุคของเพลง โดยมีความยาวเท่ากับเพลงต้นแบบในชุดข้อมูล และการร้องและฮัมจะทำด้วยเสียงของบุคคลเดียว โดยไม่มีเสียงดนตรีหรือเสียงอื่นประกอบ

วิธีการดำเนินงาน

แผนการศึกษา

1. ศึกษาการค้นคืนเพลงจากการร้องและการฮัม
2. ศึกษางานวิจัยที่เกี่ยวข้องกับการการค้นคืนเพลงจากการร้องและการฮัม
3. ศึกษาประสิทธิภาพของขั้นตอนวิธีกับคุณลักษณะต่าง ๆ ของเพลงไทย
4. ออกแบบระบบและพัฒนาโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม
5. ทดสอบประสิทธิภาพโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม
6. ประเมินและอภิปรายผล
7. จัดทำเอกสาร

ระยะเวลาที่ศึกษา

ขั้นตอน	ปี 2562					ปี 2563			
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาการค้นคืนเพลงจากการร้องและการฮัม									
2. ศึกษางานวิจัยที่เกี่ยวข้องกับการการค้นคืนเพลงจากการร้องและการฮัม									
3. ศึกษาประสิทธิภาพของขั้นตอนวิธีกับคุณลักษณะต่าง ๆ ของเพลงไทย									
4. ออกแบบระบบและพัฒนาโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม									
5. ทดสอบประสิทธิภาพโปรแกรมค้นคืนเพลงไทยจากการร้องและการฮัม									
6. ประเมินและอภิปรายผล									
7. จัดทำเอกสาร									

ประโยชน์ที่คาดว่าจะได้รับ

1. ประโยชน์ด้านความรู้และประสบการณ์ต่อนิสิต
   1.1 ได้ความรู้ ความเข้าใจเกี่ยวกับการพัฒนาระบบ
   1.2 ได้ฝึกการคิดวิเคราะห์ การทำงานอย่างมีระบบ และการทำงานเป็นกลุ่ม

1.3 ได้รับความรู้เรื่องการทำงานที่เกี่ยวข้องกับเสียงเพลงมากขึ้น

1.4 ได้รับความรู้เกี่ยวกับการเขียนโปรแกรมและการพัฒนาระบบค้นคืนเพลงไทย

2. ประโยชน์ที่ได้จากโครงงานที่พัฒนาขึ้น

2.1 ได้เครื่องมือที่ช่วยในการค้นคืนเพลงไทยได้อย่างมีประสิทธิภาพมากขึ้น

2.2 ได้เครื่องมือที่สามารถแนะนำเพลงที่ใกล้เคียงกับเพลงที่ต้องการค้นหาได้

อุปกรณ์และเครื่องมือที่ใช้

1. ฮาร์ดแวร์

คอมพิวเตอร์ส่วนตัวที่มีหน่วยประมวลผลกลางความเร็วไม่ต่ำกว่า 1.4 GHz หน่วยความจำหลักความจุไม่ต่ำกว่า 8 GB และหน่วยความจำสำรองความจุไม่ต่ำกว่า 128 GB จำนวน 2 เครื่อง

2. ซอฟต์แวร์

2.1 ระบบปฏิบัติการ Windows หรือ macOS

2.2 Google Colab Python notebook

2.3 ไลบรารี Python สำหรับการทำงานและคำนวณข้อมูล เช่น librosa, numpy, pandas, sklearn, scipy

งบประมาณ

1.	เอสเอสดีพกพา ความจุ 1 TB	2 ชิ้น 11,180	บาท
2.	กระดาษเอ4 จำนวน 500 แผ่น	120	บาท
3.	ค่าเข้าเล่มเอกสาร	300	บาท
	รวม	**11,600**	**บาท**

เอกสารอ้างอิง

[1] Shazam. [https://www.shazam.com]. เข้าถึงเมื่อ 7 พฤศจิกายน พ.ศ. 2562

[2] Soundhound. [https://www.soundhound.com/soundhound]. เข้าถึงเมื่อ 7 พฤศจิกายน พ.ศ. 2562

[3] Kichul Kim, Kang Ryoung Park, Sung-Joo Park, Soek-Pil Lee, and Moo Young Kim, “Robust Query-by-Singing/Humming System against Background Noise Environments”,IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 720-725, 2011

[4] Rifki Afina Putri and Dessi Puji Lestari, “Music Information Retrieval Using Query-by-Humming Based on the Dynamic Time Warping”, in Proceedings of the 5th International Conference on Electrical Engineering and Informatics 2015, Denpasar, Indonesia, pp. 65-70, 2015

[5] Hung-Ming Yu, Wei-Ho Tsai,and Hsin-Min Wang, “A Query-by-Singing Technique for Retrieving Polyphonic Objects of Popular Music”, in Proceedings of the Second Asia Information Retrieval Symposium, Jeju Island, Korea,
pp. 439-453, 2005

[6] Marta Jaczyńska, Piotr Bobiński and Agnieszka Pietrzak, “Music Recognition Algorithms Using Queries by Example”, in Proceedings of the 2018 Joint Conference - Acoustics, Ustka, Poland, pp. 108-111, 2018

[7] Nastaran Borjian, “Query-by-Example Music Information Retrieval by Score-Based Genre Prediction and Similarity Measure”, International Journal of Multimedia Information Retrieval, Vol. 6, No. 2, pp. 155–166, 2017

[8] Tomasz Maka and Piotr Dziurzanski, “Parallel Audio Features Extraction for Sound Indexing and Retrieval Systems”, in Proceedings of the Electronics in Marine-2013, Zadar, Croatia, pp. 185-189, 2013

[9] Yu Song,Wen-Hong Wang, and Feng-Juan Guo, “Feature Extraction and Classification for audio information in news video”, in Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition, Baoding, China, pp. 43-46, 2009

[10] Chung-Che Wang and Jyh-Shing Roger Jang, “Improving Query-by-Singing/Humming by Combining Melody and Lyric Information”, IEEE/ACM Transaction on audio, speech and language processing , Vol. 23, No. 4, pp. 798-806, 2015

ประวัติผู้เขียน

ชื่อ นางสาวขวัญชนก ศิลปพิบูลย์
รหัสนิสิต 5933606023
Email ksillapapibul@gmail.com
กำลังศึกษาหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ชื่อ นางสาวณีญาภรณ์ เพชรปาน
รหัสนิสิต 5933624323
อีเมล neeyabhorn@gmail.com
กำลังศึกษาหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย