



รายงานสหกิจศึกษาฉบับสมบูรณ์

การจำแนกประเภทเสียงโดยใช้ปัญญาประดิษฐ์
Sound Classification by using AI

นางสาวณิชามน อ้อกุล

ภาควิชา วิศวกรรมดนตรีและสื่อประสม
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาแล๓ต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อโครงการสหกิจศึกษา การจัดประเภทเสียงโดยใช้ปัญญาประดิษฐ์

ชื่อ-สกุล นักศึกษา ณิชามน อ้อกุล

คณะ วิศวกรรมศาสตร์ ภาควิชา วิศวกรรมดนตรีและสื่อประสม

ชื่อ-สกุล อาจารย์นิเทศ ดร.ณชนันท์ ชิตานนท์

ชื่อ-สกุล ผู้นิเทศงาน เจตพล สุทธิประภา

สถานประกอบการ หจก.ห้องบันทึกเสียงรามอินทรา

บทคัดย่อ

รายงานสหกิจศึกษาฉบับนี้เป็นส่วนหนึ่งของรายงานวิชาสหกิจศึกษาที่กำหนดให้นักศึกษาทำโครงการร่วมกับสถานประกอบการตลอดระยะเวลาหนึ่งภาคเรียนการศึกษา โดยสถานประกอบการที่ได้รับมอบหมายคือห้างหุ้นส่วนจำกัดห้องบันทึกเสียงรามอินทรา เป็นสถานประกอบการที่ทำงานเกี่ยวกับการทำเสียงประกอบภาพยนตร์ โฆษณา ซีรีส์ และอื่น ๆ อีกมากมาย

โดยรายงานฉบับนี้จะอธิบายถึงการจัดจำแนกประเภทของเสียงเอฟเฟค (sound effect) และเสียงประดิษฐ์ (sound design) โดยใช้ปัญญาประดิษฐ์เป็นเครื่องมือการทำงาน จุดประสงค์เพื่อเป็นต้นแบบที่สามารถนำไปพัฒนาต่อเพื่อเป็นการลดระยะเวลาการค้นหาเสียงที่ต้องการใช้งานและสะดวกเร็วในการจัดการเสียงของสถานประกอบการได้มากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Cooperative Title: Sound Classification by using AI

Student intern name: Nichamon Uekun

Faculty: Engineering

Department: Music Engineering and Multimedia

Advisor name: Dr.Natchanant Chaitanont

Mentor name: Jatetapon Suttiprapa

Company: Ramindra Recording Studio

ABSTRACT

This cooperative education report is a part of cooperative education program that assigned the student to collaborate a project with the company within one semester. The company cooperating this project is Ramindra Recording Studio, their company create sound for movies, advertise, series and a lot more.

This report will describe about classifying sound effects and sound designs by using artificial intelligent or AI. The main point of this project is being a prototype that can be developed to reduce time in classifying the sound in which category they are in .

กิตติกรรมประกาศ

การที่ข้าพเจ้าได้มาปฏิบัติงานสหกิจศึกษา ณ ห้างหุ้นส่วนจำกัดห้องบันทึกเสียงรามอินทรา ตั้งแต่ วันที่ 5 สิงหาคม พ.ศ. 2562 ถึงวันที่ 22 พฤศจิกายน พ.ศ. 2562 ส่งผลให้ข้าพเจ้าได้รับความรู้และ ประสบการณ์ต่างๆ ที่มีคุณค่ามากมาย สำหรับรายงานวิชาสหกิจศึกษานี้ สำเร็จลงได้ด้วยดีจากความ ร่วมมือและสนับสนุนจากหลายฝ่ายดังนี้

1. คุณเจตพล สุทธิประภา ซึ่งเป็น พี่เลี้ยง Sound effect, Sound design
2. คุณจิรายุทธ ฉัตรเพชร ซึ่งเป็นผู้ที่ให้คำปรึกษาเรื่องการทำงานของ AI
3. คุณคณาธิป ภูมลา ซึ่งเป็นผู้ให้คำแนะนำเรื่องการเขียนโปรแกรม
4. ดร.ณัชนันท์ ชิตานนท์ ซึ่งเป็น อาจารย์ที่ปรึกษา

รวมถึงบุคลากรท่านอื่น ๆ ที่ไม่ได้กล่าวนามทุกท่านที่ได้ให้คำแนะนำช่วยเหลือในการจัดทำรายงานในครั้งนี้

ข้าพเจ้าขอขอบพระคุณผู้มีส่วนเกี่ยวข้องทุกท่านที่มีส่วนร่วมในการให้ข้อมูล เป็นที่ปรึกษา ในการ ทำรายงานฉบับนี้จนเสร็จสมบูรณ์ ตลอดจนให้การดูแลและให้ความเข้าใจเกี่ยวกับชีวิตของการทำงานจริง ข้าพเจ้าขอขอบคุณไว้ ณ ที่นี้

ณิชามน อ้อกุล
ผู้จัดทำรายงาน
24 ตุลาคม 2562

สารบัญ

บทคัดย่อ	I
ABSTRACT	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	V
สารบัญรูป	VI
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญ	1
1.2 วัตถุประสงค์ของโครงการสหกิจ	1
1.3 ขอบเขตของโครงการสหกิจ	1
1.4 วิธีการดำเนินงาน	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 แนวคิดทฤษฎีและงานวิจัย	3
2.1 ความรู้เกี่ยวกับ ปัญญาประดิษฐ์ (Artificial Intelligence: AI).....	3
2.2 การจัดการกับข้อมูลในรูปแบบของเสียง	6
บทที่ 3 วิธีการดำเนินงาน.....	7
3.1 การเตรียมข้อมูล.....	8
3.3 การทดสอบโปรแกรม.....	11
บทที่ 4 ผลของการทำโครงการ	12
บทที่ 5 สรุปผลโครงการและข้อเสนอแนะ	13
เอกสารอ้างอิง.....	14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาแล IV ของอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ 2.1 : ตารางแสดงข้อมูลสำหรับ Supervised learning.....	4
ตารางที่ 2.2 : ตารางแสดงข้อมูลสำหรับ Unsupervised learning	5
ตารางที่ 2.3 : ตารางแสดงผลลัพธ์สำหรับ Unsupervised learning	5
ตารางที่ 3.1 : ตารางแสดงขั้นตอนการทำงาน	7



สารบัญรูป

ภาพที่ 2.1 : ภาพการเปรียบเทียบระหว่างการเขียนโปรแกรมแบบเดิมและการทำ Machine learning ...	4
ภาพที่ 2.2 : ภาพตัวกรองเมลสเกล	6
ภาพที่ 3.1 : ภาพแผนผังแสดงขั้นตอนการทำงาน	7
ภาพที่ 3.2 : ภาพตัวอย่างของคลื่นสัญญาณเสียงโทรศัพท์	8
ภาพที่ 3.3 : ภาพตัวอย่างของคลื่นสัญญาณเสียงต๋อย	9
ภาพที่ 3.4 : ภาพแผนผังการทำงานของโปรแกรม.....	9
ภาพที่ 4.1 : ภาพของผลลัพธ์การทดสอบของโปรแกรม	12



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

ห้างหุ้นส่วนจำกัดห้องบันทึกเสียงรามอินทราเป็นสถานประกอบการที่ผลิตงานด้านเสียงประกอบภาพยนตร์ คือ เสียงที่เกิดจากมนุษย์ (foley), บทพูด (dialogue) และ เสียงบรรยากาศ (ambience) และ เสียงเอฟเฟค (sound effect) รวมไปถึงการมิชเสียงทั้งหมดในภาพยนตร์ด้วย

สถานประกอบการที่ทำงานด้านเสียงในภาพยนตร์การทำงานส่วนหนึ่งคือการใส่เสียงเอฟเฟค (sound effect) และ เสียงประดิษฐ์ (sound design) การทำงานในด้านนี้จะต้องทำงานกับเสียงจำนวนมากและหลากหลายประเภทหลากหลายหมวดหมู่เพื่อนำมาประกอบให้ภาพยนตร์มีความสมจริงและสนุกมากยิ่งขึ้น จึงทำให้เกิดข้อสังเกตถึงปัญหาและจุดที่สามารถปรับปรุงในการทำงาน คือการที่เสียงที่ใช้ในการทำงานมีบางเสียงที่ยังไม่ถูกระบุชื่อว่าเป็นเสียงอะไรมีไฟล์และโพลเดอร์ที่กระจายทำให้การหยิบจับนำไปใช้งานได้ไม่เต็มประสิทธิภาพ ถ้าต้องนำไปใช้งานจะต้องฟังทีละเสียงเพื่อหาเสียงที่ต้องการแทนที่จะค้นหาจากชื่อได้เลย จึงเกิดเป็นโครงการงานสหกิจนี้ขึ้นมาเพื่อหาวิธีพัฒนาและแก้ปัญหาให้สามารถใช้งานเสียงทั้งหมดที่มีอย่างเต็มประสิทธิภาพให้ได้มากที่สุด

1.2 วัตถุประสงค์ของโครงการสหกิจ

1. สร้างต้นแบบโปรแกรมเพื่อช่วยคัดแยกหมวดหมู่เสียง
2. ลดเวลาในการจัดหมวดหมู่ของเสียง
3. สามารถนำโปรแกรมต้นแบบไปพัฒนาต่อได้ให้ละเอียดและรวดเร็วมากยิ่งขึ้น

1.3 ขอบเขตของโครงการสหกิจ

ขอบเขตของโครงการนี้คือการใช้การเรียนรู้ของเครื่อง (Machine Learning) กระบวนการหนึ่งของ ปัญญาประดิษฐ์ (Artificial Intelligence: AI) ในการแยกหมวดหมู่หรือประเภทของเสียง โดยเบื้องต้นคือการแยกเพียง 2 เสียงคือเสียงต๋อยและเสียงโทรศัพท์เพื่อเป็นการง่ายต่อการคัดแยกหมวดหมู่ (classification) เพราะสองเสียงนี้มีความแตกต่างกันอย่างชัดเจนโปรแกรมจะได้มีความแม่นยำและถูกต้องที่สุด

1.4 วิธีการดำเนินงาน

1. ศึกษาทฤษฎีและการทำงานของ ปัญญาประดิษฐ์ (Artificial Intelligence: AI)
2. เตรียมข้อมูลที่จะต้องใช้ในการดำเนินการ
3. สร้างโปรแกรม
4. ทดสอบโปรแกรมและความแม่นยำ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เข้าใจการทำงานของ ปัญญาประดิษฐ์ (Artificial Intelligence: AI)
2. เข้าใจการเขียนโปรแกรมมากขึ้น
3. เข้าใจทฤษฎีการทำงานของเสียงมากขึ้น
4. มีความรับผิดชอบต่องานที่ได้รับมอบหมายและมีการคิด วิเคราะห์ อย่างเป็นระบบ



บทที่ 2

แนวคิดทฤษฎีและงานวิจัย

ในบทนี้จะกล่าวถึงทฤษฎีแนวคิดและหลักการพื้นฐานของ ปัญญาประดิษฐ์ (Artificial Intelligence: AI) และการการจำแนกประเภทเสียง (Sound Classification) เริ่มจากความรู้เบื้องต้นเกี่ยวกับ AI (Artificial Intelligence) หรือในชื่อภาษาไทยคือ ปัญญาประดิษฐ์ ซึ่งเป็นส่วนสำคัญที่จะถูกพูดถึงและนำมาใช้งานอย่างมากในโครงงานสหกิจฉบับนี้ เมื่อมีความเข้าใจถึงทฤษฎีและหลักการของปัญญาประดิษฐ์แล้วจะสามารถเลือกขั้นตอนการทำงานที่เหมาะสมกับการทำ การการจำแนกประเภทเสียง (Sound Classification) ได้มากยิ่งขึ้น

ทฤษฎีและหลักการที่ควรรู้เพื่อนำมาใช้งานอีกส่วนหนึ่งคือทฤษฎีในการนำข้อมูลของเสียงมาใช้งานกับการจัดจำแนกหมวดหมู่ การศึกษาทฤษฎีของข้อมูลของเสียงจะทำให้สามารถจัดการข้อมูลของเสียงในรูปแบบของงานทำงานที่งานที่สุดในการนำไปใช้งานต่อ

2.1 ความรู้เกี่ยวกับ ปัญญาประดิษฐ์ (Artificial Intelligence: AI)

AI (Artificial Intelligence) หรือในชื่อภาษาไทยคือ ปัญญาประดิษฐ์ ตามความหมายของคณะกรรมการจัดทำพจนานุกรมศัพท์คอมพิวเตอร์และเทคโนโลยีสารสนเทศ มติเมื่อ 28 มิถุนายน 2562 มีบทนิยามหรืออธิบายความหมายว่า สาขาหนึ่งของวิทยาการคอมพิวเตอร์ ซึ่งเน้นเรื่องที่จะทำให้คอมพิวเตอร์ทำงานได้ใกล้เคียงมนุษย์มากขึ้น แบ่งย่อยออกเป็นสาขาต่าง ๆ เช่น การแปลภาษาด้วยเครื่อง ระบบผู้เชี่ยวชาญ วิทยาการหุ่นยนต์ การรู้จำแบบ การรับรู้เสียงมนุษย์ ฯลฯ [1] ความหมายอย่างง่ายคือ AI คือ Machine หรือคอมพิวเตอร์ที่ทำงานเลียนแบบพฤติกรรมมนุษย์โดยเฉพาะการทำงานของสมองมนุษย์ที่สามารถช่วยเหลือและอำนวยความสะดวกด้านต่าง ๆ ได้

ส่วนสำคัญของปัญญาประดิษฐ์จะถูกแบ่งเป็น 5 ส่วนหลัก ๆ คือ [2]

1. คอมพิวเตอร์วิทัศน์ (Computer vision)
เป็นการศึกษาเรื่องการมองเห็น จดจำภาพ มีสาขาย่อยเช่น การประมวลผลภาพ (Image processing)
2. การประมวลผลภาษาธรรมชาติ (Natural language processing)
เป็นการศึกษาการแปลความหมายจากภาษามนุษย์ มาเป็นความรู้ที่เครื่องจักรเข้าใจได้ สาขานี้เกี่ยวข้องกับใกล้ชิดกับ ภาษาศาสตร์เชิงคำนวณ (Computational Linguistics)
3. การแทนความรู้ (Knowledge representation)
เป็นการศึกษาด้านการเก็บความรู้ (Knowledge) ไว้ในเครื่องจักรโดยประเด็นสำคัญคือ การประหยัดหน่วยความจำ นำความรู้ไปใช้ในการให้เหตุผล การเรียนรู้ความรู้ใหม่ด้วยเทคนิคการเรียนรู้ของเครื่อง

4. ระบบผู้เชี่ยวชาญ (Expert system)

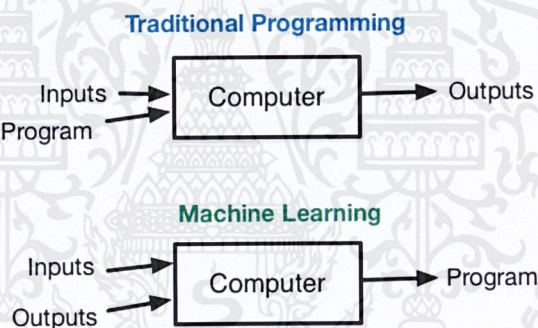
เป็นการศึกษาเรื่องสร้างระบบความรู้ของปัญหาเฉพาะอย่าง เช่น การแพทย์หรือวิทยาศาสตร์ จุดประสงค์ของระบบคือ ทำให้เสมือนมีมนุษย์ผู้เชี่ยวชาญคอยให้คำปรึกษาและคอยตอบปัญหา

5. การเรียนรู้ของเครื่อง (Machine learning)

เป็นการศึกษากระบวนการเรียนรู้ เพื่อให้เครื่องจักรสามารถเรียนรู้สิ่งใหม่ ๆ ได้คล้ายมนุษย์ โดยการให้ข้อมูลเครื่องจักรได้เรียนรู้และสามารถนำไปใช้กับข้อมูลใหม่ ๆ ได้ นี่คือนี่ส่วนที่เราจะศึกษารายละเอียดเพิ่มเติมเพื่อนำไปใช้กับโครงงานสหกิจของเราในครั้งนี้

2.1.1 การเรียนรู้ของเครื่องจักร (Machine Learning)

การเรียนรู้ของเครื่องจักร (Machine Learning) ส่วนสำคัญส่วนหนึ่งของ ปัญญาประดิษฐ์ ซึ่งคือกระบวนการเปลี่ยนแนวคิดของการเขียนโปรแกรม ถ้าหากเปรียบเทียบกับเขียนโปรแกรมแบบเดิมคือการที่เราต้องการให้โปรแกรมแสดงผลที่เราต้องการโดยที่เราป้อนข้อมูลเข้าไปแต่การทำ Machine learning คือการที่เราป้อนข้อมูลและผลลัพธ์ที่ต้องการเข้าไปเพื่อให้คอมพิวเตอร์แสดงผลที่เป็นโปรแกรมออกมา และโปรแกรมนี้อาจสามารถนำไปประยุกต์ใช้กับข้อมูลใหม่ ๆ ได้



ภาพที่ 2.1 ภาพการเปรียบเทียบระหว่างการเขียนโปรแกรมแบบเดิมและการทำ Machine learning [3]

ประเภทของ Machine learning จะถูกแบ่งตามประเภทของการเรียนรู้เป็น 2 ประเภทคือ

1. Supervised Learning: หรือ การเรียนรู้แบบมีผู้สอน ในการเรียนรู้แบบนี้คือ การสอนเครื่องจักร (machine) ว่าการให้ข้อมูล (input) แบบนี้จะได้ผลลัพธ์ (output) แบบไหน เช่น ต้องการให้เครื่องจักร (machine) เรียนรู้ว่าจะเด็กในท้องถิ่นไหนสอบตกหรือสอบผ่าน โดยมีข้อมูลของคะแนนนักเรียน (Data set) จะมีตัวอย่างดังตารางที่ 2.1

ชื่อ	คะแนน	Label
A	90	Pass
B	25	Fail
C	73	Pass
D	51	Pass
E	40	Fail
F	85	?

ตารางที่ 2.1 ตารางแสดงข้อมูลสำหรับ Supervised learning [3]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและ 4 ข้ออ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- โดยในตารางที่ 2.1 ข้อมูล (data set) จะเห็นว่ามีภาระระบุว่านักเรียนที่ได้คะแนนเท่านี้จะผ่านหรือไม่ผ่าน เครื่องจักร (machine) จะสามารถเรียนรู้ได้จากข้อมูลนี้ว่าคะแนนประมาณเท่าไหนถึงจะผ่าน และสามารถคาดเดาได้ว่านักเรียนชื่อ F ที่ได้ 85 คะแนนน่าจะได้ Pass คะแนนผ่าน
2. Unsupervised Learning: หรือ การเรียนรู้แบบไม่มีผู้สอน ในการเรียนรู้แบบนี้จะมีให้แค่ข้อมูล (input) แล้วกำหนดให้ เครื่องจักร (machine) ว่าต้องการอะไรดังตัวอย่างตารางที่ 2.2 ที่ระบุคะแนนของแต่ละเกรด

ชื่อ	คะแนน
A	90
B	25
C	73
D	51
E	40
F	85

ตารางที่ 2.2 ตารางแสดงข้อมูลสำหรับ Unsupervised [3]

โดยเพียงกำหนดว่าต้องการให้ เครื่องจักร (machine) แบ่งออกเป็น 2 กลุ่ม แต่ผู้เขียนโปรแกรมจะต้องรู้เองว่าผลลัพธ์คืออะไรเพราะคอมพิวเตอร์จะทำหน้าที่เพียงแบ่งออกเป็น 2 กลุ่มดังเช่นตัวอย่างในตารางที่ 2.3 ในช่องของ Output จะเห็นได้ว่านักเรียนจะถูกแบ่งออกเป็นสองกลุ่มโดยใช้ตัวเลข 1 และ 2 ระบุว่าใครผ่านและไม่ผ่าน

ชื่อ	คะแนน	Output
A	90	1
B	25	2
C	73	1
D	51	1
E	40	2
F	85	1

ตารางที่ 2.3 ตารางแสดงผลลัพธ์สำหรับ Unsupervised learning [3]

การทำโครงการสหกิจครั้งนี้จะใช้การทำงานของการเรียนรู้ของเครื่องจักร (Machine learning) ในรูปแบบของการเรียนรู้แบบมีผู้สอน (Supervised learning) เป็นหลักเพื่อเป็นพื้นฐานและต้นแบบที่ง่ายต่อการทำความเข้าใจต่อไปเป็นหลักการของการจัดการกับข้อมูลเสียงที่จะนำมาใช้ในการจำแนกประเภทเสียง (Sound Classification)

2.2 การจัดการกับข้อมูลในรูปแบบของเสียง

ข้อมูลที่จะใช้เป็นข้อมูลขาเข้า (input) ของการทำ Machine learning ในโครงการงานนี้คือ ข้อมูลของเสียง ในการจะคัดแยกเสียงและจัดหมวดหมู่นั้นจะต้องนำข้อมูลของเสียงมาใช้ แต่เนื่องจากเสียงเป็นข้อมูลที่เราไม่สามารถมองเห็นได้การทำ Sound Classification จึงซับซ้อนและมีขั้นตอนเยอะกว่าการทำ classification แบบอื่น ๆ สิ่งที่จะใช้ในการดึงลักษณะเด่นของเสียงออกมาคือการใช้ ค่าสัมประสิทธิ์ความถี่สเกลเมล (Mel-Frequency Cepstral Coefficient: MFCC) ถึงแม้ว่างานวิจัยส่วนใหญ่จะนิยมใช้ค่าสัมประสิทธิ์ MFCC ในวิเคราะห์ค่าลักษณะเด่นของสัญญาณเสียง แต่ค่า MFCC ก็ยังเป็นค่าสัมประสิทธิ์ที่เหมาะสมกับการวิเคราะห์ลักษณะเด่นของเสียงอื่น ๆ ด้วยเช่นกัน

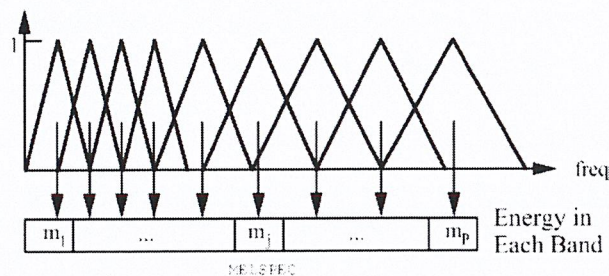
2.2.1 MFCC (Mel-Frequency Cepstral Coefficient)

MFCC คือการวิเคราะห์ลักษณะเด่นของเสียงออกมาในรูปแบบของค่าสัมประสิทธิ์ในสเกลเมล ส่วนสำคัญของตัวกรอง (filter) นี้คือการใช้ชุดตัวกรองความถี่ (Filter Bank Analysis) โดยปกติหูของมนุษย์นั้นจะมีความสามารถในการได้ยินเสียงคลื่นความถี่ที่ไม่อยู่ในลักษณะของสัญญาณเชิงเส้นได้โดยการเปลี่ยนสัญญาณดังกล่าวให้เป็นค่าสเปกตรัมของสัญญาณเสียง ซึ่งหลักการดังกล่าว ได้ถูกนำไปประยุกต์ใช้ในการสร้างค่าลักษณะเด่นของสัญญาณ เสียงพูดโดยการแปลงค่าสัญญาณที่มีลักษณะไม่เป็นเชิงเส้นได้โดยใช้ชุดตัวกรองความถี่ (Filter Bank)

สำหรับชุดตัวกรองความถี่ที่จะนำมาใช้ในงานวิจัยชิ้นนี้ จะใช้การแปลงฟูริเยร์อย่างง่ายในการออกแบบตัวกรองบนพื้นฐานของสเกลเมล (Mel-Scale) ดังที่แสดงในภาพประกอบที่ 2 ซึ่งตัวกรองที่เราใช้ในนั้นจะเป็นรูปสามเหลี่ยมและความถี่กลาง (Center Frequency) ของตัวกรองจะ กระจายตามสเกลของเมล ซึ่งนิยามไว้ดังสมการที่ 1

$$\text{Mel}(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

ในการสร้างชุดตัวกรองนั้น กรอบสัญญาณเสียงพูด (Window of Speech Data) ซึ่งจะได้ ขนาด (Magnitude) ของสัญญาณในช่วงนั้น ๆ ออกมาโดยค่าสัมประสิทธิ์ของค่าขนาดนั้นจะถูก กรองออกมาโดยผ่านตัวกรองสามเหลี่ยมในแต่ละช่วง แล้วยนำมาหาค่าเฉลี่ยกลางโดยการนำค่าที่ ผ่านตัวกรองมานั้นแปลงกับการแปลงฟูริเยร์แบบรวดเร็ว (Fast Fourier Transform: FFT) นำไปคูณ กับอัตราการขยายของตัวกรองในแต่ละช่วง สุดท้ายจึงนำค่าที่คำนวณได้มารวมกันแล้วแสดง ผลลัพธ์เป็นค่าขนาดของสเปกตรัมของช่วงตัวกรองแต่ละช่วง [4]



ภาพที่ 2.2 ภาพชุดตัวกรองเมลสเกล [4]

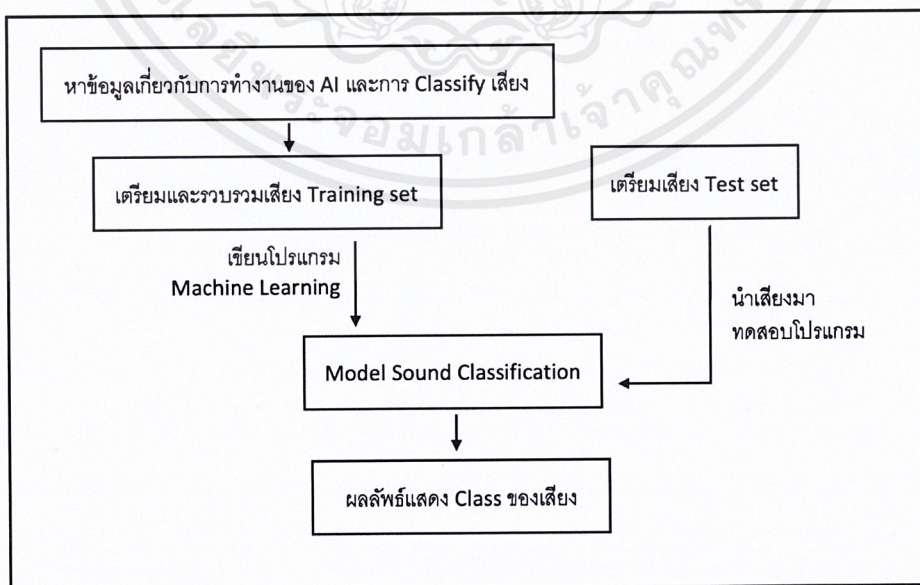
บทที่ 3 วิธีการดำเนินงาน

ขั้นตอนการดำเนินงานของโครงการสหกิจจะถูกแบ่งสัดส่วนการทำงานในระยะเวลา 4 เดือนแสดงให้เห็นตามตารางที่ 4 ดังนี้

หัวข้อ	สิงหาคม	กันยายน	ตุลาคม	พฤศจิกายน
1. หัวข้อและจุดประสงค์โครงการ	←→			
2. หาข้อมูลเกี่ยวกับ AI		←→		
3. หาข้อมูลเกี่ยวกับการจัดการข้อมูลเสียง		←→		
4. เตรียมข้อมูลเสียง		←→		
5. เขียนโปรแกรม			←→	
6. เตรียม Present และรายงาน				←→

ตารางที่ 3.1 ตารางแสดงขั้นตอนการทำงาน

ในการทำ Sound Classification ขั้นตอนที่ต้องทำจะถูกแบ่งออกเป็น 3 ส่วนคือ การเตรียมข้อมูล การเขียนโปรแกรม และการทดสอบโปรแกรม โปรแกรมนี้จะทำงานบน Colab google research และใช้ python3 ในการทำงานเพื่อสามารถใช้หน่วยประมวลผลด้านกราฟิกส์ (Graphics Processing Unit: GPU) บนอินเทอร์เน็ตได้และไม่ให้คอมพิวเตอร์ทำงานหนักเกินไป ขั้นตอนของการทำงานทั้งหมดจะถูกแสดงให้เห็นในภาพที่ 3.1



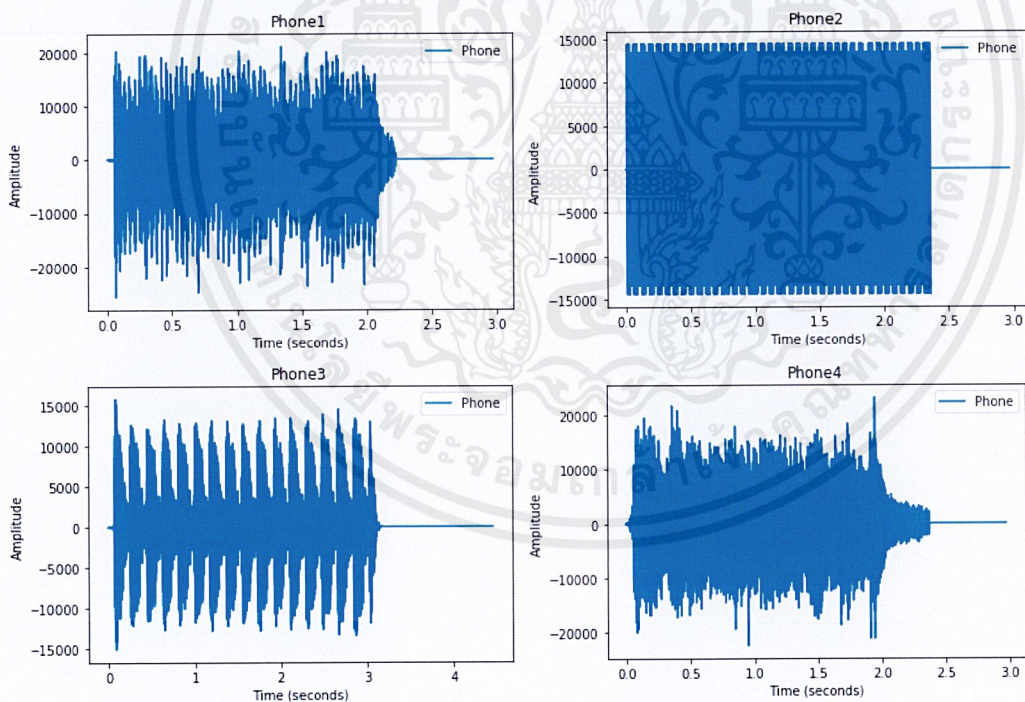
ภาพที่ 3.1 แผนผังแสดงขั้นตอนการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

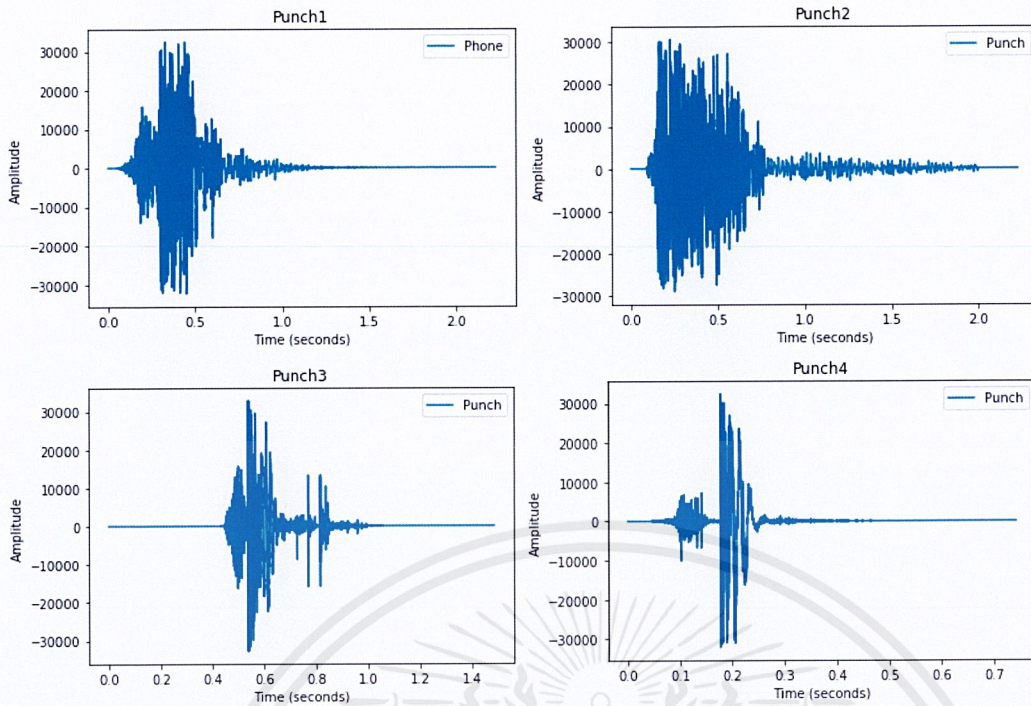
3.1 การเตรียมข้อมูล

การทำ Machine learning จะต้องมีการเตรียมข้อมูลเพื่อสอน machine หรือคอมพิวเตอร์เพื่อให้ machine เรียนรู้จุดเด่นของข้อมูล ในโครงการสหกิจครั้งนี้ข้อมูลคือไฟล์เสียงที่ถูกแบ่งเป็น 2 กลุ่มหรือในภาษาของการทำ Machine learning คือการแบ่งกลุ่มของข้อมูลออกเป็น ประเภท (class) การแบ่งเป็น class เพื่อให้ machine เรียนรู้จุดเด่นของไฟล์เสียงในแต่ละ class เพื่อที่จะสามารถตัดแยกได้ในขั้นตอนถัดไป

Class ที่เลือกนำมาทดลองทำงานนี้คือเสียงต๋อย และเสียงโทรศัพท์ เนื่องจากเป็นเสียง 2 กลุ่มที่มีความแตกต่างกันอย่างชัดเจนเพื่อให้การตัดแยกแบ่งกลุ่มเบื้องต้นแม่นยำและถูกต้องมากที่สุดเห็นตัวอย่างของคลื่นเสียงได้ในภาพที่ 3.2 และ 3.3 ในแต่ละ class จะมีไฟล์เสียง class ละ 100 เสียงเพื่อให้โปรแกรมได้เรียนรู้จุดเด่นของเสียงนั้น ๆ เสียงที่ถูกนำมาทดลองเขียนโปรแกรมนี้คือเสียงจาก freesound.org เป็นเว็บไซต์ที่สามารถนำเสียงที่ถูกอัปโหลดไว้ไปใช้ได้ฟรี เสียงที่ถูกนำมาจะเป็นเสียงที่มีรูปแบบไฟล์เดียวกันคือไฟล์ .wav มี bit depth ที่ 16 เท่ากันทั้งหมดและมี sample rate ที่ 44100 Hz เท่ากันทั้งหมด เสียงทั้งหมดจะถูกระบุไว้ว่าเป็นเสียงชนิดไหนระหว่างเสียงต๋อยและเสียงโทรศัพท์เพื่อเป็นหารทำ Machine learning แบบ Supervised learning



ภาพที่ 3.2 ภาพตัวอย่างของคลื่นสัญญาณเสียงโทรศัพท์



ภาพที่ 3.3 ภาพตัวอย่างของคลื่นสัญญาณเสียงต๋อย

เพื่อให้มีตัวแปรความแตกต่างไม่มากเกินไปแต่ไฟล์เสียงที่หลากหลายยังมีความแตกต่างมากจึงต้องนำไฟล์เสียงผ่านกระบวนการ MFCC ที่ถูกอธิบายไปในบทที่ 2 เพื่อให้โปรแกรมสามารถดึงจุดเด่นของเสียงได้มากที่สุด กระบวนการ MFCC เป็นส่วนหนึ่งของ library ของ python ที่ชื่อว่า librosa และใช้เครื่องมือ `librosa.feature.mfcc` เมื่อผ่านกระบวนการ MFCC สัมประสิทธิ์จะถูกจัดเก็บเป็นข้อมูล training set ของแต่ละ class เพื่อนำไปให้คอมพิวเตอร์ได้เรียนรู้

3.2 การเขียนโปรแกรม

ขั้นตอนนี้เป็นขั้นตอนของการเขียนโปรแกรมที่ใช้ในการให้ machine ได้เรียนรู้ training set ที่ถูกเตรียมไว้และเกิดเป็นโปรแกรมที่สามารถทดลองใช้กับเสียงใหม่ ๆ ได้ ในการเขียนโปรแกรมทำ machine learning จะต้องเลือกวิธีการเรียนรู้หรือ Algorithm ในการทำ machine learning และสำหรับ Sound Classification ในครั้งนี้จะใช้ algorithm ที่เรียกว่า Convolutional Neural Network หรือ CNN เรียนรู้ข้อมูลที่ถูกเตรียมไว้เรียกว่า training set

Convolutional Neural Network (CNN)

ความรู้เกี่ยวกับ การเรียนรู้เชิงลึก (Deep Learning) เป็นแนวคิดในการ machine learning เข้าใจข้อมูลที่ได้รับ โดยมีสถาปัตยกรรมการเรียนรู้ข้อมูลของคอมพิวเตอร์มากมายโดยจะกล่าวถึง CNN นำมาใช้เพื่อการสกัดเอาลักษณะเด่น จนกระทั่งทราบถึงผลลัพธ์ที่ตรงจวบ การทำงานของ CNN มี 4 กระบวนการด้วยกัน

3.2.1.1 คอนโวลูชัน (Convolution)

คอนโวลูชันเป็นกระบวนการที่ทำเพื่อสกัดเอาลักษณะสำคัญจากภาพออกมา โดยการใช้ค่าพิกเซล โดยค่าพิกเซล ในแต่ละรูปจะมีชุด Matrix ที่ต่างกัน และมีการกำหนด matrix อีกหนึ่งชุดขึ้นมา โดยเราจะใช้ matrix ชุดนี้เป็นตัวกรองค่าไปเก็บไว้ใน matrix ชุดที่เล็กกว่าเราเรียก matrix ชุดนี้ว่าตัวกรองค่า (Filter), เคอเนล (Kernel) หรือตัวตรวจจับลักษณะสำคัญ (Feature Detector) matrix ที่ทำหน้าที่เป็นตัวกรองค่า จะเคลื่อนที่ไปทั่วภาพและคูณค่าเก็บไว้ใน matrix ชุดใหม่ซึ่งเราเรียก matrix ชุดใหม่นี้ว่า Convolved Feature หรือ ฟีเจอร์แมพ (Feature Map) โดยผลลัพธ์ที่ได้เมื่อรูปผ่านการทำ convolution

3.2.1.2 การขจัดความเป็นเชิงเส้น (ReLU)

หลังจากได้ทำการ convolution และได้ feature map มาแล้ว จะนำปรับแต่งให้ feature map ไม่เป็นลักษณะเชิงเส้น ด้วยวิธีการ ReLU ทำการแทนที่ผลของพิกเซลที่มีค่าเป็นเชิงลบใน feature map ด้วยค่า 0 จุดประสงค์ของการทำ ReLU นั้นเพื่อให้โครงข่ายประสาทเทียมแบบ CNN เรียนรู้ข้อมูลที่ไม่เป็นเชิงเส้นจากภาพ เมื่อนำภาพเข้าสู่กระบวนการที่เป็น feature map เข้ามาทำ ReLU และค่าที่ได้จะมีเพียงค่าที่เป็นเชิงบวกเท่านั้น

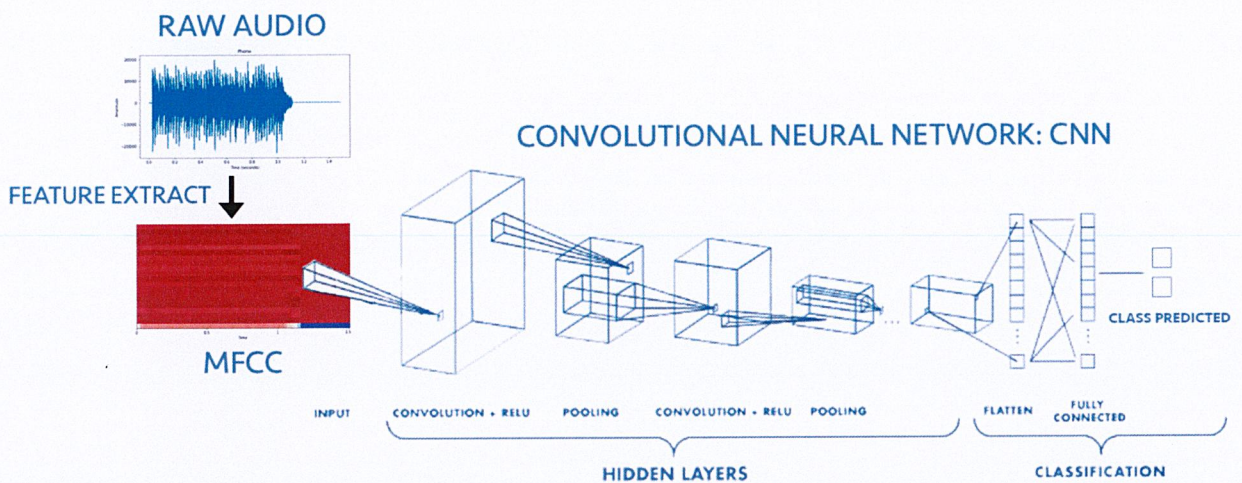
3.2.1.3 การพูลลิ่ง (Pooling)

การ Pooling ช่วยลดมิติของ feature map แต่ยังคงรักษาข้อมูลสำคัญไว้ การ Pooling สามารถจำแนกเป็นประเภทต่างได้เช่น พูลลิ่งด้วยค่าสูงสุด (Max Pooling), ค่าเฉลี่ย (Average Pooling), ผลรวม หรืออื่นๆการพูลลิ่ง ทำให้ผลลัพธ์ที่ได้มีขนาดเล็กและจัดการได้ง่ายขึ้น นอกจากนี้ยัง ลดจำนวนพารามิเตอร์และการคำนวณที่เกินจำเป็นในโครงข่ายในกรณีที่ต้องการ Pooling ด้วยค่าสูงสุด จะกำหนดหน้าต่างหนึ่งขึ้นมา ตัวอย่างในที่นี้ สมมติให้หน้าต่างมีขนาด 2x2 และหน้าต่างนี้จะทำการเคลื่อนที่ทีละ 2 พิกเซลไปจนทั่ว matrix ของ feature เพื่อทำการเก็บค่าที่สูงที่สุดในทุก ๆ 2 พิกเซล นอกจากนี้จะทำตามจำนวนแกนแนลของภาพซึ่งในกรณีภาพสีเราจะได้รับผลลัพธ์ออกมาถึงสามผลลัพธ์เมื่อ feature map ที่ผ่านการทำ ReLU เข้ามาทำการ Pooling

3.2.1.4 การเชื่อมต่อกันของแต่ละชั้น (Fully Connected Layer)

กระบวนการ Convolution , ReLU และการ Pooling กระบวนการทั้งสามกระบวนการต้องมีการทำซ้ำจนกว่าจะมีการเชื่อมต่อกันของแต่ละชั้น (Fully Connected Layer) ผลลัพธ์จาก Convolution และ Pooling นั้นให้ลักษณะเด่นในระดับสูง (High-level features) ของรูปที่รับเข้ามาจุดประสงค์ของการทำให้เชื่อมต่อกันแต่ละชั้นโดยสมบูรณ์นั้นเพื่อนำลักษณะเด่นไปทำการคัดกรองข้อมูลที่ได้รับเข้ามาให้อยู่ในรูปของคลาส (Classes) โดยผลลัพธ์ที่ได้จะแสดงค่าความมั่นใจ (Confident) ออกมา ตัวอย่างของการคัดกรองที่มีการเทรนข้อมูลไว้ 2 ประเภท เมื่อนำเข้าสู่กระบวนการทั้งหมดจะแสดงค่าผลลัพธ์การคาดเดาออกมาว่าข้อมูลนั้นอยู่ใน class ไหน

การเขียนโปรแกรมในโครงงานนี้จึงใช้ library ที่มีชื่อว่า keras ในการสร้าง model CNN สำหรับโครงงานนี้เพื่อให้ model ได้เทรนข้อมูลเสียง (training set) ที่ถูกเตรียมไว้ เมื่อสร้าง model เสร็จจึงใช้คำสั่ง `model.fit` เพื่อให้ model ได้เรียนรู้ training set ที่เตรียมไว้



ภาพที่ 3.4 ภาพแผนผังการทำงานของโปรแกรมตั้งแต่ขั้นตอนของการป้อนข้อมูลขาเข้าเป็นไฟล์เสียงจนถึงการคาดเดาของโปรแกรมว่าเป็นเสียงประเภทใด [5]

จากภาพที่ 3.4 จะแสดงภาพรวมของขั้นตอนในหัวข้อ 3.1 และ 3.2 ตั้งแต่การป้อนข้อมูลไฟล์เสียงที่ถูกกรองและคัดแยกจุดเด่น (feature extract) โดยกระบวนการ MFCC และได้ค่าสัมประสิทธิ์สเกลเมลเพื่อนำไปเป็นข้อมูลขาเข้า (input) ของการจำแนกประเภทเสียง (sound classification) โดยใช้กระบวนการ Convolutional Neural Network (CNN) เพื่อให้โปรแกรม (machine) คาดเดาประเภทของเสียงโดยวิธีในหัวข้อถัดไป การทดสอบโปรแกรม

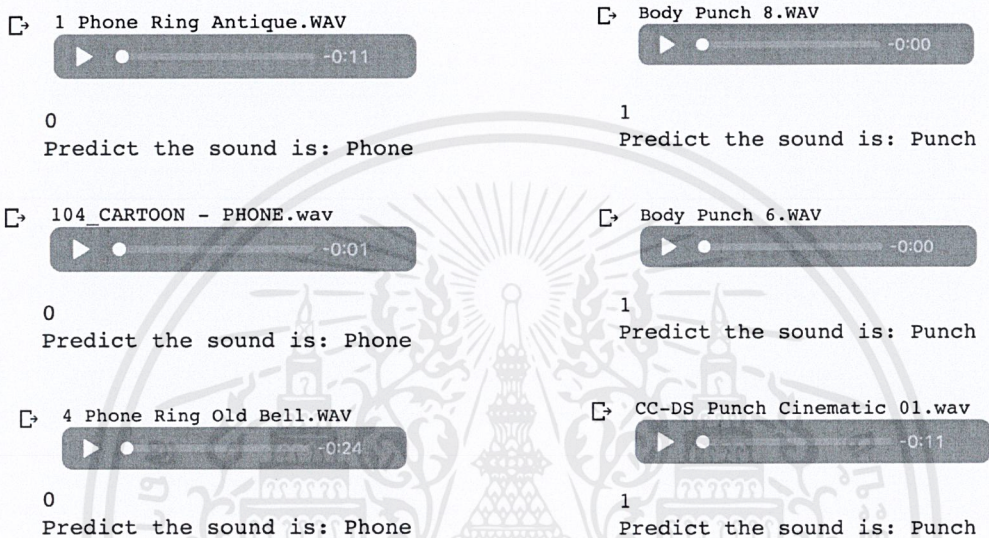
3.3 การทดสอบโปรแกรม

ในขั้นตอนของการทดสอบโปรแกรมจะมีการเตรียมเสียงใหม่ที่จะไว้ทดสอบโปรแกรมที่ถูกสร้างไว้ในหัวข้อ 3.2 เสียงที่เตรียมไว้ทดสอบจะเรียกว่า test set จะเตรียมเสียงโทรศัพท์ 5 เสียงและเสียงต๋อย 5 เสียง เพื่อที่จะทดสอบว่าโปรแกรมสามารถแยกเสียงตาม class ที่ถูกต้องและแม่นยำหรือไม่

บทที่ 4

ผลของการทำโครงการงาน

ผลของการทำโครงการงานนี้คือ โปรแกรมสามารถแยกเสียงที่ใช้ทดสอบ (test set) ได้อย่างแม่นยำ เนื่องจากกลุ่มของเสียง (Class) มีเพียงสองกลุ่มเพื่อความแม่นยำ ผลจึงออกมา 100% โปรแกรมสามารถแยกเสียงได้ทุกตัวอย่างทั้ง 10 เสียง



ภาพที่ 4.1 ภาพของผลลัพธ์การทดสอบของโปรแกรม

ในภาพที่ 4.1 จะสามารถเห็นตัวอย่างของผลลัพธ์ของโปรแกรมเมื่อทดลองโปรแกรมด้วยเสียงสองประเภท โปรแกรมสามารถระบุได้ว่าเสียงชนิดไหนเป็นเสียงที่อยู่ใน class เสียงต๋อยหรือเสียงโทรศัพท์ และเสียงที่นำมาทดสอบจะเป็นเสียงที่ไม่ซ้ำกับเสียงที่ใช้สำหรับ train โปรแกรม

บทที่ 5

สรุปผลโครงการและข้อเสนอแนะ

โครงการนี้ได้ทำให้เรียนรู้ความรู้ใหม่ ๆ มากมาย ทั้งความรู้เกี่ยวกับปัญญาประดิษฐ์ที่ไม่เคยศึกษามาก่อนเหมือนกับได้เรียนรู้วิชาใหม่ ๆ ไปพร้อมกับการทำงานนี้ การได้เขียนโปรแกรมนี้อาจทำให้ได้ทบทวนการเขียนโปรแกรมไปในตัวและได้ศึกษากระบวนการใหม่ ๆ เช่น MFCC ที่สามารถนำไปต่อยอดได้ในอนาคต ในส่วนของโปรแกรมที่ทำการ Classify เสียงนั้นสามารถนำไปต่อยอดและพัฒนาได้ให้มีประสิทธิภาพและละเอียดมากยิ่งขึ้นสามารถตัดแยกเสียงได้หลากหลายประเภทมากยิ่งขึ้นทั้งเสียงที่ต่างหมวดหมู่กันและเสียงที่มีความใกล้เคียงกันเพื่อให้เห็นผลมากยิ่งขึ้น

ข้อแนะนำสำหรับโครงการนี้คือในการทำ Machine learning ส่วนสำคัญคือ training set ที่จะต้องมีจำนวนมากเพื่อให้ machine ได้เรียนรู้และสามารถตัดแยกได้อย่างแม่นยำยิ่งถ้าเสียงมีความใกล้เคียงกันก็จะต้องมีข้อมูลให้เทรนมากขึ้นอีก



เอกสารอ้างอิง

- [1] ทองย้อย แสงสินชัย. 2019. ปัญญาประดิษฐ์ ปัญญาของราชบัณฑิตยฯ. สืบค้นเมื่อ 14 พฤศจิกายน 2019, <<http://dhamma.serichon.us/28838-2/>>
- [2] Netsakaw Wetthayanon. 2017. เรื่อง : ปัญญาประดิษฐ์ (Artificial Intelligence). สืบค้นเมื่อ 14 พฤศจิกายน 2019, <<http://mugglestar.blogspot.com/p/blog-page.html>>
- [3] DumpDataSci. 2017. Machine Learning คืออะไร? (พร้อมตัวอย่างในชีวิตจริง). สืบค้นเมื่อ 14 พฤศจิกายน 2019, <<https://medium.com/@dumpdatasci.th/machine-learning-คืออะไร-พร้อมตัวอย่างในชีวิตจริง-b3f8faf0b4da>>
- [4] โนอาห์ กิจญะวงค์. (2012). การเปรียบเทียบระบบค้นหาหลักสำหรับเสียงพูดภาษาไทย. หลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์มหาวิทยาลัยสงขลานครินทร์
- [5] Natthawat Phongchit. 2018. Convolutional Neural Network (CNN) คืออะไร. สืบค้นเมื่อ 29 พฤศจิกายน 2019, <<https://blog.datawow.io/มาลองดูวิธีการคิดของ-cnn-กัน-e3f5d73eebaa>>