



Computer Science Maejo University

เรื่อง Speech Recognition

จัดทำโดย

นาย ธนะรัชต์ ชำมยอง 6204101319

นาย บวรพจน์ แสนโซ้ง 6204101326

นาย ประมินทร์ กรกัรติการ 6204101328

เสนอ

อ.ดร.พยุงค์ดี เกษมสำราญ และ อ.ทัศนีย์ ไชยา

รายงานเล่มนี้เป็นส่วนหนึ่งของวิชา คพ 452 ปัญญาประดิษฐ์

สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยแม่โจ้

บทนำ

Speech Recognition คือระบบโปรแกรมคอมพิวเตอร์ที่สามารถแปลงเสียงพูด (Audio File) เป็นข้อความตัวอักษร (Text) โดยสามารถแจกแจงคำพูดต่างๆ ที่มนุษย์สามารถพูดใส่ไมโครโฟน โทรศัพท์หรืออุปกรณ์อื่นๆ และเข้าใจคำศัพท์ทุกคำอย่างถูกต้องเกือบ 100% โดยเป็นอิสระจากขนาดของกลุ่มคำศัพท์ ความดังของเสียง และลักษณะการออกเสียงของผู้พูด โดยระบบจะรับฟังเสียงพูดและตัดสินใจว่าเสียงที่ได้ยินนั้นเป็นคำๆใด เทคโนโลยีที่เป็นส่วนสำคัญในการทำ ASR เรียกว่า Hidden Markov Model (HMM) เทคโนโลยีชนิดนี้สามารถที่จะเข้าใจคำพูด จากการจำแนกความแตกต่างและการประมาณการถึงความเป็นไปได้ของส่วนประกอบของหน่วยที่เป็นพื้นฐานของเสียงที่อยู่ติดๆกัน โดยอาศัยหลักการที่ว่าเสียงแต่ละเสียงจะมีขอบเขตของสัญญาณและลักษณะเฉพาะที่มีความแตกต่างกัน

วัตถุประสงค์

1. เพื่อศึกษาและเข้าใจหลักการทำงานของตัว Speech Recognition
2. เพื่อศึกษาว่าสามารถนำไปประยุกต์กับการใช้งานในด้านไหนได้บ้าง
3. เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์ในอนาคต

ขอบเขต

1. สามารถพูดและให้แปลงออกมาเป็น Text ได้
2. พูดและโต้ตอบได้บางคำถามที่ได้กำหนดไว้
3. สามารถสั่งให้เปิดเว็บเพจและเปิดเพลงบน Youtube ได้

การสร้างตัว Speech Recognition

1. ทำการติดตั้ง Library ที่ต้องใช้คือตัว SpeechRecognition และ PyAudio

```
!pip install SpeechRecognition
```

Requirement already satisfied: SpeechRecognition in c:\users\yeske\anaconda3\lib\site-packages (3.8.1)

```
!pip install PyAudio-0.2.11-cp38-cp38-win_amd64.whl
```

Processing c:\users\yeske\desktop\cs452\projectai\pyaudio-0.2.11-cp38-cp38-win_amd64.whl

PyAudio is already installed with the same version as the provided wheel. Use --force-reinstall to force an installation of the wheel.

1.1 นำ Library ที่ติดตั้งเข้ามาใช้

โดยให้ตัว speech_recognition ใช้ชื่อว่า sr

```
import pyaudio
import speech_recognition as sr
```

2. ตรวจสอบไมค์ที่มีอยู่ใน Window

โดยใช้คำสั่ง sr.Microphone.list_microphone_names() จะเก็บชื่อในรูปแบบ List

```
sr.Microphone.list_microphone_names()
```

```
['Microsoft Sound Mapper - Input',
'Microphone (Realtek High Defini',
'Microsoft Sound Mapper - Output',
'Speakers (Realtek High Definiti',
'Primary Sound Capture Driver',
'Microphone (Realtek High Definition Audio)',
'Primary Sound Driver',
'Speakers (Realtek High Definition Audio)',
'Speakers (Realtek High Definition Audio)',
'Microphone (Realtek High Definition Audio)',
'Microphone (Realtek HD Audio Mic input)',
'Stereo Mix (Realtek HD Audio Stereo input)',
'Speakers (Realtek HD Audio output)']
```

2.1 เลือกไมโครโฟนที่ต้องการใช้

ใช้คำสั่ง `sr.Microphone()` และรับ argument เป็นตำแหน่งของชื่อไมโครโฟนที่ต้องการใช้ ที่อยู่ใน List ที่ได้จากคำสั่ง `sr.Microphone.list_microphone_names()` เราจะได้ object `Microphone` ที่เก็บข้อมูลเกี่ยวกับไมค์ตัวนั้นมา

```
mic = sr.Microphone(1)
mic
```

```
<speech_recognition.Microphone at 0x19e48956e80>
```

3.สร้าง Recognizer

Recognizer คือตัวโปรแกรมที่จะใช้ทำความเข้าใจเสียง และส่งออกมาเป็นตัวหนังสือ ผ่านคำสั่ง `sr.Recognizer()` ที่สร้าง Recognizer Object

```
recog = sr.Recognizer()
recog
```

```
<speech_recognition.Recognizer at 0x19e488cb8b0>
```

4.เปิดไมค์รับเสียง

เราจะใช้คำสั่งให้ recognizer ที่สร้างไว้แล้วเปิดรับเสียงจากไมค์ ผ่านคำสั่ง `recog.listen()` ที่รับ argument เป็น mic ที่เราเตรียมไว้ ข้อมูลที่ได้มาเราจะเก็บไว้ในตัวแปร `audio`

แต่ในการเปิดไมค์ จะต้องใช้คำสั่ง `with` ที่ทำการเปิดไมค์ และนำข้อมูลของไมค์เก็บไว้ในอีกตัวแปร เราจะตั้งชื่อว่า `source` ซึ่งเป็นที่มาของข้อมูลที่จะใช้กับ recognizer

```
with mic as source: #นำข้อมูลเสียงที่เราพูด(mic) ไปเก็บไว้ในตัวแปร source
    audio = recog.listen(source) #รับเสียงในตัวแปร source แล้วไปเก็บไว้ในตัวแปร audio
```

4.1 แปลงเสียงเป็นตัวอักษร

ใช้คำสั่ง `recog.recognize_google()` ที่รับ argument 2 อย่างหลักๆ คือ เสียงที่ต้องการแปลงเป็นตัวหนังสือในกรณีนี้ได้จากการเปิดไมค์ และ ภาษาที่ต้องการให้ใส่ ไปใน parameter `language` หากไม่ใส่จะจับภาษาอังกฤษ สำหรับภาษาไทยใส่ 'th' เมื่อแปลงเสียงแล้วเราจะให้แสดงผลผ่านคำสั่ง `print()`

```
with mic as source: #นำข้อมูลเสียงที่เราพูด(mic) ไปเก็บไว้ในตัวแปร source
    audio = recog.listen(source) #รับเสียงในตัวแปร source แล้วไปเก็บไว้ในตัวแปร audio
    print(recog.recognize_google(audio,language='th')) #ใช้ Google Web Speech API แปลเสียงออกมาเป็นภาษาไทย
```

สวัสดีครับสวัสดีครับ

5.วิเคราะห์ต่อเนื่อง

ใช้ while loop เพื่อสร้างการรับค่า และแปลงตัวหนังสืออย่างต่อเนื่อง เสริมด้วยตัว try except เพื่อสั่งคำสั่ง continue ใหวนลูปใหม่เมื่อเกิด error

```
with mic as source:
    while True:
        audio = recog.listen(source)
        try:|
            print(recog.recognize_google(audio,language='th'))
        except:
            continue
```

northwest Tree
1 2 3 4 5 6 7 8 9
สวัสดีครับ
ทำอะไรอยู่

นำไปประยุกต์ใช้งาน ใช้ if..else เข้ามาเพิ่ม เพื่อให้กำหนดกรอบของการทำงานให้ชัดเจน

1.ทำให้เป็นการทำงานแบบกดปุ่ม (Y/N)

```
while True:
    aw = str(input("-->คุณต้องการแปลเสียงหรือไม่? (Y-ต้องการ/N-ไม่ต้องการ): "))
    if aw == "y" or aw == "Y":
        print("*****เชิญพูดได้เลยครับ..*****")
        with mic as source:
            audio = recog.listen(source)
            print("เสียงพูด: ", recog.recognize_google(audio, language='th-TH')) #ภาษาสามารถเปลี่ยนได้ ko-KR, en-US

            #ลองเก็บคำไว้นั่นครับ ถ้าพูดว่า "เราเป็นคนไทย" จะโชว์ตอบว่า "เอ้า!! คนไทยนี่"
            sl=recog.recognize_google(audio, language='th-TH')
            if sl=='เราเป็นคนไทย':
                a1='เอ้า!! คนไทยนี่'
                print('คอมตอบ: '+a1)

            continue
    elif aw == "n" or aw == "N":
        print("Good Luck")
        break
    else:
        print("!!!โปรดเลือกแค่ Y/N")
        continue
    #aw = str(input("โปรดเลือกอีกครั้งคุณต้องการแปลเสียงหรือไม่? (Y-ต้องการ/N-ไม่ต้องการ): "))
```

```
-->:คุณต้องการแปลเสียงหรือไม่? (Y-ต้องการ/N-ไม่ต้องการ): y
*****เชิญพูดได้เลยครับ..*****
เสียงพูด: สวัสดี
-->:คุณต้องการแปลเสียงหรือไม่? (Y-ต้องการ/N-ไม่ต้องการ): y
*****เชิญพูดได้เลยครับ..*****
เสียงพูด: เราเป็นคนไทย
คอมตอบ: เอ้า!! คนไทยนี่
-->:คุณต้องการแปลเสียงหรือไม่? (Y-ต้องการ/N-ไม่ต้องการ): n
Good Luck
```

2.ทำให้เป็นการทำงานแบบพูดอย่างเดียว

```
with mic as source:
    while True:
        print("*****เชิญพูดได้เลยครับ..****")
        audio = recog.listen(source)
        try:
            print("เสียงพูด: ",recog.recognize_google(audio,language='th'))
        except:
            continue

        sl=recog.recognize_google(audio,language='th-TH')

        if sl=='ไม่พูดแล้ว':
            print('คอมตอบ: '+ 'โอเค นายยยยจ้าาา...')
            break
        elif sl=='สวัสดี':
            print('คอมตอบ: '+ 'สวัสดีจ้า')
        elif sl=='ทานข้าวหรือยัง' or sl=='ทานข้าวหรือยัง' or sl=='ทานข้าวหรือยัง':
            print('คอมตอบ: '+ 'เราไม่กินข้าว เรากินไฟฟ้า')
```

```
*****เชิญพูดได้เลยครับ..****
*****เชิญพูดได้เลยครับ..****
เสียงพูด: สวัสดี
คอมตอบ: สวัสดีจ้า
*****เชิญพูดได้เลยครับ..****
เสียงพูด: ทานข้าวหรือยัง
คอมตอบ: เราไม่กินข้าว เรากินไฟฟ้า
*****เชิญพูดได้เลยครับ..****
เสียงพูด: ไม่พูดแล้ว
คอมตอบ: โอเค นายยยยจ้าาา...
```

3.ใช้ API ในการเก็บคำถาม-คำตอบ และสร้างฟังก์ชันเก็บคำสั่งไว้

3.1 เรียกใช้ API ที่สร้างไว้เก็บคำถามและคำตอบในรูปแบบ list

```
import requests

url=('https://pond-api-answer.anonymous91.repl.co/qanda_thanarat_Api')
r=requests.get(url)
j=r.json()
```

3.2 นำ Library ที่ต้องใช้เข้ามา และสร้าง Function เพื่อเก็บคำสั่ง

```
import webbrowser #ใช้เปิดหน้าเว็บ/
import time #ใช้ รอเวลา เพื่อแสดงชื่อเพลงใน fn ถ้าไม่ได้อยู่ในชื่อเพลงแสดงไม่ทัน
import os, glob #ลบไฟล์ขยะ
```

```

def yt_open(namemusic):
    print("*****เชิญพูดได้เลยครับ..*****")
    with mic as source:
        audio = recog.listen(source)
    nm=recog.recognize_google(audio,language='th')
    print('คอมตอบ: ชื่อเพลง '+nm)
    time.sleep(1)
    webbrowser.open("https://www.youtube.com/results?search_query="+nm)

def fb_open():
    time.sleep(1)
    webbrowser.open("https://web.facebook.com/")

def de_temp():
    print('*****กำลังลบ...*****')
    dir = 'C:/Windows/Temp/'
    dir2 = 'C:/Users/Yeske/AppData/Local/Temp/'
    for file in os.listdir(dir2):
        try:
            os.remove(file.path)
        except:
            continue

    for file in os.listdir(dir1):
        try:
            os.remove(file.path)
        except:
            continue
    print('*****ลบเสร็จสิ้น*****')

```

3.3 Code

```

with mic as source:
    while True:
        print("*****เชิญพูดได้เลยครับ..*****")
        audio = recog.listen(source)
        try:
            print('เสียงพูด: ',recog.recognize_google(audio,language='th'))
        except:
            continue

        s1=recog.recognize_google(audio,language='th-TH')

        try:
            if s1=='ไม่พูดแล้ว':
                print('คอมตอบ: '+ 'โอเค นายยยยจ้าาา...')
                break

            elif j[s1]==j[s1] and s1!='เปิดเพลงบน YouTube หน่อย' and s1!='เปิด Facebook หน่อย' and s1!='ลบไฟล์ขยะ window':
                print('คอมตอบ: '+j[s1])
            elif s1=='เปิดเพลงบน YouTube หน่อย' or s1=='เปิด Facebook หน่อย' or s1=='ลบไฟล์ขยะ window':
                print('คอมตอบ: '+j[s1])
                break

        except:
            print('คอมตอบ: '+ 'ผมไม่เข้าใจคำถามบบ ลองพูดประโยคอื่น')
            continue

if s1=='เปิดเพลงบน YouTube หน่อย':
    yt_open(s1)
elif s1=='เปิด Facebook หน่อย':
    fb_open()
elif s1=='ลบไฟล์ขยะ window':
    de_temp()

```


*****เชิญพูดได้เลยครับ.*****

เสียงพูด: สวัสดี

คอมตอบ: สวัสดี มนุษย์

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: มีแฟนหรือยัง

คอมตอบ: ผมไม่เข้าใจคำถาม ลองพูดประโยคอื่น

*****เชิญพูดได้เลยครับ.*****

*****เชิญพูดได้เลยครับ.*****

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: ฉันรักคุณนะ

คอมตอบ: ผมไม่เข้าใจคำถาม ลองพูดประโยคอื่น

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: ฉันขอยืมเงินหน่อย

คอมตอบ: ฉันไม่แนะนำนะ ดอกเบี้ยมันสูง

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: xxx

คอมตอบ: ผมไม่เข้าใจคำถาม ลองพูดประโยคอื่น

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: YouTube หน่อย

คอมตอบ: ผมไม่เข้าใจคำถาม ลองพูดประโยคอื่น

*****เชิญพูดได้เลยครับ.*****

เสียงพูด: เปิดเพลงบน YouTube หน่อย

คอมตอบ: เอาเพลงอะไรดีล่ะ นอกแค่อัปเดตเพลงนะ

*****เชิญพูดได้เลยครับ.*****

คอมตอบ: ชื่อเพลง ดวงเดือน

ขอขอบคุณ

- <https://replit.com/>



บรรณานุกรม

Ultimate Python. (2564). *ถอดเสียงเป็นข้อความ real-time ใน 10 บรรทัด! Python Speech to Text*. สืบค้น 10 ตุลาคม 2564. จาก <https://www.ultimatepython.co/post/speech-to-text-python>

Techiedelight. (2564). *Delete all files in a directory in Python*. สืบค้น 10 ตุลาคม 2564. จาก <https://www.techiedelight.com/delete-all-files-directory-python/>

Mike Driscoll. (2562). *Python sleep(): How to Add Time Delays to Your Code*. สืบค้น 10 ตุลาคม 2564. จาก <https://realpython.com/python-sleep/#reader-comments>

Codegrepper. (2563). *how to open a new tab in chrome using python*. สืบค้น 10 ตุลาคม 2564. จาก <https://www.codegrepper.com/code-examples/python/how+to+open+a+new+tab+in+chrome+using+python>

Patiphan Phengpao. (2564). *สร้าง API ของตัวเองง่ายๆ แบบเขียนโค้ดไม่เกิน 30 บรรทัด* 🤖 WOW!!  . สืบค้น 10 ตุลาคม 2564. จาก <https://youtu.be/xGY9qrk2GnU>