嵌入式系統總整與實作

曾煜棋、吳昆儒

National Yang Ming Chiao Tung University

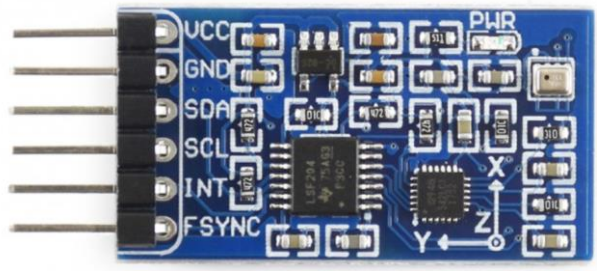
日期	主題
2/21	0. 課程介紹
2/28	梅竹賽!!
3/7	1. 嵌入式開發板 - 樹莓派介紹與設定 (headless)
3/14	2. 連接感測器 (GPIO, I2C)
3/21	3. 整合感測資訊 (IMU)
3/28	4. 整合音訊資料 (麥克風, 語音識別)
4/4	清明節放假
4/11	5. 整合視覺資料 (攝影機, 影像辨識)
4/18	期中考Midterm, Project分組
4/25	6. 嵌入式模型 (mediapipe, video, audio, text)
5/2	7. 喚醒詞原理 (by 台灣樹莓派)
5/9	Final Project – Proposal 分組報告 (online?)
5/16	8. 樹莓派核心編譯 (Cross compile, Kernel)
5/23	9. 嵌入式套件編譯
5/30	端午節放假
6/6	Final project準備周, Q&A (學期考試周)
6/13	**Final Project demonstration**

期中考周
(4/7-4/11)

期末考周
(6/2-6/6)

Last week

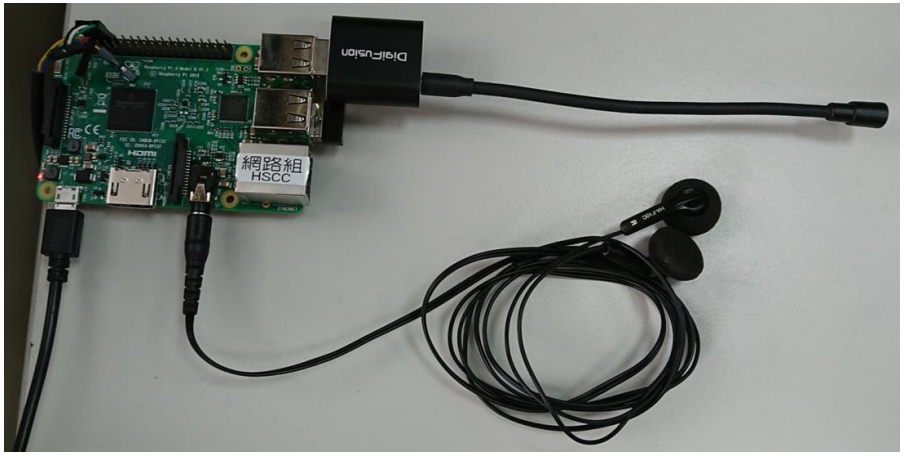
- Process sensing data
  - The information of IMU
  - Datasheet and code configuration
  - IMU applications
    - Calibration
    - Calculate distance, rotation angle and heading
    - Fall detection



IMU: ICM20948
(Inertial measurement unit)

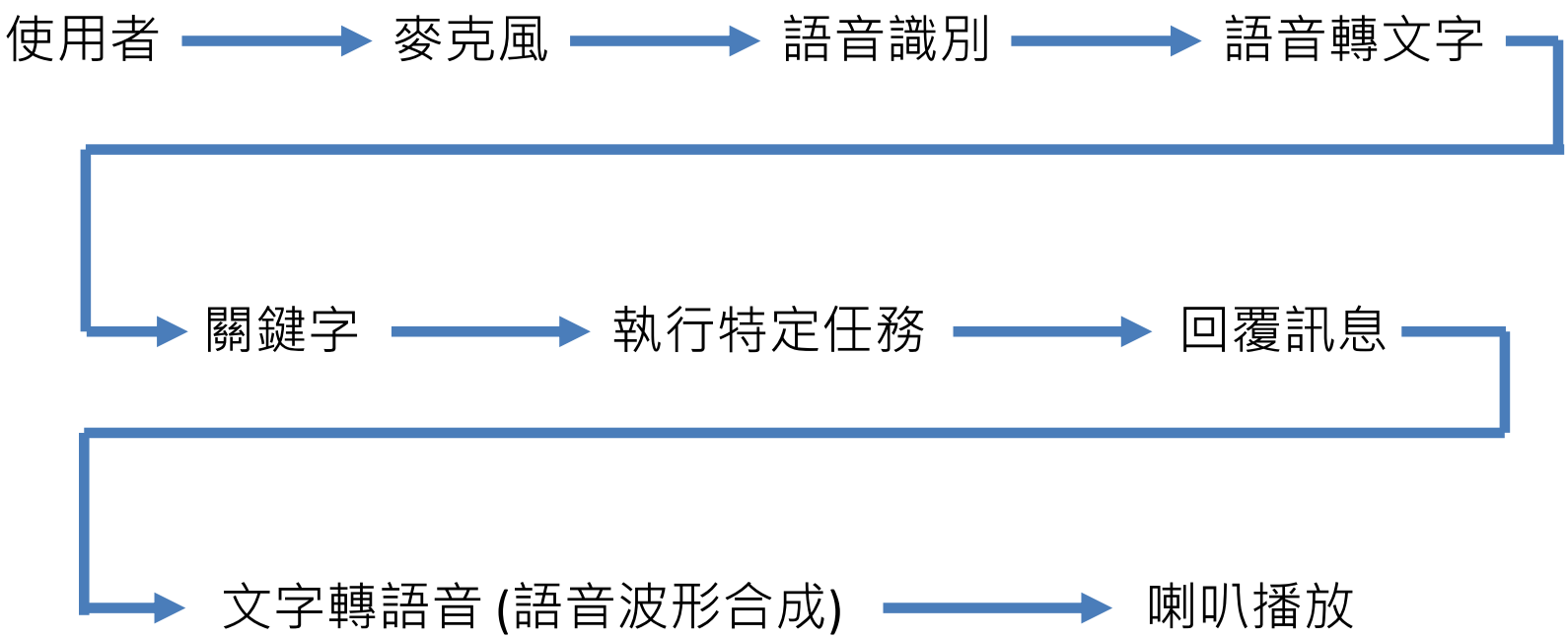
This week

- 嵌入式應用: 語音識別
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)



- 語音識別 (Speech recognition)
  - 自動語音辨識 (Automatic Speech Recognition, ASR)
  - 語音轉文字識別 (Speech To Text, STT)
  - 自然語言處理 (Natural Language Processing, NLP)
  - 語音合成 (Text to Speech, TTS)
  - 大語言模型 (Large Language Model, LLM)

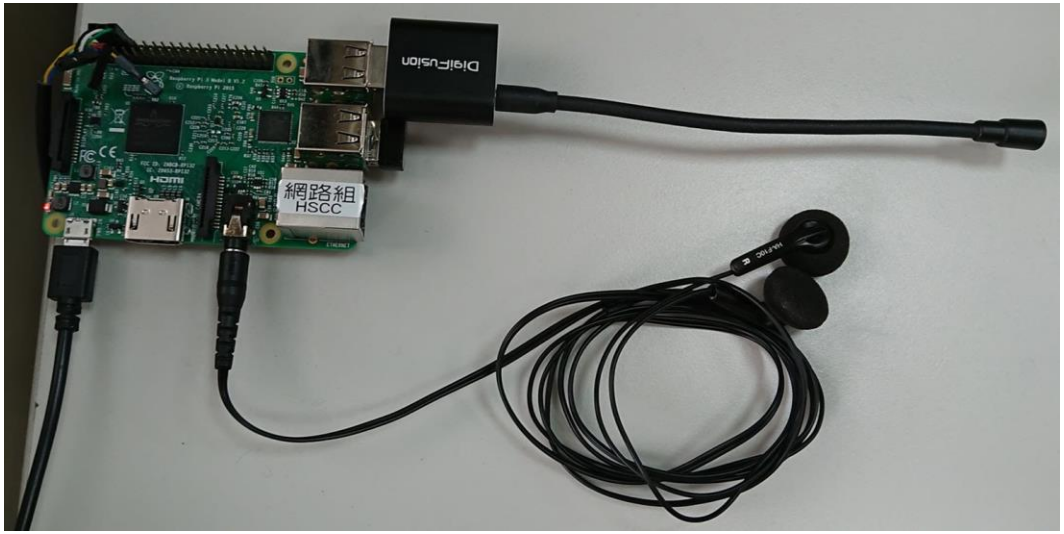
語音識別流程



- Speech recognition
- Speech to text
- Text to speech

Install Dependency

- pip install SpeechRecognition
- pip install gTTS
- sudo apt install libasound2-dev
- sudo apt install python3-pyaudio
- sudo apt install flac



Check your device

- aplay -l

```
pi@raspberrypi:~ $ aplay -l
**** List of PLAYBACK Hardware Devices ****
card 0: Headphones [bcm2835 Headphones], device 0: bcm2835 Headphones [bcm2835 Headphones]
  Subdevices: 8/8
  Subdevice #0: subdevice #0
  Subdevice #1: subdevice #1
  Subdevice #2: subdevice #2
  Subdevice #3: subdevice #3
  Subdevice #4: subdevice #4
  Subdevice #5: subdevice #5
  Subdevice #6: subdevice #6
  Subdevice #7: subdevice #7
card 1: vc4hdmi0 [vc4-hdmi-0], device 0: MAI PCM i2s-hifi-0 [MAI PCM i2s-hifi-0]
  Subdevices: 1/1
  Subdevice #0: subdevice #0
card 2: vc4hdmi1 [vc4-hdmi-1], device 0: MAI PCM i2s-hifi-0 [MAI PCM i2s-hifi-0]
  Subdevices: 1/1
  Subdevice #0: subdevice #0
card 3: Device [USB Audio Device], device 0: USB Audio [USB Audio]
  Subdevices: 1/1
  Subdevice #0: subdevice #0
```

- arecord -l

```
**** List of CAPTURE Hardware Devices ****
card 3: Device [USB Audio Device], device 0: USB Audio [USB Audio]
  Subdevices: 1/1
  Subdevice #0: subdevice #0
```

https://linux.die.net/man/1/arecord

Test and play microphone

- Check device
  - aplay -l
  - arecord -l
- Record your voice
  - arecord -f cd Filename.wav
  - arecord -f cd -d 2 Filename.wav

use “ctrl + c” to stop recording

record 2 seconds

```
pi@raspberrypi:~$ arecord  -f cd Filename.wav
Recording WAVE 'Filename.mp3' : Signed 16 bit Little Endian, Rate 44100 Hz, Stereo
^CAborted by signal Interrupt...

pi@raspberrypi:~$ file Filename.wav
Filename.wav: RIFF (little-endian) data, WAVE audio, Microsoft PCM, 16 bit, stereo
44100 Hz
```

Play record file

Play audio

- Method 1: Use build-in cmd
  - aplay Filename.wav
- Method 2: Use media player
  - sudo apt-get install sox libsox-fmt-all
  - play Filename.wav
- Method 3: Use python
  - See next page

Python sample code

```
import wave
import pyaudio

def play_wav(file_path):
    CHUNK = 1024
    wf = wave.open(file_path, 'rb')
    p = pyaudio.PyAudio()
    stream = p.open(format=p.get_format_from_width(wf.getsampwidth()),
                    channels=wf.getnchannels(),
                    rate=wf.getframerate(),
                    output=True)
    data = wf.readframes(CHUNK)
    while data:
        stream.write(data)
        data = wf.readframes(CHUNK)
    stream.stop_stream()
    stream.close()
    p.terminate()

play_wav('Filename.wav')
```

可是會噴太多警告訊息...

```
pi@raspberrypi:~ $ python pythonwav1.py
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.front.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM front
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.rear
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.center_lfe
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.side
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround21
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround21
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround40.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround40
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround41
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround50
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround51
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround71.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround71
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.iec958.0:CARD=0,AES0=4,AES1=130,AES2=0,AES3=2'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM iec958
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.iec958.0:CARD=0,AES0=4,AES1=130,AES2=0,AES3=2'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM spdif
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.iec958.0:CARD=0,AES0=4,AES1=130,AES2=0,AES3=2'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM spdif
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.hdmi
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.hdmi
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.modem
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.modem
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.phoneline
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.phoneline
Cannot connect to server socket err = No such file or directory
Cannot connect to server request channel
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
Cannot connect to server socket err = No such file or directory
Cannot connect to server request channel
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
ALSA lib pcm_oss.c:377:(_snd_pcm_oss_open) Unknown field port
ALSA lib pcm_oss.c:377:(_snd_pcm_oss_open) Unknown field port
ALSA lib pcm_a52.c:823:(_snd_pcm_a52_open) a52 is only for playback
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.iec958.0:CARD=0,AES0=6,AES1=130,AES2=0,AES3=2'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM iec958:{AES0 0x6 AES1 0x82 AES2 0x0 AES3 0x2  CARD 0}
ALSA lib pcm_usb_stream.c:486:(_snd_pcm_usb_stream_open) Invalid type for card
ALSA lib pcm_usb_stream.c:486:(_snd_pcm_usb_stream_open) Invalid type for card
```

(下面其實還有)

Python sample code 2

(Less warning messages)

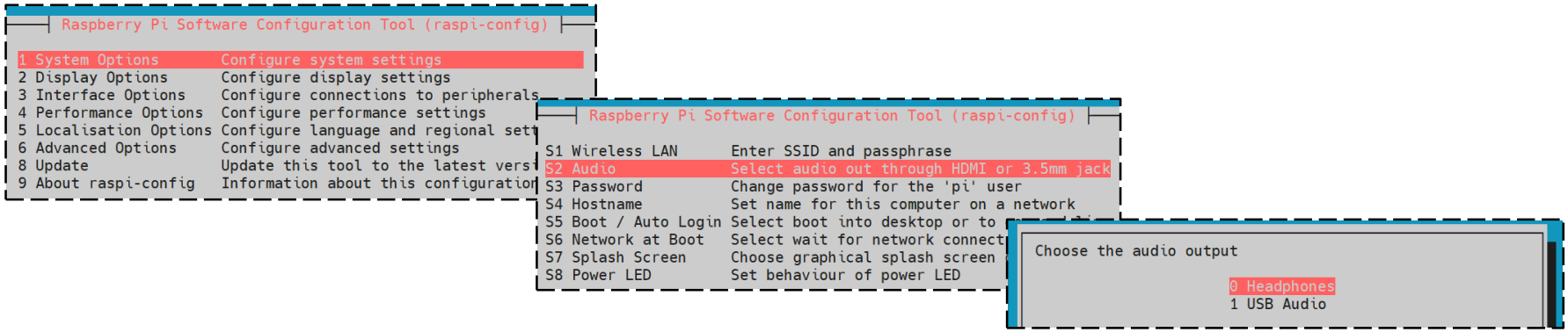
```
  GNU nano 5.4                                        1_playwav.py
from ctypes import *
from contextlib import contextmanager
import pyaudio
import wave

ERROR_HANDLER_FUNC = CFUNCTYPE(None, c_char_p, c_int, c_char_p, c_int, c_char_p)

def py_error_handler(filename, line, function, err, fmt):
    pass

c_error_handler = ERROR_HANDLER_FUNC(py_error_handler)

@contextmanager
def noalsaerr():
    asound = cdll.LoadLibrary('libasound.so')
    asound.snd_lib_error_set_handler(c_error_handler)
    yield
    asound.snd_lib_error_set_handler(None)


def play_wav(file_path):
    with noalsaerr():
        CHUNK = 1024
        wf = wave.open(file_path, 'rb
        p = pyaudio.PyAudio()
        stream = p.open(format=p.get_
                    channels=wf.getnc
                    rate=wf.getframer
                    output=True)
        data = wf.readframes(CHUNK)
        while data:
            stream.write(data)
            data = wf.readframes(CHUN
        stream.stop_stream()
        stream.close()
        p.terminate()

play_wav('fox.wav')
```

sudo apt install libasound2-dev

```
pi@raspberrypi:~ $ python pythonwav.py
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Cannot connect to server socket err = No such file or directory
Cannot connect to server request channel
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
Cannot connect to server socket err = No such file or directory
Cannot connect to server request channel
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
Cannot connect to server socket err = No such file or directory
Cannot connect to server request channel
jack server is not running or cannot be started
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
pi@raspberrypi:~ $
```

疑難排解

- **Q: 播放錄製的檔案沒有聲音?**
- 1. 檢查PI音效輸出設定
  - sudo raspi-config => 1 System Options => S2 Audio
- 2. 把錄製檔案傳到電腦上播放確認內容
  - 如果播放後發現沒聲音, 請參考下方第三點
- 3. 把USB音效卡跟麥克風接到電腦上測試 or 更換一組設備
  - (機率極低) 音效卡的麥克風孔位可能故障
    or 麥克風的阻抗很大, 收音效果很差



Outline

- 嵌入式應用: 語音識別
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)

- 語音識別 (Speech recognition)
  - 自動語音辨識 (Automatic Speech Recognition, ASR)
  - 語音轉文字識別 (Speech To Text, STT)
  - 自然語言處理 (Natural Language Processing, NLP)
  - 語音合成 (Text to Speech, TTS)
  - 大語言模型 (Large Language Model, LLM)

Audio feature extraction

- Short-term feature extraction:
  - It splits the input signal into short-term windows (frames) and computes a number of features for each frame.
- Mid-term feature extraction:
  - extract a number of statistics (e.g. mean and standard deviation) over each short-term feature sequence.

Feature ID	Feature Name	Description
1	Zero Crossing Rate	The rate of sign-changes of the signal during the duration of a particular frame.
2	Energy	The sum of squares of the signal values, normalized by the respective frame length.
3	Entropy of Energy	The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes.
4	Spectral Centroid	The center of gravity of the spectrum.
5	Spectral Spread	The second central moment of the spectrum.
6	Spectral Entropy	Entropy of the normalized spectral energies for a set of sub-frames.
7	Spectral Flux	The squared difference between the normalized magnitudes of the spectra of the two successive frames.
8	Spectral Rolloff	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
9-21	**MFCCs**	**Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.**
22-33	**Chroma Vector**	**A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).**
34	Chroma Deviation	The standard deviation of the 12 chroma coefficients.

Mel-Frequency Cepstral Coefficients

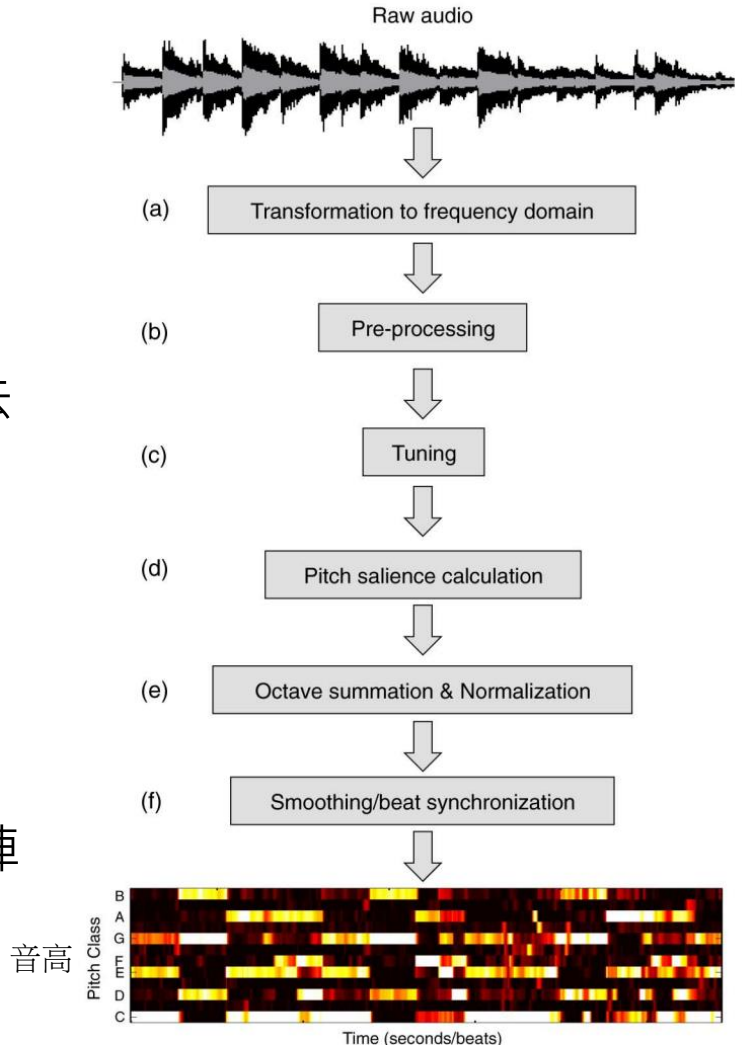
- MFCCs are commonly used as features in speech recognition systems, such as the systems which can automatically recognize numbers spoken into a telephone.

- MFCC(梅爾倒頻譜係數)
  1. Take the Fourier transform of a signal (with sliding window)
  2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
  3. Take the logs of the powers at each of the mel frequencies.
  4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
  5. The MFCCs are the amplitudes of the resulting spectrum.

- Application: music information retrieval
  - audio similarity measures

Chromagram feature (色度圖)

- 計算Chromagram流程
- 1. 將原始音訊從時間域轉成頻率域 (STFT, 短時傅立葉轉換)
- 2. Pre-processing: 取出音高資訊, 需去除背景頻譜、裝飾音、諧波...等
- 3.將音高調整為標準音高
- 4. 平滑處理
- 5. 計算音調, 每個時間檢的音高強度, 可獲得音高的能量分布
- 6. 進行normalization, 可獲得特徵矩陣



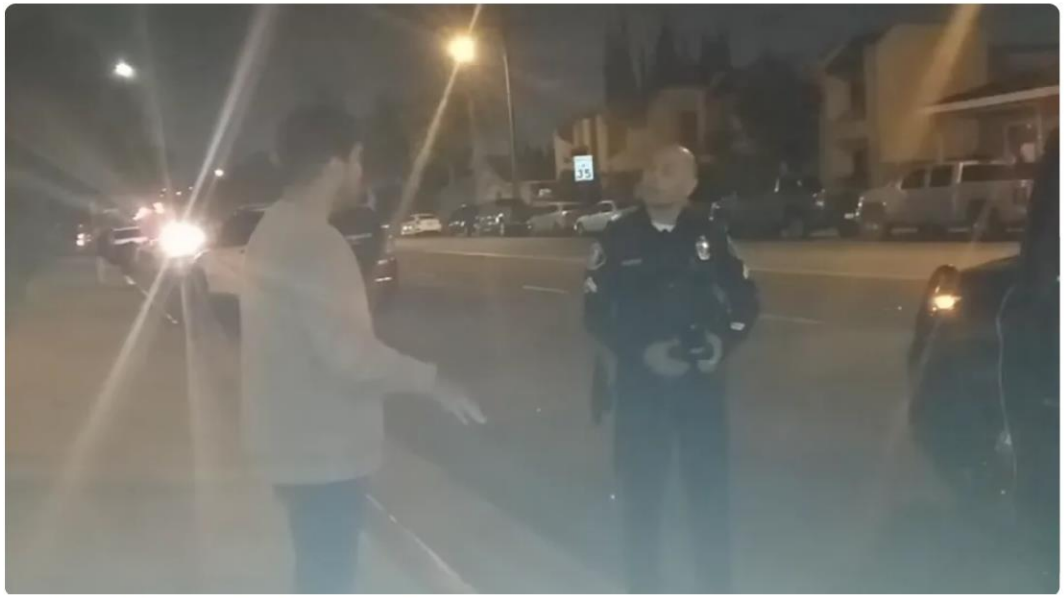
Chromagram

https://tw.news.yahoo.com/disney-074750859.html

YouTube的版權音樂條款

- **為防止民眾拍攝執法現場，美國警察公開播放「迪士尼」版權音樂**

根據美國全國廣播公司（NBC）於 4 月 8 日的報導，當時數輛警車包圍一輛白色汽車，正在調查一起汽車盜竊案。現場一名警察發現有人拿起手機打算錄影時，便打開了巡邏車上的音響，開始大聲播放迪士尼音樂（其中一首音樂為《玩具總動員》You've Got A Friend In Me）。這個行為很顯然是在刻意播放版權音樂，如果相關影片被上傳到各大社群媒體，會因包含受版權保護的音樂而遭刪除。



為防止民眾拍攝執法過程，警察在大半夜大聲播放迪士尼版權音樂

https://www.hackster.io/roni-bandini/reggaeton-be-gone-e5b6e2

Reggaeton Be Gone

- Consider this scenario: Your wall-to-wall neighbor loves to blast Reggaeton music at full volume through a Bluetooth speaker every morning at 9 am. You have two options:
  - A. Knock on their door and politely ask them to lower the volume.
  - B. Build an AI device that can handle the situation more creatively.

Reggaeton redemption: Balvin, Bunny top Latin Grammy nods

At last year's Latin Grammy Awards, popular reggaeton and Latin trap musicians such as J Balvin, Bad Bunny and Ozuna were dismissed in the show's top categories

Via AP news wire • Tuesday 29 September 2020 14:15 BST

Music-Latin Grammy Nominations

Disclaimer: Reggaeton Be Gone is an experimental project. Before deploying it, check your local laws and regulations. Use it only with your own Bluetooth speakers for educational purposes. Also, keep in mind that you need to be close enough to the speaker for the RPI BT to reach the speaker, and not all Bluetooth speakers are vulnerable. Last but not least: I have nothing against this or other music genres. It was purelycoincidental.

pyAudioAnalysis

- Install
  - git clone https://github.com/tyiannak/pyAudioAnalysis.git
  - cd pyAudioAnalysis
  - pip install -r ./requirements.txt
  - pip install -e .

- Chromagram visualization
  - python pyAudioAnalysis/audioAnalysis.py fileChromagram -i your_file

修改 requirements.txt

```
1   matplotlib>=3.4.2
2   simplejson>=3.16.0
3   scipy>=1.6.3
4   numpy>=1.20.3
5   hmmlearn>=0.2.5
6   eyeD3>=0.9.6
7   pydub>=0.25.1
8   scikit_learn>=0.24.2
9   tqdm>=4.52.0
10  plotly>=5.3.1
11  pandas>=1.2.4
12  imblearn
```

numpy==1.23.5

plotly==1.23.5

因為最新版的語法有調整, 請安裝舊版的套件

audioAnalysis.py 功能

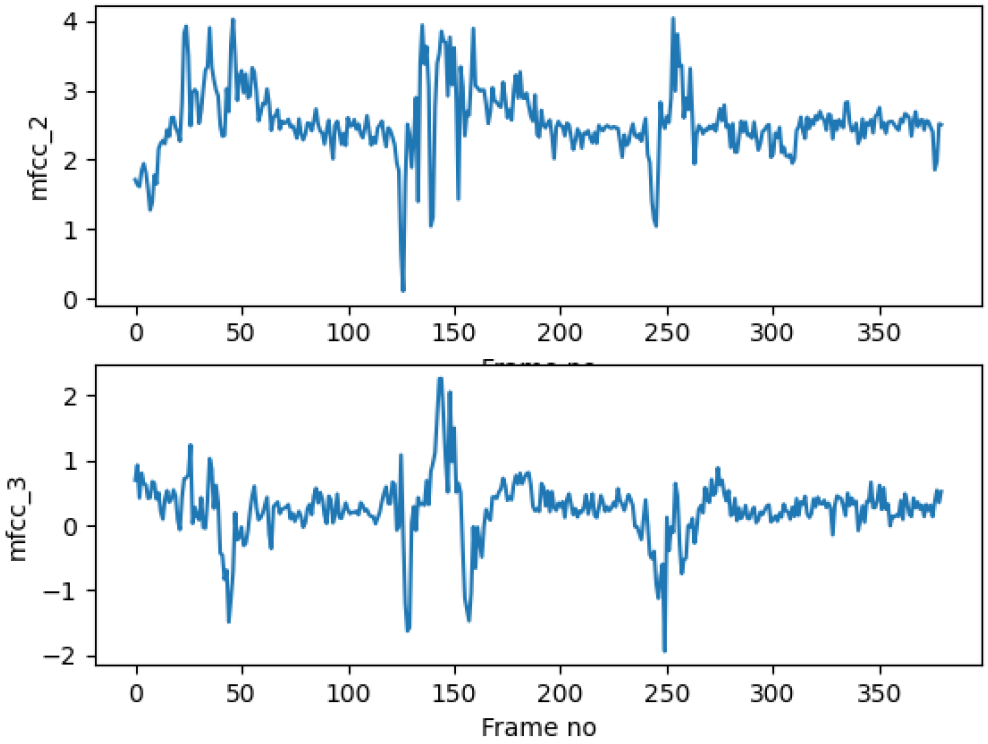
- 檔案格式轉換
  - dirMp3toWavWrapper, dirWAVChangeFs
  - 將MP3轉換為 WAV；調整WAV採樣率。
- 特徵提取
  - featureExtractionFileWrapper, featureExtractionDirWrapper, beatExtractionWrapper
  - 對單一/目錄中的WAV提取音訊特徵；提取節拍特徵。
- 特徵視覺化
  - featureVisualizationDirWrapper: 進行特徵可視化。
- 分類和回歸
  - trainClassifierWrapper, classifyFileWrapper, trainRegressionWrapper, regressionFileWrapper, classifyFolderWrapper, regressionFolderWrapper
  - 訓練分類器；使用訓練好的模型進行分類；訓練回歸模型；應用回歸模型；對文件進行分類與回歸。

audioAnalysis.py 功能

- 段落識別
  - trainHMMsegmenter_fromfile, trainHMMsegmenter_fromdir: 訓練 HMM 模型進行段落識別
  - segmentclassifyFileWrapper, segmentclassifyFileWrapperHMM: 對 WAV 文件進行段落分類
- 靜音去除與說話者辨識
  - silenceRemovalWrapper: 檢測音軌的非靜音段落
  - speakerDiarizationWrapper: 執行說話者辨識
- 影音縮略圖
  - thumbnailWrapper: 幫音軌生成縮圖

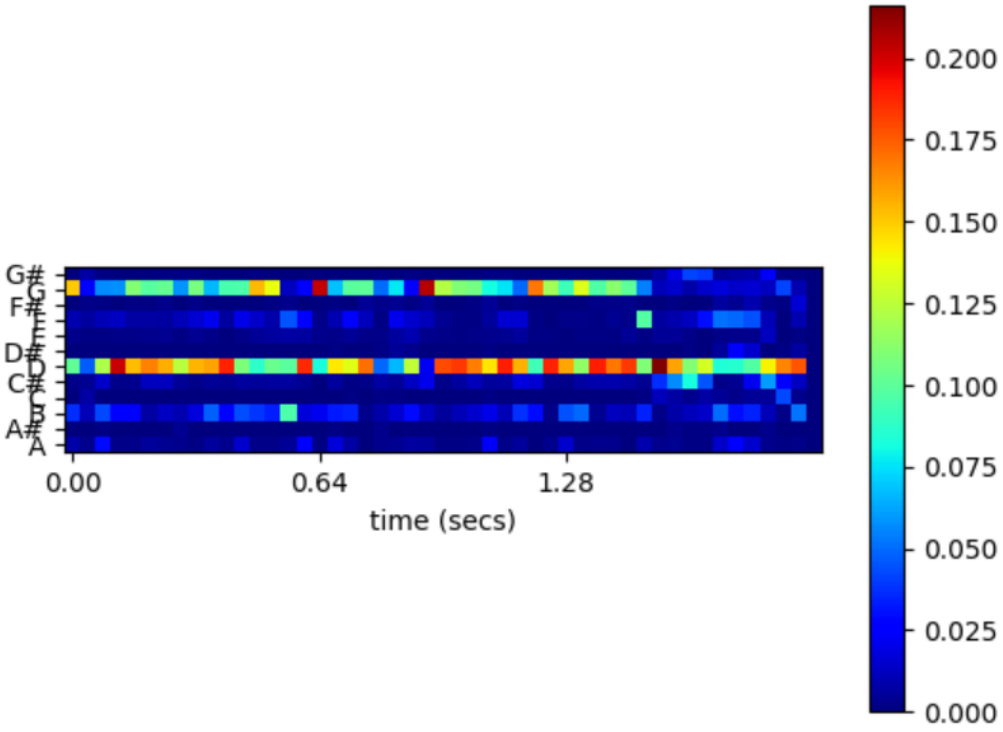
MFCC

Feature ID	Feature Name	Description
9-21	**MFCCs**	**Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.**



Chromagram visualization

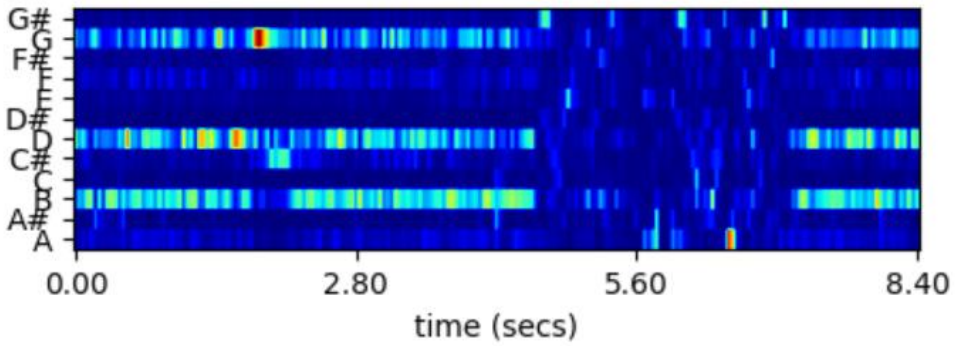
Feature ID	Feature Name	Description
22-33	**Chroma Vector**	**A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).**

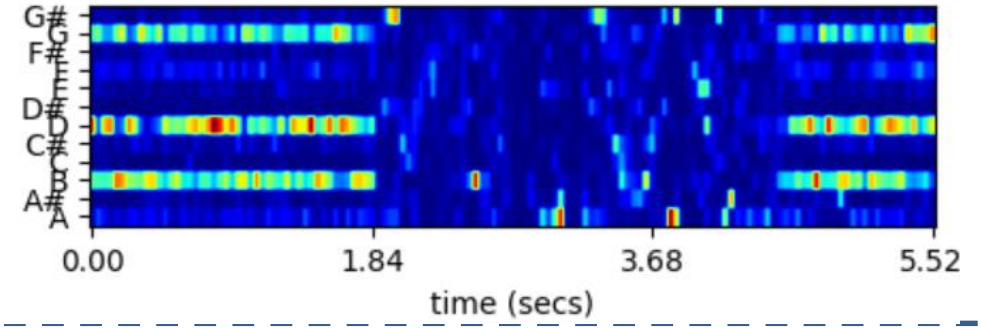


Discussion 1

- Try to say the same sentence twice, then plot the chromagram for both result.

Example: (The quick fox jumps over the lazy dog)





Outline

- 嵌入式應用: 語音識別
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)

  - 語音識別 (Speech recognition)
    - 自動語音辨識 (Automatic Speech Recognition, ASR)
    - 語音轉文字識別 (Speech To Text, STT)
    - 自然語言處理 (Natural Language Processing, NLP)
    - 語音合成 (Text to Speech, TTS)
    - 大語言模型 (Large Language Model, LLM)

Speech Recognition

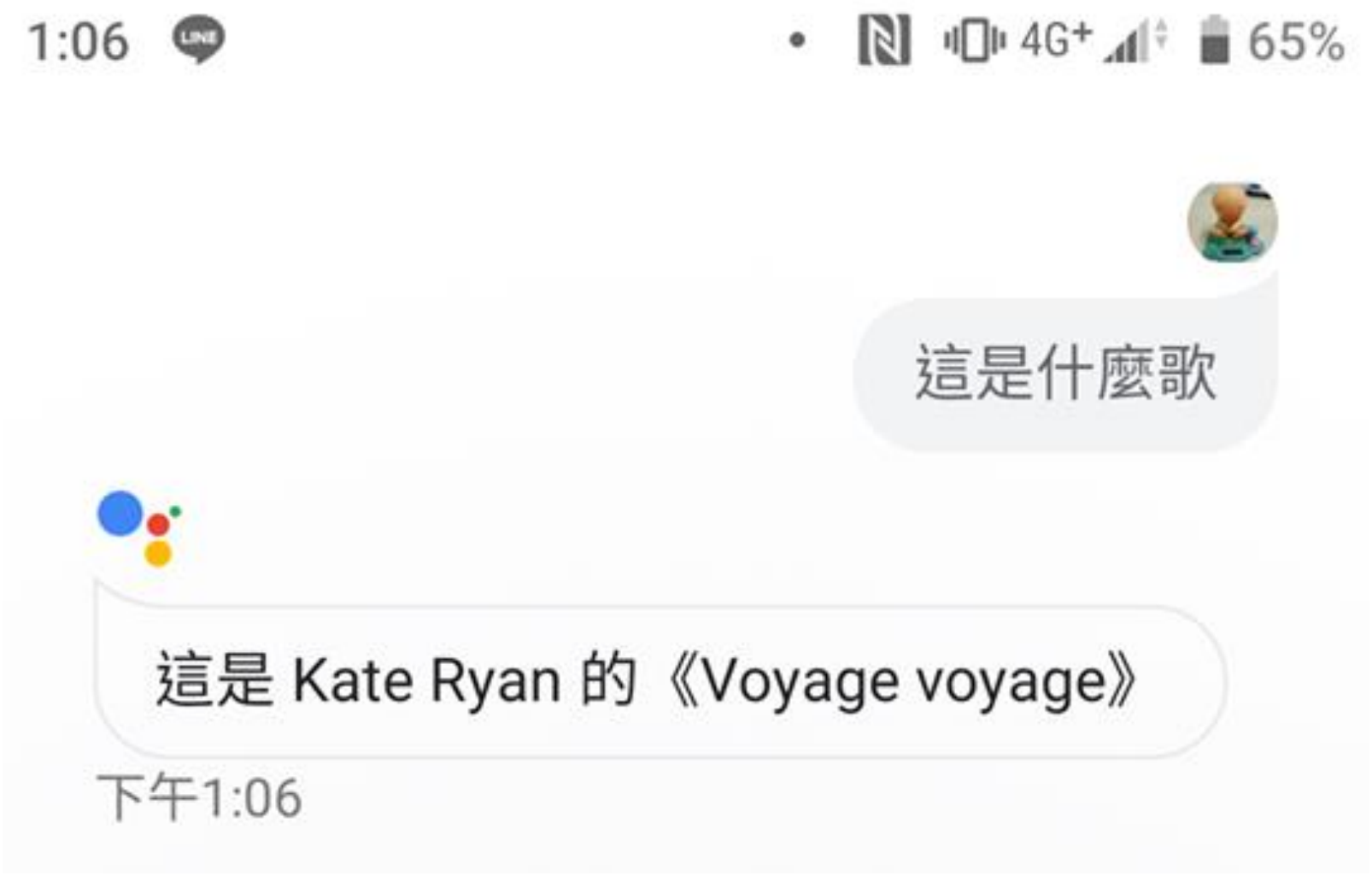
- 聲音被麥克風收集並轉換成數位信號，然後送入語音辨識系統中，利用演算法和模型來**分析聲音特徵**，並將其轉換成文字或指令。
- 自然語言處理是讓機器能夠像人類一樣理解語言。包含語言中的詞彙、句法、語義和語境，並從中提取有用的信息。也可以用來生成自然語言，包含機器翻譯、對話系統等應用。
- 大型語言模型是一種基於深度學習的人工智慧模型，具有驚人的能力來理解和生成自然語言。這些模型通常由數十億到數百億個參數組成，使用大量的文本進行訓練，能夠捕捉語言中的豐富訊息。
  - 代表作品：GPT（Generative Pre-trained Transformer）

https://www.ithome.com.tw/news/160091

LLM與電腦算力

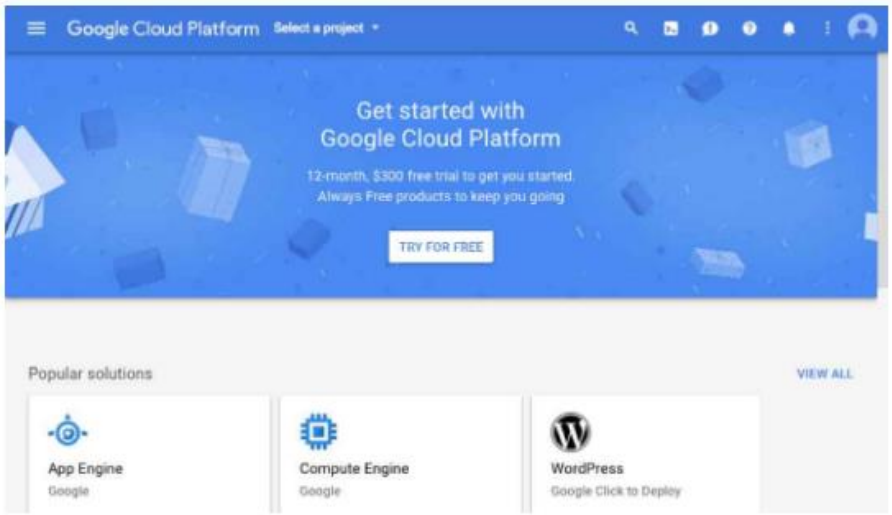
- (國網中心) 臺灣杉2號採V100 GPU，專為AI模型開發和推論而設計，運算效能可達9 PFLOPS。
- 國網中心主任張朝亮指出，以Meta開源模型Llama 2為例，它有70億參數（7B）、130億參數（13B）和700億參數（70B）版本，在標準條件下，進行7B、13B模型預訓練和全參數微調，臺灣杉2號都能應付。但若是70B參數的模型預訓練，國網中心算力可能就不太夠了。
- Meta從無到有訓練Llama 2時，需要上千甚至上萬片A100 GPU($100萬)，所需時間大約為6個月，而臺灣杉2號採用相對低階的V100 GPU($30萬)，效能約為1：3。若以臺灣杉2號進行70B模型預訓練，可能得花上9個月至1年。

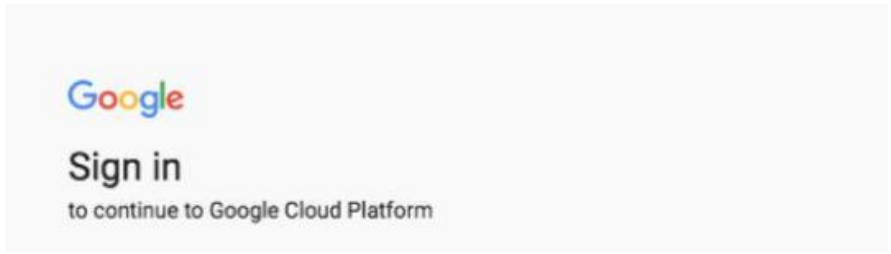
Google assistant



Voice Kit

- Do-it-yourself intelligent speaker. Experiment with voice recognition and the Google Assistant.





1. Head to the Google Cloud Platform

In order to make use of the Google Assistant and Cloud Speech APIs, you need to get credentials from Google's developer console.

On your computer (not the Raspberry Pi), go to https://console.cloud.google.com/.

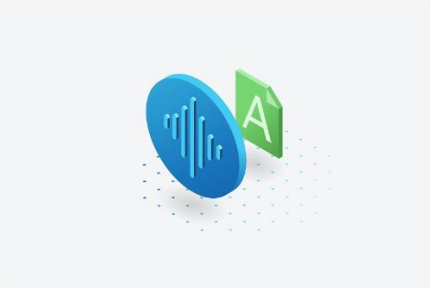
2. Login using your Google account

Log in using your Google account.

Don't have a Google account? Sign up for one here.

https://aiyprojects.withgoogle.com/voice/

Azure



語音轉換文字

快速而精確地將音訊轉譯成 100 種以上的語言及各種不同版本。透過話務中心的謄寫、改善具備語音功能之助理程式的體驗、掌握會議中的重要議題等等，以取得客戶的見解。

深入了解語音轉換文字 >

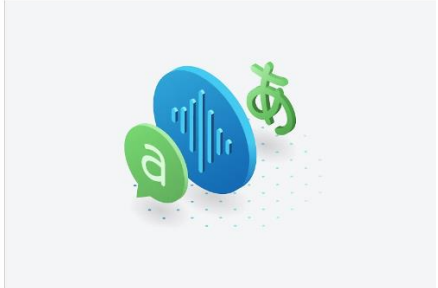
立即開始轉譯語音 >

為您的應用程式賦予聲音

使用文字轉換語音來建立可交談的應用程式和服務。建立自然發音的音訊內容、利用大聲朗讀的功能來改善協助工具，及建立自訂語音助理。

深入了解文字轉換語音 >

了解將文字轉換成音訊的方法 >

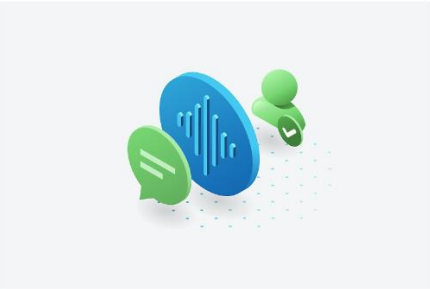


即時翻譯語音

全部依照您常用的程式設計語言，翻譯來自 30 多種語言的音訊，並可為您組織的專用術語自訂翻譯。

深入了解語音翻譯 >

開始進行即時翻譯語音 >

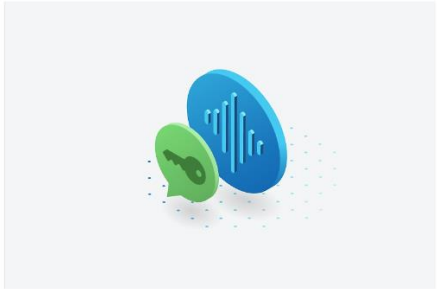


驗證和辨識說話者

將說話者的驗證和識別加入至您的應用程式中，藉以確認某人的身分識別，或於會議中辨識說話者。

深入了解說話者辨識 >

了解如何在應用程式中辨識說話者 >

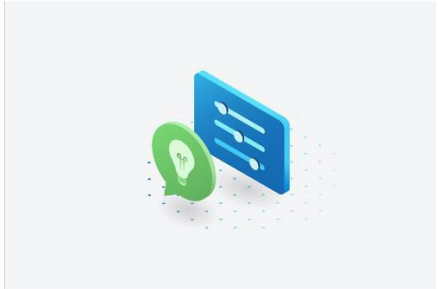


使用自訂關鍵字來啟動您的助理或 IoT 裝置

為 IoT 裝置和具備語音功能的助理建立自訂的關鍵字，可讓您打造更個人化、更有吸引力且更安全的特色品牌。

了解如何建立自訂關鍵字 >

開始建立語音控制的應用程式 >



新增適用於免持聽筒案例的語音命令

建立免觸控的語音優先體驗，以提升安全性並支援返回工作案例。

深入了解自訂命令 >

開始新增自訂命令 >

https://azure.microsoft.com/zh-tw/services/cognitive-services/speech-services/

OpenAI

Speech to text

Copy page

Learn how to turn audio into text.

The Audio API provides two speech to text endpoints:

- `transcriptions`
- `translations`

Historically, both endpoints have been backed by our open source Whisper model (`whisper-1`). The `transcriptions` endpoint now also supports higher quality model snapshots, with limited parameter support:

- `gpt-4o-mini-transcribe`
- `gpt-4o-transcribe`

All endpoints can be used to:

- Transcribe audio into whatever language the audio is in.
- Translate and transcribe the audio into English.

File uploads are currently limited to 25 MB, and the following input file types are supported: `mp3` , `mp4` , `mpeg` , `mpga` , `m4a` , `wav` , and `webm` .

https://platform.openai.com/docs/guides/speech-to-text

https://pypi.org/project/SpeechRecognition/

SpeechRecognition

- Library for performing speech recognition, with support for several engines and APIs, online and offline.
- Speech recognition engine/API support:
  - CMU Sphinx (works offline) (卡內基大學)
  - **Google Speech Recognition**
  - Google Cloud Speech API
  - Wit.ai (Meta)
  - Microsoft Azure Speech
  - Houndify API (SoundHound,音樂識別平台)
  - IBM Speech to Text
  - Snowboy Hotword Detection (works offline)
  - Tensorflow
  - Vosk API (works offline)
  - OpenAI whisper (works offline)
  - **OpenAI Whisper API**
  - Groq Whisper API

More examples: https://github.com/Uberi/speech_recognition/tree/master/examples

Outline

- 嵌入式應用: 語音識別
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)

  - 語音識別 (Speech recognition)
    - 自動語音辨識 (Automatic Speech Recognition, ASR)
    - 語音轉文字識別 (Speech To Text, STT)
    - 自然語言處理 (Natural Language Processing, NLP)
    - 語音合成 (Text to Speech, TTS)
    - 大語言模型 (Large Language Model, LLM)

Usage: python 2.stt_microphone.py

Speech to text (microphone)

```
import speech_recognition as sr

#obtain audio from the microphone
r=sr.Recognizer()

with sr.Microphone() as source:
    print("Please wait. Calibrating microphone...")
    #listen for 1 seconds and create the ambient noise energy level
    r.adjust_for_ambient_noise(source, duration=1)
    print("Say something!")
    audio=r.listen(source)

# recognize speech using Google Speech Recognition
try:
    print("Google Speech Recognition thinks you said:")
    print(r.recognize_google(audio))
except sr.UnknownValueError:
    print("Google Speech Recognition could not understand audio")
except sr.RequestError as e:
    print("No response from Google Speech Recognition service: {0}".format(e))
```

Noise Suppression

Usage: python 2.stt_file.py

Speech to text (audio file)

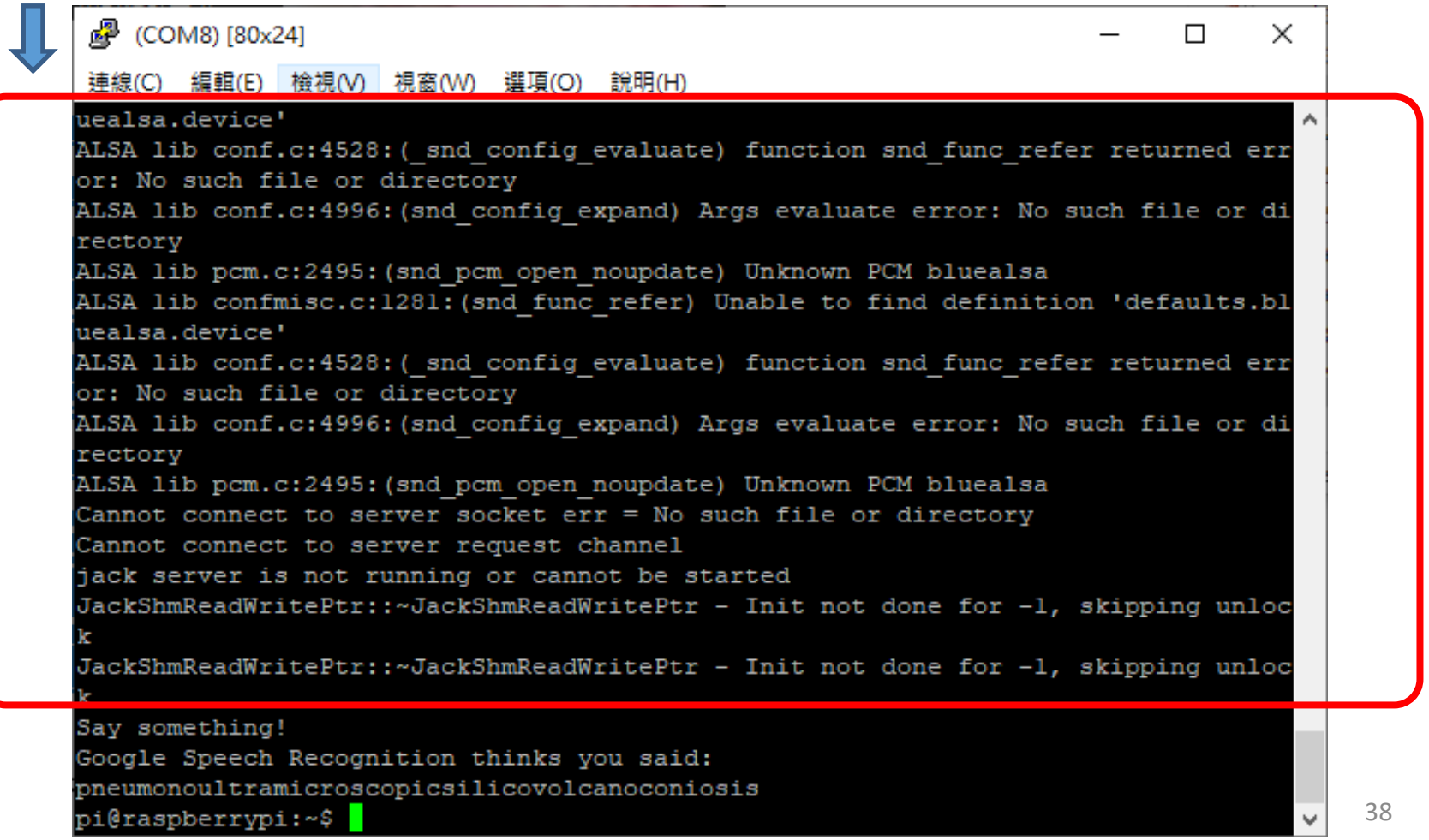
```
import speech_recognition as sr

#obtain audio from the microphone
r=sr.Recognizer()

myvoice = sr.AudioFile('hello.wav')
with myvoice as source:
   print("Use audio file as input!")
   audio = r.record(source)

# recognize speech using Google Speech Recognition
try:
   print("Google Speech Recognition thinks you said:")
   print(r.recognize_google(audio))
except sr.UnknownValueError:
   print("Google Speech Recognition could not understand audio")
except sr.RequestError as e:
   print("No response from Google Speech Recognition service: {0}".format(e))
```

Input format: PCM WAV, AIFF/AIFF-C, or Native FLAC

If you want to use MP3, you might need: **ffmpeg -i input.mp3 output.flac**

Speech to text (result)

You can ignore the ALSA warning messages when using microphone



https://pypi.org/project/SpeechRecognition/
https://github.com/Uberi/speech_recognition/blob/master/examples/audio_transcribe.py

SpeechRecognition

Speech recognition engine/API support:

- CMU Sphinx (works offline)
- **Google Speech Recognition**
- Google Cloud Speech API
- Wit.ai
- Microsoft Bing Voice Recognition
- …
- **OpenAI Whisper API**

- r.recognize_sphinx(audio)
- **r.recognize_google(audio)**
- r.recognize_google_cloud(audio, credentials_json=GOOGLE_CLOUD_SPEECH_CREDENTIALS)
- r.recognize_wit(audio, key=WIT_AI_KEY)
- r.recognize_azure(audio, key=AZURE_SPEECH_KEY)
- r.recognize_bing(audio, key=BING_KEY)
- r.recognize_houndify(audio, client_id=HOUNDIFY_CLIENT_ID, client_key=HOUNDIFY_CLIENT_KEY)
- r.recognize_ibm(audio, username=IBM_USERNAME, password=IBM_PASSWORD)

Speech Recognition Library Reference
https://github.com/Uberi/speech_recognition/blob/master/reference/library-reference.rst

r.recognize_google(audio)

- recognizer_instance.recognize_google(audio_data: AudioData, key: Union[str, None] = None, language: str = "en-US", , pfilter: Union[0, 1], show_all: bool = False) -> Union[str, Dict[str, Any]]

- The Google Speech Recognition API key is specified by **key**. **If not specified, it uses a generic key that works out of the box**. This should generally be used for personal or testing purposes only, as it may be revoked by Google at any time.
- Note that the **API quota for your own keys is 50 requests** per day.
- The recognition language is determined by language, an IETF language tag like "en-US" or "en-GB", **defaulting to US English**. A list of supported language tags can be found on the website. Basically, language codes can be just the language (en), or a language with a dialect (en-US).

Speech Recognition Library Reference
https://github.com/Uberi/speech_recognition/blob/master/reference/library-reference.rst

STT by Whisper API

Require: pip install openai

```
import speech_recognition as sr
Import os

#obtain audio from the microphone
r=sr.Recognizer()

with sr.Microphone() as source:
    print("Please wait. Calibrating microphone...")
    #listen for 1 seconds and create the ambient noise energy level
    r.adjust_for_ambient_noise(source, duration=1)
    print("Say something!")
    audio=r.listen(source)

# recognize speech using Whisper API
OPENAI_API_KEY = "INSERT OPENAI API KEY HERE"
os.environ["OPENAI_API_KEY"] = OPENAI_API_KEY
try:
    print(f"OpenAI Whisper API thinks you said {r.recognize_openai(audio)}")
except sr.RequestError as e:
    print(f"Could not request results from OpenAI Whisper API; {e}")
```

(We will provide API key to everyone!)

Quiz 1

- There are too many warning messages when running “Speech to text (microphone)”
- Please combine the previous sample code to reduce the warning messages.
- Set your own openai API key

```
pi@raspberrypi:~ $ python 2.stt_microphone.py
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
Expression 'GetExactSampleRate( hwParams, &defaultSr )' failed in 'src/hostapi/alsa/pa_linux_alsa.c', line: 895
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.front.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM front
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.rear
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.center_lfe
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM cards.pcm.side
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
ALSA lib conf.c:4745:(_snd_config_evaluate) function snd_func_refer returned error: No such file or directory
ALSA lib conf.c:5233:(snd_config_expand) Evaluate error: No such file or directory
ALSA lib pcm.c:2660:(snd_pcm_open_noupdate) Unknown PCM surround21
ALSA lib confmisc.c:1281:(snd_func_refer) Unable to find definition 'cards.bcm2835_hdmi.pcm.surround51.0:CARD=0'
```

(下方省略)

Outline

- 嵌入式應用: 語音識別
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)

- 語音識別 (Speech recognition)
  - 自動語音辨識 (Automatic Speech Recognition, ASR)
  - 語音轉文字識別 (Speech To Text, STT)
  - 自然語言處理 (Natural Language Processing, NLP)
  - 語音合成 (Text to Speech, TTS)
  - 大語言模型 (Large Language Model, LLM)

Text to speech

```
from gtts import gTTS
import os

tts = gTTS(text='hello', lang='en')
tts.save('hello.mp3')

os.system('play hello.mp3 > /dev/null 2>&1')
```

The output format is mp3!

Parameters:

- **text** (*string*) – The text to be read.
- **lang** (*string, optional*) – The language (IETF language tag) to read the text in. Defaults to 'en'.
- **slow** (*bool, optional*) – Reads text more slowly. Defaults to `False`.
- **lang_check** (*bool, optional*) – Strictly enforce an existing `lang`, to catch a language error early. If set to `True`, a `ValueError` is raised if `lang` doesn't exist. Default is `True`.

https://pypi.org/project/gTTS/
https://gtts.readthedocs.io/en/latest/

gTTS (Google Text-to-Speech)

- An interface to Google Translator's Text-to-Speech API.

gTTS (`gtts.gTTS`)

class **gtts.tts.gTTS**(*text, lang='en', slow=False, lang_check=True, pre_processor_funcs=[<function tone_marks>, <function end_of_line>, <function abbreviations>, <function word_sub>], tokenizer_func=<bound method Tokenizer.run of re.compile('(?<=\?).|(?<=\!).|(?<=\ ?).|(?<=\ !).|(?<!\.[a-z])\. |(?<!\.[a-z])\, |(?<!\d)\:|\;|\—|\ ، |\...|\ ։|\n|\¿|\(|\ ، |\[|\)|\¡|\؟|\..|\]|\ ۔', re.IGNORECASE) from: [<function tone_marks>, <function period_comma>, <function colon>, <function other_punctuation>]>*) [source]

Parameters:

- **text** (*string*) – The text to be read.
- **lang** (*string, optional*) – The language (IETF language tag) to read the text in. Defaults to 'en'.
- **slow** (*bool, optional*) – Reads text more slowly. Defaults to `False`.
- **lang_check** (*bool, optional*) – Strictly enforce an existing `lang`, to catch a language error early. If set to `True`, a `ValueError` is raised if `lang` doesn't exist. Default is `True`.

Offline text-to-speech

- pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline.

- Installation
  - sudo apt install espeak ffmpeg
  - pip install pyttsx3

```
import pyttsx3

engine = pyttsx3.init()
engine.say("hello")
engine.runAndWait()
```

Changing Voice , Rate and Volume :

```
"""VOICE"""
voices = engine.getProperty('voices')          #getting details of current voice
#engine.setProperty('voice', voices[0].id)     #changing index, changes voices. o for male
engine.setProperty('voice', voices[1].id)      #changing index, changes voices. 1 for female

""" RATE"""
rate = engine.getProperty('rate')              # getting details of current speaking rate
print (rate)                                   #printing current voice rate
engine.setProperty('rate', 125)                # setting up new voice rate

"""VOLUME"""
volume = engine.getProperty('volume')          #getting to know current volume level (min=0 and max=1)
print (volume)                                 #printing current volume level
engine.setProperty('volume',1.0)               # setting up volume level  between 0 and 1

"""Saving Voice to a file"""
# On linux make sure that 'espeak' and 'ffmpeg' are installed
engine.save_to_file('Hello World', 'test.mp3')
engine.runAndWait()
```

TTS by OpenAI

```
from pathlib import Path
from openai import OpenAI
import os

OPENAI_API_KEY = "your_API_key"
os.environ["OPENAI_API_KEY"] = OPENAI_API_KEY

client = OpenAI()
speech_file_path = Path(__file__).parent / "speech.mp3"
response = client.audio.speech.create(
  model="gpt-4o-mini-tts",
  voice="coral",
  input="Today is a wonderful day to build something people love!",
  instructions="Speak in a cheerful and positive tone.",
)
response.with_streaming_response().save_to_file(speech_file_path)
```

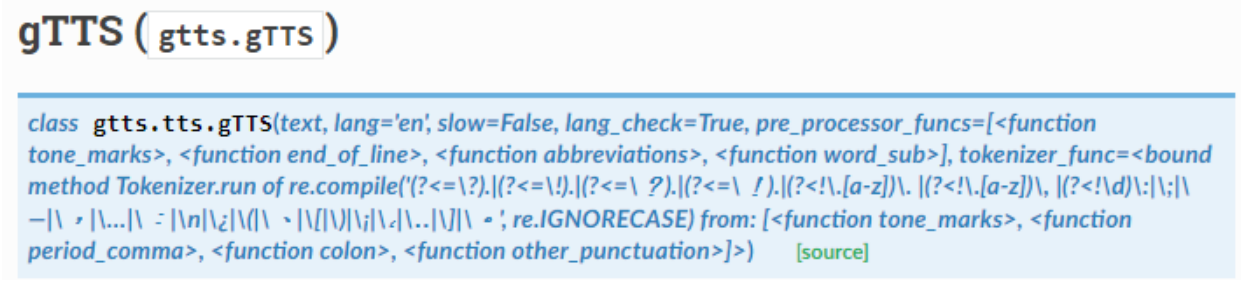
Read more: https://platform.openai.com/docs/guides/text-to-speech

TTS by OpenAI

- Text-to-speech models
  - Newest: gpt-4o-mini-tts
  - Our other text-to-speech models are tts-1 and tts-1-hd. The tts-1 model provides lower latency, but at a lower quality than the tts-1-hd model.
- Voice options
  - Voices are currently optimized for English.
- Supported output formats
  - MP3, WAV, AAC, FLAC, PCM …

Discussion 2

- **Q: How to make the model to speak other language?**
- A. For gTTS: how to speak other language?



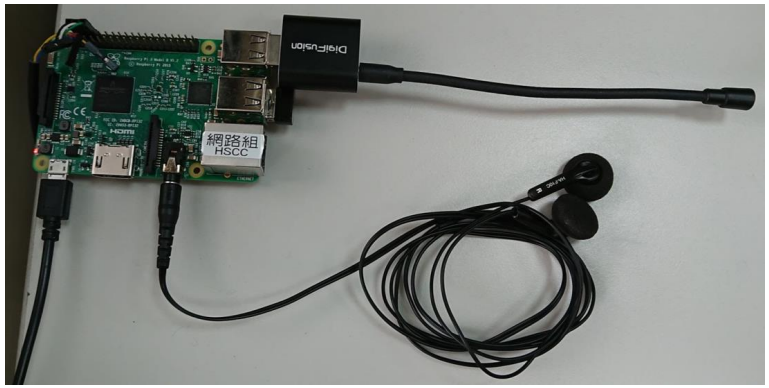
gTTS (`gtts.gTTS`)

class **gtts.tts.gTTS**(*text, lang='en', slow=False, lang_check=True, pre_processor_funcs=[<function tone_marks>, <function end_of_line>, <function abbreviations>, <function word_sub>], tokenizer_func=<bound method Tokenizer.run of re.compile('(?<=\?).|(?<=\!).|(?<=\ ?).|(?<=\ !).|(?<!\.[a-z])\. |(?<!\.[a-z])\, |(?<!\d)\:|\;|\—|\ ، |\...|\ ：|\n|\¿|\(|\ ، |\[|\)|\¡|\،|\..|\]|\ ۔', re.IGNORECASE) from: [<function tone_marks>, <function period_comma>, <function colon>, <function other_punctuation>]>*) [source]

- B. For pyttsx3: how to make this offline model to speak other language?
- C. For openai: how to speak other language?

Quiz 2

- Say a custom command to Raspberry PI, it will start to measure the distance. (provided by HC-SR04)
  - Input could be microphone or audio file
    - gTTS can be used to generate an audio file as input
- After measuring distance, use gTTS (Google Text-to-Speech) to speak out the result.
  - Ex: the distance is 30 cm



Quiz 2 (cont.)

- Input options (chose one):
  1. Microphone: talk to microphone directly

  2. Audio file: recode your voice on PC, then send it to Raspberry PI. The file should be PCM WAV, AIFF/AIFF-C, or Native FLAC.

  3. gTTS: generate the audio file from text
     - The default output is mp3. Use the following command to convert
     - ffmpeg -i input.mp3 output.wav

補充資料: openwakeword

- An open-source audio wake word (or phrase) detection framework with a focus on performance and simplicity.
- openWakeWord has four high-level goals, which combine to (hopefully!) produce a framework that is simple to use *and* extend.
  1. Be fast *enough* for real-world usage
  2. Be accurate *enough* for real-world usage.
  3. Have a simple model architecture and inference process.
  4. Require **little to no manual data collection** to train new models.

補充資料: spaCy

- spaCy is a library for **advanced Natural Language Processing** in Python and Cython. It's built on the very latest research, and was designed from day one to be used in real products.

Edit the code & try spaCy spaCy v3.7 · Python 3 · via Binder

```
# pip install -U spacy
# python -m spacy download en_core_web_sm
import spacy

# Load English tokenizer, tagger, parser and NER
nlp = spacy.load("en_core_web_sm")

# Process whole documents
text = ("When Sebastian Thrun started working on self-driving cars at "
        "Google in 2007, few people outside of the company took him "
        "seriously. “I can tell you very senior CEOs of major American "
        "car companies would shake my hand and turn away because I wasn’t "
        "worth talking to,” said Thrun, in an interview with Recode earlier "
        "this week.")
doc = nlp(text)

# Analyze syntax
print("Noun phrases:", [chunk.text for chunk in doc.noun_chunks])
print("Verbs:", [token.lemma_ for token in doc if token.pos_ == "VERB"])

# Find named entities, phrases and concepts
for entity in doc.ents:
    print(entity.text, entity.label_)
```

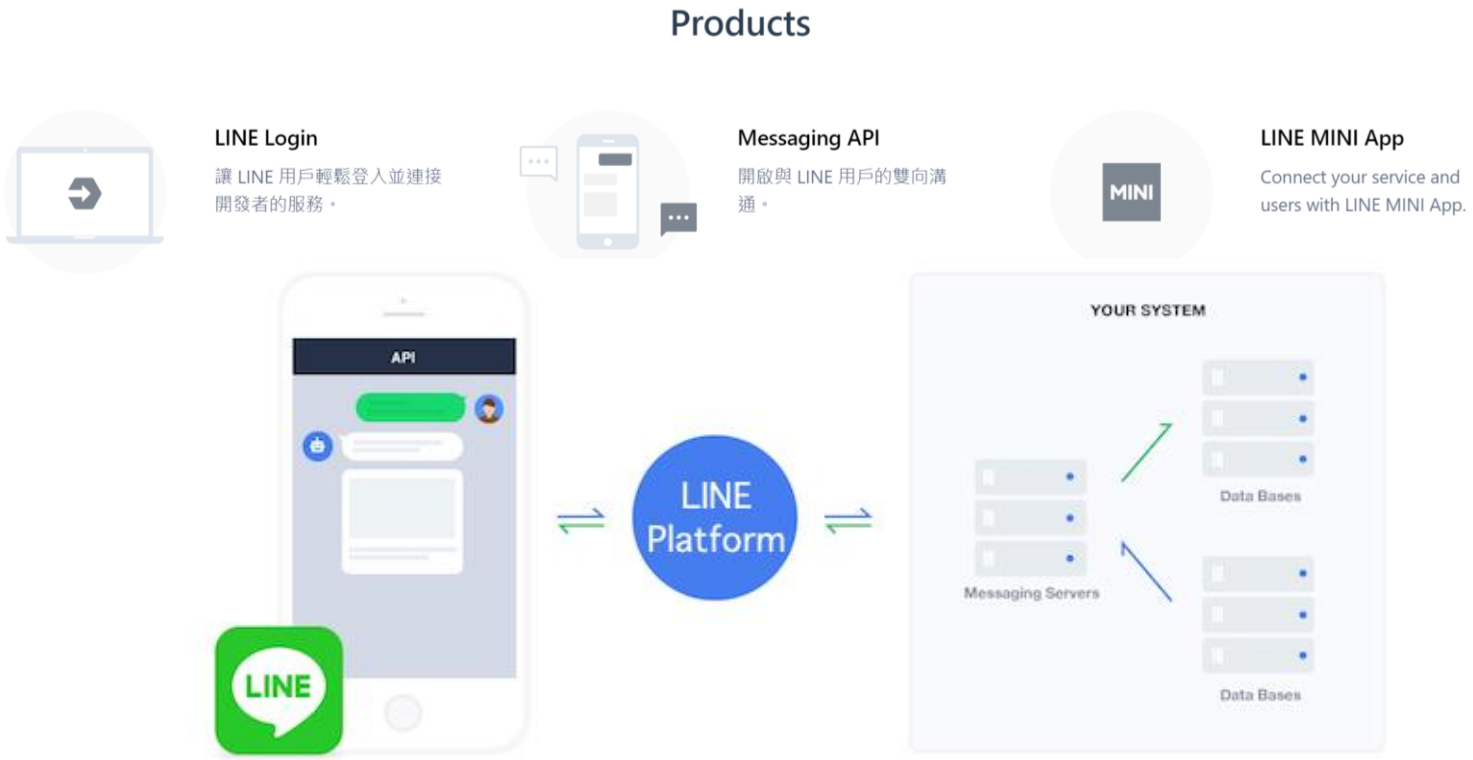
RUN

Features

- Support for **75+ languages**
- **84 trained pipelines** for 25 languages
- Multi-task learning with pretrained **transformers** like BERT
- Pretrained **word vectors**
- State-of-the-art speed
- Production-ready **training system**
- Linguistically-motivated **tokenization**
- Components for **named entity** recognition, part-of-speech tagging, dependency parsing, sentence segmentation, **text classification**, lemmatization, morphological analysis, entity linking and more
- Easily extensible with **custom components** and attributes
- Support for custom models in **PyTorch**, **TensorFlow** and other frameworks
- Built in **visualizers** for syntax and NER
- Easy **model packaging**, deployment and workflow management
- Robust, rigorously evaluated accuracy

補充資料: Chatbot

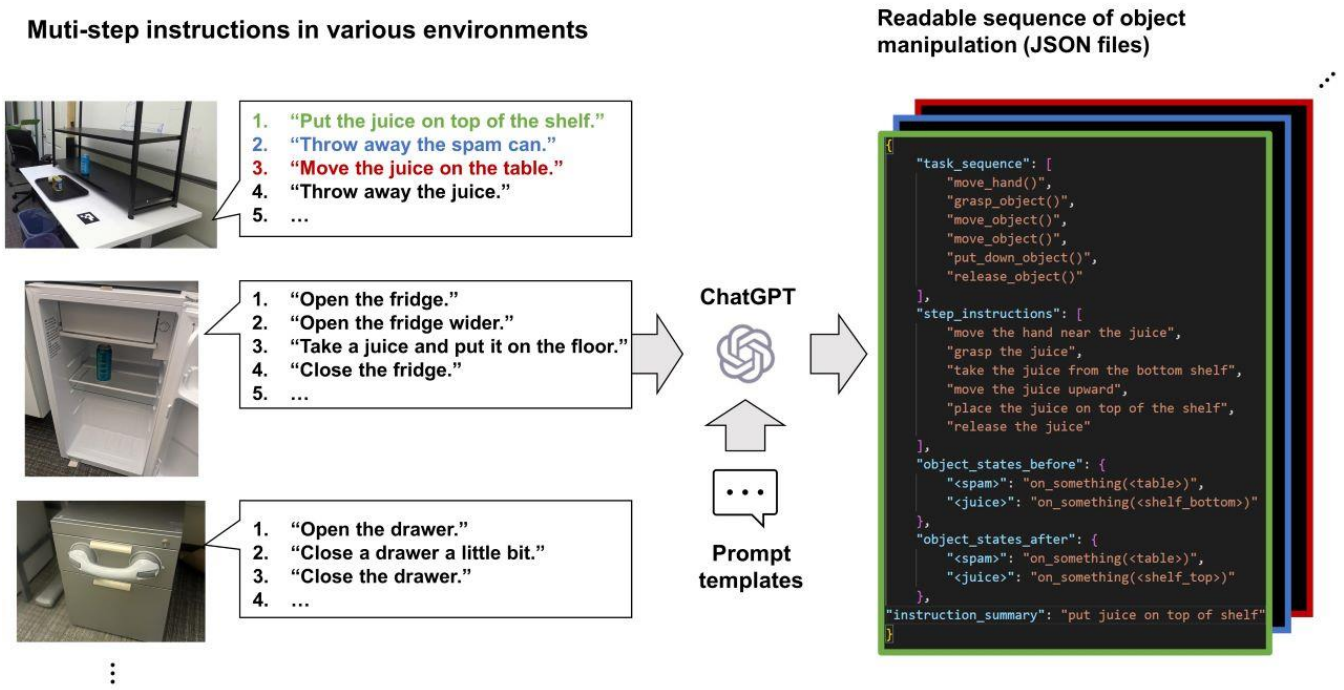
- Line bot (https://developers.line.biz/zh-hant/)



- 其他還有: discord bot, Telegram Bot…等

補充資料: LLM

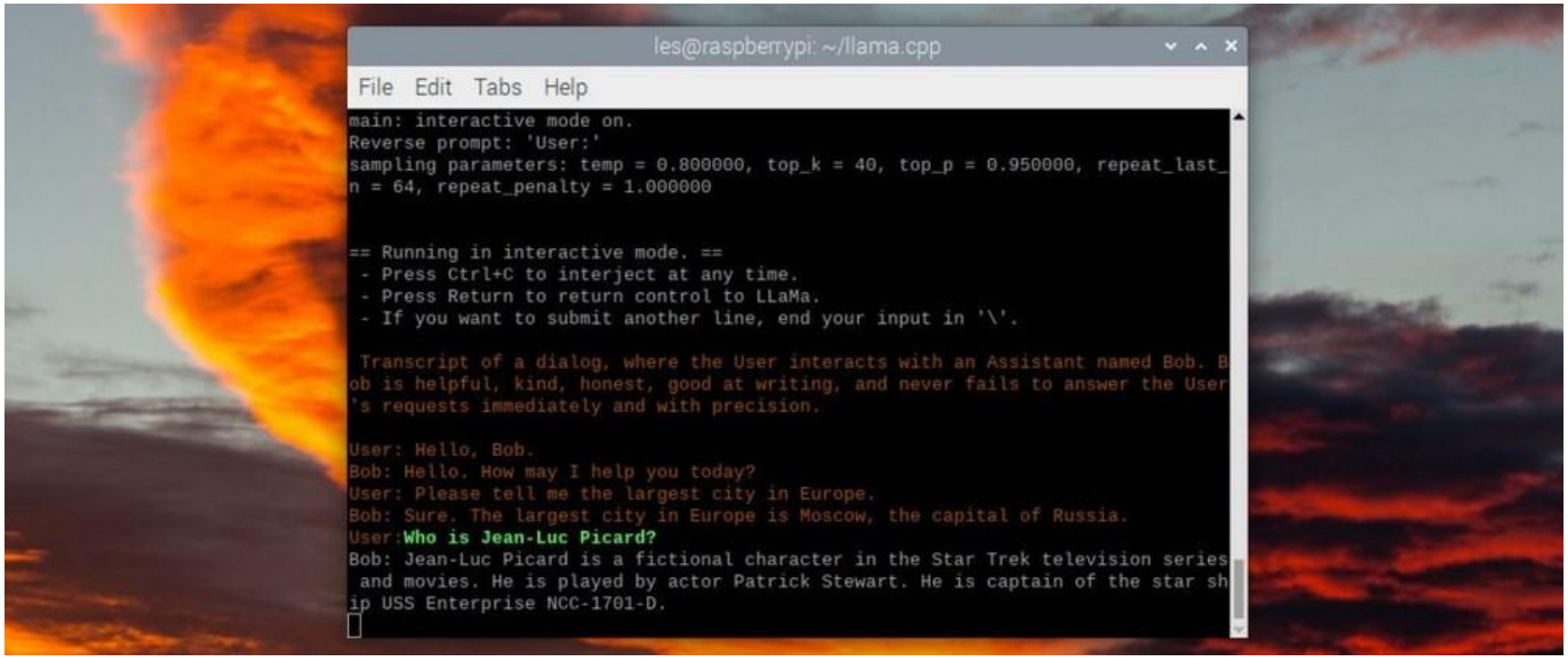
- LLM: large language model (大語言模型)
  - Ex: ChatGPT, Copilot, Gemini…等
- ChatGPT-Robot-Manipulation-Prompts



https://www.dfrobot.com/blog-13412.html

補充資料:
Deploy and run LLM on Raspberry Pi 4B

- this article runs these LLM models (LLaMA, Alpaca, LLaMA2, ChatGLM) on a Raspberry Pi 4B, as well as how to build your own AI Chatbot Server on these devices.



Summary

- Practice Lab
  1. Audio feature
  2. Speech to text (STT)
  3. Text to speech (TTS)

- Write down the answer for discussion 1-2
  - Discussion1: plot the speech feature using chromagram
  - Discussion2: how to make model to speak other language?

- Quiz1-2:
  - 1. combine code to reduce the warning messages
  - 2: Say a custom command to Raspberry PI, it will start to measure the distance and speak the results