

嵌入式系統總整與實作

曾煜棋、吳昆儒

National Yang Ming Chiao Tung University

嵌入式系統總整與實作

日期	主題
2/17	0. 課程介紹
2/24	梅竹賽!!
3/3	1. 嵌入式開發板 - 樹莓派介紹與設定 (headless)
3/10	2. 連接感測器 (GPIO, I2C) + 2 topic sharing
3/17	3. 處理感測資訊 (valuable data) + 2 topic sharing
3/24	4. 網路攝影機 IP cam
3/31	5. 語音互動
4/7 (線上)	6. 嵌入式 + AI模型: 邊緣裝置影像辨識
4/14	期中考Midterm, Project分組
4/21	專案檢索分享 (分組報告)
4/28	Final Project – Proposal (分組報告)
5/5	7. 嵌入式 + AI模型: 語音模型 (台灣樹莓派)
5/12	8. 網路應用: 推播廣告
5/19	9. 樹莓派核心編譯 (Cross compile, Kernel)
5/26	Final Project checkpoint, Q&A, 補demo
6/2	Final Project demonstration (學期考試周)
6/9, 16	彈性補充周

期中考問 (4/7-4/13)

期末考周 (5/27-6/2)



Last week

- 嵌入式應用:網路攝影機
 - Raspberry Pi Camera intro.
 - Python + OpenCV
 - 1. Calculate FPS
 - 2. 建立網路串流
 - 3. 影像辨識 (opencv)
 - A. 人臉識別
 - B. 人臉輪廓識別



This week

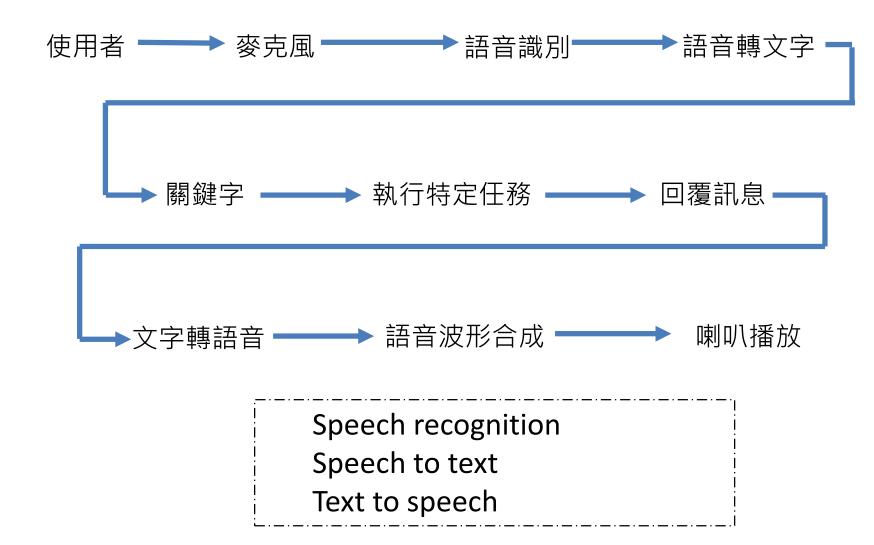
- 嵌入式應用: 語音識別 (語音助理)
 - 1. Mel-Frequency Cepstral Coefficients
 - 2. Speech to text (STT)
 - Text to speech (TTS)



- 語音識別 (Speech recognition)
 - 自動語音辨識 (Automatic Speech Recognition, ASR)
 - 電腦語音識別 (Computer Speech Recognition)
 - 語音轉文字識別 (Speech To Text, STT)
 - 自然語言處理 (Natural Language Processing, NLP)
 - 讓電腦擁有理解人類語言的能力



語音助理流程





Check your device

• aplay -l

```
(COM8) [80x24]
                                                                               ×
                                                                         連線(C) 編輯(E) 檢視(V) 視窗(W) 選項(O) 說明(H)
oi@raspberrypi:~$ aplay -1
*** List of PLAYBACK Hardware Devices ****
card 0: ALSA [bcm2835 ALSA], device 0: bcm2835 ALSA [bcm2835 ALSA]
 Subdevices: 7/7
 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
card 0: ALSA [bcm2835 ALSA], device 1: bcm2835 ALSA [bcm2835 IEC958/HDMI]
 Subdevices: 1/1
 Subdevice #0: subdevice #0
card 1: Device [USB Audio Device], device 0: USB Audio [USB Audio]
 Subdevices: 1/1
 Subdevice #0: subdevice #0
pi@raspberrypi:~$
```

arecord -l

```
● (COM8) [80x24] - □ × 連線(C) 編輯(E) 檢視(V) 視窗(W) 選項(O) 說明(H)

pi@raspberrypi:~$ arecord -1

**** List of CAPTURE Hardware Devices ****

card 1: Device [USB Audio Device], device 0: USB Audio [USB Audio]

Subdevices: 1/1

Subdevice #0: subdevice #0

pi@raspberrypi:~$
```

Test and play microphone

- Check device
 - aplay -l
 - arecord -l
- Record your voice
 - arecord -f cd Filename.mp3 # use "ctrl + c" to stop recording
 - arecord -f cd -d 2 Filename.mp3 # record 2 seconds

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

- Play audio
 - omxplayer -o local -p Filename.mp3





Outline

- 嵌入式應用: 語音識別 (語音助理)
 - 1. Mel-Frequency Cepstral Coefficients
 - Speech to text (STT)
 - Text to speech (TTS)
 - 語音識別 (Speech recognition)
 - 自動語音辨識 (Automatic Speech Recognition, ASR)
 - 電腦語音識別 (Computer Speech Recognition)
 - 語音轉文字識別 (Speech To Text, STT)
 - 自然語言處理 (Natural Language Processing, NLP)
 - 讓電腦擁有理解人類語言的能力

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



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。



10

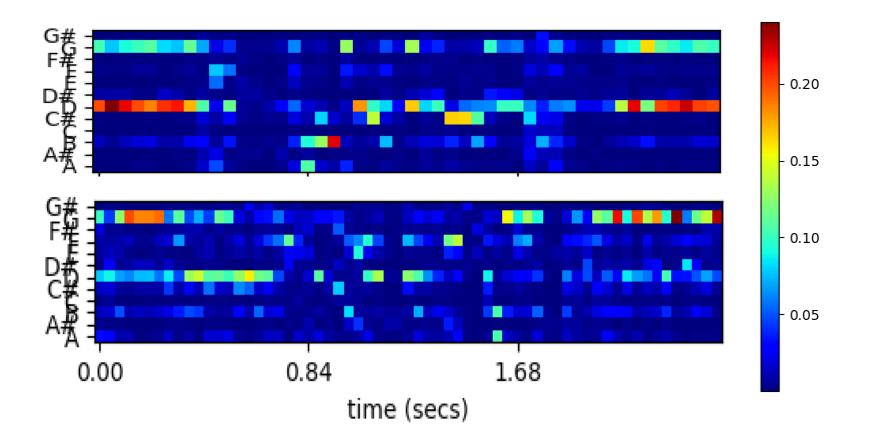


pyAudioAnalysis

- # Install
- git clone https://github.com/tyiannak/pyAudioAnalysis.git
- cd pyAudioAnalysis
- pip3 install -r ./requirements.txt
- pip3 install -e .
- # usage
- python3 pyAudioAnalysis/audioAnalysis.py fileChromagram -i your_file

Chromagram visualization

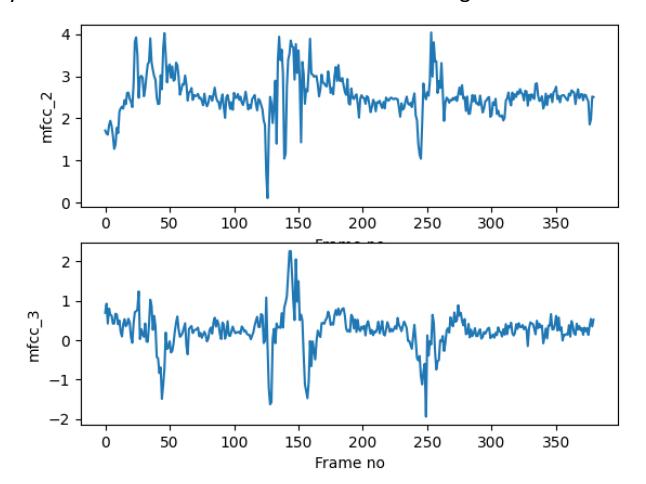
The chromagram is a chroma-time representation, similar to the spectrogram. The following command computes the chromagram of a signal stored in a WAV file:





MFCC

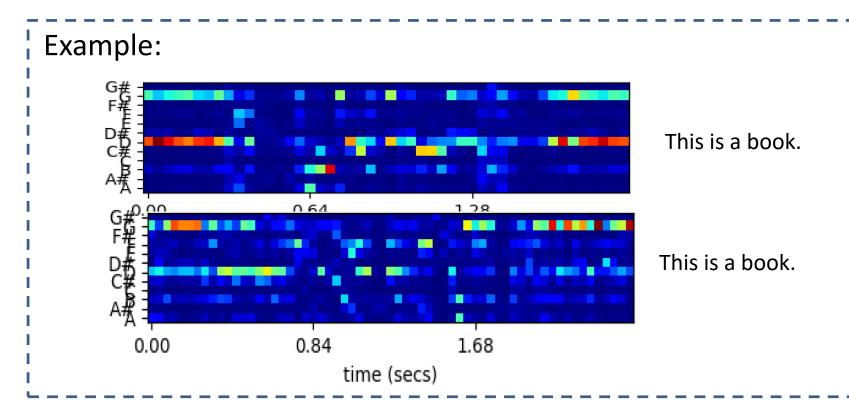
Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.





Discussion 1

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





Outline

- 嵌入式應用: 語音識別 (語音助理)
 - 1. Mel-Frequency Cepstral Coefficients
 - 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)
 - 電腦語音識別 (Computer Speech Recognition)
 - 語音轉文字識別 (Speech To Text, STT)
 - 自然語言處理 (Natural Language Processing, NLP)
 - 讓電腦擁有理解人類語言的能力



1:06





這是什麼歌

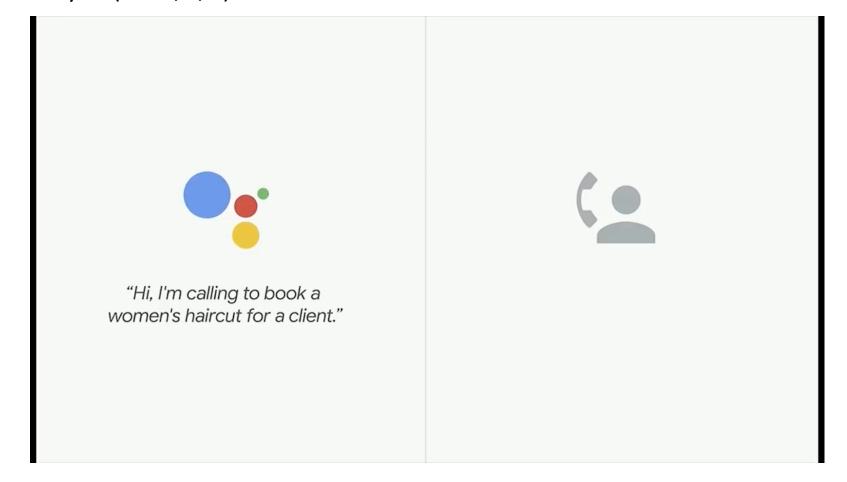


這是 Kate Ryan 的《Voyage voyage》

下午1:06

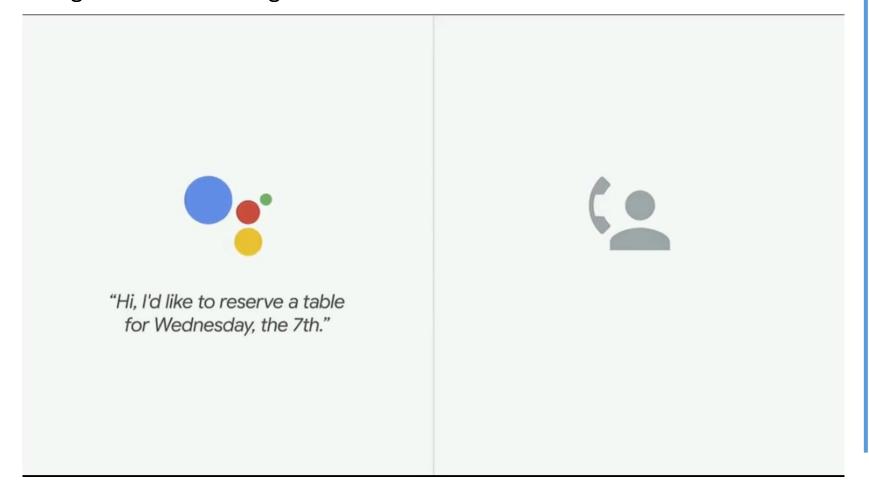


 Google Assistant will soon be able to call restaurants and make a reservation for you (2018/5/9)





Google Assistant calling a restaurant for a reservation



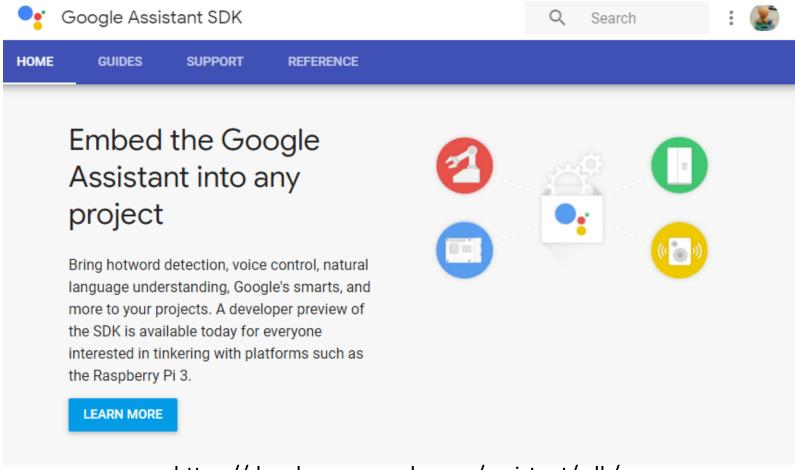


Google IO 2019 Next Gen Google Assistant (2019/5/7)





Google assistant SDK

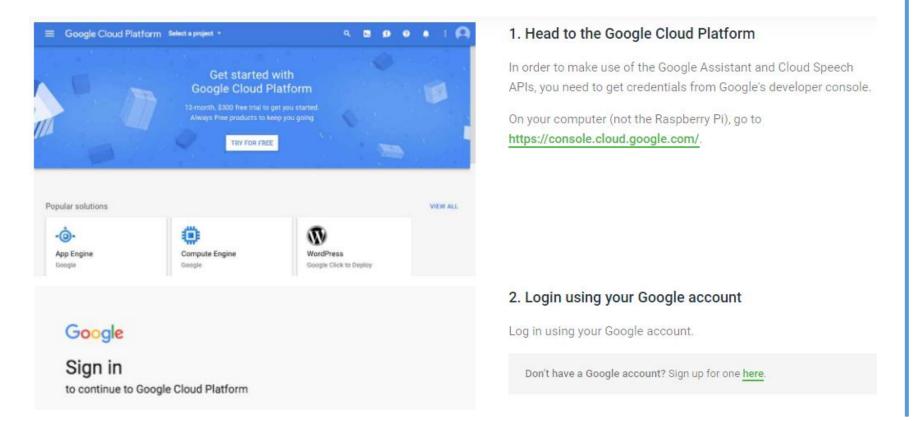


https://developers.google.com/assistant/sdk/





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



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



Azure

語音轉換文字 - 將語音轉換成文字以取得直覺式互動

輕鬆將即時語音轉換文字的功能新增到您的應用程式之中,以應用在語音命令、即時轉譯、自動會議記錄,或是話務中心 的記錄分析等等。





深入了解`

文字轉換語音 - 為您的應用程式提供自然語音

建置智慧型應用程式和服務,使用文字轉換語音服務自然地與使用者交談。近乎即時地將文字轉換成音訊,並根據說話速度、音調、音量等變化進行調整。

使用自訂語音模型,為您的應用程式提供獨特且可辨識的品牌語音。只要錄製並上傳定型資料,服務就會建立專為您的錄音調整的獨特語音效果。



深入了解

語音翻譯

為您的應用程式提供任何支援語言的即時語音翻譯功能,並接收文字或語音翻譯。語音翻譯模型是以尖端語音辨識和神經機器翻譯系統 (NMT) 技術為基礎。這些模型已經過最佳化,能夠理解人們在真實生活中的說話方式,並產生絕佳品質的翻譯。



深入了解



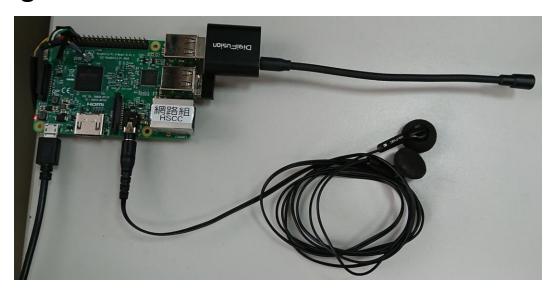
SpeechRecognition

- Library for performing speech recognition, with support for several engines and APIs, online and offline.
- Speech recognition engine/API support:
 - CMU Sphinx (works offline) (卡内基大學)
 - Google Speech Recognition
 - Google Cloud Speech API
 - Wit.ai (Facebook, Messenger ChatBot)
 - Microsoft Bing Voice Recognition
 - Houndify API (SoundHound,音樂識別平台)
 - IBM Speech to Text
 - Snowboy Hotword Detection (works offline)



Install Dependency

- sudo pip3 install SpeechRecognition
- sudo pip3 install gTTS
- sudo apt-get install libasound2-dev
- sudo apt-get install python3-pyaudio
- sudo apt-get install flac





Outline

- 嵌入式應用: 語音識別 (語音助理)
 - 1. Mel-Frequency Cepstral Coefficients
 - Speech to text (STT)
 - Text to speech (TTS)
 - 語音識別 (Speech recognition)
 - 自動語音辨識 (Automatic Speech Recognition, ASR)
 - 電腦語音識別 (Computer Speech Recognition)
 - 語音轉文字識別 (Speech To Text, STT)
 - 自然語言處理 (Natural Language Processing, NLP)
 - 讓電腦擁有理解人類語言的能力

Usage: sudo python3 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))
```

Usage: sudo python3 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.flac') # not mp3, not mp3, not mp3!!
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))
```

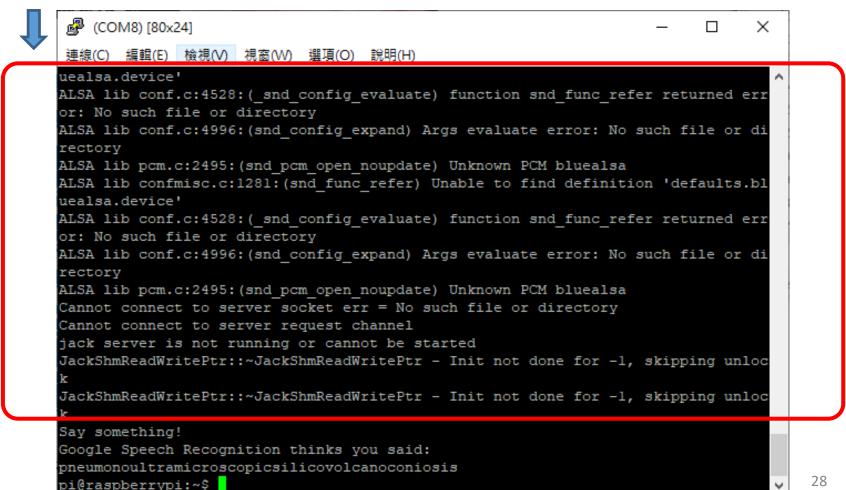
Input format: PCM WAV, AIFF/AIFF-C, or Native FLAC

You might need: ffmpeg -i input.mp3 output.flac



Speech to text (result)

You can ignore the ALSA warning messages





SpeechRecognition

Speech recognition engine/API support:

- CMU Sphinx (works offline)
- Google Speech Recognition
- Google Cloud Speech API
- Wit.ai
- Microsoft Bing Voice Recognition
- Houndify API
- IBM Speech to Text

- 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



Outline

- 嵌入式應用: 語音識別 (語音助理)
 - 1. Mel-Frequency Cepstral Coefficients
 - Speech to text (STT)
 - Text to speech (TTS)
 - 語音識別 (Speech recognition)
 - 自動語音辨識 (Automatic Speech Recognition, ASR)
 - 電腦語音識別 (Computer Speech Recognition)
 - 語音轉文字識別 (Speech To Text, STT)
 - 自然語言處理 (Natural Language Processing, NLP)
 - 讓電腦擁有理解人類語言的能力

Usage: sudo python3 3.tts.py



Text to speech

```
from gtts import gTTS import os

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

os.system('omxplayer -o local -p 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.

gTTS (Google Text-to-Speech

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

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.



Discussion 2

- gTTS: An interface to Google Translator's Text-to-Speech API.
- how to make gTTS speak other language?

```
gTTS (gtts.gtts)
```



Discussion 3

- Say a custom command to Raspberry PI, it will start to measure the temperature. (provided by GY801)
 - Input could be microphone or audio file
 - gTTS can be used to generate an audio file

```
Say something!

Google Speech Recognition thinks you said:

pneumonoultramicroscopicsilicovolcanoconiosis

pi@raspberrypi:~$
```



```
pi@raspberrypi:~/gy801$ python 4baro.py
Baro:
    Temp: 25.200000 C (77.360000 F)
    Press: 1007.310000 (hPa)
    Altitude: 49.237740 m s.l.m
```



Discussion 3 (cont.)

- Input
 - Microphone: talk to microphone directly
 - 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.
 - gTTS: generate the audio file from text
 - The default output is mp3. Use the following command to convert
 - ffmpeg -i input.mp3 output.flac



Discussion 3 (cont.)

- After measuring temperature, use gTTS (Google Text-to-Speech) to speak out the result.
 - Ex: the temperate is 26 degree

```
COM6 - PuTTY

pi@raspberrypi:~/gy801$ python 4baro.py

Baro:
   Temp: 25.200000 C (77.360000 F)
   Press: 1007.310000 (hPa)
   Altitude: 49.237740 m s.l.m
```

IMU (BMP085) can provide temperature!

- Record a video clip to show the whole procedure
 - Say a custom command, and then listen the result



Summary

- Practice Lab (Speech feature, STT and TTS)
- Write down the answer for discussion 1-3
 - Discussion1: plot the speech feature by chromagram
 - Discussion2: how to make gTTS speak other language?
 - Discussion3: Say a custom command to Raspberry PI, it will start to measure the temperature
 - 書面問答, 請上傳到e3
 - Discussion3請用錄製影片的方式來呈現結果
- Next week (4/7): 嵌入式 + AI模型: 邊緣裝置影像辨識
 - 線上課程