

數位語音處理概論 期末 project 報告

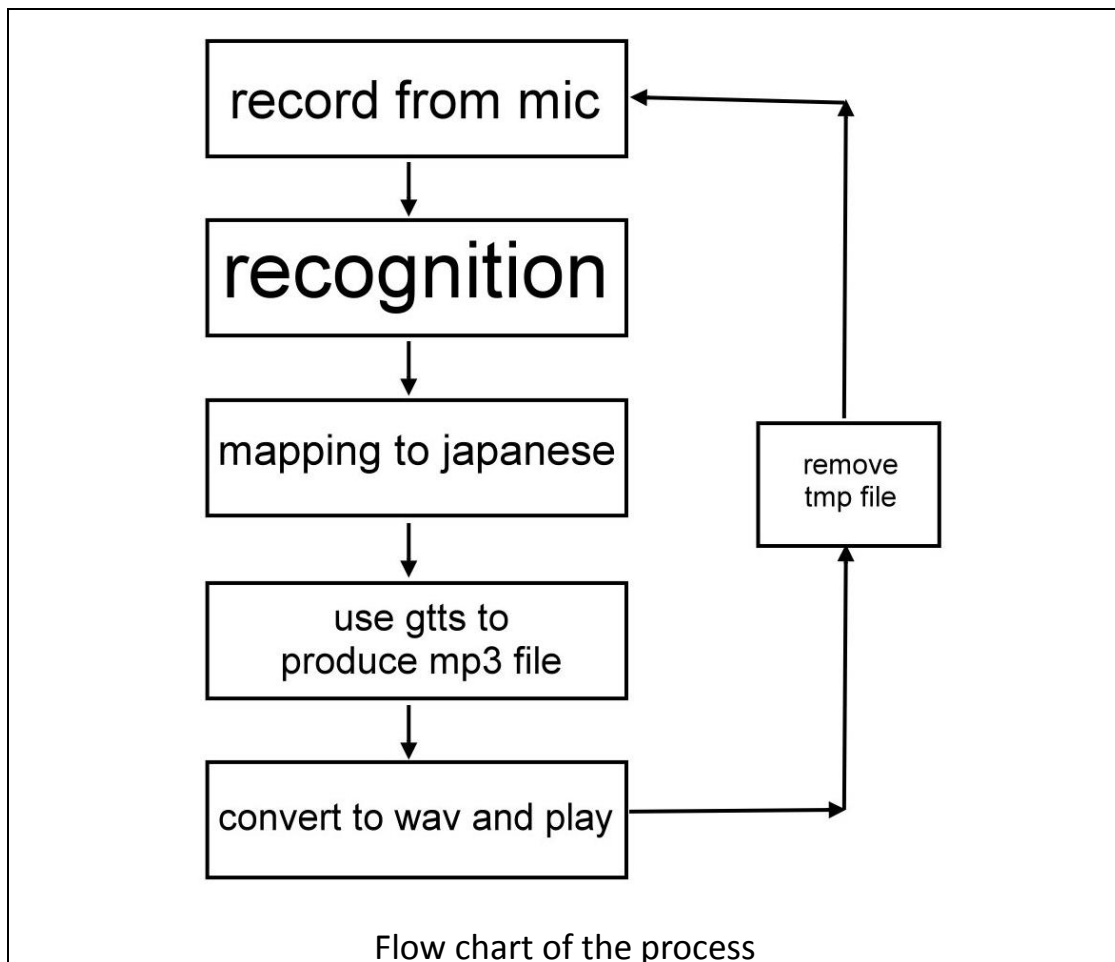
B04902028 資工三 洪浩翔 B04902116 資工三 蘇景耀

1. project 理念及內容:

我們的想法是透過一個程式，此程式會錄使用者所說的英文句子作為輸入，並轉換為日語發音的英文句子念出。主要就是做英語至日語發音英文的 transliteration。

2. 實作方法:

我們的實作方法大致可分為 3 部分，第一部分是輸入，主要是依賴 python 中的 speech_recognition 套件對筆電的 mic 做連接並直接讀取語音輸入，錄音過程會持續 5 秒，完成後並對其進行辨識轉為 text。轉為 text 後，會將其透過 mapping function “w2k” 將英文的 text 轉成日文的 phoneme，主要表示為片假名。最後一部分為輸出，將片假名送至 google text-to-speech 套件(gTTS)進行輸出為 mp3，並透過 ffmpeg 轉為 wav 檔案，輸出檔案會透過 Linux command line “play music.wav” 的方式進行撥放。待播放完成，會進行下一輪的錄音。



3. 需求套件:

測試環境:Ubuntu 16.04.3 LTS in virtual box

Python	
speech_recognition	record and recognition
subprocess	to call command line
gtts	convert text to speech

Ubuntu	
sox	play music in command line
libsox-fmt-mp3	Allow to play mp3. This tool is optional
ffmpeg	Convert mp3 to wav

4. 執行方法:

先確定網路連線，再以 `python 2` 執行程式，並等待麥克風設置。當程式出現 `start record` 後開始錄音，最多只能錄 5 秒。錄製完畢後，等待程式進行轉換，程式會自動以 `google` 小姐的方式播出轉換後的語音。等待程式移除暫存資料後並等待重置，即可開始下一輪錄音。

5. 成果展示:

<pre>Please wait. Remind that you have 5 sec to record your query. start record. Please say something. done record Recognition :is it good to drink</pre>	Start record and recognize.
<pre>Recognition :イズ イト グード ツー ドリンク</pre>	Mapping to Japanese.
<pre>ffmpeg version 2.8.11-0ubuntu0.16.04.1 Copyright (c) 2000-2017 the FFmpeg developers built with gcc 5.4.0 (Ubuntu 5.4.0-6ubuntu1~16.04.4) 20160609</pre>	Call ffmpeg to convert sound file.
<pre>Input #0, mp3, from 'speech.mp3': Duration: 00:00:02.64, start: 0.000000, bitrate: 32 kb/s Stream #0:0: Audio: mp3, 24000 Hz, mono, s16p, 32 kb/s Output #0, wav, to 'tmp.wav':</pre>	Convert mp3 file to wav file.
<pre>tmp.wav: File Size: 127k Bit Rate: 384k Encoding: Signed PCM Channels: 1 @ 16-bit Samplerate: 24000Hz Replaygain: off Duration: 00:00:02.64 In:100% 00:00:02.64 [00:00:00.00] Out:63.4k [] Clip:0 Done. start new record</pre>	Play sound and restart a new record.

6. 例外狀況:

在進行播放時，有時會發現播放的 hardware device 會莫名被 block，使得音訊無法順利播放，而解決方法為使用 htop 開啟所有程式的列表，將

```
5499 et 9 -11 407M 10860 7996 S 1.4 0.2 0:01.55 /usr/bin/pulseaudio --start --log-target=syslog
```

此行程式終止，若還是無法，則將其他在 pulseaudio 的程式也終止，並重新執行原本的程式進行錄音，應就可順利運行。

7. 分工:

B04902028 洪浩翔	B04902116 蘇景耀
輸入、輸出套件處理	英文至羅馬拼音 mapping
報告編寫	羅馬拼音至日文片假名 mapping
圖片處理	日文 phoneme 對應英文 phoneme 研究

8. Reference:

speech_recognition	Codesofinterest, python-speech-recognition-pocketsphinx www.codesofinterest.com/2017/03/python-speech-recognition-pocketsphinx.html
	python Package Index, SpeechRecognition https://pypi.python.org/pypi/SpeechRecognition/2.1.3
W2k transform	Japanesetransliteration, English to Japanese transliteration japanesetransliteration.com/
Play sound.wav	Askubuntu, Play sound in command line https://askubuntu.com/questions/920539/how-to-play-a-sound-from-terminal
Google text to speech transform	Pythonprogramminglanguage, text-to-speech transform https://pythonprogramminglanguage.com/text-to-speech/
	python Package Index, gTTS https://pypi.python.org/pypi/gTTS