

Online VC improvement

Student : Sian-Yi Chen

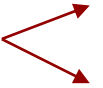
Advisor : Tay-Jyi Lin and Chingwei Yeh

● Online VC improvement

- 改善目標：“數值轉換問題”，讓數值轉換時，不要帶單位 (已完成)
- 改善方法：
 - ◆ Google 提供解決方案：**speech adaptation boost** Improve transcription accuracy
在轉換程式中可以通過給予**提示參數**，使特定單詞或是專有名詞轉換準確率提高，
如果轉換強度不夠，還可以給予**增強係數**來提高準確性
 1. 提示參數：adaptation
 - weather 與 whether 皆讀作 **/'weð.ə/**，可選擇轉換詞彙
 2. 強度係數：boost
 - 在 **"/'weð.ə/ is good** “ 這句話中，轉換成 weather 信心度為 0.6，而 whether 轉換信心度為 0.3
 - 通過提示參數也不一定能轉換成 whether，因此需要再加上 boost 參數，係數範圍從 0.0 ~ 20.0

● Offline VC 規劃

- 何謂離線版本
- Online 與 Offline 的相互關係
- 實作步驟

/'weð.ə/  weather
whether

Offline VC 規劃

- 何謂離線版本

1. 在電梯、地下室等離線環境時的使用模式，使用者可自行切換至離線模式下使用

- Online 與 Offline 的相互關係

1. offline 版本為：YMDVC (尚未定名)
2. 初次使用 offline 版本前，需在有網路的環境下錄製設置的 320 句，等待訓練完成後，才能下載 offline 模型 (如果讓使用者隨意錄製語句作為訓練語料，無法清楚定義字數或句數，無法保證能訓練出完整模型)
3. 在以上基礎下，使用線上模式錄製與轉換後的語料也會加入訓練
4. 而離線模式則是先保存使用者音訊，待連線後，將語料一併上傳至雲端做轉換，再加入訓練

- 實作步驟

使用兩台電腦，一台當作裝置 (A)，一台當作雲端 (B)

1. 使用 B 電腦錄製 320 句當作初次使用模型
2. 將模型下載到 A 電腦中並離線轉換
3. 將線上與離線模式錄製的語音上傳到 B 電腦，再次訓練模型

附錄－API 改善過程

speech_adaptation_beta.py 官方版本：

[https://github.com/googleapis/python-](https://github.com/googleapis/python-speech/blob/e1eae214c98e777516457d619bd14bd13c244127/samples/snippets/speech_adaptation_beta.py)

[speech/blob/e1eae214c98e777516457d619bd14bd13c244127/samples/snippets/speech_adaptation_beta.py](https://github.com/googleapis/python-speech/blob/e1eae214c98e777516457d619bd14bd13c244127/samples/snippets/speech_adaptation_beta.py)

```
parser = argparse.ArgumentParser()
parser.add_argument(
    "--storage_uri",
    type=str,
    # default="gs://cloud-samples-data/speech/brooklyn_bridge.mp3",
    # default="gs://chullin_bucket/offline_converter/weather.wav",
    default="gs://chullin_bucket/offline_converter/1234.wav",
    # default="gs://chullin_bucket/offline_converter/2021430.wav",
    # default="gs://chullin_bucket/offline_converter/負三.wav",
)
parser = argparse.ArgumentParser()
parser.add_argument(
    "--storage_uri",
    type=str,
    # default="gs://cloud-samples-data/speech/brooklyn_bridge.mp3",
    # default="gs://chullin_bucket/offline_converter/weather.wav",
    default="gs://chullin_bucket/offline_converter/1234.wav",
    # default="gs://chullin_bucket/offline_converter/2021430.wav",
    # default="gs://chullin_bucket/offline_converter/負三.wav",
)
```

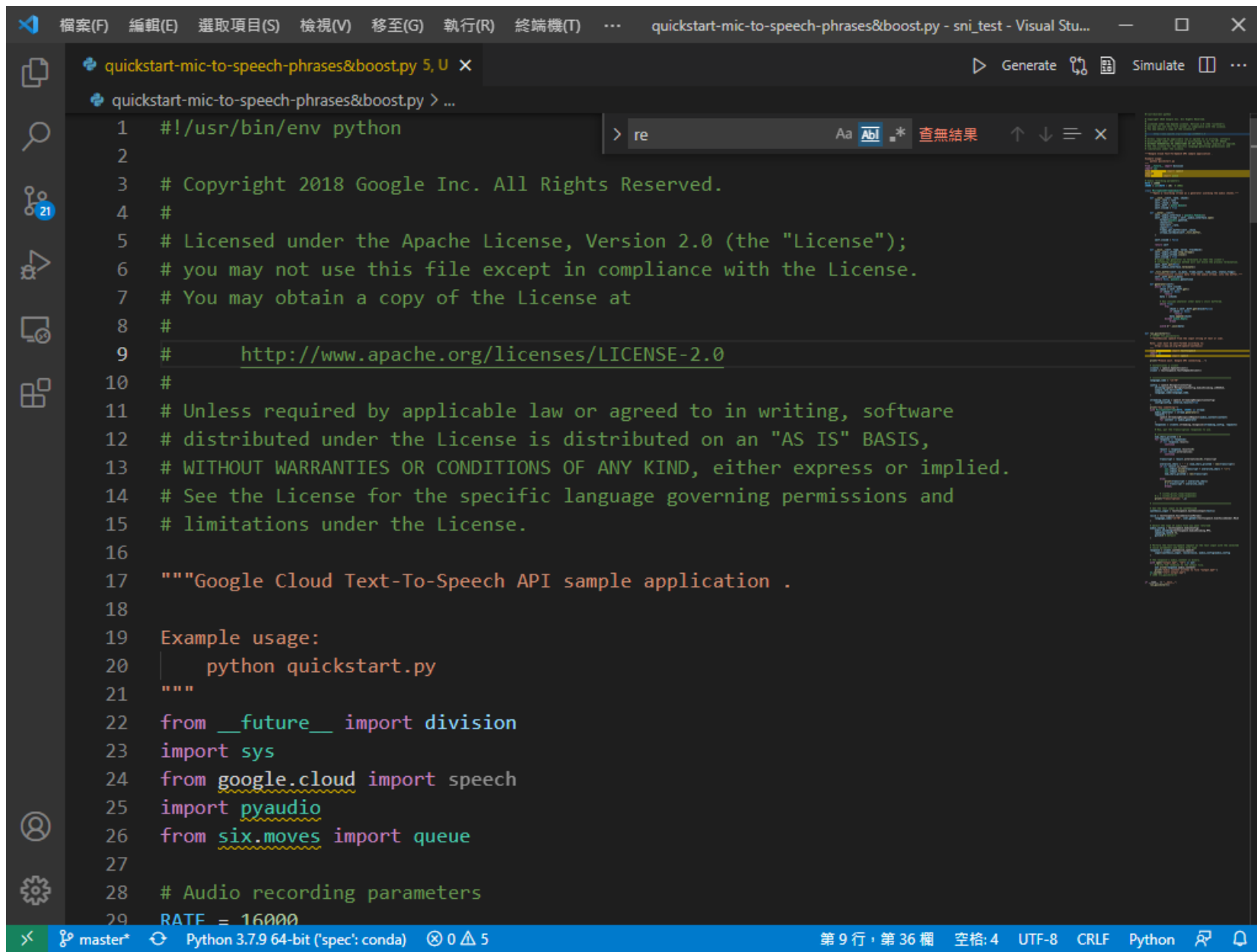
Google 有兩種表示方式

1. 直接用 JSON 表示
2. 一般函數寫法

串聯的麥克風版本不能用 JSON 格式，先讓 boost 版本先改成一般函數寫法

speech to text 中的 `transcribe_context_classes` 是一般版本
但 `sample_adaptation_beta` 卻是 beta 版
必須改使用

```
from google.cloud import speech_v1p1beta1 as speech
```



```
quickstart-mic-to-speech-phrases&boost.py 5, U x
quickstart-mic-to-speech-phrases&boost.py > ...

1 #!/usr/bin/env python
2
3  # Copyright 2018 Google Inc. All Rights Reserved.
4  #
5  # Licensed under the Apache License, Version 2.0 (the "License");
6  # you may not use this file except in compliance with the License.
7  # You may obtain a copy of the License at
8  #
9  # http://www.apache.org/licenses/LICENSE-2.0
10 #
11 # Unless required by applicable law or agreed to in writing, software
12 # distributed under the License is distributed on an "AS IS" BASIS,
13 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 # See the License for the specific language governing permissions and
15 # limitations under the License.
16
17 """Google Cloud Text-To-Speech API sample application .
18
19 Example usage:
20 | python quickstart.py
21 """
22 from __future__ import division
23 import sys
24 from google.cloud import speech
25 import pyaudio
26 from six.moves import queue
27
28 # Audio recording parameters
29 RATE = 16000
```

<https://cloud.google.com/speech-to-text/docs/how-to?hl=zh-TW>

自由決定要轉換成 whether 或是 weather :

<https://cloud.google.com/speech-to-text/docs/speech-adaptation>

<https://cloud.google.com/speech-to-text/docs/boost>

<https://cloud.google.com/speech-to-text/docs/class-tokens>

<https://cloud.google.com/speech-to-text>

https://console.cloud.google.com/storage/browser/_details/chullin_bucket/offline_converter/%E8%B2%A0%E4%B8%89.wav?authuser=1&hl=zh-TW&organizationId=319425611965&project=dev-acolyte-307405