

# Online VC 改善 & personalized offline VC 規劃

---

Student : Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

# Outline

個人化離線語音裝置-線上版本

## ● Online VC 改善成效

- 已成功將使數字轉換不會帶有單位，但整體效果變較差，

原本是只有純數字會變成帶單位轉換，現在是只要是數字就一律不帶單位轉換，錯誤的例子為：

( e.g. 唸 2021年5月18日，原本讀作：二零二一年五月十八日；現在讀作：二零二一年五月一八日)

( e.g. 唸 一萬兩千三百四十五，原本讀作：一萬兩千三百四十五；現在讀作：一二三四五)

- 問題 & 目前改善方法：

轉換過程中包含 1.轉文字 2.轉語音

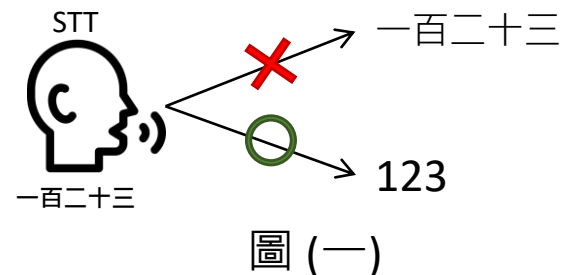
1. 轉文字：Google 提供 speech adaptation 方法解決此類似問題，可以改善對指定單詞的識別，但在中文並沒有支援數字轉換不帶單位的模型，Google 傾向將數字轉換成阿拉伯數字表示，而非中文，如圖 (一)
2. 轉語音：支援著一種 SSML (Speech Synthesis Markup Language) 的特殊語法，可以透過將文字加上標籤，使一段文字有不同的語音合成方式，也是目前的改善方式

( e.g. `<say-as interpret-as="characters">can</say-as>`，原本讀作：/kæn/ or /kən/；現在讀作：C A N)  
`</speak>`

## ● 下週規劃

- 離線版本實作

- 環境建置 (使用 Linux 環境架設 YMDVC 執行環境)



# ■ 技術細節

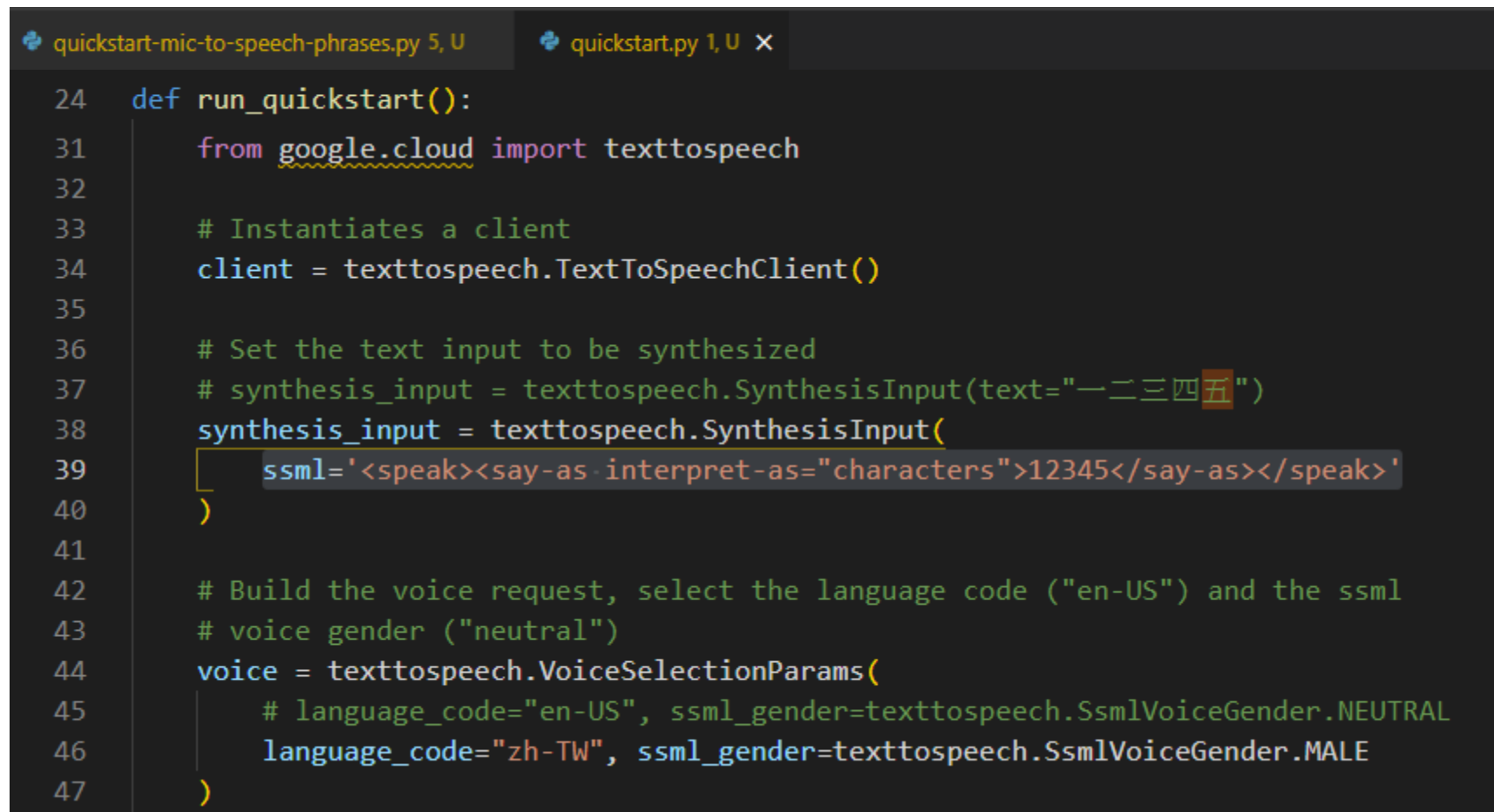
- 現在將“數字唸成數值”問題，整理成為字串處理問題，著重在 TTS (text to speech) 的部分，以下為實現步驟：
  1. 收到一段文字
  2. 判斷字串是否包含數字
  3. 使用正規表達式將文字與數字分開
  4. 將數字加上 SSML 標籤
  5. 將文字與更改後有分類標籤的數字和為一字串
- 目前方法的衍生問題：
  - 只要字串中包含數字，不管是否有帶單位，轉換後一律不帶單位
- 衍生問題的解決方法：
  - 再多加判斷式，判斷字串中是否包含單位，或是年份，將純數字與有單位量詞的句子分開 (e.g. 判斷字串中是否包含“個”、“張”、“元”、“歲”、“年”、“月”、“萬”...等等)
- 但無法解決根本問題：與說話者轉換一模一樣的聲音
  - (e.g. 讀：一七與十七，皆會轉換成相同文字“17”，無法解決)

# 附錄－API 改善過程

目前發現 STT 對於數字都是傾向使用數值表示，而非中文  
像是 12345 會用 '12345' 表示，而非 '一二三四五'  
所以嘗試一輪後決定往 TTS 找方法  
如果使用 SSML 語法，就可以很容易解決目前的問題

語音合成標記語言（**SSML**）：

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



```
24 def run_quickstart():
31     from google.cloud import texttospeech
32
33     # Instantiates a client
34     client = texttospeech.TextToSpeechClient()
35
36     # Set the text input to be synthesized
37     # synthesis_input = texttospeech.SynthesisInput(text="一二三四五")
38     synthesis_input = texttospeech.SynthesisInput(
39         ssml='<speak><say-as interpret-as="characters">12345</say-as></speak>'
40     )
41
42     # Build the voice request, select the language code ("en-US") and the ssml
43     # voice gender ("neutral")
44     voice = texttospeech.VoiceSelectionParams(
45         # language_code="en-US", ssml_gender=texttospeech.SsmlVoiceGender.NEUTRAL
46         language_code="zh-TW", ssml_gender=texttospeech.SsmlVoiceGender.MALE
47     )
```

D:\Work\_Space\off\_line\_speech\_converter\Quickstarts\python-texttospeech\samples\sni\_test\quickstart.py

所以目前可以讓 '12345' 念成 '一二三四五'  
接著就是在 麥克風轉文字再轉換成語音 的版本中  
可能用個判斷，如果是數字，就將它包上 SSML 語法

文字前加上 <say-as interpret-as="characters">  
文字後加上 </say-as></say>

另外一個感覺是問題的就是 text 標籤要更改成 ssml

python 怎麼判斷輸入的值是否為數字？

<https://www.html.cn/qa/other/19601.html>

如何在 Python 中檢查字串是否包含數字：

<https://www.delftstack.com/zh-tw/howto/python/how-to-check-a-string-contains-a-number-or-not-in-python/#map-%E5%87%BD%E5%BC%8F>

在包含數字的字串中，

1. 首先使用 正規表達式 將數字與文字分開  
假設 ' 我是12345' => '我是' , '12345'
2. 加標籤
3. 最後在用 str.join 合在一起

```
1  import re
2
3  str1 = "python1"
4  str2 = "nonumber"
5  str3 = "12345"
6
7  print(any(chr.isdigit() for chr in str1))
8  print(any(chr.isdigit() for chr in str2))
9  print(any(chr.isdigit() for chr in str3))
10
11
12
13  s='3a4b5cdd7e'
14  result = re.findall(r'[0-9]+|[a-z]+',s)
15
16  print(result)
17  print(type(result))
18
19  str = ' '
20  print(str.join(result))
```

```
<say>
    我是
    <say-as interpret-as="characters">
        12345
    </say-as>
</say>
```

## Python 正規表達式：

<https://yanwei-liu.medium.com/python%E6%AD%A3%E8%A6%8F%E8%A1%A8%E9%81%94%E5%BC%8F-fbffb92972dc>

## [Finding the index of an item in a list](https://stackoverflow.com/questions/176918/finding-the-index-of-an-item-in-a-list)：

<https://stackoverflow.com/questions/176918/finding-the-index-of-an-item-in-a-list>

## Day11 Python 基礎 - 列表的使用：

<https://ithelp.ithome.com.tw/articles/10185694>

```
24 str4 = "我是12345"
25 # 如果字串中有包含數字就會進 if 迴圈
26 if any(chr.isdigit() for chr in str4) == True :
27     print("會進來嗎 \n")
28     result2 = re.findall(r'[0-9]+|[a-z]+|[\u4e00-\u9fa5]+', str4) # 利用正規表達式將字串拆成 1
29     print(result2) # ['我是', '12345']
30     print("陣列第0個位置 = " + result2[0]) # 我是
31     print(type(result2[0])) # <class 'str'>
32
33     for i in result2:
34         if any(chr.isdigit() for chr in i) == True :
35             new = print('<say-as interpret-as="characters">' + i + '</say-as>' + "\n")
36             new = '<say-as interpret-as="characters">' + i + '</say-as>'
37             # print(new) # 印出 <say-as interpret-as="cha
38             # print(result2.index('12345')) # .index 可以印出位置
39             # result2.insert(result2.index('12345'), new) # 對第幾個位置插入
40             result2[result2.index(i)] = new
41             print(result2)
42
43 else:
44     print("錯瞞")
45
46 result2.insert(0, "<speak>")
47 result2.append("</speak>")
48 print(result2)
```

```
命令提示字元 - "C:\Users\User\Anaconda3\condabin\conda.bat" activate ...
['3', 'a', '4', 'b', '5', 'cdd', '7896', 'e']
<class 'list'>

會進來嗎

['我是', '12345']
陣列第0個位置 = 我是
<class 'str'>
<say-as interpret-as="characters">12345</say-as>

['我是', '<say-as interpret-as="characters">12345</say-as>']
['<speak>', '我是', '<say-as interpret-as="characters">12345</say-as>', '</speak>']

<speak> 我是 <say-as interpret-as="characters">12345</say-as> </speak>

3 a 4 b 5 cdd 7896 e

(offline_converter) D:\Work_Space\off_line_speech_converter\Quickstarts\
python-texttospeech\samples\sni_test>
```

D:\Work\_Space\off\_line\_speech\_converter\Quickstarts\python-texttospeech\samples\sni\_test\judg\_num.py

```
str1 = "nonumber"
str3 = "我是12345"
如果沒有數字，就要回傳 text
    text = str1
如果有數字，就要回傳 ssml
    ssml =
    <say-as interpret-as="characters">
        + str3 +
    </say-as>
```

```
quickstart-mic-to-speech-phrases.py 5, U
quickstart-judg_num.py 1, U
have_number.py U X
i_test > have_number.py > ...
> x Aa Abi .* No results ↑ ↓ ≡ ×

1 def have_number(str):
2     import re
3
4     if any(chr.isdigit() for chr in str) == False :
5         text = str
6         return text
7     if any(chr.isdigit() for chr in str) == True :
8         result2 = re.findall(r'[0-9]+|[a-z]+|[\u4e00-\u9fa5]+', str)
9
10        for i in result2:
11            if any(chr.isdigit() for chr in i) == True :
12                # new = print('<say-as interpret-as="characters">' + i + '</say-as>' + "\n")
13                new = '<say-as interpret-as="characters">' + i + '</say-as>'
14                result2[result2.index(i)] = new
15                # print(result2)
16
17        result2.insert(0, "<speak>")
18        result2.append("</speak>")
19        # print(result2)
20        # print("\n")
21        str = ' '
22        str.join(result2)
23        # print(result2)
24
25        ssml = result2
26        return ssml
27
28
29 if __name__ == "__main__":
30
31     str2 = "nonumber"
32     str4 = "我是12345"
33     str = ' '
34     ssml = str.join(have_number(str4))
35     print(ssml)
36     print(type(ssml))
37
```