Title:

以雲端語音合成技術為基礎的音文同步有聲書之建立系統

Author:

張朝凱 ,長庚大學資訊工程。

Abstract:

本系統主要是藉助於 Google 的語音合成的 API，來幫助我們將文字的內容轉換成語音，再運用語音結合文字，搭配上文字背景的變化來表達出同步化的效果。擷取出語音中每個斷點發生的位置，與每個字在語音訊號中的起始位置及結束位置，來推算語音的發音開始時間及發音結束時間來做判斷，再取得螢幕畫面中每個文字的位置運用背景高亮的方式將文字顯示出來。

Description:

近年來語音合成技術逐漸發展成熟被應用於許多的平台上如 Apple 的 SIRI，Google 的 TRANSLATE 甚至利用在許多語言的學習上，其主要目的是針對使用者所需的文字內容提供所需要的語音內容。

雖著智慧型手機及平板電腦的普及許人們的閱讀習慣也逐漸的遭到改變，有鑑於傳統的紙本書太重，太占空間，製造過程砍伐樹木會造成全球暖化，攜帶亙是不便，於是漸漸有了把科技和書本結合再一起的構想，因而催生了電子書的產生。其優勢具有攜帶方便，減少污染等，人們閱讀的平台由一般的書籍逐漸轉移到智慧型裝置上。為了克服運用行動行智慧裝置閱讀電子書時的不便，例如:字體太小所造成的閱讀不便時。運用語音合成的技術將文字的內容合成出語音並將語音及文字做結合來輔助使用者來閱讀電子書的內容。

本系統主要是藉助於 Google 的語音合成的 API 來幫助我們將文字的內容轉換成語音，再運用語音結合文字及配合上文字背景的變化來表達出同步化的效果。其程式執行流程為:

Step1. 準備一個文字檔(txt)。
Step2. 將文字檔的內容做文字分析。
Step3. 經由文字分析後將文章內的每個字給分割出來。
Step4. 將每個字上傳至 Google 的 API 合成語音。
Step5. 結合語音及文字及配合上文字背景的變化來表達出同步化的效果。
以上步驟，圖示如下圖所示(Figure1)。

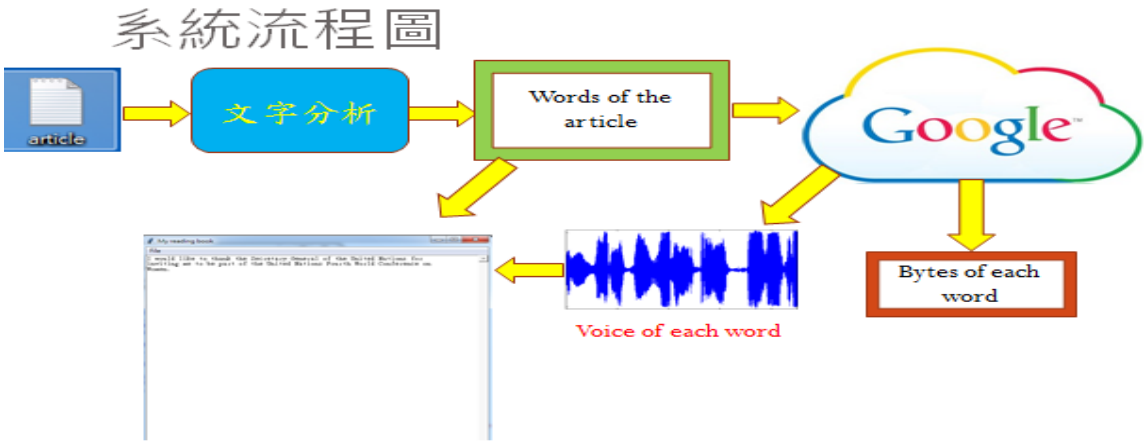


Figure 1.系統流程圖

為了達到語音成同步的效果我們必須先去求出每個字在語音中的發音位置，在此我們運用了 pygame.mixer.music.get_pos( )來擷取出語音中每個斷點發生的位置。並運用從 google 取得每個字的 byte 數，來估測每個字在語音訊號中的起始位置及結束位置，來推算語音的發音開始時間及發音結束時間，其估測的公式為:

$$\text{Bytes of Pos}_i \text{ occur} = \text{Pos}_i/\text{Length}*\text{sumT}$$

Pos:運用 pygame.mixer.music.get_pos( )所取得語音中的斷點(單位:ms)。
Length:運用 mutagen 所取出語音的總長(單位:s)。
sumT:將所有文字的 byte 做相加取得語音檔案的大小。($\sum_1^n word\ of\ byte$)

經由上述的公式，我們可以輕易地求出文章中每個字和字之間停頓發生的時間點，及每個字的語音訊號在整段語音中的第幾個 byte 到第幾的 byte。運用剛才的結果來做判斷當公式，算出的結果大於語音訊號起始位置或小於語音訊號結束位置時則將文字做特殊的背景處理。

有了每個字在聲音訊號中的起始位置及結束位置後，接下來我們必須去取得螢幕畫面中每個文字的位置。
其搜尋文字位置的流程為:

Step1. 設定開始搜尋的位置為 1.0 其中 1.0 為文字在畫面上的 X 座標，1.0 為文字在畫面上的 Y 座標。
Step2. 在螢幕畫面中插入我們所要閱讀的文字內容。
Step3. 在每次插入文字時計算每個文字是由多少字元所構成例如:word(包含 4 個字元)、get(包含 3 個字元)等。
Step4. 插入文字後從畫面位址 1.0 開始搜尋文字在畫面中的開始位置。
Step5. 將搜尋到畫面中的文字起始位址和文字字元做加總來產生文字在畫面中的結束位置。
以上步驟，圖示如下圖所示(Figure2)。

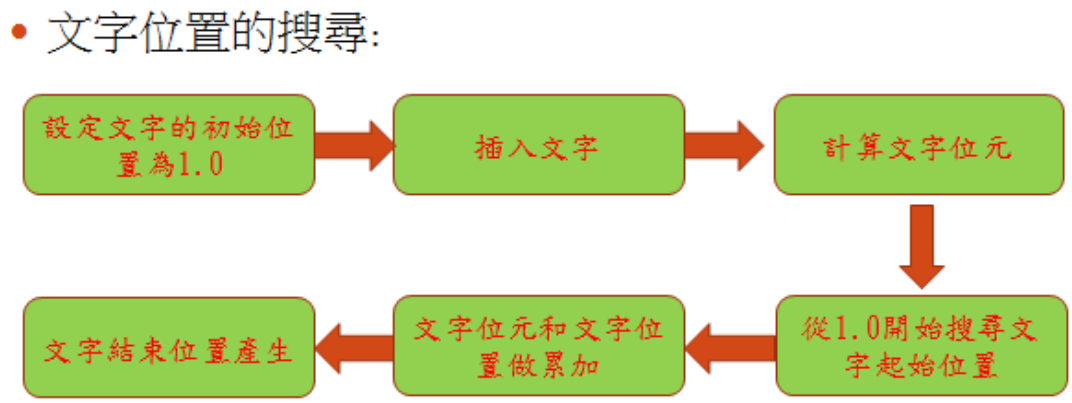


Figure 2.文字位置的搜尋

有了文字在畫面中的起始位置和結束位置後運用背景高亮的方式將文字顯示出來。圖示如下圖所示(Figure3)

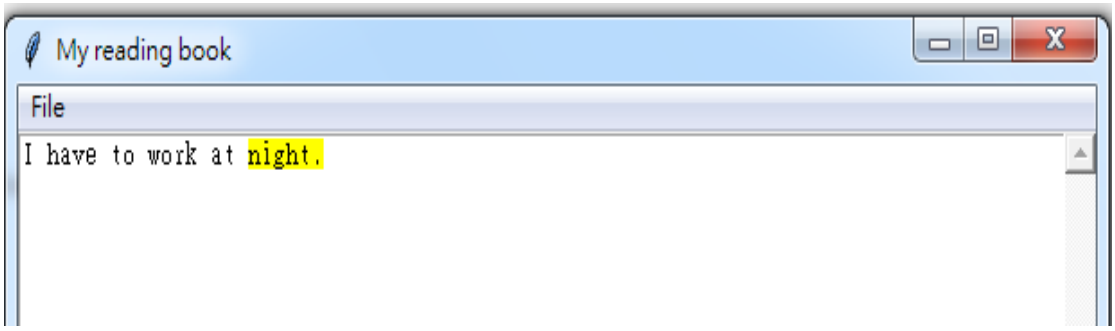


Figure 3.文字背景高亮展示

語音和文字同步化的過程為:

Step1. 插入文字於畫面中。

Step2. 搜尋畫面中每個文字在畫面中的起始位置及結束位置。

Step3. 播放文字的語音內容。

Step4. 擷取語音內的斷點。

Step5. 以斷點做判斷。

Step6. 當斷點的位置大於語音訊號起始位置或小於語音訊號結束位置則顯示出文字高亮的背景，當斷點的位置小於語音訊號起始位置或大於語音訊號結束位置則繼續保持原始的文字狀態。

以上步驟，圖示如下圖所示(Figure4)。

- 語音同步化流程圖:

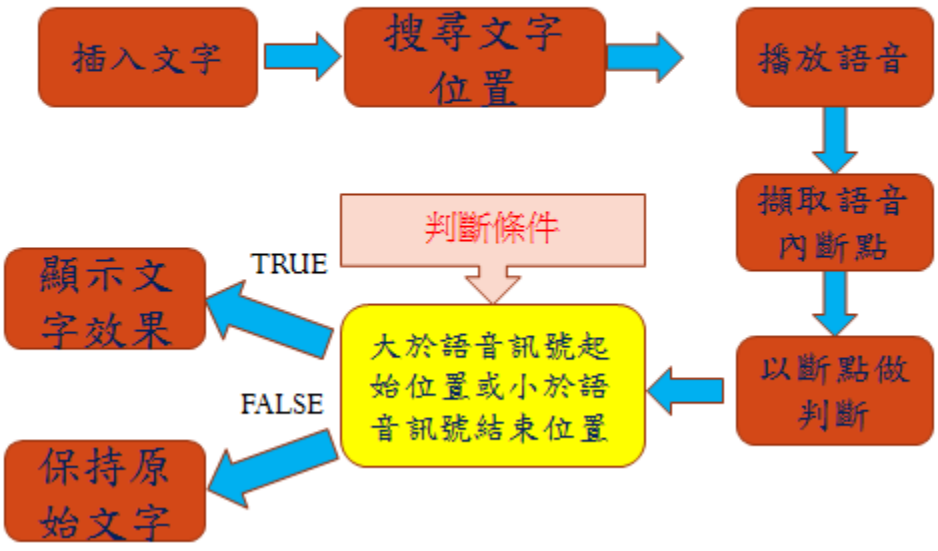
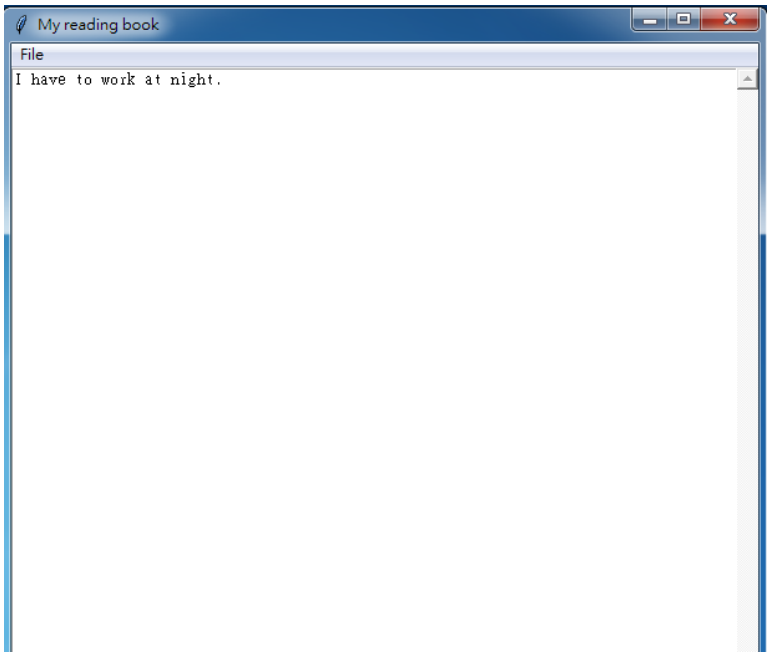
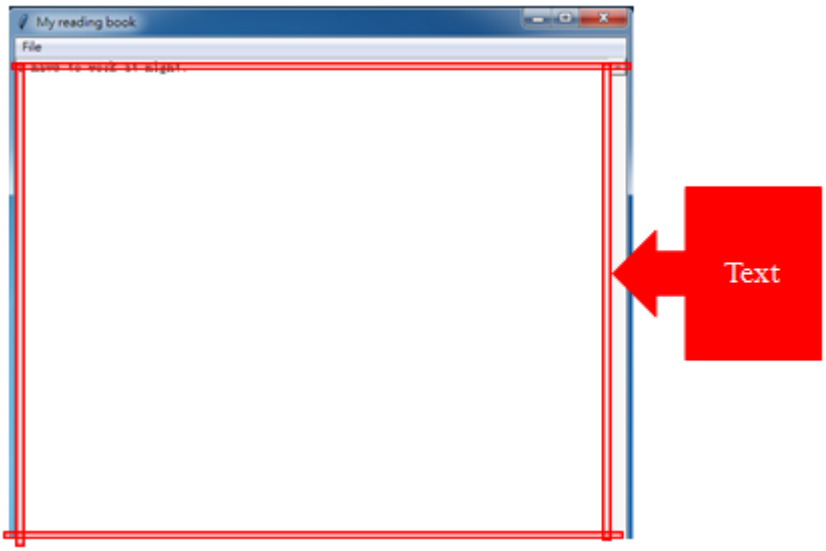


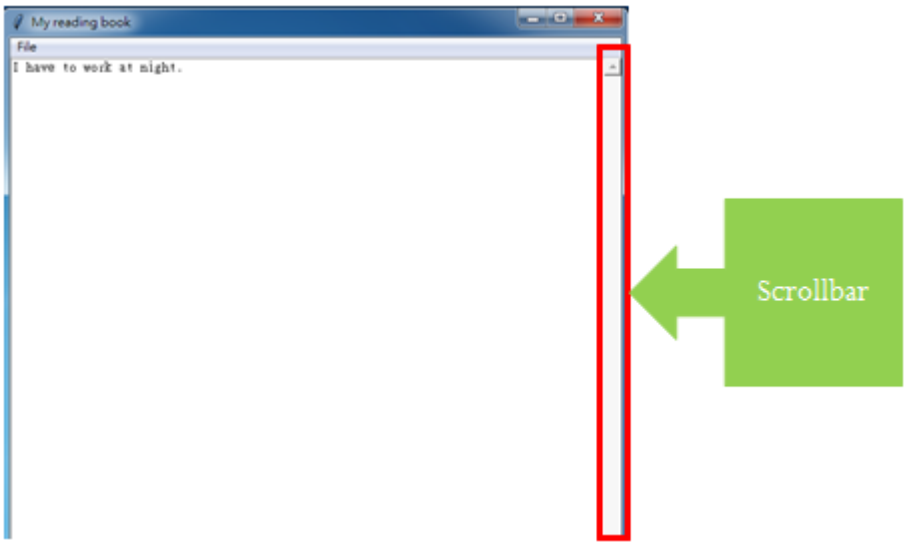
Figure 4.語音同步化流程

在使用者的介面上主要是運用 python 的 Tkinter 作為版面設計的工具。使用的物件有 Text、Scrollbar 及 menubar。



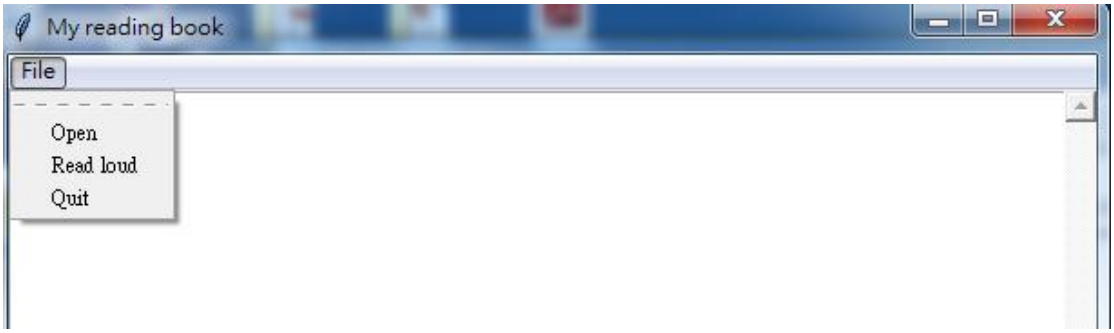


Text:主要是用來顯示插入的文字內容。



Scrollbar:當插入的文字內容過多時可用 Scrollbar 做上下的移動。

Menubar:為了簡化畫面因此運用 Menubar 將所有的功能加入 Menubar 來供一般的使用者來做選擇。



Reference:

---

http://www.tutorialspoint.com/python/python_gui_programming.htm
http://docs.python-requests.org/en/latest/
http://pygame.org/news.html
https://docs.python.org/3/library/tokenize.html
https://github.com/leo-labs/gTTS
http://tkinter.unpythonic.net/wiki/tkFileDialog
https://bitbucket.org/lazka/mutagen