走音救星：把歌詞轉成唱歌的軟體

組員：陳羿豐

1. 動機與目的

我很喜歡唱歌，也很喜歡聽同學唱歌，然而同學總是說他們不會唱，於是我就想著，能不能錄下同學們說話的聲音，然後用剪接軟體把說話轉成唱歌。我認為這是做得到的，因為我曾經用Audacity和Praat把Google翻譯的語音轉成歌聲，只是受限於我能取得的工具，當時花了非常久才完成。於是我決定開發只要唸出歌詞，就可以合成出歌聲的App。

這個專題原本的名稱為「開發把朗讀變唱歌的軟體」，但是我覺得名字不夠吸引人，所以就換名稱。也許同學是怕走音才不敢唱歌吧。

1. 現有相關研究比較

市面上已經有多種合成歌聲的軟體：

1. Praat：是語音分析軟體，可用於分析音高、音量及共振峰等語音參數。[1] Praat還提供操縱語音音高的功能，雖然合成歌聲不是其設計目的，但可用來合成歌聲。
2. VOCALOID：專業的歌聲合成軟體，初音未來就是用這個軟體創造的。然而音色只能由廠商提供，使用者不能創造音色檔。
3. UTAU：以剪接音檔來合成歌聲的軟體，可自行錄製音色。

以上的軟體中，只有Praat有開放原始碼，並且提供研究論文，因此本專題參考Praat合成語音的方法來實作。

1. 原理

Praat合成歌聲的方式是TD-PSOLA (Time Domain Pitch Synchronous Overlap and Add)，這個方法會把現有的錄音檔轉為歌聲。以下列出其步驟：

* 1. 找出錄音檔(或音色檔)在各個時間的周期
  2. 把聲音按照周期切成多個區塊
  3. 改變這些區塊的間距，並將區塊中互相重疊的部分相加，即可以改變聲音檔的周期。由於音高是周期的倒數，因此可以改變音高。
  4. 第3步改變區塊的間距，會導致聲音的長度改變，為了使音長不變，可以內插區塊，或者刪除多餘的區塊。

圖1是TD-PSOLA的運作方式。每個區塊用不同的顏色標記，在改變區塊間距時，相同的顏色表示重複的區塊。



找出週期

把聲音切成區塊

改變區塊間距

圖1、PSOLA的示意圖

每個區塊實際上包含兩個周期的聲波，因為如果區塊只有一個周期，則在區塊邊界上會有斷點產生，合成結果會有雜音。在取出區塊時，為了消除斷點，還要把區塊乘上漢明窗，把邊緣處柔化。

把聲音按照周期切割的原因是，可以假設人聲是由聲帶振動產生的脈衝，經由共鳴器官得到的結果，脈衝的頻率是音高，而共鳴器官的脈衝響應就是區塊。

為了做出TD-PSOLA，必須要能夠偵測聲音的周期，然而偵測聲音的周期(或音高)是很難的問題。Praat的論文提到，可以用自相關函數來取得周期和音高。[3]自相關函數就是自身與延遲後訊號的相似度，由公式(1)計算出。計算聲音的自相關函數後，可尋找其最大值，通常這就是周期。

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

由公式(1)計算的自相關函數，在週期很大的時候會失真，因為累加的項數減少了。Praat的論文提出的方法是，把自相關函數除以窗函數的自相關函數，就可以修正。[3]

即使做了以上的修正，有時候還是會受到雜訊的影響，使得最大值的位置不是周期，這時可以計算每個局部最大值的位置作為音高的機率，然後用維特比演算法取得最佳音高序列。不是所有的聲音都有音高，比如說錄音的開頭和結尾、語氣停頓的地方、還有子音，因此使用維特比演算法還要把「無音高」狀態加進來考慮。當音量很小，或者自相關函數很小時，就有可能是子音，在這些情況下，要把「無音高」的機率提高。[3]

在合成歌聲時，我假設錄音檔裡只有子音和母音，母音有音高，可以用TD-PSOLA處理，但是子音沒有音高，無法以音高切割聲音。我的處理方法是直接把子音部分疊加到合成波形上，因為子音不會改變長度。萬一子音部分太長，超過音符長度，才會嘗試縮短子音，這時會把聲音切成長度0.008~0.012秒的區塊，然後刪除部分的區塊，直到長度滿足要求。

1. 系統實作

本專題所開發的App是網頁，以JavaScript來撰寫，這樣就不需要發佈到App Store，而且也省下編譯的麻煩。

程式的流程如圖2所示。使用此程式需要先錄下唸歌詞的聲音，然後從錄音檔裡面圈選出每個單字的範圍，最後輸入歌詞的音高和音符。

錄音

record.html

從錄音圈選單字

segment.html

分析音高

pitch.js

輸入樂譜、歌詞

makesong.html

選擇錄音檔

makesong.html

合成波形

makesong.js

錄音資料庫

storage.js

圖2、系統架構圖

程式由主畫面和數個子程式構成。主畫面提供五個功能：

* + Record in browser：錄音程式
  + Change pitch by changing playback speed：變更播放速度，以改變音高
  + Detect pitch：偵測音高程式的展示
  + Segment each word：圈選單字的程式
  + Make song：輸入音符和歌詞後，就可以合成歌聲的程式

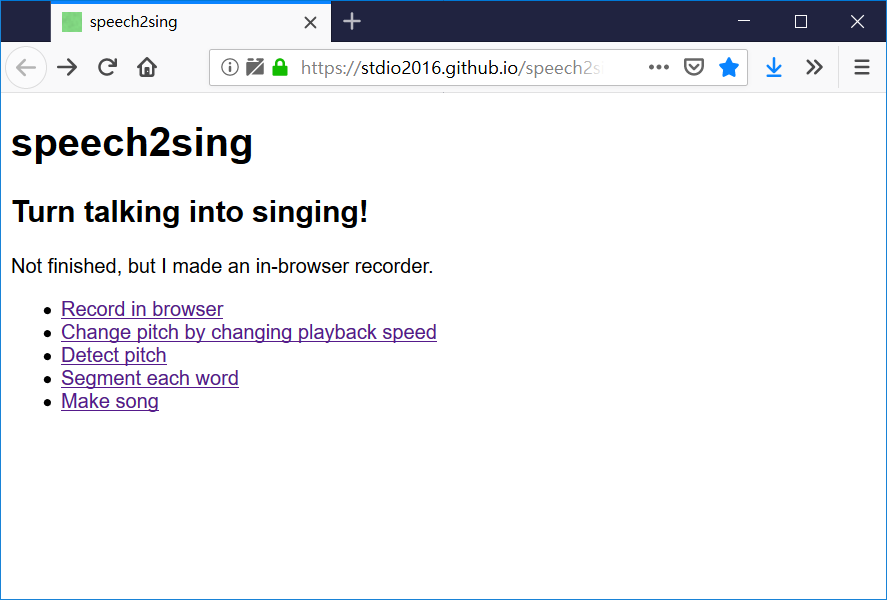


圖3、程式主畫面

以下說明各個子程式的介面

* + Record in browser：錄音程式

網頁名稱為record.html，如圖4，UI的程式為js/record.js。

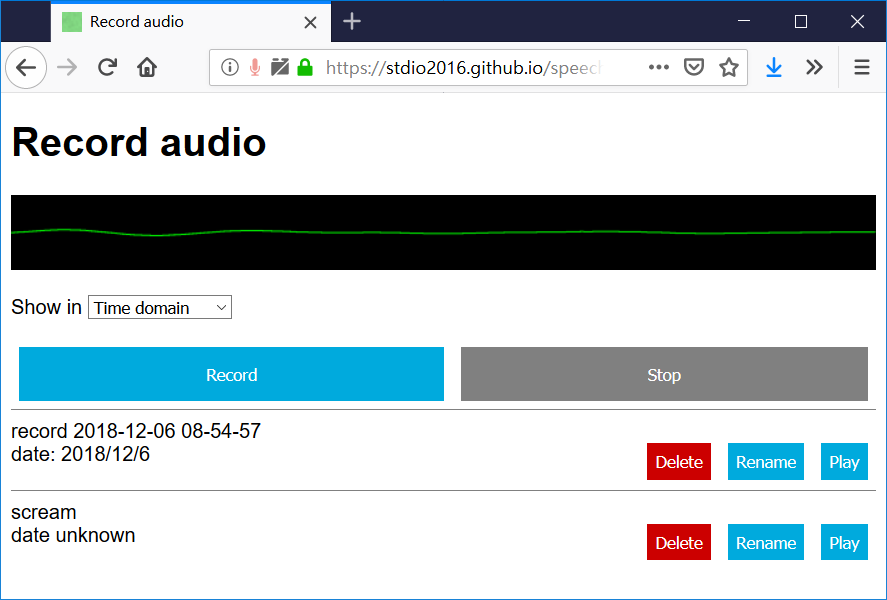


圖4、Record in browser的介面

按下Record鍵可以錄音，再按Stop可以結束錄音，這時系統會詢問錄音檔的名稱，輸入後就可以存檔。如果不指定名稱，則預設以現在時間命名。

已儲存的錄音檔會列出來，可以播放錄音、更改檔名或刪除錄音。

錄音和播放的聲音可以視覺化顯示，提供Time domain (時間域)、Frequency domain (頻率域) 和Autocorrelation (自相關) 輸出，可從Show in下拉式選單選取。Time domain顯示波形，Frequency domain顯示0~5000Hz的頻譜圖，Autocorrelation顯示正規化後的自相關函數。

* + Change pitch by changing playback speed：變更播放速度，以改變音高

網頁名稱為changepitch.html，如圖5，UI的程式內嵌於網頁裡。

這個程式可以用重新取樣的方式來改變音高，使用方法是：調整pitch拉桿，然後在下方的檔案列表中按下Play按鈕，就可以播放。pitch拉桿在正中央時，聲音不變，拉桿越往左，聲音越低，速度也變慢，拉桿越往右，聲音越高，速度也變快。用重新取樣改變音高的特點是，音長也會一起改變，且音長和音高成反比。

這個程式沒有在流程圖上，功能也非常侷限，因為我寫這個程式只是為了測試瀏覽器的API功能，在實作出合成器之前，我可以先知道調整音高後聲音會變怎樣。

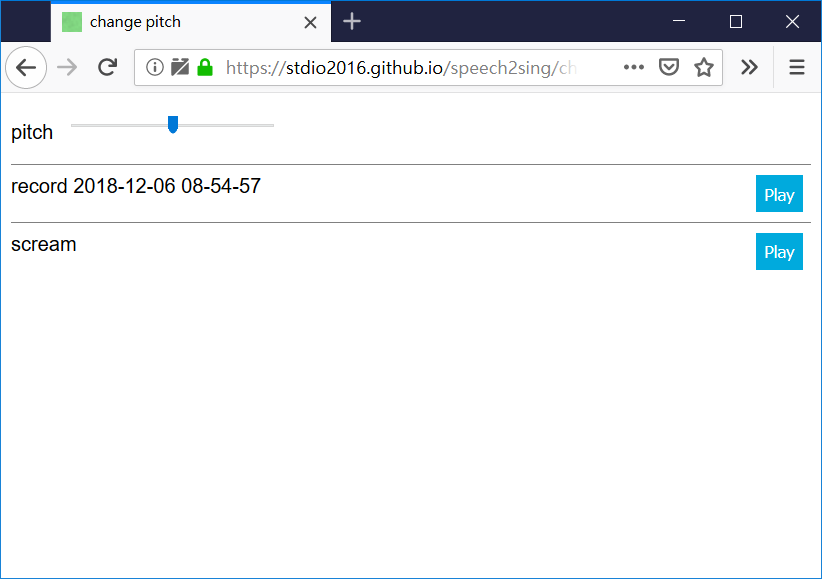


圖5、Change pitch by changing playback speed的介面

* + Detect pitch：偵測音高程式的展示

網頁名稱為pitch.html，UI的程式為js/pitch\_ui.js。

* + Segment each word：圈選單字的程式

網頁名稱為segment.html，UI的程式為js/segment.js。

* + Make song：輸入音符和歌詞後，就可以合成歌聲的程式

網頁名稱為makesong.html，UI的程式為js/makesong.js。

1. 成果

程式已放到網路上，網址為<https://stdio2016.github.io/speech2sing/index.html>

1. 結論

這個專題只有我一個人做，而且同時是資訊工程專題和互動式音訊處理導論的專題。

24235235你好365AA。結論寫不出來，GG！

1. 參考文獻

[1] Paul Boersma & David Weenink (2017): Praat: doing phonetics by computer [Computer program]. Version 6.0.23, retrieved 11 November 2017 from http://www.praat.org/

[2] 吳銘冠、陳嘉平（2013）。基於時域上基週同步疊加法之歌聲合成系統。載於 Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013) (頁76-89)。中華民國計算語言學學會

[3] Paul Boarsma (1993): Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA Proceedings 17: 97-110.