

智慧型自我情緒判定之研究與實作

專題學生：金子恩、羅煒傑、林逸融、劉泯錚

指導教授：吳永基 教授

長榮大學資訊工程學系

2019 年 11 月 28 日

摘要：

語音為人們重要的交流媒介之一，語言信號中包含了許多豐富的情感資訊。本專題開發一個自我評估情緒的電腦系統程式與 APP，研究語言及音調與情緒之關聯，透過語言語料庫的建立以及音調的分析方式，來推算、評估個人的心理狀態，包含了憂鬱、焦慮、悲傷、害怕與生氣。設計一款用語音辨識自我有無負面情緒的系統。透過與專業的心理諮商師合作，將諮商師與患者談話所得到的寶貴經驗，包含說話內容與音調特徵，來當作本專題評估使用者當時的心理狀態之準則。使用者可以使用本系統了解自我情緒，能長期觀察自身情緒之變化，做到自我判斷，當發現有狀況時，也可以尋求協助，防患未然。

關鍵詞：情緒、語音、情感辨識、語料庫。

1. 前言

1.1 研究動機與目的

隨著社會經濟高度成長，科技帶給人們高度的文明，卻也造成人們精神方面疾病的大幅增加[1]。根據健保署資料統計，截至 2016 年底，國內經醫師處方服用抗憂鬱藥物的人數為 127 萬 2566 人，相較五年前 98 萬 3490 人增加 23%，人數明顯逐年增加[2]，使得憂鬱症已成為常見且嚴重的文明疾病，在全球已有超過三億名患者。但在外顯行為沒有明顯變化下，人們較難察覺自己是否有憂鬱症狀，而是當這些症狀開始嚴重影響生活時，人們才肯開始面對並接受治療。但憂鬱症的治療週期很長，從諮詢認識、諮商晤談、服藥治療、心理治療等非常辛苦。如果能夠及早得知、偵測出徵兆並且加以預防或是提早就醫，可以帶來很大的效益[3]。目前國內憂鬱症的主要檢測還是為引導式問卷、各式憂鬱量表、或到身心診所進行諮商，但不論是諮商行為或是問卷的蒐集、資料處理前／後，皆需要耗時大量的時間與人力成本問題。

而利用語料庫做語言研究和計算是未來的趨勢。但隨著研究議題的多元，使用程式來處理語料，並建立語料庫是很重要的一件事。根據前述背景與動機，本專題研究與實作智慧型自我情緒判定之研究與實作。除了減少在處理問卷資料時所耗費的大量時間，也能夠幫助使用者更了解當下自身的情緒事件程度，若發現有長時間處於悲傷情緒的跡象，則需要多加留意自身狀況進而預防憂鬱症的發生；擷取與心理師諮商過程和固定式問題裡未被發現的情緒和被忽略的語調特徵。

本專題目前透過與心理諮商師合作，具有豐富經驗的心理諮商師，經過無數次的看診，從患者說話中的語調變化與文字內容所得到的寶貴經驗，當作本系統評估使用者談話時心理狀態的準則。透過經驗、講話內容、聲音的特徵來分析

情緒與語音的關聯性，並將兩者間的相關性以數據化方式呈現情緒狀況給使用者。

1.2 競爭對手與困難點

現有市場上有各式各樣情緒檢測、憂鬱情緒檢測壓力紅綠燈 APP，但這些 APP 主要的檢測都是透過各個面向的問題，如身心健康、個人價值、心理安全、家庭健康來統計使用者的平均精神健康，準確率的高低也和題目數量成正比。但語音中卻已包含了豐富的情感訊息，將信號與情緒的特徵值比較相似程度，能夠彈性的提問題，也能根據回答的前後文意，判斷當下情緒狀態。

聲音轉換成文字後，整段文字並沒有任何邊點符號。為了將各個有意義的單字、句子區隔，將搭配斷詞語料庫做文字的斷詞，負面情緒語料庫做情緒的分類。

在這節中探討的三個困難點：

- (1)語音與情緒的特徵值分析。
- (2)如何正確的將使用者語音的文字檔案斷詞，例如：一句「馬路很難過」，在此場景，難過並不是情緒狀態用詞。
- (3)斷詞語料庫、負面情緒語料庫的建立。

1.3 使用效益

使用者使用本系統後，藉由音調及語詞和經驗的判斷，能夠了解自身的情緒狀況，並且做到自我判斷。比起傳統的紙本問卷，本系統可以更快地了解狀況，並做長期觀察，能降低因情緒狀況不好而造成的遺憾事件，藉由推廣本系統，讓大家能更了解與重視憂鬱症，能多關心身邊的人。

2.文獻探討與使用技術

2.1 與市面上類似產品不同之處：

本系統與市面上的憂鬱量表、問卷／壓力檢測 APP 比較差異（表 1）：

表 1 差異比較表

	智慧型自我情緒判定系統	憂鬱量表、問卷／壓力檢測APP
檢測方式	<ol style="list-style-type: none"> 擷取使用者說話中的語調變化與文字內容。 將語調變化、自身常模數值，透過與諮商師討論後所建置的負面情緒規則分析比較；文字內容則利用負面情緒語料庫去計算句子所代表的權重分數。 最後輸出五種負面情緒數值。 	<ol style="list-style-type: none"> 透過作答憂鬱、壓力情緒篩檢量表來檢視自己的情緒狀態。 受測者依據選項程度之輕重，給予分數，所得分數越高，代表憂鬱、壓力感的程度越高。
主要功能	<ol style="list-style-type: none"> 利用手機錄製使用者情緒語句，藉由音調及字句和經驗的判斷，能夠了解自身情緒狀況，做到自我判斷、提前預防。 能長期觀察自身情緒之變化，當發現有狀況時，也可以尋求協助，防患未然。 利用 GPS 與建立身心診所地標，當使用者自覺不對勁時，能夠隨時查詢最近的關懷單位，以提供協助。 	<ol style="list-style-type: none"> 透過作答憂鬱量表題目、壓力量表題目，統計使用者受憂鬱、壓力之苦的程度。

2.2 App 端

本專題主要開發於行動裝置，其主要功能有

三項。分別為負面情緒語句分析、關懷系統、圖表紀錄如圖 1。

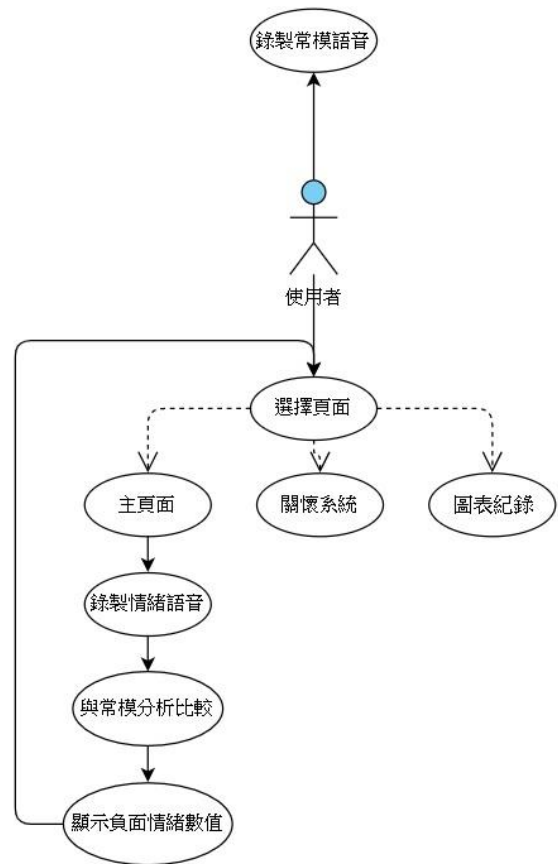


圖 1.App 使用流程圖

初次使用本系統時，為了建立使用者常模語音數值，將錄製使用者平時正常說話時的數值，作為常模數值的依據。

當本系統已擁有使用者的常模語音數值後，則能夠正常使用負面情緒語句分析、關懷系統、圖表紀錄三項功能。

2.3 使用技術

2.3.1 Android Studio

在熱門的行動裝置作業系統 Android 與 iOS 兩者間，我們選擇了 Android 做為系統平台，原因有以下幾點：

- 市場占有率第一的智慧型手機作業系統。
- 開發成本較低（iOS 開發需使用 macOS）。
- 開發完成後易於發布。

考慮了上述幾點原因後，在選擇 App 的系統環境，最終選擇了 Android 做為系統平台。使用的語言為 Java，並透過 Android Studio 作為 App 的開發工具。

2.3.2 情緒分類

人類的情緒非常複雜的，依據情緒的不同，每個人在表達以及陳述各項感受時，也都各有差

異。W.Parrot 舉出了 Philipp Shaver 等人蒐集了 213 組不同的情緒表達詞彙，加以分類成六組集合，並且以樹狀表的方式做為分類法，以三種要素來維繫：Primary emotions, secondary emotions, tertiary emotions [5][6]，而我們會針對主要情緒的 Anger、Sadness、Fear 作為情緒分類之參考如圖 2。

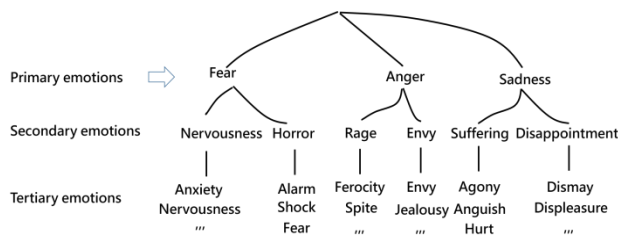


圖 2. 情緒分類圖 [7]

2.3.3 語音特徵

研究中，探討在聲音中可能影響情緒的特徵值，在 [8] 的研究中指出 pitch(fundamental frequency)是作為情緒識別的主要因素、其他因素包含在 vocal energy、frequency spectral features、formants、temporal features。另一種特徵提取的方法是加強或增進整組特徵，藉由考慮一些特徵的衍生，比如 Linear Predictive Coefficients (LPC)、Zero Crossing Rate (ZCR)。每種情緒都有其代表鑑別度的語音特徵，因此將這些語音特徵合併使用，以此達到更好的辨識率。

2.3.4 數位取樣 (Sampling rate)

人的發聲系統，主要是聲帶、聲道、鼻腔和嘴唇四個器官所構成，氣流由人體的肺部被擠壓出，經過氣管到達聲帶，聲帶振動所產生之氣流，在發音腔道內共振，產生聲波，因而形成語音訊號。聲音是一種波形，隨著時間做連續的變化，是一種連續的類比訊號。但是電腦卻無法處理類比訊號，因此必須將類比訊號取樣，轉換為數位訊號。

2.3.5 音量

「音量」代表聲音的強度，又稱為「力度」、「強度」(Intensity) 或「能量」(Energy)，可由一個音框內的訊號震幅大小做計算，一般有兩種計算方法，其中 s_i 為該音框的第 i 個採樣點， n 為該音框的採樣總點個數。

- 每個音框的絕對值的總和

第一種是取每個音框振幅的絕對值總和，如公式 (1)，這種方式雖然計算量小，但不太符合人的聽覺感受。

$$\text{Volume} = \sum_{i=1}^n |s_i| \quad (1)$$

- 振幅平方和的常數對數的十倍

第二種則是取每個音框振幅平方和的常數對數的十倍，如公式 (2)，單位為分貝，是一個對數強度值，比較符合人耳對聲音大小的感覺。[9]

$$\text{Volume} = 10 * \log_{10} \sum_{i=1}^n s_i^2 \quad (2)$$

2.3.6 過零率 (Zero Crossing Rate: ZCR)

過零率是在每個音框中音訊通過零的次數，是一種輕鬆計算的基本聲學特徵[10]。其具有下列特性：

- (1) 一般而言，雜訊及氣音的過零率均會大於正常說話聲音。
- (2) 雜訊和氣音較難從過零率來分辨。計算方式如公式 (3)，定義訊號 s 的第 i 個值以 s_i 表示。

$$\text{zcr} = \sum_{i=2}^n |\text{sgn}(s_i) - \text{sgn}(s_{i-1})|$$

$$\text{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3)$$

2.3.7 適應性取樣

在數位取樣中，通常使用固定之取樣頻率取樣類比訊號，雖然這種取樣方式可以把一個音檔完整的紀錄起來，但其實判斷資料特徵所需用到的資料並不需要這麼多，因此可以用適應性取樣把特徵比較不明顯的資料做刪減，假設有一音檔訊號如圖 3，首先用斜率將每個點與其相鄰點的相對位置求出如圖 4，接下來比較相鄰的兩斜率，如果兩斜率數值差異不大，那麼就會把兩斜率中間的訊號給剔除如圖 5，最後得出所需的訊號如圖 6，這樣不僅可以減少程式運算量，且對於找尋特徵值的影響不大。

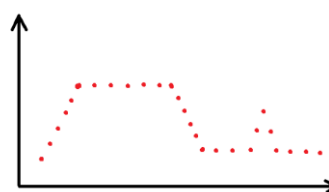


圖 3. 原始訊號

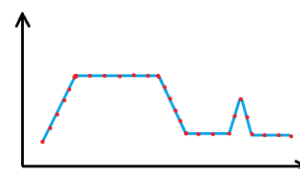


圖 4. 比較斜率

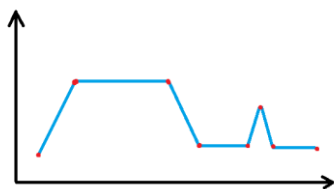


圖 5.適應性取樣

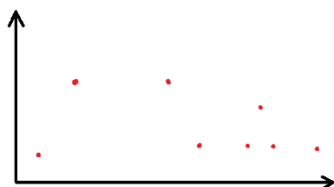


圖 6.取樣後訊號圖

2.3.8 音框

在語音訊號中，短時間的變化是很緩慢的，因此在分析語音訊號特性時，可將其分段處理。音框的切取方式是由 N 個點組成一個音框，由起點開始每隔 M 個點取一個新的音框，重複所有信號都至少包含在一個音框裡，如圖 7。

若 $M < N$ ，則相鄰的兩個音框將會重疊，音框間的銜接也較平滑；若 $M > N$ ，則相鄰的音框則不會重疊，會造成一些位被包含於音框中的資料被遺失。

為了避免資料的遺失，通常會將音框做重疊 (Overlap)，音框長度通常取 10~40ms，而寬度通常為音框的 $1/3$ 到 $1/2$ 之間。這樣的音框長度適合計算取得語音的特徵參數，本專題根據使用 32ms 的音框長度，重疊 16ms 的音框長度。

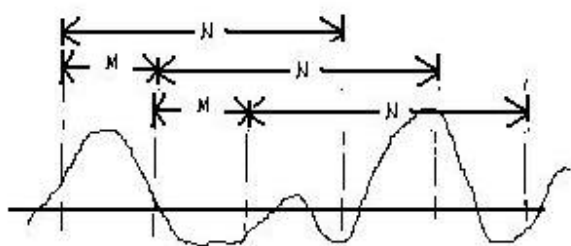


圖 7.音框切分

2.3.9 語料庫

語料庫是指詞語料資料庫，通常包含了大量經過分析整理的文本，具有既定的格式與標記，藉以提供語言學相關研究做分析與統計，語料庫的來源的詞語與文件越多越好，例如最具有代表性的機讀語料庫，是第一個平衡語料庫。舉世聞名的美國的布朗語料庫中便包含了一百多萬種

詞語，[11]包含了一千多萬的詞語，是世上第一個擁有完整詞類標記的中文語料庫。

在本專題的語料庫中參考了憂鬱自我評估量表[12]、小鬱亂入[13]、簡式健康表 (BSRS-5) [14]、一分鐘測出微笑抑鬱症[15]、憂鬱症邊緣人性格焦慮症自我檢測[16]、憂鬱症檢測[17]、別再叫我加油好嗎?[18]等有關研究主題的各式量表、文章與書籍內容來放入我們的語料庫。

再者，我們也與心理諮商師溝通討論來檢測我們放入的詞彙是否符合，以免變成只有製作人的判斷來增加可信度。

2.3.10 中文斷詞

在拼音文字的系統中，由於構成詞語的字母與詞語彼此之間的距離不一樣，因此可以明顯判別出每個詞語的界線；與其不一樣的，中文屬於表意文字，在構成詞語的字與詞語之間的距離是相同的，因為中文此種特別構成方式，使得斷詞成為學習中文時的挑戰，但並不是每一個句子都能斷的如此順遂，例如「我們在野生動物園玩」可以被斷成「我們／在野／生動／物／園／玩」或是「我們／在／野生／動物園／玩」，可是我們比較希望可以斷成後面的句子，所以下面有一些方法可以幫忙解決這個問題。

第一個方式使用正向最大匹配法，系統會有一個語料庫，將句子在語料庫中由前向後比對，一一比對最長詞的匹配結果，就可以斷出「我們／在野／生動／物／園／玩」。第二個方法是使用逆向最大匹配法，我們將句子在語料庫中由後向前比對，一一比對最長詞的匹配結果，就可以斷出「我們／在／野生／動物園／玩」。第三種方法則是正向跟逆向都算一遍，取每組句子切割的詞語中最長的，因此我們會斷出「我們／在／野生／動物園／玩」這個結果。

不過上面的做法都有一個共同缺點，只要語料庫沒有收錄句子中的詞，那效果可能不會很好。因此現在大部份比較好的斷詞系統都是使用全切分方法，切分出與詞庫匹配的所有可能，再運用統計模型決定最好的切分結果。例如 Jieba 中文斷詞程式[19]就是使用類似的方法完成的。

而為了避免語料庫太過龐大，避免程式運行過慢，所以放進語料庫的資料經過整理，使用了基於前綴辭典如表 2，實現高效能的掃描，生成每一個句子中詞彙所有可能成詞所構成的有向無環圖 (DAG)，加快執行速度。

表 2.中文前綴

單字	可能字詞
時	時
	時薪
	時薪制
	時裝
	時裝劇
	時裝史
	時裝屋

2.3.11 Edit distance

稱 Levenshtein Distance[20]，是指兩個字串之間，由一個轉成另一個所需的最少編輯操作次數。許可的編輯操作包括將一個字元替換成另一個字元，插入一個字元，刪除一個字元。一般來說，編輯距離越小，兩個字串的相似度越大。

編輯距離可以用在自然語言處理中，例如拼寫檢查可以根據一個拼錯的字和其他正確的字編輯距離，判斷哪一個（或哪幾個）是比較可能的字。DNA 也可以視為用 A、C、G 和 T 組成的字符串，因此編輯距離也用在生物信息學中，判斷二個 DNA 的類似程度，如圖 8。

		f	a	s	t
	0	1 1	2 2	3 3	4 4
c	1	1 2	2 3	3 4	4 5
	1	2 1	2 2	3 3	4 4
a	2	2 2	1 3	3 4	4 5
	2	3 2	3 1	2 2	3 3
t	3	3 3	3 2	2 3	2 4
	3	4 3	4 2	3 2	3 3
s	4	4 4	4 3	2 3	3 3
	4	5 4	5 3	4 2	3 3

圖 8.Edit distance

3.研究流程

3.1 組內分工

組員名稱	工作內容
金子恩	撰寫分析語調程式、整合程式碼、APP 開發
林逸融	撰寫分析語調程式
羅煒傑	撰寫語料庫程式
劉泯錚	協助程式開發、文書處理、影片製作、資料搜集與整理、會議記錄

3.2 研究進度甘特圖

圖 9 表示組內的研究進度甘特圖。

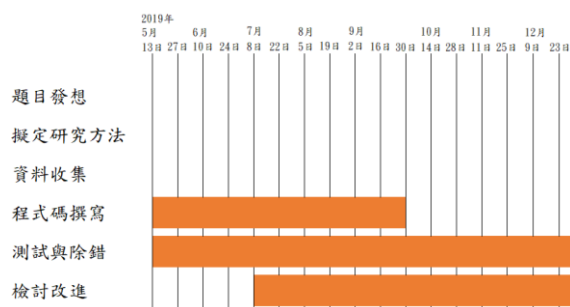
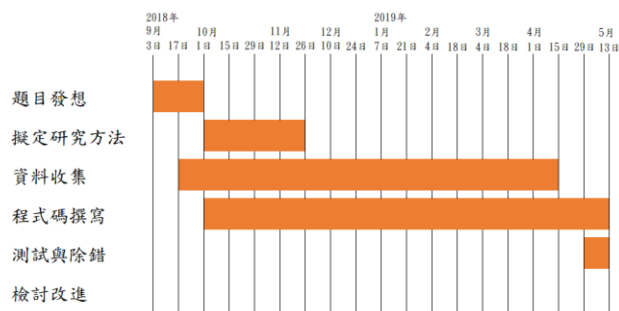


圖 9.研究進度甘特圖

3.3 現階段研究結果

本專題的系統架構之流程如圖 10 所示，架構中的每項階段和步驟，將會依序詳細說明。

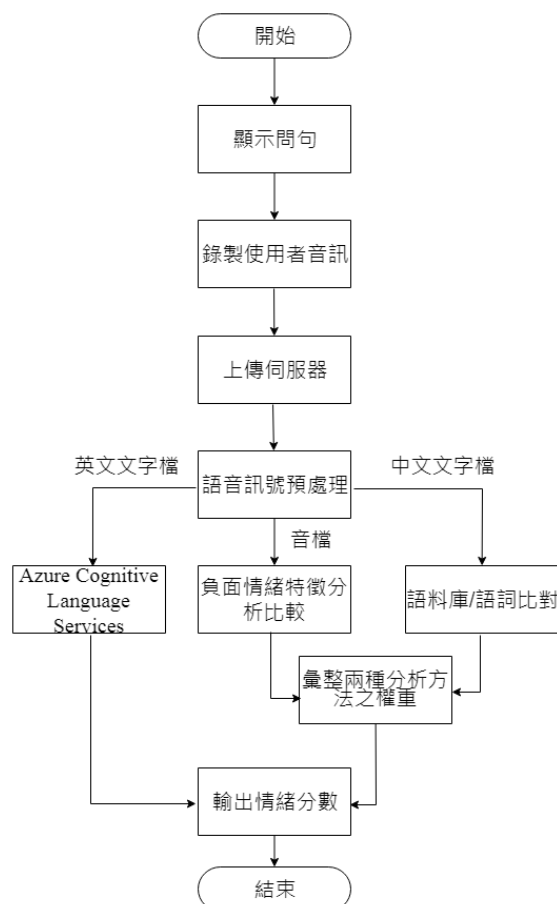


圖 10.智慧型自我情緒判定系統架構之流程

3.3.1 語音訊號預處理

將錄製好的音訊做預處理，在預處理的過程中，依序分為數位取樣、語音轉文字、音量計算、過零率計算、適應性取樣、端點偵測、切割音框，流程如圖 11 所示。

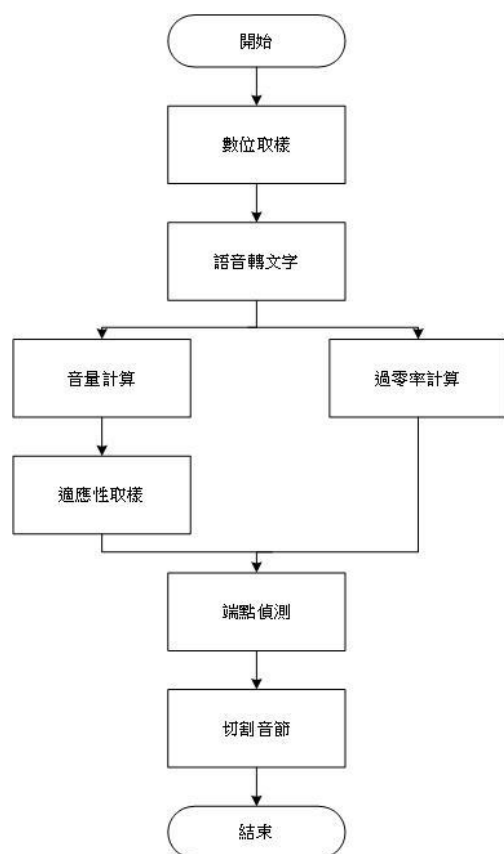


圖 11.語音訊號預處理

3.3.1.1 數位取樣

本專題經由數位錄音裝置，錄製使用者之語音訊號，並以特定編碼率(bit rate)：16 位元、取樣率(sample rate)：8KHz、單聲道轉換為離散時間序列。

3.3.1.2 語音轉文字

透過 Python 中的 SpeechRecognition 套件。使用 Google 語音識別 API 將語音訊號轉換為文字檔，以利用文字內容做後續之情緒特徵判別。

3.3.1.3 音量計算

音量數值能做為端點偵測的其中一項考量，本專題先將原始波形切割成每秒鐘 62.5 個音框，再根據的音量公式 (1)，求得所需的音量數值，圖 12 為音檔原始波形，圖 13 為此音檔音量，從時域圖能夠觀察到，在 10 秒~12 秒、12 秒~14 秒間有兩段句子。

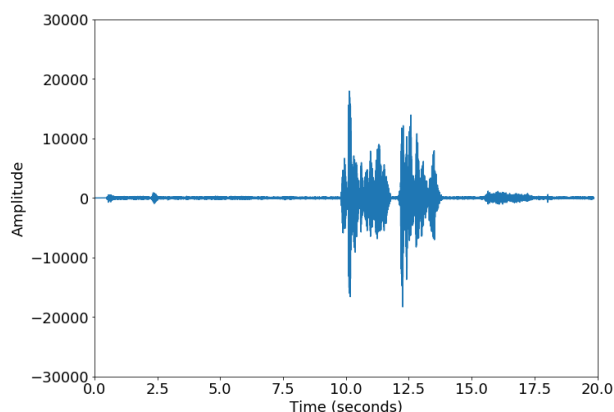


圖 12.音檔原始波形

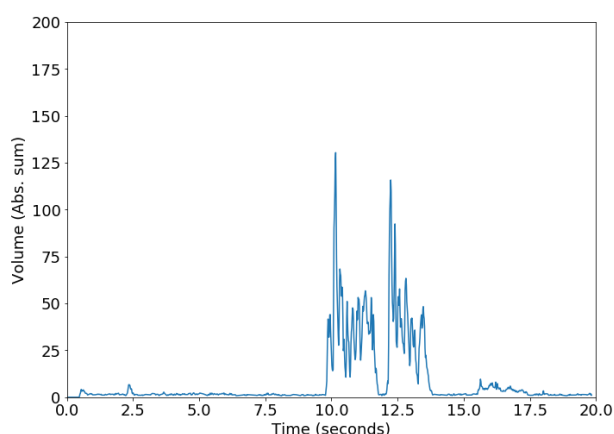


圖 13.音檔音量

3.3.1.4 過零率計算

圖 14 為音檔之過零率計算，在音量圖中能夠區分環境音與說話聲音的不同，但若是音量較大的氣音則沒辦法比較、分辨。

於是本組參考了雜訊與氣音在過零率中均大於說話聲音的特性，藉由實驗得到氣音的平均值與正常音的平均值，分別落在 83 與 57 左右，再將兩個值取平均來當作判斷氣音的標準值，若該段音節中的點有半數超過氣音標準值，則該段可能是氣音或環境音。由此標準可觀察在時域圖中 1~2 秒、11~12 秒、16~18 秒段落間過零的次數較高，此段可能為氣音片段。

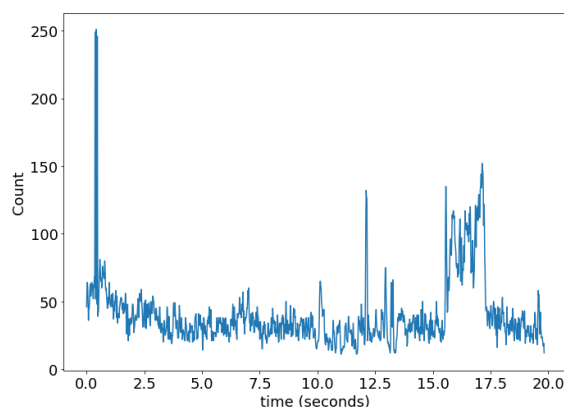


圖 14.過零率計算

3.3.1.5 端點偵測

開發此項目時，根據端點偵測的研究資料[21]，實作了利用音量數值判斷端點的基礎版本，當音量小於某個門檻值如圖 15，將認定是靜音或雜訊，若是音訊乾淨，雜訊較小，將會得到不錯的效果，如圖 16，端點時間如表 3。門檻值設為：
 (最大音量數值*0.05)+(最小音量數值*5.0)

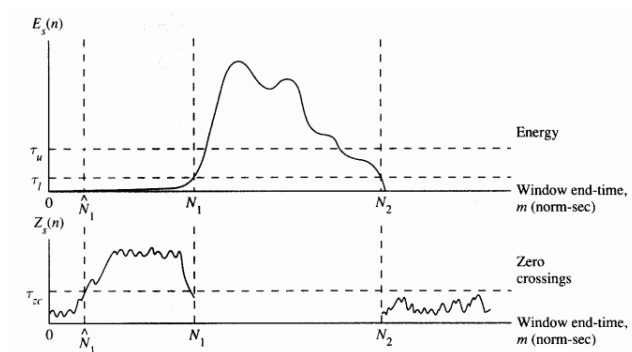


圖 15.門檻值示意圖

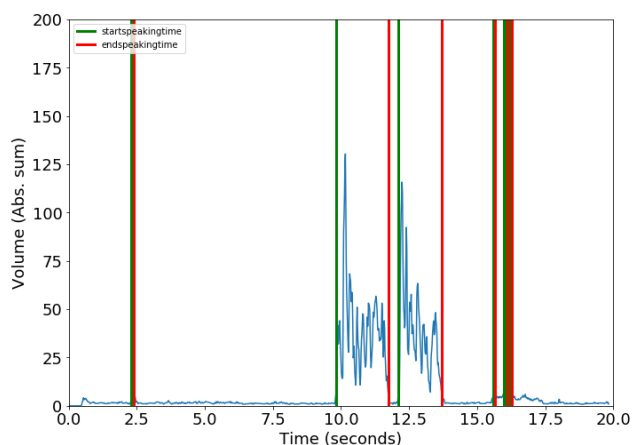


圖 16.端點偵測音量圖

表 3.端點時間圖

Total time : 19.824 (秒)		
	Start speaking time (秒)	End speaking time (秒)
1.	2.304	2.384
2.	9.808	11.744
3.	12.096	13.712
4.	15.6	15.664
5.	15.984	16.048
6.	16.048	16.096
7.	16.192	16.24
8.	16.256	16.288

但若是遇到以下問題，則結果將會與預期不符。

- 雜訊過大。
- 氣音過大。
- 音量變化量過大。

因此為了有效識別氣音、雜訊、字句間的停頓時間，將利用預處理後的資料協助判斷端點。

(1).將氣音、雜訊的音節過濾

透過音量為主，過零率為輔的方式，本組將音量所偵測出來的端點時間，計算區間的平均過零率，若是高於氣音、雜訊的標準值，則剷除此音節，如圖 17，端點時間如表 4。

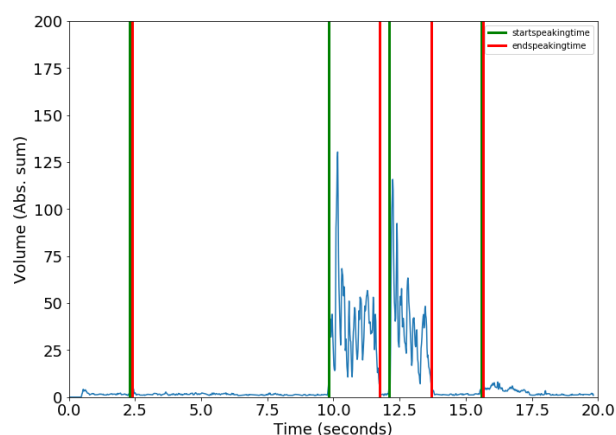


圖 17.端點偵測音量圖

表 4.端點時間圖

Total time : 19.824 (秒)		
	Start speaking time (秒)	End speaking time (秒)
1.	2.304	2.384
2.	9.808	11.744
3.	12.096	13.712
4.	15.6	15.664

(2).字句間的停頓時間

透過實驗數據去實測一句話內字與字的間隔時間，進而用此標準去判斷該間隔是兩句話之間的間隔還是一句話裡字與字的間隔。藉由實驗數據裡得知一個字最少要花多少時間，再以此標準刪除時間太短的音節，最後得出我們所需要的端點，如圖 18，端點時間如表 5。

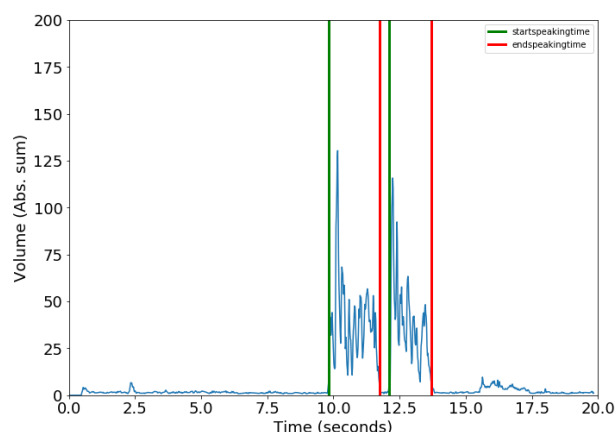


圖 18.端點偵測簡易版

表 5.端點偵測結果

Total time : 19.824 (秒)		
	Start speaking time (秒)	End speaking time (秒)
1.	9.808	11.744
2.	12.096	13.712

3.3.1.6 輸出結果

表 6 為音檔預處理後輸出的結果及各項數值。

表 6.預處理後輸出結果

完整音量平均	34.21007958
完整音量變異數	964.27320216
完整語速	2.07353699
斷句音量平均	34.47533888
斷句音量變異數	1206.78503734
斷句音量語速	0.0
斷句沉默時間	0.0

3.3.2 時域中，負面情緒與聲音特徵的關係

為了取得負面情緒與聲音特徵的關係，本組透過與心理諮商師討論，藉由諮商師與病人晤談的經驗和準則，將負面情緒分類為：憂鬱、焦慮、悲傷、害怕與生氣五類，並整理出以下的假設：

在時域、文字中為了取得負面情緒與聲音特徵的關係，本組透過與心理諮商師討論，藉由諮商師與病人晤談的經驗和準則，將負面情緒分類為：憂鬱、焦慮、悲傷、害怕與生氣五類，並整理出以下的規則：Vol 為平均音量、Var 為音量變異數、S 為語速、M 為沉默時間、T 為文字數。

- 生氣情緒時的平均語速、音量變異數、平均音量，皆會高於常模。

S_angry>S_normal
Var_angry>Var_normal
Vol_angry>Vol_normal

- 悲傷情緒時平均語速、音量變異數、平均音量皆低於常模，且文字量少，沉默時間較長。

S_sad<S_normal
Var_sad<Var_normal
Vol_sad<Vol_normal
T_sad<T_normal
M_sad>M_normal

- 焦慮情緒時平均語速會高於常模、平均音量與音量變異數則低於常模。

S_anxious>S_normal
Vol_anxious<Vol_normal
Var_anxious<Var_normal

- 害怕情緒時平均語速與平均音量會低於常模、音量變異數則高於常模。

S_scared<S_normal
Vol_scared<Vol_normal

Var_scared>Var_normal

害怕和焦慮屬於同一向度，較難區分，通常搭配字句去判斷。

為了驗證以上的假設，本組找來 4 男 4 女共 8 位同學，設計情境和語句請測試者揣摩並記錄過程，再根據音訊最大音量、最小音量、平均值、語速、變異數等特徵值與常模觀察並比較數據結果，如表 7。

表 7.情境設計

問題	觀察常模與各個負面情緒之差異。
情緒	憂鬱、焦慮、悲傷、害怕、生氣
範例語句	
生氣	你真的很煩欸！講過幾次了，就叫你東西不要亂丟要收好，你有沒有在聽啊
	門都不鎖，你乾脆立個牌子歡迎人家來偷！
	為什麼是我？為什麼是我生病？
害怕	那裡有很多狗在叫欸，我不敢自己走過去啦！
	全世界都討厭我，我要怎麼活下去？
	是不是每天都有人要我的命。
悲傷	我根本沒資格在這裡。
	我又造成別人的麻煩了。
	我不夠好，沒資格跟他們競爭。
焦慮	怎麼辦？我好累，我快活不下去了。
	我是不是很沒用，什麼忙都幫不上。
	我的存在根本就沒有意義啊。
憂鬱	其實我還挺期待死的那一天的。
	至少還有人會覺得我還活著。
	我想悄悄的死去，我會很開心。

3.3.3 負面情緒語音特徵分析

圖 19 為負面情緒語音分析架構圖，目前已完成部分如下：

將使用者錄製的常模音檔數值與之前做實驗得出的數值來去做調整，藉此得到一個屬於使用者個人的判斷標準，例如使用者的平均音量在實驗得出的平均音量範圍裡是比較低的，那麼我們會將實驗裡面每種情緒的平均音量調低，來當作本系統來判斷使用者情緒狀況的一個標準，再將使用者錄的情緒語句數值丟到使用者個人的判斷標準，最後會得出使用者每種情緒的分數。

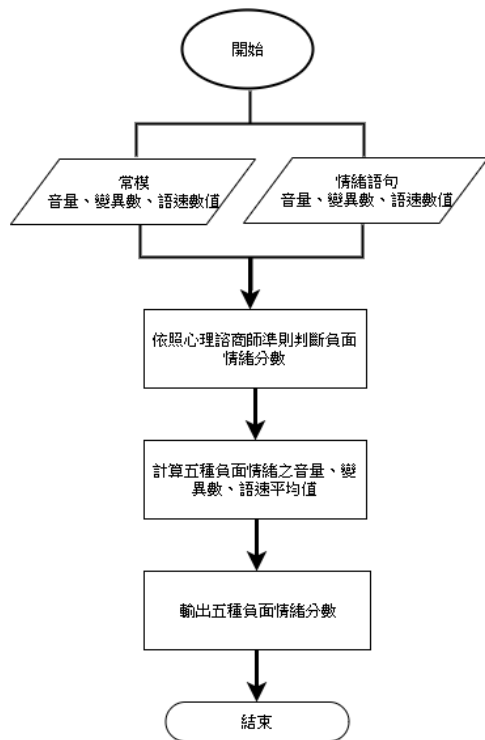


圖 19.負面情緒語音分析架構圖

3.3.3.1 判斷標準

我們藉由心理諮商師提供的規則，搭配上一節所得到的實驗音檔去尋找各個數值落在哪個區段是可以代表甚麼情緒，結果如表 8。

表 8. 心理諮商師提供的規則數據

normal range	volume	34.21~40.91 (秒)
	variance	806~1516
	speed of speech	2.5~3.7 (字/秒)
angry range	volume	43.51~68.0 (秒)
	variance	2034~3127
	speed of speech	2.8~6.5 (字/秒)
sad range	volume	26.0~32.9 (秒)
	variance	341~830
	speed of speech	0.8~3.4 (字/秒)
anxiety range	volume	28.8~40.91 (秒)
	variance	512~1948
	speed of speech	2.6~4.8 (字/秒)
scared range	volume	27.5~36.8 (秒)
	variance	926~2110
	speed of speech	2.2~3.4 (字/秒)

3.3.3.2 數據計算方式

我們這邊有做過兩種計算方式，第一種是我們一開始採用的計分方式，其操作方式為判斷情緒語句數值是否有過每種情緒的門檻值，如果有

超過，則在該情緒分數加分，分數加完後，再把每種情緒的分數做標準化，讓使用者可以看到他這句情緒語句的情緒成分比例為多少，但缺點是如果使用者說了比較不帶情緒的語句時，雖然他的每種情緒分數較低，他還是會因為標準化的關係而顯示出每種情緒比例，這會讓使用者有誤解如表 9、表 10，第一張為生氣音檔的數值分數，第二張為正常說話的數值分數，但是從這兩張圖看到，分數比例是相差不多的，這會讓使用者有混淆的狀況發生。

表 9. 生氣音檔的數值分數

生氣	34.57
悲傷	29.90
焦慮	18.74
害怕	16.79

表 10. 正常說話的數值分數

生氣	9.63
悲傷	22.37
焦慮	38.59
害怕	29.41

第二種方式是我們目前採用的方式是以線性的方式去計算分數，操作方式有兩種，第一種為判斷情緒的大小門檻標準高於使用者常模數值時，如果使用者的數值沒有超過情緒的最小門檻的話，則分數為 0 分，使用者數值介於最大門檻與最小門檻之間的話，則使用公式 4，當過了最大門檻的話，則分數就會到達滿分 100 分。第二種為判斷情緒的大小門檻標準低於使用者常模數值時，如果使用者的數值沒有低於情緒的最大門檻的話，則分數為 0 分，使用者數值介於最大門檻與最小門檻之間的話，則使用公式 5，當低於最小門檻的話，則分數就會到達滿分 100 分，計算完後會得出每種情緒各項數值的分數如表 11。接著將各個情緒裡的所有數值的分數相加取平均，會得到每種情緒的分數如表 12。這種方式算出來的分數會隨著使用者情緒起伏做調整，會讓使用者看起來較為協調。

$$\frac{(\text{情緒分數}-\text{最小門檻數值})}{(\text{最大門檻}-\text{最小門檻})} \times 100 \quad (4)$$

$$\frac{(\text{最大門檻數值}-\text{情緒數值})}{(\text{最大門檻}-\text{最小門檻})} \times 100 \quad (5)$$

圖 20. 語料庫／語詞比對系統架構圖

表 11. 各數值的情緒分數表

生氣平均音量分數	0
生氣音量變異數分數	0
生氣語速分數	0
難過平均音量分數	0
難過音量變異數分數	0
難過語速分數	24.868
焦慮平均音量分數	18.268
焦慮音量變異數分數	69.903
焦慮語速分數	6.9703
害怕平均音量分數	0
害怕音量變異數分數	42.0449
害怕語速分數	78.8827

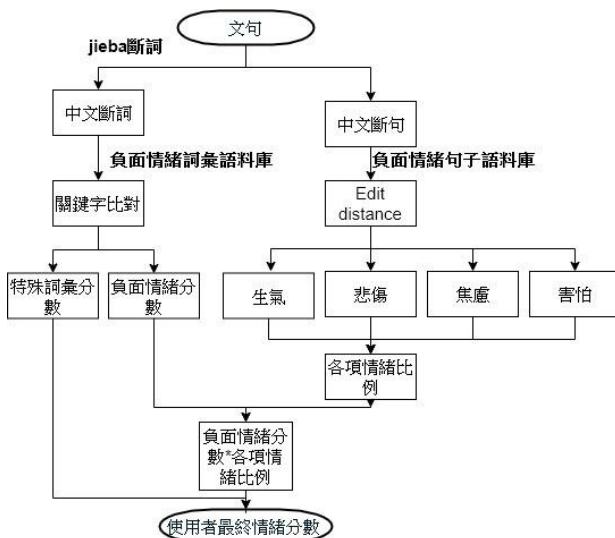
表 12. 每種情緒分數表

生氣	0.00
悲傷	8.29
焦慮	31.71
害怕	40.31

3.3.4 語料庫／語詞比對

圖 20 為語料庫／語詞比對系統架構圖，目前已完成部分如下：

將語音訊號預處理所得出的中文文字檔跟中文斷句檔，把中文文字檔進行 jieba 斷詞來斷詞，再跟負面情緒詞彙語料庫比對關鍵字，劃分為負面情緒分數跟特殊詞彙分數，負面情緒分數依照關鍵字權重來計算得出，特殊詞彙分數以比對到「自殺」等高危險詞給予特殊分數，並把中文斷句檔跟負面情緒句子語料庫使用 Edit distance 來判斷兩個句子的最短編輯距離，得出各項情緒比例，將負面情緒分數 x 各項情緒比例得出使用者最終的情緒分數。



3.3.4.1 文字檔斷詞

利用預處理過後的音訊文字檔斷詞。

例如：

「對於自殺不能解決問題不要把痛苦留給愛妳的人想想身邊的親朋好友這些話語真的非常反感」

斷詞為：

「／對於／自殺／不能／解決／問題／不要／把／痛苦／留給／愛／妳／的／人／想想／身邊／的／親朋好友／這些／話語／真的／非常／反感／」

3.3.4.2 情緒表分數

將斷詞後的單詞與與心理諮商師討論所得出的負面情緒字詞權重表做統計分數及次數，繪製成長條圖來辨識哪個情緒佔比例較重；另一步驟是將一整串話斷成一小句，以上的例句。

舉例：

「對於自殺不能解決問題／不要把痛苦留給愛妳的人／想想身邊的親朋好友／這些話語真的非常反感」

若只把一段話斷成一個一個詞彙，可能會因為新詞或太過口語而造成斷詞不夠好，因此本組從心理諮商師推薦的書籍中，將內容裡以負面情緒口吻的句子加入到口語的語料庫，並用 Edit distance 比較的斷句最短距離有多少，以此直接判斷這一句歸類在哪種情緒分類，最後統計成長條圖，以提醒使用者了解自己近期的情緒狀況。

3.3.5 彙整兩整分析方法之權重

我們根據心理諮商師的意見，決定將語音情緒分數跟語料庫情緒分數做整合，將他們兩個做標準化後，再算其平均，得出最後輸出的情緒分數，如表 13，整合後的數值相較於未整合的數值較容易讓使用者去閱讀。

表 13. 整合後的情緒分數表

生氣	0.00
悲傷	0.30
焦慮	0.40
害怕	0.79

3.3.6 Azure Cognitive Language Services

Azure Cognitive Language Services 是一項自然語言處理(NLP)服務，使用機器學習來發現文本中的洞察訊息。將語音轉為文字後的字句，透過文字情緒分析 API，辨識使用者所表達的情緒分數。以作為辨識負面情緒的一項參考，情緒分數介於 0~1 之間，越接近 0 的分數則表示負面情緒，分數為 0.5 則表示缺乏情緒，例如事實的陳述。

如圖 21 為中文文字檔轉換英文文字檔，使用 Text Analytics API 後所得出之情緒分數。

```
{
  "documents": [
    {
      "language": "cn",
      "id": "1",
      "text": "How can I do it, I am so tried, I can't live anymore."
    },
    {
      "language": "cn",
      "id": "2",
      "text": "You are really annoyed!"
    },
    {
      "language": "cn",
      "id": "3",
      "text": "I think my sister is very good to me."
    }
  ]
}

{
  "documents": [
    {
      "id": "1",
      "score": 0.0064101815223693848
    },
    {
      "id": "2",
      "score": 0.011714428663253784
    },
    {
      "id": "3",
      "score": 0.82522678375244141
    }
  ],
  "errors": []
}
```

圖 21.情緒分數

3.3.7 期初進度與期中進度表

表 14 為期初進度與期中進度之比較。

表 14.期初、期中進度表

期初進度	期中進度
錄製語音。	計算原始波形之過零率並表示。
計算原始波形轉之音量並表示。	端點偵測。
計算整段音檔之特徵結果。 ● 最大音量 ● 最小音量 ● 平均音量 ● 語速 ● 變異數	針對端點偵測後的句子計算其結果。 ● 最大音量 ● 最小音量 ● 平均音量 ● 語速 ● 變異數
建立中文斷詞資料庫及權重。	實作適應性取樣，減少訊號資料量，但保留其特徵。
語音轉文字。	新增詞語至中文斷詞資料庫。
將斷詞後的結果擷取負面情緒關鍵字，例如：	建立中文斷句資料庫及權重。

我很難過、我想哭。

利用 Edit distance 加快資料搜尋速度。

4.系統介面設計

圖 22 為 App 系統架構圖，使用者將透過行動裝置錄製情緒語句，錄製完畢後，系統會將音檔上傳至 Server 端執行分析程式，最後將分析後的數值儲存至 Client 端。

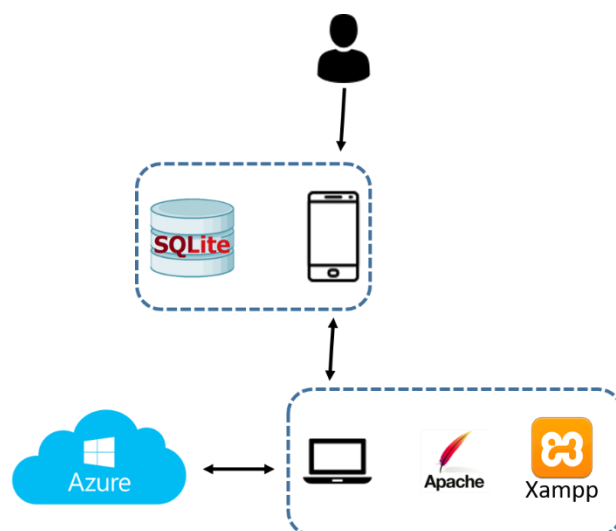


圖 22.App 系統架構圖

- 使用者透過手機使用本系統。
- 電腦透過 Xampp 架設的 Apache 建立區域網路進行傳輸。
- 手機端將音檔透過 PHP 上傳至 Server，並發送參數執行前處理、分析程式碼。
- Server 端程式透過呼叫 Azure Text Analysis API 分析使用者文字中所包含的情緒。
- 分析結束後將結果回傳至手機端 SQLite 資料庫儲存。

4.1 介面 (UI)、使用者體驗 (UX) 設計圖

圖 23 為 App Logo，圖 24 為進場動畫，圖 25~31 為錄製常模語音頁面，圖 32 為上傳過場動畫，圖 33 為 App 主頁面，圖 34 為負面情緒檢測功能說明，圖 35 為負面情緒檢測頁面，圖 36~38 為錄製負面情緒語句頁面，圖 39 為音檔上傳動畫，圖 40 為分析後負面情緒結果，圖 41~42 為關懷系統頁面，圖 43 為記錄圖表。



圖 23.App Logo



圖 24.App Logo

初次使用系統，系統為了建立使用者常模數值，將會錄製使用者常模語音，如圖 25。

點擊紅色圓形開始錄音如圖 26，初次使用將會要求麥克風及檔案存取權限，如圖 27。

接著頁面將會出現範例句子，按下紅色圓形按鈕將會開始錄音，在一次點擊則停止錄音，如圖 28。錄製三次後點選右方向鍵後會將常模音檔上傳至伺服器分析並進入到主頁，如圖 29~33。



圖 25.錄製常模語音



圖 26.錄製常模語音



圖 27.錄製常模語音



圖 28.錄製常模語音



圖 29.錄製常模語音



圖 30.錄製常模語音



圖 31.錄製常模語音



圖 32.上傳過場動畫

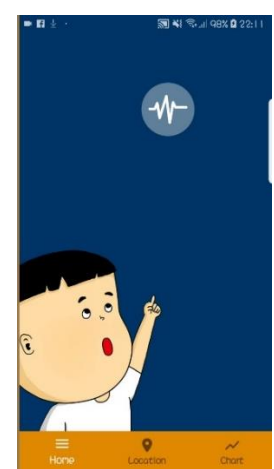


圖 33.App 主頁面

點選聲波圖形介紹則開始介紹情緒檢測功能，如圖 34，按下右方向鍵則入情緒檢測頁面，如圖 35。

點選嘴巴圖形則進入到錄音介面，如圖 36，此頁面上方會隨機句子詢問使用者近況，點擊紅色圓形按鈕則開始錄音，如圖 37，再次點擊則停止錄音，如圖 38，按下右方向鍵則將情緒語句上傳至伺服器分析，如圖 39。

分析後的數值將會儲存至手機資料庫並顯示數值點選房子圖形則回到主頁面，如圖 40。

圖 41 為關懷系統，此頁面將會顯示使用者

目前位置，愛心圖形為全台所有心靈診所位置地標，使用者可選擇離自身最近的心靈診所尋求協助、諮商服務。

點選愛心圖形後則出現心靈診所名稱及電話，右下角地圖圖示可導航到目的地，如圖 42。

圖 43 為紀錄圖表，將會顯示使用者各個負面情緒之數值，點選不同表情顯示不同情緒之圖表。



圖 34 情緒檢測功能說明



圖 35 為負面情緒檢測頁面

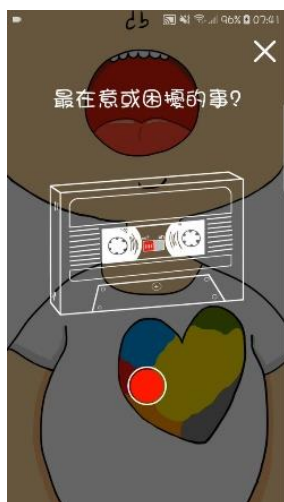


圖 36 錄製負面情緒語句

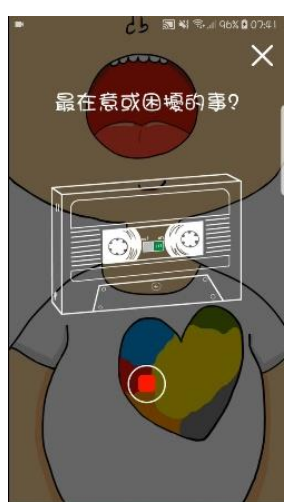


圖 37 錄製負面情緒語句

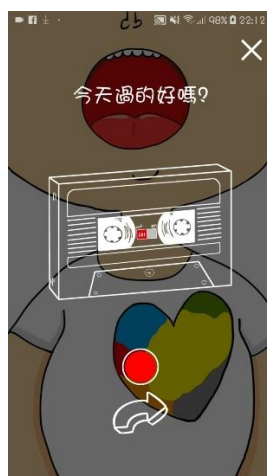


圖 38 錄製負面情緒語句

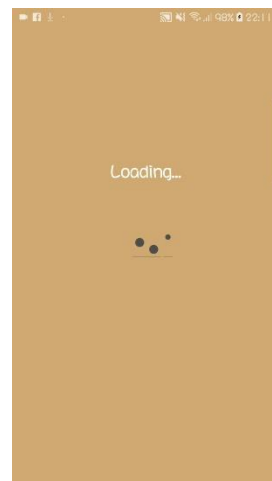


圖 39 音檔上傳動畫

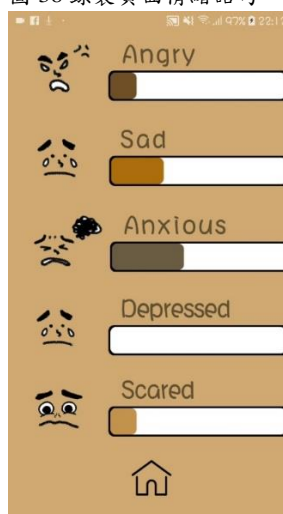


圖 40 分析後負面情緒結果



圖 41 為關懷系統頁面



圖 42 為關懷系統頁面

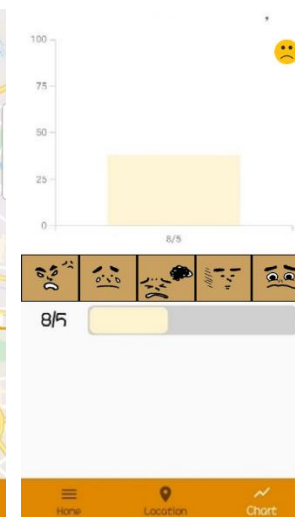


圖 43.紀錄圖表

5.未來工作

- 找測試人員協助測試，改善 App 流程及分析方式。
- 將 Server 架設到雲端
- 將此 App 上架至 Play 商店

附註:本專題合作對象為：陳姝瑤 臨床心理諮商師。

參考文獻

- [1] 余文章、洪偉欽 (2009)，運動與憂鬱症之探討，嘉大體育健康，8 (1)，228-234。
- [2] 衛生福利部統計處，(2018)，Available：
<https://dep.mohw.gov.tw/DOS/cp-1720-9734-113.html>
- [3] 楊東麟、洪明傳 (2012)，研究與開發一個具備影像與聲音辨識之行動雲端憂鬱症醫療探勘系統。
- [4] Capturing the Sound of Depression in the Human Voice，
[Online]，Available:
<https://www.kqed.org/futureofyou/435986/capturing-the-sound-of-depression-in-the-human-voice>
- [5] Parrott, W, " Emotions in Social Psychology, Psychology",
(2001)
- [6] 鄭穎懋、包蒼龍 (2011)，以語音情緒辨識技術為基礎之情緒感知行動通訊服務。
- [7] "Basis emotions," [Online]. Available:
<http://changingminds.org/explanations/emotions/basic%20emotions.htm>. [Accessed FEB. 10, 2019].
- [8] V. Petrushin, " Emotion in speech: Recognition and application to call centers", (1999) .
- [9] 语音信号处理之时域分析-音量及其 Python 实现，
[Online]，Available:
<http://ibillxia.github.io/blog/2013/05/15/audio-signal-process-time-domain-volume-python-realization/>
- [10] Zero Crossing Rate (過零率)，[Online]，
Available:[http://mirlab.org/jang/books/audioSignalprocessing/basicFeatureZeroCrossingRate.asp?title=5-3%20Zero%20Crossing%20Rate%20\(%B9L%B9s%B2v\)](http://mirlab.org/jang/books/audioSignalprocessing/basicFeatureZeroCrossingRate.asp?title=5-3%20Zero%20Crossing%20Rate%20(%B9L%B9s%B2v))
- [11] 中央研究院現代漢語標記語料庫 4.0 版(Sinica Corpus 4.0)，[Online]，Available:<http://asbc.iis.sinica.edu.tw/>
- [12] 台灣人憂鬱症量表，[Online]，
Available:<https://www.jtf.org.tw/overblue/taiwan1/>
- [13] 小鬱亂入，[Online]，
Available:<http://depressytrouble.tw/index.php/quiz/>
- [14] 臺灣憂鬱症防治協會簡式健康表(BSRS-5)，[Online]，
Available:<http://www.depression.org.tw/health/health.asp>
- [15] 一分鐘測出微笑憂鬱症，[Online]，
Available:<https://blog.xuite.net/breeze6moon9/blog/235165344-%E4%B8%80%E5%88%86%E9%90%98%E6%B8%AC%E5%87%BA%E5%BE%AE%E7%AC%91%E6%8A%91%E9%AC%B1%E7%97%87>
- [16] 羅吉斯心理諮商所，[Online]，
Available:<http://www.rogersc.com/check.php>
- [17] kenkon 健康網，[Online]，
Available:http://www.kenkon.com.tw/melancholia_survey.php?act=list
- [18] 張閔筑。別再叫我加油，好嗎。三采出版社。2018 年。
- [19] 結巴中文分詞，[Online]，Available:
<https://github.com/fxsjy/jieba>
- [20] 編輯距離，[Online]，
Available:<https://zh.wikipedia.org/wiki/%E7%B7%A8%E8%BC%AF%E8%B7%9D%E9%9B%A2>
- [21] EPD in Time Domain (端點偵測：時域的方法)，
[Online]Available:[http://mirlab.org/jang/books/audioSignalprocessing/epdTimeDomain.asp?title=6-2%20EPD%20in%20Time%20Domain%20\(%BA%DD%C2I%B0%BB%B4%FA%A1G%AE%C9%B0%EC%AA%BA%A4%E8%AAk\)&language=chinese](http://mirlab.org/jang/books/audioSignalprocessing/epdTimeDomain.asp?title=6-2%20EPD%20in%20Time%20Domain%20(%BA%DD%C2I%B0%BB%B4%FA%A1G%AE%C9%B0%EC%AA%BA%A4%E8%AAk)&language=chinese)