

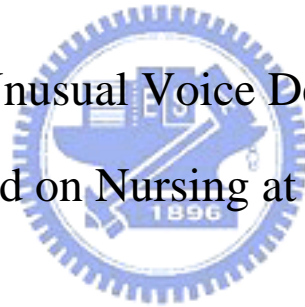
國立交通大學

資訊科學系

碩士論文

一個基於居家看護的異常聲音偵測法

An Unusual Voice Detector
Based on Nursing at Home



研究生：王朝君

指導教授：陳玲慧 教授

中華民國九十四年六月

一個基於居家看護的異常聲音偵測法

An Unusual Voice Detector Based on Nursing at Home

研 究 生：王朝君

Student : Chao Chun Wang

指導教授：陳玲慧

Advisor : Dr. Ling-Hwei Chen

國 立 交 通 大 學

資 訊 科 學 系

碩 士 論 文



A Thesis

Submitted to Department of Computer and Information Science

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer and Information Science

June 2005

Hsinchu, Taiwan, Republic of China

中華民國九十四年

一個基於居家看護的異常聲音偵測法

學生：王朝君

指導教授：陳玲慧 博士

國立交通大學資訊科學學系(研究所)碩士班

摘要

由於工業社會的快速發展和高齡化的社會結構產生越來越多的獨居老人，使得獨居老人缺乏妥善照顧。於是有發展居家看護系統的想法。就是在獨居老人的週遭裝設偵測儀器，即時傳送老人的生理狀況，進而掌握老人的健康情形。目前的看護系統大多是利用視訊或電子相關儀器來偵測被看護者之狀態，而本篇論文主要是從被看護者發出的聲音來偵測是否被看護者的健康情況有疑慮。首先我們定義四種危急聲音，分別為：劇烈咳嗽聲、呻吟聲、喘息聲和求救聲。當被看護者發出以上之危急聲音時，我們判定他的健康情況有疑慮，需有人實際去關注。為偵測上述的四種危急聲音，我們從聲波圖上擷取區域內聲音區段的數目（Number of the Segment Parts）、波形持續時間（Duration of Waveform）、平均音量（Mean of Volume）、正規化零越率（Zero Crossing Rate）和聲音區段彼此之間的相關性（Correlation）五種具區別率的特徵為我們偵測之依據。實驗結果顯示，對四種危急聲音的偵測率高達 94%~97%，且假警報只有 0.08% 的錯誤率。

An Unusual Voice Detector Based on Nursing at Home

student : Chao Chun Wang

Advisor : Dr. Ling-Hwei Chen

Institute of Computer and Information Science
National Chiao Tung University

ABSTRACT

Because fast development of the industrial society and structure of aging society, there are more and more solitary people. Thus, developing a nursing system at home is necessary. That is some instruments of detection are installed in the solitary people's home, the physiological states of the old man are conveyed immediately, and the health situation of the old man is under control. Most nursing systems use video information or electronic instrument to detect the healthy state of the person. In this method, we hope to know whether the health condition of the person nursed has a doubt by detecting from the voice that the person nursed emitted. We define four kinds of unusual voices at first, including cough, groan, wheeze and cry for help. When the person nursed sends out the above four kinds of unusual voices, we judge that his health condition have a doubt, and need someone to pay attention actually. In order to detect four kinds of unusual voices, we extract five features on audio waveform, including the number of segmented parts, duration of waveform, mean of volume, zero crossing rate and correlation. Experimental results show that the detection rate can be up to 94%~97% for these four kinds of unusual voices. In false alarm, there are only 0.08% of wrong rates.

誌謝

首先感謝指導老師 陳玲慧教授在我兩年碩士生涯的教導，使我能學習到許多專業知識與技能，還有許多處理事情的應有態度。還要感謝實驗室瑞祥、民全、萱聖、文超、惠龍學長們的熱心指導與幫忙。當然還有在這兩年時光中一起學習的同學，感謝大家的互相激勵使我受益良多。也感謝學妹的精神支持。是大家使我這兩年的學業生活多采多姿。此外還要謝謝我的口試委員陳朝欽教授、賴尚宏教授、陳佑冠教授們的建議和指導使我的論文更加完善。

最後我要感謝我的父母和家人，是他們在我背後給予我的默默支持，是我最大動力來源和奮而不懈的原因，所以我將這篇論文獻給我最親愛的家人。



目 錄

| | |
|--|-----|
| 摘要 | I |
| ABSTRACT | II |
| 誌 謝 | III |
| 目 錄 | IV |
| 圖目錄 | V |
| 表目錄 | VI |
| 第一章 簡介 | 1 |
| 1.1 動機 | 1 |
| 1.2 文獻探討 | 2 |
| 1.3 論文架構 | 5 |
| 第二章 特徵的擷取及選擇 | 6 |
| 2.1 前言 | 6 |
| 2.2 分割 | 9 |
| 2.3 特徵分析 | 14 |
| 2.4 特徵統計與選擇 | 15 |
| 第三章 異常聲音偵測法 | 23 |
| 3.1 波形相關性 (Waveform Correlation) | 23 |
| 3.2 各種聲音之判定條件及系統流程 | 27 |
| 第四章 實驗結果 | 30 |
| 第五章 結論 | 33 |
| 文獻參考 | 34 |

圖目錄

| | |
|--------------------|----|
| 圖一:求救聲聲波圖及分割區段 | 7 |
| 圖二:呻吟聲聲波圖 | 7 |
| 圖三:咳嗽聲聲波圖 | 8 |
| 圖四:喘息聲聲波圖 | 8 |
| 圖五:人對話聲聲波圖 | 8 |
| 圖六:電視聲聲波圖 | 9 |
| 圖七:平滑化及去雜訊後之求救聲聲波圖 | 10 |
| 圖八:平滑化及去雜訊後之呻吟聲聲波圖 | 11 |
| 圖九:平滑化及去雜訊後之咳嗽聲聲波圖 | 11 |
| 圖十:平滑化及去雜訊後之喘息聲聲波圖 | 11 |
| 圖十一:求救聲及其分割區段 | 12 |
| 圖十二:呻吟聲及其分割區段 | 13 |
| 圖十三:咳嗽聲及其分割區段 | 13 |
| 圖十四:喘息聲及其分割區段 | 14 |
| 圖十五:重新取樣聲波圖 | 24 |
| 圖十六:求救聲子類別一聲波圖 | 25 |
| 圖十七:求救聲子類別二聲波圖 | 26 |
| 圖十八:求救聲子類別三聲波圖 | 26 |
| 圖十九:呻吟聲子類別一聲波圖 | 26 |
| 圖二十:呻吟聲子類別二聲波圖 | 27 |
| 圖二十一:呻吟聲子類別三聲波圖 | 27 |
| 圖二十二:無線藍牙麥克風 | 30 |
| 圖二十三:USB 藍牙接收器 | 30 |
| 圖二十四:實際運用狀況 | 31 |

表目錄

| | |
|----------------------------------|----|
| 表一 特徵分析統計表..... | 16 |
| 表二 咳嗽聲的平均音量之平均、變異數和特徵臨界值..... | 17 |
| 表三 咳嗽聲的正規化零越率之平均、變異數和特徵臨界範圍..... | 18 |
| 表四 求救聲的持續時間之平均、變異數和特徵臨界範圍..... | 19 |
| 表五 呻吟聲的持續時間之平均、變異數和特徵臨界範圍..... | 20 |
| 表六 喘息聲的持續時間之平均、變異數和特徵臨界範圍..... | 21 |
| 表七 特徵臨界範圍統計表..... | 22 |
| 表八 偵測結果與偵測率..... | 32 |



第一章

簡介

1.1 動機

由於目前工業社會的快速發展和社會結構的變化，使得人口老年化問題越來越嚴重，加上社會福利越來越注重身體有殘缺的弱勢族群，所以針對目前人力不足且需要有效利用人力之情況下，發展居家看護系統是有必要的，以便減少社會人力成本的浪費。在過去的看護系統主要都是藉由視訊（Video）或電子儀器來觀察被看護者的健康情況是否正常，而在顧慮到被看護者的隱私和希望減少與沉重的電子儀器直接接觸，於是我們擷取被看護者發出的聲音訊號，經過處理之後來加以判斷此被看護者是否處於危險狀態。在此，我們將被看護者發出的危急聲音視為音訊關鍵字，在我們系統中，當被看護者發出音訊關鍵字時，系統會發出警訊告知相關單位需要注意此被看護者。在此，我們定義四種可能的危急聲音，分別為：劇烈咳嗽聲、呻吟聲、喘息聲和求救聲。

在本論文我們將發展一套以異常聲音(危急聲音)偵測為基礎的看護系統。當系統執行時，藉由被看護者身上掛的無線聲音接收器接收被看護者所發出的聲音，經過處理後加以判斷目前被看護者的狀況是否正常。如有異常，則發出警訊通知相關醫護人員處理。

1.2 文獻探討

目前有關於居家看護系統的研究，大多是利用視訊或電子相關儀器來偵測被看護者的狀態。Huang and Huang [1] 是利用視訊來偵測被看護者所做出之特別動作，如手掌握拳或伸出食指比出一等手勢，以顯示被看護者正在主動發出警訊，傳達自己健康狀況有疑慮之訊息，需要看護人員關注。然而此方法是不太可行的，因為當一個人生理有狀況需要援助時，自己是否真的能夠發出求救之手勢是值得關注的，且在視訊的偵測當中，更無個人隱私可言。Huang et. al[2]則在被看護者的身上裝設簡單電子儀器，主要用來偵測被看護者的心電圖，利用心電圖呈現出來的資訊來判斷被看護者健康狀況是否有疑慮，同時將判斷結果傳回醫護中心，但如此一來被看護者完全處於被動情況，無法自己發出危急訊號。

我們提出的方法是利用音訊分類（Audio Classification）的技巧，來研判被看護者是否需要被進一步的照顧。而音訊分類為目前針對音訊方面研究最為常見的類型之一，其在資訊安全和資料檢索等方面都有很好的應用[3-4]。音訊分類的方法大致可分為兩類：兩類分類法（Two-Way Classification）[5-13]和多類分類法（Multi-Way Classification）[14-16]。兩類分類法即是將一般的音訊訊號分為兩大類，通常為語音（speech）和非語音（non-speech），而多類分類法則是分更多更細的類別。E. Zwicker and H. Fastl[10]指出一般人類的聲音頻率集中在低頻，所以當一段音訊訊號之頻率

偏低時，則為語音，反之則為非語音。而 Saunders[11]則利用正規化零越率 (Zero Crossing Rate) 當作特徵即可初略地將語音和非語音分出。

Abu-El-Quran and Goubran [13]另外提出利用音調 (Pitch) 計算音調率 (Pitch Ratio) 的方法，將音訊分為語音與非語音。其中音調是在聲波圖上波形震動的頻率。而他們提出的方法描述大致如下：首先將一段音訊重疊 (Overlap) 分割成 N 個框 (Frame) ，並計算每個框之音調，如果計算出來的音調是在七十赫 (Hz) 到兩百八十赫 (Hz) 的範圍內，則屬於有標記的框，反之則為無標記的框。而音調率的計算方法為：音調率=NP/NF ，其中 NP 為有標記的框數，而 NF 為所有的框數。當一段音訊訊號之音調率大於等於零點三時則判定為語音類別，反之則為非語音類別。由於兩類分類法在實際應用中是不足夠的，所以最近越來越多關於多類分類法的研究。

L. Wyse and S. Smoliar [14]則是將一般的音訊訊號分類為音樂、語音和其他類別。有別於之前的兩類分類法，將音樂從音訊中分類出。D. Kimber and L. D. Wilcox [15]更進一步將音訊訊號分類為語音、靜音、笑聲和非語音，可說是將兩類分類法中的語音部份分類的更細。最後在 Lin and Chen[16]提出的方法中，則是將音訊分類為五大類，分別為：純語音、音樂、人清唱之歌聲、背景聲音為音樂的語音和背景為週遭環境噪音的語音。其所利用的方法為利用各類別在頻譜圖 (spectrogram) 上能量分布的方向及頻率高低的不同來直接擷取各類別的代表特徵。此方法雖然可以有效將音訊分類為純

語音、音樂、人清唱之歌聲、背景聲音為音樂的語音和背景為週遭環境噪音的語音五大類，但是卻無法將純語音的部分在予以細分。

因為藉由人所發出的聲音，可以直接或間接的傳達此人想表示之訊息，所以我們將利用此特性來發展一套基於語音偵測的看護系統，即系統會藉由被看護者發出的聲音來判斷此被看護者的健康狀況是否有疑慮，同時也確保個人之隱私。

在本論文中，我們提出的方法則屬於多類分類法，主要是將語音部分再細分為咳嗽聲、呻吟聲、喘息聲、求救聲和其它不屬於前四類之聲音，以應用於看護系統。在我們的方法中將危急聲音定義為劇烈咳嗽聲、呻吟聲、喘息聲和求救聲四類，而不屬於這四類的聲音則判定為正常聲音。為偵測上述的四種危急聲音，我們從聲波圖上擷取區域內聲音區段的數目，簡稱分割區段，還有波形持續時間、平均音量、正規化零越率和聲音區段彼此之間的相關性這五種具區別率的特徵來當作我們要偵測之依據。其中分割段數即為在五秒音訊檔中被分割出之有聲區段數，波形持續時間則為所分割出有聲區段之時間長短，平均音量大小為五秒音訊檔中，所有被分割出有聲區段之平均音量大小，正規化零越率則為所有有聲區段在聲波圖上單位時間波型震盪通過零點之次數，而相關性則是比對分割出來有聲區段之波型和我們定義的關鍵字波形區段之相似程度。在實驗結果顯示，我們的方法對這四種危急聲音的偵測率分別可以高達 94%~97%，在假警報方面只

有 0.08% 的錯誤率。

1.3 論文架構

在第一章中，我們描述本論文的研究動機、過去相關的研究成果和我們的方法簡介。接下在第二章，我們將定義四種危急聲音(求救、呻吟、咳嗽、喘息)的特徵向量以及他們的擷取方法。而在第三章中，我們將詳細介紹如何利用這些特徵來進行聲音的分類。最後在第四章和第五章則是我們的實驗結果與結論。



第二章

特徵的擷取及選擇

2.1 前言

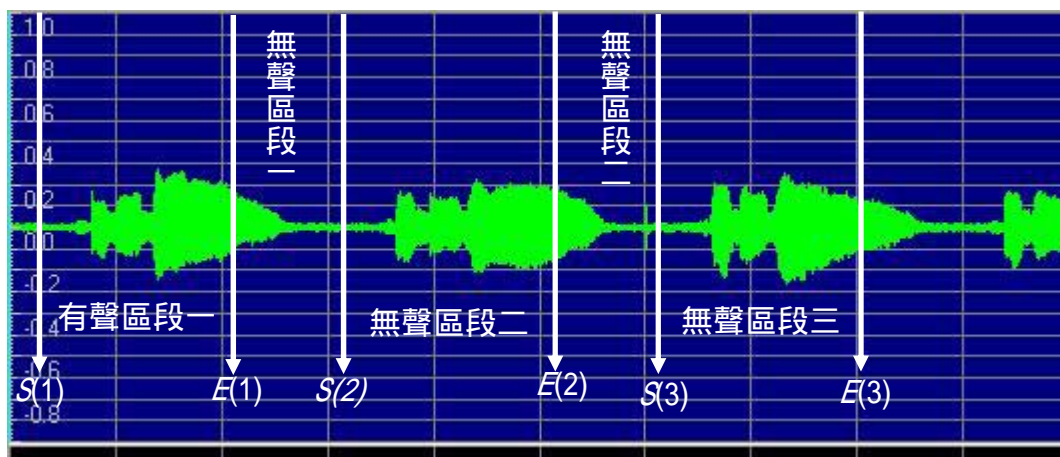
如前所言我們定義四種危急聲音為音訊關鍵字，分別為：求救聲、呻吟聲、劇烈咳嗽聲和喘息聲。為了尋找具有區別率的特徵，我們預先錄製這四種危急聲音作為樣本資料庫，其中每種類別皆錄製 23 個樣本，共 92 個。每個樣本長度皆為 5 秒，取樣頻率為 16 kHz，解析度為 16 bits 的單聲道數位音訊訊號。完成樣本資料後，我們接著從音訊的聲波圖上分割出人聲之波形區段並尋找可能之特徵再加以分析統計，最後選擇出適當特徵。在我們所錄製危急聲音樣本資料中，各種危急聲音的聲波圖依次列於圖一至圖四。

藉由觀察圖一至圖四的四種聲波圖，我們初步選擇五種可能可以做為特徵之資訊，分別為：分割段數、平均音量、波形持續時間、相關性以及正規化零越率，並對樣本資料庫內所有樣本進行以上五種可能特徵的數據統計，並作為之後選擇特徵之依據。

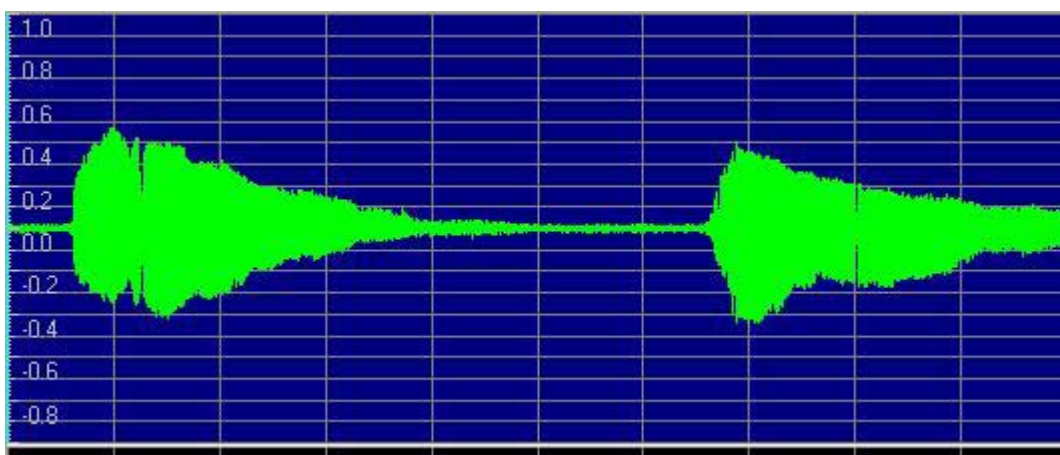
為方便接下來的描述，我們先定義聲波圖上的橫座標即時間方向用 x 代表，而縱座標為對應到 x 軸的音量大小用 $f(x)$ 表示。

聲波圖上每個分割出來的連續波形部分，稱為一個區段，有聲音的區段稱為有聲區段，無聲音的區段則為無聲區段。每個有聲區段的起點和終點

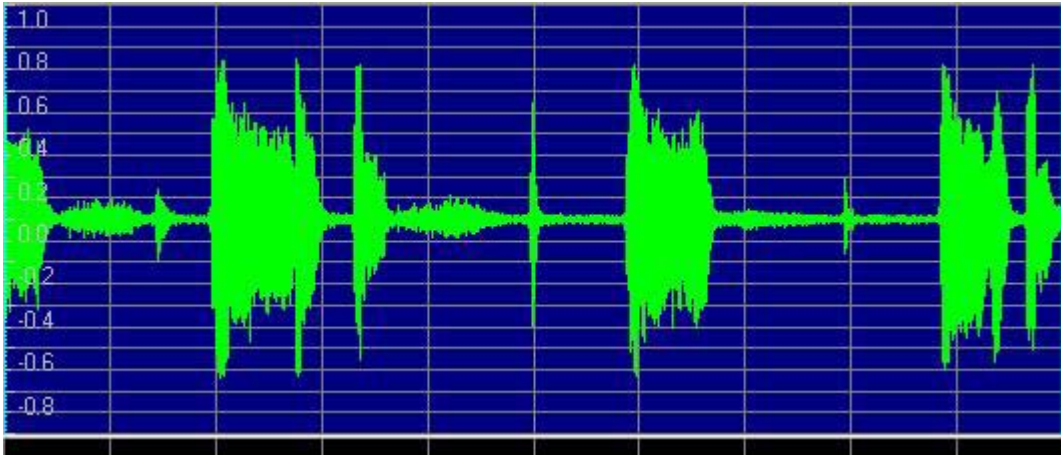
分別用 $S(n)$ 和 $E(n)$ 表示(請參考圖一), 其中 n 表第 n 個分割出來的區段。



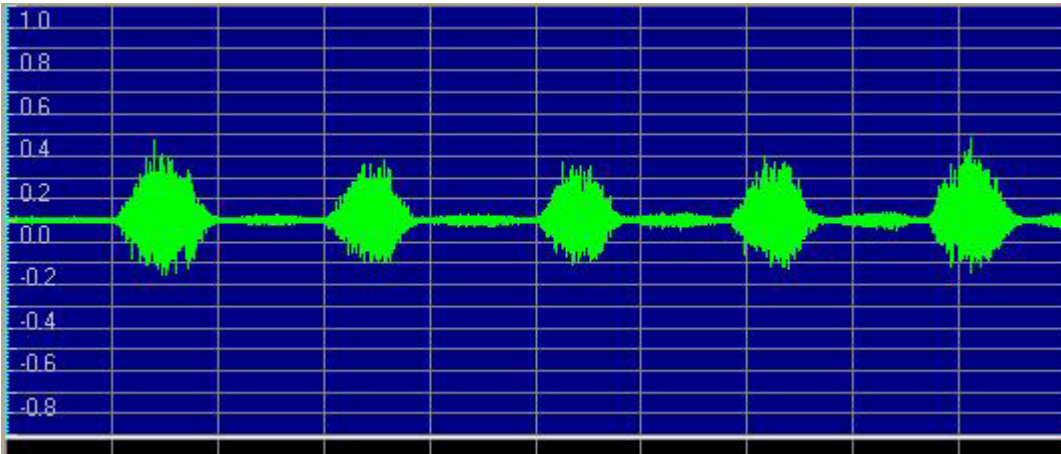
圖一：求救聲聲波圖及分割區段



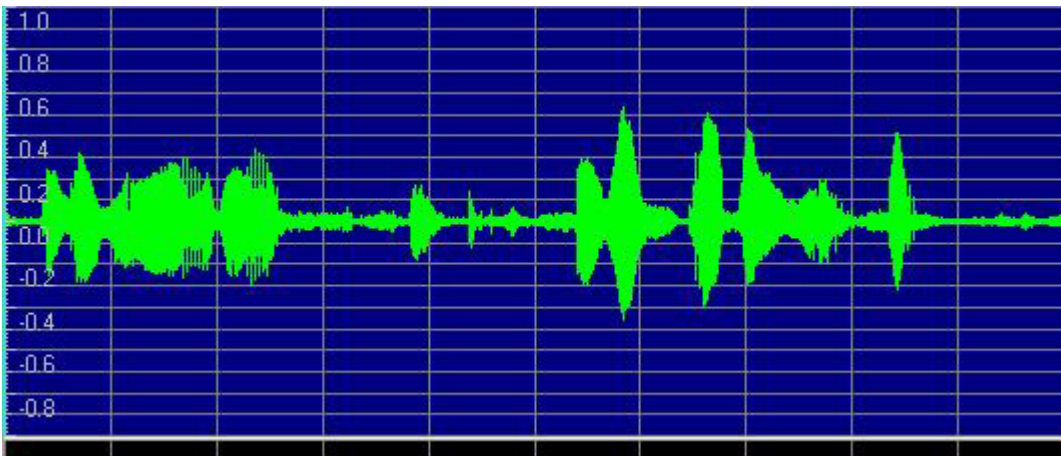
圖二：呻吟聲聲波圖



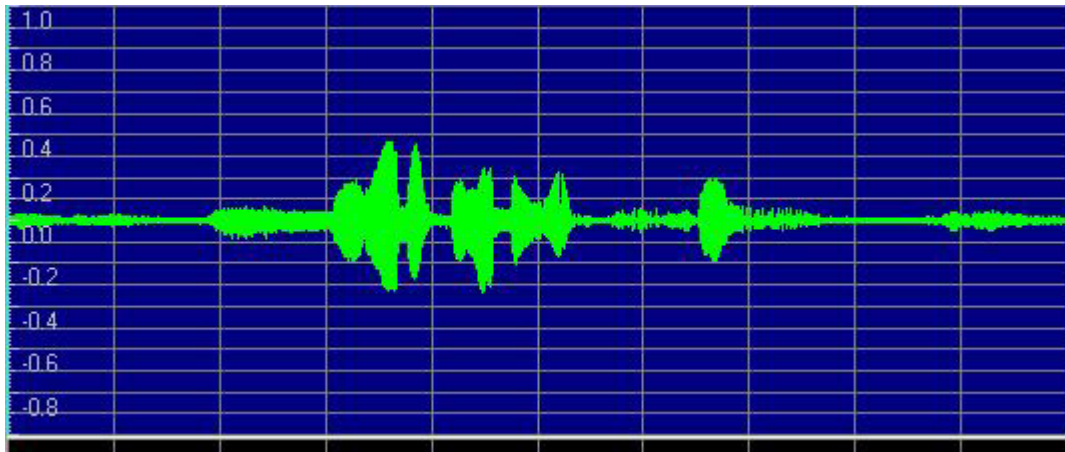
圖三：咳嗽聲聲波圖



圖四：喘息聲聲波圖



圖五：人對話聲聲波圖



圖六：電視聲聲波圖

2.2 分割

圖一至圖四的聲波圖明顯顯示出，被看護者發出之危急聲音除劇烈咳嗽聲外，通常具有週期性，這和一般在被看護者週遭常接收到的聲音如：人的對話聲(如圖五)、電視聲音(如圖六)等有明顯的區別，而通常人的對話聲與電視聲音是沒有週期性。所以為了能有效偵測聲音的週期性，我們首先進行前處理（Preprocessing）以及分割（Segmentation），藉此分離出各個連續音訊區間，即有聲區段，並對單一有聲區段進行分析。音訊的前處理包含平滑化（Smoothing）與去雜訊（Noise Removing）兩個動作，主要目的是為了方便接下來的分割動作。

首先對聲波圖做平滑化。詳細做法如下：對聲波圖上的每一個取樣點（Sample） x ，往其左右各 16 個點張開一個視窗（window）。然後將視窗內

的所有音量值($f(x)$)，取絕對值並計算其平均，將之作為平滑化後 $f(x)$ 值之大小。平滑化後的 $f(x)$ 值之正負號則與平滑化前相同。如 2-1 所示：

$$\overline{f(x)} = \text{sign}(f(x)) \times \frac{1}{33} \sum_{i=-16}^{16} |f(x+i)| \quad (2-1)$$

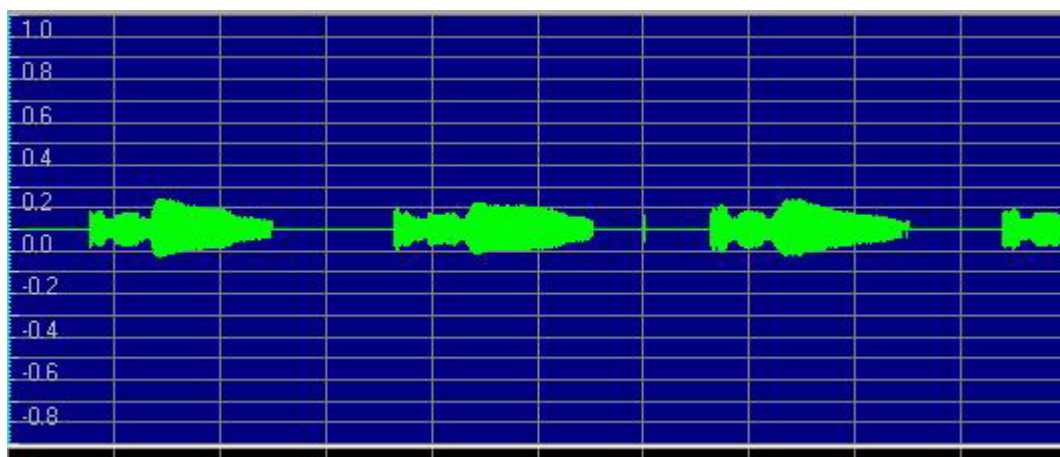
其中 $f(x)$ 為原始音量訊號， $\overline{f(x)}$ 為平滑化後的音量訊號

接著對平滑化後的訊號 $\overline{f(x)}$ 進行去雜訊。首先找出平滑化後整體五秒訊號中最大的絕對值 $L = \max(|\overline{f(x)}|)$ ，接著對每一個 $\overline{f(x)}$ 進行測試。若 $\overline{f(x)} < 1500$ 或 $\overline{f(x)} < L/6$ ，則將其視為雜訊並去除(訊號設為零)，如式子 2-2 所示

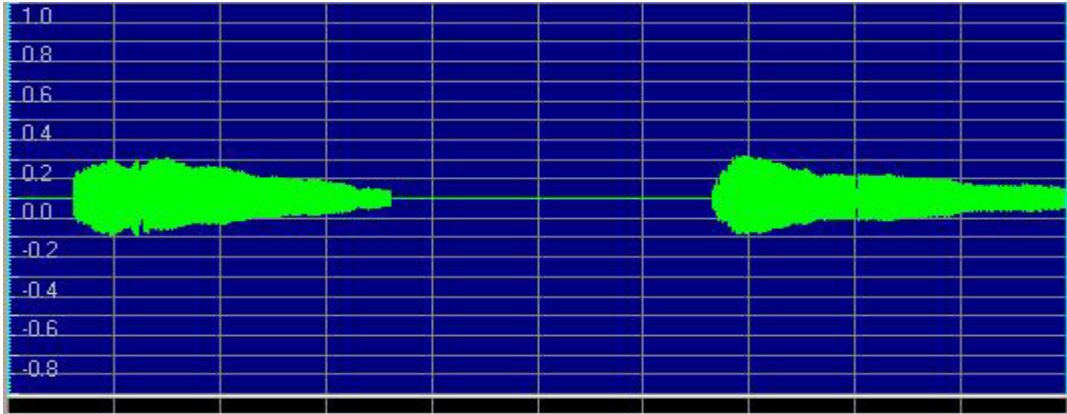
$$\overline{\overline{f(x)}} = \begin{cases} 0 & \text{if } |\overline{f(x)}| \leq 1500 \text{ or } |\overline{f(x)}| \leq (L/6) \\ \overline{f(x)} & \text{otherwise} \end{cases} \quad (2-2)$$

其中 $\overline{\overline{f(x)}}$ 為去雜訊後訊號

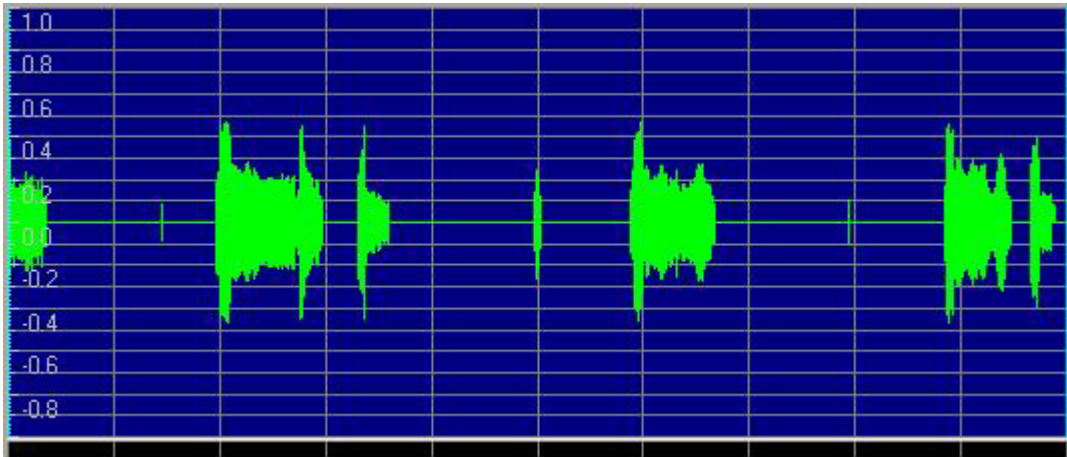
平滑化及去雜訊後可以突顯聲波圖中我們欲分割出之有聲波形區段，以便之後的分割程序。做完平滑化及去雜訊後之聲波圖如圖七至圖十。



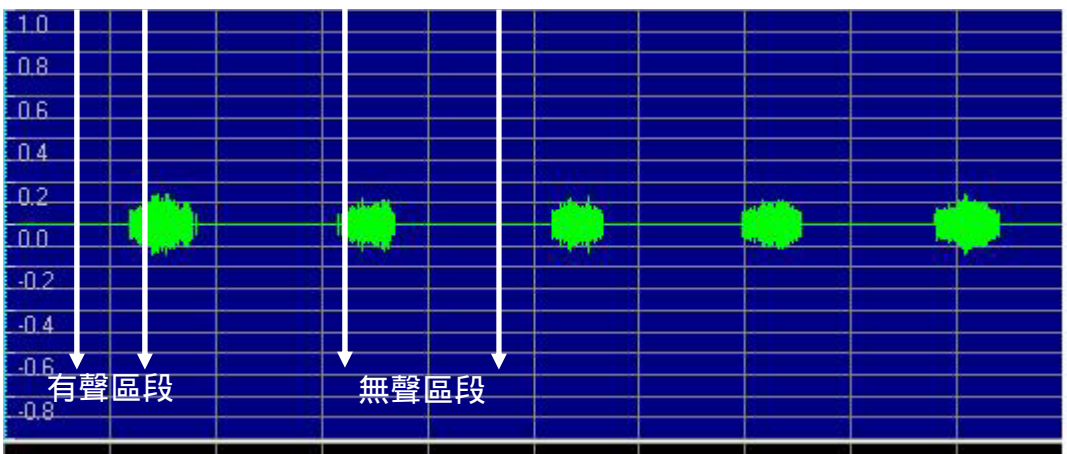
圖七:平滑化及去雜訊後之求救聲聲波圖



圖八：平滑化及去雜訊後之呻吟聲聲波圖



圖九：平滑化及去雜訊後之咳嗽聲聲波圖

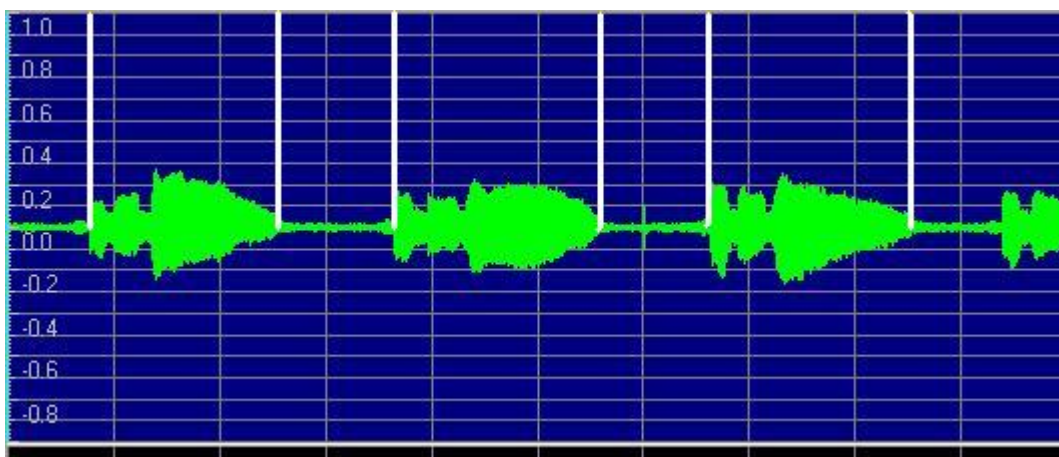


圖十：平滑化及去雜訊後之喘息聲聲波圖

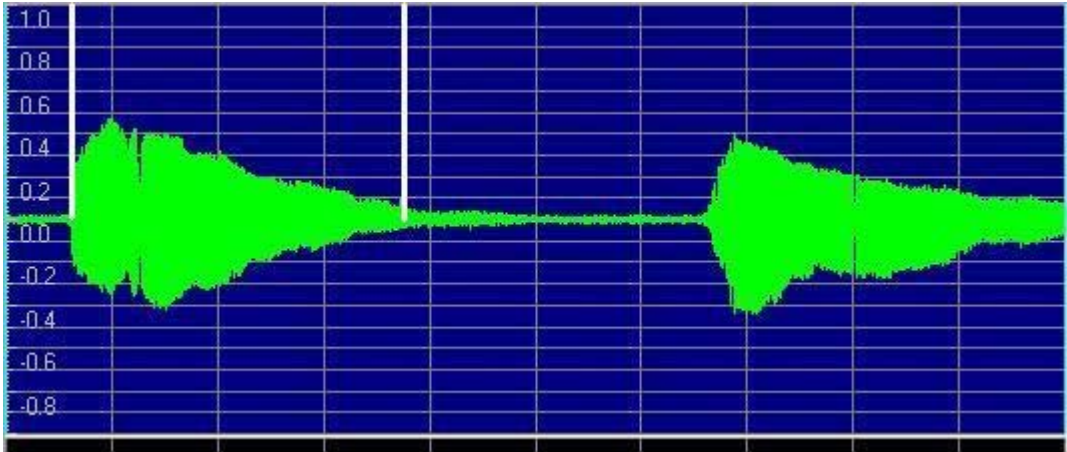
如圖七至圖十顯示，在平滑化及去雜訊後，原本的雜訊部分已被去除，在五秒音訊訊號中之有聲區段(參考圖十)和無聲區段(參考圖十)也清楚的分開。因為無聲區段為一段音量大小皆為零之值，所以我們利用此一特性將無聲區段給擷取出來，剩下的則為有聲區段部分，即為我們欲分割出之有聲波形區段。而這個有聲區段即是我們要分析的區段。

由於在一般情況下偶爾會接收到突發的短暫噪音，為消除突發噪音。我們假設有聲波形區段至少會持續 $1/16$ 秒，少於這個長度且聲音很大的區域將被視為突發噪音，而被刪除不處理，詳細做法如下：系統設定一個臨界值 (threshold) 為 1000 個取樣點，檢查每一個分割出來有聲區段的時間長度。若此時間長度小於該臨界值，則判定為噪音並予以去除。

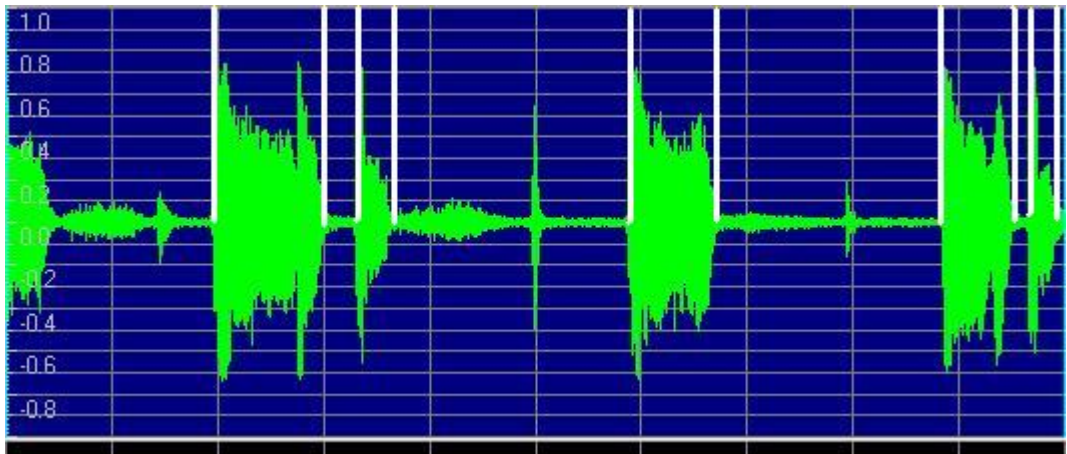
圖十一至圖十四為分割後之音訊資料，每一個有聲區段之起始點及終點以白線標示。



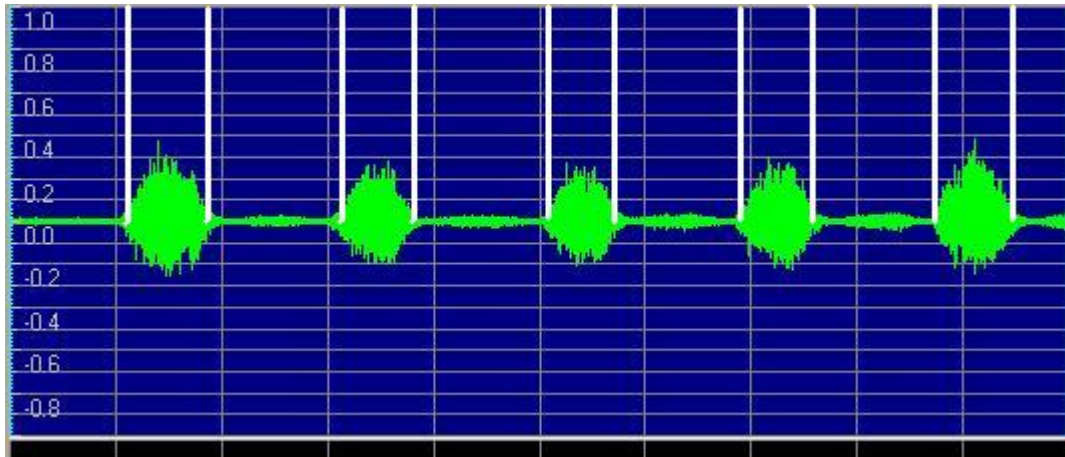
圖十一：求救聲及其分割區段



圖十二：呻吟聲及其分割區段



圖十三：咳嗽聲及其分割區段



圖十四：喘息聲及其分割區段

2.3 特徵分析

觀察求救、呻吟、咳嗽和喘息四種危急聲音分割後之聲波圖，可以發現求救、呻吟和喘息這三種類別其分割出來的每個有聲區段非常相似且具周期性，即有聲區段樣式（pattern）很固定。而在咳嗽聲之聲波圖中，發現每個有聲區段的平均音量大小相對於其他類別偏高。加上一般音訊分類常用的正規化零越率，我們初步歸納出五種可能的特徵分別定義如下：

- 分割段數(N)：在五秒之音訊中，有聲區段的數目。
- 平均音量(M)：將分割出來的有聲區段，取所有絕對值的平均。算法如式子 2-3：

$$M = \frac{1}{\sum_{n=1}^N (E(n) - S(n))} \sum_{n=1}^N \sum_{x=S(n)}^{E(n)} |f(x)| \quad (2-3)$$

其中 N 為五秒訊號中之分割段數， $S(n)$ 和 $E(n)$ 分別為第 n 個有聲區段的

起點和終點，單位為取樣點， $f(x)$ 為音量大小。

- 持續時間(D_n)：對每個有聲區段，算它持續的時間。如方程式 2-4 所示：

$$D_n = E(n) - S(n) \quad (2-4)$$

其中 n 為第 n 個有聲區段，單位為取樣點。

- 相關性(C)：將分割出來的有聲區段波形和各種聲音的代表區段波形做相關性計算。至於如何取得各種聲音的代表區段波形及相關性計算，之後會詳細介紹。
- 正規化零越率(ZCR)：對聲波圖中分割出來的所有有聲區段，計算其上下震動的波形穿越過零點的次數再除以所有有聲區段的總持續時間。如

方程式 2-5 所示：

$$ZCR = \frac{ZCRN}{\sum_{n=1}^N (E(n) - S(n))} \quad (2-5)$$



其中 $ZCRN$ 為五秒之音訊檔中所分割出來的有聲區段其波形上下震動穿越過零點之次數。

2.4 特徵統計與選擇

在我們所建的資料庫中對每個聲音檔，擷取上述所列各個特徵，並做統計之後，我們可以得到表一。

表一 特徵分析統計表

| | 分割段數 N (個) | 平均音量 M | 持續時間 D_n (取樣點) | 正規化零越率 ZCR |
|-----|---------------|-----------|---------------------|---------------|
| 求救聲 | 3~4 | 1000~1250 | 11800~14900 | 0.08~0.12 |
| 呻吟聲 | 1~2 | 800~2200 | 17800~29200 | 0.11~0.115 |
| 咳嗽聲 | 3~11 | 5500~7500 | 1200~18400 | 0.14~0.19 |
| 喘息聲 | 5~6 | 400~1000 | 3500~6500 | 0.21~0.24 |

由表一的統計資料，我們可以觀察到咳嗽聲在其平均音量和正規化零越率明顯的與其他類別不同，因此我們利用這兩個資訊當作其代表特徵。對於求救聲、呻吟聲和喘息聲這三類音訊，其分割段數相較穩定，且聲音較具規律性而具區別率，所以我們取分割段數和持續時間作為其代表特徵。至於對各類別代表特徵如何訂定其臨界值（Threshold）和臨界範圍（Threshold Range），及如何決策各類別的特徵，詳細說明如表二至表六所示：

表二 咳嗽聲的平均音量之平均、變異數和特徵臨界值

| | 平均音量 | |
|---------------------|----------------------------------|--|
| | 平均 $Mean_M$ | 變異數 $Variance_M$ |
| 咳嗽聲 | $\frac{\sum_{f=1}^{23} M_f}{23}$ | $\left(\frac{\sum_{f=1}^{23} (M_f - Mean_M)^2}{23} \right)^{1/2}$ |
| 臨界值 $Threshld_M$ | $Mean_M - Variance_M$ | |

表二為針對咳嗽聲類別計算其平均音量之平均值和變異數，且制定其臨界值。表中 f 為資料庫中咳嗽聲的檔案索引， M_f 為第 f 個檔案的平均音量。 $Threshld_M$ 為咳嗽聲的平均音量之臨界值。日後當系統接收到的五秒音訊訊號之平均音量大於此臨界值，就判定為咳嗽的特徵之一。

表三 咳嗽聲的正規化零越率之平均、變異數和特徵臨界範圍

| | 正規化零越率 | |
|-------|--|--|
| | 平均 $Mean_{ZCR}$ | 變異數 $Variance_{ZCR}$ |
| 咳嗽聲 | $\frac{\sum_{f=1}^{23} ZCR_f}{23}$ | $\left(\frac{\sum_{f=1}^{23} (ZCR_f - Mean_{ZCR})^2}{23} \right)^{1/2}$ |
| 臨界值範圍 | $Mean_{ZCR} - Variance_{ZCR} \leq TR_{ZCR} \leq Mean_{ZCR} + Variance_{ZCR}$ | |

表三為針對咳嗽聲類別計算其正規化零越率之平均值和變異數，且制定其臨界範圍。表中 f 為資料庫中咳嗽聲的檔案索引， ZCR_f 為第 f 個檔案的正規化零越率。 TR_{ZCR} 為咳嗽聲的正規化零越率之臨界範圍。日後當系統接收到的五秒音訊訊號之正規化零越率在此臨界範圍內，就判定為咳嗽的特徵之一。

表四 求救聲的持續時間之平均、變異數和特徵臨界範圍

| | 持續時間 | |
|------|---|--|
| | 平均 $Mean_{HD}$ | 變異數 $Variance_{HD}$ |
| 求救聲 | $\frac{\sum_{f=1}^{DN} HD_f}{DN}$ | $\left(\frac{\sum_{f=1}^{DN} (HD_f - Mean_{HD})^2}{DN} \right)^{1/2}$ |
| 臨界範圍 | $Mean_{HD} - Variance_{HD} \leq TR_{HD} \leq Mean_{HD} + Variance_{HD}$ | |

表四為針對求救聲類別計算其所有有聲區段持續時間之平均值和變異數，且制定其臨界範圍。表中 f 為資料庫中所有求救聲的有聲區段索引， HD_f 為第 f 個有聲區段的持續時間， DN 為所有有聲區段數。 TR_{HD} 為求救聲的持續時間之臨界範圍。日後當系統接收到的五秒音訊訊號中，有一個有聲波形區段之持續時間在此臨界範圍內，就判定此五秒音訊訊號為求救的特徵之一。

表五 呻吟聲的持續時間之平均、變異數和特徵臨界範圍

| | 持續時間 | |
|------|---|--|
| | 平均 $Mean_{GD}$ | 變異數 $Variance_{GD}$ |
| 呻吟聲 | $\frac{\sum_{f=1}^{DN} GD_f}{DN}$ | $\left(\frac{\sum_{f=1}^{DN} (GD_f - Mean_{GD})^2}{DN} \right)^{1/2}$ |
| 臨界範圍 | $Mean_{GD} - Variance_{GD} \leq TR_{GD} \leq Mean_{GD} + Variance_{GD}$ | |

表五為針對呻吟聲類別計算其所有有聲區段持續時間之平均值和變異數，且制定其臨界範圍。表中 f 為資料庫中所有呻吟聲的有聲區段索引， GD_f 為第 f 個有聲區段的持續時間， DN 為所有有聲區段數。 TR_{GD} 為呻吟聲的持續時間之臨界範圍。日後當系統接收到的五秒音訊訊號中，有一個有聲波形區段之持續時間在此臨界範圍內，就判定此五秒音訊訊號為呻吟的特徵之一。

表六 喘息聲的持續時間之平均、變異數和特徵臨界範圍

| | 持續時間 | |
|------|---|--|
| | 平均 $Mean_{WD}$ | 變異數 $Variance_{WD}$ |
| 喘息聲 | $\frac{\sum_{f=1}^{DN} WD_f}{DN}$ | $\left(\frac{\sum_{f=1}^{DN} (WD_f - Mean_{WD})^2}{DN} \right)^{1/2}$ |
| 臨界範圍 | $Mean_{WD} - Variance_{WD} \leq TR_{WD} \leq Mean_{WD} + Variance_{WD}$ | |

表六為針對喘息聲類別計算其所有有聲區段持續時間之平均值和變異數，且制定其臨界範圍。表中 f 為資料庫中所有喘息聲的有聲區段索引， WD_f 為第 f 個有聲區段的持續時間， DN 為所有有聲區段數。 TR_{WD} 為喘息聲的持續時間之臨界範圍。日後當系統接收到的五秒音訊訊號中，有一個有聲波形區段之持續時間在此臨界範圍內，就判定此五秒音訊訊號為喘息的特徵之一。

表七為從表二至表六的式子計算其臨界值範圍的實際數據：

表七 特徵臨界範圍統計表

| | 分割段數 N (個) | 平均音量 M | 持續時間 D_n (取樣點) | 正規化零越率 ZCR |
|-----|---------------|--------|---------------------|-------------|
| 求救聲 | 3~4 | | 12256~14892 | |
| 呻吟聲 | 1~2 | | 18863~28903 | |
| 咳嗽聲 | | 5498 | | 0.152~0.189 |
| 喘息聲 | 5~6 | | 3755~6292 | |

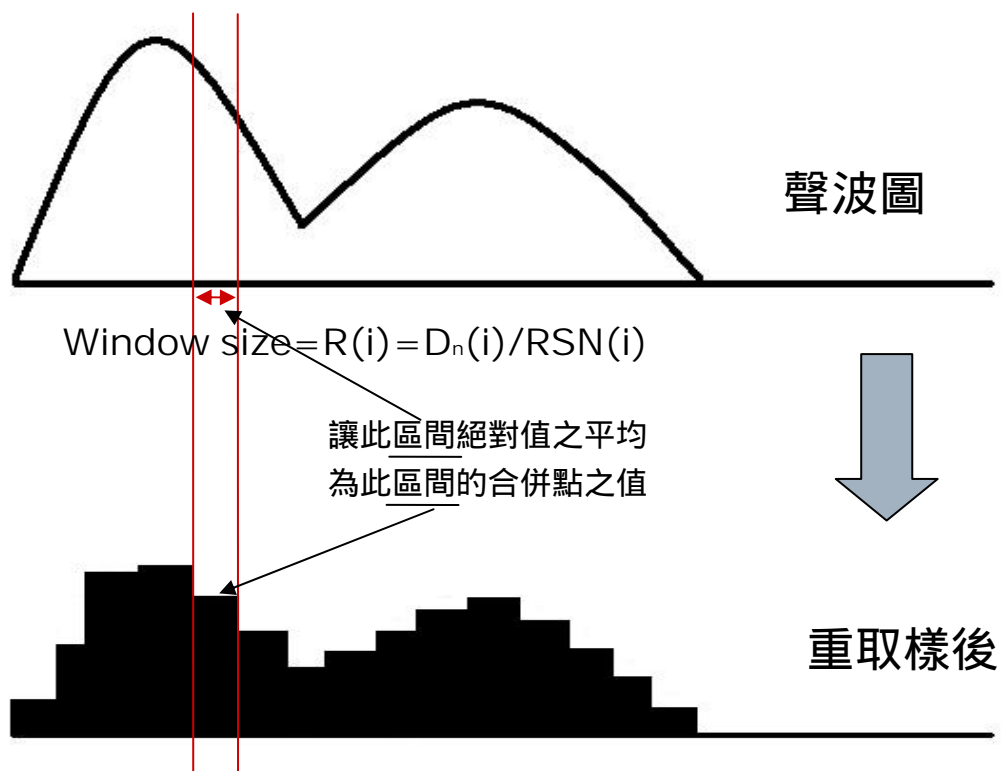
表七為從表一針對各類別的特徵計算其臨界值的結果。在分割段數方面和表一相同：求救聲範圍為 3~4 個，呻吟聲範圍為 1~2 個，喘息聲範圍為 5~6 個，平均音量部分：咳嗽聲之臨界值為 5498 單位量，有聲區段持續時間範圍部分：求救聲為 12256~14892 個取樣點，呻吟聲為 18863~28903 個取樣點，喘息聲為 3755~6292 個取樣點，正規化零越率範圍部分：咳嗽聲為 0.152~0.189。下一章我們將根據表七所得的數據進行對被看護者的聲音檢測。

第三章

異常聲音偵測法

3.1 波形相關性 (Waveform Correlation)

因為求救、呻吟和喘息三種波形有聲區段之樣式較固定，因此波形的相關性成為判定求救、呻吟和喘息三種危急訊號之重要依據。我們預先經由訓練 (Training) 的過程取出求救、呻吟和喘息三種標準有聲波形區段樣式，置於系統當中當作系統在執行時跟分割出來之有聲波形區段比對的標準。而訓練標準有聲波形區段樣式的方法如下，從我們資料庫中的每種類別分別去做，根據資料庫之統計，求救、呻吟和喘息三種聲波之有聲波形區段其平均持續時間分別為 13574 取樣點、23881 取樣點、5024 取樣點。由於在計算波形相關性時，每個有聲波形區段必須有相同的點數，且為了加速運算速度，我們將對每個有聲波形區段重新取樣，使得求救聲用 13 個取樣點 呻吟聲用 23 個取樣點 喘息聲用 5 個取樣點來表示。接下來設 $RSN(i)$ 為類別之取樣點數，而 $D_n(i)$ 為類別中聲波圖上第 n 個有聲區段之持續時間，令 $R(i) = \frac{D_n(i)}{RSN(i)}$ ，其中 $i=1,2,3$ 為類別(1 為求救，2 為呻吟，3 為喘息)則取樣方法為將聲波圖上每 $R(i)$ 個點合併成一點，且將此 $R(i)$ 個點之絕對值平均設定為合併點之值，如圖十五所示。



圖十五:重新取樣聲波圖

根據我們的觀察，發現在求救聲當中所分割出來的有聲區段波形，依其相似程度又可分為三個子類別，如圖十六至圖十八的第一個有聲波形區段所示。呻吟聲亦有相同的狀況，所以同樣分成三個子類別，如圖十九至圖二十一的第一個有聲波形區段，而喘息聲因為有聲波形區段最為穩定，所以不再分類，自己為一個子類別即可。針對每一個子類別，我們將取其平均有聲波形區段，作為系統執行時的比對標準有聲區段。平均有聲波形區段之詳細計算方法如下：

每個子類別中所有有聲波形區段: $X^1 X^2 \dots X^i \dots X^n$,

其中 n 為有聲波行區段數

$X^i = (x_1^i, x_2^i \dots x_j^i \dots x_{RSN(k)}^i)$, $1 \leq i \leq n$, x_j^i 為重新取樣點 , k 為類別

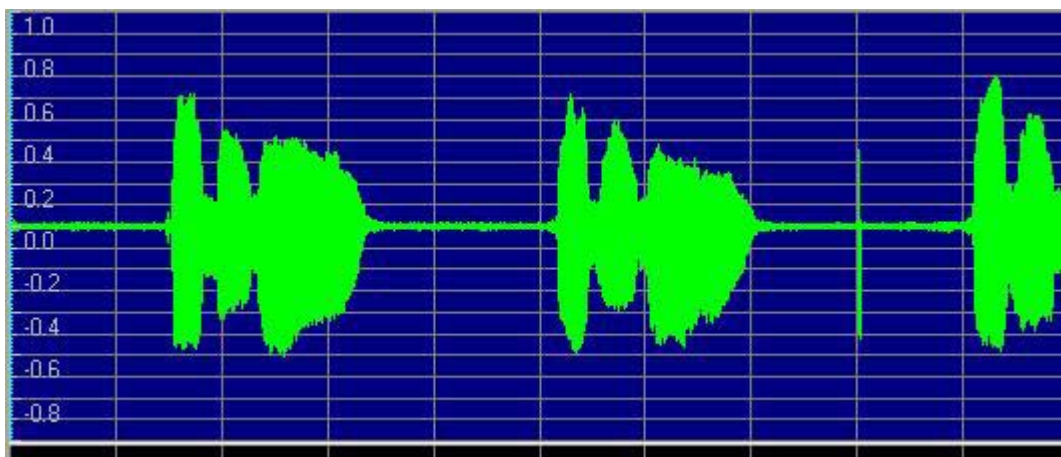
平均有聲波形區段 $X = (x_1, x_2 \dots x_l \dots x_{RSN(k)})$, $1 \leq l \leq RSN(k)$

其中 , $x_l = MEAN(x_l^1, x_l^2, \dots x_l^n)$

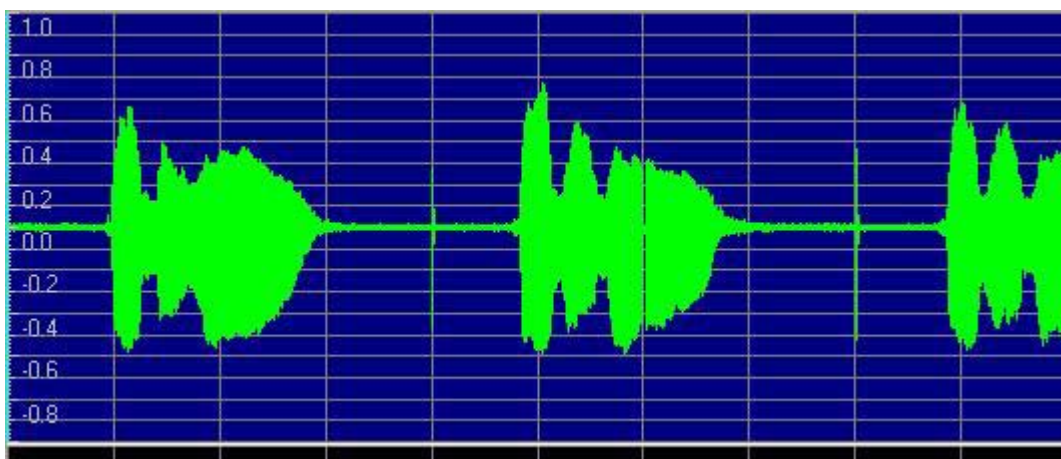
當程式執行時 , 系統會將分割出來之有聲波形區段分別重新取樣成 $RSN(i)$ 個取樣點 , 與系統中第 i 個類別的平均有聲波形區段做相關性計算 , 而相關性公式 3-1 方程式所表示 , 其中 X 為平均有聲波形區段 , Y 為目前系統分割出之有聲波形區段重新取樣後。

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \quad (3-1)$$

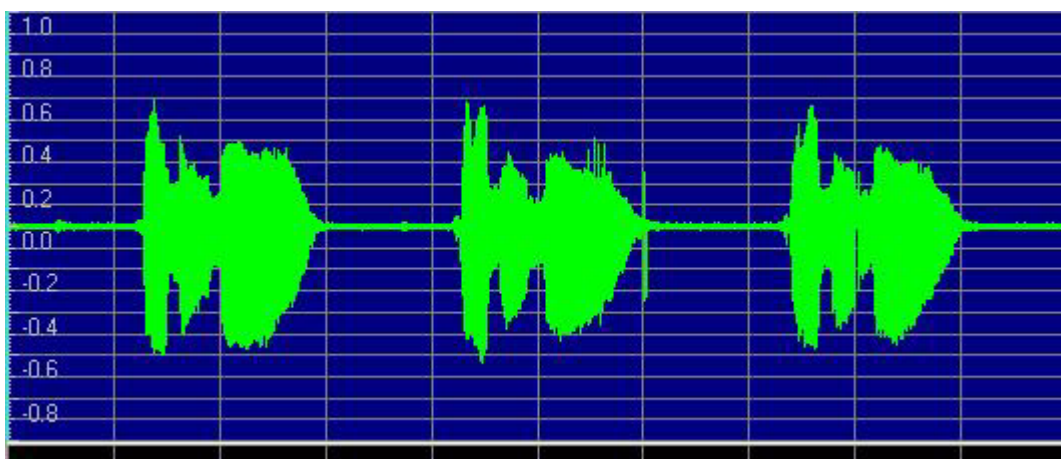
然後依值最高的判定此被分割出來之有聲波形區段屬於求救、呻吟和喘息的那個類別。



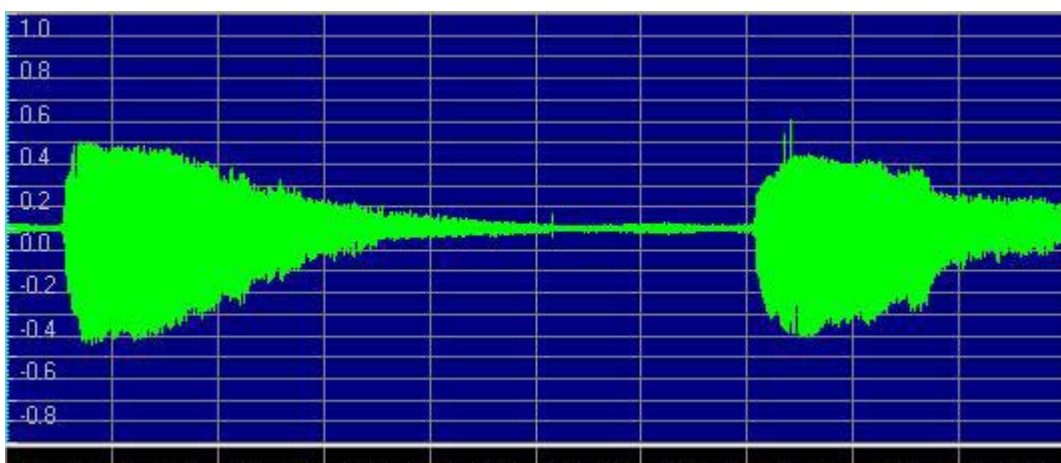
圖十六: 求救聲子類別一聲波圖



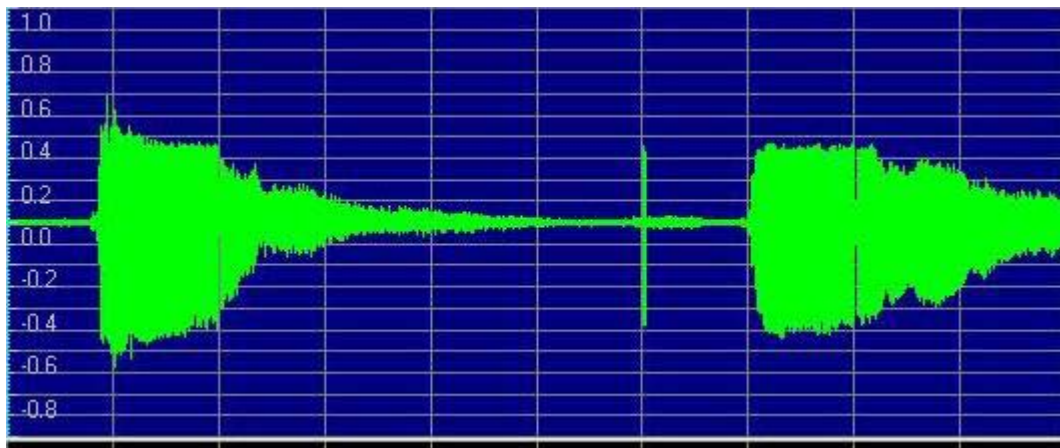
圖十七: 求救聲子類別二聲波圖



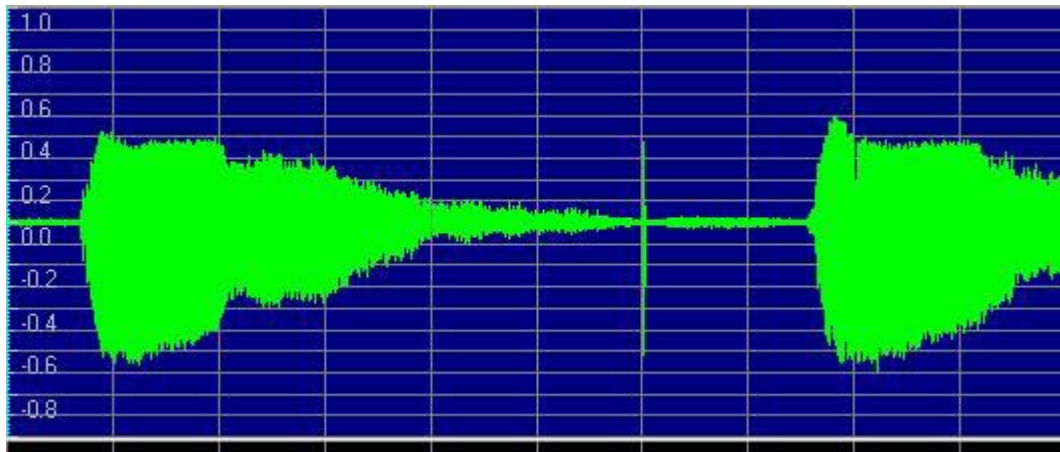
圖十八: 求救聲子類別三聲波圖



圖十九: 呻吟聲子類別一聲波圖



圖二十：呻吟聲子類別二聲波圖

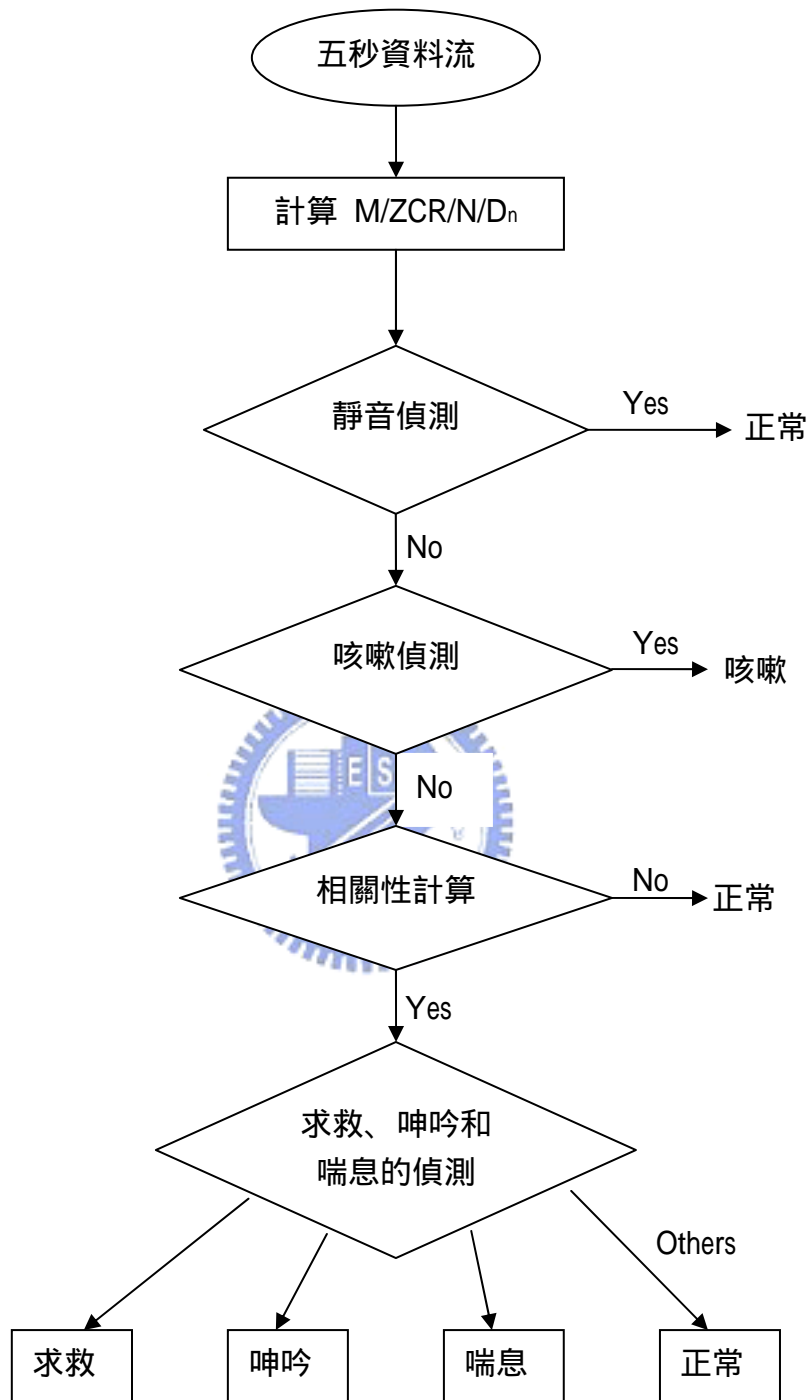


圖二十一：呻吟聲子類別三聲波圖

3.2 各種聲音之判定條件及系統流程

針對各種類別所選定之特徵，作為系統執行時對各種危急聲音之偵測依據，如果條件皆不符合表示並無危急情況視為正常。綜合前面聲波分析的結果(請參考表七)以及我們計算出的平均有聲波形區段跟資料庫裡面的各個類別做相關性計算後所得到之結果。我們可以為各種危急聲音類別決定

其代表的特徵和最佳的臨界值。整個系統的主要判斷流程將如下：



- 靜音偵測： $N=0$ ，即當一段五秒音訊訊號在處理過後，其分割段數為零，就是無聲靜音狀態。
- 咳嗽聲偵測： $(M > 5500)$ and $(0.15 < ZCR < 0.2)$ ，即當一段五秒音訊訊號

在處理過後，其平均音量跟正規化零越率在其臨界範圍內，則判定為咳嗽聲。

- 相關性計算：當一段五秒音訊訊號，所分割出來的有聲波形區段和系統內求救、呻吟、喘息之標準平均有聲波形區段計算相關性。當其中一個有聲波形區段和求救相關性大於 0.7 或和呻吟相關性大於 0.9 或和喘息相關性大於 0.8 則繼續下列的判斷條件。
- 求救聲偵測： $(N=3 \text{ or } 4)$ and $(D_n=12000\sim 15000)$ ，即當一段五秒音訊訊號在符合求救聲相關性計算條件後，其分割段數和其中一個有聲波形區段之波形持續時間在其臨界範圍內，則判定為求救聲。
- 呻吟聲偵測： $(N=1 \text{ or } 2)$ and $(D_n=17000\sim 29000)$ ，即當一段五秒音訊訊號在符合呻吟聲相關性計算條件後，其分割段數和其中一個有聲波形區段之波形持續時間在其臨界範圍內，則判定為呻吟聲。
- 喘息聲偵測： $(N=5 \text{ or } 6)$ and $(D_n=3500\sim 6500)$ ，即當一段五秒音訊訊號在符合喘息聲相關性計算條件後，其分割段數和其中一個有聲波形區段之波形持續時間在其臨界範圍內，則判定為喘息聲。

第四章

實驗結果

硬體設備：本系統使用 Intel Pentium 4 2.8 GHz，1G 記憶體之機器和無線藍牙麥克風和 USB 藍牙接收器，如圖二十二至圖二十四所示：



圖二十二：無線藍牙麥克風



圖二十三：USB 藍牙接收器



圖二十四：實際運用狀況

音訊格式：取樣頻率 16 KHz，解析度 16 bits，單聲道。

實驗環境和軟體平台：將圖二十二之無線藍牙麥克風掛於被看護者脖子上，以接收被看護者發出之聲音。搭配 Microsoft Direct Show 元件所撰寫的音訊擷取程式，讓系統實際執行 150 分鐘，每一次處理之資料長度為五秒且每隔一秒處理一次，所以有四秒的重疊，最後總共處理 8996 筆資料 ($150 \times 60 - 4$)，其中實驗過程包含一般人講話的聲音、電話鈴聲、電視聲音和週遭的噪音，同時求救聲、呻吟聲、咳嗽聲和喘息聲各發生 150 次，其結果和偵測率如表八：

表八 偵測結果與偵測率

| 類別 (數量) | 偵測結果(偵測率) | | | | |
|--------------|----------------|--------------|----------------|----------------|------------------|
| | 求救聲 | 呻吟聲 | 咳嗽聲 | 喘息聲 | 正常 |
| 求救聲 (150) | 145 (96.7%) | 0 (0%) | 0 (0%) | 0 (0%) | 5 (3.3%) |
| 呻吟聲 (150) | 0 (0%) | 144 (96%) | 0 (0%) | 0 (0%) | 6 (4%) |
| 咳嗽聲 (150) | 0 (0%) | 0 (0%) | 143 (95.3%) | 0 (0%) | 7 (4.7%) |
| 喘息聲 (150) | 0 (0%) | 0 (0%) | 0 (0%) | 142 (94.7%) | 8 (5.3%) |
| 正常 (8396) | 0 (0%) | 1 (0.01%) | 4 (0.05%) | 2 (0.02%) | 8389 (99.92%) |

在系統執行 150 分鐘當中，求救聲、呻吟聲、咳嗽聲和喘息聲各發生 150 次，並被正確辨識出 145、144、143、及 142 次。應被判定為沒事的狀況則有 8396 次，當中發生 7 次的假警報。這七次假警報當中，以咳嗽聲四次最多，其原因為麥克風收到我們音量很大之對話聲和期間之正規化零越率剛好符合咳嗽聲之條件，造成誤判。而呻吟和喘息則因為五秒音訊檔中所分割出有聲區段之波形與呻吟聲和喘息聲之標準有聲波形區段相似且波形持續時間和分割段數剛好符合之情況下，造成誤判。然而在異常聲音判斷不出來的情況，則都是因為在系統執行中分割出來的有聲波形區段和系統內的標準平均有聲波形區段之相關性值低於臨界值，故產生錯誤。

第五章

結論

本論文提出一個可即時 (Real-Time) 偵測異常聲音之居家看護系統。即在被看護者之身上掛一個輕巧的無線藍牙聲音接收器，隨時將被看護者週遭的聲音加以接收並經由我們的系統處理，然後由系統判定所處理過後之音訊訊號是否為危急聲音。是危急聲音時，則通知相關單位確認此被看護者之實際狀況，正常聲音時則不發出任何警訊。實驗結果顯示對求救聲、呻吟聲、劇烈咳嗽聲和喘息聲四種危急狀態之偵測準確率高達 94%~97% ，在假警報方面，僅只有 0.08% 的發生率。本系統較以往以視訊為基礎之看護系統更可以確保個人隱私，因為我們只需要在偵測到異常聲音時才需要去注意此被看護者。否則一般正常情況時，我們不會主動監看被看護者此時從事之行為，保護被看護者的隱私權。

未來的研究將繼續加強危急聲音偵測的準確率和進一步降低假警報率，甚至希望可以偵測到更多的聲音類別，以確切落實一套實用的居家看護系統可以百分之百的偵測出被看護者發出之危急聲音，同時希望不要太常發出假警報之訊息，以免增加相關醫護人員處理次數之頻率。

文獻參考

- [1] C. L. Huang and W. Y. Huang, "Sign Language Recognition using Model-Based Tracking and 3-D Hopfield Neural Network", *Machine Vision and Application*, vol.10, no. 5, 1998, pp. 292-301.
- [2] T. Y. Huang, M. L. Wu, T. Y. Chiang, Y. D. Lin and H. W. Chung, "Design and Implementation of a Fascimile Electrocardiographic System for Convenient Remote Monitoring", *National Taiwan University Engineer Magazine*, vol.88, 2003, pp. 43-50.
- [3] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search, and retrieval of audio", *IEEE Multimedia*, vol. 3, no. 3, Fall 1996, pp. 27-36.
- [4] L. Guojun and T. Hankinson, "A technique towards automatic audio classification and retrieval", in *Proc. Int. Conf. Signal Processing*, vol. 2, 1998, pp. 1142-1145.
- [5] E. Scherier and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator", in *Proc. Int. Conf. Acoustics, Speech, Signal Processing'97*, Munich, Germany, April 1997, pp. 1331-1334.
- [6] S. Rossignol, X. Rodet, and J. Soumagne et. al, "Feature extraction and temporal segmentation of acoustic signals", in *Proc. ICMC*, Ann Arbor, Michigan, 1998, pp. 199-202.
- [7] D. Li, I. K. Sethi, N. Dimitrova, and T. McGee, "Classification of general audio data for content-based retrieval", *Pattern Recognition Letters*, vol. 22, no. 5, April 2001, pp. 533-544.
- [8] T. Zhang and C.-C. J. Kuo, "Hierarchical classification of audio data for archiving and retrieving", in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, 1999, pp. 3001-3004.
- [9] T. Zhang and C. C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification", *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, May 2001, pp. 441-457.

- [10] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Model* Springer, 1990
- [11] J. Saunders, "Real-time discrimination of broadcast speech/music", in *Proc. Int. Conf. Acoustics, Speech, Signal Processing'*, vol. 2, Atlanta, GA, May 1996, pp. 993-996.
- [12] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Model*. Springer, 1990
- [13] A. Abu-El-Quran, and R. Goubran, "Pitch-based feature extraction for audio classification", *Proc. of the 2nd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications*, 2003. Ottawa, Canada, September 2003, pp.43-47.
- [14] L. Wyse and S. Smoliar, "Toward content-based audio indexing and retrieval and a new speaker discrimination technique", in *Proc. ICJAI'95*, Singapore, December 1995.
- [15] D. Kimber and L. D. Wilcox, "Acoustic segmentation for audio browsers", in *Proc. Interface Conf.*, Sydney, Australia, July 1996.
- [16] R. S. Lin and L. H. Chen, "A new approach for classification of generic audio data", in *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 19, No. 1, 2005, pp.63-78.