

行政院國家科學委員會專題研究計畫 成果報告

睡眠研究中鼾聲之聲學分析 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 95-2221-E-004-012-
執行期間：95年08月01日至96年07月31日
執行單位：國立政治大學資訊科學系

計畫主持人：廖文宏
共同主持人：楊建銘
計畫參與人員：碩士班研究生-兼任助理：林裕凱、紀煜豪

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中華民國 96 年 10 月 31 日

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

睡眠研究中鼾聲之聲學分析

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 95-2221-E-004-012-

執行期間：95年 8月 1日至96年 7月 31日

計畫主持人：廖文宏

共同主持人：楊建銘

計畫參與人員：林裕凱、紀煜豪

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份

執行單位：國立政治大學資訊科學系

中華民國 96年 10月 31日

睡眠研究中鼾聲之聲學分析

Acoustic Analysis of Snoring Sounds in Sleep Studies

計畫編號：NSC 95-2221-E-004-012

執行期限：95 年 8 月 1 日至 96 年 7 月 31 日

主持人：廖文宏

計畫參與人員：林裕凱、紀煜豪

執行機構及單位名稱：國立政治大學資訊科學系

一、中文摘要

本研究探討的內容主要關於人聲分類研究。在第一個層次我們將家庭環境中的聲音分為說話聲、人聲非說話聲以及環境音三種。為了達到此目標，我們尋找了許多特徵並找出合適的幾個做為判別的參數。

在夜間睡眠研究的部分，我們希望可以將整晚的睡眠資料分為鼾聲與非鼾聲兩部分，針對鼾聲的部分再深入去探討是否有呼吸中止症的情況產生。

最後，我們也結合視訊資料，自動判斷睡姿與翻身狀況，並探討睡眠時姿勢與鼾聲間的關係。

關鍵詞：音訊分類、鼾聲分析、睡眠呼吸中止症。

Abstract

In this research, we describe the classification of audio signals in a smart home environment and in all-night sleep studies. In a home environment, our objective is different from most audio scene analysis projects in that we are mainly concerned with the distinction of human and non-human sounds. Toward this goal, we identify appropriate features to be extracted from audio files and discuss the rationale behind choosing a particular feature.

In all-night sleep recording, we describe the classification of audio signals recorded in all-night sleep studies. Our objective is to separate the episodes into snoring sounds and non-snoring sounds. We perform further analysis of the extracted snoring sounds to

check if the testee has apnea.

Finally, we have also developed algorithms to automatically process video captured during sleep in an attempt to reveal the relationship between posture and the characteristics of snoring sounds.

Keywords: Audio classification, snoring analysis, OSA.

二、緣由與目的

睡眠是人類生命歷程中重要的一環，隨著現代工商社會生活型態的改變，睡眠問題也日益嚴重，據統計，台灣地區約有超過 20% 的人口有程度不一的睡眠障礙 [1,2]，也因此近來有諸多醫療或研究機構建立睡眠中心，進行睡眠品質的評估以及睡眠問題的診斷與治療。

在出現睡眠問題的案例中，有若干比例是為鼾聲所困擾，導致其本身以及同住者之睡眠品質受到直接的影響。鼾聲的成因主要有上呼吸道的阻塞、呼吸道某些地方的肌肉鬆弛、咽喉黏膜出現異常、氣道過於狹小等，而發生的時機大多是在睡覺的時候，此時患者往往需要增強他們呼吸的強度來獲得足夠的氧氣。當呼吸道完全阻塞時，此人已經不再單純只是打鼾，阻塞型睡眠呼吸中止症 (Obstructive Sleep Apnea Syndrome, OSAS) 的情況也會跟著產生(見圖一)，患者常常會因為這樣的症狀造成血氧濃度過低，以至於在睡眠的過程中需要耗費更多的力氣呼吸，此時睡眠往往會被打斷，患者在夜間會不斷的醒來目前的治療或紓解方案，進而使得睡眠品質低落，情況嚴重者白天的生活作息將嚴重的受到影響，在工作上無法集中精神，更

會對家人甚至鄰居產生甚多困擾。

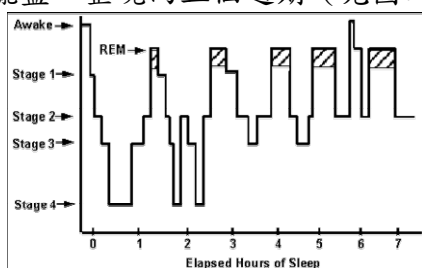


圖一 (a) 因呼吸道阻塞造成嚴重鼾聲



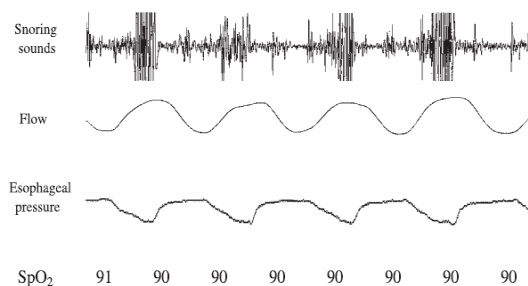
圖一 (b) 產生呼吸中止現象

以國內之醫療院所而言，對於鼾聲的診斷與治療，目前集中在耳鼻喉科、胸腔科與睡眠研究中心，而檢測的方式，則多採用多頻道睡眠生理記錄(polysomnography, PSG)。PSG 主要是根據腦波(electroencephalogram, EEG)、眼電圖(electroculogram, EOG)以及下顎的肌電圖(electromyogram, EMG)，將整夜的睡眠區分為階段一至階段四的睡眠，以及快速動眼睡眠(rapid eye movement sleep, REM)五個階段，其中階段三及階段四的睡眠又稱為慢波睡眠(slow wave sleep, SWS)或深度睡眠(deep sleep)，階段一至階段四的睡眠又稱為 NREM(non-rapid eye movement sleep)，整夜的睡眠會在 REM 及 NREM 睡眠間擺盪，整晚約五個週期（見圖二）。



圖二 睡眠週期示意圖

國內外針對鼾聲的分析，多集中醫學領域，也因此其研究的重點在於診斷、治療方法與療效之評估，而聲音也只是眾多可能指標中的一項(如圖三)。近年來由於電腦化自動分析的成本漸趨降低，因此使用數位化技術，包含線上監控 (online monitoring) 與離線分析 (off-line analysis) 的工具陸續推出。例如: SNAP[3]是採取離線分析的方式，主要是使用氣體流量與聲波來判別睡眠時的呼吸型態，受測者將儀器帶回家中做整晚紀錄後，再交回 SNAP 實驗室作後續的資料分析;國內元智大學老人福祉中心也發展了適用於居家環境的攜帶式鼾聲監控系統[4]，除了即時監控功能，亦保留整晚分析紀錄，但並未儲存原始聲音資料，整體的準確率(65%-79%)也有改善空間。



圖三 與呼吸型態有關的 PSG 測量[5]

基於以上緣由，本計劃針對鼾聲形成的聲學模型，鼾聲聲波之基本特性，鼾聲感測裝置的效能，整晚鼾聲型態之分析，鼾聲種類等議題，進行一全面性之探討，並回答下列問題:

- 是否能單純藉由聲音訊號，判別鼾聲之嚴重性與是否患有呼吸中止症?
- 收音之品質與判別的精確度之間關係為何?
- 家庭環境中收音效果是否足以提供有效之判別依據?

關於人類的發聲原理，多數研究集中在語音部分[6]，對於鼾聲的成因則僅限於病理的觀察與解釋，而欠缺物理模型。由

於鼾聲可能與上呼吸道的阻塞有關，與講話時的發聲方式有顯著差異，因此對現有的發聲模型做適度的修正，發展一個符合鼾聲發聲原理之模型，有助於了解鼾聲聲波之基本特性。

鼾聲的感測方式，大略可分為聲音、震動及呼吸氣流三大類，其中聲音屬非接觸式，而且儀器價格相對低廉，因此被多數醫療院所採用，也有較多研究成果發表。然而審視相關論文，多數侷限於時域(time domain)方面的分析，有的針對振幅的相對大小進行比較[7]，有的著重周期或間隔的規律性[8]，少數觸及頻域(frequency domain)的研究則僅統計主要頻率範圍[9]，或顯示手術前後頻譜圖的差異[10]，缺乏量化的分析，而且前述研究所取樣本數量均極有限，亦甚少完整利用整夜監控的資料。相較之下，近來在語音或音樂訊號的處理、分析與比對技術上有甚多進展，鼾聲的發聲原理雖然有所不同，但也屬音訊的一種，應可利用數位訊號處理的相關技術，擷取鼾聲之特徵，並比較不同受測者間鼾聲之異同。

在錄音設備與收音方式(如麥克風之規格、擺放位置)方面，目前並未制定通用之標準，因此本計劃也針對不同規格與等級之收音設備(如錄音筆、一般電腦用麥克風、微型麥克風、監視型麥克風、麥克風陣列等)，在不同環境(包含實驗室、一般住家臥室)、方位、距離之收音效果，進行系統化之測試，以確實評估收音品質對後續信號處理階段之影響，並藉此建立一套通用的收音準則(protocol)。同時我們也利用多軌同步錄音裝置，研究貼在喉嚨部位的震動感應器(vibration sensor)所得資訊與聲波訊號之關聯度，以評估震動感應器之必要性。

整晚睡眠的鼾聲分析，也是先前研究中較少觸及的，由於訊號量龐大，因此有賴於自動化處理，這個課題牽涉到以下步驟：

(1) 前處理:雜訊之消除、聲音訊號之偵測與自動切割(segmentation)。

- (2) 訊號過濾:必須將不屬於鼾聲的聲響過濾，以免干擾後續的分析。
- (3) 鼾聲的分群:必須能辨識同一個段落(episode)的鼾聲，加以分群，並尋求其共通特性。
- (4) 鼾聲型態之判別:必須能對整晚不同段落(episode)之鼾聲進行相似度之比較，推測鼾聲特性是否因人而異，或者與睡眠階段、睡姿、甚至其他生理訊號之關係較為密切。

目前睡眠研究雖使用近紅外線攝影的裝置，但極少對影像做後續的分析與處理，因此目前這方面的文獻不多。2005年7月，在加拿大醫學協會期刊(Canadian Medical Association Journal, CMAJ)發表的教學案例[11]中，運用了PSG以及攝影機等儀器記錄下患有睡眠呼吸疾病(Sleeping breathing disorders)病患睡眠的影像與各項生理測量數據。在分析報告中，諸多生理數據的分析都是有賴於PSG儀器的記錄，然而影像的分析，卻是透過人工調閱的方法觀察病患是否有試圖在睡眠中將氣道打開的舉動。假使能讓電腦自動判斷出病患是否有這些舉動，勢必能降低人工調閱的麻煩。除了PSG的記錄之外，尚須有聲音及影像的全程記錄，配合病患平常上床的時間，記錄病患整晚的睡眠情況直到病患起床為止。研究人員再依據PSG、聲音與影像的記錄判斷病患的病因，以找出治療之道。[12]

三、研究方法與結果

3.1 鼾聲的基本聲學分析

本期之研究主要在實驗室環境進行，在此系統發展階段選取若干正常睡眠的受試者以及若干有睡眠呼吸異常(sleep-related breathing disorders)及週期性肢體運動異常(periodic limb movement disorder)的病患至政大心理系睡眠研究室進行整晚的睡眠記錄。除標準PSG記錄以及臨床檢測所需的肢體活動、呼吸氣流、胸部及腹部的呼吸運動、以及由指尖記錄

血氧濃度外，並將執行音訊(分析打鼾及呼吸的型態)、視訊(近紅外線夜視影像分析活動狀態)、活動狀態(actigram)、體溫、心跳等非接觸式的行為及生理記錄。

PSG 資料將依據睡眠階段的國際標準，以每 30 秒為一個時段進行分析；並根據美國睡眠醫學學會的標準，分析睡眠當中擾醒(arousal)、呼吸中止或部分中止、以及肢體活動的發生。分析結果將與聲波信號相互參照，以驗證單純使用鼾聲型態是否確能評斷受測者打鼾之嚴重性。在聲波的分析與處理方面，比對正常睡眠的受試者以及若干有睡眠呼吸異常患者之呼吸型態與產生之聲響，以確立鼾聲的操作型定義；接著將以人工方式建立包括打鼾的次數、聲音的大小、每一段落的總和時間與時間間隔，段落之間的變異程度等數據庫，作為爾後評估自動化分析效能的參考標準。

PSG 試驗中，常見在受測者的喉嚨部位貼上震動感應器(見圖四)，以測量該處震動的幅度，然於此部位量得訊號與鼾聲之關連性為何，黏貼部位不同是否造成感應結果有明顯之差異，甚至是否可由此處量測之資料反推回阻塞的位置，均欠缺相關之研究，也是本計劃在此階段欲探討的重點。實際的作法將透過多軌同步錄音裝置，同時錄製不同種類之訊號源，並與 PSG 的其他資料交叉比對，據以評估使用震動訊號之準確度與必要性。



圖四 震動偵測器

有了基本的資料後，下一步我們將針對鼾聲聲學的基本特性進行探討，此一階段的研究將著重在兩個層面：第一個層面

是聲波訊號的前處理與特徵抽取，第二個層面是鼾聲發聲成音的建模(modeling)，而此兩項問題都可借助語音訊號處理的相關技術，例如 MPEG-7 中關於音訊的基本特徵向量、以及語音合成中的 source-filter 模型。

下一階段我們將評估錄音設備與收音方式對收音效果與鼾聲訊號後續分析的影響，希望根據實驗結果訂定標準的收音程序及建議的硬體規格。我們將先參考相關文獻中有關量測鼾聲的建議方式擺設收音裝置，並測試、比較在不同角度、距離、方向的收音效果，為了排除其他可能干擾因素(如不同受測者、同一受測者不同時間、不同錄音裝置等)，我們將採用同類型的高感度監視用麥克風，並使用多軌同步錄音卡進行數位錄音，必要之時將參照受測者姿勢與肢體活動情形，修正相關之數據。此外，由於鼾聲的頻率集中在低頻帶，因此我們也將就錄音時的不同取樣率(如 8KHz, 16KHz, 22.05KHz, 44.1KHz 等)進行紀錄與分析，以決定適當的取樣率。

收音效果除了與設備置放位置有關，與錄音裝置本身的規格也有密切關係，初步實驗發現具內建訊號擴大器的監視用麥克風收音效果最為理想，但是由於本計劃中鼾聲的頻率集中在較低頻處，過度靈敏的感應設備是否適得其反？或者在原始資料收集(raw data acquisition)步驟應該盡可能保留最多資訊，資料的價值應交由後續處理程序來判定？也因此我們將針對市面上常見收音裝置，包含錄音筆、一般電腦用麥克風、微型麥克風、監視型麥克風，甚至麥克風陣列(microphone array)等逐一進行測試，並比較其收音效果。

綜合以上步驟分析所得的結果，我們將制定一套鼾聲收音的標準程序，針對錄音裝置的數量、擺放位置、硬體規格、頻率響應、取樣率等作出明確的規範，以解決目前在鼾聲聲波分析研究中各家做法不同所可能造成的問題。

3.2 一般環境中之人聲分類

第一階段的工作主要在探討鼾聲的基本聲學特性，通常在實驗室中進行，因此有較理想的收音環境，或可以藉由人工方式過濾不相干的雜訊，但如果要發展適用於居家環境的自動化鼾聲辨識、監控與分析系統，則須考量實際收音時存在許多可能的干擾因素，例如磨牙(bruxism)、咳嗽聲、說話聲、汽車聲、記錄儀器發出的聲響、環境噪音等，因此首要的任務，就是必須能在各種可能出現的聲音類型中區辨出哪些是真正屬於鼾聲的段落，哪些是無關的雜訊，為了達到這個目標，除了利用第一階段研究中鼾聲的特徵外，還要針對所有可能出現的干擾音源進行訊號特性的分析，以找出足以分辨鼾聲與其他聲音的特徵向量。

我們採用的是階層式的分類法則(hierarchical classification)，先將聲音分為人聲(human sounds)與非人聲(non-human sounds)，再從人聲的類別中區分語音(speech)及非語音(non-speech)，最後在非語音的類別中再篩選出鼾聲的段落，而每個階層考慮使用不同組合的特徵向量，以簡化計算的複雜度。選用的特徵包含基頻、頻譜質心、音質等。在經過越零率等參數分離出人聲與非人聲之後，可利用頻譜質心及基頻等參數加以判斷是否為鼾聲，由於鼾聲多集中在低頻，且發生部位與說話時不同，因此頻譜質心與基頻可以幫助分類人聲中鼾聲與非鼾聲，以下就採用的特徵與其原因做簡短的說明。

基頻(Fundamental frequency):基頻代表著聲音發生共振時的最低頻率，亦可稱為 F_0 。由於基頻的計算並不複雜，也是 MPEG-7 的 audio descriptor 之一，因此可知，在某種程度上它對於聲音訊號具有一定的代表性，不同類型的聲音檔案其基頻必定也不相同。故本研究採用此特徵當作分類的參數之一，將其運用在區分人聲與非人聲這個層面。

越零率(Zero-crossing rate, ZCR):越零

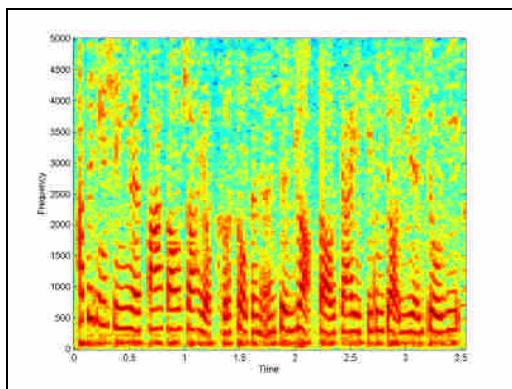
率代表著聲音訊號通過零點的次數，在大部分情況雜訊的越零率較非雜訊的越零率高，因此有部分的研究是利用這個特徵來切割出聲音訊號的起始與結束，本計劃則擬使用這個特徵來輔助區辨人聲與非人聲。

相關度(Autocorrelation):相關度主要是度量訊號的前後一致性，因此當訊號的前後一致性高時，相關度就會較高，反之則否。家庭環境中的風扇聲或吹風機的聲音，由於訊號的起伏變化極小，相較於人的說話聲是較為前後一致的。本計畫嘗試將人的說話聲與這類的聲音訊號做比較，其方法是將算出的相關度繪出得一連續曲線，當相關度高時訊號的變化不大，曲線也較為緩和，反之曲線的高低起伏便會極為明顯，因此可以預期的是，類似風扇聲這類訊號的高低起伏應該較人聲的起伏緩和，因此可以訂出一個公式，用以量測相關度曲線的起伏明顯程度(peak 的明顯程度)值越大表示起伏越明顯。

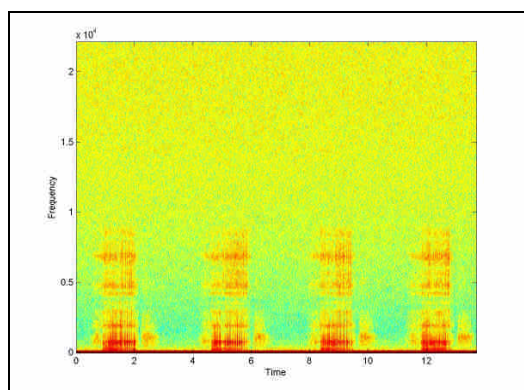
頻譜質心(Spectral centroid):訊號的頻譜質心是用來描述聲音檔案在頻譜圖(spectrogram)中頻率的質心，此特徵亦屬於 MPEG-7 的 audio descriptor 之一。在一般家庭環境中，人聲除了說話聲之外，往往還會夾雜著其他各種不同的聲音，例如：哭聲、笑聲、叫聲和鼾聲，這些聲音與說話聲的頻譜質心都不盡相同，由於笑聲與叫聲是屬於較高頻的聲音檔案，因此計算出來的質心應較說話聲的值為高，同理，鼾聲多集中在低頻的地方，故其計算出來的質心會較說話聲的值為低。

熵(Entropy):頻譜圖(spectrogram)是用來描繪頻率強度的圖形(如圖五(a))，其橫軸為時間，縱軸為頻率，深淺則代表強度，紅色代表的強度最強，依次為橙、黃、綠、藍、紫，強度代表能量的多寡。因此，不

同的聲音訊號在頻譜圖中所呈現的趨勢也就不一樣，相對於一些較為低頻的訊號(如圖五(b))，人的說話聲所呈現的頻率變化極為明顯，頻率的分布沒有集中在某範圍的現象。



圖五 (a) 為人類說話聲的頻譜圖



圖五 (b) 為鼾聲的頻譜圖

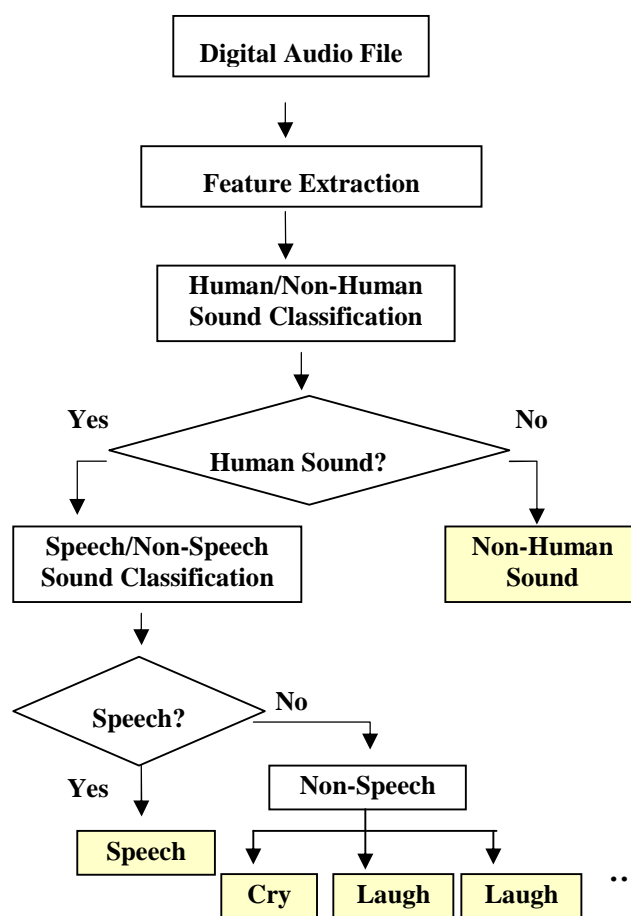
由於不同的聲音訊號在頻譜圖上的分布也不同，因此可以選用頻譜圖中訊號強弱的集中度作為其中一個參數，也就是頻譜圖的熵。熵是用來描述“亂度”，當資料的一致性較高時，則亂度較小，反之則亂度較高，以圖五而言，圖五(a)的亂度值較圖五(b)的值高，這是因為圖五(b)的能量多集中在低頻的地方，而圖五(a)的能量則幾乎在各個頻率範圍都有出現。亂度的計算方法為：

$$H(X) = -\sum_x P(x) \log_2 P(x)$$

音質(Formant frequency): 音質的高低與發聲部位到嘴唇這段距離有關。人聲中

除了說話聲之外，也有可能是笑聲或哭聲等其他聲音，而這些聲音所造成的共振不盡相同，因此在本計劃中則是利用音質的不同，針對人聲再加以分類，判別此人聲檔案是否為說話聲。其中，本研究選用第一個共振頻率作為代表，將人的說話聲與鼾聲的共振頻率加以比較。

有了前面的幾個特徵作為參數之後，便可以開始將聲音檔案分類，其類別為：人聲中的說話聲、人聲中的非說話聲、非人聲。其分類流程如圖六所示



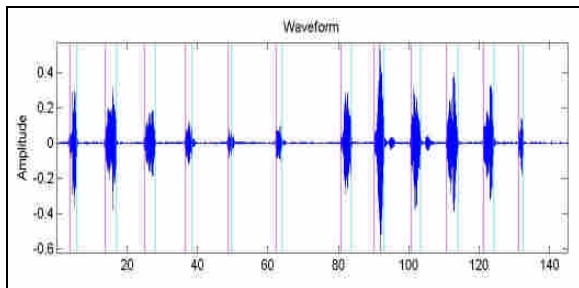
圖六 音訊分類的流程圖

3.3 睡眠時期的鼾聲監測

完成自動分類的工作後，我們就可以特別鎖定鼾聲的段落進行處理與分析，一般而言，在單純的實驗室中錄製的音訊干擾較少，大致上只有儀器聲、冷氣聲、咳嗽聲，以及一開始時的說話聲，如果是在

家庭環境中，可能就多了一些狗吠聲和車聲，所以處理上面較為複雜。在這一階段所牽涉的工作內容包含：聲音切割、鼾聲偵測與分群。

鼾聲切割與偵測：屬於前處理的部份，一開始必須將整晚的聲音檔案等分切割，針對這些檔案，在去雜訊的處理之後，必須找出有聲音的部份，也就是進行 segmentation 的步驟，所用到的方法，大致上是利用能量的變化，找出音訊的起點與終點，其結果如圖七所示。



圖七 切割過後的聲音檔案，紅線代表起點，綠線代表終點

經過上述的步驟之後，還要深入去判別這些音訊是否為鼾聲。在錄製整晚的聲音檔中，經過前處理去雜訊的步驟之後，留下來的為一開始入睡之前的說話聲，而說話聲跟鼾聲有一個很大的區別，就是說話聲是沒有週期性的，而鼾聲是較具有週期規律的，也就是說，在每個鼾聲之間至少要有一秒左右的間隔，並且每個切割出來的鼾聲長度不會超過五秒，這樣的規則可以濾掉大部分的說話聲，再加上鼾聲大多集中在低頻的部份(如圖五(b))，因此透過 spectrogram，我們可以很清楚的分出哪些段落是屬於鼾聲。

鼾聲的分群：這部分主要是想探討當有打鼾的情況產生時，這樣的打鼾是屬於一般型(simple snore)的或是阻塞型(OSA)的，一般型的打鼾其信號前後的一致性較高，而阻塞型的打鼾會伴隨呼吸中止的現

象，因此可能中間幾秒會完全沒有鼾聲，或者會有幾個鼾聲音量特別大，這是因為阻塞發生時，受試者需要用更大的力氣去獲得足夠的氧氣所造成，藉由這樣的分類方法便可以將鼾聲作簡單的分群。

分類的依據主要是照鼾聲信號中的差異度，先將頻譜圖在時域和頻域上做切割，時間軸切割成 n 等分、頻域切割成 m 等分，針對每等份算出頻譜圖中強度的總和，由於頻譜圖中的數值為複數型態，所以必須取絕對值以表示其大小。經過上述步驟之後，便可得到 $n*m$ 個數值(以 A_{ij} 表示)，將這些數值正規化，使得這 $n*m$ 個 A_{ij} 數值總和為 1，也就是 $A_{ij}=A_{ij}/\text{sum}(A)$ ，如此便可將頻譜圖轉為一個機率分布(p.d.f)，稱之為 $p(x)$ 。若要判斷兩個鼾聲訊號的相似度，則可透過上述的方法，得到兩個近似機率分布的函數 $p_1(x)$ 、 $p_2(x)$ ，計算差異的方法擬採用 K-L distance，定義如下：

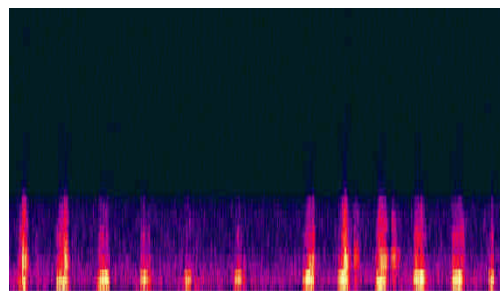
$$D_{KL}(p_1(x), p_2(x)) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx$$

由於 $D_{KL}(P_2(x), P_2(x))$ 與 $D_{KL}(P_2(x), P_1(x))$ 並非對稱，因此 $p_1(x)$ 與 $p_2(x)$ 的距離將取 $D_{KL}(p_1(x), p_2(x))$ 與 $D_{KL}(p_2(x), p_1(x))$ 兩者的平均來計算。

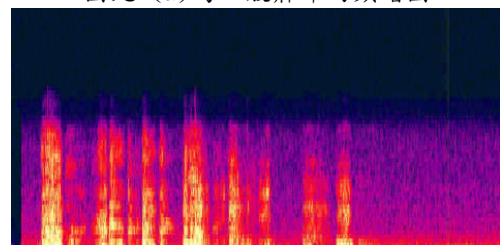
至於要如何定出鼾聲之間的差異，主要可以分為兩個層面：時域(time domain)和頻域(frequency domain)。在時域上的作法，主要是找出鼾聲是否有忽大忽小的情況產生，一般的鼾聲其聲音較為規律(如圖八(a))，大小較為一致，有呼吸中止症的病人其鼾聲大小會有所差異(如圖八(b))，因此在一開始的時候要先選定檔案中的第一個鼾聲片段(snoring episode)作為基準，分別算出後續的鼾聲與這個基準的音量相差多少，若差異變化很大，便可得知這個時段之內可能有呼吸中止症的產生。

在頻域上，由於呼吸中止症的病人阻塞情形較為嚴重，因此在鼾聲中會夾雜很多的氣音，氣音在頻率上的分佈較鼾聲廣泛，而單純的鼾聲卻只會集中在低頻的頻帶，且呼吸中止症的患者會有特定幾個鼾聲比較長，為要得到足夠的氧氣。因此在頻譜圖(spectrogram)中，一般打鼾者的亂度(entropy)會比較小，且每個鼾聲之間在頻譜圖上的差異度較低，也就是 K-L distance 會較小，相反的，呼吸中止症患者者在頻譜圖中的亂度較大，每個鼾聲片段之間的差異度也較大，以下兩張圖是兩種鼾聲在頻譜圖上的表現，圖九(a)是一般鼾聲的頻譜圖，圖九(b)是呼吸中止症患者的頻譜圖。

結合時域與頻域特性是必要的，由於受試者並非整晚固定睡姿，因此不同方向的側睡可能會導致不同的收音結果，所以在聲波的振幅上必定會有大小的差別，但如果是屬於同一段落的鼾聲，則在頻譜圖上的表現會較為相似，綜合時域與頻域特徵的結果，應可以較準確地判斷出受試者此時鼾聲上的差異是否源自於睡姿上的改變，或是確有呼吸中止現象產生。



圖九 (a)為一般鼾聲的頻譜圖



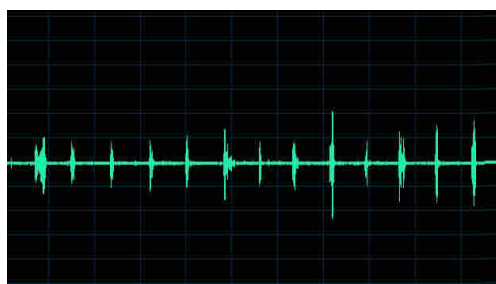
圖九 (b)為呼吸中止症患者的頻譜圖

以上的討論是針對音訊訊號，受試者在專業的實驗室或者醫療院所中受測，再配合上其他的生理訊號做對照，受試者配有鼾聲偵測器(平貼在喉嚨上偵測因打鼾發出的震動)，因此可以進一步去比較有鼾聲的同時也有振動的相關度。配合上腦波圖找出睡眠階段(sleep stage)，亦可以比較鼾聲的大小是否與睡眠階段有關，以及睡眠呼吸中止症患者在哪个睡眠階段比較容易造成阻塞，經過其他生理訊號的交叉比對，或許可以提供更多的資訊與研究方向。

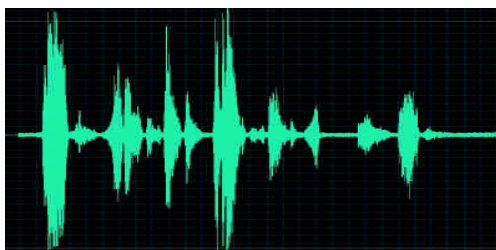
3.4 睡姿和鼾聲的關係

本研究的場景為睡眠環境，故影像中只包含單一物件——人體。此外，影像的背景固定為床的範圍，且翻身範圍也有限，只侷限於床的範圍。實驗中，由於受試者若在中途起來上廁所，會導致拍攝環境的改變，例如：場景中的人物突然增多、減少以及光線增多導致紅外線攝影機轉成彩色模式等情況。因此我們將區分出因外界因素影響的範圍，排除此異常時段、不與 PSG 作對照，以確保資料分析的正確性。

考慮受測者的隱私，以及睡眠實驗室空間上的限制，故無法將攝影機裝置在天花板上(此實驗低畫質的拍攝-640*480)，以理想的角度拍攝實驗場景。在經過多種角度的試驗後，我們選出放置攝影機最適



圖八 (a)為單純的打鼾情況，也就是所謂的 simple snoring

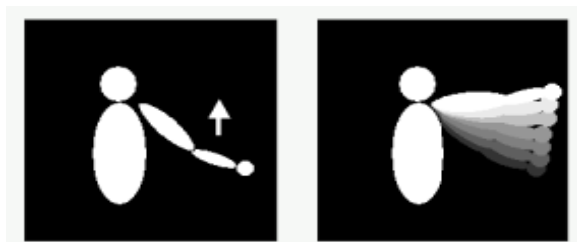


圖八 (b)為睡眠呼吸中止症病人的鼾聲，可以看出兩者間(a)較有規律

當的角度為床頭(此實驗高畫質的拍攝-160*120)。實驗中，我們以 PSG 監控各種生理訊號，並使用近紅外線攝影機紀錄睡眠活動。

3.4.1. 計算移動量和方向

將睡眠影片中每秒的畫面細分成數個畫格(frame)，再利用時間(t-1)的 frame 和時間(t)的 frame 互相做影像相減，藉此得到每個 frame 間的差異，累加出每秒的差異量後，即可得知在每個時間點的移動量。判斷出睡眠影片中移動的物件或是人物之後，將最近發生移動的部份設定為最高值，而相對於較早移動的部份則設為較低值，直到超出我們所要找出最近移動的時間範圍以後，其餘的階設為零，以此類推，產生階層式的運動歷史圖像(Motion History Image, MHI)如圖十：



圖十 左圖為方向預計向上的剪影，所產生的 MHI 為右圖。

得到運動歷史圖像之後，利用 Sobel 將移動的部分作邊緣化處理，而 Sobel 邊緣偵測原理就是對於一張暗色背景下有亮條紋的影像，由水平掃描線的灰階剖面圖及其一階導數圖可發現一個邊緣。得知 $S_x(x,y)$ 、 $S_y(x,y)$ ，再算出物件移動的方向和大小(圖十一左)。

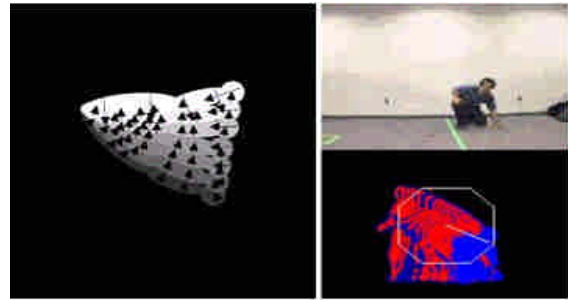
方向：

$$\arctan(S_y(x, y)/S_x(x, y))$$

大小：

$$(S_x^2(x, y) + S_y^2(x, y))^{1/2}$$

將具有方向和大小的部份整理出方向的統計圖，計算方向之間的最小差異，最後統計出結果，找出最多的方向，配合最小差異，算出整體方向 global motion(圖十一右) [13]。



圖十一 左圖算出大小方向、右圖算出整體運動方向

3.4.2. 門檻值與角度設定

由於處理影像的畫質不一，所以必須依照輸入影片的不同解析度來調整 threshold。睡眠影片中，就算是場景不變，但每張 frame 多多少少會受雜訊的影響而有差異，設定 threshold 的目的即是過濾移差異較小的部份，省略不必要的計算，找出真正有效的移動部份。此實驗為例，所輸入資料分別為 640*480 和 160*120 的整晚睡眠影片，利用 openCV 之函式庫，先對低畫質的影片調成與高畫質相同的大小，再將高畫質的 threshold 調為 60，低畫質調為 6。

此外，也注意判定翻身的角度，攝影機架設位置會影響到我們所定義的左翻或右翻。因此角度依照攝影機的架設位置判定並由人工設定。此實驗的高畫質影片將正負 45 度的方向定義為右翻，135 度至 225 度的方向定義為左翻；低畫質的影片由於受試者所躺位置非水平，將 50 度至 140 度定義為右翻，220 度至 320 度定義為左翻。(圖十二)



圖十二 左圖為 640*480，右圖為 160*120

3.5 實驗結果

3.5.1. 一般環境中之人聲分類

表格三 最後分類結果

我們收集了許多不同類型的聲音來測試所提出分類方法的正確性和準確度，說話聲部分包含男、女聲且錄製的時間長度不固定，人聲部分包含笑聲、哭聲、尖叫聲、鼾聲和哈欠聲，環境聲音則包含電風扇、冷氣、電話、除草機、微波爐和吹風機運作時發出的聲音。

第一階段先進行人聲和非人聲分類，結果如表一所示：

Sound Type	Total Number	Result	
		Human Sound	Environment Sound
Human Sound	60	57	3
Environment Sound	30	2	28

表一 人聲和環境聲音分類

第二階段更進一步將人聲分類為語音和非語音部分，假若經分類器運算後的特徵值結果未落在語音或非語音的範圍，則我們將其歸類為環境聲音，分類結果如表格二所示，值得注意的是其中有一個語音訊號被誤判為非語音訊號是由於該檔案的基頻範圍並未落在一般正常語音的範圍內，推測可能的原因為測試者說話時用雙手托住下巴。

Sound Type	Total Number	Result	
		Speech	Non-Speech
Speech	30	29	1
Non-Speech	27	1	26
Environment Sound	2	0	2

表二 語音和非語音訊號的分類

表三列出表一和表二合併後的分類結果：

Sound Type	Result		
	Speech	Non-Speech	Environment Sound
Speech	96.67%	3.33%	0%
Non-Speech	3.33%	86.67%	10%
Environment Sound	0%	3.33%	96.67%

3.5.2. 睡眠時期的鼾聲分類

我們收集了整晚睡眠的錄音來驗證所提出分鼾聲分類方法的正確性和準確性，資料的時間長度約為 7 小時 4 分鐘，我們以 17 個非鼾聲片段和 18 個鼾聲片段做為訓練資料，測試資料方面則有 80 個非鼾聲片段和 1213 個鼾聲片段。

第一階段先將鼾聲從整個晚上的錄音檔中擷取出。藉由上個章節所討論的特徵值，我們給予不同的權重和門檻值即可以訂出兩階段式分類，透過該階層式分類將聲音資料先分為人聲和非人聲，再更進一步將人聲分類為鼾聲和非鼾聲，分類的結果如表格四所示：

Sound Type	Result		
	Snoring sounds	Non-snoring sounds	No sounds
Snoring sounds	96.5%	3.5%	0%
Non-snoring sounds	16.7%	83.3%	0%
No sounds	0%	0%	100%

表四 鼾聲和非鼾聲的分類

將訊號分類為鼾聲和非鼾聲後，我們可以將其再細分為有呼吸中止(apnea)和無呼吸中止(Non-apnea)症狀的鼾聲，我們以 50 個具有呼吸中止症狀的鼾聲片段和 61 個不具呼吸中止症狀的鼾聲片段做為訓練資料，能量和頻譜亂度的 KL divergence 為特徵值，並分別給予兩特徵值不同的權重，權重的選定可來自於訓練其他鼾聲資料求得，最後根據兩特徵值的總和我們可得到一個門檻值，在測試階段，假若該鼾聲片段的分數低於門檻值則歸類為不具呼吸中止症狀的鼾聲，反之則歸類為具呼吸中止症狀的鼾聲，分類結果如表五所示：

Sound Type	# Data Segments	Result	
		Apnea	Non-apnea
Apnea	50	94%	6%
Non-apnea	61	18%	82%

表五 鼾聲的分類

3.4.3 睡姿和鼾聲的關係

解決影片大小不一和角度問題之後，分別對高解析度和低解析度的影片執行運算。實驗結果顯示利用高解析度的影片，可以正確判斷大致上受試者移動的方向，並可局部的動作分析。紅色圈出的地方代表移動的部位，紅圈裡的半徑角度代表移動的方向。由圖四可知，此位受試者試圖向右翻身，但也有例外發生，如圖四右圖的頭部移動方向為向上，但實際上影片中的受試者頭部並沒有出現向上的移動，因此可能為光線影響到計算移動量的程度（圖十三）：



圖十三 左圖找出大致上方向後右圖仍有例外



圖十三 例外放大圖

低解析度的影片，所偵測出的移動部分較少，降低 threshold 的目的就是更容易找到細微變化的部份，實驗結果發現，判定出連續的同方向中，偶爾會出現不同方向的情況發生，扣除受試者本身的移動，可能是雜訊的影響，因此可利用規則判斷出左翻還是右翻。例如連續幾個左翻身之中若有單一的右翻身出現，則自動忽略此右翻，直接判斷此時段受試者為左翻身。部分數據如圖十四：

angle	direction	time(s)	frames
182.000000	left	8884.000000	8883th
91.000000	right	8885.000000	8884th
213.000000	left	8886.000000	8885th
218.000000	left	8887.000000	8886th
217.000000	left	8888.000000	8887th
210.000000	left	8889.000000	8888th

圖十四：紅色標示角度為 91 度在低畫質中判定為右翻身，可是其前後之角度均在左翻身的範圍內，因此可直接省略。

四、結論

本研究著重於人聲的分類，首先我們提出了新的方法去分類一般環境中的人聲，目標希望將聲音訊號自動分類成語音、非語音和環境聲音，我們正確地找出合適的特徵值達成聲音訊號的分類，實驗的結果明確的顯示出我們所提出的演算法是有效的。

目前所收集的環境聲音侷限於機器所產生的聲音，在未來我們將試著加入更多類型的環境聲音，例如：將數種聲音混合在一起，另外希望更進一步細分非語音人聲部分，使其能分辨哭聲、笑聲、尖叫聲等聲音。藉由合併基本的生理訊號例如：皮膚電流反應(Galvanic skin response, GSR)、呼吸率和心跳率，將有效提升分類的程度和準確率。

在睡眠時期的鼾聲分類部分，我們著重於鼾聲訊號的分析，目標希望可由整晚的聲音訊號中自動分類出鼾聲和非鼾聲，我們已經正確地找出合適的特徵值達成鼾聲分類，實驗的結果也顯示我們所提出的演算法是有效的。

在睡姿和鼾聲的關係的研究部分，目前睡眠研究中的影像資料雖極少被用來做後續的分析與處理，然而根據此資料仍可提供有用的資訊提供專業人員參考。利用自動化分析影像資訊，比起人工判讀不但可以更省時省力，亦可增加精確度。由目前的實驗數據，我們可推知睡眠處於睡或

醒的階段與睡眠時翻身的程度等移動量有相當大程度的關聯，在腦波處於清醒階段時，翻身程度相當顯著、次數也明顯增加；而腦波若處於睡眠階段，無論是階段一、階段二、階段三或階段四，其翻身程度都較低。為了驗證以上的假設，必須再收集更多的實驗資料，或與 Acti-watch 計算移動量的手錶型計數器做比較，以更精確的分析睡眠狀態和做更進一步的資料分析。

總結而言，本計畫的執行除了加強與本校心理系睡眠實驗室的跨領域研究，建立合作關係，並已開始學術論文的協同撰寫，具體產出的成果為一篇碩士論文[14]，兩篇國內會議論文[15,16]與一篇國際會議論文[17]，目前仍有一篇期刊論文投稿中。

五、參考文獻

- [1] 醫療快訊
<http://www.libertytimes.com.tw/2002/new/mar/2/today-m1.htm>
- [2] 黃席珍: 睡眠忠告, "睡一個好覺", 商周出版。
- [3] SNAP laboratories:
<http://www.snaplab.com/home.htm>
- [4] 陳明周, 鼾聲與呼吸中止症監測裝置設計與製作", 元智大學機械工程研究所碩士論文, 民國 93 年。
- [5] Hattori H.; Kawakatsu K.; Hattori C.; Nishimura Y.; Suzuki K.; Yagisawa M.; Nishimura T. "Acoustic Analysis of Pediatric Snoring", International Congress Series Vol. 1257, pp. 219-225, 2003.
- [6] Vocal tract acoustics:
<http://www.phys.unsw.edu.au/speech/>
- [7] K. Wilson, R. A. Stoohs, T. F. Mulrooney, L. J. Johnson, C. Guilleminault, Z. Huang, "The Snoring Spectrum", Chest, March 1, 115(3): 762 - 770, 1999.
- [8] R. Beck, M. Odeh, A. Oliven, N. Gavriely, "The Acoustic Properties of Snores", European Respiratory Journal, 8, pp. 2120-2128, 1995.
- [9] A.R.A. Sovijärvi, L.P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmaso, C. Sacco, M. Rossi, J.E. Earis, "Characteristics of Breath Sounds and Adventitious Respiratory Sounds", European Respiratory Review, 10: 77, 591-596, 2000.
- [10] T. M. Jones, A. C. Swift, P. M. A. Calverley, M. S. Ho and J. E. Earis, "Acoustic Analysis of Snoring Before and After Palatal Surgery", European Respiratory Journal, 25:1044-1049, 2005.
- [11] CMAJ:

- <http://www.cmaj.ca/cgi/content/full/173/1/28>
- [12] 葉世彬, "睡眠檢查",
<http://www.stm.org.tw/yeh/exam.htm>
- [13] OpenCV documentation:
<http://www710.univ-lyon1.fr/~bouakaz/OpenCV-0.9.5/docs/>
- [14] 蘇以暄, "Analysis and Classification of Human Sounds", 碩士論文, 國立政治大學資訊科學系, 2006.
- [15] 黃思瑜、廖文宏, "睡眠研究中的視訊處理", Multimedia and Networking Systems Conference, Dec. 2006.
- [16] 黃思瑜、廖文宏、楊建銘、蔡明潔: "睡眠研究中的視訊分析: 與 Actiwatch 的比較", 台灣睡眠醫學學會年會, Mar. 2007.
- [17] Wen-Hung Liao, Yi-Syuan Su, "Classification of Audio Signals in All-Night Sleep Studies," pp. 302-305, 18th International Conference on Pattern Recognition (ICPR'06) Volume 4, 2006

六、計畫成果自評

就研究內容與原計畫相符程度來評估，廣度部分超越了原先的設定，特別是在視訊處理的部分有更多的著墨，至於深度部分也大抵符合原來設定的目標，對於主要的議題均有相對的研究，唯一較需加強的是樣本收集的數量與在醫療院所實際的個案研究。由於整晚睡眠資料蒐集部分程序較為繁複，雖然我們著重在音訊處理的部分，但為了與其他生理訊號作參考比對參考，仍需同時進行多重生理紀錄儀實驗，因此相當耗時耗力，所得樣本數目與原先規劃稍有差距，但是分析結果仍有一致性，不致影響研究結論。此外，多數資料仍在政大心理系睡眠實驗室收集，或許有不夠多樣性的問題，若將目前技術推廣至實際應用，仍需在醫療院所多蒐集個案來驗證可行性。

研究之部分成果部分已經發表至國內外會議，目前正持續整理中，準備投稿相關的期刊。此外，我們也針對演算法的強健性 (robustness) 進行測試，並準備導入實務之應用。

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

睡眠研究中鼾聲之聲學分析
Acoustic Analysis of Snoring Sounds in Sleep Studies

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 95-2221-E-004-012-

執行期間：95年8月1日至96年7月31日

計畫主持人：廖文宏

共同主持人：楊建銘

計畫參與人員：林裕凱、紀煜豪

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：國立政治大學資訊科學系

中華民國 96年 10月 31日

可供推廣之研發成果資料表

 可申請專利 可技術移轉

日期：96年10月31日

國科會補助計畫	計畫名稱：睡眠研究中鼾聲之聲學分析 計畫主持人：廖文宏 計畫編號：NSC 95-2221-E-004-012 學門領域：資訊
技術/創作名稱	睡眠研究中鼾聲之分析與分類
發明人/創作人	廖文宏
技術說明	中文： 我們提出了睡眠研究中鼾聲的自動分類方法。此項技術的主要目標是將整晚的錄音中分出鼾聲與非鼾聲的片段，並就鼾聲的段落再進行分析。我們使用階層分類法，利用基礎聲學特徵將錄音區分為人聲與非人聲，再用更多的特徵將人聲的部分分為鼾聲與非鼾聲，接著並就鼾聲的特性再細分為正常與睡眠呼吸終止症的鼾聲，實驗結果驗證了所提出方法的效能與準確性。
	英文： We developed an automatic process for the classification of audio signals recorded in all-night sleep studies. Our objective is to separate the episodes into snoring sounds and non-snoring sounds. To begin with, we employ hierarchical classification schemes to classify sounds into human sounds and non-human sounds. We then attempt to organize human sounds into snore and non-snore segments based on their acoustic properties. We perform further analysis of the extracted snoring sounds to check if the testee has apnea. Experimental results have validated the efficacy of the proposed method.
可利用之產業及可開發之產品	<ul style="list-style-type: none"> ● 醫療院所之睡眠檢驗室。 ● 研究機構之睡眠實驗室。 ● 一般患者之自我檢測。
技術特點	<ul style="list-style-type: none"> ● 利用電腦自動分析並分類鼾聲，加速處理的程序與增加準確度。 ● 所使用之技術可處理一般廉價錄音筆所錄製之音訊檔，減低成本。
推廣及運用的價值	<ul style="list-style-type: none"> ● 相較於多重生理紀錄儀(polysomnogram, PSG)，睡眠研究中的音訊分析屬於簡易、非接觸式且成本低廉的技術，可有效應用於患者的自我篩檢或前測。

- ※ 1. 每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。
- ※ 2. 本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。
- ※ 3. 本表若不敷使用，請自行影印使用。

出席國際學術會議心得報告

計畫編號	NSC 95-2221-E-004-012
計畫名稱	睡眠研究中鼾聲之聲學分析
出國人員姓名 服務機關及職稱	廖文宏 國立政治大學資訊科學系 助理教授
會議時間地點	Hong Kong, 8/20/2006-8/24/2006
會議名稱	18 th International Conference on Pattern Recognition
發表論文題目	1. Classification of Audio Signals in All-Night Sleep Studies(睡眠研究之音訊分類) 2. A CAPTACHA Mechanism by Exchanging Image Blocks(植基於圖像內涵之自動化人機區辨機制)

一、參加會議經過及與會心得

2006年8/20-24日在香港國際會議中心舉行的國際圖型識別研討會(International Conference on Pattern Recognition)，每兩年舉辦一次，至今已第十八屆，邁入三十六年，是圖形識別領域歷史最為悠久，也是最重要的會議之一，本次會議分為五大主題(tracks)，共有超過兩千篇投稿，接受的比例在五成上下，實際與會的人數達一千多人，可以說規模盛大，而報告內容也是包羅萬象，吸引來自不同領域與國家的研究學者。

此次會議的五大主題為：

1. Computer Vision and Image Analysis
2. Pattern Recognition and Basic Technologies
3. Signal, Speech and Image Processing
4. Systems, Robotics and Applications (with Associated Theme : Biometrics)
5. Cognitive Approaches & Soft Computing

由於有五場論文發表同步進行，加上海報部份也在相鄰場地展示，因此實際參與的場次以個人發表(兩個場次)、Keynote 演講以及 1,3,5 類與個人研究相關者為主。

第一場 Keynote 是由本屆 K. S. Fu Prize 得主 Prof. Josef Kittler 發表的演講，題為”On Context, Modeling, Dimensionality and Small Sample Size in Pattern Recognition”，主要在於回顧得獎者在圖型識別領域之研究成果，特別是在 training sample 數目不多或者每個類別的 sample 數不均衡時的解決方案，溫故知新之餘，也對 Kittler 教授與其實驗室的研究有了更深入的了解。

第二場 Keynote 由 Michigan State University 的 Anil Jain 教授報告指紋辨識方面的最新進展(於 http://biometrics.cse.msu.edu/Presentations/AnilJain_Fingerprints_ICPR06.pdf 下載簡報)，

Prof. Jain 是生物辨識(biometrics)方面的專家，在指紋辨識領域有卓越的研究成果，因此報告內容豐富，理論與實際並重，任何對此一領域有興趣的讀者，都不應錯過。

另外一個頗有意思的演講為來自 MIT 的 Alex Pentland 教授報告的”A Computational Model of Social Signaling”，利用人類不自覺的聲音與姿態上細微變化，來觀察並推估特定情境下的決策模式，與 Picard 教授的情意運算(affective computing)有本質上的差異，卻有著十分有趣的應用。由於本人亦參與另一國科會的「智慧生活空間」跨領域研究計畫，對於新型態的人機介面，如 Attentive UI, Affective UI 的發展保持密切的注意，因此此一觀點勢將引起團隊成員的熱烈討論。

其餘參加的場次由於論文數目繁多，因此不再一一詳述，論文光碟作成可搜尋的形式，可以很方便的檢索。

值得一提的是今年大會新增了 Aggarwal award，表揚圖型識別領域的年輕學者，而 Dr. Aggarwal 正是個人博士班的指導教授，在香港再度見面，格外令人興奮；而我也藉此機會探詢 Aggarwal 教授到本校進行短期訪問之可能性，並得到正面的回應，因此將透過系上進行正式的邀請。

- 攜回資料名稱與內容

18th ICPR Proceedings 光碟