

子計畫三：導盲機器人之聽覺系統與人機溝通介面研製(I)
Hearing and Human-machine Communication System of a Guidance
Robot for Blind Pedestrians

計畫編號：NSC88-2213-E-009-107

執行期限：87.8.1-88.7.31

主持人：林進燈 國立交通大學 教授

執行機構：國立交通大學電機與控制工程研究所

一、摘要

本研究計畫主要是建立導盲機器人的聽覺系統與口語人機溝通界面，以提供盲胞更方便的操作。在聽覺系統方面，目前我們已經研發出抗雜訊的口語辨識器使其在高噪音的環境中能有良好的辨識能力。我們提出利用遺傳演算法則的改良式二維倒頻譜參數(GA-based MTDC)。此參數使用 temporal filter 來去除噪音干擾所造成的 dc 成分後再經由我們利用遺傳演算法則選取抗雜訊參數位置，確實發現 GA-based MTDC 具有抗雜訊能力。

二、目前研究進度：

在眾多語音參數中，二維倒頻譜(Two-Dimensional Cepstrum—TDC)同時將語音信號局部的特性與整體的變化包含在一個係數矩陣中。根據分析，在乾淨音下，位於係數矩陣較低階區域的參數為重要的語音參數。因此只需採用小部份的係數來形成參數向量，代表該語音。因為每一個語音信號是用一個參數向量而不是一串參數向量代表，所以所需儲存參數空間較少且計算量少。但我們的實驗發現在噪音環境下，利用 TDC 的辨識率會急速下降。我們由實驗分析出一個現象，雜訊會對各頻帶造成直流效應，這現象使 TDC 矩陣易受雜訊干擾。在上半年中，我們提出改良式二維倒頻譜參數(Modified TDC —MTDC)來解決此問題。MTDC 主要觀念為利用一個 4 階

IIR 高通濾波器依音框方向濾除噪音成份。此參數可以有效減少環境噪音對其干擾。MTDC 之求取共分成四個步驟：

<步驟 1>：求出 20 個 Mel 刻度頻帶的語音輸出，如圖 1 所示。

<步驟 2>：對每個頻帶的輸出加上時間軸濾波器並且利用全波整流器(Full-wave rectifier)去除雜訊干擾部份。

<步驟 3>：對全波整流器輸出的語音波形處理，取餘弦轉換(cosine transform)得到 MFC (Mel-frequency cepstrum)參數。

<步驟 4>：收集所有的 MFC 參數，然後沿著時間軸方向進行反離散複利葉轉換(Inverse Discrete Fourier Transform)。最後選取實數部分當作語音辨識參數。

根據人類感官聽力的實驗，我們知道人耳對低頻部分為線性，但高頻部份則呈對數(log-linearly)刻度分布。因此我們使用 mel 頻率刻度，而其頻率定義為：

$$m = 2595 (1 + f / 700) \quad (1)$$

其中 m 為 mel 頻率刻度而 f 為 Hz，我們用(1)式設計濾波頻帶和 mel 刻度加權參數。我們可以對頻譜乘上加權參數：

$$X_{nm} = \sum_{k=0}^{N-1} X_{nmk} f_{nk} \leq i \leq 20 \quad n \leq N$$

其中 i 是濾波器編號，k 為頻譜編號，f(i, k)是第 i 個濾波器頻譜加權參數。步驟 3 的過程如下式所示。

$$X(m, l) = \begin{cases} \log|X(m, i)| & 1 \leq l \leq D, 0 \leq m \leq M-1 \\ 0 & \text{if } X(m, i) \leq 1 \end{cases}$$

$$C(m, n) = \sum_{i=1}^{20} \hat{X}(m, i) \cos\left(\frac{n(i-0.5)\pi}{20}\right)$$

(3)

其中 $C(m, n)$ 是 MFC 係數。步驟 4 即是 MTDC 處理程式(見圖 2)。在此步驟中，我們收集每個音框的 MFC 係數，沿著時間軸累積成 MFC 矩陣，然後對矩陣作 Inverse discrete Fourier transform (IDFT) 並取其實數部分產生 MTDC 矩陣。步驟 4 的過程可以下式來表示

$$C_{rn} v_{rn} = \frac{1}{M'} \sum_{m=0}^{M-1} C_{rn} m_{rn} W_M^{-rnv}$$

$$\text{where } C_{rn} = \begin{cases} F & \text{if } v \leq M' \\ MF & \text{if } v > M' \end{cases},$$

$1 \leq n \leq D$

$$MTDC_{nm} = \text{Re}[C_{nm}] \quad (4)$$

由上面步驟我們可獲得 MTDC 矩陣。但如前所述，僅少數的 MTDC 係數即可產生高辨識率。如何選擇最具代表性的 MTDC 係數是一個研究重點。尤其在噪音環境下，由於不同噪音對不同的 MTDC 係數有不同程度的影響，故我們希望利用遺傳演算法則去求出最具抗雜訊能力的參數位置。在下半年的研究中，我們提出基於基因演算法的最佳係數選取。我們利用遺傳演算法則選取 MTDC 抗雜訊參數位置。而此新參數稱為 GA-based MTDC。

基因演算法是基於自然界物競天擇、遺傳、突變的過程來搜尋最佳解。它具有整體最佳解搜尋的能力並且被應用在兩個領域：最佳化(optimization)及機械學習(machine learning)。但在語音辨識的領域中，遺傳演算法卻很少被使用。我們嘗試結合遺傳演算法到語音辨識。我們經由人為選取的 M_TDC 矩陣中的係數位置，雖然無雜訊環境下有好的辨識率，但在噪音環境下卻辨識率不佳。原因是我們人為選取的係數位置可能易受噪音干擾。為了尋找 M_TDC 較抗雜訊係數的位置，我們使用遺傳演算法中物競天擇的進化概念來達到這個

目的。使用遺傳演算法中物競天擇的進化概念來達到這個目的。在傳統的遺傳演算法中，它可分成三部分運作：繁殖、交配、和突變。詳述如下：

•編碼：

將參數編碼成基因，如圖 3。我們先將 M_TDC 矩陣取出非重覆性的有意義部分形成一小矩陣，它位於 M_TDC 矩陣的左下角。此部分含有語音辨識的重要訊息。我們將此小矩陣中的每一個參數設定代碼(index)。後我們對此代碼作 7 位元長的編碼。例如：字“1110101”代表代碼為 117。我們使用 30 個係數代表一個語音訊號，故我們希望從小矩陣中選出 30 個抗雜訊的係數位置。

•初始值：

在開始進入基因演算法參數選取之前，我們需要先選取某些參數。這些被始設定尚有族群大小 N ，突變機率 P_m ，交配機率 P_c ，和最大的疊代終止次數。

•合適函數(Fitness function)：

合適值(fitness value)我們取用辨識率。詳細步驟如下：等一我們將語音資料分成兩部分。第一部分訓練語音辨識器。此部份為使用乾淨音。第二部分為用來測無噪音及噪音下的辨識率。這些包含在 SNR 為 5dB, 10dB, 15dB, 20dB, 30dB 及 ∞ dB。各自代表不同程度噪音干擾。因此我們可以利用辨識率來代表遺傳演算法中的合適值。

•繁殖：

各個編碼後的基因根據各自的合適值來決定繁殖能力。有一種稱輪盤法(roulette wheel parent selection)被我們採用。但這方法有一個缺點，即擁有最大合適值者也可能被犧牲掉(只是機率比較小)。為克服此機率上的缺失，我們用“精英保留”(elitist strategy)。“菁英保留”會保留合適值高者的繁殖力故能夠加速收速的速度，也因此提昇基因演算法的能力。

•交配：

繁殖雖然增加了個體數，卻未增加任何新的成員。新成員的加入有靠交配的運作。首先兩個基因碼交配的機率為 P_c ，且兩刀切割法被用於此。

•突變：

突變是一種使全新的基因碼產生的步驟。此全新的基因碼是隨機產生的。突變機率必須取很小值，因它代表隨機搜尋之意義。如果突變機率設定值太大則遺傳演算法則相當於隨機搜尋最佳解。我們系統中設定突變機率值為 $P_m = 0.001$ 。

我們將基因演算法則用在抗雜訊語音參數位置選取之步驟流程圖如圖 4 所示。

三、實驗測試

噪音信號取自 NATO Research Study Group on Speech Processing (RSG.10) Noise-ROM-0。它們來自於 24 種各式各樣的噪音種類。我們從此噪音資料庫選出 5 種代表性噪音訊號。這 5 種噪音分為別“人群喧嘩聲”，“驅逐艦工作室”，“工廠噪音”，“戰鬥機聲”，“白色雜訊(white noise)”。原始音檔取樣頻率為 19.98KHz 儲存在 16bits 整數型態。在我們實驗中它們被轉成 16KHz 來使用。我們使用中文 10 個數字來用作實驗。這 10 個數字由 10 位男性所發音錄製，每個數字唸 20 遍，取樣頻率為 16KHz，每個音被用手動切割選出正確語音訊號。目前此部份語音資料庫已經搭配我們所提出的 GA-based MTDC 建立成一套抗雜訊的語音辨識系統。圖 5 顯示此系統的主功能畫面。為了實驗測試需要，我們可以利用圖 6 的環境設定選項來設定語音辨識環境。語音辨識結果則顯示在圖 7。

我們在實驗中比較了 GA-based M_TDC 方法和 TDMFC(two-dimension mel-frequency cepstrum)及 HMM 方法。TDC 方法所用參和所用相同。在 HMM 中所用的參數為 MFCC (mel-frequency cepstrum coefficient)。視窗長度(window)為 25ms 且重疊部分為 10ms。預強調係數為 0.97。每個音框的係數共 26 個，包含 MFC, delta MFC 及 energy, delta energy 等等。每個音都被模組成 5 個狀態(state) 每個狀態被分成兩個流(stream) 變異數(Covariance)(即只要算平均值和變異數)結果在 Fig. 8~12。從實驗結果中我們可以發現 GA-base M_TDC 具有最佳的抗雜訊能力。

四、結論與討論

在本年度中我們提出利用遺傳演算法則的改良式二維倒頻譜參數(GA-based MTDC)。此參數使用使用 temporal filter 來去除噪音干擾所造成的 dc 成分。我們再利用遺傳演算法來找出 M_TDC 矩陣中較具抗雜訊的參數。GA-based MTDC 能夠同時表現出許多包含於語音波形中的重要消息及包含靜態與動態參數，並且每個語音只需某些參數代表而不需用到所有參數。因此，它具有低計算量和低儲存需求的優點，故可以實現到真實世界。在實驗數據中，我們經由 5 種不同噪音源測試，我們發現 GA-based M_TDC 在噪音干擾時的辨識率確實是大於 TDC 在相同環境下的辨識率。在本年度我們的重點工作是成功地解決在噪音環境下提昇語音辨識率的問題。

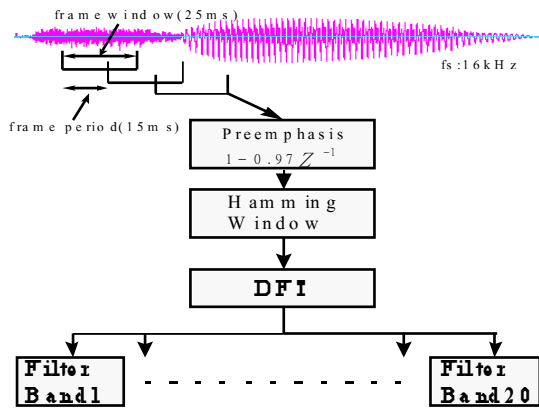


圖 1：求取 MTDC 過程之步驟 1 的方塊圖。

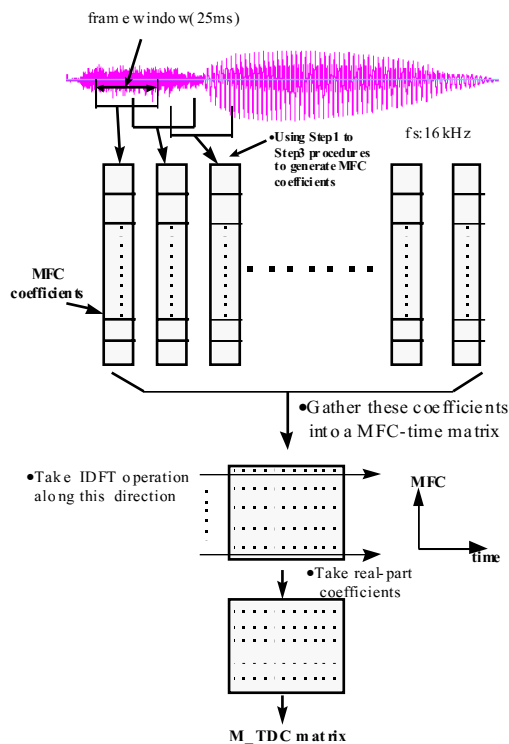


圖 2：求取 MTDC 過程之步驟 4 方塊圖。

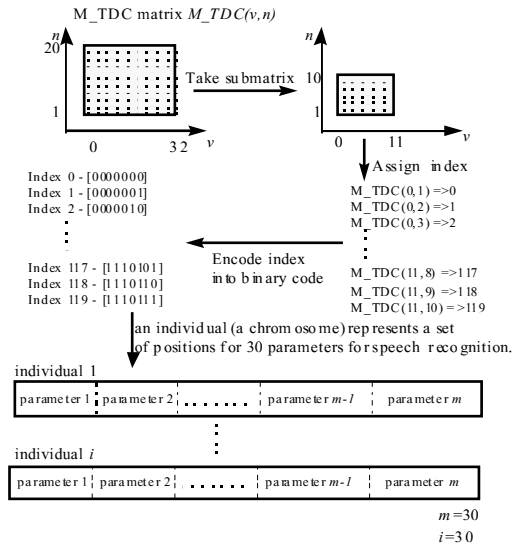


圖 3：將語音參數編碼成基因之示意圖。

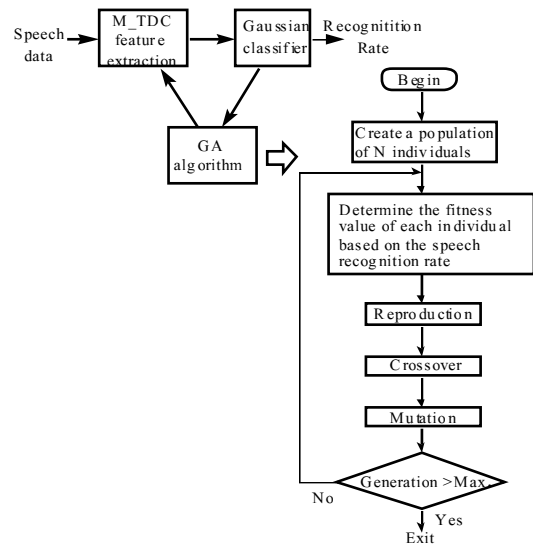


圖 4：將基因演算法則用在抗雜訊語音參數位置選取之步驟流程圖。



圖 5：語音辨識系統主功能畫面。



圖 6：語音辨識系統功能選擇畫面

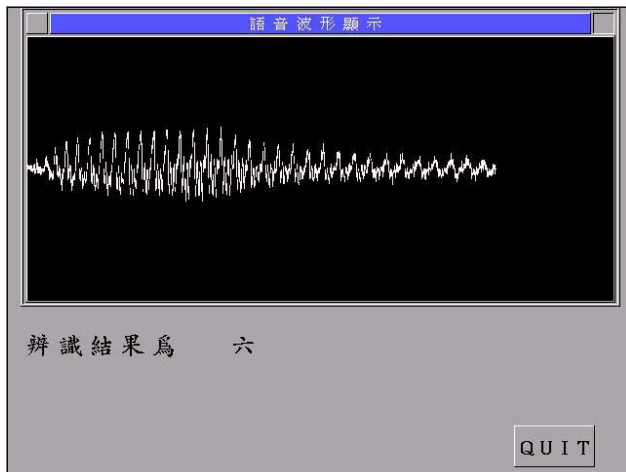


圖 7：語音辨識系統辨識結果畫面

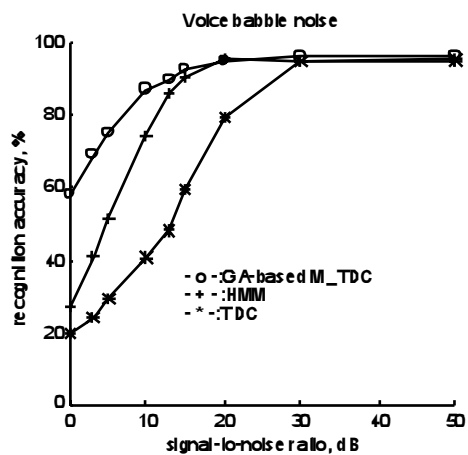


圖 8：在人群喧鬧聲中之辨識率實驗結果。

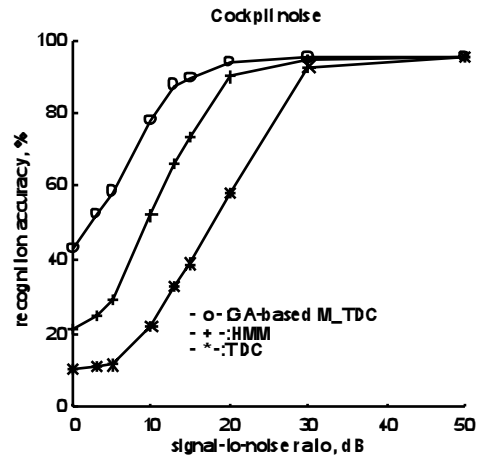


圖 9：在戰鬥機聲中之辨識率實驗結果。

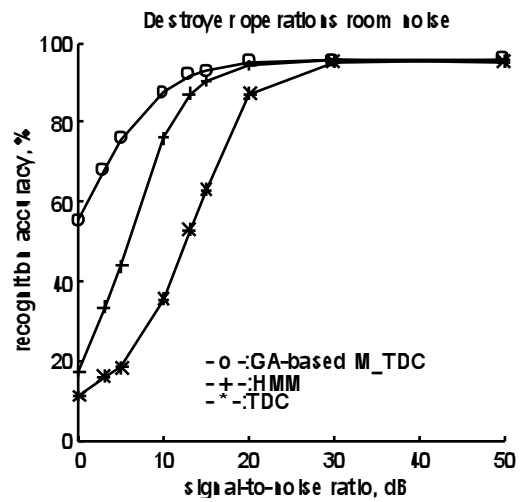


圖 10：在驅逐艦聲中之辨識率實驗結果。

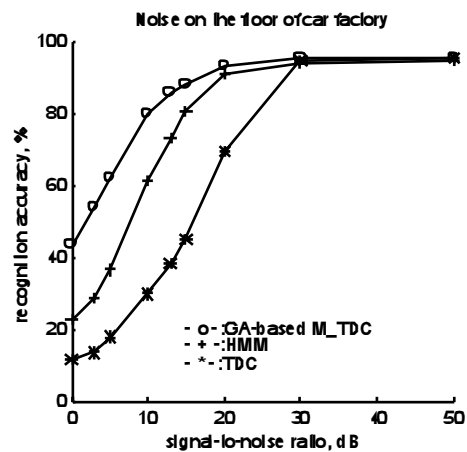


圖 11：在工廠聲中之辨識率實驗結果。

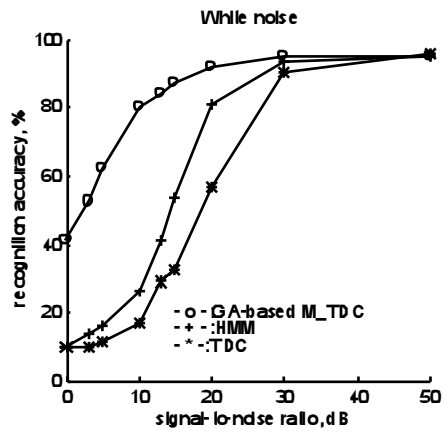


圖 12: 在白色雜訊干擾下之辨識率實驗結果。