The 2015 Conference on Computational Linguistics and Speech Processing
ROCLING 2015, pp. 123-135

語音增強基於小腦模型控制器

朱皓駿 Hao-Chun Chu， 李仲溪 Jung-Hsi Lee
方士豪 Shih-Hau Fang， 林志民 Chih-Min Lin
元智大學電機工程學系
Department of Electrical Engineering, Yuan Ze University
david4633221@gmail.com
{eejlee, shfang, cml}@saturn.yzu.edu.tw

張雲帆 Yun-Fan Chang， 曹昱 Yu Tsao
中央研究院資訊科技創新研究中心
Research Center for Information Technology Innovation, Academia Sinica
{she2113, yu.tsao}@citi.sinica.edu.tw

摘要

本文提出了一個小腦模型控制器(Cerebellar Model Articulation Controller, CMAC)應用於語音增強系統(Speech Enhancement System)，所提出的 CMAC 使用歸一化梯度下降法(Normalized Gradient Descent Method) 增加 CMAC 參數的自適應學習速度，具有比傳統類神經網路方法更快的學習速度、體積小且良好的泛化，因此更適合做高速的訊號處理。實驗方面，使用 CMAC 與 MMSE 做比較，為了比較性能，我們用了三種語音評估方法來做 CMAC 消除雜音及 MMSE 消除雜音後的數值比較，分別為(Perceptual Evaluation of Speech Quality, PESQ)、(Segmental Signal-to-Noise Ratio, SSNR)以及(Speech Distortion Index, SDI)。由實驗結果可知，在三種評估方法，CMAC 皆能達到較佳的結果。

Abstract

Traditionally, cerebellar model articulation controller (CMAC) is used in motor control, inverted pendulum robot, and nonlinear channel equalization. In this study, we investigate the capability of CMAC for speech enhancement. We construct a CMAC-based supervised speech enhancement system, which includes offline and online phases. In the offline phase, a paired noisy-clean speech dataset is prepared and used to train the parameters in a CMAC model. In the online phase, the trained CMAC model transforms the input noisy speech signals to enhanced speech signals with reduced noise components. To test the CMAC-based speech enhancement system, this study adopted three speech objective evaluation metrics, including perceptual evaluation of speech quality (PESQ), segmental signal-to-noise ratio (SSNR) and speech distortion index (SDI). A well-known traditional speech enhancement approach, minimum mean-square-error (MMSE) algorithm, was also tested performance for comparison. Experimental results demonstrated that CMAC provides superior performances to the MMSE method for all of the three objective evaluation metrics.

關鍵詞：小腦模型控制器，語音增強，最小均方誤差

Keywords: CMAC, Speech Enhancement, MMSE

一、簡介

語音訊號會由於背景雜音造成語音品質降低，語音增強系統(Speech Enhancement System)主要目的是減少雜音成分，從而提高訊雜比(SNR)。從吵雜語音中估計出乾淨語音是許多實際應用中非常重要的語音技術，如自動語音識別(Automatic Speech Recognition, ASR)和助聽器(Hearing Aids) [1, 2]等應用。語音增強算法大致分為兩類，即非監督(Unsupervised)和監督(Supervised)算法，非監督語音增強算法優點在於需要很少甚至不需要事先準備數據，一個好的非監督語音增強算法是利用頻譜恢復 [3]，頻譜恢復方法的目標是在頻域中估計出增益函數，以用來降低雜音，頻譜恢復的方法包括譜減法(Spectral Subtraction, SS) [4]和溫尼濾波器(Wiener Filtering) [5]，與他們的各種延伸 [6-9]。此外，另一些頻譜恢復的方法是推導出語音訊號和帶雜音訊號的概率模型(Probabilistic Models)，成功的例子包括最小均方誤差(MMSE)頻譜估計 [10-14]、最大事後頻譜振幅(Maximum A Posteriori Spectral Amplitude, MAPA)估計器 [15-18]和最大可能頻譜振幅(Maximum Likelihood Spectral Amplitude, MLSA)估計器 [19, 20]等。目的是用雜訊追蹤法(Noise Tracking)估計出雜訊的功率頻譜，常見的雜訊追蹤法如語音活動檢測(Voice Activity Detection, VAD)、最小統計法(Minimum Statistic, MS) [21, 22]等。得到雜訊功率頻譜後，即可得到事前訊雜比(a priori SNR)與事後訊雜比(a posteriori SNR)，根據這兩種訊雜比可以算出增益函數(Gain Function)，利用此增益函數做語音增強，即可估計出乾淨語音訊號頻譜。而監督語音增強算法需要事先混合雜音和乾淨語料，以便處理在線(Online)語音增強，成功的例子包括 Deep Neural Network(DNN) [23]、Deep Denoising Autoencoder(DDAE) [24]、Sparse Coding [25]及 Nonnegative Matrix Factorization(NMF) [26]語音增強算法等。本文提出的 CMAC 語音增強是採用監督算法。

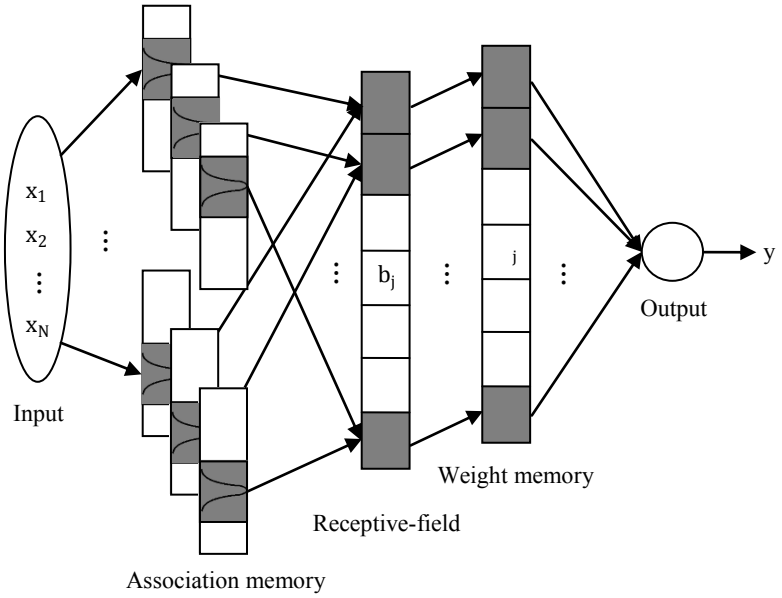
近年來在語音增強系統(Speech Enhancement System)上有許多機器學習(Machine Learning)方法，如: DNN、Sparse Coding 及 NMF 等。本論文則使用 CMAC，近年 CMAC 較常應用在馬達控制 [27]、倒單擺機器人 [28]、MIMO [29]控制等，而在訊號處理方面，非線性信道均衡(Nonlinear Channel Equalization)以及雜訊消除(Noise Cancellation)系統上均有良好的效果 [30]，我們則研究此方法在語音增強系統(Speech Enhancement System)上的效果。由於在降噪的過程中可能會造成語音訊號失真，這會嚴重降低語音的品質，因此我們使用 SDI 評估方法來決定 CMAC 參數的調整，最後使用 SSNR 與 PESQ 評估語音訊雜比與語音品質。

小腦模型控制器(CMAC)被列為是非完全連接感知機(non-fully connected perceptron-like)聯想記憶網路(associative memory network)重疊接受域(receptive-field) [31]。它可以解決規模快速增長(fast size-growing)的問題，還有現有神經網路學習上的困難。傳統的 CMAC 使用局部性(local)固定二進制接受域(receptive-field)基礎函數，缺點是輸出中每個量化的狀態不變，不保留衍生的信息。學習時 CMAC 的輸入為帶雜訊語音，輸出為增強後乾淨語音，我們會記錄學習完成後的 CMAC 內的所有參數，在測試時直接使用這個 CMAC Model 亦可以把雜音消除。

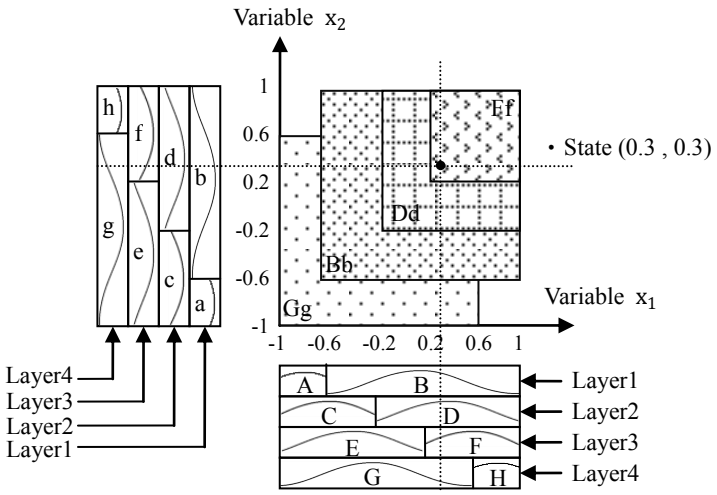
本論文第二章介紹 CMAC 主架構，第三章介紹 CMAC 參數的自適應學習算法，第四章介紹實驗與評估方法，音檔處理過程以及 CMAC 消除雜音步驟，再探討 CMAC 各參數的設置會造成什麼影響，第五章結論。

二、CMAC 結構

CMAC 架構圖示於圖一(A)，是由一個輸入空間(Input space)，聯想記憶空間(Association memory space)，接受域空間(Receptive-field space)，權重儲存空間(Weight memory space)，輸出空間(Output space)組成，圖一(B)是由一個由二維情況下的圖解法。



(A)



(B)

圖一、(A)CMAC 架構圖 (B)二維 CMAC 示意圖

1. 輸入空間(Input space)：輸入$x_i = [x_1, x_2, \cdots, x_N]^T \in R^N$，其中 N 是輸入維數，在圖一(B)上$N = 2$，$x_i$可以被量化到離散區域$N_e$，$N_e$被稱為元素數(Elements)，也可稱為分辨率，上界(Upper bound) = 1，下界(Lower bound) = -1，N_e在上界及下界中分割成 5 等份({-1, -0.6},{-0.6, -0.2},{-0.2, 0.2},{0.2, 0.6},{0.6, 1})，所以$N_e = 5$。本文 CMAC 架構設計時，元素數(N_e)除層數(Layer)需要餘 1。

2. 聯想記憶空間(Association memory space)：多個元素(Elements)可以累積為一個塊(N_B)，$N_B = \text{ceil}(N_e/\text{Layer})$，ceil 代表餘數無條件進位，通常$N_B \geq 2$，在圖一(B)上$N_B = 8$ (A, B, C, D, E, F, G, H)。N_A表示聯想記憶空間個數($N_A = N \times N_B$)。在每個塊(N_B)空間中，需要放入一個連續有界函數，它可以定義為三角形函數或小波函數或任意連續有界函數，在這裡聯想記憶函數是採用高斯函數，它可以表示為(1)式

$$\varphi_{ij} = \exp\left[\frac{-(x_i - m_{ij})^2}{\sigma_{ij}^2}\right] \quad \text{for} \quad j = 1,2,\cdots,N_B \quad \text{and} \quad i = 1,2,\cdots,N \tag{1}$$

其中m_{ij}和σ_{ij}分別為聯想記憶函數內第 i 個輸入的第 j 個塊的平均值及變異數，c_i是輸入訊號。

3. 接受域空間(Receptive-field space)：多個聯想記憶空間可以組成一個接受域空間，在本文中$N_B = N_R$，如圖一(B)是由兩個聯想記憶空間內相對應的兩個塊(N_B)組成一個接受域(N_R)，如 A 塊和 a 塊組成一個接受域(Aa)。第 j 個接受域函數表示為(2)式

$$b_j = \prod_{i=1}^{N} \varphi_{ij} = \exp\left[-\left(\sum_{i=1}^{N} \frac{(x_i - m_{ij})^2}{\sigma_{ij}^2}\right)\right] \tag{2}$$

接受域函數可以用向量的形式表示，如(3)式

$$\underline{b} = [b_1, b_2, \cdots, b_{N_R}]^T \tag{3}$$

4. 權重儲存空間(Weight memory space)：在接受域空間中的每個位置的權重調節值可表示為(4)式

$$\underline{\quad} = [\ _1,\ _2, \cdots,\ _{N_R}]^T \tag{4}$$

5. 輸出空間(Output space)：CMAC 的輸出是(3)式(4)式內的每個值相乘，最後加總起來，並表示為(5)式

$$y = \underline{\quad}^T \underline{b} = \sum_{j=1}^{N_R} \ _j b_j \tag{5}$$

如圖一(B)中，(State 點)的輸出值是接受域(Bb, Dd, Ff, Gg)乘上相對應的權重的總和。

三、自適性 CMAC 的學習算法

CMAC 的學習算法是考慮如何獲得梯度向量，在每個調節值的學習算法被定義為目標函數(Objective function)相對於輸入參數的導數，目標函數表示為(6)式

$$En(k) = \frac{1}{2}(d(k) - y(k))^2 = \frac{1}{2}e^2(k) \tag{6}$$

其中誤差訊號$e(k) = d(k) - y(k)$，表示所希望的響應$d(k)$和濾波器輸出$y(k)$之間的誤差。在使用目標函數En時，根據歸一化梯度下降法可以衍生(7)式，使用連鎖律(Chain rule)方法獲得。

$$s(k+1) = s(k) + \mu_s e(k) P_s(k) \tag{7}$$

其中μ_s是學習率(Learning rate)，在(7)式中s可替換成 $, m, \sigma$，分別代表是權重、平均值、變異數的更新法，$P_s(k)$在(7)式中可以替換為

$$P_w(k) = \frac{\partial y}{\partial \_{j}} = [\frac{\partial y}{\partial \_{1}}, \cdots, \frac{\partial y}{\partial \_{j}}, \cdots, \frac{\partial y}{\partial \_{N_R}}]^T \tag{8}$$

$$P_m(k) = \frac{\partial y}{\partial m_{ij}} = [\frac{\partial y}{\partial m_{11}}, \cdots, \frac{\partial y}{\partial m_{N1}}, \cdots, \frac{\partial y}{\partial m_{1j}}, \cdots, \frac{\partial y}{\partial m_{Nj}}, \cdots, \frac{\partial y}{\partial m_{1N_R}}, \cdots, \frac{\partial y}{\partial m_{NN_R}}]^T \tag{9}$$

$$P_\sigma(k) = \frac{\partial y}{\partial \sigma_{ij}} = [\frac{\partial y}{\partial \sigma_{11}}, \cdots, \frac{\partial y}{\partial \sigma_{N1}}, \cdots, \frac{\partial y}{\partial \sigma_{1j}}, \cdots, \frac{\partial y}{\partial \sigma_{Nj}}, \cdots, \frac{\partial y}{\partial \sigma_{1N_R}}, \cdots, \frac{\partial y}{\partial \sigma_{NN_R}}]^T \tag{10}$$

最後$P_s(k)$可以推導成以下式子

$$\frac{\partial y}{\partial \_{j}} = b_j \tag{11}$$

$$\frac{\partial y}{\partial m_{ij}} = \_{j} b_j \frac{2(x_i - m_{ij})}{(\sigma_{ij})^2} \tag{12}$$

$$\frac{\partial y}{\partial \sigma_{ij}} = \_{j} b_j \frac{2(x_i - m_{ij})^2}{(\sigma_{ij})^3} \tag{13}$$

四、實驗與評估

(一)、評估方法

在評估方面，我們用了三種語音評估方法來做 CMAC 及 MMSE 消除雜音的數值比較，分別為(Perceptual Evaluation of Speech Quality, PESQ)、(Segmental Signal-to-Noise Ratio, SSNR)以及(Speech Distortion Index, SDI)。

首先將簡單介紹這三種評估方法：

1. Perceptual Evaluation of Speech Quality (PESQ)的評價方法是以國際電信聯盟(ITU-T)標準為基礎，為一套客觀評價語音品質的方法，比較方法是比較"增強後語音"與"原始乾淨語音"之間的差異，PESQ 的分數範圍為 0.5 到 4.5 分，分數越高代表越接近原始乾淨語音。在本實驗是將"增強後語音的 PESQ"與"帶雜訊語音的 PESQ"相減，觀察語音品質的增加量，即分數越高越好。PESQ 可以表示為(14)式

   $$\Delta PESQ = PESQ_{en} - PESQ_{noise} \tag{14}$$

   其中$PESQ_{en}$是增強後語音的 PESQ，$PESQ_{noise}$是帶雜訊語音的 PESQ。
2. Segmental Signal-to-Noise Ratio (SSNR)為分段式訊號功率與雜訊功率的比，即點對點的差。本實驗是將"增強後語音的 SSNR"與"帶雜訊語音的 SSNR"相減，觀察語音

SNR 增加量，即分數越高越好。SSNR 可以表示為(15)式

$$\Delta SSNR = \frac{P_{clean}}{P_{en}} - \frac{P_{clean}}{P_{noise}} = \frac{A_{clean}^2}{A_{en}^2} - \frac{A_{clean}^2}{A_{noise}^2} \quad (15)$$

其中P_{clean}為乾淨語音功率，P_{en}為增強後語音功率，P_{noise}為帶雜訊語音功率，A_{clean}為乾淨語音振幅，以此類推。

3. Speech Distortion Index (SDI)是比較"增強後語音訊號"與"原始乾淨語音訊號"的能量差值，即計算增強後語音的失真量，本實驗是將"帶雜訊語音的 SDI"與"增強後語音的 SDI"相減，觀察語音失真值減少量，即分數越高越好。SDI 可以表示為(16)式

$$\Delta SDI = \frac{E[(S_{clean}[n] - S_{noise}[n])^2]}{E[S_{clean}^2[n]]} - \frac{E[(S_{clean}[n] - S_{en}[n])^2]}{E[S_{clean}^2[n]]} \quad (16)$$

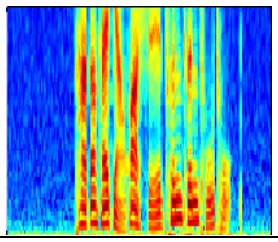
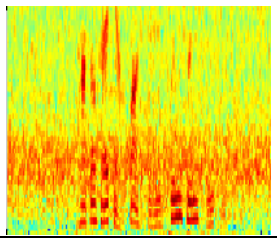
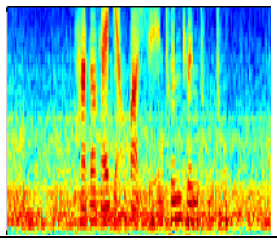
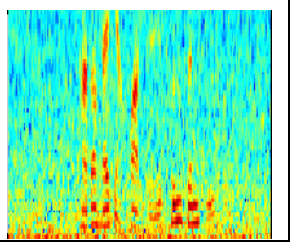
其中$S_{clean}[n]$為原乾淨語音訊號，$S_{noise}[n]$為帶雜訊語音訊號，$S_{en}[n]$為增強後語音訊號。

(二)、實驗方法

在音源庫方面，我們用了三種不同的環境配合六種不同的訊雜比(SNR)，環境分別有仿人聲雜音(SSN)、車子雜音(Car)、粉紅雜音(Pink)，訊雜比分別有-5dB,0dB,5dB,10dB,15dB,20dB，總共 18 種不同的環境做語音增強。乾淨語音方面，我們使用 300 個相同語者而不同語音內容的音檔。且帶雜訊語音與乾淨語音每個音檔有 3 秒鐘，取樣率(Sampling rate)均為 8K。在製做語音時，我們把乾淨語音及帶雜訊語音先做正規化，如需 5dB 時，就把乾淨語音能量增強 5dB 與帶雜訊語音做結合；如需 10dB 時，就把乾淨語音能量增強 5dB 與帶雜訊語音能量降低 5dB 做結合。最後有 18×300 個帶雜訊語音音檔及 300 個乾淨語音音檔。

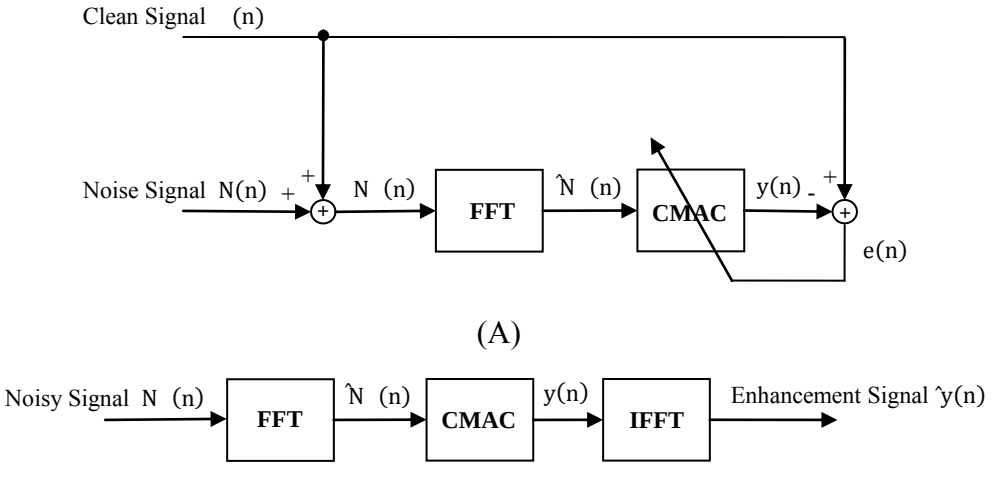
三種雜音類型：

1. 仿人聲雜音(SSN)：能量分佈平均，但在中頻有較高的能量分佈。
2. 車子雜音(Car)：在低頻有較高的能量，越往高頻能量分佈遞減。
3. 粉紅雜音(Pink)：能量分佈平均，但在低頻有較高的能量分佈。

頻譜圖(0dB)			
乾淨語音(Clean)	仿人聲雜音(SSN)	車子雜音(Car)	粉紅雜音(Pink)
			

圖二、實驗語音頻譜圖(0dB)，顏色為藍色代表無能量，顏色越紅代表能量越大。

在實驗方面，先把乾淨語音及帶雜訊語音取音框(Framing)，因為語音訊號是連續時變(Time-varying)，取完音框後，可將語音訊號視為一個固定週期的訊號，以利於處理，本實驗中我們的音框是 32 毫秒(256/8K)。而後將每個音框的訊號乘上一個固定長度的視窗(Hamming Window)，主要的目的為強調視窗中間的主要訊號，並壓抑視窗兩側的訊號。之後將帶雜訊語音訊號做快速傅立葉轉換(Fast Fourier Transform, FFT)，得到 256 個值，256 個值再轉到梅爾頻率域(Mel-frequency domain)上壓縮成 80 個值。實驗時，使用其中 250 個帶雜訊語音做訓練語料(Training)，另外 50 個帶雜訊語音做測試語料(Test)。訓練時，將 250 個帶雜訊語音串在一起(同環境且同訊雜比的音檔)成數據庫，而後隨機抽取其中 80000 點當訓練數據，因為語音是二維的，所以總共有 80(頻域)×80000(訓練數據)點的訓練數據，乾淨語音則沒有不同訊雜比的狀況，但處理亦相同，同樣有 80(頻域)×80000(訓練數據)點的訓練數據，帶雜訊語音及乾淨語音的所有點是互相對應，點對點做學習，每一個頻率學習出一組 CMAC Model，總共學習出 80 組 CMAC Model，每一組 CMAC Model 用 80000 筆訓練數據學習，CMAC Model 內的資訊有高斯函數的平均值(m_{ij})、變異數(σ_{ij})以及權重值($_j$)。測試時，50 個帶雜訊語音同樣使用快速傅立葉轉換(Fast Fourier Transform, FFT)，頻域壓縮成 80，而後輸入對應頻率上的 CMAC Model 後將會消除雜音還原成乾淨訊號，再經由快速傅立葉逆轉換(Inverse Fast Fourier Transform, IFFT)轉回時域上，即可得到增強後的乾淨語音訊號。圖三為 CMAC 語音增強系統方塊圖，N (n)為乾淨語音訊號加上雜訊訊號，$\hat{N}$ (n)為帶雜訊語音訊號經由 FFT 後的訊號，y(n)為 CMAC 輸出訊號，e(n) = (n) − y(n)為誤差訊號，如果e(n)為零代表 CMAC 輸出訊號等於乾淨語音訊號，$\hat{y}$(n)為 CMAC 輸出訊號經由 IFFT 後還原的訊號。



圖三、CMAC 語音增強系統方塊圖 (A)訓練 (B)測試

(三)、CMAC 與 MMSE 方法比較

將實驗中處理效果最好的 CMAC 設定與 MMSE 方法做比較。
在本實驗中 CMAC 特徵如下：

1. 層數(Layer)：3(Layer)
2. 上界(Upper bound)：6；下界(Lower bound)：-6
3. 一層內的塊數(N_B)：$\text{ceil}(106N_e/3\text{Layer}) = 36$
4. 接受域數(N_R)：塊數(N_B)
5. 聯想記憶空間函數：$\varphi_{ij} = \exp\left[-(x_i - m_{ij})^2/\sigma_{ij}^2\right]$ for $i = 1$ and $j = 1, \cdots, N_R$

其中 ceil 代表餘數無條件進位。上界和下界需要包含所有語音訊號參數，事前要先偵測語音訊號參數的範圍。高斯函數的平均值初始值(m_{ij})設置是自動調整在每塊(N_B)的正中間，變異數初始值(σ_{ij}) $= 1$，權重初始值($_j$) $= 0$。學習率$\mu_s = \mu_w = \mu_m = \mu_\sigma = 0.05$。
表一至表三為 CMAC 與 MMSE 方法使用三種語音評估方法比較效果。

表一、 CMAC 方法及 MMSE 方法的ΔPESQ 效果比較

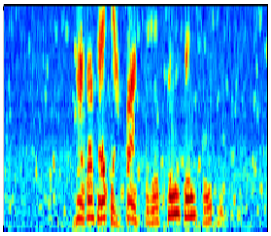
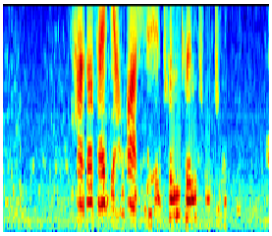
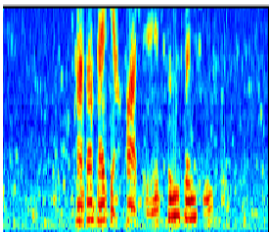
Evaluations	ΔPESQ					
SNR (dB)	SSN noise		Car noise		Pink noise	
	CMAC	MMSE	CMAC	MMSE	CMAC	MMSE
-5	**0.141**	0.006	**0.492**	0.228	**0.442**	0.061
0	**0.387**	-0.263	**0.511**	0.291	**0.679**	0.329
5	**0.678**	-0.308	**0.552**	0.315	**0.788**	0.467
10	**0.808**	0.041	**0.576**	0.339	**0.800**	0.484
15	**0.852**	0.152	**0.559**	0.352	**0.776**	0.456
20	**0.820**	0.123	**0.532**	0.289	**0.687**	0.415
Ave.	**0.614**	-0.042	**0.537**	0.302	**0.695**	0.369

表二、CMAC 方法及 MMSE 方法的ΔSSNR 效果比較

Evaluations	ΔSSNR					
SNR (dB)	SSN noise		Car noise		Pink noise	
	CMAC	MMSE	CMAC	MMSE	CMAC	MMSE
-5	**14.618**	4.635	**13.221**	7.703	**12.619**	8.021
0	**13.523**	4.344	**12.728**	7.173	**10.919**	7.400
5	**11.550**	3.811	**10.528**	6.262	**8.843**	6.437
10	**9.536**	3.015	**8.806**	5.034	**7.583**	5.066
15	**8.061**	2.020	**7.125**	3.597	**5.957**	3.429
20	**6.228**	0.940	**6.145**	2.014	**4.134**	1.716
Ave.	**10.586**	3.128	**9.759**	5.297	**8.343**	5.345

表三、 CMAC 方法及 MMSE 方法的ΔSDI 效果比較

Evaluations	ΔSDI					
SNR (dB)	SSN noise		Car noise		Pink noise	
	CMAC	MMSE	CMAC	MMSE	CMAC	MMSE
-5	**1.680**	0.110	**1.223**	0.118	**1.008**	0.051
0	**0.717**	0.063	**0.715**	0.046	**0.381**	0.017
5	**0.244**	0.023	**0.216**	-0.008	**0.120**	-0.009
10	**0.070**	0.003	**0.060**	-0.023	**0.030**	-0.022
15	**0.016**	-0.004	**0.012**	-0.022	**0.003**	-0.022
20	**-0.001**	-0.005	**-0.001**	-0.017	**-0.005**	-0.018
Ave.	**0.454**	0.032	**0.371**	0.016	**0.256**	-0.001

頻譜圖(Enhancement 0dB)		
仿人聲雜音(SSN)	車子雜音(Car)	粉紅雜音(Pink)
		

圖四、增強後語音頻譜圖(0dB)

在ΔSDI 評估方法中，可以觀察到三種雜音在 20dB 處有失真量增大的情形，但在ΔSSNR 評估方法中，20dB 處能有效提升訊雜比，由此可知每種評估方法量測的準則不一樣。CMAC 方法皆能達到比 MMSE 較佳的結果，在品質(ΔPESQ)中仿人聲雜音(SSN)平均提升 0.656、車子雜音(Car)平均提升 0.235、粉紅雜音(Pink)平均提升 0.326。在訊雜比(ΔSSNR)中仿人聲雜音(SSN)平均提升 7.458dB、車子雜音(Car)平均提升 4.462dB、粉紅雜音(Pink)平均提升 2.998dB。在失真量(ΔSDI)中仿人聲雜音(SSN)平均減少 0.422、車子雜音(Car)平均減少 0.355、粉紅雜音(Pink)平均減少 0.257。圖四為 SNR 在 0dB 時，語音增強後的頻譜圖，比較圖二可以看出 CMAC 語音增強系統在對付噪音有明顯改善。

(四)、同時學習所有訊雜比

本實驗目的在於實際應用中，我們無法得知當下環境的訊雜比(dB)，所以本實驗是同時學習同個環境中所有訊雜比(dB)的雜音，觀察在三種環境中的語音增強效果。
在本實驗中 CMAC 特徵如下：

1. 層數(Layer)：3(Layer)
2. 上界(Upper bound)：6；下界(Lower bound)：-6
3. 一層內的塊數(N_B)：$\text{ceil}(106N_e/3\text{Layer}) = 36$
4. 接受域數(N_R)：塊數(N_B)
5. 聯想記憶空間函數：$\varphi_{ij} = \exp\left[-(x_i - m_{ij})^2/\sigma_{ij}^2\right]$ for $i = 1$ and $j = 1, \cdots, N_R$

其中 ceil 代表餘數無條件進位。上界和下界需要包含所有語音訊號參數，事前要先偵測語音訊號參數的範圍。高斯函數的平均值初始值(m_{ij})設置是自動調整在每塊(N_B)的正中間，變異數初始值(σ_{ij}) $= 1$，權重初始值($_j$) $= 0$。學習率$\mu_s = \mu_w = \mu_m = \mu_\sigma = 0.05$。表四至表六為 CMAC 方法的三種語音評估數據。

表四、三種雜音的ΔPESQ 效果比較

ΔPESQ			
SNR(dB)	SSN noise	Car noise	Pink noise
-5	-0.185	0.570	0.375
0	0.134	0.539	0.568
5	0.299	0.478	0.566
10	0.295	0.333	0.449
15	0.156	0.112	0.284
20	-0.006	-0.169	0.102
Ave.	0.116	0.310	0.391

表五、三種雜音的ΔSSNR 效果比較

ΔSSNR			
SNR(dB)	SSN noise	Car noise	Pink noise
-5	11.919	11.920	9.216
0	12.325	12.196	9.626
5	11.217	9.908	8.637
10	8.932	6.829	6.375
15	5.438	3.183	3.216
20	1.268	-0.825	-0.543
Ave.	8.517	7.202	6.088

表六、三種雜音的ΔSDI 效果比較

ΔSDI			
SNR(dB)	SSN noise	Car noise	Pink noise
-5	1.567	1.197	0.957
0	0.645	0.690	0.362
5	0.198	0.169	0.097
10	0.007	-0.035	-0.014
15	-0.094	-0.125	-0.058
20	-0.165	-0.169	-0.084
Ave.	0.360	0.288	0.210

在 SDI 評估方法中，可以觀察到在較高 SNR 情況下有失真的情形，明顯降低處理效果，因為背景雜音差異量太大，CMAC 無法適應到所有 SNR(dB)均適合的轉移函數，但比較表一至表三中 MMSE 方法的實驗數據，CMAC 方法還是略贏 MMSE 方法。

五、結論

本文我們提出一個 CMAC 語音增強系統，以消除語音訊號的背景雜音，在此我們研究 CMAC 方法在不同類型的環境雜音中的處理能力，以及 CMAC 架構中數值設定的規範。根據歸一化梯度下降法增加了 CMAC 參數學習速度。為了更穩定加快學習速度，如自適性的學習率將是我們今後的研究。在低訊雜比(dB)的情況下，ΔPESQ、ΔSSNR 及ΔSDI 語音頻估方法均可以看出有較佳的處理效能。我們進一步與 MMSE 相比，在不同類型的環境雜音中，CMAC 方法均有較佳的結果。

參考文獻

[1] T. Venema, *Compression for Clinicians*, Thomson Delmar Learning, 2006, Chapter 7.

[2] H. Levitt, "Noise reduction in hearing aids: An overview," *Journal of Rehabilitation Research and Development*, vol. 38, pp. 111-121, 2001.

[3] J. Chen, *Fundamentals of Noise Reduction*, Springer Handbook of Speech Processing, 2008, Chapter 43.

[4] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions, Acoustics, Speech and Signal Processing*, vol. 27, pp. 113-120, 1979.

[5] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *Proceedings ICASSP*, pp. 629-632, 1996.

[6] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *ELSEVIER, Speech Communication*, vol. 50, pp. 453-466, 2008.

[7] J. Li, S. Sakamoto, S. Hongo, M. Akagi and Y. Suzuki, "Adaptive β-order generalized spectral subtraction for speech enhancement," *ELSEVIER, Signal Processing*, vol. 88, pp. 2764-2776, 2008.
[8] Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions, Speech and Audio Processing*, vol. 3, pp. 251-266, 1995.
[9] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Transactions, Speech and Audio Processing*, vol. 8, pp. 159-167, 2000.
[10] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions, Acoustics, Speech and Signal Processing*, vol. 32, pp. 1109-1121, 1984.
[11] I. Y. Soon, S. N. Koh and C. K. Yeo, "Improved noise suppression filter using self-adaptive estimator of probability of speech absence," *ELSEVIER, Signal Processing*, vol. 75, pp. 151-159, 1999.
[12] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Transactions, Speech and Audio Processing*, vol. 13, pp. 845-856, 2005.
[13] J. H. L. Hansen, V. Radhakrishnan and K. H. Arehart, "Speech enhancement based on generalized minimum mean square error estimators and masking properties of the auditory system," *IEEE Transactions, Audio, Speech, and Language Processing*, vol. 14, pp. 2049-2063, 2006.
[14] D. Malah, R. V. Cox and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement non-stationary noise environments," *Proceedings ICASSP*, pp. 789-792, 1999.
[15] E. Plourde and B. Champagne, "Auditory-based spectral amplitude estimators for speech enhancement," *IEEE Transactions, Audio, Speech, and Language Processing*, vol. 16, pp. 1614-1622, 2008.
[16] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP, Applied Signal Processing*, vol. 7, pp. 1110-1126, 2005.
[17] S. Suhadi, C. Last and T. Fingscheidt, "A data-driven approach to a priori SNR estimation," *IEEE Transactions, Audio, Speech, and Language Processing*, vol. 19, pp. 186-195, 2011.
[18] Z. Xin, P. Jancovic, L. Ju and M. Kokuer, "Speech signal enhancement based on MAP algorithm in the ICA space," *IEEE Transactions, Signal Processing*, vol. 56, pp. 1812-1820, 2008.

[19] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions, Acoustics, Speech and Signal Processing*, vol. 28, pp. 137-145, 1980.
[20] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," *Proceedings EUSIPCO*, pp. 295-299, 2012.
[21] R. Martin, "Spectral subtraction based on minimum statistics," *Proceedings EUSIPCO*, pp. 1182-1185, 1994.
[22] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions, Speech and Audio Processing*, vol. 9, pp. 504-512, 2001.
[23] Y. Xu, J. Du, L.-R. Dai and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Processing Letters*, vol. 21, pp. 65-68, 2014.
[24] X. Lu, Y. Tsao, S. Matsuda and C. Hori, "Speech enhancement based on deep denoising autoencoder," *Interspeech 2013*, pp. 436-440, 2013.
[25] C. D. Sigg, T. Dikk and J. M. Buhmann, "Speech enhancement using generative dictionary learning," *IEEE Transactions, Audio, Speech, and Language Processing*, vol. 20, pp. 1698-1712, 2012.
[26] K. Wilson, B. Raj, S. Paris and A. Divakaran, "Speech denoising using nonnegative matrix factorization with priors," *Proceedings ICASSP*, pp. 4029-4032, 2008.
[27]R.-J. Wai, C.-M. Lin and Y.-F. Peng, "Adaptive hybrid control for linear piezoelectric ceramic motor drive using diagonal recurrent CMAC network," *IEEE Transactions, Neural Networks*, vol. 15, pp. 1491-1506, 2004.
[28]C.-M. Lin and T.-Y. Chen, "Self-Organizing CMAC Control for a Class of MIMO Uncertain Nonlinear Systems," *IEEE Transactions, Neural Networks*, vol. 20, pp. 1377-1384, 2009.
[29]C.-M. Lin, L.-Y. Chen and C.-H. Chen, "RCMAC hybrid control for MIMO uncertain nonlinear systems using sliding-mode technology," *IEEE Transactions, Neural Networks*, vol. 18, pp. 708-720, 2007.
[30]C.-M. Lin, L.-Y. Chen and D. S. Yeung, "Adaptive filter design using recurrent cerebellar model articulation controller," *IEEE Transactions, Neural Networks*, vol. 19, pp. 1149-1157, 2010.
[31]J. S. Albus, "A new approach to manipulator control: the cerebellar model articulation controller (CMAC)," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 97, pp. 228-233, 1975.