

行政院國家科學委員會補助專題研究計畫

期中精簡報告

免持聽筒汽車套件之迴聲與噪音控制系統

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 97-2221-E-009-010-MY3

執行期間： 97 年 8 月 1 日至 100 年 7 月 31 日

計畫主持人：白明憲

共同主持人：

計畫參與人員： 林家鴻、廖國志、劉嫻婷

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

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：國立交通大學 機械工程研究所

中文摘要

本計劃第二年度提出一種能夠實現在電信通訊系統中的麥克風陣列技術運用聲學信號處理方法，此技術稱為聲源等值反濾波器設計演算法。單進多出聲源等值反濾波器設計演算法(SIMO-ESIF)的目的為在充滿迴響的環境裡能夠重建語音訊號，此系統能夠達到兩個重要的目標：抑制殘響和消除噪音。其適用的電信通訊系統如車內免持聽筒的系統，在密閉的車子環境裡所收到的語音通常需要被改善因其夾雜著許多背景噪音，此演算法結合提出的 GSC 演算法是為了進一步在更嚴重迴響的環境裡改善噪音消除的效果。主觀測試的結果用變異數分析方法來做為分析的工具，進一步使用 Fisher's LSD 分析法來證明新提出的方法在改善含有噪音的語音訊號上效果有明顯的進步並且提供更棒的音質。

英文摘要

New microphone array techniques are proposed for acoustic signal processing in telecommunication application. These endeavors are based on the central idea of Equivalent Source Inverse Filtering (ESIF). The single input multiple output equivalence source imaging (SIMO-ESI) algorithms are suggested to reconstruct the speech signal in a reverberant environment. Specifically, the system serves two purposes: dereverberation and noise reduction. It has promise in telecommunication application such as the automotive hands-free system, where noise-corrupted speech signal often needs to be enhanced. In order to further improve the noise reduction performance in spatial filtering and robustness against system uncertainties, the SIMO-ESIF algorithm is combined with an adaptive Generalized Side-lobe Canceller (GSC). The system is implemented on an NI-PXI platform and evaluated experimentally in car environment. As indicated by several performance measures in noise reduction and speech distortion, the proposed microphone array algorithm proved effective in reducing noise in human speech without significantly compromising the speech quality. The results of subjective tests were processed by using analysis of variance (ANOVA) to justify the statistic significance. A post-hoc test Fisher's LSD was conducted to further assess the pairwise difference between the NR algorithms.

一、文獻探討

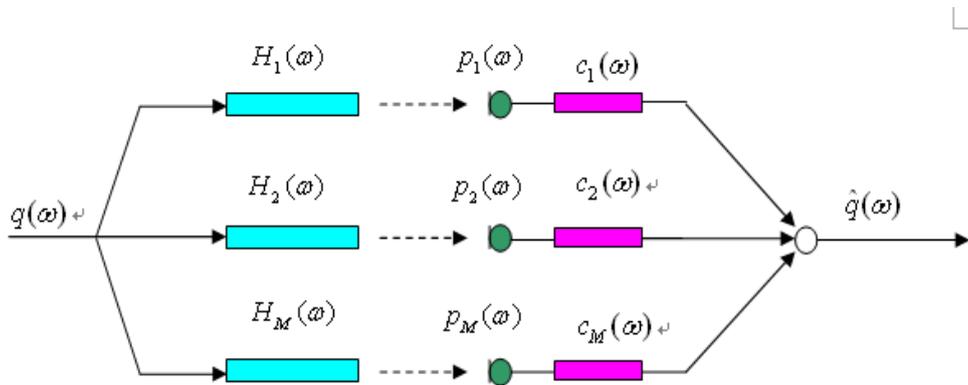
近年來，麥克風陣列被廣泛的應用於通訊系統、語音辨識、語音訊號增強等，在這些應用中，能在噪音環境中有效率的通訊變成的非常重要，也是目前關注的問題。傳統的延遲相加法(delay-and-sum-beamformer, DAS)，被廣泛的使用於辨識率系統和減噪等應用中，但是此種 DAS 方法只能對抗和訊號無關的噪音，若是有關聯的噪音則效果便不佳。而標準的超指向性波束形成器方法是另一對抗此種問題的技術，其對於擴散噪音(diffuse noise)的表現較佳，但是以上兩種方法都是針對減噪部份，對於殘響的效果非常有限。

在某些環境中(如車上)，語音訊號常常被背景噪音以及嚴重的殘響所遮蔽，對於此種情形適應性的波束形成器(adaptive beamforming)是有希望能將其改善的。此種方法早在 1960 年代便被推薦用於聲納、雷達等訊號處理，以改善其效果。此後很快的便發展出許多不同的公式來最佳地偵測和評估問題，其基礎的概念是使用量測的背景空間相關特徵來抑制噪音和干擾，進而增進輸出的訊噪比(signal-to-noise ratio, SNR)。廣義側葉消除器(generalized sidelobe canceller, GSC)是其中一種適應性的波束形成器，其能使用很少數量的麥克風、在一很小的空間中也能有很好的干擾減少效果，它對於房間的殘響、誤差等非常的敏感，這些因素會使目標訊號被消除或失真，因此利用 steering vector 來接收從設計方向過來的訊號。目前很多強健的適應性波束形成器技術已經被應用於防止訊號被消除。Griffiths-Jim beamformer (GJBF)是一種在 GSC 基礎下的適應性波束形成器，但其目標訊號在 steering-vector 有誤差時會被消除，上述誤差是由於麥克風位置、麥克風增益值、殘響、和目標位置造成的。

在本研究中，一種新的麥克風陣列技術被推薦用於在通訊系統應用中的聲音訊號處理。其構想是從等效聲源反濾波器(equivalent source inverse filtering, ESIF)中所得到的，此種方法是用於抑制噪音。將這種技術用於單近多出(single-input-multiple-output, SIMO)結構上，加上其濾波的設計是基於量測從聲源

到麥克風之間的系統響應，此種量測的響應包含了直接傳遞而來的聲源和由邊界反射的訊號，能夠同時達到減噪和消除殘響的效果。此技術另一個特徵是利用 GSC 來更進一步的增進其表現效果。此外，本研究亦將利用一多聲道的訊號系統在車上實現，使用客觀的目標函數來測試評估此演算法，並做主觀性的聆聽實驗來評估其表現，並利用 MANOVA 來分析數據。

二、等效聲源反濾波器演算法 (ESIF)



圖一、SIMO-ESIF 架構圖

本章節介紹 SIMO-ESIF 演算法的主要概念。圖一即為此為單進多出(single input multi-output)聲源等值反濾波器(Equivalent source inverse filtering)演算法的基本架構圖，在圖中有 M 個麥克風接收從遠場傳遞而來的聲源，其中 $q(\omega)$ 為在空間中一固定的聲源， $H(\omega)$ 為聲源到麥克風之間的系統響應(plant)，其系統響應可經由頻譜分析儀所量測得到。實驗的架構為一聲源經由 plant 後被麥克風陣列 $p(\omega)$ 所收到，在頻率域的代表法如下，

$$\mathbf{p} = \mathbf{H}q \quad (1)$$

其中 $\mathbf{p} = [p_1(\omega) \cdots p_M(\omega)]^T$ 是聲壓向量的轉置矩陣， $\mathbf{H} = [H_1(\omega) \cdots H_M(\omega)]^T$ 是一傳遞矩陣，在此主要的目標是利用量測到的聲壓和經由一組反濾器來估計聲源訊號 $q(\omega)$ ，此組反濾器為

$$\mathbf{C} = [c_1(\omega) \cdots c_M(\omega)]^T \quad (2)$$

反濾器必須滿足式子 $\mathbf{C}^T \mathbf{H} \approx \mathbf{I}$ ，因此

$$\hat{q} = \mathbf{C}^T \mathbf{p} = \mathbf{C}^T \mathbf{H}q \approx q \quad (3)$$

也就是說，聲源可由求解以下最佳化的問題而估測出來，

$$\min_q \|\mathbf{p} - \mathbf{H}q\|_2^2 \quad (4)$$

其中 $\|\cdot\|_2$ 是代表向量的 2-norm，這是一個 over-determined 的問題，因此其最小平方解為：

$$\hat{q} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{p} = \frac{\mathbf{H}^H \mathbf{p}}{\|\mathbf{H}\|_2^2} = \mathbf{c}^T \mathbf{p} \quad (5)$$

其中“ H ”代表 hermitian transpose，比較(3)式和(5)式可得到如下最佳的反濾波器：

$$\mathbf{c}^T = \frac{\mathbf{H}^H}{\|\mathbf{H}\|_2^2} \quad (6)$$

假如我們省略分母項 $\|\mathbf{H}\|_2^2$ ，則最佳的反濾波器在頻率域上來看就變成了“相位共軛 (phase-conjugated)”濾波器，而在時間域上可視為將脈衝響應做翻轉而成為“時間翻轉 (time-reversal)”濾波器。明確地說，對於一個在自由聲場的點聲源來說，其可寫成：

$$\|\mathbf{H}\|_2^2 = \sum_{m=1}^M \frac{1}{r_m^2} \quad (7)$$

其中 r_m 代表第 m 個麥克風到聲源之間的距離，由於此式和頻率無關，因此可了解到反濾波器和時間翻轉濾波器之間只有一個常數項的倍數關係。另外，在一有殘響的環境中，這兩種濾波器則不相同。由於這種演算法使用量測到的系統響應，包含一些殘響的特徵，因此和傳統的 DAS 方法比起來對抗干擾的效果甚佳。

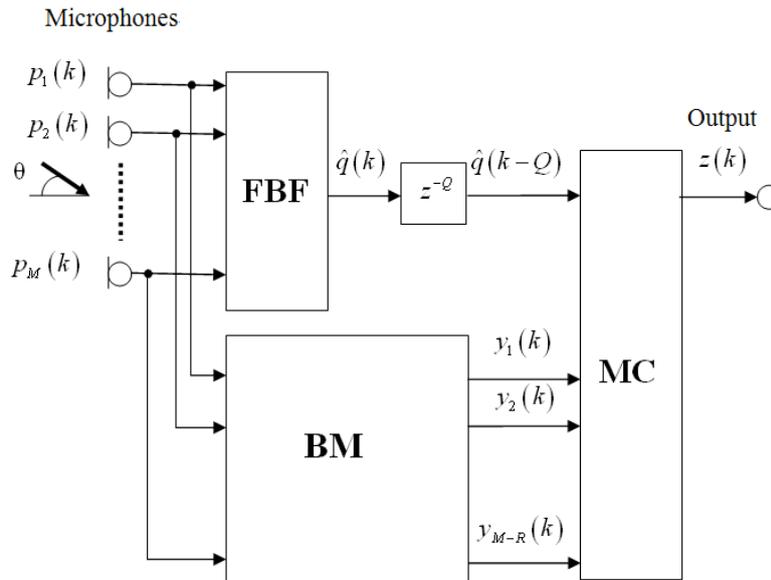
在實際做實驗時，這些計算出來的反濾波器必須轉回時域的 FIR 濾波器。因此必須將頻率域上設計出最佳的反濾波器，經由反快速傅立葉轉換(IFFT)及適當的平移而得到。如此一來聲源訊號便可由下列式子計算出：

$$\hat{q}(k) = \mathbf{c}^T(k) * \mathbf{p}(k), \quad (8)$$

其中 k 代表離散時間參數， $\mathbf{c}(k)$ 是反濾波器的脈衝響應，“*”代表迴旋積分。

三、SIMO-ESIF 結合 GSC (Generalized Side-lobe Canceller)

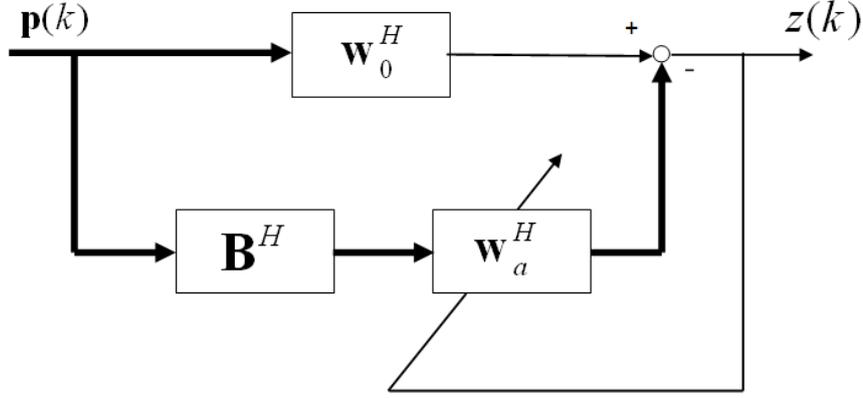
上一章節所提到的 SIMO-ESIF 演算法可以經由加上 GSC 至系統中來更進一步的提昇其效能，使用此方法有雙倍的好處：利用抑制由 side-lobe 洩漏造成的干擾，能增加陣列的指向性，在面對不確定因素以及擾動的時候，能強健麥克風陣列，GSC 的基本架構圖如圖二所示：



圖二、GSC 基本架構圖

SIMO-ESIF 結合 GSC 的設計結構包含下列三個項目，fixed beamformer (FBF)，a multiple-input canceller (MC)，和 a blocking matrix (BM)。FBF 的設計目的是要在向著聲源的方向合成一個波束 (beam)，能讓主要的訊號通過並減低其餘的干擾雜訊。其中 $p_m(k)$ 是第 m 個麥克風收到的訊號， $\hat{q}(k)$ 是 FBF 在時間函數 k 時的輸出。MC 是由多個適應性濾波器 (adaptive filter) 所組成的，其目的在於設計一個近似於干擾的濾波器。經過 BM 之後的輸出訊號 $y_m(k)$ 被從 FBF 而得的延遲輸出訊號 $\hat{q}(k-Q)$ 相減， Q 則代表延遲的數目。另一方面，BM 的設計目的正好與 FBF 相反，BM 為一個抑制器將主要的訊號消除而使其餘的訊號通過。

1. Blocking matrix 的公式



圖三、SIMO-ESIF 結合 GSC 結構圖

GSC 的目標為使陣列輸出的能量最小，而在主軸方向的值則保持不變。其示意圖如圖三所示，其中 $x(k)$ 是麥克風陣列收到的訊號， w_0 是前一章節所設計出來的最佳反濾波器，也就是 FBF； B 是 blocking matrix， w_a 是適應性濾波器， $y(k)$ 是輸出的訊號。上述條件可寫成下列式子：

$$\min_{\mathbf{w}} E\{ |z|^2 \} = \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{pp} \mathbf{w} \quad (9)$$

$$\text{其中 } \mathbf{H}^H \mathbf{w} = 1 \quad (10)$$

在上述式子裡 $\mathbf{R}_{pp} = E\{\mathbf{p}\mathbf{p}^H\}$ 表示 correlation matrix， \mathbf{H} 是聲源到每個麥克風之間的頻率響應， \mathbf{w} 是陣列濾波器的向量係數，這種有限制的最佳解問題能利用將最佳濾波器 \mathbf{w} 分解成兩個線性獨立且相互垂直的組成而將其變成無限制的問題。 \mathbf{w} 被分解成 constraint range space $R(\mathbf{H})$ 和 orthogonal null space $N(\mathbf{H}^H)$ ：

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{v}, \quad (11)$$

其中 $\mathbf{w}_0 \in R(\mathbf{H})$ 是一個固定濾波器， $\mathbf{v} = \mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H)$ 則是適應性濾波器。其中

$$\mathbf{h}^H \mathbf{w} = \mathbf{h}^H (\mathbf{w}_0 - \mathbf{B}\mathbf{w}_a) = \mathbf{h}^H \mathbf{w}_0 - \mathbf{h}^H \mathbf{B}\mathbf{w}_a \approx 1 \quad (12)$$

固定濾波器 \mathbf{w}_0 保證波束組成器中必要的部份被保存下來，此濾波器設計是 off-time 的，和 correlation matrix 無關，如此最小值便能在不影響上述限制條件

下計算出來。

傳統是使用 ad hoc blocking matrices，其是基於遠場平面波聲源的假設，也就是 $\mathbf{H} = [1 \ 1 \dots 1]\mathbf{H}$ ，因為 $\mathbf{H}^H \mathbf{B} = \mathbf{0}$ ，如果 \mathbf{B} 的各欄之和為 0，就能保證有 blocking 的效果。但在殘響環境中，這種 ad hoc blocking matrices 方法就不適合了。目前推薦的方法和一般方法最主要的區分為，現在這種 blocking matrix 能應用於更真實的環境。

為了達成 $\mathbf{B}\mathbf{w}_a \in N(\mathbf{H}^H) \Leftrightarrow \mathbf{H}^H \mathbf{B}\mathbf{w}_a = \mathbf{0}$ 的目標， \mathbf{B} 的各個欄必須由 $N(\mathbf{H}^H)$ 組成，使得 $\mathbf{H}^H \mathbf{B} = \mathbf{0}$ ，

$$\text{Let } \mathbf{H} = [a_1, a_2, \dots, a_n]^H, \quad \mathbf{x} = [x_1, x_2, \dots, x_n] \in N(\mathbf{H}^H) \quad (13)$$

$$\mathbf{H}^H \mathbf{x} = 0 \Rightarrow a_1 x_1 + a_2 x_2 + \dots + a_n x_n = 0 \quad (14)$$

$$\text{If } a_1 \neq 0, \quad x_1 = -\frac{a_2}{a_1} x_2 - \frac{a_3}{a_1} x_3 - \dots - \frac{a_n}{a_1} x_n \quad (15)$$

$$\text{Let } x_2 = \alpha_2, \quad x_3 = \alpha_3, \dots, x_n = \alpha_n \quad (16)$$

$$\Rightarrow x_1 = -\frac{a_2}{a_1} \alpha_2 - \frac{a_3}{a_1} \alpha_3 - \dots - \frac{a_n}{a_1} \alpha_n \quad (17)$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} = \alpha_2 \begin{bmatrix} -\frac{a_2}{a_1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} -\frac{a_3}{a_1} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + \alpha_n \begin{bmatrix} -\frac{a_n}{a_1} \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (18)$$

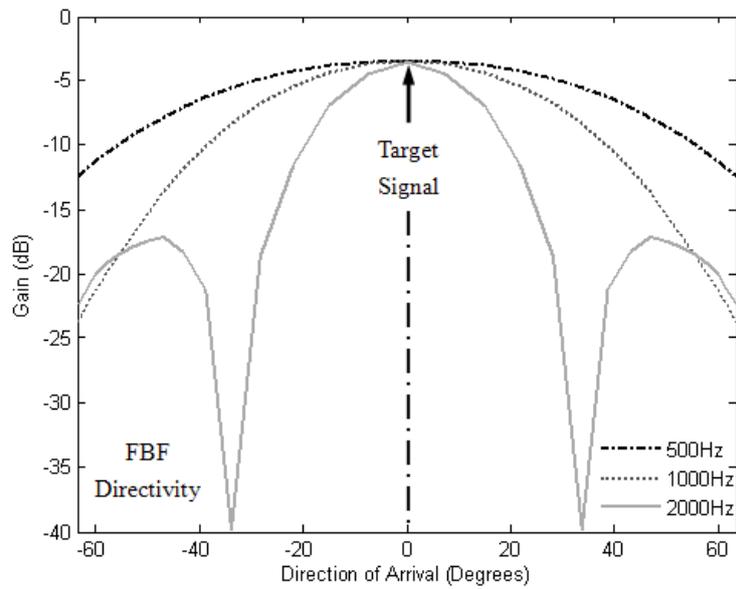
$\mathbf{v}_2 \qquad \mathbf{v}_3 \qquad \mathbf{v}_n$

由上式不難發現 $\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n$ 是線性獨立，且是由 null space $N(\mathbf{H}^H)$ 組成，因此矩陣

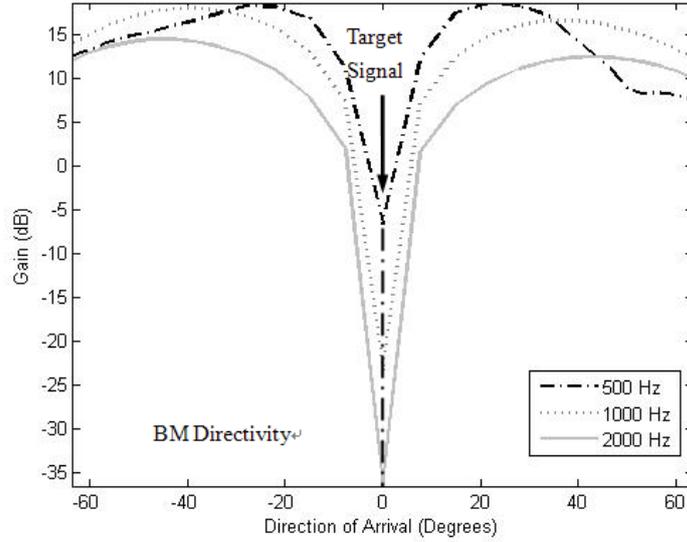
$\mathbf{B} = [\mathbf{v}_2 \ \mathbf{v}_3 \ \dots \ \mathbf{v}_n]$ 可用來當作 blocking matrix，即

$$\mathbf{B} = \begin{bmatrix} -\frac{a_2}{a_1} & \frac{a_3}{a_1} & \dots & -\frac{a_n}{a_1} \\ 1 & 0 & & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (19)$$

圖四是 FBF 和 BM 的 beam patterns，使用三種 sine 波訊號 500Hz, 1kHz 和 2kHz 來比較 BM 和 FBF 之效果。在目標方向 FBF 形成一主波束(mainlobe)，而 BM 則是有一個波谷，因此能進一步抑制從側葉(sidelobe)方向過來的噪音和雜訊，在實際實驗時，blocking matrix \mathbf{B} 需要利用 IFFT 和平移轉成脈衝響應。



圖四(a) FBF 的 beam pattern



圖四(b) BM 的 beam pattern

2. Multiple-Input Canceller

通常在 GSC 會使用比固定濾波器更強健的適應性濾波器，原本需要計算的 correlation matrix \mathbf{R}_{pp} 利用此演算法之後就不需考慮。例如洩漏適應性濾波器 (leaky adaptive filters, LAF) 就可用來當作 MC。LAFs 把 $\hat{q}(k-Q)$ 內和 $y_n(k)$, ($n = 0, \dots, N$) 有關的組成減去，其中 Q 是使系統 causality 的延遲， $\mathbf{w}_n(k)$ 和 $\mathbf{y}_n(k)$ 為第 n 個 LAF 的係數以及訊號向量，MC 的輸出則可寫成

$$z(k) = \hat{q}(k-Q) - \sum_{n=0}^{N-1} \mathbf{w}_n^T(k) \mathbf{y}_n(k) \quad (20)$$

$$\mathbf{w}_n(k) \triangleq [w_{n,0}(k), w_{n,1}(k), \dots, w_{n,M_2-1}(k)]^T \quad (21)$$

$$\mathbf{y}_n(k) \triangleq [y_n(k), y_n(k-1), \dots, y_n(k-M_2+1)]^T \quad (22)$$

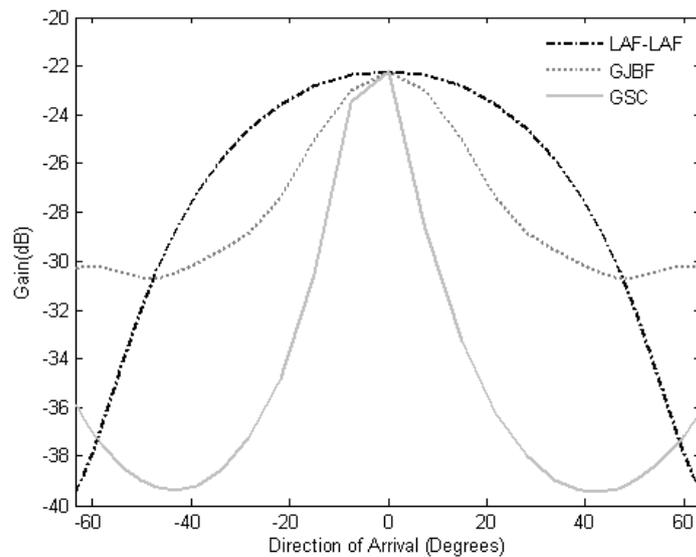
而適應性濾波器 MC 則運用 LMS (Least Mean Square method) 演算法來實現。其演算法如下式，

$$\mathbf{w}_n(k+1) = \mathbf{w}_n(k) + \mu z(k) \mathbf{y}_n(k) \quad (23)$$

M_2 為適應性濾波器的長度， μ 為 step size。

在圖五中，比較本研究所推薦的適應性 GSC 演算法和另外兩種常見的驗算法

—GJBF 以及 LAF-LAF 演算法在 500Hz 時的 beam pattern。GJBF 演算法是將相鄰的兩個麥克風收到的訊號相減來當成其 BM，而 LAF-LAF 演算法則是利用適應性濾波器來將阻止目標訊號。這兩種方法均是利用如式(18)的適應性演算法來獲得 MC。從圖五中可發現，本研究所推薦的適應性 GSC 演算法擁有最好的效果，其 beam 非常的陡峭且 sidelobe 非常的小。



圖五、比較 GJBF, LAF-LAF 和 SIMO-ESIF-GSC 在 500Hz 下的 beam pattern

四、客觀評估參數

在本章節中，將定義用來評估陣列表現的客觀量測參數。利用第一個麥克風當作參考，則輸入雜訊比(SNR)定義為：

$$SNR_1(\text{dB}) = 10 \log \frac{E\{x_1(k)^2\}}{E\{v_1(k)^2\}}, \quad (24)$$

其中 k 為離散時間因子， $x_1(k)$ 和 $v_1(k)$ 分別為第一個麥克風收到的語音訊號以及雜訊。而輸出雜訊比則可定義為

$$SNR_A(\text{dB}) = 10 \log \frac{E\{|\mathbf{c}(k)^T * \mathbf{x}(k)|^2\}}{E\{|\mathbf{c}(k)^T * \mathbf{v}(k)|^2\}}, \quad (25)$$

其中 $\mathbf{c}(k)$ 是反濾波器的脈衝響應，“*”表示迴旋積分。

因此，雜訊比增益值(SNR gain)可利用將輸出雜訊比減去輸入雜訊比而獲得：

$$SNRG(\text{dB}) = SNR_A - SNR_1 \quad (26)$$

由 $SNRG(\text{dB})$ 大小值可得知經過最佳濾波器處理後的訊號消噪程度為多少， $SNRG(\text{dB})$ 越大，則消除噪音的效果就越好。除了關心消噪程度的好壞，我們對於語音的失真與否也很在乎，因此定義了一個語音失真的指標 (speech-distortion index) 如下，

$$SDI(\text{dB}) = 10 \log \frac{E\{x_1^2\}}{E\{|x_1 - \mathbf{c}^T * \mathbf{x}|^2\}} \quad (27)$$

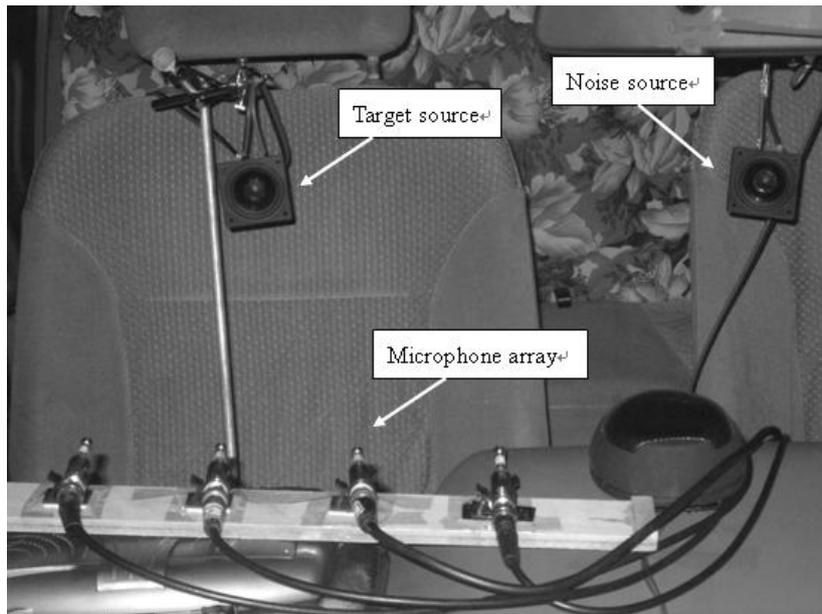
其值越大，則代表語音失真的程度越少。由於上述兩種參數是不可能同時獲得最大值的，因此陣列的目標為將這兩種參數折衷。

五、主客觀實驗評估

本研究所推薦的演算法將在一個 2000c.c.的汽車中作實驗，圖六為實驗配置圖片。陣列訊號處理演算法均在 National Instruments Labview 8.6 以及 NI-PXI 8105 系統下實現，取樣頻率為 8kHz，聲壓參數則利用四個線性等間距陣列麥克風(PCB 130D20)來接收，每個麥克風之間的距離為 8 公分，其中一揚聲器放在麥克風陣列中心前方 40 公分的地方，播放一男性語音訊號來當作目標訊號，而另一個揚聲器則放在距離麥克風陣列 30 公分，和主軸夾角 53 度的地方，播放白噪音(white noise)訊號來當作干擾。



圖六(a) 實驗車外觀



圖六(b) 實驗車內部架設圖

本研究使用主、客觀測試兩種方法來評估效果。使用 SIMO-ESIF 演算法來當作 FBF，而在 MC 所使用的適應性濾波器則選擇長度 512 點以及 step size $\mu = 0.001$ 。在 SIMO-ESIF 演算法中依據響應模式(plant model)以及在 FBF 中的濾波器模型(filtering model)不同而有許多種模式，歸納在表 I：

algorithm	acronym	description
SIMO-ESIF	PIF	Point source model-based inverse filtering
	MIF	Measured plant-based inverse filtering
	MTR	Measured plant-based time reversed filtering
SIMO-ESIF-GSC	GSC-PIF	Point source model-based inverse filtering
	GSC-MIF	Measured plant-based inverse filtering
	GSC-MTR	Measured plant-based time reversed filtering

表 I、各種 SIMO-ESIF 模式

其中有兩種 plant model，自由聲場中的點聲源以及在車中量測到的系統響應被使用來計算反濾波器；有兩種濾波器模式：反濾波器和時間翻轉濾波器被用來作為 FBF 的設計。此外，三種不同的 GSC 方法亦在表 I 中清楚表示。

1. 客觀評估

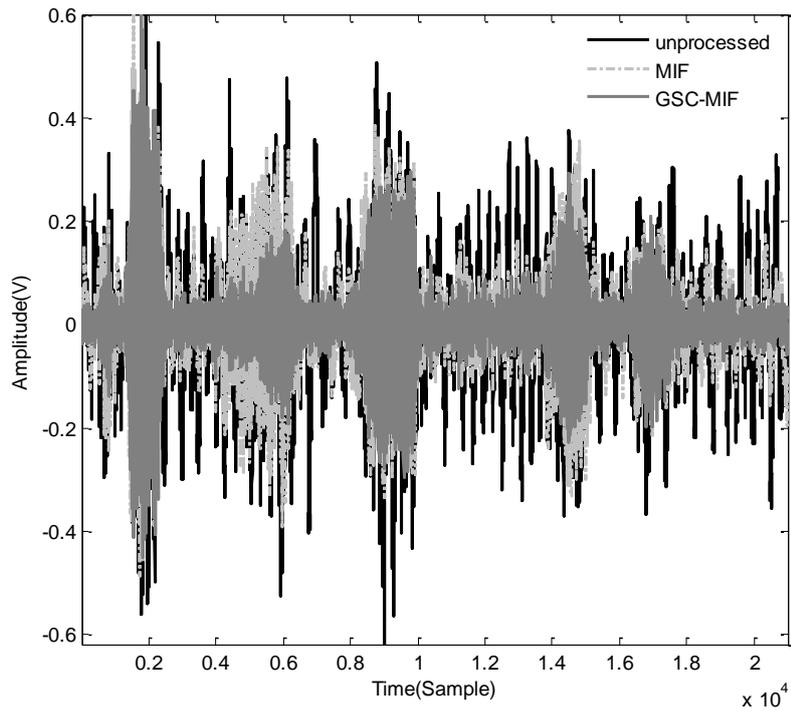
前面章節所提及之 SNR_1 , SNR_A , $SNRG$ 和 SDI 四種客觀量測參數被使用來評估表 I 中的六種被推薦的演算法，實驗結果歸納在表 II 中：

SIMO-ESIF	PIF		MIF		MTR	
	without	with	without	with	without	with
SNR_1 (dB)	3.79	3.79	3.79	3.79	3.79	3.79
SNR_A (dB)	12.96	15.28	15.56	19.19	13.58	13.66
$SNRG$ (dB)	9.16	11.49	11.77	15.41	9.78	9.87
SDI (dB)	2.87	2.60	1.72	1.59	0.86	1.56

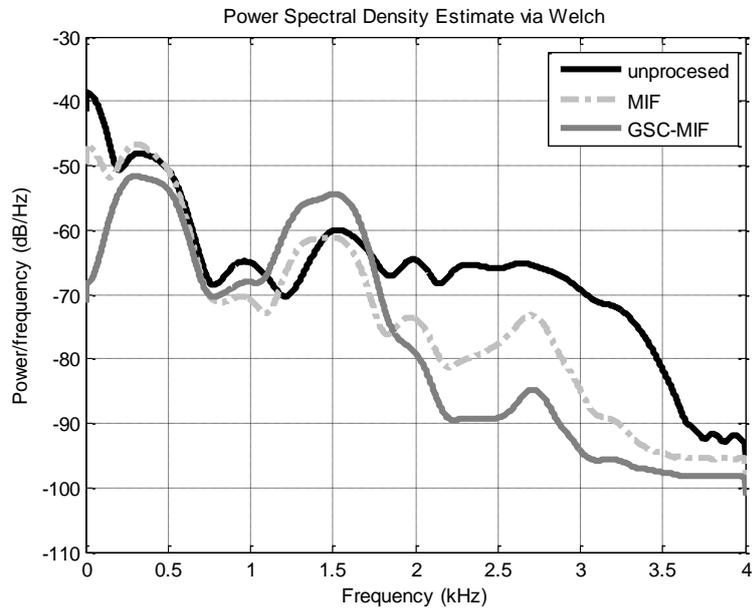
表 II、各種 GSC 演算法之客觀參數比較

結果顯示三種不同的濾波器設計法都能有效的提升輸出的 SNR，這代表著此兩種演算法對於消噪及抑止殘響都有相當大的成效，而再進一步比較兩種演算法，可以看出 SIMO-ESIF 結合 GSC 演算法在 SNR gain 的部分又高於 SIMO-ESIF 演算法，如此表示此種演算法能更進一步的提升消噪及抑止殘響的效果。時間翻轉濾波器在一般情形下表現均略差於反濾波器演算法，反濾波器演算法搭配量測到的系統響應之效果比點聲源模式擁有更好的效果，如表中所示：GSC-MIF 的 $SNRG = 15.41$ dB 而 GSC-PIF 的 $SNRG = 11.49$ dB。此結果也顯示出反濾波器演算法搭配量測的系統響應擁有最佳的抑制殘響之效果。最後來看 SDI 的部分，其值大概落在 1.7 左右，這表示語音經過處理過後有一定程度的失真，但此失真度並不太大，仍然在我們可接受的範圍之內。

圖七比較在時域中的 SIMO-ESIF 演算法有加上 GSC 以及沒有加的語音訊號，從圖中可得知本研究所介紹的 GSC 演算法的確能有效的增加在噪音環境中的陣列表現。



圖七(a)、比較 MIF 和 GSC-MIF 演算法時域訊號



圖七(b)、比較 MIF 和 GSC-MIF 演算法的 power spectral density functions

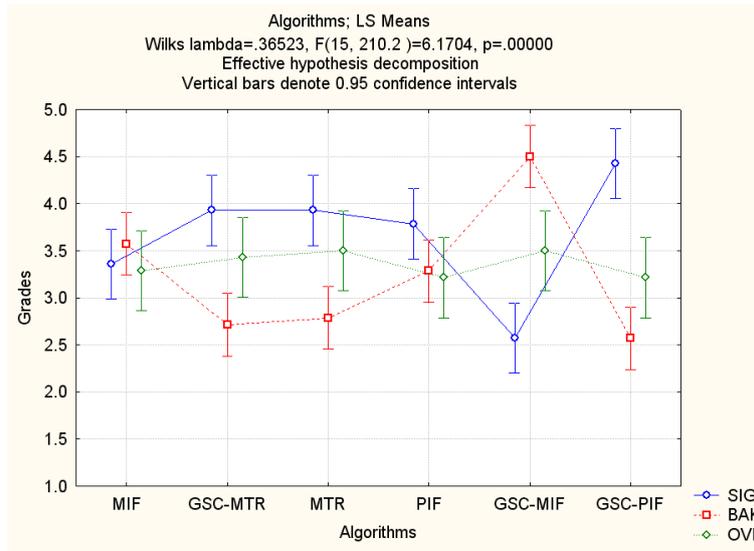
表 III 比較本研究所推薦的適應性 GSC 演算法以及其他兩種常見的演算法—GJBF 和 CCAF。GJBF 演算法利用將相鄰麥克風訊號相減來作為其 BM，而 CCAF 演算法是使用有限制的適應性濾波器來將目標訊號抑制。此兩種演算法均是利用適應性濾波器來作為 MC 部份。其結果可看出 SIMO-ESIF 結合 GSC 的演算法效果比單用 SIMO-ESIF 演算法來的好，在所有 GSC 演算法中，本研究所推薦的 GSC 擁有最高的訊噪比增益值(SNRG)，也就是對於噪音的抑制擁有最佳的效果。

Objective index	MIF			
	ESIF	ESIF-GSC	GJBF	CCAF
SNR_1 (dB)	-1.04	-1.04	-1.04	-1.04
SNR_A (dB)	6.20	12.72	10.27	9.92
$SNRG$ (dB)	7.24	13.76	11.31	10.96
SDI (dB)	1.86	1.42	2.49	1.90

表 III、比較 ESIF、ESIF-GSC、GJBF 以及 CCAF 之客觀參數實驗結果

2. 主觀評估

和客觀參數非常不同的是使用根據 ITU-R BS1116 之聆聽測試實驗來評估演算法效果，聆聽測試能同時評估減噪效果以及語音訊號是否失真。本實驗使用三種評估參數：訊號失真(signal distortion, *SIG*)、背景干擾(background intrusiveness, *BAK*)以及總品質(overall quality, *OVL*)；共有十四個受過聆聽測試訓練之測試者參與，主觀評估分數範圍為 1 至 5 分。測試者將在聆聽之後必須回答問題，總共有六種被推薦的演算法使用於聆聽實驗中，測試訊號以及情況和客觀參數評估相同，參考訊號(reference signal)以及 anchor signal 是參造 ITU-R BS1116。在本實驗中，reference signal 是使用第一個麥克風收到的未經過處理的訊號，而 anchor signal 是使用過低通濾波器的訊號。圖八為聆聽測試分數分佈圖：



圖八、六種 SIMO-ESIF 演算法的聆聽測試結果，使用三種評估參數：*SIG*、*BAK* 以及 *OVL*

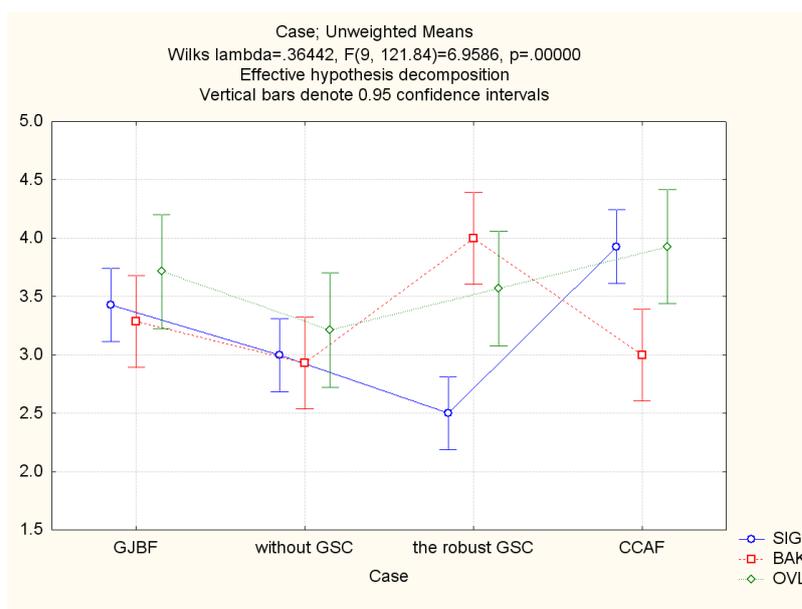
表 IV 為將聆聽實驗的結果使用統計的方法來觀察結果，本實驗使用 MANOVA 法，若 significance levels 在 0.05 以下則表示各種方法的區別是非常顯著的，在表中明顯的發現，*SIG* 和 *BAK* 有非常顯著的差異，使用回歸分析來找尋 *OVL* 和 *SIG* 以及 *BAK* 的關係，則此三種參數可寫成 $OVL = 1.71 + 0.2 * SIG + 0.28 * BAK$ ，由此式可得知 *SIG* 對於 *OVL* 的影響略高於 *BAK*，而 *SIG*、*BAK* 這兩種參數是一種 trade-off 的關係，因此從 *OVL* 來看各種演算法是沒有顯著差異的。

Noise type	Significance value <i>p</i>		
	SIG	BAK	OVL
White noise	0.000	0.000	0.847

表 IV、六種演算法的聆聽實驗結果，significance value *p* 若小於 0.05 表示演算法之間有顯著差異

在使用 MANOVA 之後，亦使用 post-hoc Fisher's LSD 測試來觀察比較各種方法，在圖八中，和客觀測試結果不同的是 GSC-MIF 演算法在 *SIG* 方面不如預期的好，這是由於此演算法對於噪音消除有顯著效果，因此語音訊號有些許失真，使得測試者認為其效果沒有這麼好。從 *SIG* 來看，post hoc 測試的結果顯示

GSC-PIF 較其他的方法來的佳，也就是反濾波器演算法在消除噪音及殘響方面非常成功。雖然在 *SIG* 表現良好，但是這種 PIF 方法在 *BAK* 分數上卻略低，此結果亦和客觀測試結果吻合。另一方面，GSC-PIF 演算法在 *SIG* 分數上略高於 PIF 演算法，意思是 GSC 演算法的確會提昇 SIMO-ESIF 表現。但是，在 MTR 和 GSC-MTR 此兩種演算法中，*SIG* 和 *BAK* 這兩種參數卻有極大不同。比較 *BAK* 分數後發現，本研究所推薦的方法均表現的比參考訊號來的佳。



圖九、比較不同 GSC 演算法的聆聽測試結果

圖九比較推薦的 GSC 演算法以及 GJBF、CCAF 三種演算法，從圖中可看出，本研究所推薦的演算法其 *BAK* 分數最高，而 *SIG* 分數則略低於其他演算法。顯然地，此方法能在略微損失一點語音訊號情況下有最佳的減噪效果。

参考文献

- ¹ J. Benesty, J. Chen and Y. Huang, *Microphone arrays signal processing* (Springer, 2008).
- ² J. Bitzer, K. U. Simmer and K. D. Kammeyer, “Multi-microphone noise reduction techniques for hands-free speech recognition –a comparative study-,” in *Robust Methods for Speech Recognition in Adverse Conditions (ROBUST99)*, 171–174, Tampere, Finland, May 1999.
- ³ J. Bitzer, K. D. Kammeyer and K. U. Simmer, “An alternative implementation of the superdirective beamformer,” Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 1999.
- ⁴ H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
- ⁵ H. Cox, R. M. Zeskind and M. M. Owen, “Robust adaptive beamforming,” IEEE Trans on acoustics., **ASSP-35**(10) (1987).
- ⁶ N. L. Owsley, “Source location with an adaptive antenna array,” Technical Report, Naval Underwater Systems Center, National Technical Information Service, Springfield, Virginia, January 1971.
- ⁷ L. J. Griffiths and C. W. Jim, “An alternative approach to linear constrained adaptive beamforming,” IEEE Trans. Antennas Propagat., **AP-30**, 27-34 (1982).
- ⁸ J. Bitzer, K. U. Simmer and K. D. Kammeyer, “Multichannel noise reduction –algorithms and theoretical limits-,” in Proc. EURASIP European Signal Proc. Conference (EUSIPCO), **1**, 105-108, Rhodes, Greece, Sep. 1998.
- ⁹ M. R. Bai, and J. H. Lin, “Source identification system based on the time-domain nearfield equivalence source imaging: fundamental theory and implementation,” J. Sound Vib. **307**, 202–225 (2007).
- ¹⁰ M. Brandstein and D. Ward, *Microphone arrays* (Springer, New York, 2001).

- ¹¹ O. Hoshuyama, A. Sugiyama and A. Hirano “A robust adaptive beamformer for microphone array with a blocking matrix using constrained adaptive filters,” *IEEE Trans Signal Processing*, **47**(10) (1999).
- ¹² Y. Grenier, “A microphone array for car environment,” *Speech Communication*, **12**(1), 25-39 (1993).
- ¹³ O. L. Frost , III, “An algorithm for linearly-constrained adaptive array processing,” *Proc. IEEE*, **60**(8), 926-935 (1972).
- ¹⁴ G. F. Franklin, M. L. Workman and D. Powell, *Feedback Control of Dynamic Systems, 2nd edition* (Addison-Wesley, Boston, 1993).
- ¹⁵ I. Claesson and S. Nordholm, “A spatial filtering approach to robust adaptive beamforming, ” *IEEE Trans. Antennas Propagat.*, 1093-1096 (1992).
- ¹⁶ O. Hoshuyama and A. Sugiyama, “A robust generalized sidelobe canceller with a blocking matrix using leaky adaptive filters,” *Electron Communicat. Japan*, **80**(8), 56-65 (1997).
- ¹⁷ S. Sharma, *Applied multivariate techniques* (John Wiley, New York, 1996).
- ¹⁸ National Instruments, <http://sine.ni.com/nips/cds/view/p/lang/zht/nid/202630> (date last viewed 7/17/09).
- ¹⁹ ITU-R Rec. BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,” (International Telecommunications Union, Geneva, Switzerland, 1994-1997).

計畫成果自評

本計畫提出一種能夠使用於吵雜的汽車環境的 SIMO-ESIF 麥克風陣列技術，語音通訊品質也由於減噪和消除殘響而提昇，GSC 使用經過特別設計的 BM 之後，SIMO-ESIF 擁有更佳的效果。經過一連串的主客觀測試，證明本研究所推薦的演算法是能夠使用於真實環境的，結果也顯示使用 SIMO-ESIF 技術能消除殘響以及雜訊。此方法也展示了對於消除噪音表現和語音失真品質不同的角度，MIF 和 GSC-MIF 這兩種演算法似乎對於上述兩種表現達到了最佳的妥協。以上所有結果都能看出 SIMO-ESIF-GSC-MIF 演算法能在不失去語音品質的情況下有效的減少噪音和干擾。