

# 發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：

96146698

※申請日期：

96.12.7

※IPC 分類：G10L 2/02 (2006.01)

一、發明名稱：(中文/英文)

智慧型語音純化系統及其方法

二、申請人：(共1人)

姓名或名稱：(中文/英文)

國立交通大學

代表人：(中文/英文) 吳重雨

住居所或營業所地址：(中文/英文)

新竹市大學路 1001 號

國籍：(中文/英文) 中華民國 TW

三、發明人：(共2人)

姓名：(中文/英文)

1. 胡竹生

2. 楊佳興

國籍：(中文/英文)

中華民國 TW (皆同)

#### 四、聲明事項：

主張專利法第二十二條第二項第一款或第二款規定之事實，其事實發生日期為： 年 月 日。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

## 五、中文發明摘要：

本發明係提供一種智慧型語音純化系統及其方法，其係利用智慧型語音純化系統純化受干擾之語音訊號與消除回音的干擾。智慧型語音純化包括以 USB 為傳輸介面的數位麥克風陣列語音擷取系統與語音純化技術。相較於先前技術，本發明能增加整體電路的抗雜訊能力並縮小電路體積。另外，本發明採用常見的 USB 傳輸介面，藉此提升語音擷取系統的方便性與擴充性。在語音純化系統方面，本發明能適用於各種不同干擾源的吵雜環境並且不需考慮麥克風匹配問題以降低製作成本。

## 六、英文發明摘要：

**七、指定代表圖：**

(一)本案指定代表圖為：第(一)圖。

(二)本代表圖之元件符號簡單說明：

12 智慧型語音純化系統

14 數位麥克風陣列語音訊號擷取裝置

16 語音純化裝置

**八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：**

## 九、發明說明：

### 【發明所屬之技術領域】

本發明係一種智慧型語音純化系統及其方法，其係特別關於一種麥克風陣列訊號擷取與語音強化技術的智慧型語音純化系統及其方法。

### 【先前技術】

隨著視訊會議系統、網路電話與語音命令操作技術的蓬勃發展，像是IBM的ViaVoice與Nuance的Naturally Speaking都是語音命令操作的發展廠商，而微軟更在新一代作業系統Windows Vista中支援麥克風陣列語音純化系統。因此，具有高穩定度與高擴充性的語音純化系統將愈來愈重要與普遍。

於現實環境中，語音辨識器易受到環境噪音、回音反射繞射等而影響其辨識效果。在現今的技術中，已有各種利用麥克風陣列的方法來達到抑制噪音源、增強語音品質並改善語音辨識率的方法。雖然這些方法在某些條件下能達到一定的效果，然而，這些方法卻較少同時考慮到製作成本、現實環境噪音種類與空間濾波器參數自動調整正確時機等問題。舉例來說，許多演算法須假設每顆麥克風的頻率響應為完全匹配，但此假設對製作成本是一極大的負擔。以參考訊號為基礎的技術已提出並解決麥克風之間校正問題，但此技術對自動適應性環境時機較不穩健。另外在硬體方面，傳統麥克風陣列的製作是利用類比麥克風，而類比麥克風的抗雜訊能力較差，易受射頻和電磁干擾，並且會有放大電路體積龐大與串音干擾等缺點，對演算法的效果也會有相對的影響。在已知的技術中，並未對上述的問題同時加以考慮，但這些問題會使目前已知技術效能降低，導致應用層面將

受到侷限。

由 Malvar 等人於 2006 年發表的美國專利 No. 321954 [1-1]，提出用一記憶體儲存雜訊統計特性與另一記憶體儲存訊號+雜訊統計特性，並利用適應性演算法與此兩記憶體所存特性來設計以迴授為調整機制的空間濾波器。但此方法無法自動適應性環境訊號，只能消除雜訊無法消除回音，且並無語者方位的資訊，使用者移動後，則需人為調整記憶體儲存資訊。

由 Lambert 等人於 2005 年發表的美國專利 No. 197817 [1-2]，提出用麥克風陣列中某一顆麥克風當作參考麥克風，並利用適應性有限長度脈衝響應濾波器的設計，使其它麥克風線性輸出總和能趨近於預期信號，而適應性演算法調整時間是由語音活動偵測所決定，若語音活動偵測判別錯誤，則有限長度脈衝響應濾波器參數會不適於現階段環境，另外，此系統亦無聲源偵測判別，使用者位置將受到限制。

由 Chol 等人於 2004 年發表的美國專利 No. 757994 [1-3]，提出用固定空間濾波器與純雜訊擷取器，搭配適應性迴授演算法，達到消去固定空間濾波器輸出雜訊的部分。此法即是一 Generalized Sidelobe Canceller 架構，若純雜訊擷取器輸出並非純雜訊，即含有預期訊號，則會發生預期訊號相銷的情況並降低 SNR。

由 Attias 等人於 2004 年發表的美國專利 No. 183267 [1-4]，提出利用模型方式來達到語音純化的效果，此方法會建立純語音模型與雜訊模型，並利用 Expectation Maximization (EM) 演算法離線訓練純語音模型，而雜訊模型則為離線或線上訓練。此方法過程較為複雜，並沒有語者方位

的資訊，則純語音模型則無法自行適應調整。

由 Gannot 等人於 2004 年所發表的文獻中[2-1]，提出一用於 GSC(Generalized Sidelobe Canceller)中設計封鎖矩陣(Blocking Matrix)的新架構，此文獻利用最小平方法求出麥克風之間的傳遞含數比(Transfer Function Ratio)，並再搭配多種後濾波器(Postfilter)與適應性演算法進而求出語音純化的結果。然而，此方法理論架構完整，但整體架構過於複雜，用於實際環境，若傳遞含數比若稍有誤差則效果會大打則扣。

由 Nordholm 等人於 1999 年所發表的文獻中[2-2]，提出一以參考訊號為基礎的語音純化系統架構，此方法須先於無雜訊環境下錄製固定聲源與噪音源，並再搭配適應性演算法求出空間濾波器參數，達到語音純化的效果。次方法於純化系統架構上極有貢獻度，但缺少自動適應性環境能力，需人為設定空間濾波器參數調整時間，若使用者改變方位或預設噪音源位置改變則語音純化能力會大為降低。

由 Hoshuyama 等人於 1999 年所發表的文獻中[2-3]，提出一用於 GSC 中有適應性能力的封鎖矩陣，此方法利用固定空間濾波器(Fixed Beamformer)的輸出作為封鎖矩陣每個通道訊號輸入的參考訊號，並配合適應性演算法求出兩組適應性濾波參數，達到與固定空間濾波器輸出中雜訊相銷的效果。此方法適用於實際環境中，但在封鎖矩陣中適應性演算法是以單一通道為架構，則封鎖訊號源效果有限，若固定空間濾波器輸出包含太多雜訊，則此方法效能勢必下降。為此，本發明提出一種智慧型語音純化系統及其方法，以改善上述缺失。

**【發明內容】**

本發明之主要目的在提供一種智慧型語音純化系統及其方法，其係利用數位麥克風陣列擷取語音訊號裝置擷取訊號，再利用語音純化裝置純化吵雜環境中的語音訊號。

本發明之另一目的在提供一種智慧型語音純化系統及其方法，其係可對複雜環境的各類噪音來做出適應性調整。

本發明之又一目的在提供一種智慧型語音純化系統及其方法，其係具有低成本、抗雜訊能力佳、體積小、功能彈性與適應環境能力佳等優點。

本發明係提供一種智慧型語音純化系統及其方法，首先錄製一使用者的語音參考訊號(無噪音理想訊號)、判斷聲源方向，同時儲存語音參考訊號；擷取具有環境噪音的一聲音訊號，並儲存該聲音訊號，將此聲音訊號通過一空間濾波器後，再通過一語音活動偵測裝置，如語音活動偵測判定為有語音，則將語音輸出，如語音活動偵測判定為無語音，則進入一更新參數機制更新適應性空間濾波器參數。進入該更新參數機制後，將語音參考訊號與聲音訊號組合，進入適應性空間濾波器，搭配適應性演算法得到新的空間濾波器參數，得到後再將新的空間濾波器參數傳遞給空間濾波器；若使用者位置沒有變動或沒有更換使用者則反覆執行系統反覆執行更新參數機制。若使用者位置變動或更換使用者則重新實施本方法。

整體而言，本發明改善傳統類比麥克風陣列亦受雜訊干擾與體積龐大的問題，並整合了語音純化技術，使系統能依據不同環境穩健地自行調整空間濾波器參數，達到一高穩定與高效能的語音純化系統。

底下藉由具體實施例配合所附的圖式詳加說明，當更容易瞭解本發明



之目的、技術內容、特點及其所達成之功效。

### 【實施方式】

本發明利用數位麥克風陣列體積小抗射頻和電磁干擾能力佳的優點來設計一全數位式電路解決傳統麥克風陣列類比電路易受雜訊與串音干擾，陣列放大器電路體積龐大等問題。除此之外，本發明提出一更新參數機制(更新空間濾波器參數)與利用參考語音訊號為基礎的架構解決語音強化技術中常見的問題，包括期望信號消除(desired signal cancellation)、通道效應(channel effect)、空間濾波器參數自動更新時機與理想環境假設等。

為能詳細說明本發明，請參照第一圖，本發明之智慧型語音純化系統係由 USB 為傳輸介面的數位麥克風陣列語音訊號擷取裝置及語音純化裝置組成。本發明能應用之產業包括 IC 設計(IC Design)，個人電腦(PC)，視訊會議(Video Conference)，家庭自動化(Home Automation, Robotics)，車內環境(Vehicular filed)。本發明能應用之產品包括車內環境(Car environment)，桌上型或膝上型電腦(Desktop & Laptop computers)，筆記型電腦(Notebook)，行動電話(Cell phones)，個人數位助理(PDA)，家庭環境(Home environment)，視訊會議(Video conference product)，機器人(Robot)。

在此先介紹智慧型語音純化系統 12 的結構，再說明智慧型語音純化系統 12 的純化語音的方式。

首先介紹數位麥克風陣列語音訊號擷取裝置 14 的組成元件，請參照第二圖，其包括：

1. 作為感應器的數位麥克風陣列 18，其中數位麥克風陣列 18 可以擷取聲音訊號。
2. 與數位麥克風陣列 18 連接之抽樣濾波器 20(Decimation Filter)，抽樣濾波器 20 接收來自數位麥克風陣列 18 傳遞之序列資料，並將序列資料轉為並行資料，其中抽樣濾波器 20 亦包含通代濾波器(Bandpass Filter)的效果。
3. 與抽樣濾波器 20 連接之多工器 22，可將上述並行資料逐一傳輸給後端的 USB 控制器 24。
4. USB 控制器 24 利用 USB 傳輸介面，將資料傳輸給語音純化裝置 16。  
再來說明語音純化裝置 16 的內部組成元件，請參照第三圖：

1. 方位判別器 26，與上述 USB 傳輸介面連接，方位判別器 26 可以接收來自的 USB 傳輸介面的資料後，判別使用者所在位置；
2. 第一記憶體 28，與方位判別器 26 連接，第一記憶體 28 係用來儲存語音參考訊號；
3. 第二記憶體 30，與該方位判別器 26 連接，用來儲存環境聲音；
4. 空間濾波器 32，與該方位判別器 26 連接，用以純化受干擾之語音訊號；
5. 語音活動偵測裝置 34，與空間濾波器 32 連接，語音活動偵測裝置 34 接收之空間濾波器 32 輸出之語音訊號，以提升語音活動偵測判別準確性與改善空間濾波器參數自動更新時機。
6. 適應性空間濾波器 36 的輸入端與第一記憶體 28 及第二記憶體 30 連接，適應性空間濾波器 36 能接收該語音參考訊號與該環境噪音，並更新空間

濾波器參數。

依據上述介紹之智慧型語音純化系統 12，繼續說明智慧型語音純化系統 12 的純化語音的方式，請參照第四圖：首先進行步驟 S12，第一次使用智慧型語音純化系統 12 時，智慧型語音純化系統 12 會先錄製使用者的一語音參考訊號，並判斷聲源方向，同時儲存語音參考訊號；其中，錄製使用者的一語音參考訊號係利用作為感應器的數位麥克風陣列 18，數位麥克風陣列 18 可以擷取聲音訊號。擷取聲音訊號後，數位麥克風陣列 18 連接之抽樣濾波器 20 能將取樣的聲音訊號解碼與濾波，並將單位元(如 16 位元)數據串流轉換為並行資料，再將並行資料傳輸到多工器 22。USB 控制器 24 係與連接多工器 22，USB 控制器 24 可透過多工器 22 將解碼後的資料逐一讀取，再利用 USB 傳輸介面作為輸出介面輸出到方位判別器 26；錄製後的語音參考訊號由方位判別器 26 判斷聲源方向後，會儲存到語音純化裝置 16 內的第一記憶體 28 中。此步驟中儲存於第一記憶體 28 中的語音參考訊號屬於理想(無噪音)的聲音訊號。由於數位麥克風陣列語音訊號擷取裝置 14 整體皆為數位化，相較於傳統的類比放大電路，本發明之數位麥克風陣列語音訊號擷取裝置 14 有體積小、抗雜訊能力佳等優點。

再來進行步驟 S14，再次擷取一聲音訊號，並儲存聲音訊號，並將此聲音訊號通過一空間濾波器 32 後，再通過語音活動偵測裝置 34，如語音活動偵測判定為有語音，則將語音輸出，如語音活動偵測判定為無語音，則進行步驟 S16 進入一更新參數機制以適應性空間濾波器 36 參數。

在本步驟中，擷取之聲音訊號同樣是由作為感應器的數位麥克風陣列

18 擷取聲音訊號。擷取聲音訊號後，數位麥克風陣列 18 連接之抽樣濾波器 20 能將取樣的聲音訊號解碼與濾波，將單位元數據串流轉換為並行資料，再將並行資料傳輸到多工器 22。USB 控制器 24 透過多工器 22 將解碼後的資料逐一讀取，再利用 USB 傳輸介面作為輸出介面輸出到方位判別器 26；錄製後的語音參考訊號不須由方位判別器 26 判斷聲源方向，會直接儲存到語音純化裝置 16 內的第二記憶體 30 中。此步驟中的聲音訊號屬於具有環境噪音的聲音訊號，與儲存在第一記憶體 28 中的理想(無噪音)的聲音訊號不同。

接著進行步驟 S16，進入一個更新參數機制，更新參數機制開始後，語音純化裝置 16 會將語音參考訊號  $S[n]$  與具有環境噪音的聲音訊號  $N[n]$  組合出  $X[n]$ ，並進入適應性空間濾波器 36，搭配適應性演算法得到一較適合現在環境的空間濾波器參數，再將此參數傳遞給空間濾波器 32。其中，適應性空間濾波器 36 會將其輸出訊號與語音參考訊號於  $e[n]$  比對，使輸出訊號趨近未受干擾的訊號參數後，再將空間濾波器參數傳遞給空間濾波器 32。

在啟動更新參數機制，變更適應性空間濾波器 36 參數的時機方面，傳統方式大多利用將環境聲音直接通過語音活動偵測的結果來決定空間濾波器 32 是否進入適應性調整，若有語音，則使語音通過空間濾波器 32 達到純化的效果，反之，則進入空間濾波器 32 適應性參數調整狀態。但完美的語音活動偵測是不存在的，因此，本發明提出一適應性空間濾波器 36 的更新參數調整機制，即本系統將語音活動偵測設置於空間濾波器 32 之後，則語音活動偵測的輸入為一經過語音純化的訊號，此設計除了可增加語音活

動偵測的準確性外，亦可提升更新參數的適當時機，使更新空間濾波器參數時機較穩健於不同聲場環境。

接著進行步驟 S18，若使用者位置變動或更換使用者則重新進入步驟 S12，若使用者位置沒有變動或沒有更換使用者則反覆執行步驟 S14 及步驟 S16。

承上所述，本發明應用了數位麥克風陣列 18 體積小抗雜訊的優點；採用的介面為 USB 傳輸介面，具有高度的擴充性與便利性，另外，語音純化上不需麥克風之間互相匹配，能用於吵雜環境，有效降低製作成本。

再者，本發明針對現實環境所出現的問題皆納入考量，能便利地使用於多種環境，不易受噪音源種類與環境複雜度等影響，適用於桌上型或筆記型電腦。

為能更清楚比較出本發明與先前技術之不同，特提供以下表單說明本發明與上揭先前技術在技術層面與運用層面的差異：

技術	回音消除	演算法複雜度	麥克風匹配性需求	自行適應性環境能力	雜訊環境之穩健性
[1-1]	否	簡單	高	低	低
[1-2]	可	適中	低	低	低
[1-3]	否	適中	高	低	低
[1-4]	可	複雜	高	中	低
[2-1]	可	複雜	高	高	中

[2-2]	可	簡單	低	中	低
[2-3]	否	適中	高	低	低
本發明	可	適中	低	高	高

整體而言，本發明改善傳統類比麥克風陣列亦受雜訊干擾與體積龐大的問題，並整合了語音純化技術，使系統能依據不同環境穩健地自行調整空間濾波器參數，達到一高穩定與高效能的語音純化系統。

唯以上所述者，僅為本發明之較佳實施例而已，並非用來限定本發明實施之範圍。故即凡依本發明申請範圍所述之形狀、構造、特徵及精神所為之均等變化或修飾，均應包括於本發明之申請專利範圍內。

### 【圖式簡單說明】

第一圖係本發明之智慧型語音純化系統之架構圖。

第二圖係本發明之數位麥克風陣列語音訊號擷取裝置之架構圖。

第三圖係本發明之語音純化裝置之架構圖。

第四圖係本發明之實施流程圖。

### 【主要元件符號說明】

12 智慧型語音純化系統

14 數位麥克風陣列語音訊號擷取裝置

16 語音純化裝置

18 數位麥克風陣列

20 抽樣濾波器

22 多工器

24 USB 控制器

26 方位判別器

200926150

28 第一記憶體

30 第二記憶體

32 空間濾波器

34 語音活動偵測裝置

36 適應性空間濾波器

## 十、申請專利範圍：

1. 一種智慧型語音純化的系統，包括下列步驟：

一數位麥克風陣列語音訊號擷取系統，擷取一聲音訊號；

一方位判別器，與該數位麥克風陣列語音訊號擷取系統連接，可判斷該聲音訊號之聲源方向；

一第一記憶體，與該方位判別器連接，儲存一語音參考訊號，該語音參考訊號係該聲音訊號無噪音時的訊號；

一第二記憶體，與該方位判別器連接，儲存一聲音訊號，該聲音訊號係有噪音時的訊號；

一適應性空間濾波器，其輸入端與該第一記憶體及該第二記憶體連接；

一空間濾波器，與該方位判別器連接，接收來自該方位判別器的聲音訊號；

一語音活動偵測系統，與該空間濾波器連接，接收來自該空間濾波器的聲音訊號，且該語音活動偵測系統會判定是否有語音，如是，則將語音輸出，如否則進入一更新參數機制更新適應性空間濾波器的參數，適應性空間濾波器的參數會再回傳給該空間濾波器後，該空間濾波器再次將該聲音訊號處理後經由該語音活動偵測系統輸出。

2. 如申請專利範圍第1項所述之智慧型語音純化的系統，其中該數位麥克風陣列語音訊號擷取系統包括：

一數位麥克風陣列，能取樣一聲音訊號，

一抽樣濾波器，與該數位麥克風陣列連接，該抽樣濾波器將取樣後的該聲



音訊號解碼與濾波，將單位元數據串流轉換為並行資料；

一多工器，與該抽樣濾波器連接，可接收並行資料；

一 USB 控制器，與該多工器連接，該 USB 控制器可透過多工器將解碼後的資料逐一讀取，最後透過 USB 傳輸介面將一資料傳輸給一方位判別器。

3、如申請專利範圍第 2 項所述之智慧型語音純化的系統，其中該該單位元係 16 位元。

4、一種智慧型語音純化的方法，包括下列步驟：

(1) 錄製一使用者的一語音參考訊號，並判斷聲源方向，同時儲存該語音參考訊號；

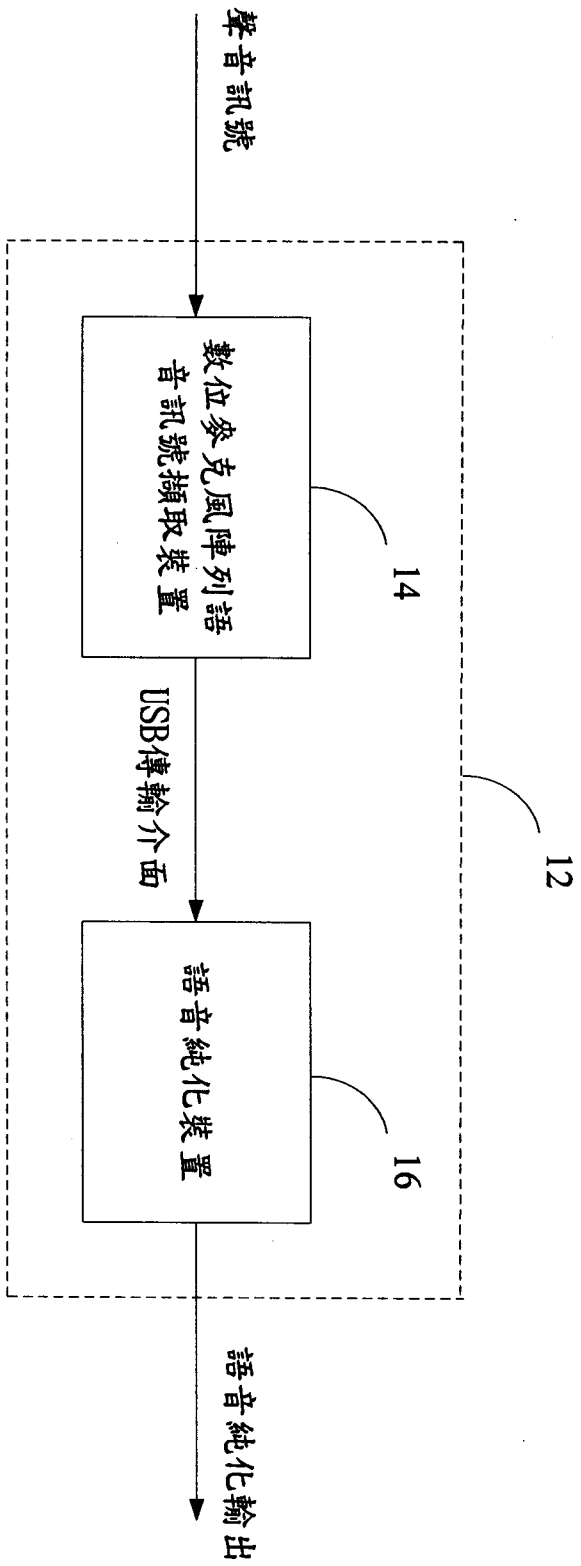
(2) 擷取一聲音訊號，並儲存該聲音訊號，將該聲音訊號通過一空間濾波器後，再通過一語音活動偵測裝置，如語音活動偵測判定為有語音，則將語音輸出，如語音活動偵測判定為無語音，則進入一更新參數機制更新適應性空間濾波器的參數；

(3) 進入該更新參數機制後，將該語音參考訊號與聲音訊號組合，並進入一適應性空間濾波器，搭配一適應性演算法得到新的空間濾波器參數，得到後再將新的空間濾波器參數傳遞給該空間濾波器；以及

(4) 系統反覆執行(2)和(3)。

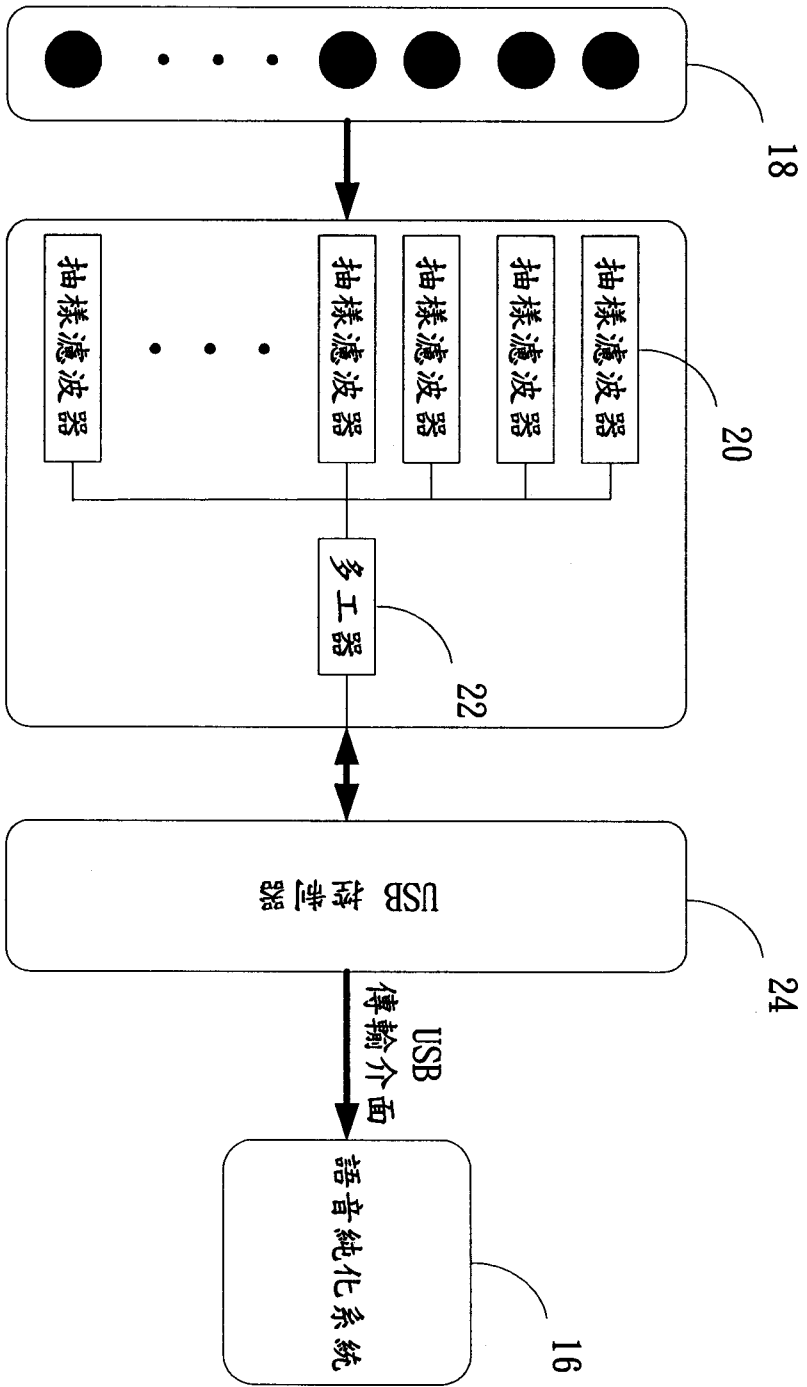
5、如申請專利範圍第 4 項所述之智慧型語音純化的方法，其中若使用者位置變動或更換使用者則系統進入步驟(1) 重新執行。

- 6、如申請專利範圍第 4 項所述之智慧型語音純化的方法，其中判斷聲源方向係利用一方位判別器。
- 7、如申請專利範圍第 4 項所述之智慧型語音純化的方法，其中儲存該語音參考訊號係利用一第一記憶體。
- 8、如申請專利範圍第 4 項所述之智慧型語音純化的方法，其中儲存該聲音訊號係利用一第二記憶體。
- 9、如申請專利範圍第 4 項所述之智慧型語音純化的方法，其中錄製該使用者的語音參考訊號及擷取該聲音訊號係利用一數位麥克風陣列語音訊號擷取裝置。
- 10、如申請專利範圍第 9 項所述之智慧型語音純化的方法，其中該數位麥克風陣列語音訊號擷取裝置包括：
  - 一數位麥克風陣列，能取樣一聲音訊號，
  - 一抽樣濾波器，與該數位麥克風陣列連接，該抽樣濾波器將取樣後的該聲音訊號解碼與濾波，將單位元數據串流轉換為並行資料；
  - 一多工器，與該抽樣濾波器連接，可接收並行資料；
  - 一 USB 控制器，與該多工器連接，該 USB 控制器可透過多工器將解碼後的資料逐一讀取，最後透過 USB 傳輸介面將一資料傳輸給一方位判別器。
- 11、如申請專利範圍第 10 項所述之智慧型語音純化的方法，其中該該單位元係 16 位元。

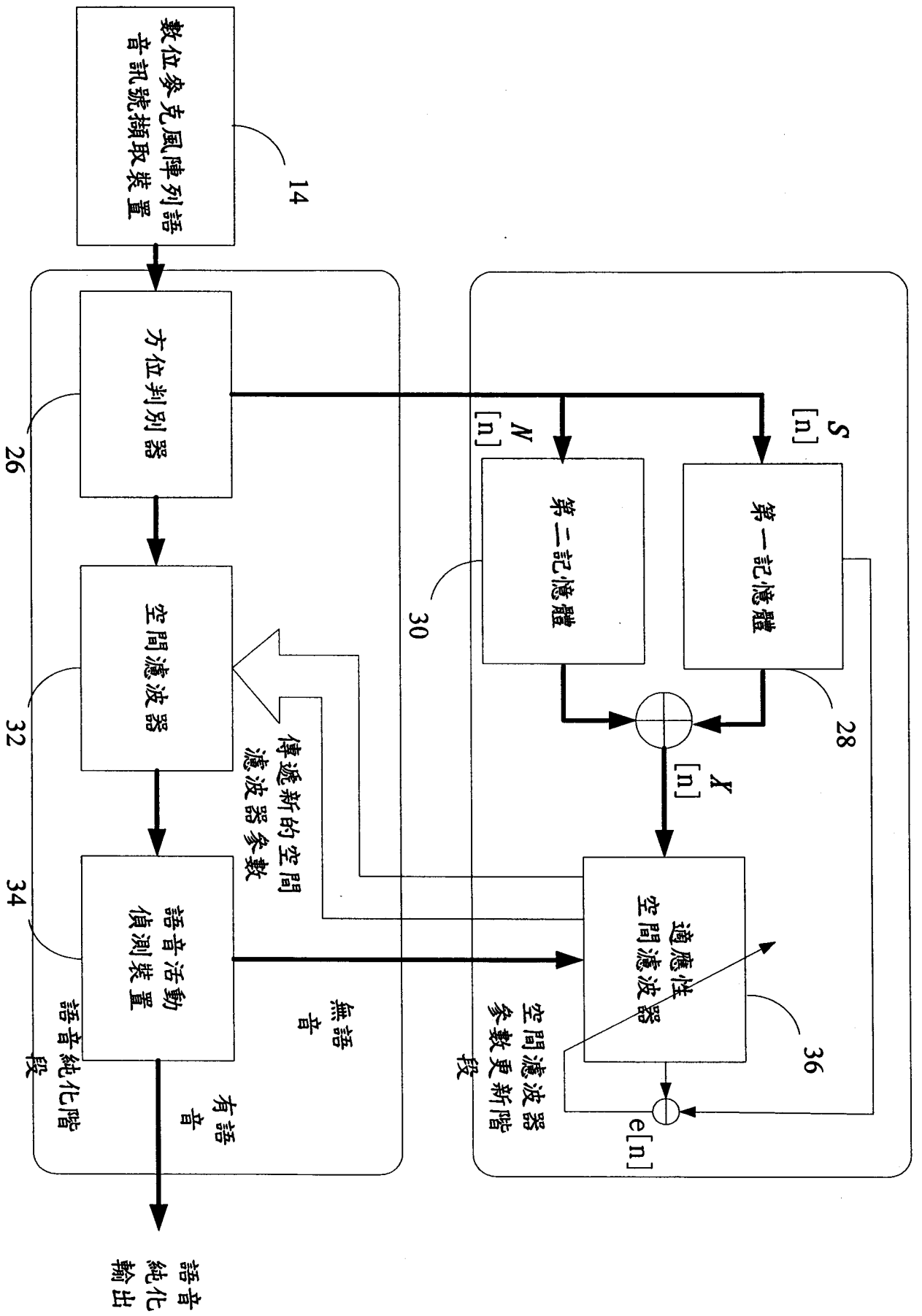


第一圖

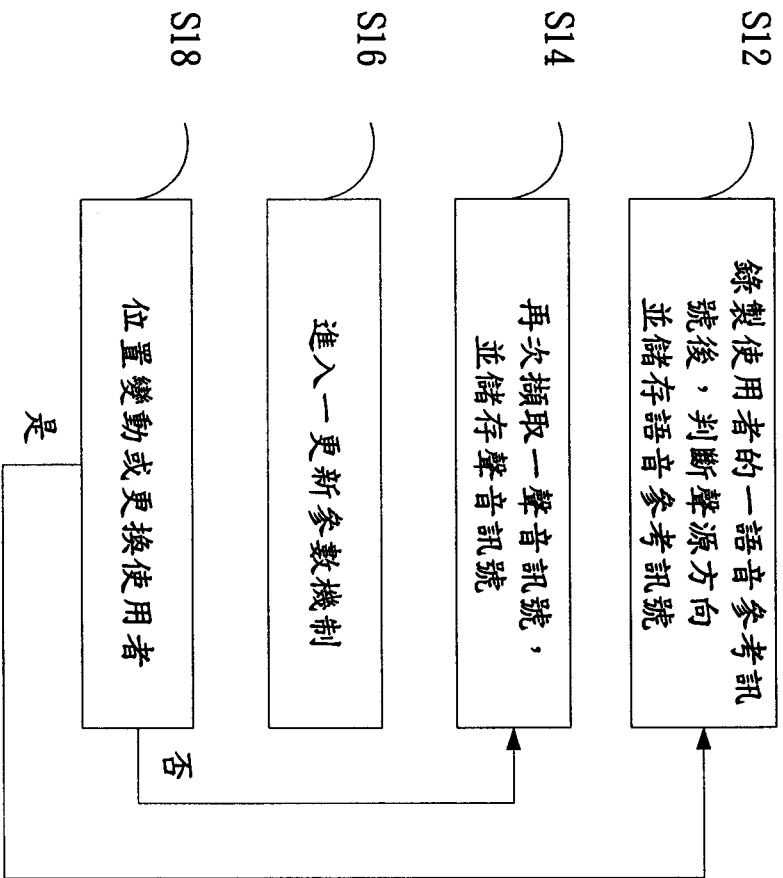
數位麥克風陣列



第二圖



第三圖



第四圖