



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I412023 B

(45) 公告日：中華民國 102 (2013) 年 10 月 11 日

(21) 申請案號：099143712

(22) 申請日：中華民國 99 (2010) 年 12 月 14 日

(51) Int. Cl. : **G10L21/02 (2013.01)**(71) 申請人：國立交通大學 (中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)
新竹市大學路 1001 號

(72) 發明人：白明憲 BAI, MINGSIAN R. (TW) ; 陳俊宏 CHEN, CHUN HUNG (TW)

(74) 代理人：林火泉

(56) 參考文獻：

EP 1590800B1

US 7577262B2

審查人員：涂淑惠

申請專利範圍項數：17 項 圖式數：2 共 0 頁

(54) 名稱

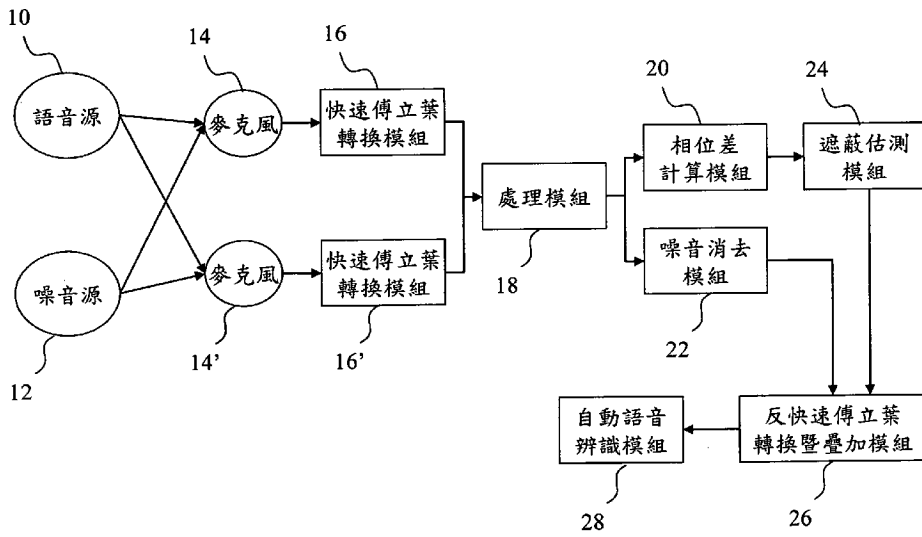
可消除噪音且增進語音品質之麥克風陣列架構及其方法

A MICROPHONE ARRAY STRUCTURE AND METHOD FOR NOISE REDUCTION AND
ENHANCING SPEECH

(57) 摘要

本發明提供一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，利用至少二麥克風接收含有噪音訊號及語音訊號之麥克風訊號，利用快速傅立葉轉換將麥克風訊號轉換至頻率域，接著計算麥克風訊號中噪音訊號及語音訊號之夾角，並依據此夾角選擇使用相位差演算法、噪音消去法或二者合併使用，若使用相位差演算法，則再計算麥克風訊號之間的相位差以獲得時間域及空間域之遮蔽訊號，再將遮蔽訊號乘上麥克風訊號之平均而得到麥克風訊號中之語音訊號，去除噪音部分，增進語音品質。

The present invention provides a microphone array structure and method for noise reduction and enhancing speech, which uses at least two microphones to receive microphone signals comprising noise signal and speech signal. The microphone signals are transformed into frequency domain by fast Fourier transform (FFT). Calculate the angle of noise signal and speech signal, and select phase difference estimation algorithm, noise reduction algorithm or both according to the angle. If phase difference estimation algorithm is selected, the phase difference between microphone signals is calculated to get the time-frequency domain mask signal. The mask signal then multiplied with the average of microphone signals to get the speech signal and remove noise signal, and enhance the speech quality.



第 1 圖

- 10 . . . 語音源
- 12 . . . 噪音源
- 14、14' . . . 麥克風
- 16、16' . . . 快速傅立葉轉換模組
- 18 . . . 處理模組
- 20 . . . 相位差計算模組
- 22 . . . 噪音消除模組
- 24 . . . 遮蔽估測模組
- 26 . . . 反快速傅立葉轉換暨疊加模組
- 28 . . . 自動語音辨識模組

發明專利說明書

(本說明書格式、順序，請勿任意更動，※記號部分請勿填寫)

※申請案號： 99/43712

※申請日： 99.12.14

※IPC 分類： G10L 21/02 (2006.01)

一、發明名稱：(中文/英文)

可消除噪音且增進語音品質之麥克風陣列架構及其方法 / A
microphone array structure and method for noise reduction and enhancing
speech

二、中文發明摘要：

● 本發明提供一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，利用至少二麥克風接收含有噪音訊號及語音訊號之麥克風訊號，利用快速傅立葉轉換將麥克風訊號轉換至頻率域，接著計算麥克風訊號中噪音訊號及語音訊號之夾角，並依據此夾角選擇使用相位差演算法、噪音消去法或二者合併使用，若使用相位差演算法，則再計算麥克風訊號之間的相位差以獲得時間域及空間域之遮蔽訊號，再將遮蔽訊號乘上麥克風訊號之平均而得到麥克風訊號中之語音訊號，去除噪音部分，增進語音品質。

三、英文發明摘要：

● The present invention provides a microphone array structure and method for noise reduction and enhancing speech, which uses at least two microphones to receive microphone signals comprising noise signal and speech signal. The microphone signals are transformed into frequency domain by fast Fourier transform (FFT). Calculate the angle of noise signal and speech signal, and select phase difference estimation algorithm, noise reduction algorithm or both according to the angle. If phase difference estimation algorithm is selected, the phase difference between microphone signals is calculated to get the time-frequency domain mask signal. The mask signal then multiplied with the average of microphone signals to get the speech signal and remove noise signal, and enhance the speech quality.

四、指定代表圖：

(一)本案指定代表圖為：第 (1) 圖。

(二)本代表圖之元件符號簡單說明：

- | | |
|------------------|------------------|
| 10 語音源 | 12 噪音源 |
| 14、14' 麥克風 | 16、16' 快速傅立葉轉換模組 |
| 18 處理模組 | 20 相位差計算模組 |
| 22 噪音消去模組 | 24 遮蔽估測模組 |
| 26 反快速傅立葉轉換暨疊加模組 | |
| 28 自動語音辨識模組 | |

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

六、發明說明：

【發明所屬之技術領域】

本發明係有關一種消除麥克風噪音之技術，特別是指一種可消除噪音且增進語音品質之麥克風陣列架構及其方法。

【先前技術】

按，麥克風接收聲音訊號之方式可分為單通道及雙通道，單通道之消噪方式需要估算消噪比，而雙通道感應多是利用波束形成法 (beam forming) 以陣列方式產生有方向性之麥克風系統，對人聲的敏感度較高而指向人的位置接收聲音訊號，對背景的噪音則較不敏感，但兩個麥克風所形成之波束相當大，指向性不足。

目前用於車內或一般室內之行動電話通訊噪音消除裝置大多使用為數眾多的麥克風、各種濾波器與龐大的矩陣運算，在如此沉重的運算量、巨大的記憶體空間與眾多的麥克風下，對於硬體的成本實為一大負擔。且由於指向性不足，目前無論是市面上的產品或有關麥克風陣列的專利及文獻都無法在存有噪音的環境下有效的消除噪音且不讓語音失真。

因此，本發明即提出一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，將語音訊號分離出提升語音品質，以克服上述該等問題，具體架構及其實施方式將詳述於下。

【發明內容】

本發明之主要目的在提供一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，其係提供相位差演算法及噪音消去法兩種消噪方法，藉由判斷語音及噪音之夾角為零度或不為零度之狀況，選擇使用不同之消噪

方法以得到最佳音質。

本發明之另一目的在提供一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，其係利用黃金比例搜尋法尋找最佳的耳間時間差閾值，使每個角度之語音訊號皆可得到最好的語音品質。

為達上述之目的，本發明提供一種可消除噪音且增進語音品質之麥克風陣列架構，包括至少二麥克風、至少二快速傅立葉轉換模組、一處理模組、一相位差計算模組、一遮蔽估測模組以及一反快速傅立葉轉換暨疊加模組，其中麥克風接收含有噪音訊號及語音訊號之至少二麥克風訊號，快速傅立葉轉換模組將麥克風訊號轉換至頻率域；處理模組計算麥克風訊號中噪音訊號及語音訊號之夾角，並依據此夾角選擇使用相位差演算法配合遮蔽估測、噪音消去法或二者合併使用；相位差計算模組計算麥克風訊號之相位差及耳間時間差，並找出不同之夾角所對應之耳間時間差的最佳閾值；遮蔽估測模組依據此閾值利用一遮蔽法則得到一遮蔽訊號，再將遮蔽訊號乘上麥克風訊號之平均而得到麥克風訊號中之語音訊號；反快速傅立葉轉換暨疊加模組將語音訊號由頻率域轉為時間域。

本發明另提供一種可消除噪音且增進語音品質之麥克風陣列方法，包括下列步驟：接收至少二麥克風訊號，並分別利用一快速傅立葉轉換模組轉至頻率域；計算麥克風訊號中語音訊號及噪音訊號之夾角，並依據此夾角選擇使用相位差演算法配合遮蔽估測、噪音消去法或二者合併使用以將麥克風訊號中之噪音訊號去除；計算麥克風訊號之相位差，以進一步找出一耳間時間差；利用一黃金比例搜尋法找出對應不同夾角時耳間時間差最佳之一閾值；依據一遮蔽法則及閾值得到一遮蔽訊號，將麥克風訊號之平

均與遮蔽訊號相乘得到麥克風訊號中之語音訊號；以及將語音訊號利用一反快速傅立葉轉換暨疊加模組轉至時間域輸出。

底下藉由具體實施例詳加說明，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】

本發明提供一種可消除噪音且增進語音品質之麥克風陣列架構及其方法，利用兩麥克風之間的相位差以獲得麥克風訊號在時間域及頻率域之遮罩，消除噪音，以增進語音品質。

請參考第 1 圖，其為本發明消除噪音且增進語音品質之麥克風陣列架構，包括至少二麥克風 14、14'、至少二快速傅立葉轉換模組 16、16'、一處理模組 18、一相位差計算模組 20、一噪音消去模組 22、一遮蔽估測模組 24、一反快速傅立葉轉換暨疊加模組 26 以及一自動語音辨識模組 28，其中，語音源 10 及噪音源 12 之聲音傳送出去後，麥克風 14、14' 接收同時含有噪音訊號及語音訊號之麥克風訊號，快速傅立葉轉換模組 16、16' 用以將麥克風訊號轉換至頻率域；處理模組 18 用以計算麥克風訊號中噪音訊號及語音訊號之夾角為何，並依據此夾角選擇使用相位差演算法配合遮蔽估測、噪音消去法或二者合併使用；相位差計算模組 20 計算麥克風訊號之相位差及耳間時間差，並找出不同之夾角所對應之耳間時間差的最佳閾值；遮蔽估測模組 24 依據閾值利用一遮蔽法則得到一遮蔽訊號，再將遮蔽訊號乘上麥克風訊號之平均而得到麥克風訊號中之語音訊號；噪音消去模組 22 利用噪音消去法 (noise reduction) 將麥克風訊號中之噪音訊號去除；反快速傅立葉轉換暨疊加模組 26 用以將語音訊號由頻率域轉為時間域；自動語音辨識

模組 28 用以接收反快速傅立葉轉換暨疊加模組 26 所輸出之語音訊號，並進行語音辨識。

本發明所提供可消除噪音且增進語音品質之麥克風陣列方法如第 2 圖之流程圖所示，在步驟 S10 中，噪音訊號及語音訊號經由麥克風接收後，經漢明窗（Hamming window）和快速傅立葉轉換（FFT）轉至頻率域，其二麥克風訊號 $P_1(k,l)$ 及 $P_2(k,l)$ 如下式(1)、(2)所示：

$$P_1(k,l) = X(k,l) + \sum_{i=1}^V N_i(k,l) \quad (1)$$

$$P_2(k,l) = X(k,l) + \sum_{i=0}^V e^{-j\omega_k d_i(k,l)} N_i(k,l) \quad (2)$$

其中 (k,l) 代表第 k 個頻率，第 l 個畫框， X 代表語音訊號， N_i 代表第 i 個噪音源， P_m 是第 m 個麥克風收到之訊號， $\omega_k = 2\pi k/N$ ， $0 \leq k \leq N/2-1$ ， N 是快速傅立葉轉換之長度。

接著在步驟 S12 中，計算此二麥克風訊號 $P_1(k,l)$ 及 $P_2(k,l)$ 中噪音訊號及語音訊號之夾角，亦即語音源及噪音源之間的夾角，以選擇使用相位差演算法配合遮蔽估測或噪音消去法，亦可將二者合併使用。

在步驟 S14 中判斷夾角是否為 0，若否，則步驟 S16 計算噪音訊號及語音訊號之相位差及耳間時間差（interaural time difference, ITD）之閾值。

一般而言，假設語音訊號在麥克風正前方，則其耳間時間差為 0，其他方向來的噪音則用 $d_i(k,l)$ 來表示其耳間時間差，耳間時間差和時間及頻率有關。若有一時-頻域 $\text{bin}(k_j, l_j)$ 是由一最強干擾所支配，則上式(1)、(2)可簡化為下式(3)、(4)：

$$P_i(k_j, l_j) \approx N_n(k_j, l_j) \quad (3)$$

$$P_2(k_j, l_j) \approx e^{-j\omega_{k_j} d_n(k_j, l_j)} N_n(k_j, l_j) \quad (4)$$

此時的耳間時間差可經由計算兩麥克風訊號之間的相位差而得到，如下式

(5)：

$$|d_n(k_j, l_j)| \approx \frac{1}{|\omega_{k_j}|} \min_r |\angle P_1(k_j, l_j) - \angle P_2(k_j, l_j) - 2\pi r| \quad (5)$$

由於接下來在步驟 S18 中會應用到耳間時間差之閾值 (ITD threshold)，因此在本發明步驟 S16 中更提供搜尋最佳閾值之方法，係利用黃金比例搜尋法 (GSS) 來找尋對應各個夾角的最佳閾值 τ 。假設一函數 $f(x)$ 在 $[a, b]$ 內是連續的且只有一最小值，在 $[a, b]$ 內選取兩點 c 和 d ，其關係如下式(9)：

$$\frac{\overline{ca}}{\overline{ba}} = \frac{3 - \sqrt{5}}{2} \quad (9)$$

其中 d 為 c 在 \overline{ab} 線段上的對稱點，比較 $f(c)$ 和 $f(d)$ 的大小，若 $f(c) < f(d)$ ，則新的搜尋點變成 $[a, d]$ ，否則變成 $[c, b]$ ，然後在新的範圍內再取一點，再次比較內部兩點之大小，重複此步驟不斷把範圍縮小，當範圍小到可接受的地步時，就將其當作函數 $f(x)$ 在 $[a, b]$ 區間的最小值，根據泰勒理論，函數 $f(x)$ 靠近 x_m 時，其值近似於：

$$f(x) \approx f(x_m) + \frac{1}{2} f''(x_m) (x - x_m)^2 \quad (10)$$

若 $f(x)$ 夠靠近 $f(x_m)$ ，則後面二次微分項小到可忽略，因此公式(10)可表示為如下式(11)：

$$\frac{1}{2} f''(x_m) (x - x_m)^2 < \varepsilon |f(x_m)| \quad (11)$$

其中 ε 為 10^{-3} 。使用語音失真度，消噪程度與整體語音品質做為黃金比例搜

尋法中函數的參數，可得到夾角對 τ 值的函數如下式(12)：

$$\tau = -0.000056\theta^2 + 0.0108\theta - 0.0575 \quad (12)$$

其中 θ 為語音訊號與噪音訊號之間的夾角，在此 θ 所對應的 τ 可以使經過處理的訊號有最佳的語音品質。

得到最佳之耳間時間差的閾值後，接著在步驟 S18 中依據遮蔽法則 (binary mask principle) 由下式(6)估計出麥克風訊號之遮蔽訊號：

$$B(k_j, l_j) = \begin{cases} 1, & \text{if } |d_n(k_j, l_j)| \leq \tau \\ 0.01, & \text{otherwise} \end{cases} \quad (6)$$

其中，只有耳間時間差比 τ 小的訊號會被認為是目標語音訊號。

最後的語音訊號 $S(k, l)$ 可經由將二麥克風訊號之平均 $\bar{P}(k, l)$ 及遮蔽訊號 $B(k_j, l_j)$ 相乘而得，如下式(7)及下式(8)：

$$\bar{P}(k, l) = \frac{1}{2} \{P_1(k, l) + P_2(k, l)\} \quad (7)$$

$$S(k, l) = B(k, l) \bar{P}(k, l) \quad (8)$$

當步驟 S18 將語音訊號與噪音訊號分離之後，步驟 S22 此頻率域之語音訊號再經過反快速傅立葉轉換 (IFFT) 及重疊相加法 (OLA) 來轉為時域訊號輸出；最後，步驟 S24 自動語音辨識 (Automatic Speech Recognition, ASR) 對輸出之語音訊號進行辨識。

若在步驟 S14 中判斷夾角為 0，則在步驟 S20 中利用噪音消去法 (noise reduction) 去除麥克風訊號中之噪音訊號，保留語音訊號，接著步驟 S22 此頻率域之語音訊號再經過反快速傅立葉轉換及重疊相加法來轉為時域訊號輸出；最後，步驟 S24 自動語音辨識對輸出之語音訊號進行辨識。

綜上所述，本發明提供之可消除噪音且增進語音品質之麥克風陣列架

構及其方法，藉由判斷語音及噪音之夾角是否為零，若為零度選擇噪音消去法，若不為零度則選擇相位差演算法，並在相位差演算法中提供最佳的耳間時間差閾值，以在各個角度皆能達到最佳之消噪效果與整體音質。

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 1 圖為本發明可消除噪音且增進語音品質之麥克風陣列架構之方塊圖。
- 第 2 圖為本發明可消除噪音且增進語音品質之麥克風陣列方法之流程圖。

【主要元件符號說明】

- 10 語音源
- 12 噪音源
- 14、14' 麥克風
- 16、16' 快速傅立葉轉換模組
- 18 處理模組
- 20 相位差計算模組
- 22 噪音消去模組
- 24 遮蔽估測模組
- 26 反快速傅立葉轉換暨疊加模組
- 28 自動語音辨識模組

七、申請專利範圍：

1. 一種可消除噪音且增進語音品質之麥克風陣列架構，包括：

至少二麥克風，接收含有噪音訊號及語音訊號之至少二麥克風訊號；

至少二快速傅立葉轉換模組，將該等麥克風訊號轉換至頻率域；

一處理模組，計算該等麥克風訊號中噪音訊號及語音訊號之一夾角，並

依據該夾角選擇使用一相位差演算法配合遮蔽估測、一噪音消去法或

二者合併使用；

一相位差計算模組，計算該等麥克風訊號之一相位差及一耳間時間差，

並找出不同之該夾角所對應之該耳間時間差中最佳之一閾值；

一遮蔽估測模組，依據該閾值及一遮蔽法則得到一遮蔽訊號，再將該遮

蔽訊號乘上該等麥克風訊號之平均而得到該等麥克風訊號中之語音訊

號；以及

一反快速傅立葉轉換暨疊加模組，將該語音訊號由頻率域轉為時間域。

2. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，其中該閾值係利用一黃金比例搜尋法找出。

3. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，更包括一噪音消去模組，該夾角為零時於該噪音消去模組中使用噪音消去法。

4. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，其中該相位差計算模組係於該夾角大於零計算該相位差及該耳間時間差。

5. 如申請專利範圍第 3 項所述之可消除噪音且增進語音品質之麥克風陣列

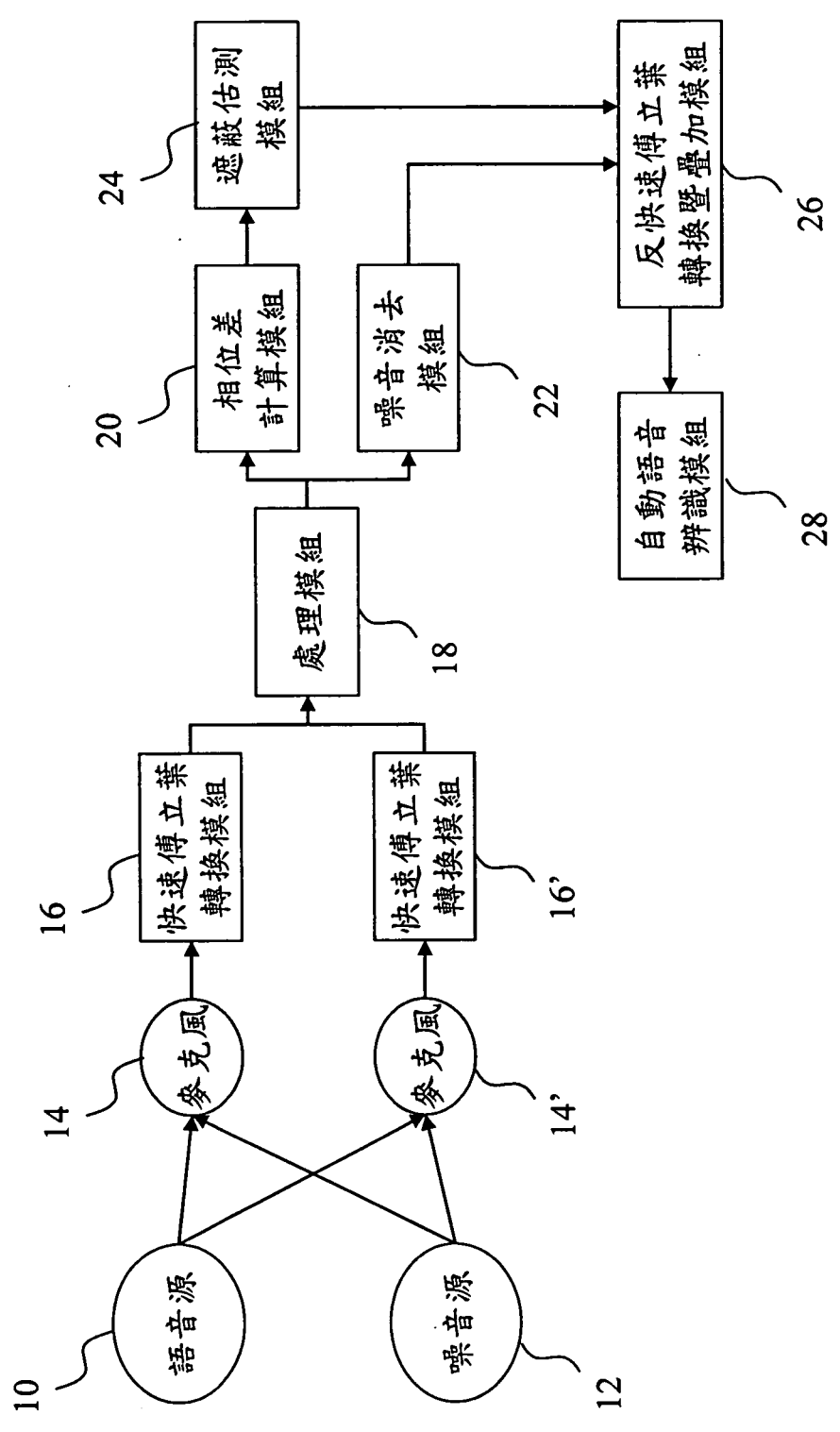
架構，其中該噪音消去模組與該相位差計算模組同時連接至該處理模組。

6. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，其中該反傅立葉疊加模組包含快速反傅立葉轉換及重疊相加法。
7. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，其中該語音訊號位於該等麥克風之正前方時，該耳間時間差為零。
8. 如申請專利範圍第 1 項所述之可消除噪音且增進語音品質之麥克風陣列架構，更包括一自動語音辨識模組，接收該反快速傅立葉轉換疊加模組輸出之該語音訊號，以進行語音辨識。
9. 一種可消除噪音且增進語音品質之麥克風陣列方法，包括下列步驟：
接收至少二麥克風訊號，並分別利用一快速傅立葉轉換模組轉至頻率域；
計算該等麥克風訊號中語音訊號及噪音訊號之一夾角，並依據該夾角選擇使用一相位差演算法配合遮蔽估測、一噪音消去法或二者合併使用以將該等麥克風訊號中之噪音訊號去除，保留該語音訊號；以及
將該語音訊號利用一反快速傅立葉轉換疊加模組轉至時間域輸出。
10. 如申請專利範圍第 9 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該反快速傅立葉轉換疊加模組係以反快速傅立葉轉換以及重疊相加法將頻率域之該語音訊號轉為一時域訊號。
11. 如申請專利範圍第 9 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該相位差演算法係於該夾角大於零時使用，更包含下列步驟：
計算該等麥克風訊號之相位差，以進一步找出一耳間時間差；
利用一黃金比例搜尋法找出對應不同之該夾角時該耳間時間差最佳之一閾值；以及

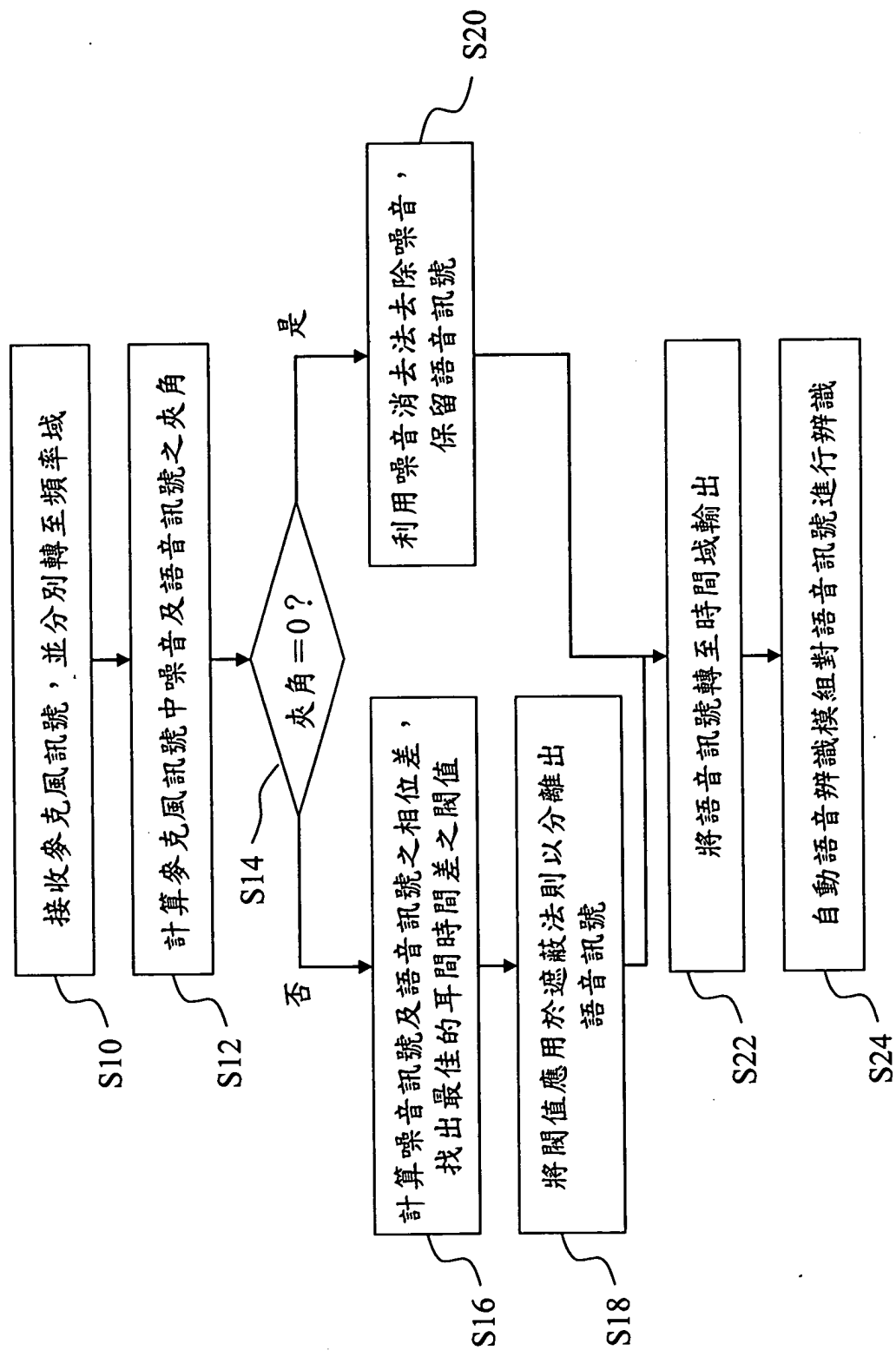
依據一遮蔽法則及該閾值得到一遮蔽訊號，將該等麥克風訊號之平均與該遮蔽訊號相乘得到該等麥克風訊號中之語音訊號。

- 12.如申請專利範圍第 11 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該語音訊號位於該等麥克風之正前方時，該耳間時間差為零。
- 13.如申請專利範圍第 9 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該夾角為零時使用該噪音消去法將該等麥克風訊號中之噪音訊號消除。
- 14.如申請專利範圍第 11 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該黃金比例搜尋法係在一連續範圍內任選兩點，比較該兩點之一函數值大小以將該連續範圍縮小，並重複任選兩點及比較函數值之步驟以將該連續範圍繼續縮小，找出該連續範圍內該函數值之一最小值。
- 15.如申請專利範圍第 14 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該閾值可利用該最小值搭配泰勒理論求得。
- 16.如申請專利範圍第 11 項所述之可消除噪音且增進語音品質之麥克風陣列方法，其中該耳間時間差小於該閾值時，將該麥克風訊號視為該語音訊號。
- 17.如申請專利範圍第 11 項所述之可消除噪音且增進語音品質之麥克風陣列方法，更包括利用一自動語音辨識模組接收該反快速傅立葉轉換暨疊加模組輸出之該語音訊號，以進行語音辨識。

八、圖式：



第 1 圖



第 2 圖