



(21)申請案號：100123330

(22)申請日：中華民國 100 (2011) 年 07 月 01 日

(51)Int. Cl. : **G10L21/02 (2006.01)**

(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)；劉嫻婷 LIOU, YING TING (TW)；桂振益 KUEI, CHEN YI (TW)；徐偉智 HSU, WEI CHIH (TW)

(74)代理人：林火泉

申請實體審查：有 申請專利範圍項數：20 項 圖式數：6 共 26 頁

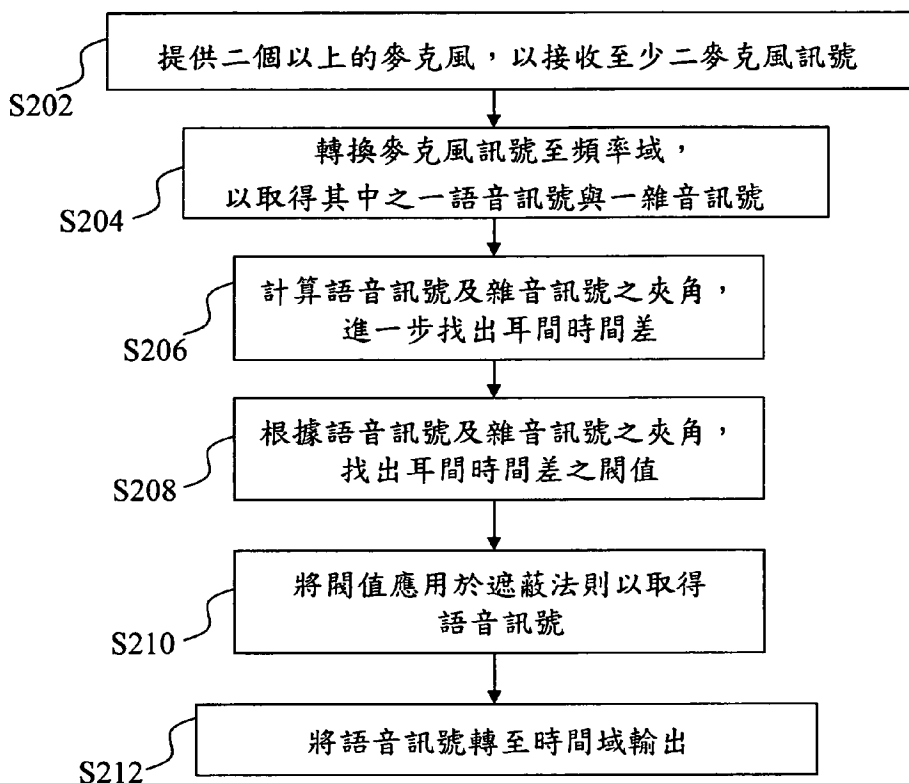
(54)名稱

可消除雜音且增進語音辨識率之方法

A METHOD FOR NOISE REDUCTION AND SPEECH ENHANCEMENT

(57)摘要

一種可消除雜音且增進語音辨識率之方法，包括以下步驟：提供二個以上的麥克風，用以接收至少二麥克風訊號；將麥克風訊號利用快速傅立葉轉換至頻率域，以取得其中之語音訊號與雜音訊號；計算語音訊號與雜音訊號之夾角，並使用一相位差演算法，進一步找出一耳間時間差；依據該夾角計算出耳間時間差之一閾值；依據耳間時間差與其閾值，使用一遮蔽法則，以取得語音訊號並去除雜音訊號；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，藉此去除聲源中的雜音部分，以增進語音品質。



# 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：100123330

※申請日：

※IPC 分類：G10L 21/02 (2006.01)

## 一、發明名稱：(中文/英文)

可消除雜音且增進語音辨識率之方法 / A method for noise reduction and speech enhancement

## 二、中文發明摘要：

一種可消除雜音且增進語音辨識率之方法，包括以下步驟：提供二個以上的麥克風，用以接收至少二麥克風訊號；將麥克風訊號利用快速傅立葉轉換至頻率域，以取得其中之語音訊號與雜音訊號；計算語音訊號與雜音訊號之夾角，並使用一相位差演算法，進一步找出一耳間時間差；依據該夾角計算出耳間時間差之一閾值；依據耳間時間差與其閾值，使用一遮蔽法則，以取得語音訊號並去除雜音訊號；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，藉此去除聲源中的雜音部分，以增進語音品質。

## 三、英文發明摘要：

A method for noise reduction and speech enhancement includes steps of providing at least two microphones to receive at least two microphone signals, transferring the two microphone signals into frequency domain by fast Fourier transform (FFT) to obtain a speech signal and a noise signal therein, calculating an angle between the speech and noise signal, and further obtaining an ITD by applying a phase difference estimation algorithm to the angle, calculating a threshold value of the ITD corresponding to the angle, applying a binary mask principle to the ITD and its threshold value to purify the speech signal out of the noise signal, and transferring the speech signal back into time domain to obtain a purified speech signal, thereby removing the noise signal and enhancing the speech quality.

四、指定代表圖：

(一)本案指定代表圖為：第 ( 2 ) 圖。

(二)本代表圖之元件符號簡單說明：

無。

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】

本發明係有關於一種消除麥克風雜音之方法，特別是一種可消除雜音並降低通訊中迴音干擾，以有效提升語音辨識率的方法。

### 【先前技術】

一般而言，麥克風接收聲音訊號之方式主要分為單通道及雙通道兩種，其中，單通道之消噪方式需要估算消噪比，而雙通道感應則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。

此種麥克風系統對人聲的敏感度較高，因而指向人的位置接收聲音訊號，而對背景的噪音則較不敏感。但此種麥克風系統由於包含有兩個或兩個以上的麥克風，其所形成之波束相當大，容易造成指向性不足的問題。

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數量眾多的麥克風、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。

其次，由於指向性不足的問題，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雜音的環境下，有效的消除雜音並不讓語音失真。

除此之外，一般的行動電話或車內通訊裝置也常具有在通話過程中因迴聲（echo）太大，而影響到通訊品質的問題。

因此，如何提出一種可有效消除環境中雜音，並增進語音品質之麥克風收音方法，係為熟習此項技術領域者亟需解決的問題之一。

### 【發明內容】

本發明之主要目的係在提供一種可消除雜音且增進語音辨識率之方法，其係利用黃金比例搜尋法搭配泰勒理論，計算得到最佳的耳間時間差閾值，使得每一個角度的語音訊號皆可得到最佳的語音品質。

本發明之另一目的係在提供一種可消除雜音且增進語音辨識率之方法，其係利用一複合式迴聲消除系統，濾除語音訊號之主要聲學迴音及環境擾動，藉此消去通訊過程中所產生的迴音，進一步地提高語音品質。

為達到上述之目的，本發明係有關於一種可消除雜音且增進語音辨識率之方法，其包括以下步驟：提供二個以上的麥克風，用以接收至少二麥克風訊號；將該些麥克風訊號利用快速傅立葉轉換至頻率域，以取得其中之一語音訊號與一雜音訊號；計算語音訊號與雜音訊號之夾角，並使用一相位差演算法，進一步找出一耳間時間差；依據語音訊號與雜音訊號之夾角計算出耳間時間差之一閾值；依據耳間時間差與閾值，使用一遮蔽法則，以取得語音訊號，去除雜音訊號；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。

本發明另有關於一種可消除雜音且增進語音辨識率之方法，其包括以下步驟：提供二個以上的麥克風，用以接收至少二麥克風訊號；將該些麥克風訊號利用快速傅立葉轉換至頻率域，以取得麥克風訊號中之一語音訊號與一雜音訊號；計算語音訊號與雜音訊號之夾角，並依據該夾角使用一相位差演算法配合遮蔽估測，以取得麥克風訊號中之語音訊號，去除雜音訊號；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；以及串聯一複合式迴聲消除系統於轉換回時間域之語音訊號後，以濾除語音訊號之聲學擾動。

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】

本發明提供一種可消除雜音且增進語音辨識率之方法，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，消除雜音，以增進語音品質。

請參考第 1 圖，係為根據本發明實施例可消除雜音且增進語音辨識率之麥克風陣列之示意圖，其包含有一麥克風陣列（包括至少二麥克風 14、14'）、至少二快速傅立葉轉換模組 16、16'、一運算模組 18、一遮蔽估測模組 20、以及一反快速傅立葉轉換暨疊加模組 22。

請參閱第 2 圖，係為根據本發明實施例之可消除雜音且增進語音辨識率之方法的步驟流程圖。以下關於此一實施例之實施方式的說明，請一併參照第 1 至 2 圖所示。

如步驟 S202 所示，語音源 10 及雜音源 12 之聲音傳送出去後，麥克風 14、14' 接收同時含有雜音訊號及語音訊號之麥克風訊號。

之後，如步驟 S204 所示，快速傅立葉轉換模組 16、16' 用以將麥克風 14、14' 收到之麥克風訊號轉換至頻率域，以取得麥克風訊號中之語音訊號與雜音訊號。

之後，如步驟 S206 所示，運算模組 18 連接麥克風 14、14'，用以計算麥克風訊號中語音訊號及雜音訊號之夾角為何。藉此，運算模組 18 依據此夾角，使用相位差演算法進一步找出耳間時間差（interaural time difference，ITD）。

如步驟 S208 所示，在運算模組 18 找出耳間時間差後，運算模組 18 進一步地計算出該耳間時間差對應每一雜音訊號與語音訊號之夾角的閾值為何。

之後，如步驟 S210 所示，遮蔽估測模組 20 依據算出的耳間時間差與閾值，利用一遮蔽法則，以取得語音訊號，去除雜音訊號。

最後，如步驟 S212 所示，反快速傅立葉轉換暨疊加模組 22 用以將語音訊號由頻率域轉回時間域，以得到去除雜音後具有較高語音辨識率之語音訊號。

其中，在步驟 S204 中，雜音訊號及語音訊號經由麥克風 14、14' 接收後，由快速傅立葉轉換模組 16、16' 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  $i$  個雜音源， $P_m$  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 S206 中，運算模組 18 計算此二麥克風訊號  $P_1(k,l)$  及  $P_2(k,l)$  中雜音訊號及語音訊號之夾角，亦即語音源 10 及雜音源 12 之間的夾角，以進一步找出耳間時間差 (ITD)。

一般而言，假設語音訊號在麥克風的正前方，則其耳間時間差為 0，其

他方向來的雜音則用  $d_i(k,l)$  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)：

$$P_1(k_j, l_j) \approx N_n(k_j, l_j) \quad (3)$$

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

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)：

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

之後，在步驟 S208 中，運算模組 18 係進一步地計算出耳間時間差對應雜音訊號與語音訊號之夾角的閾值為何。根據本發明之實施例，運算模組 18 計算最佳閾值的方法，係利用黃金比例搜尋法 (Golden-Section Search, GSS) 搭配泰勒理論，來找尋對應各個夾角的最佳閾值  $\tau$ 。

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  $\varepsilon$  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  $\tau$  值的函數如下式(12)：

$$\tau(i) = (-7.76 * 10^{-5})i^2 + (1.69 * 10^{-2})i - (5.45 * 10^{-2}) \quad (12)$$

其中  $i$  為語音訊號與雜音訊號之間的夾角，在此夾角  $i$  所對應的閾值  $\tau$  可以使經過處理的訊號有最佳的語音品質。

因此，在得到耳間時間差之最佳閾值  $\tau$  後，在步驟 S210 中，遮蔽估測模組 20 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  $\tau$  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 S210 取得語音訊號，以與雜音訊號成功分離之後，在步驟 S212

中，反快速傅立葉轉換暨疊加模組 22 將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，以得到去除雜音後具有較高語音辨識率之語音訊號。

請參照第 3 圖，其係為根據本發明另一實施例可消除雜音且增進語音辨識率之麥克風陣列之示意圖。如第 3 圖所示，在本發明提出之架構下，反快速傅立葉轉換暨疊加模組 22 更可連接有一自動語音辨識模組 24，用以接收反快速傅立葉轉換暨疊加模組 22 所輸出之語音訊號，以進行語音辨識。

其次，考量到若聲源位置不在麥克風陣列正前方時，本發明更提出一種波束轉向 (beam-steering) 的技術，其藉由將不同的延遲 (delay) 加入各個麥克風，來控制麥克風之波束轉向角度，使其轉至聲源位置。

假設轉向角度為  $\theta_M$ ，則波束轉向之濾波頻率因子可如下式(13)所示：

$$w_n = e^{-j n k d \sin \theta_M} = e^{\frac{-j \omega f_s n d \sin \theta_M}{c}} \quad (13)$$

其中  $n$  表第  $n$  個麥克風， $\omega$  是頻率因子， $f_s$  是取樣頻率， $d$  是麥克風間距。則在時域上，此濾波器即可依下式(14)所示，而寫成一延遲：

$$delay = \frac{f_s n d \sin \theta_M}{c} \quad (14)$$

由於上式延遲不是整數，因此必須使用拉格朗內插法 (Lagrange interpolation) 來使其更容易達成，此內插法可利用如下式(15)所示之無限脈衝響應系統 (Infinite Impulse Response Filter) 簡單的達成：

$$h(n) = \prod_{\substack{k=0 \\ k \neq n}}^N \frac{D-k}{n-k} \quad \text{for } n=0,1,2,\dots,N \quad (15)$$

其中  $N$  是此濾波器的階數，在此使用一階， $D$  則是延遲小數部份。

根據本發明之實施例，波束轉向的角度包括 0 度至 180 度。也就是說，在麥克風陣列接收到麥克風訊號之後，麥克風首先進行全方位（0°~180°）的波束轉向，並在每一次波束轉向後，進行頻譜分析計算耳間時間差，再通過如上式(6)的遮蔽法則，保留目標聲源並且抑制干擾。經過了上述語音純化的過程後，最後計算各麥克風在每一轉向角度之波束能量，以進行語音源方位的偵測（Direction of arrival estimation，DOA）。

其原因在於，當麥克風轉向到實際的聲源方位的時候應可得最大的能量（因為目標聲源的能量皆能通過上式(6)的遮蔽法則），以藉此判斷正確的聲源方向。其可如下式(16)來計算其能量大小：

$$\arg \max_{\lambda} \{power[\bar{P}(k,l) * e^{-jk\lambda}]\} \quad (16)$$

其中  $\bar{P}(k,l)$  為雙聲道訊號經過相位差演算法純化後的訊號； $(k,l)$  分別為頻率及時間的指數； $e^{-jk\lambda}$  是為頻率函數的波束轉向濾波器；而  $\lambda$  必須要如下式(17)所示，介於最大及最小的延遲時間內：

$$\frac{f_s d}{c} \leq \lambda \leq -\frac{f_s d}{c} \quad (17)$$

請參考第 4 圖，係為根據本發明另一實施例可消除雜音且增進語音辨識率之麥克風陣列之示意圖，其包括有麥克風陣列（包括至少二麥克風 14、14'）、至少二快速傅立葉轉換模組 16、16'、運算模組 18、遮蔽估測模組 20、反快速傅立葉轉換暨疊加模組 22、一固定式濾波器 26 與一適應性濾波器 28。

請參閱第 5 圖，係為根據本發明另一實施例之可消除雜音且增進語音辨識率之方法的步驟流程圖。以下關於此實施例之說明，請一併參照第 4

至 5 圖所示。

其中，步驟 S502 至步驟 S508 係同本發明前一實施例之步驟 S202 至步驟 S212 所示，故在此不再重述。值得注意的是，在此實施例中，本發明更包括步驟 S510：串聯一複合式迴聲消除系統（固定式濾波器 26 與適應性濾波器 28）於轉換回時間域之語音訊號後，以利用此複合式迴聲消除系統濾除掉語音訊號之聲學擾動。

詳細而言，當使用者與一遠端的第三者進行通話，且系統中具有一揚聲器 30，遠端第三者產生一遠端音訊 32 時，揚聲器 30 與麥克風 14、14' 將形成一固定的迴音路徑（echo path）。本發明係在反快速傅立葉轉換暨疊加模組 22 後串聯一固定式濾波器（fixed filter）26 與適應性濾波器（adaptive filter）28，以產生複合式迴聲消除系統。

在此一實施例中，固定式濾波器 26 係用以濾除語音訊號主要的聲學迴音，適應性濾波器 28 則用以濾除語音訊號於周遭環境中擾動所產生的問題。舉例而言，固定式濾波器在加入系統前可先以離線方式（off-line）將系統動態特性辨識出來。且此複合式系統的適應性演算法可以是但不限於 filterd-x LMS 演算法，其包括有：如下式(18)所示之整段迴聲消除路徑：

$$\hat{h}(n) = f(n) * [\delta(n) + \Delta w(n)] \quad (18)$$

$f(n)$  為固定式濾波器， $w(n)$  為適應性濾波器，其中  $w_0(n)$  為 1， $\Delta w(n) = [w_1(n) \ w_2(n) \ \cdots \ w_{L-1}(n)]$ ， $\delta(n)$  為單位脈衝數列；

如下式(19)所示之計算誤差訊號：

$$e(n) = d(n) - y(n) = d(n) - \mathbf{w}^T(n) [f(n) * \mathbf{x}(n)] \quad (19)$$

$d(n)$  為麥克風收入訊號， $y(n)$  為濾波器輸出訊號，

$\mathbf{w}(n)=[w_1(n) w_2(n) \cdots w_{L-1}(n)]^T$  為在時間  $n$  時的適應性濾波器係數組成向量，  
 $\mathbf{x}(n)=[x(n) x(n-1) \cdots x(n-L+1)]^T$  為在時間  $n$  時的輸入訊號向量；

以及依據下列公式(20)使用 FXLMS 演算法來更新適應性濾波器，

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}'^T(n)\mathbf{x}'(n)} e(n)\mathbf{x}'(n) \quad (20)$$

$$\mathbf{x}'(n) = f(n) * \mathbf{x}(n)$$

然而，值得注意的是，當遠端第三者產生遠端音訊 32 時，此時麥克風 14、14' 所收到的訊號將不只是系統產生的聲學迴音，於此將會造成適應性濾波器 28 的發散。有鑑於此，如第 6 圖所示，在確定更新適應性濾波器 28 之前，本發明另包括有步驟 S602 至 S608。

如步驟 S602 至步驟 S604 所示，系統中係包括有一雙邊對話偵測器 (double talk detector, DTD) 用以偵測語音源 10 產生的語音訊號與遠端第三者產生的遠端音訊 32 是否同時發生。之後，如步驟 S606 所示，若二者同時發生時 (意即使用者與遠端第三者同時說話)，則停止更新適應性濾波器 28。否則，如步驟 S608 所示，若二者未同時發生時 (意即使用者與遠端第三者未同時說話)，則繼續持續地更新適應性濾波器 28。

詳細而言，本發明主要比較麥克風 14、14' 收到的訊號與適應性濾波器 28 輸出的訊號。由於比較能量大小會造成適應性濾波器 28 開關太劇烈，因此便依據下式(21)至(23)所示，計算麥克風訊號  $d(n)$ 、固定式濾波器輸出訊號  $\mathbf{x}'(n)$  及適應性濾波器輸出訊號  $y(n)$  所形成的波封  $v_d(n)$ 、 $v_x(n)$  及  $v_y(n)$ ，  
 $\alpha = 0.99$ 。

$$v_x(n) = \alpha v_x(n-1) + (1-\alpha)|x(n)| \quad (21)$$

$$v_d(n) = \alpha v_d(n-1) + (1-\alpha)|d(n)| \quad (22)$$

$$v_y(n) = \alpha v_y(n-1) + (1-\alpha)|y(n)| \quad (23)$$

再依據下式(24)與(25)，由  $v_d(n)$ 、 $v_x(n)$  及  $v_y(n)$  求得偵測函數 (detection function)  $\xi(n)$  及動態門檻函數 (dynamic threshold function)  $T(n)$ 。當偵測函數  $\xi(n)$  大於動態門檻函數  $T(n)$  時便代表遠端第三者產生遠端音訊 32，適應性濾波器 28 也隨之停止更新。

當偵測函數  $\xi(n)$  小於動態門檻函數  $T(n)$  時，適應性濾波器 28 才繼續做更新。 $\gamma$  由實驗可得最佳值為 0.05，加入小正實數  $\beta$  係為防止偵測錯誤所預留的範圍。

$$\xi(n) = \frac{v_d(n)}{v_x(n) + \gamma} \quad (24)$$

$$T(n) = \frac{v_y(n)}{v_x(n) + \gamma} + \beta \quad (25)$$

綜上所述，本發明提出一種可消除雜音且增進語音辨識率之方法，其可將聲學訊號處理方法實現在電信通訊系統中。此種方法不僅可利用兩麥克風之間的相位差，獲得聲源角度進而決定波束開口大小，以增進語音辨識率，更可透過波束轉向自動偵測聲源位置。此外，利用波束轉向技術更可解決語音訊號不在主軸位置上的情況。

本發明提出之可消除雜音且增進語音辨識率之方法，並可應用於語音打斷 (barge in) 系統，結合複合式迴聲消除系統，有效地降低迴聲對辨識率的干擾。此系統適用於需要使用到語音辨識系統的手機、智慧型玩具等儀器內，使辨識系統即便在雜音及殘響嚴重的空間內，也能擁有不錯的辨識率。

以上所述之實施例僅係為說明本發明之技術思想及特點，其目的在使

熟習此項技藝之人士能夠瞭解本發明之內容並據以實施，當不能以之限定本發明之專利範圍，即大凡依本發明所揭示之精神所作之均等變化或修飾，仍應涵蓋在本發明之專利範圍內。

**【圖式簡單說明】**

第 1 圖係為根據本發明實施例之可消除雜音且增進語音辨識率之麥克風陣列之示意圖。

第 2 圖係為根據本發明實施例之可消除雜音且增進語音辨識率之方法的步驟流程圖。

第 3 圖係為根據本發明另一實施例之可消除雜音且增進語音辨識率之麥克風陣列之示意圖。

第 4 圖係為根據本發明另一實施例可消除雜音且增進語音辨識率之麥克風陣列之示意圖。

第 5 圖係為根據本發明另一實施例之可消除雜音且增進語音辨識率之方法的步驟流程圖。

第 6 圖係為根據第 5 圖在更新適應性濾波器之前的步驟流程圖。

**【主要元件符號說明】**

10 語音源

12 雜音源

14、14' 麥克風

16、16' 快速傅立葉轉換模組

18 運算模組

20 遮蔽估測模組

- 22 反快速傅立葉轉換暨疊加模組
- 24 自動語音辨識模組
- 26 固定式濾波器
- 28 適應性濾波器
- 30 揚聲器
- 32 遠端音訊



## 七、申請專利範圍：

1. 一種可消除雜音且增進語音辨識率之方法，包括以下步驟：

提供二個以上的麥克風，用以接收至少二麥克風訊號；

將該些麥克風訊號利用快速傅立葉轉換至頻率域，以取得該些麥克風訊號中之一語音訊號與一雜音訊號；

計算該語音訊號與該雜音訊號之一夾角，並使用一相位差演算法，進一步找出一耳間時間差；

依據該語音訊號與該雜音訊號之該夾角計算出該耳間時間差之一閾值；

依據該耳間時間差與該閾值，使用一遮蔽法則，以取得該語音訊號，去除該雜音訊號；以及

將該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。

2. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中該閾值的計算係利用一黃金比例搜尋法（Golden-Section Search）搭配泰勒理論求得。
3. 如請求項 2 所述之可消除雜音且增進語音辨識率之方法，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，該閾值係利用該最小值搭配泰勒理論求得。
4. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中使用該遮蔽法則時更包括步驟：  
比較該耳間時間差與該閾值之大小，以得到一遮蔽訊號；以及

將該等麥克風訊號之平均與該遮蔽訊號相乘，以取得該些麥克風訊號中之該語音訊號。

5. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中該反快速傅立葉轉換暨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。
6. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中該語音訊號位於該些麥克風之正前方時，該耳間時間差為零。
7. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，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。
8. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中該些麥克風係排列為一陣列。
9. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，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。
10. 如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中在接收該些麥克風訊號後，更包括步驟：
  - 針對各該麥克風加入一延遲，以控制該些麥克風之波束轉向角度。
11. 如請求項 10 所述之可消除雜音且增進語音辨識率之方法，其中在針對各該麥克風加入該延遲後，更包括步驟：
  - 計算各該麥克風在每一轉向角度之波束能量，以判斷該語音訊號的聲源方向。
12. 如請求項 11 所述之可消除雜音且增進語音辨識率之方法，其中該些麥克

風之波束轉向角度包括 0 度至 180 度。

13.如請求項 1 所述之可消除雜音且增進語音辨識率之方法，其中在轉換該

語音訊號至時間域之後，更包括步驟：

串聯一複合式迴聲消除系統於轉換回時間域之該語音訊號後，以濾除該語音訊號之聲學擾動。

14.如請求項 13 所述之可消除雜音且增進語音辨識率之方法，其中該複合式

迴聲消除系統包括一固定式濾波器與一適應性濾波器，該固定式濾波器濾除該語音訊號之聲學迴音，該適應性濾波器濾除該語音訊號於環境中所造成之擾動。

15.如請求項 14 所述之可消除雜音且增進語音辨識率之方法，更包括步驟：

利用一 FXLMS 演算法更新該適應性濾波器。

16.如請求項 15 所述之可消除雜音且增進語音辨識率之方法，其中在更新該

適應性濾波器前，更包括步驟：

提供一雙邊對話偵測器（double talk detector，DTD）；

偵測該語音訊號與一遠端音訊是否同時發生；以及

當該語音訊號與該遠端音訊同時發生時，停止更新該適應性濾波器。

17.一種可消除雜音且增進語音辨識率之方法，包括以下步驟：

提供二個以上的麥克風，用以接收至少二麥克風訊號；

將該些麥克風訊號利用快速傅立葉轉換至頻率域，以取得該些麥克風訊號中之一語音訊號與一雜音訊號；

計算該語音訊號與該雜音訊號之一夾角，並依據該夾角使用一相位差演算法配合遮蔽估測，以取得該些麥克風訊號中之該語音訊號，去除該

雜音訊號；

將該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；以

及

串聯一複合式迴聲消除系統於轉換回時間域之該語音訊號後，以濾除該

語音訊號之聲學擾動。

18.如請求項 17 所述之可消除雜音且增進語音辨識率之方法，其中該複合式迴聲消除系統包括一固定式濾波器與一適應性濾波器，該固定式濾波器濾除該語音訊號之聲學迴音，該適應性濾波器濾除該語音訊號於環境中所造成之擾動。

19.如請求項 18 所述之可消除雜音且增進語音辨識率之方法，更包括步驟：利用一 FXLMS 演算法更新該適應性濾波器。

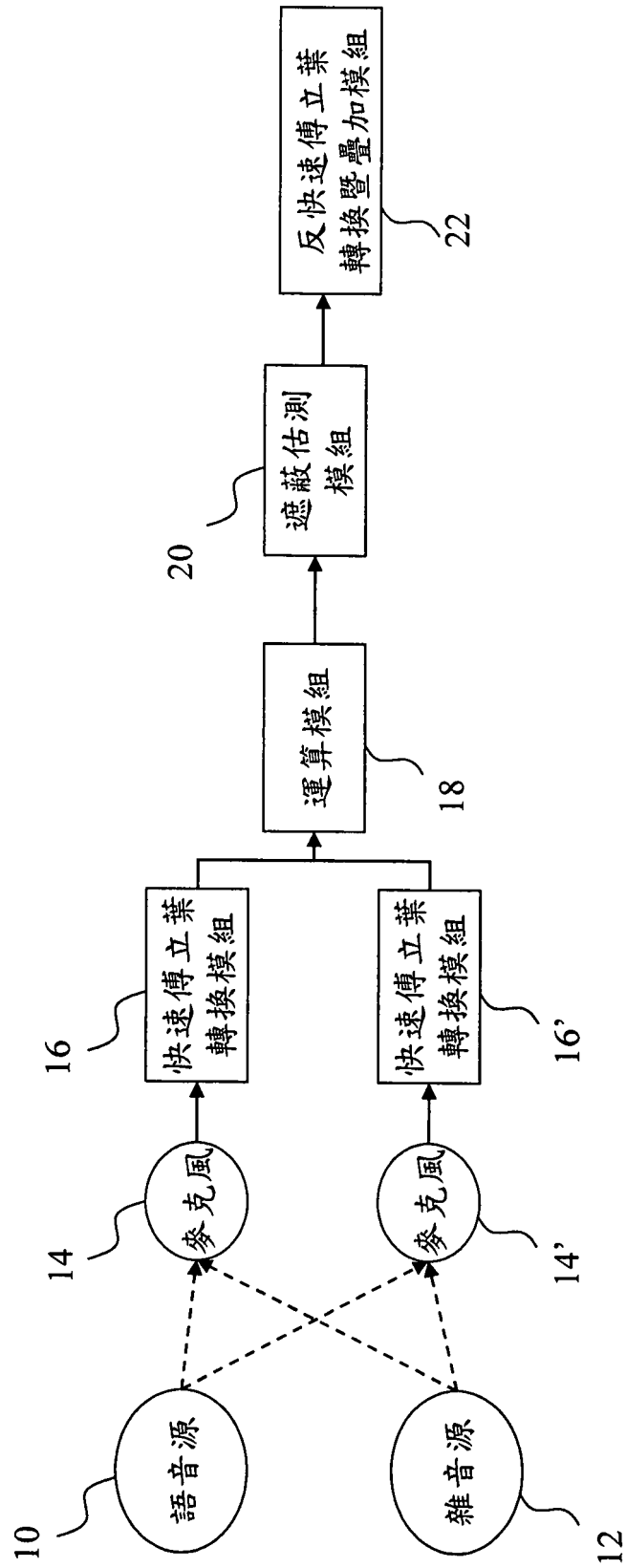
20.如請求項 19 所述之可消除雜音且增進語音辨識率之方法，其中在更新該適應性濾波器前，更包括步驟：

提供一雙邊對話偵測器 (double talk detector, DTD)；

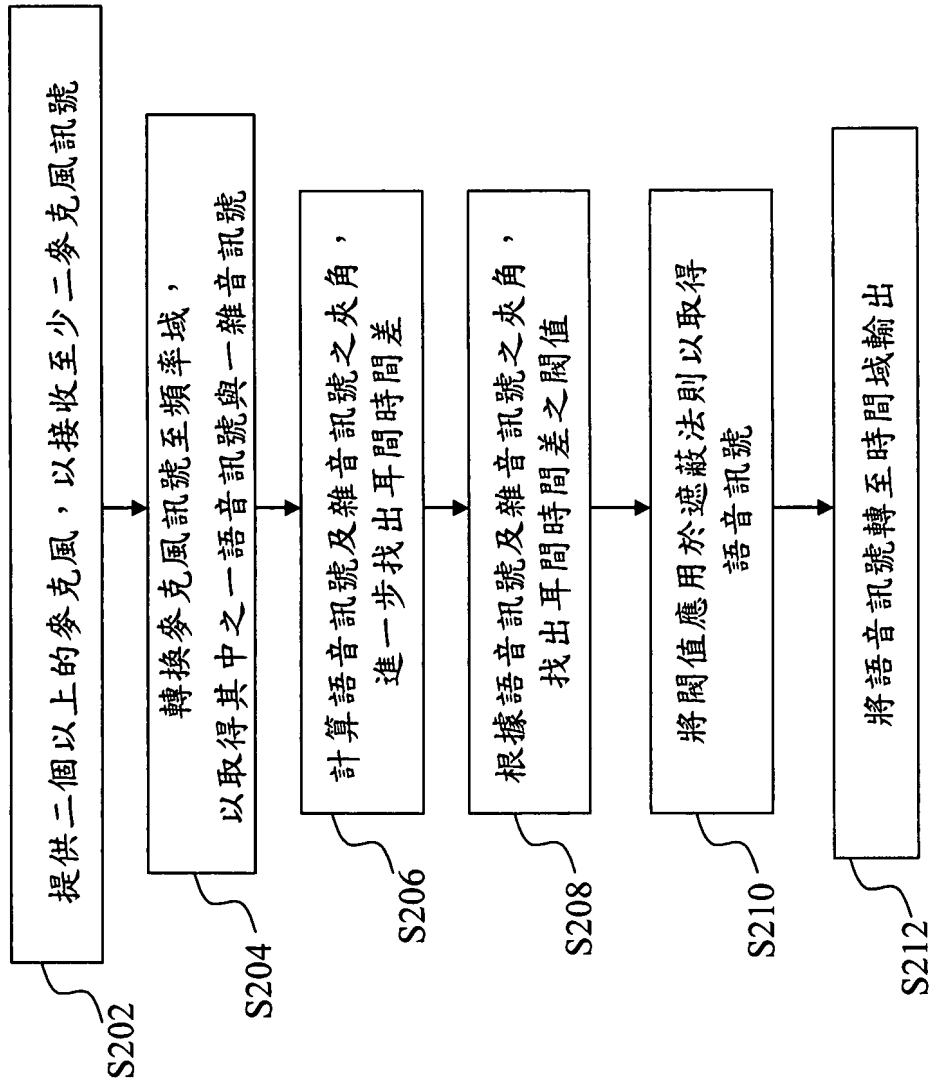
偵測該語音訊號與一遠端音訊是否同時發生；以及

當該語音訊號與該遠端音訊同時發生時，停止更新該適應性濾波器。

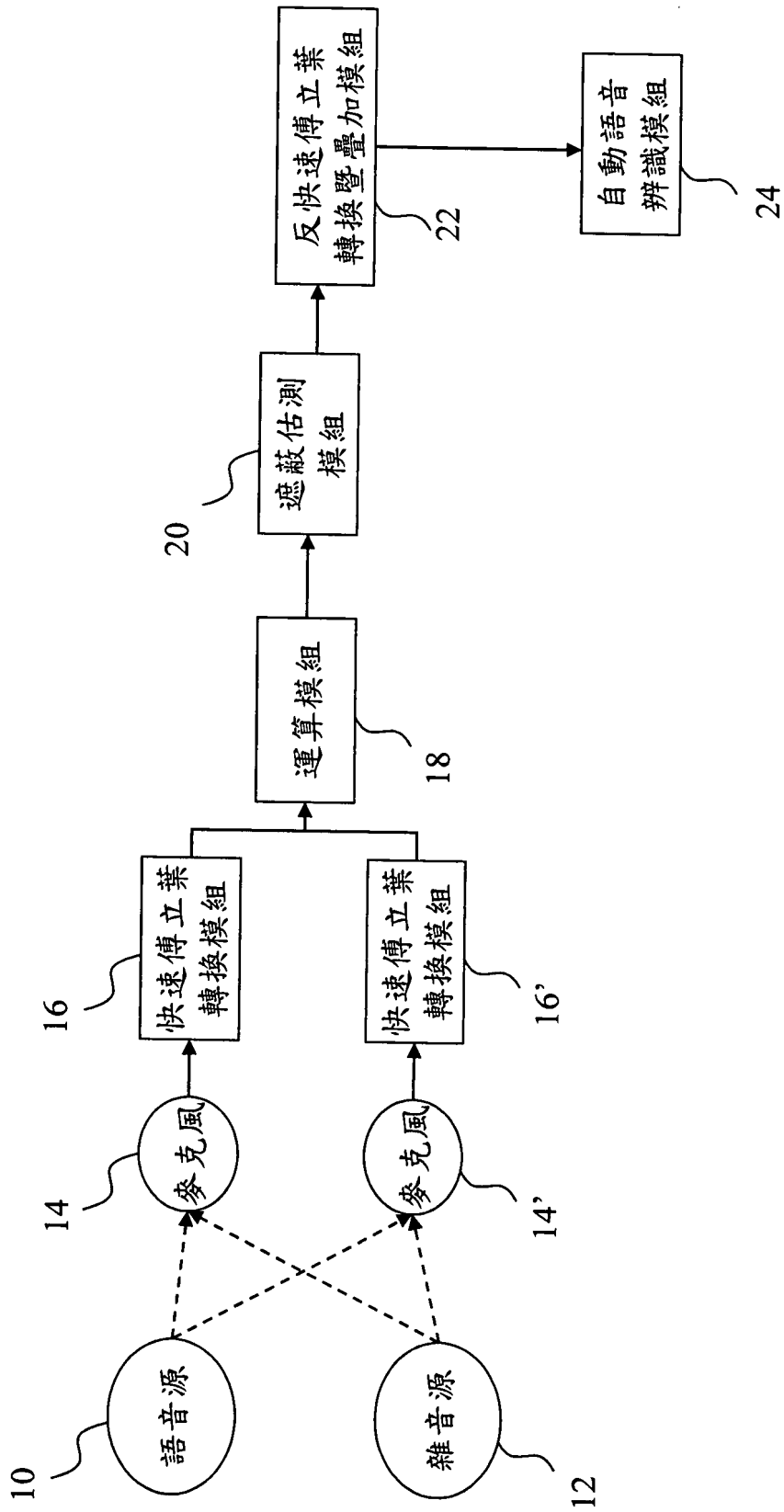
八、圖式：



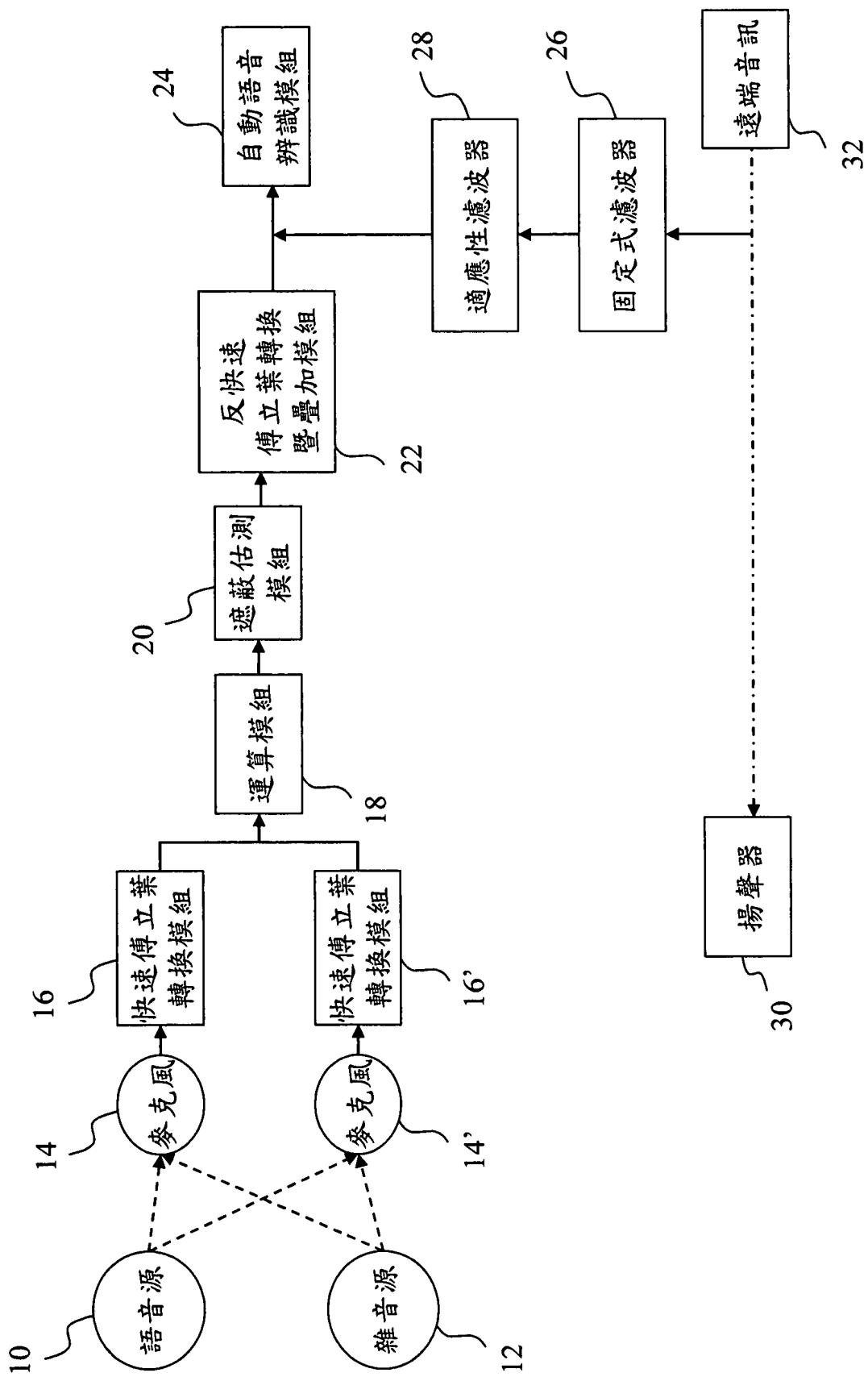
第 1 圖



第 2 圖

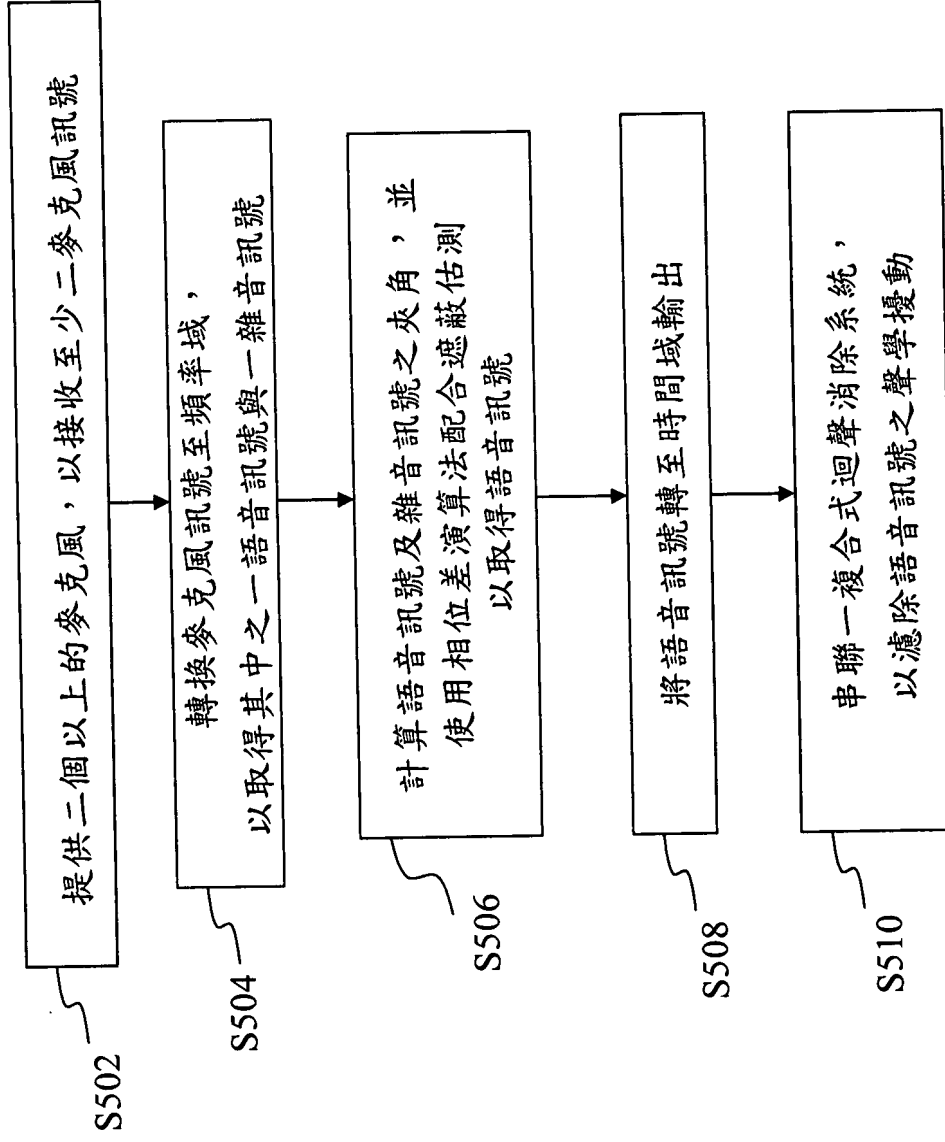


第 3 圖

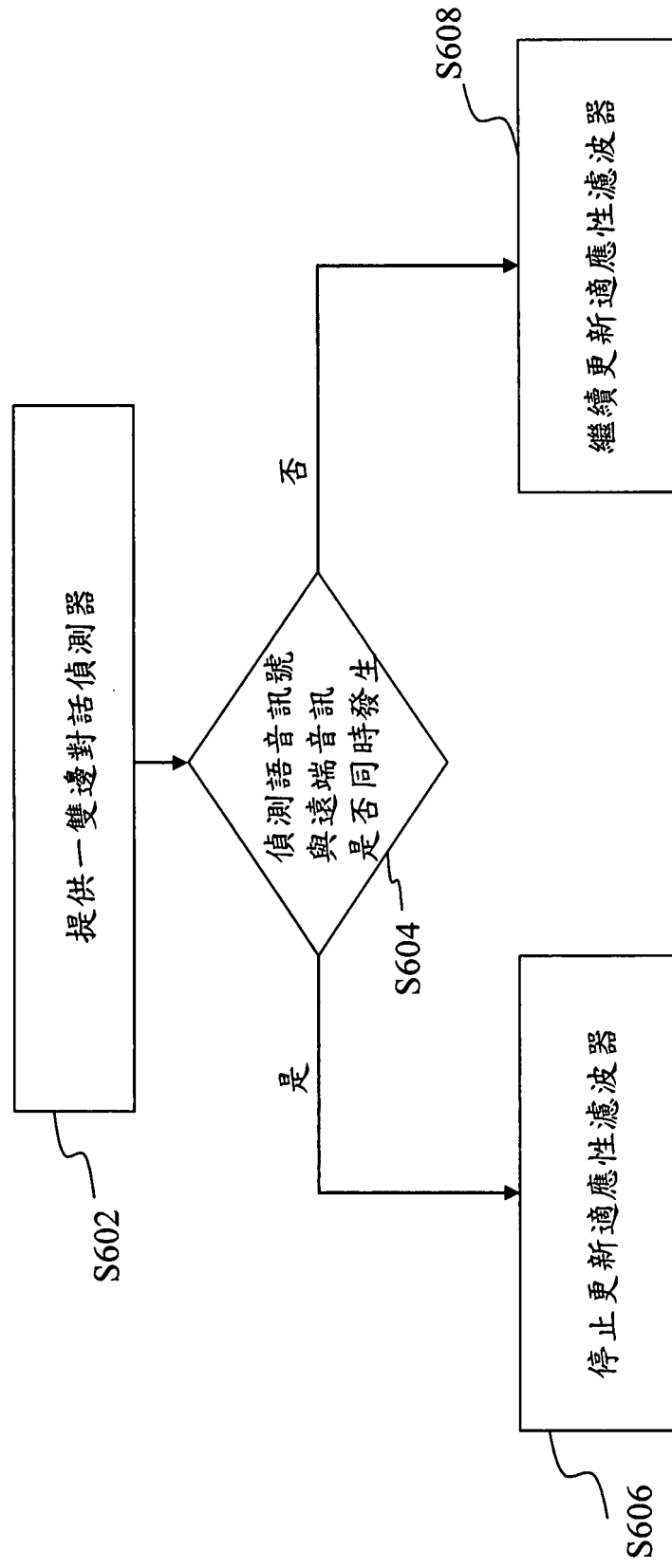


第 4 圖





第 5 圖



第 6 圖