發明專利說明書 200636561

（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：

※ 申請日期：

※IPC 分類：

一、發明名稱：（中文/英文）

語音定位方法與系統

二、申請人：（共 1 人）

姓名或名稱：（中文/英文）

國立交通大學 / National Chiao Tung University

代表人：（中文/英文）張俊彥 / Chun-Yen Chang

住居所或營業所地址：（中文/英文）

新竹市大學路 1001 號 / 1001 Ta Hsueh Rd.,Hsinchu, Taiwan

國　籍：（中文/英文）中華民國

三、發明人：（共 3 人）

姓　名：（中文/英文）

1.胡竹生

2.劉維瀚

3.鄭价呈

國　籍：（中文/英文）

1.中華民國

2.中華民國

3.中華民國

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：

本發明係為一語音定位方法與系統，將聲音訊號藉由麥克風陣列接收並傳送至語音偵測系統，以決定系統之運作狀態流程，再將聲波訊號傳送至環境參數訓練子系統藉由加成性原理模擬產生訓練訊號，並將該訓練訊號經由特徵之抽取及描述此特徵之分佈狀況以產生一統計參數，最後將麥克風陣列所擷取之聲音訊號及統計參數驅動語音位置偵測子系統以偵測語者位置。

六、英文發明摘要：

七、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:

麥克風陣列 1

語音偵測系統 2

環境參數訓練子系統 3

語者參考訊號之記憶體 31

環境噪音訊號之記憶體 32

合併器 33

相位差特徵抽取模組 34

特徵統計參數化模組 35

語音位置偵測子系統 4

位置偵測模組 41

統計參數更新模組 42

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】

本發明係提供一語音定位方法與系統，特別是本發明係由一麥克風陣列、一語音偵測系統、一環境參數訓練子系統及一語音位置偵測子系統組成，可運用於積體電路設計、個人電腦、視訊會議室、車內空間、機器人、保全監控系統及家庭自動化等相關產業中。

【先前技術】

一般習用之語音定位技術已有各種不同的方式藉由單一麥克風陣列（兩顆以上之麥克風）來達到語者定位的目的，雖然習用之技術在某些條件下可達到一定的效果，但由於該單一麥克風陣列之定位技術大多以角度來區分不同的語者，當遇到語者與麥克風間有障礙物存在，或同方向有兩個以上語者時，即使在理想的環境中，該定位技術仍然無法成功的分辨語者的差異，此外，由於現實狀況中尚有許多不理想之特性（如：近場效應、麥克風間的不匹配、環境噪音、迴音反射與語音訊號在空氣中傳遞時的聲場暫態反應等），習用之技術並未對上述的問題同時加以考慮，而使其效果降低，造成應用層面相對受到侷限。

請參閱『表 3』所示，係習用之語音定位技術及本發

明之語音定位方法與系統之比較表。該習用之語音定位技術之說明如下：

1.美國專利公告第6,826,284號由Benesty等人在提出利用適應性特徵值分解理論（AEDA－Adaptive Eigenvalue Decomposition Algorithm）配合最小均方演算法（LMS）來估測通道轉移函數（channel transfer function），再結合時間差聲源角度定位法（TDOA）之計算方式解決有物體擋住及高度反射的問題，但由於該通道必須假設為線性和時間不變性（time invariant），然而，上述之假設無法辦法解決暫態響應的問題，且必須在無噪音干擾及麥克風間互相匹配下才可以使用。

2.美國專利公報（Pub. No.：US 2004/0013275）中由 Balan 等人提出利用求取共變異矩陣（covariance matrix）求取角度，但由於必須要假設噪音本身為高斯分布且其期望值為零才可將噪音的影響降到最低，然而，該假設通常是不切實際的，所以估測會有一定的不準確性，且不能解決有物體擋住和同方向但不同距離的問題，並必須在麥克風間互相匹配下才可使用。

3.美國專利公告第6,449,593號由Valve提出利用波束形成器（beamformer）在想要判斷的角度上形成波束，即可壓抑周遭環境之噪音，並針對那些角度增強聲

源訊號的能量，最後比較經過該波束形成器後的能量來決定聲源的角度。雖上述之方法可以在噪音環境下使用，但仍無法解決有物體擋住和同方向但不同距離的問題，且必須在麥克風間互相匹配下才可使用。

4.美國專利公告第 6,243,471 號由 Brandstein 等人提出利用三個以上之麥克風作為一個群組，再利用簡單的幾何關係找出三維空間的資訊，因此，只要有複數個群組就可產生複數個三維空間資訊，即便可估測出聲源方位，亦不會有物體擋住和同方向但不同距離的問題，但上述之方法若使用在複雜的環境下，其所需的麥克風陣列數目則相當可觀，且由於利用時間差聲源角度定位法（TDOA），在高反射和暫態的時候就會產生角度上的誤差，而利用簡單的幾何關係求解會因不準而有更大的誤差，雖可利用了求取變異數（variance）配合高斯分布的假設來製造出不同的權重以減少誤差，但若有雜訊存在時，該高斯分布的假設就不能同時適用於雜訊和聲源同時存在的狀況，其估測出來的角度會產生誤差而不準，且必須在麥克風間互相匹配下才可使用。

5.美國專利公告第 5,778,082 號由 Chu 等人提出利用簡略的聲源偵測系統對是否發生需要估測之聲音加以分辨，以求找出雜訊區段，並預先估測雜訊之共相關矩

陣（crosscorrelation matrix），而再利用該共相關矩陣與欲估測之聲源的共相關矩陣作相減動作，以達到消去雜訊所造成之影響的效果。但由於上述之方法並沒有針對語音偵測設計，因此無法穩定的偵測出語音發生，此外，若雜訊之共相關矩陣估測不準確，則會造成估測結果之誤差，亦無法對被擋住的物體作分辨，且必須使用匹配的麥克風。

6.美國專利公告第5,465,302號由Lazzari等人提出利用複數個麥克風並以兩兩計算出時間延遲（time delay）後，再利用非平面波假設原理計算出聲源與麥克風陣列之相對位置。上述之方法雖可估測出聲源之位置，但需要麥克風間互相匹配，且對於環境噪音、反射和聲源與麥克風陣列間有障礙物等問題則無法解決。

7. 美國專利公告第4,333,170號由Mathews等人提出利用相位差異斜率（phase difference slope），並加以分群以計算出聲源對麥克風陣列之角度關係，並利用訊號頻譜之能量強度找出適合之頻率來加以分群，由於上述之方法在聲音與雜訊包含同樣之頻帶時會發生誤差，因而導致估測結果不準確，同時，該方法亦需麥克風間互相匹配，且不能分辨被擋住之聲源或同樣角度，距離不同之聲源。

8. 由 Lo 等人於 IEEE TRANSACTIONS ON INSTRUMENTION AND MEASUREMENT, VOL.53,NO.4,AUGUST 2004 所發表的 Robust Joint Audio-Vvideo Localization in Video Conferencing Using Reliability Information 中提出一種同時利用影像與聲音訊號的方法來達到語者定位的效果，該方法的聲音訊號定位部分只利用了簡單的延遲加成波束型成器（delay and sum beamformer）來計算不同區域（section）的聲音能量，並尋找最大能量的區域來當作可能的語者位置，再與該影像之部分判斷融合，以達到語者定位的目標，但由於該方法需要一組圓形麥克風陣列放置在語者的中間，才能夠分出不同區域的聲音能量，而該區域由麥克風陣列的圓心放射狀區分，並不方便使用，若要達到較佳之穩定結果時，需要影像的整合，該方法之系統架構複雜，且價格昂貴，不利使用。

9. IEEE TRANSACTIONS ON NEUWORAL NETWORKS, VOL.11,NO.4,JULY 2000 由 Guner Arslan 等人所發表的 A Unified Neural-Network-Based Speaker Localization Technique 中使用了以類神經網路（neural-network）為基礎的技術來作聲源定位，當訊號雜訊比（SNR）高過20dB的時候，即使是大角度的定位

依然有很好的效果，且可用於近場（near-field）和遠場（far-field）的應用中，但無法使用在周遭環境之噪音大的時候，且也無法解決物體擋住和同方向但不同距離的問題，並必須在麥克風間互相匹配下才可使用。

10. IEEE TRANSACTIONS ON SPEECH AND AUDIO PRESSING, VOL.8, NO.2, MARCH 2000 由 James G. Ryan 等人所發表的 Aarray Optimization Applied in the Near Field of a Microphone Array 中提出當麥克風間距離等於一半波長($d=\frac{\lambda}{2}$)的時候會有最好的效果，因此在作定位的時候特別將$d=\frac{\lambda}{2}$和$d<\frac{\lambda}{2}$這兩個情況分開討論，且上述之方法最好的運用情況是聲源在近場（near-field），而雜訊是在遠場（far-field）的時候，該方法也無法解決物體擋住和同方向但不同距離的問題，且必須在麥克風間互相匹配下才可使用。

11.Huang 等人在 IEEE TRANSACTIONS ON INSTRUMENTION AND MEASUREMENT, VOL.44,NO.3,JUNE 1995 所發表的 A Biomimetic System for Localization and Separation of Multiple Sound Sources 中提出一種以「抵達時間差」（Arrival Temporal Disparities, ATD）的方式來計算聲源角度，該方法必須偵測聲源的起始點（onset），並利用該起始點

短時間內沒有被迴音污染的特性來達到比較精確的效果，然而，該方法卻相當依賴該起始點估測的穩定性，當聲源起始前若沒有安靜區間，則偵測結果將會發生誤判，而導致後面的ATD統計發生誤差，而使得效果下降，此外，該方法無法對付近場（near-field）聲源、麥克風特性誤差等問題，且只能分辨出不同方向之聲源，對距離不同則無分辨能力。故，一般習用者係無法符合使用者於實際使用時之所需。

【發明內容】

因此，本發明之主要目的係在於提供一可解決近場效應、麥克風間的不匹配、迴音反射與語音訊號在空氣中傳遞時的聲場暫態反應問題之語音定位方法與系統。

為達上述之目的，本發明係提供一語音定位方法與系統，係由一麥克風陣列、一語音偵測系統、一環境參數訓練子系統及一語音位置偵測子系統組成，當一聲音訊號藉由該麥克風陣列接收並傳送至該語音偵測系統，利用該語音偵測系統決定系統之運作狀態流程，再將該聲波訊號傳送至該環境參數訓練子系統藉由加成性原理模擬產生一訓練訊號，並將該訓練訊號經由特徵之抽取及描述此特徵之分佈狀況以產生一統計參數，最後將該麥克風陣列所擷取之聲音訊號及該統計參數驅動該語音位置偵測子系統以偵測語者位置，可適應不同之環境及

抵抗雜訊所造成的影響，具有高準確性之偵測結果、低成本、彈性與自動適應環境調整等優點，可運用於積體電路設計、個人電腦、視訊會議室、車內空間、機器人、保全監控系統及家庭自動化等相關產業中。

【實施方式】

請參閱『第１圖』所示，係本發明之語音定位方法與系統示意圖。如圖所示：本發明之語音定位系統係由一麥克風陣列１、一語音偵測系統２、一環境參數訓練子系統３及一語音位置偵測子系統４組成，該麥克風陣列１包含至少2顆以上之麥克風，該語音偵測系統２包含一語者參考訊號之記憶體３１、一環境噪音訊號之記憶體３２、一合併器３３、一相位差特徵抽取模組３４及一特徵統計參數化模組３５，而該語音位置偵測子系統４由一位置偵測模組４１及一統計參數更新模組４２組成。當一聲音訊號藉由該麥克風陣列１接收並傳送至該語音偵測系統２，利用該語音偵測系統２決定系統之運作狀態流程，再藉由該語者參考訊號之記憶體３１及該環境噪音訊號之記憶體３２利用聲波訊號之加成性原理以模擬真實之訓練訊號，再利用該合併器３３將該訓練訊號產生並傳送至該相位差特徵抽取模組３４，藉由該相位差特徵抽取模組３４抽取出一與語者無關但與位置有關之特徵向量資訊，再將該特徵向量資訊傳送至該特徵統計參數化模組３５用以描述特性分佈狀況並產生一

統計參數，將該統計參數與該麥克風陣列１所擷取之聲音訊號傳送至該位置偵測模組４１，藉由該位置偵測模組４１得到一語者位置之偵測結果，並將該偵測結果傳送至該統計參數更新模組４２中，以判斷是否具有足夠的資訊可以提供統計參數而進行更新的動作，如此便可針對複雜環境與噪音作出適應調整，使該語音定位系統具有穩健偵測之特性而能在各種場合中運用。

本發明之語音定位方法與系統係可分為該環境參數訓練子系統３及該語音位置偵測子系統４兩個子系統，當聲音訊號經由該麥克風陣列１接收後，藉由該語音偵測系統２將系統之運作狀態流程區分為三個運作狀態，請參閱『表１』所示，係本發明之語音定位方法與系統之運作狀態對應表。該運作狀態之說明如下：

（a） 運作狀態1：

該麥克風陣列１只接收語音訊號，亦即在無環境噪音且只有語音存在之狀態中，將更新該語者參考訊號之記憶體３１之語者參考訊號 $S(n)$，以作為語者空間特徵的參考指標。

（b） 運作狀態2：

該麥克風陣列１所擷取到之訊號為環境之噪音，亦即在無語音存在狀態中，將更新該環境噪音訊號之記憶體３２之環境噪音訊號 $N(n)$，並將該語者參考訊號和該環境噪音訊

號藉由該合併器３３組合出一訓練訊號$X(n)$，再傳送至該相位差特徵抽取模組３４及該特徵統計參數化模組３５（如：混合高斯模型（GMM）、核心基礎模型（Kernel based model））中抽取出一與語者無關但與位置有關之特徵向量資訊，並描述其混合相位差特性分佈之狀況，進而產生一統計參數。該該相位差特徵抽取模組３４已將空間訊號交疊效應考慮在內，且以使用不同之麥克風對組合來分配不同頻帶之訊號（如表２所示），若使用之麥克風陣列為線性均勻排列方式，且麥克風之間的距離為d，麥克風個數為M，則可分出$M-1$個頻帶。

（c） 運作狀態3：

在具有語音存在的一般使用狀態下，將上述之統計參數傳送至該位置偵測模組４１中，且將該麥克風陣列１所收到之聲音訊號導入該位置偵測模組４１中得到一語者位置之偵測結果（如：選取模型中後驗機率最大者），並將該偵測結果傳送至該統計參數更新模組４２中，藉以判斷該偵測結果是否具有足夠的資訊可以提供統計參數進行更新的動作，若該判斷之結果為正面的，則將進行

統計參數更新以利進一步的適應空間環境變化，並將更新後之統計參數重新導入該位置偵測模組４１中（如：使用遺忘因子或期望最大化理論更新混合高斯模型），以進行下一次的語者位置偵測。

本發明之語音定位方法與系統係利用預先錄製之語音訊號配合相位差特徵抽取模組及特徵統計參數化模組，可在同樣角度但不同距離，或麥克風與語者間有障礙物的狀況下維持位置偵測之高準確性，並可解決近場效應、麥克風間的不匹配、迴音反射與語音訊號在空氣中傳遞時的聲場暫態反應問題，並藉由環境參數訓練子系統可使得系統適應不同之環境，抵抗雜訊所造成的影響，且配合統計參數更新模組使得預錄之語音訊號在使用過程中可即時更新，以達到適應語者位置些微變動仍可有高準確性的偵測結果，該語音定位方法與系統具有低成本、彈性與自動適應環境調整之優點，可運用於積體電路設計、個人電腦、視訊會議室、車內空間、機器人、保全監控系統及家庭自動化等相關產業中。

綜上所述，本發明之語音定位方法與系統，以「環境參數訓練子系統及一語音位置偵測子系統」作為主要策略，可有效改善習用方法之種種缺點，使其可解決近場效應、麥克風間的不匹配、迴音反射與語音訊號在空氣中傳遞時的聲場暫態反應問題，適應不同之環境及抵

抗雜訊所造成的影響，並具有高準確性之偵測結果、低成本、彈性與自動適應環境調整等優點，進而使本發明之產生能更進步、更實用、更符合使用者之所需，確已符合發明專利申請之要件，爰依法提出專利申請，尚請貴審查委員撥冗細審，並盼早日准予專利以勵創作，實感德便。

惟以上所述者，僅為本發明之較佳實施例而已，當不能以此限定本發明實施之範圍；故，凡依本發明申請專利範圍及發明說明書內容所作之簡單的等效變化與修飾，皆應仍屬本發明專利涵蓋之範圍內。

【圖式簡單說明】

第１圖，係本發明之語音定位方法與系統示意圖。

表１，係本發明之語音定位方法與系統之運作狀態對應表。

表２，係本發明之語音定位方法與系統之不同麥克風與頻帶對應表。

表３，係習用之語音定位技術及本發明之語音定位方法與系統之比較表。

【主要元件符號說明】

麥克風陣列　1

語音偵測系統　2

環境參數訓練子系統　3

語者參考訊號之記憶體　３１

環境噪音訊號之記憶體　３２

合併器　３３

相位差特徵抽取模組　３４

特徵統計參數化模組　３５

語音位置偵測子系統　4

位置偵測模組　４１

統計參數更新模組　４２

表1

	語者參考訊號之記憶體	環境噪音訊號之記憶體	相位差特徵抽取模組	特徵統計參數化模組	位置偵測模組	統計參數更新模組
運作狀態1	更新	X	X	X	X	X
運作狀態2	X	更新	執行	執行	X	X
運作狀態3	X	X	X	X	執行	執行

200636561

表 2

頻帶編號	麥克風對組合	麥克風對組合數	各頻帶所屬之頻率區間
頻帶 1 $(b=1)$	$(m, m+M-1)$ 其中 $m=1$	$J_1=1$	$0 \le f \le \frac{c}{2(M-1)d}$
頻帶 2 $(b=2)$	$(m, m+M-2)$ 其中 $1 \le m \le 2$	$J_2=2$	$\frac{c}{2(M-1)d} < f \le \frac{c}{2(M-2)d}$
⋮	⋮	⋮	⋮
頻帶 $M-1$ $(b=M-1)$	$(m, m+1)$ 其中 $1 \le m \le M-1$	$J_{M-1}=M-1$	$\frac{c}{4d} < f \le \frac{c}{2d}$

200636561

表 3
（前案）

	近場與遠場之應用	反射環境之穩健性	雜訊環境之穩健性	麥克風匹配性需求	同方向但不同距離之語者判別能力	有遮蔽之語者判別能力
US 6,826,284	可	高	低	高	低	低
2004　US 0,013,275	否	高	中	高	低	低
US 6,449,593	否	中	高	高	低	低
US 6,243,471	否	低	低	高	高	高
US 5,778,082	否	低	中	高	低	低
US 5,465,302	可	低	低	高	低	低
US 4,333,170	否	低	低	高	低	低
Lo 等人提出	否	中	低	高	低	低
Guner Arslan 等人提出	可	中	低	高	低	低
James G. Ryan 等人提出	可	中	中	高	低	低
Huang 等人提出	否	中	低	高	低	低
本發明	可	高	高	低	高	高

200636561

十、申請專利範圍：

1. 一種語音定位系統，其係包含：

一麥克風陣列，係用以接收聲音訊號；一語音偵測系統，係用以決定系統之運作狀態流程；一環境參數訓練子系統，係用以將聲波訊號藉由加成性原理模擬產生一訓練訊號，並將該訓練訊號經由特徵之抽取及描述此特徵之分佈狀況以產生一統計參數；及一語音位置偵測子系統，係用以藉由該麥克風陣列所擷取之聲音訊號及該統計參數偵測語者位置，並利用估測結果，經由統計參數更新模組來更新統計參數化模組。

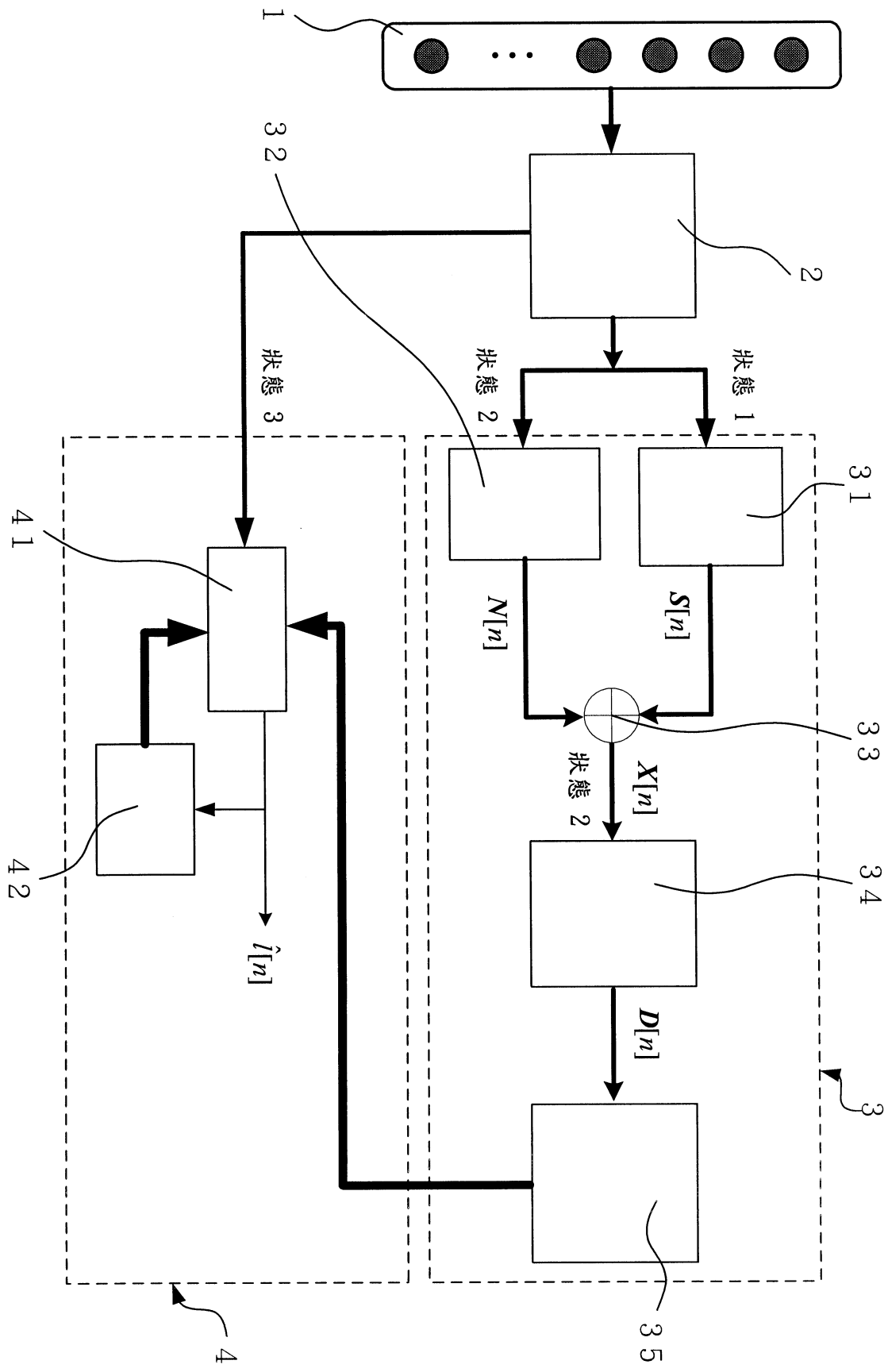
2. 依據申請專利範圍第 1 項所述之語音定位方法與系統，其中，該麥克風陣列係包含至少 2 顆以上之麥克風。

3. 依據申請專利範圍第 1 項所述之語音定位方法與系統，其中，該環境參數訓練子系統係包含一語者參考訊號之記憶體、一環境噪音訊號之記憶體、一合併器、一相位差特徵抽取模組及一特徵統計參數化模組。

4. 依據申請專利範圍第 3 項所述之語音定位方法與系統，其中，該特徵統計參數化模組係為混合高斯模型(GMM)及核心基礎模型(Kernel based model)中擇其一。

5. 依據申請專利範圍第 1 項所述之語音定位方法與系統，其中，該語音位置偵測子系統係包含一位置偵測模組及一統計參數更新模組。

200636561



第 1 圖