

## 混合聲音事件驗證在家庭自動化之應用

林昶宏 Chang-Hong Lin

國立中央大學資訊工程學系

Department of Computer Science and information Engineering  
National Central University

西雅恩 Ernestasia Siahaan

國立中央大學資訊工程學系

Department of Computer Science and information Engineering  
National Central University

陳伯煒 Bo-Wei Chen

國立成功大學電機工程學系

Department of Electrical Engineering  
National Cheng Kung University

莊祥瓏 Hsiang-Lung Chuang

國立中央大學資訊工程學系

Department of Computer Science and information Engineering  
National Central University

謝琇棋 Wen-Chi Hsieh

國立中央大學資訊工程學系

Department of Computer Science and information Engineering  
National Central University

王家慶 Jia-Ching Wang

國立中央大學資訊工程學系

Department of Computer Science and information Engineering  
National Central University

[jcw@csie.ncu.edu.tw](mailto:jcw@csie.ncu.edu.tw)

### 摘要

在本篇論文中，我們提出了一個在家庭自動化系統中，基於無線感測網路之混合聲音事件驗證的問題。在聲音分離階段，我們建構一個旋積盲訊號源分離系統，我們發展混合矩陣的估計方法，此混合矩陣可以用來重建分離的聲音來源。在驗證階段，我們使用從訊號小波包分解推導出的梅爾倒頻譜係數和費舍爾分數來當作支持向量機的特徵

參數。實驗結果顯示了此系統在基於無線感測網路家用環境下的混合聲音驗證中具有強健性及可行性。

關鍵詞：盲訊號源分離，家庭自動化，聲音驗證，支持向量機，無線感測網路

## 一、簡介

近年來由於無線技術的蓬勃發展，對於感測網路的應用產生許多正面的影響。無線感測網路已被應用於不同的領域，Talantzis *et al.* [1] 使用聲音及視覺訊號來追蹤在擁擠的室內環境中活躍的語者。Nishimura and Kuroda [2] 用圖形、聲音和加速訊號來實現在環境監測中的多功能辨認或人類行為識別。無線感測網路的實際應用已經得到學術界廣泛的關注，特別是有關家庭自動化科學的研究 [3-5]。在 2007 年，Baker *et al.* [3] 發展了一個居家健康照護系統。透過在房子周圍佈署聲音感測器來讓聽障者感知周遭環境。相關的研究可以參考 [4], [5]。

在本篇論文中，我們專注在無線感測網路的訊號擷取及處理。更一步來說，本研究特別強調用於家庭自動化中，聲學無線感測網路的聲音驗證。雖然在家庭自動化的設定中有許多聲音辨認的研究 [25-30]，現今的文獻仍缺乏在家庭自動化中，對混合聲音驗證的討論。

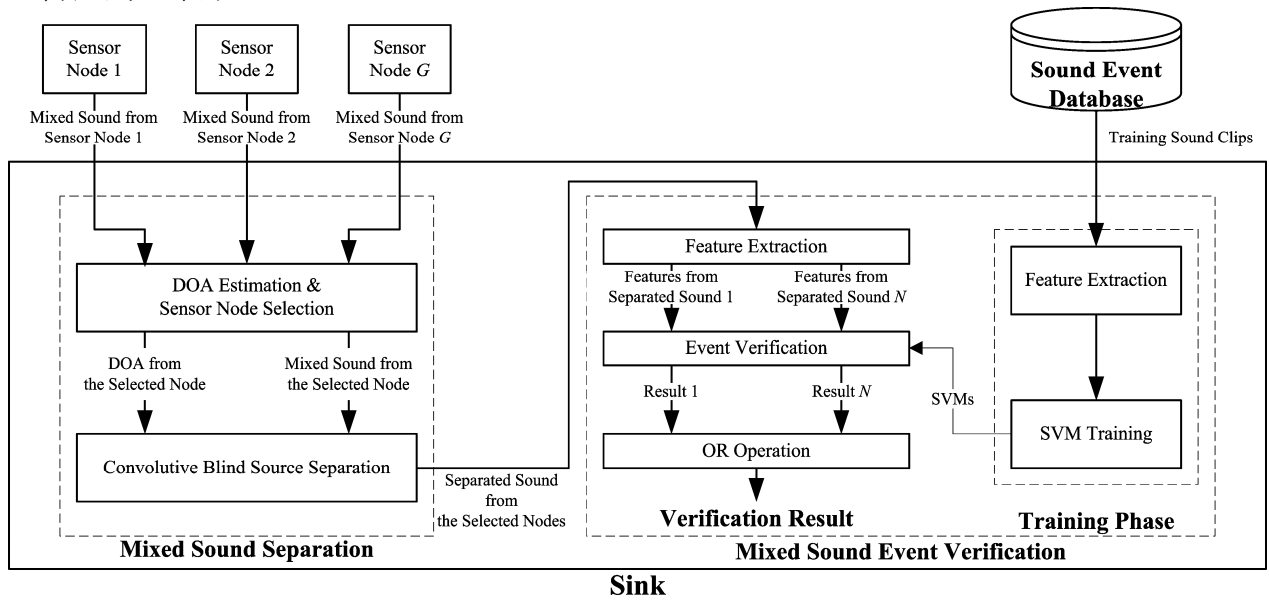
混合聲音的分離在近年來受到極大的關注。盲訊號源分離 (Blind Source Separation, BSS) 嘗試在大多數的來源資訊和混合過程是未知的狀況下，將來源從混合訊號中分離。這些限制使得盲訊號源分離成為一個具挑戰性的研究議題。Chen *et al.* [7-9] 提出了三個用來做盲訊號源擷取的無線感測網路結構，包含基於分群 (Cluster-Based)、無關分群 (Cluster-free) 和串接網路 (Concatenated Networks)。這些研究替應用於無線環境的訊號源分離演算法的發展提供了有價值的成果。

在此篇論文中，我們敘述了一個基於時頻分群的旋積盲訊號源分離 (Convolutional Blind Source Separation, CBSS) 方法 [10]。在我們的方法中，我們並不預先假設聲音來源的數量。我們實作了聲音來源數量的估計，並且用相位補償技術來估計混合矩陣，此混合矩陣將會用來重建分離的聲音來源。在感測網路系統中，聲音訊號經過擷取以及驗證後會觸發一個服務或是回應，因此被擷取訊號的驗證扮演了一個重要的角色。在我們的研究中，我們使用梅爾倒頻譜係數 (Mel Frequency Cepstral Coefficients, MFCCs) 和費舍爾分數 (Fisher Scores) [14] 當作聲音的特徵值，將這些特徵值送入預先定義聲音事件模組的支持向量機中 (Support Vector Machines, SVM)，用以決定某些事件是否發生了。

## 二、在無線感測網路上的聲音事件驗證

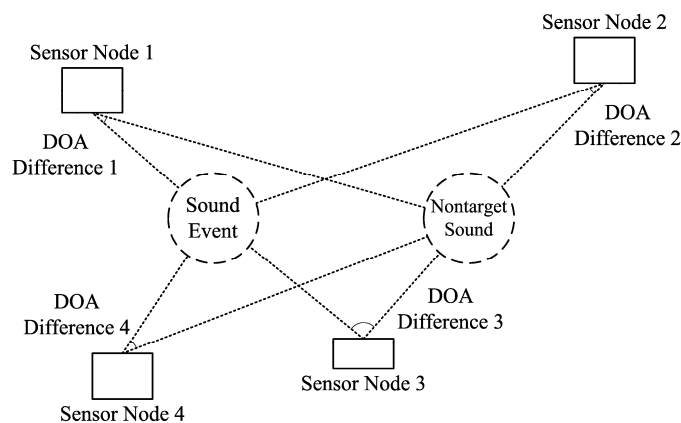
在此篇論文中，我們提出一個混合聲音事件驗證的無線感測網路架構，此網路架構包含許多無線感測節點以及一個匯集點。無線感測節點捕捉在房間內同時產生的聲音，每個感測節點附有一個小的麥克風陣列。每個感測節點中的麥克風陣列接收混合訊號並

且傳送到匯集點。



圖一、由匯集點 (Sink) 所執行的任務。

圖一顯示匯集點所執行的任務。第三章第四小節說明了估計從感測節點接收到的未知訊號到達方位 (Directions Of Arrival, DOA) 的方法。我們的觀察指出，當 DOA 差異增加時，CBSS 具有較好的分離表現。根據這個發現，系統被設計為選取擁有最大平均 DOA 差距的感測節點，同時，使用和此感測節點相關的分離訊號來執行聲音驗證。圖二為一個關於每個感測節點 DOA 差距的例子。在此例子中，感測節點 3 所接收的混合訊號將被用來執行聲音分離和驗證。我們的驗證階段的目的是驗證輸入訊號是否包含我們所感興趣的聲音。我們採用梅爾倒頻譜係數和費舍爾分數作為聲音特徵值。最後，我們使用支持向量機分類器來執行聲音驗證。



圖二、關於每個感測節點 DOA 差距的例子。

### 三、混合聲音分離

#### (一)、混合矩陣表示式

這篇論文考量到一個旋積混合型態的模型。我們利用下面的數學表示方式來描述此模型。

$$x_q(t) = \sum_{k=1}^N \sum_{l=0}^{L-1} h_{qk}(l) s_k(t-l) \quad (1)$$

其中  $x_q$  是感測器  $q$  對應的混合訊號， $s_k$  為源訊號  $k$ ， $h_{qk}$  則是語者  $k$  到麥克風  $q$  的脈衝響應，並且令這個濾波器 (Filter) 的型式為一個  $L$  階 ( $L$ -Tap) 的有限脈衝響應 (Finite Impulse Response, FIR) 濾波器。由於語音在時間域上的稀疏特性並不明顯，所以我們採用短時傅利葉轉換 (Short Time Fourier Transform, STFT)，以取樣頻率  $f_s$  將時間域上的混合訊號  $x_q(t)$  轉換成頻率域上的時間序列  $x_q(f, \tau)$ ，其中  $f$  是某個頻帶， $\tau$  為短時傅利葉轉換音窗的指標 (Frame Index)。在時頻域上執行盲訊號源分離的另一個好處是我們可以將旋積混合過程單純視為各個頻帶的瞬時混合型式，即如同以下之敘述。

$$X(f, \tau) = H(f)S(f, \tau) = \sum_{k=1}^N H_k(f)S_k(f, \tau) \quad (2)$$

, where  $X(f, \tau) \in C^{M \times 1}, S(f, \tau) \in C^{N \times 1}, H(f) \in C^{M \times N}$

其中  $X(f, \tau)$  和  $S(f, \tau)$  分別代表混合訊號以及來源訊號在時頻域上的成份。 $H(f)$  則是某一個頻帶的混合矩陣。然而，假設在一個時頻點上，只有一個來源訊號在活動，我們令  $H_k(f)$  是  $H(f)$  的第  $k$  個行向量，則可將式子(2)簡化為：

$$X(f, \tau) = H_k(f)S_k(f, \tau), k \in \{1, \dots, N\} \quad (3)$$

所謂的波束形成 (Beamforming)，即為一種空間上之濾波器，它利用訊號的空間關係，希望能夠對不同方向的訊號做出不同的增益，以達到空間濾波的效果，藉以分離空間中不同方向聲源的訊號。依波束形成定理，我們靠著麥克風陣列的來源訊號方向和時間延遲去近似混合過程。因此當頻率為  $f$  時，語者  $k$  到麥克風  $q$  的混合係數可表示為：

$$h_{qk}(f) = g_{qk} e^{j2\pi f c^{-1} d_q \cos \theta_k} \quad (4)$$

其中  $g_{qk}$  為訊號  $k$  至麥克風  $q$  的增益值， $d_q$  表感測器  $q$  與麥克風陣列中心之間的距離， $\theta_k$  是源訊號  $k$  對應到麥克風陣列的角度。我們可利用式子(4)，將混合矩陣表現成下面的形式，往後有關混合矩陣的推導過程，多數都是建立在這個預設形式之上。

$$H(f) = \begin{bmatrix} h_{11}(f) & \cdots & h_{1N}(f) \\ \vdots & \ddots & \vdots \\ h_{M1}(f) & \cdots & h_{MN}(f) \end{bmatrix} \quad (5)$$

## (二)、特徵值選取

我們定義了兩個混合訊號的特徵參數 (Level-Ratio 和 Phase-Difference) [11]。利用觀察資料的二階範數對混合訊號的絕對值頻譜 (Magnitude Spectrum) 做正規化，我們稱之為 Level-Ratio，這邊用  $\psi_q^L(f, \tau)$  表示；至於 Phase-Difference 被定義成與一個指定的混合訊號之間的相位角度差，以  $\psi_q^P(f, \tau)$  來表示。它們的表示式分別顯示如下：

$$\psi_q^L(f, \tau) = \frac{|x_q(f, \tau)|}{|X(f, \tau)|_2} \quad (6)$$

$$\psi_q^P(f, \tau) = \phi[x_q(f, \tau)] - \phi[x_1(f, \tau)] \quad (7)$$

其中  $\phi$  為相位的運算子。然後利用一個複數表示式 (Complex Representation) 來表現這兩個特徵參數。

$$\psi_q(f, \tau) = \psi_q^L(f, \tau) \times \exp[j\psi_q^P(f, \tau)] \quad (8)$$

於是我們得到了一個新的樣本型態 (Sample Form)，由  $M$  個 Level-Ratio 和 Phase-Difference 組成的複數值所構成。將原先的觀察資料轉換成這樣的資料型式後，我們即可使用這些新建立的樣本，做後續的處理和訊號分析，包括估計源訊號個數以及混合矩陣。令  $T$  為向量的轉置，則樣本型態表示如下：

$$\Psi(f, \tau) = [\psi_1(f, \tau) \cdots \psi_M(f, \tau)]^T \quad (9)$$

### (三)、混合矩陣估測

利用著名且應用廣泛的分群方法 K-Means 演算法，將樣本型態分割到  $N$  個群聚  $C_1, \dots, C_N$  中，並且利用下面的式子獲得混合向量：

$$h_i = \frac{1}{|C_i|} \sum_{\Psi \in C_i} \Psi, \quad i \in \{1, \dots, N\} \quad (10)$$

其中  $|C_i|$  代表第  $i$  個群聚擁有的樣本數。然而每個混合向量都會對應到一個來源訊號。因為我們是根據每個頻帶上的時間序列去估測混合矩陣，所以各個頻帶執行過 K-Means 演算法後都會回傳  $N$  個群聚，並求出代表的  $h_i$ 。最後，如何確認  $h_i$  在矩陣中的位置也是一個很重要的問題。

根據式子(4)，可以得知

$$\frac{h_i(r)}{h_i(s)} = \frac{g_{ri}}{g_{si}} e^{j2\pi f c^{-1}(d_r - d_s) \cos \theta_i} \quad (11)$$

所以經推導後，DOA 可以由下式獲得

$$\theta_i = \cos^{-1} \frac{\phi\left(\frac{h_i(r)}{h_i(s)}\right)}{2\pi fc^{-1}(d_r - d_s)} \quad (12)$$

其中  $r$ 、 $s$  是麥克風陣列中兩支距離最近的，它們在混合向量  $h_i$  中所對應到的指標； $d$  表示  $r$ 、 $s$  兩支麥克風之間的距離。因為我們對所有  $h_i(i=1, \dots, N)$  都偵測 DOA，所以共得到了  $N$  個角度值。最後根據這個結果確定  $h_i$  在混合矩陣中所對應的行索引。

假設混合訊號的某個時頻點  $X(f, \tau)$ ，只有源訊號  $k$  為非零的值。透過式子(3)和式子(4)，可將  $X(f, \tau)$  可表現為：

$$\begin{aligned} X(f, \tau) &= \begin{bmatrix} g_{1k} e^{j2\pi fc^{-1} d_1 \cos \theta_k} \\ \vdots \\ g_{Mk} e^{j2\pi fc^{-1} d_M \cos \theta_k} \end{bmatrix} \times g_{s_k}(f, \tau) e^{j\phi[s_k(f, \tau)]} \\ &= \begin{bmatrix} g_{1k} g_{s_k}(f, \tau) \times e^{j(2\pi fc^{-1} d_1 \cos \theta_k + \phi[s_k(f, \tau)])} \\ \vdots \\ g_{Mk} g_{s_k}(f, \tau) \times e^{j(2\pi fc^{-1} d_M \cos \theta_k + \phi[s_k(f, \tau)])} \end{bmatrix} \end{aligned} \quad (13)$$

因為本論文要對由 Level-Ratio 以及 Phase-Difference 所組成的樣本作群聚分割。所以，得出了混合訊號樣本在極度稀疏的情形下表現的型式後，我們將式子(13)代入式子(6)和式子(7)，看看若利用這種形態的樣本去定義 Level-Ratio 和 Phase-Difference 這兩種特徵參數， $\psi_q^L(f, \tau)$  和  $\psi_q^P(f, \tau)$  分別為：

$$\psi_q^L(f, \tau) = g_{qk} / \text{norm}([g_{1k} \cdots g_{Mk}]^T) \quad (14)$$

$$\begin{aligned} \psi_q^P(f, \tau) &= (2\pi fc^{-1} d_q \cos \theta_k + \phi[s_k(f, \tau)]) \\ &\quad - (2\pi fc^{-1} d_1 \cos \theta_k + \phi[s_k(f, \tau)]) \\ &= 2\pi fc^{-1} (d_q - d_1) \cos \theta_k \end{aligned} \quad (15)$$

然後，同樣的將上述兩個特徵參數用複數表示形態來敘述。最後，樣本  $\Psi(f, \tau)$  會以下的樣子呈現。

$$\Psi(f, \tau) = \begin{bmatrix} \psi_1^L(f, \tau) \times e^{j2\pi fc^{-1} (d_1 - d_1) \cos \theta_k} \\ \psi_2^L(f, \tau) \times e^{j2\pi fc^{-1} (d_2 - d_1) \cos \theta_k} \\ \vdots \\ \psi_M^L(f, \tau) \times e^{j2\pi fc^{-1} (d_M - d_1) \cos \theta_k} \end{bmatrix} \quad (16)$$

其中，第一項為一實數值。藉由上式，我們可以說，當語音具有極度稀疏的性質時，只會因為主導的源訊號不同造成  $\theta_k$  的改變而產生  $N$  種型式的  $\psi(f, \tau)$ 。所以當結束分群演算法估計混合矩陣之行向量的程序，並且解決了排列問題後，在最理想的情況下，也就是當極度稀疏的條件成立時，混合矩陣會長成：

$$\begin{bmatrix} \psi_{11}^L & \dots & \psi_{1N}^L \\ \psi_{21}^L e^{j2\pi f c^{-1}(d_2-d_1)\cos\theta_1} & \dots & \psi_{2N}^L e^{j2\pi f c^{-1}(d_2-d_1)\cos\theta_N} \\ \vdots & \ddots & \vdots \\ \psi_{M1}^L e^{j2\pi f c^{-1}(d_M-d_1)\cos\theta_1} & \dots & \psi_{MN}^L e^{j2\pi f c^{-1}(d_M-d_1)\cos\theta_N} \end{bmatrix} \quad (17)$$

盲訊號源分離在欠定的條件下，根據式子(2)，源訊號  $S$  可以有無限多個解，所以我們利用最小化  $\ell_1$  的範數以及  $X=HS$  作為限制式，此最佳化問題的解即為所求。如下列式子所示：

$$\min_S \sum_i |S_i|, \quad i = 1, \dots, N, \quad s.t. \quad HS = X \quad (18)$$

從這無限多組解中選取一個適當的答案。恢復源訊號的步驟就是依靠這個以最大後驗機率(Maximum a Posteriori, MAP)為基礎的方法 [12]。

#### 四、聲音驗證

##### (一)、支持向量機

支持向量機 (Support Vector Machines, SVM) [13] 是一個藉著建立最佳超平面 (Hyperplane) 使得兩類之間的margin最大的一種二元分類器。假設最佳分離超平面  $(w \cdot x) + b = 0$  最大化margin  $2/\|w\|^2$ ，其中  $w \in \mathbb{R}^d$  且  $b \in \mathbb{R}$ 。根據決策函數 (Decision Function) 資料點  $x$  被標記成  $y \in \{1, -1\}$

$$f(x) = \text{sign}((w \cdot x) + b) \quad (19)$$

我們可以在 SVM 中使用核方法 (Kernel Methods)。首先，將分離超平面函數表示成資料點  $x$  的內積，則決策函數可以寫成下面這個式子：

$$f(x) = \text{sign}\left(\sum_{i=1}^m \alpha_i x_i \cdot x + b\right) \quad (20)$$

其中  $\alpha$  為拉普拉斯乘數 (Lagrange Multiplier)， $i$  為向量的數量。此向量乘積可被核函數 (Kernel Function)  $k(x, x_i)$  所取代，

$$f(x) = \text{sign}\left(\sum_{i=1}^m \alpha_i k(x, x_i) + b\right) \quad (21)$$

藉由使用Mercer's理論，我們可以引入一個映射函數（Mapping Function） $\varphi(x)$ ，使得  $k(x_j, x_i) = \varphi(x_j)\varphi(x_i)$ 。藉由將原始輸入空間 $\mathbb{R}^d$ 投影到其他空間，此方法提供了處理非線性資料的能力。

## （二）、聲音特徵值擷取

### 1、梅爾倒頻譜係數

為了從聲音訊號中擷取梅爾倒頻譜係數 [14]，我們將聲音訊號切成短時窗（Short-Time Window）並對每個短時窗進行快速傅立葉轉換（Fast Fourier Transform）。然後將頻譜所包含的能量映射到使用三角濾波器頻帶的梅爾尺度上。在每個梅爾濾波頻帶上，我們計算對數能量，最後用離散餘弦轉換（Discrete Cosine Transform）來得到 MFCCs。

### 2、費舍爾分數空間

費舍爾分數使用一組參數生成模型，此模型將一串序列映射到固定維度空間的單一點上，例如：分數空間 [15]。藉由生成模型的似然值（Likelihood）分數來執行此映射。在這項研究中，我們推導出費舍爾分數用來將節點的機率分佈對映到聲音的小波包分解。根據經驗，每個節點被視作一種單一高斯分佈。

給定一個參數集為 $\theta$ 的生成模型  $p(X|\theta)$ ，我們可以藉由下列的式子得到相關的分數空間。

$$\Psi_{F,f}(X) = F(f(p(X|\theta))) \tag{22}$$

在上式中， $\Psi$ 表示分數向量； $f(p(X|\theta))$ 表示分數參數，此分數參數為本篇論文中生成模型的對數似然函數； $F$ 表示將分數參數對應到分數空間的分數運算。

## 五、實驗結果

我們實驗的第一個目標在證明我們的聲音分離階段的表現。接著，我們藉由比較混合和分離的聲音的驗證結果來展示聲音分離對聲音驗證的貢獻。在此研究中，我們利用三種我們感興趣的聲音訊號種類來作為目標聲音，例如：門鈴響、玻璃破裂和敲門聲。我們還定義了四個不希望得到的聲音，包含貓叫、狗吠、彈鋼琴及人說話來當作非目標聲音。在訓練階段，我們用從這些類別的乾淨聲音取出的特徵值來訓練 SVM 分類器。此外，20 個從目標類別選出的乾淨聲音檔案及 30 個從非目標類別選出的聲音片段也被用來訓練 SVM 分類器。我們的系統中使用的聲音長度是 1 秒，取樣率為 8 kHz。

我們用目標聲音和人說話聲的混合訊號來測試系統。在第一個實驗中，我們用不同的 DOA 差異值來評估聲源數量估計和分離的表現。我們採用訊號干擾比（Signal-To-Interference Ratio, SIR）[16] 來評價分離的表現。



表一、不同 DOA 差異值的平均 SIR

DOA Difference	40°	80°	160°
SIR of Separated Sound	17.5794 db	17.7916 db	18.7986 db

表一列出了三種不同的 DOA 差異值（40°, 80°, 和 160°）的平均 SIR。實驗結果顯示越大的 DOA 差異通常導致越高的 SIR，進而保證了較佳的分離訊號。這樣的測試支持了系統選擇具有最大 DOA 值的感測節點來進行訊號分離。

在驗證系統方面，我們比較了混合訊號和分離訊號的驗證表現。在分離訊號方面，系統會檢視在兩個訊號中，是否有從屬於目標類別的分離程序所產生的訊號。若這兩個訊號皆不為目標聲音，則將此聲音訊號歸屬在非目標類別。對每個音檔，我們從每個聲音訊號的音框取出 13 維 MFCCs，以及所有音框的 MFCCs 的平均值及標準差來作為聲音特徵。同時，我們從每個訊號的 3 階小波包分解樹( Three-Level Wavelet Packet Decomposition Tree )中取出 16 個費舍爾分數，得到總數為 42 維的特徵向量。我們使用 F-Score [17]量測方法來評測我們的系統。

表二、基準系統和提出的系統在驗證表現的比較

Sound Class	F-score		
	Mixed Sounds	Separated Sounds (Baseline)	Separated Sounds (Proposed System)
Doorbell Ringing	0.00	0.46	0.92
Glass Breaking	0.82	0.80	0.92
Door Knocking	0.23	0.00	0.32

表二為三個目標聲音類別的混合聲音及分離聲音的驗證結果比較。我們可以看出在驗證系統中，用分離訊號比用混合訊號的效果來的好很多。此外，我們提出的 CBSS 系統表現也比基準系統優異。

## 六、結論

在本篇論文中，我們描述了一個在家庭自動化中，基於無線感測網路之混合聲音事件分離和驗證系統。我們說明了旋積盲訊號源分離可以被用來分離混和聲音事件訊號。除了混合聲音分離，我們採用支援向量機來進行聲音驗證。所使用的特徵集包含 MFCCs 和 Fisher Scores。我們的實驗顯示，所提出的混合聲音分離架構顯著地增進了聲音驗證的結果。

## 參考文獻

- [1] F. Talantzis, A. Pnevmatikakis, and A. G. Constantinides, “Audio-visual active speaker racking in cluttered indoor environments,” *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 38, pp. 799-807, Jun. 2008.
- [2] J. Nishimura and T. Kuroda, “Versatile recognition using Haar-like feature and cascaded

- classifier,” *IEEE Sensors Journal*, vol. 10, pp. 942-951, May 2010.
- [3] C. R. Baker, K. Armijo, S. Belka, M. Benhabib, V. Bhargava, N. Burkhart, A. Der Minassians, G. Dervisoglu, L. Gutnik, M. B. Haick, C. Ho, M. Koplow, J. Mangold, S. Robinson, M. Rosa, M. Schwartz, C. Sims, H. Stoffregen, A. Waterbury, E. S. Leland, T. Pering, and P. K. Wright, “Wireless sensor networks for home health care,” in *Proc. 21st Int. Conf. Advanced Information Networking and Applications Workshops*, Niagara Falls, Canada, 2007, May 21–23, pp. 832–837.
- [4] A. Sleman and R. Moeller, “Integration of wireless sensor network services into other home and industrial networks using device profile for web services (DPWS),” in *Proc. 3rd Int. Conf. Information and Communication Technologies: From Theory to Applications*, Damascus, Syria, 2008, Apr. 07–[12p. 1–5.
- [5] H. Yan, H. Huo, Y. Xu, and M. Gidlund, “Wireless sensor network based E-health system—Implementation and experimental results,” *IEEE Trans. Consumer Electronics*, vol. 56, no. 4, pp. 2288–2295, Nov. 2010.
- [6] J.-C. Wang, H.-P. Lee, J.-F. Wang, and C.-B. Lin, “Robust environmental sound recognition for home automation,” *IEEE Trans. Automation Science and Engineering*, vol. 5, no. 1, pp. 25–31, Jan. 2008.
- [7] H. Chen, C. K. Tse, and J. Feng, “Source extraction in bandwidth constrained wireless sensor networks,” *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 55, no. 9, pp. 947–951, Sep. 2008.
- [8] H. Chen, C. K. Tse, and J. Feng, “Impact of topology on performance and energy efficiency in wireless sensor networks for source extraction,” *IEEE Trans. Parallel and Distributed Systems*, vol. 20, no. 6, pp. 886–897, Jun. 2009.
- [9] B. Bloemendal, J. v. d. Laar, and P. Sommen, “Blind source extraction for a combined fixed and wireless sensor network,” in *Proc. 20th European Signal Processing Conference*, Bucharest, Romania, 2012, Aug. 27–31, pp. 1264–1268.
- [10] A. Aissa-El-Bey, K. Abed-Mraim, and Y. Grenier, “Blind separation of underdetermined convolutive mixtures using their time-frequency Representation,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 1540-1550, Jul. 2007.
- [11] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” *Signal Processing*, vol. 87, pp. 1833-1847, Feb. 2007.
- [12] S. Winter, W. Kellermann, H. Sawada, and S. Makino, “MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and  $l_1$ -norm minimization,” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 24717, 12 pages.
- [13] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.
- [14] J. C. Wang, J. F. Wang, and Y. S. Weng, “Chip design of MFCC extraction for speech recognition,” *Integration, the VLSI journal*, vol. 32, pp. 111–131, 2002.
- [15] V. Wan and S. Renals, “Speaker verification using sequence discriminant support vector machines,” *IEEE Trans. Speech and Audio Processing*, vol. 13, pp.203-210, Mar. 2005.

- [16] E. Vincent, R. Gribonval and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, pp. 1462-1469, 2006.
- [17] T. Kemp, M. Schmidt, M. Westphal, and A. Waibel, “Strategies for automatic segmentation of audio data,” in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 2000, pp. 1423–1426.