

# 在無線感測網路上實現音訊(語音)融合與事件判斷技術發展之可行性構想研究

Ing-Jr Ding

Networks and Multimedia Institute, Institute for Information Industry, Taipei, Taiwan

ingjr@nmi.iii.org.tw

## 摘要

在無線感測網路的環境中，系統會根據感測器的感測，自動感知到偵測環境中的情境而發出警訊或提供服務。傳統的感知網路，多利用視覺感測器做環境的感測與監控，然而，其有關隱私權的考量，一直是資訊化社會持續受到關切的議題。基於此，本文將論述若不利用視覺感測器，而使用較無隱私權問題之音訊感測器以在特定的無線感測網路環境中實現音訊(語音)融合與事件判斷技術之可行性分析研究。

**關鍵詞：**無線感測網路、感測器、音訊資料融合、異常事件判斷、音源追蹤

## 1. 前言

隨著數位科技快速演進，人類生活各層面，包括食衣住行育樂等活動的內涵，將產生質變，構成所謂數位生活的型態。在這個新的生活型態中因數位技術而衍生出新的生活質素。而所謂的數位生活，既是傳統的遞嬗，也是創新的轉型。它蘊涵著一個巨大而快速的變遷，從技術變遷進而激發社會性的變遷、文化性的變遷、教育性的變遷、藝術性的變遷、法律性的變遷、娛樂性的變遷等等，其影響將甚為深遠。

假如數位生活及其可提供的服務及其可提供的服務即將來臨，那麼我們便有必要充分準備，以迎接它的到來，並且嘗試將其導引到正確的發展，以為人類建構一個未來的前景。事實上，這個前景雖可預見，但其路途仍然艱辛，仍有許多課題須待克服。然而，隨著微機電技術的進步與無線傳輸技術之發展，無線感測網路的應用範圍越來越廣，而這也是達成上述數位生活的美好前景的主要關鍵技術應用之一。無線感測網路發展至今，仍有許多可以應用的環境待開發，也仍有許多技術上的問題需要克服，值得各界學者及研究團隊繼續深入地研究。許多國內外的研究機構與廠商亦紛紛投入發展行列，因此，實值得在此領域做更深入的實現[1][2]。而欲將無線感測網路實質帶入人類生活中，感測資料的處理、整合、與後段的處理則扮演了一個重要的角色。而感測資料如環境中的聲音、光線、溫度、溼度等因子，又如人的影像、講話聲音

等因子，其中，聲音資料的呈現與取得又最為廣泛與容易。

無線感測網路的應用為乃數位生活最核心的環節之一，此刻要精準而完整地描述這類的藍圖並不容易。但是，一些可以預見的情節已呼之欲出。本處先簡述下列幾個這類可能情節為引子，說明一個新生活型態的重要機能和運作特性，繼而從中歸納出支援此類生活型態應該實現的核心技術課題，並據以規劃在無線感測網路的環境中實現音訊(語音)融合與事件判斷技術之可行性分析研究。

家庭安全防護方面：居家生活的安全是非常重要的。資訊和通信科技改善了傳統保全系統的限制和缺陷，提供較完善的 home security service。在一個完善的 home security service 居家環境中，一個好的安全防護系統若能偵測到環境中各式奇特怪異的聲音而據此判斷家中異常的事件發生，並立即通知家中、或在外面的使用者立即處理，相信可以減少家中意外事件或暴力事件頻頻發生的機率。例如，可以偵測家中小孩子的哭聲和吵鬧聲以判定是否有裸母虐待小孩的事件、偵測激烈吵架的聲音以判定是否有丈夫摑打老婆的家暴事件、偵測熱水燒開的聲音或浴缸放水的聲音以通知使用者立即關閉瓦斯爐及立即關閉水龍頭等等。而此能夠提供家中發生異常時立即通知的智慧化安全防護系統，其所具備的功能包括：(1)它由一個裝有 audio sensors 和 voice sensors 的網路所構成的 home security detection network。(2)它具有高度 pattern recognition 和分析推論的能力，能快速且精準地判定居家的 security condition，並據以歸納出異常的現象。(3)它最好能具有強力的通訊能力，能使得在家中的家人或甚至外出的家人都可以得到異常訊息通知的功能，以隨時掌握家中的狀況，並立即處理緊急狀況。(4)它具有 high fault tolerance 及 availability 特性，此 home security detection network 系統可在緊急或遭破壞時，仍可維持基本運作。

汽車防盜方面：目前，不僅在台灣，且在世界各地，汽車被盜的案件頻繁發生。而正如一諺語所說的「道高一尺，魔高一丈」，而為了保護車主的財產安全並且為了有效打擊汽車盜竊案之日益猖狂，很多研究汽車防盜的高科技產品也隨之增加。然而，若能在汽車被竊前的當下，立即通知車主，則必能有效減少汽車遭竊的事件。基於此，在優質的感測網路的數位生活中也必須提供一個能協助車主防範汽車被盜的機制，例如利用感測網路偵測

汽車周遭環境的聲音：人的腳步聲、動物(狗或貓的叫聲)、汽車玻璃擊破的聲音、汽車開門或關門的聲音、汽車引擎發動的聲音、及汽車警報的聲音等等，而從這些偵測到的聲音資訊加以判斷汽車被盜的風險，而能立即通知車主做最快的反應。此汽車被盜的機制需要具備下列的機能：(1)它由一個裝有 audio sensors 的網路所構成的 car security detection network。(2)它必須具有 automatic audio information classification 的能力。(3)它必須具有 audio data fusion and intelligent fusion 的能力，以準確判斷車竊的行為，而減少 false alarm 的發生機率。(4)它具有訊息通知的能力，能使得車主立即得知車子即將遭竊的訊息。

其上述的二種情境中，若以家庭安全防護之情境為例以做音訊(語音)融合與事件判斷技術進行可行性分析研究，我們可以發現：資訊的擷取、分析、過濾與傳遞在後端平台不斷的進行，透過長時間觀察與多項資料比對解析，現代人在不自覺的情況下常在公共場所被觀察、紀錄已是常態，但人們是否容許在最後的堡壘：「家」中也發生同樣的狀況，是否隱私權的考量會凌駕於安全性的考量？顯然答案仍待吾人探索。因此，若不考量隱私權的問題，則實可針對家庭環境再進行技術上實現之可行性的分析。

## 2. 系統發展建置

在發展在感測網路上利用感測聲音資訊以判斷特定事件發生的技術必須思考下列課題，並且提出完整而有效的解決方案，才真能呈現感測網路之生活願景的價值：

### 2.1 感測網路的架設是否考量使用者的心理感受？

若就感測網路的架設地點而言，家庭環境和辦公環境對使用者的心理而言，其所有的心理感受中，由安全感所造成的隱私問題，是必須加以考量的[3]。隱私問題使得感測器的鋪設必須優先考量對於使用者造成的威脅感(threat)有多大，例如視訊儘管是一個重量級的輸入媒介，但使用者是否願意在家中各處都裝上攝影機？如果不行的話，那麼是否有其他的媒介，可以達到相同的功效、卻不會造成這麼大的威脅感？或者使用者認為隱私權至上，寧願犧牲監控可帶來的服務品質來換取個人隱私？系統賦予使用者調控的彈性空間有多大？惟有在潛在威脅和預期利益之間取得平衡的設計，才能夠讓使用者接受。

因此，我們可以瞭解到，在其他環境中的可行技術，不一定可以直接套用到家庭環境及辦公室中。有時候，甚至部份技術因此而不適用以取得使用者的信任及增加他們的安全感，只擷取、儲存其

所需的最低限度資訊(如僅為擷取、儲存資料的特徵而非原始的資料)，以防惡意入侵者竊取、散布攸關個人隱私的資料。

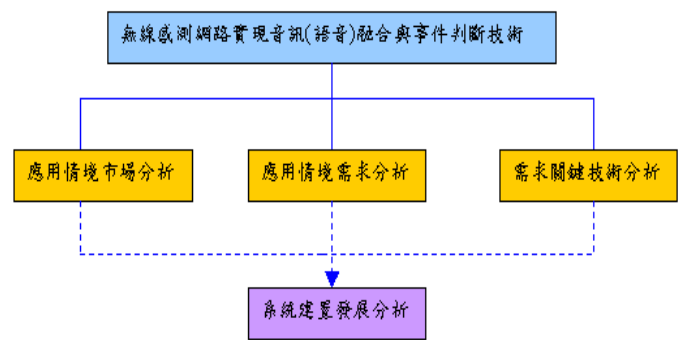
事實上，視訊裝置感測器儘管是感測網路架設的一個重量級的輸入媒介，但若就個人隱私問題而言，聲音裝置感測器卻是一個較佳的選擇。

### 2.2 聲音感測器間的協力合作能力能否得到提昇？

過去的感測應用型態僅在於單個感測器的運用，但是一個完整的特定環境的監控常常需要運用多個感測器才能完成。大量的感測器，依需求散布在待感測區域，感測器可蒐集各種環境的聲音資料。而在感測網路中，各式的聲音感測器位置該如何散布，而每個聲音感測器在安裝前該如何事先確認自己與其他大量聲音感測器之間的相對位置，即是聲音感測器間彼此之間該如何協力合作而自動組織出一個通訊網路，使得在感測區域中的所有感測資料，都能透過自我組態所建立的網路，將資料送出。

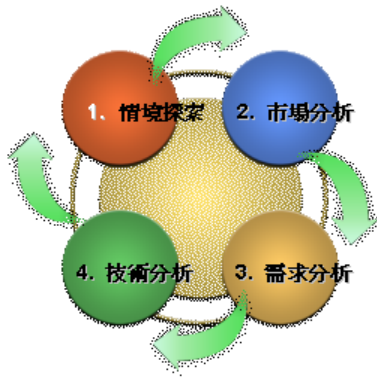
### 2.3 系統發展建置前分析

若要針對特定之情境實現無線感測網路上之音訊(語音)融合與事件判斷技術必須著重於應用情境之應用情境市場分析、應用情境需求分析、可能的關鍵問題與技術解決方案分析，並強調多元技術的整合，才能評估如在本文前言中所描述的情節是否得以藉此實現，如下圖所示：



圖一、分析架構圖

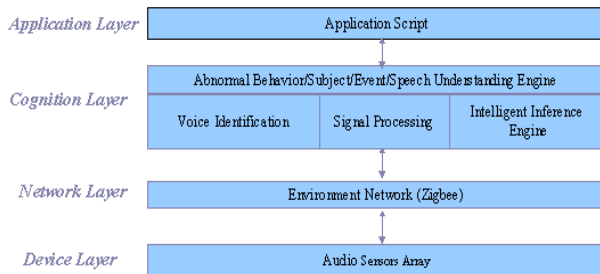
而整個應用情節之是否可行與市場的應用價值，將可以如下的階段圖，周而復始，據以評估：



圖二、技術指標階段圖

在階段圖中的第一步即是先選定在無線感測網路上實現音訊(語音)融合與事件判斷技術之可能的應用情境的探索，在選定情境後，即進行此情境之相關的市場分析，如市場規模分析、運用與買單意願業者分析、傳統解決方案分析、無線感測的價值分析等等，在進行完應用情境的市場分析之後，即進行第三步的使用者需求分析，在此步驟中，即是進行在選定的應用情境中之所有可能的使用者的需求分析，最後一步是實現此選定情境之所需用到的技術需求分析，如此即完成了一個應用情境分析的循環，並繼續下一個循環。

其中，在技術分析階段，可以利用如下之階層式的技術分層方式來分析是否可行：



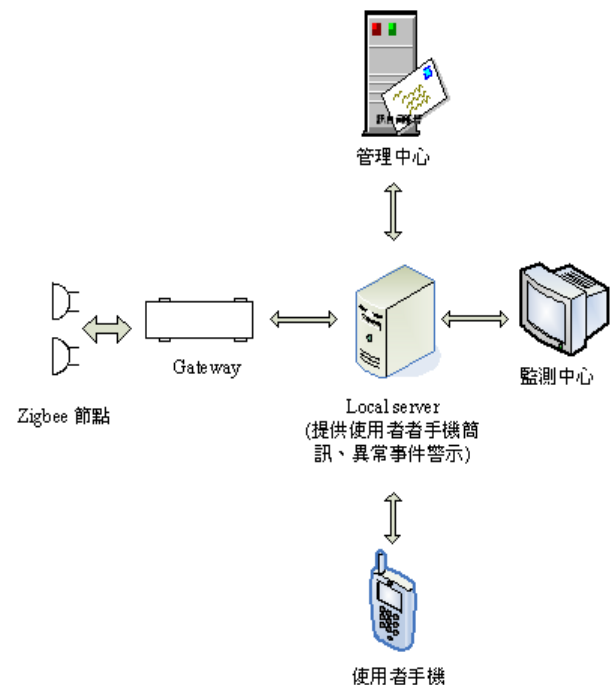
圖三、技術分析分層圖

在圖三中，關於各層次之技術分析的簡述說明如下：

在感知器層方面，此層為監測待測環境的網路實體層的技術分析，由各種能偵測各類環境聲音的感知器所組成。本計劃將以現行能偵測各類環境聲音的感知器之佈建方式進行相關技術的分析，如陣列訊號的技術分析、發聲音源偵測的技術分析等等，而並不是直接探索底層各種實際聲音的感知器，而是以整合現有的感知器產品，加強佈建及設置方式之技術分析，以能達到感測資料足以進行後端智慧型推論分析為主。

在網路層方面，此層負責聲音感測器連接及感知聲音資料內容的傳輸之技術分析。為提供全域

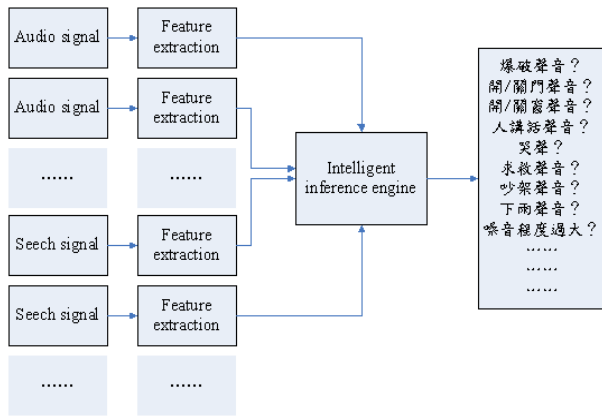
服務，本計劃之分析擬在假設待監測的特定環境內的進行 ZigBee 基礎平台的建置，而基礎建置主要是於此特定環境內預先設置所有 ZigBee 節點(與其電源)與 ZigBee 轉乙太網路閘道器等兩種裝置。ZigBee 節點即是各式的聲音感測器，如一般常見的麥克風、玻璃破碎感測器等等。ZigBee 轉乙太網路閘道器的功用在於接收來自前述 ZigBee 節點的聲音感測器所接收的聲音訊號並將訊號作融合分析後的結果後傳回去監控中心的伺服器上。如圖二所示，由所有 ZigBee 裝置形成的 ZigBee 無線網路，並會透過 ZigBee 轉乙太網路閘道器將訊號處理過後的結果往伺服器送。每一區 ZigBee 局部網路都必須設置一個 ZigBee 轉乙太網路閘道器，而此閘道器再與 Local Server 做連接。Local Server 為整個系統的伺服端，提供所有事件資料的儲存，其會再將此事件訊息後送到服務中心、使用者手機、或管理中心。



圖四、異常事件警示系統

在 Cognition Layer 方面，在此層的技術分析，則將利用各式的聲音感測器所接收到的聲音資訊之做融合判定的相關技術分析以判定是否有特定異常事件發生的任務。而此任務可以是在 ZigBee 轉乙太網路的閘道器中完成的，亦即在此閘道器中可完成的功能項目為：(1)Audio (speech) signal feature extraction procedure。(2)Intelligent inference engine of specific abnormal events。

如下之圖五所示：



圖五、異常事件或行為之判斷系統

而在閘道器中輸出的即是特定異常事件，並將此異常事件的訊息後送至後端的 Local Server。而在所有各層次的技術分析中，以在此層的技術分析為主要的分析，其包括從感知器端接收的感測資料(audio signal sensor array 設置、microphone array 設置)，該如何進行資料處理、如何進行相關特徵參數擷取、如何進行融合智慧型判斷推論等等之有關的技術[4][5]。

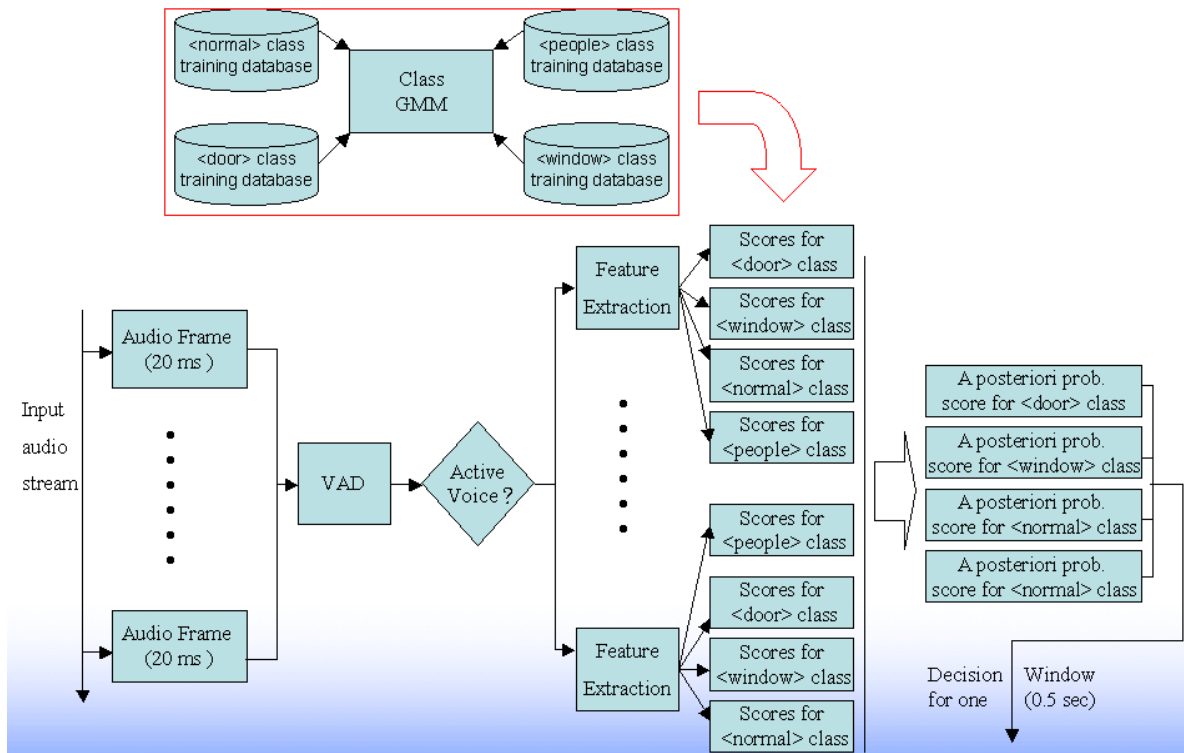
在應用層方面，此層為進行一個無所不在的 abnormal event instant message delivery engine 的技術分析。在 ZigBee 轉乙太網路的閘道器中若判定有異常事件發生，則會將此異常事件的訊息傳送至 Local Server，而 Local Server 一旦接收到來自閘道器的異常事件的訊息，則會再將此訊息透過 MMS、

IMPS、SIP 等相關技術，即時地轉送至監看的 client 端用戶、監測中心、或管理中心等位置。

## 2.4 系統建置

在經過上述之各項需求或技術之分析後，本文提出了一個可以偵測開/關門聲、開/關窗戶聲、及人聲的一個簡單的建置系統，如下之圖六所示，針對數種特定的聲音事件：開/關門聲音、開/關窗戶聲音、小孩的哭聲等等聲音事件的偵測進行系統建置，而系統置基本上可分為數個子議題，預計將各子議題的產出加以整合即可成為一個有效率的聲音事件偵測系統。

聲音訊號陣列的設置議題方面：基本上此聲音訊號陣列的設置之目的在於做聲音訊號的音源定位及聲音訊號的強度放大等兩個功用。前者可以做為小孩的哭聲之聲源的定位，除了判定此聲音的種類外，亦可約略判定發出哭聲之小孩的位置；後者將聲音訊號的強度放大可以有利後緒聲音訊號的處理、分類等等[6][7]。



圖六、無線感測網路環境中實現音訊(語音)融合與事件判斷技術系統建置架構圖

聲音活動的偵測議題方面：為了偵測真正“有意義之聲音”的活動，則必須使用此聲音活動的偵測動作以進行活動聲音的擷取，而此動作也是在平凡的周遭環境聲音中擷取突發的片斷之活動聲音。此外，此步驟亦進行環境中噪音的移除，利用此噪音的移除動作，可將擷取的片斷之活動聲音再進一步進行淨化處理。

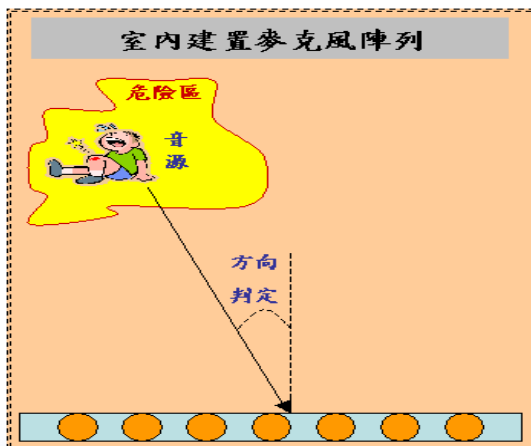
聲音分類模型的訓練議題方面：此步驟進行將聲訊訊號的特徵參數建立分類模型，而所用的分類模型是以 GMM(Gaussian Mixture Model)為主。利用資料庫相對應的聲音資料以訓練 GMM 模型直至其收斂為止，而此步驟即著重在反覆的模型訓練程序。另一方面，在 GMM 模型的設定方面，如高斯混合數的選擇亦必須仔細考量，此部份須配合聲音資料庫的資料量多寡、聲訊訊號的特徵參數特性、及實際的測試等等一併做考慮。

聲音特徵參數的選取議題方面：由感測裝置擷取到的聲音訊號經過適當地處理後，即可將此訊號進一步進行其特徵值的選取，而關於此聲音訊號之特徵值的選取，則必須依據聲音訊號的特性，選擇最合適的特徵參數。另一方面，關於特徵參數的維度設定方面，也必須謹慎考量，此部份必須配合聲音資料庫的資料量多寡及實際的測試以做最佳考量。

聲音資料庫的管理議題方面，必須針對欲偵測的數種特定的聲音事件：開/關門聲音、開/關窗戶聲音、小孩的哭聲等等聲音事件進行聲音資料庫的搜集或上述聲音的錄制動作。另一方面，針對所搜集的聲音資料庫亦必須進行整理、確認、分類等等人工處理的工作。

## 2.5 系統應用

透過以上之系統的建置，則針對家中的環境若設置此系統，則當小孩獨自在家中，若發生危險，則此音訊(語音)融合與事件判斷技術系統即可以判定此哭聲事件並同時偵測其約略的位置，而對外出的父母或監控中心即時發出警訊通知，以降低其受害的程度。



圖七、系統建置之應用

## 3. 結論

本文提出以無線感測網路實現音訊(語音)融合與事件判斷技術服務之可行性構想研究，其研究之成果可做為智慧型社區、企業服務供應商(如保全、居家照護等等)，或者感測器製造商等等相關業者欲發展相關監控軟體之重要的參考依據。本文以在做系統發展建置前，先實行應用情境市場分析、應用情境需求分析、可能的關鍵問題與技術解決方案分析等等的分析以確保在複雜多變的自然環境中能找出適合發展無線感測網路之情境，並以所提出之一個系統建置的雛形加以評估在各項分析中所歸納出之可行的應用情境在技術上再做實際是否可以實現發展之可行性的分析，而透過此一完整的運作分析模式，則亦可找出感測網路之未來的關鍵性應用及發展的趨勢。

## 參考文獻

- [1] A. K. Dey, "Providing Architectural Support for Building Context-Aware Applications," *College of Computing*, Georgia Institute of Technology, 2000.
- [2] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing. Computer," IEEE Computer Society Press, Vol. 34, No. 8, pp. 57-66, 2001.
- [3] Meyer and Rakotonirainy, "A survey of research on context-aware homes," Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003, Adelaide, Australia, pp. 159-168, February 2003.
- [4] D. Sturim, H. Silverman, and M. Brandstein, "Tracking Multiple Talkers Using Microphone-Array Measurements," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, 1997.
- [5] J. K. Aggarwal and Sangho Park, "Human Motion: Modeling and Recognition of Actions and Interactions," 3D Data Processing, Visualization, and Transmission, 2nd International Symposium on (3DPVT'04), pp. 640-647, September 2004.
- [6] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," Proc. ICASSP, Vol. II, pp. 273-276, 1994.
- [7] J. Chen, J. Benesty, and Y. A. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," IEEE Trans. Speech Audio Proc., Vol. 11, No. 6, pp. 549-557, November 2003.