



## 計算機組

### 林敬舜 助理教授

美國南加州大學電機博士

研究領域：沉浸式系統、多媒體應用、空間音效合成、多重聲道音訊處理、虛擬技術、影像與圖形間轉換、軟式計算

關鍵字：嵌入式多媒體、沉浸式系統、虛擬麥克風/揚聲器、視覺聽覺融合、軟式計算、生物醫學圖像辨識、機器學習

網頁：<http://homepage.ntust.edu.tw/chingshl/>

電子郵件：[CHINGSHL@mail.ntust.edu.tw](mailto:CHINGSHL@mail.ntust.edu.tw)

電話：886-2-27376372(Voice), 886-2-27376424(Fax)

#### 一、研究主題與目標

開發沉浸式系統的目的為讓身處不同的區域的人們有著對周遭環境相同的感受。為了達到這個目標，必須就人們在這個環境下所能看得到的及聽得到的資訊做擷取及再生。本實驗室的主要研究方向為嵌入式多媒體系統的開發，尤其將著重在沉浸式聲響系統技術的發展及應用：

- 嵌入式多媒體系統開發
- 機器學習應用於視覺-聽覺融合
- 虛擬技術在網路上的應用
- 基於圖形建構之影像重建
- 擴增實境與沉浸式系統建置
- 多重聲道合成與可視化
- 虛擬聲效在高解析音訊傳遞的應用
- 空間聲響學/等化/校準
- 適應性訊號處理在免持式防噪耳機與麥克風之應用
- 生物醫學圖像辨識
- 軟式計算
- 以及這些技術在消費性電子的應用

沉浸式聲響系統的應用包含有視訊會議、航空交通控管、駕駛員警告與導引、遠距教學、專業音效編輯、以及對視障者或聽障者的協助。

#### 二、最近研究題目

##### (1)基於總體經驗模態分解法之近場音源成像系統：

傳統的麥克風陣列傅立葉近場聲全像術，透過快速傅立葉轉換(Fast Fourier Transform)能夠高效率的重建聲場，並且準確的獲得噪音源的分佈情形，但實際應用上因傅立葉轉換的特性會造成量測的誤差，且需根據聲源訊號的頻譜資訊，才能選擇主要的頻率來觀看聲場全像圖。有鑑於此，在滿足經驗模態分解法條件下，我們利用總體經驗模態分解法(Ensemble Empirical Mode Decomposition, EEMD)的自適應性以及低模態混雜(Mode Mixing)等特性在空間域上對多個音源訊號進行分解，且利用獲得的本質模態函數(Intrinsic

Mode Function, IMF)來求得音源相對應的瞬時頻率，在不需事先選擇主要頻譜資訊即可同步觀察各音源在聲場空間中的分佈情形。除此之外，EEMD 音源成像技術亦能結合近場等效音源影像系統(Near-field Equivalent Source Imaging, NESI)中的虛擬麥克風技術來提高聲場影像的解析度。下圖為利用二維 EMD 演算法鑑別 1kHz 與 450Hz 訊號的結果。

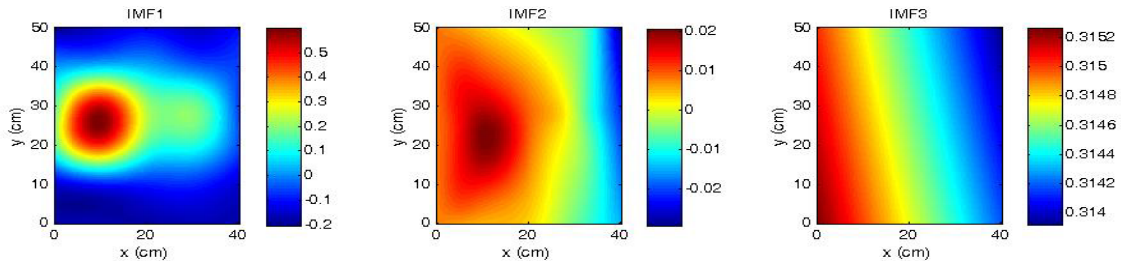


Fig. 1. EEMD成像術對1.2kHz與0.5kHz訊號於重建平面上之聲場重建結果

(2)以PAT Tree索引DTW參考區間之音樂檢索:

音樂檢索近年來受到許多的注意，其中使用哼唱的查詢方式(Query by Humming, QBH)令使用者感到直覺方便。然而哼唱歌曲片段與樣本音樂在特性上的差異以及使用者是否哼唱得精確增加了比對的困難度。除了檢索正確率之外，檢索的時間花費也是重要的考量。在本文中，我們使用兩階段的搜尋方式。第一階段使用簡單的歌曲相對音高差序列當成索引的關鍵特徵符號串，因 PAT Tree(變化自 Patricia Tree, Practical Algorithm To Retrieve Information Coded In Alphanumeric)在處理符號子字串匹配問題上有優異的表現，因此用其來快速找出歌曲中候選的特徵序列區間。此外，在音高差異序列的索引上常發生的多音、少音與走音的錯誤仍可用 PAT Tree 找到部份符合的子符號串，並可得到較多的候選參考位置。第二階段再對這些候選的特徵序列區間進行較精確的 Dynamic Time Warping (DTW)比對。在不失容錯力下，我們的方式避免了完全使用 DTW 比對所帶來的耗時計算。

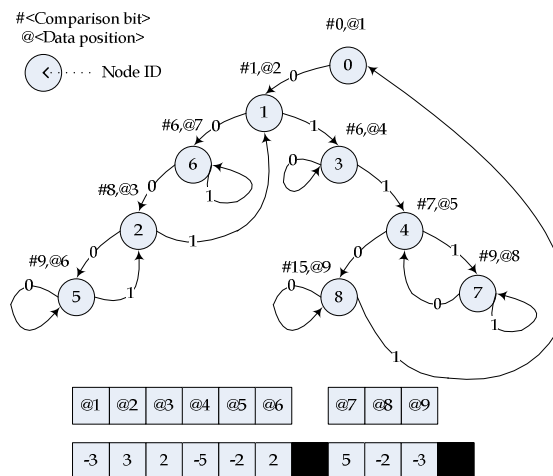


Fig. 2. PAT Tree 結構及其搜尋

(3)以低階本質模態函數為基礎之即時扭曲區間計算：

使用非線性刻度的技巧對訊號進行分析、量測與建模被廣泛地用在訊號處理、地震預測、結構測試與故障診斷中。基本上有兩種方法常被用來近似非線性解析度，其一是使用加權的方式在高頻段容許更多的誤差，並將重點擺在能量集中的低頻段上。另一種方式則是對訊號在時域、頻域或空間域以扭曲轉換(Warping Transformation)做預處理，並在整個系統的後端加上反扭曲函數，如此一來可避免過多的變數導入原本主要的計算，並可有效地處理能量多集中在相對低頻的多媒體訊號。我們嘗試以此更直覺與充分定義的機制來達成對訊息粒度(Information Granules)的語意分析，為了要適應顆粒描述(Granular Descriptors)的複雜性，我們結合低階本質模態函數(Low-level Intrinsic Mode Function, LIMF)能迅速將資料中的高頻變化取出之特性來設計整個系統，以提供一種具自適應性的瞬時歸屬函數計算方式。此非均勻間隔區段計算法將有助於時變訊號的分析，而此模糊集合的適切性將可透過即時偏移與改變歸屬函數的形狀來調整，如此一來可改善原本這些集合本質上的不連續與離線性所帶來的資料錯誤建模。

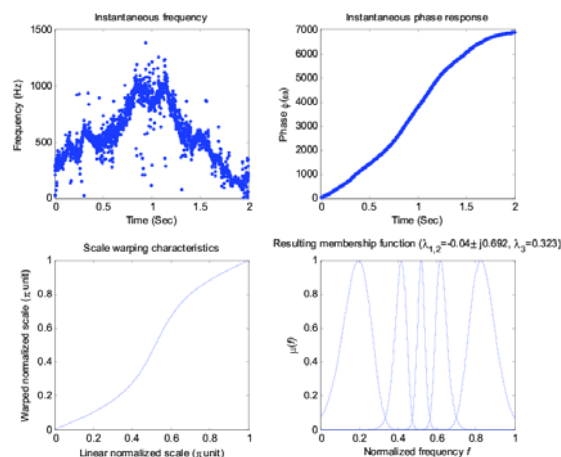


Fig. 3. 非對稱Chirp資料經3階全通濾波器轉換後所產生之歸屬函式

(4)利用 Shaped Noise 實現語音相關訊號之快速總體經驗模態分析：

對於訊號分析，經驗模態分解法是一個很好用的方法，它具有自適應性(Adaptive)，並能夠處理非線性(Nonlinear)與非穩定(Nonstationary)的訊號，且能在時域或空間域上直接進行訊號拆解，但其主要的缺點之一在於會有模態混雜(Mode Mixing)的問題。為了解決此問題，Wu與Huang利用白雜訊(White Noise)提出了總體經驗模態分解法(Ensemble Empirical Mode Decomposition, EEMD)，但此方法最後的結果是透過多次迭代而產生的，因此運算時間相當的冗長。有鑑於此，針對大部分能量分佈在低頻的語音訊號，我們使用Shaped Noise取代白雜訊來此降低EEMD迭代的次數，以增加EEMD的收斂速度。Shaped Noise或稱為訊號頻譜相依雜訊(Signal-Spectrum Dependent Noise, SSDN)，它可以隨著輸入訊號的頻譜特性而隨機改變，且具有自適應性的特性，相較於對訊號加入白雜訊，它可以節省對細微能量

的頻譜的運算，藉此降低運算時間與資源。在此研究中，我們也發現，針對語音訊號，我們亦可以利用與其特性較相近的雜訊如粉紅雜訊(Pink Noise)與棕雜訊(Brown Noise)等來取代白雜訊，在不影響收斂結果的情況下同樣能達到降低EEMD迭代次數的效果。

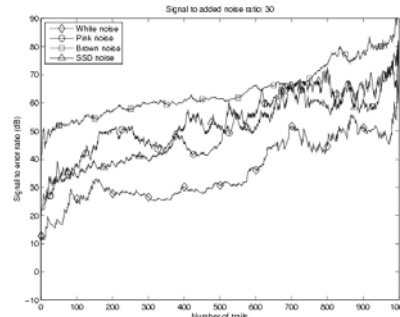


Fig. 4. 使用Shaped Noise實現語音相關訊號之快速總體經驗模態分解之收斂速度比較

#### (5)基於參考簡化模式之聲音識別:

聲音識別首重特徵選取，篩選的標準一般取決於分類訊號是否正確地被表示，以及之後的識別是否較為容易。我們採取可減低對資料庫依賴的訊號表示法，同時也簡化了在識別過程中比對需要的編碼方法，在此提出的方法主要分為三部分。在第一個步驟中，我們採用了相關頻譜轉換-感知線性預測法(RASTA-PLP)，也就是利用基於心理聲學的線性預測法所獲得的資訊來表示數位訊號的頻譜封包。其次，為了保留音色變化時的自然逼真性，輸入訊號與已知資料的特徵配對與對齊就變得不可或缺。傳統的聲音辨識方式使用基於頻譜振幅與調和音域攫取出的特徵向量來達到這個目的，但對於需要在時間與頻率同時精確定位的頻譜圖來說，藉由比較音框間的內積大小在資料庫裏找尋與輸入訊號最相似的頻譜圖將是更好的方法。在最後的步驟中，我們提出間歇性的碎形空間填充曲線(或稱虛擬Hilbert曲線)來對頻譜圖上的資訊進行區域編碼，因該輸出序列較短，使得利用高斯混合模型與存放在資料庫中的序列做比較時更有效率。與完全參考模型不同的是，參考簡化模式只使用了少量的比對資料，這使得此法更適合用在手持式的設備上。

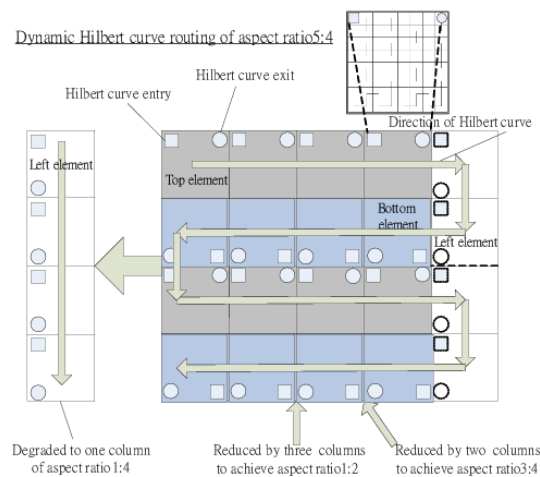


Fig. 5. 虛擬Hilbert曲線的繞行方式

(6) 多重聲道合成:

多重聲道音訊將成為音樂再製的下一個重點。增加聲道的數量將提供聆聽者更愉悅與真實的感受。然而，相較於現有的音樂資料，只有一小部分的音樂被進一步的多聲道化，許許多多的舊時代所錄製的唱片也就被擱置在單聲道或雙聲道的狀態。解決這個問題的其中一個方法就是將這些現有的錄音加上其他來自不同方位的訊息，基於這個想法我們發展出利用成對參考訊號的頻譜圖形變來合成其他的聲道。這個方法類似於圖學的形變，使得原始訊號不僅僅可轉成多重聲道訊號，其中亦保留了原來錄音場地的空間聲響特性。

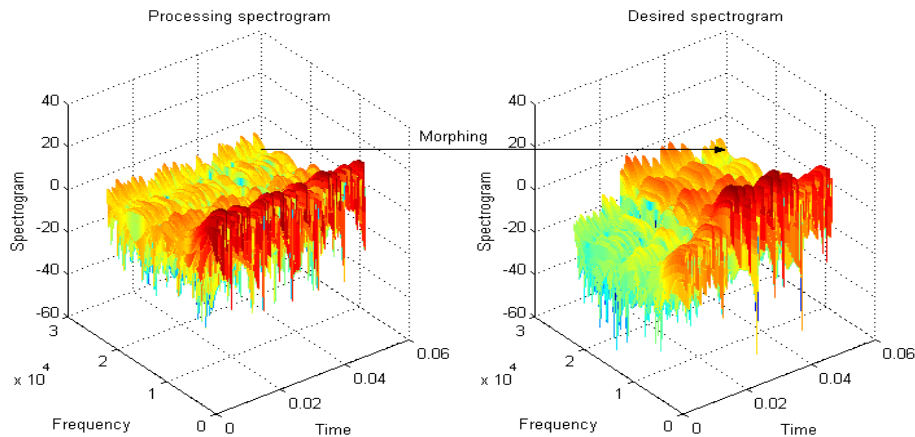


Fig. 6. 基於頻譜圖形變之多聲道音效合成

(7) 多頻段雜訊抑制:

背景音雜訊消除長久以來一直是被廣泛討論的議題，相關的文獻多半是針對雜訊本身做探討，往往忽略了播放系統與環境對整體播放效果的影響，加上有些雜訊濾除必須以雜訊特性作參考，這也使得將這些理論應用於訊號易失真的寬廣播放空間與一連串電聲系統時顯得不切實際。針對這個問題，我們提出了可對整個空間雜訊做偵測與評量的模組，透過這個設計，使得背景音雜訊的特性可以是未知，以往以控制學為主的加強式學習法也可順利地應用在數位訊號處理上。

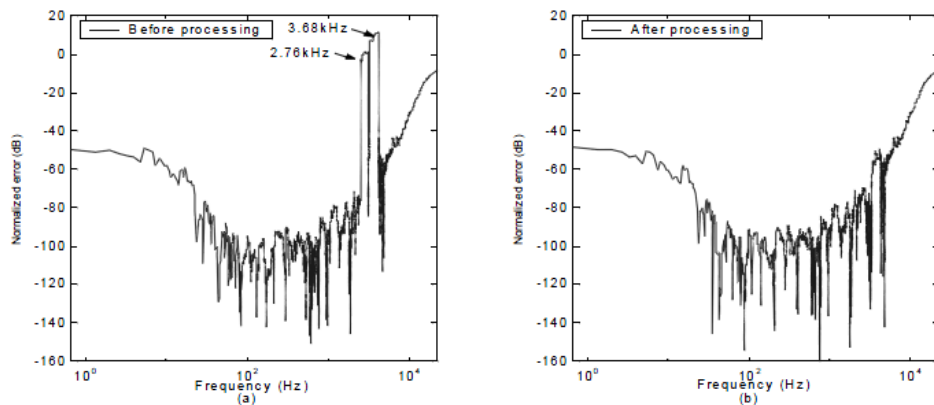


Fig. 7. 基於 1/3 octave smoothing 之比較 (a) 雜訊濾除前之正規化誤差 (b) 雜訊濾除後之正規化誤差



### 三、主要的研究成果與所執行的計劃

- [1] **C. S. Lin** and Y. C. Chao, "Empirical mode decomposition based near-field equivalence source imaging system for sound identification," *8th International Conference on Information, Communications and Signal Processing*, Singapore, Dec. 2011. (EI)
  - [2] **C. S. Lin**, S. J. Hong, D. L. Shih, and Z. C. Cheng, "Instantaneously warping interval computation using low level intrinsic mode functions," *IEEE International Conference on Granular Computing*, Kaohsiung, Taiwan, Nov. 2011.
  - [3] **C. S. Lin**, J. S. Wang, and Z. C. Cheng, "Fast ensemble empirical mode decomposition for speech-like signal analysis using shaped noise addition," *4th International Conference on Interaction Sciences: IT, Human and Digital Content*, pp. 112-117, Busan, Korea, Aug. 2011. (EI)
  - [4] **C. S. Lin** and D. R. Wang, "Spectrogram image encoding based on dynamic Hilbert curve routing," *International Conference on Image Processing Theory, Tools, and Applications*, Paris, France, Jul. 2010.
  - [5] **C. S. Lin** and C. Kyriakakis, "Synthesizing multichannel recordings based on adaptive spectrogram morphing," *4th International Conference on Innovative Computing, Information and Control*, pp. 156-159, Kaohsiung, Taiwan, Dec. 2009.
  - [6] **C. S. Lin** and C. Kyriakakis, "Removing additive noise via neuro-fuzzy-based reinforcement learning," *Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1026-1037, Aug. 2008. (SCI, EI)
- [1] 國科會 — 基於高斯粒化合成之音訊強化與修復
  - [2] 國科會 — 音視訊號融合之情緒偵測系統
  - [3] 國科會 — 基於虛擬聲響學之視聽輔助系統開發
  - [4] 三軍總醫院與台灣科技大學 — 數位化腸音聽診器的發展與臨床應用