

行政院國家科學委員會專題研究計畫成果報告

利用共極點 ARMA 模式作 3D 音效之研究

Study on common-pole ARMA modeling of 3D sound

計畫編號：NSC 88-2218-E-009-031

執行期限：87 年 8 月 1 日至 88 年 7 月 31 日

主持人：謝世福 博士 交通大學電信系 副教授

一、中文摘要

近年來電腦多媒體已成為一般大眾視聽娛樂必然的趨勢。本計畫書是針對虛擬實境中，對於如何有效利用 HRTF 來產生 3D 立體聲之研究。在產生 3D 立體聲的過程中，頭部轉移函數(Head Related Transfer Function)決定了虛擬音源的位置，藉由不同的頭部轉移函數，將可提供放送系統不同位置的虛擬音源，以達虛擬實境的目的。

在量測到了 3D 中數個位置的 HRTFs 後，我們可對這些 HRTF 做簡化的工作，而不失其原本具有的音源位置辨識功能。為了降低乘法器的數目，由於耳朵內的共振現象可以由適當的極點模式加以補捉，所以有效的 ARMA 模式可以逼近 HRTF。我們也會考慮更高階的即點模式，來擴增近來研究指出的低階頭部/耳翼模式來更逼近 HRTF。

我們相信在空間裡頭不同方位的 HRTFs 應該仍然具有共通的極點。典型的系統識別方法如 Prony、Shanks...等等，可以輕易的提供單一脈衝響應的 ARMA 近似模式，但是對於多個脈衝響應，如何提供一個共同極點，多個不同零點的 ARMA 來分別近似，就是本計畫的目標。

此外，建立群組 ARMA 模式也是本計畫的研究議題，最終吾人會利用電腦模擬分析並測聽所提模式的適用性。

關鍵詞：3D 音效、頭部轉移函數、ARMA 模式

Abstract

As the virtual reality plays an important role in multimedia applications, a true 3D sound using the HRTFs has an edge over the conventional audio in consumer electronics. This project will focus an efficient modeling of HRTFs, which represent the head-related transfer functions that affect human spatial hearing.

Since the amount of these measured HRTFs is long in time and large in space, an effective modeling of them is important and necessary in practice. Due to the resonance of sound when arriving at human ears, we aim to use a common-pole ARMA model for these HRTFs, which is a generalization of a recent research assuming the simple one-pole head model and a pinna model for the HRTF.

We believe that different HRTFs from different directions should still preserve some common poles, at least for the neighboring HRTFs. The well-known system identification methods for ARMA modeling, such as Prony and Shanks,...can easily solve modeling of one single impulse response. However, to approximate a set of HRTFs using the ARMA models with common poles and different zeros is the goal of this project.

Besides, a clustered ARMA model will also be considered. Finally, computer simulations and listening tests will be performed to justify the usefulness of our

proposed models.

Keywords: 3D sound, HRTF, ARMA model

二、緣由與目的

在人類的聽覺系統中，藉由聲源到達聽者兩耳的聲音之間的差異，人可以判別出聲源實際所在位置[1, 2, 8]。但是只由兩耳之間音量大小的差異和到達兩耳的時間差這兩參數所建構出來的播放系統，其虛擬出來的音源並不能達到 3D 的要求，因此我們引入 HRTFs 的觀念。所謂 HRTFs，是藉由一個真人或假人頭兩耳所測量到的響應。

在量測到了 3D 中數個位置的 HRTFs 後，我們可對這些 HRTF 做簡化的工作，而不失其原本具有的音源位置辨識功能[3, 4, 5]，做此步驟的目的在於減少運算量以及 HRTF 所占有的記憶體空間。

近來有人將 HRTF 簡化成頭部與耳翼模式，前者為一階的 IIR 低通濾波器，後者為五個係數的 FIR 濾波器，並經實驗確認其適用性[7]。吾人覺得此種方式太過直覺，又嫌過度簡化，我們覺得應該可以考慮更高階的極點模式，來逼近 HRTF。

此外，我們相信在空間裡頭不同方位的 HRTFs 應該仍然具有共通的極點[9]。典型的系統辨識方法如 Prony、Shanks...[10] 等等，可以輕易的提供單一脈衝響應的 ARMA 近似模式，但是對於多個脈衝響應，如何提供一個共同極點，多個不同零點的 ARMA 來分別近似，就是本計劃的目標。

三、研究方法與成果

HRTFs data set 為實現 3D 虛擬音訊的基本資料集合，裡面包含不同方位至左右耳的脈衝響應。對於特定方位的 HRTF，其本身即為一長度高達 226 的脈衝響應(如圖 1)，而縮短聲訊脈衝響應的傳統方法，通常是利用 ARMA 模式(pole/zero 模式)來對其作分析，希望能藉由較少的 pole/zero 參數，取代原本冗長的脈衝響應。

在以 pole/zero 模式對 HRTF 作分析前，我們應該先定義一誤差項，將其當作 pole/zero 模式下合成的與原有的 HRTF 差距的指標。一個最直接的定義如下：

$$e(n) = h(n) - \hat{h}(n)$$

其中 $e(n)$ 為直接誤差、 $h(n)$ 為真實的 HRTF、 $\hat{h}(n)$ 為合成的 HRTF。所以決定 pole/zero 參數的基準便是讓 $e(n)$ 最小。

然而若直接針對 $e(n)$ 做微分以求出 pole/zero 參數，這將導致解非線性方程式的難題。因此，我們將利用可以避免掉解非線性方程式的以下三種方法來求 pole/zero 參數。

● Prony's method

既然直接對前面定義的直接誤差以微分方式最小化求出 pole/zero 參數會陷入解非線性方程式的窘境，Prony's method 轉而對一 filtered error 做最小化，這樣的改變，將可避免掉解非線性方程式。其中 filtered error 的示意圖如圖(2)。對 filtered error 做最小化，可得以下 pole/zero 參數

$$\Rightarrow \begin{cases} \bar{a}_{pr} = (H^T H)^{-1} H^T \bar{g} \\ \bar{b}_{pr} = \bar{g}_0 + H_0 (H^T H)^{-1} H^T \bar{g} \end{cases}$$

的解：

其中

$$H, H_0, \bar{g}, \bar{g}_0$$

由 $h(n)$ 所組成。

● Shanks' method

其 pole 係數的決定與 Prony 相同，但對於 zero 係數的決定，Shanks' method 的準則又回歸於直接誤差的最小化，如圖(3)。由於直接誤差是我們最標準的效能指標，所以 Shanks' method 將會比 Prony's method 有更好的效果。

● Iterative prefiltering

不同於前面兩種方法，Iterative prefiltering 並非在最小化一妥協的誤差項；相反地，此方法避免非線

$$[F^{(i)}(z)] = \left[\frac{A^{(i)}(z)}{A^{(i-1)}(z)} F(z) - \frac{B^{(i)}(z)}{A^{(i)}(z)} \right]$$

性方程式產生的著眼點在於最小化的程序。此處最小化的誤差項是之前最標準的直接誤差，取代原本微分的最小化程序，此處以迭代法來尋求誤差極小值。將直接誤差的定義以 z 轉換的形式改寫如下：

其中 i 代表第 i 次的迭代。讓誤差項逼近零，可得下列線性迭代方程組，其中 pole/zero 參數可由以上方程組利用最小方根法不斷迭代求出。

圖(4)為以上三種方法合成 HRTFs 的模擬結果。觀察之後可以發現，iterative prefiltering 的效果明顯較好。圖(5)則可發現 ARMA 模式下的三種方法都要比 FIR 的方式來得好。

● 共極/零點模式

由於耳翼具有類似共振腔的結構，HRTF data sets 中不同方位的 HRTFs 可能具有共同的極點，因此對於這些 HRTFs，我們假設它們共享一組極點，但卻擁有個別的零點。

延續之前三種決定極/零點的方式，套用於共極/零點模式，其示意圖分別如圖(6)、圖(7)及圖(8)。而模擬結果分別如圖(9)、圖(10)及圖(11)，觀察後發現，在共極/零點模式下，以 iterative prefiltering 方法合成的效果最好，且與極/零點模式下的合成曲線十分接近，由此我們可說共極/零點模式適用於 HRTF data sets。

● 分群

儘管 HRTF data sets 適用共極/零點模式，但是若是將過多或是不適合的 HRTFs 套用於共極/零點模式，那麼就很有可能造成合成效果變差。於是為了降低誤差，我們必須在決定共極/零點前，採取分群的策略。圖(12)為分群的示意圖(此處我們假設分為三群)。這裡我們應用 LBG 演算法進行了兩種分群法，一針對 HRTF 的距離，另一則是針對極點係數。如圖(13)所示。

四、結論與討論

在本篇報告中，我們利用共極/零點模式對 HRTF data sets 模式化。同時，我們也將原本運用在極/零點模式下求極/零點係數的三種方法延伸至共極/零點模式。經過電腦模擬發現，利用 iterative prefiltering 的方法，不論是在極/零點或共極/零點模式下，皆能決定出較正確的極/零點係數並合成與真實 HRTF 相似的波形。

在分群方面，我們分別針對極點與 HRTF 做分群，結果發現經由分群後的 HRTFs 的確較未分群的 HRTFs 能決定適合的極/零點係數。

參考文獻

- [1] D.R. Begault, 3D Sound for Virtual Reality and Multimedia, AP Professional, 1994.
- [2] W. Gaik, "Combined evaluation of interaural time and intensity differences Psychoacoustic result and computer modeling" JASA, 94(1), 1993. July.
- [3] J. Chen, B.D. Van Veen, and H.E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," J. Acoust. Soc. Am. 97(1), pp. 439-452.
- [4] C.C. Chuang, "Study on HRTF Clustering and Synthesis with 3D Sound Applications," Master thesis, National Chiao Tung University, 1995.
- [5] C. P. Brown and R.O. Duda, "An efficient HRTF model for 3-D Sound", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House.
- [6] J. Blauert, "Spatial hearing," MIT Press, Cambridge, MA, 1997.
- [7] Y. Haneda and Y. Kaneda, "Interpolation and extrapolation of common transfer functions based on common acoustical poles and their residues," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House.
- [8] J. G. Proakis and D. G. Manokakis, "Digital Signal Processing: Principles, Algorithms, and Applications," third ed., Prentice Hall, 1996.

五、圖表

