

個人化 HRTF 合成機制之研究

莊盛宇

大同大學 資訊工程學系
e1348ms29@gmail.com

張嘉銘

大同大學 資訊工程學系
cmchang@ttu.edu.tw

摘要—近年來 3D 立體音效技術漸漸被運用在各種應用之中。要產生 3D 立體音效，就必須探討音源定位的問題。由於每個人先天上的外觀差異，例如軀幹、肩膀、頭型、耳廓大小的不同，造就每個人獨特的 HRTF。HRTF 包含了聲波從音源到達耳膜間，所有頻率響應的總和，和音源定位有相當大的關係。要進行音源定位，就必需對個人化 HRTF 進行準確的調整。

本篇論文提出了一個運用其他測試者已量測的 HRIR 資料，來進行調整、合成出個人化 HRIR 的演算法。首先，我們利用樣本資料的 HRIR 去製作半環繞音效。接著讓測試者去聆聽這些半環繞音效，找出其中定位感覺不太平順的部份，利用多頻率聲音測試其他樣本同一位置的 HRIR 資料，若有較佳的結果即替換之。將整個 HRIR 測試且代換完畢後，讓測試者進行最後的聆聽測試，若有定位不準處再重覆替換測試，直到完成調整。

利用這樣的一個系統，就可以不用經由繁瑣複雜，且還有危險性的侵入式測量，而得到每個人獨特的 HRTF 資料。

關鍵詞 — Head related impulse response (HRIR)，Head related transfer function (HRTF)

一、緣由與目的

隨著近年來 DSP 技術不停的發展以及多媒體的發達，3D 立體音效的技術在這幾年也蓬勃發展起來。所謂的 3D 音效，主要來說就是透過兩個喇叭，或是一對耳機，就可以感受到三度空間的聲音效果。換句話說，我們可以透過 3D 立體音效，感受到有如親臨現場般的真實聽覺感受。

在這幾年中，3D 立體音效已經漸漸的運用在許多的應用方面之中。例如在科學，商業，以及娛樂方面，或多或少都可以看到 3D 音效被運用在其中。

目前，有許多研究都有提到如何產生 3D 立體音效，其中可以歸類為兩大類，分別是多聲道，以及雙聲道[4]。

多聲道的方法主要是將實際空間中的音源位置，透過多個擴音器傳達給收聽者，例如杜比 5.1 環繞音效。這種方法可以有效的傳達音源的位置，但是為了要達到精準的音源定位，便需要為數不少的擴音器，如此需要花上一筆昂貴的費用[5]。而且對於某些應用來說，多聲道並不方便使用。

雙聲道的方法主要是透過耳機，讓使用者達到同樣達到有如收聽 3D 立體音效的效果。此種方法只要考慮耳機左右兩聲道，不需要昂貴的裝備，且容易實現在許多應用中。因此，本論文是以使用雙聲道的耳機來進行研究。

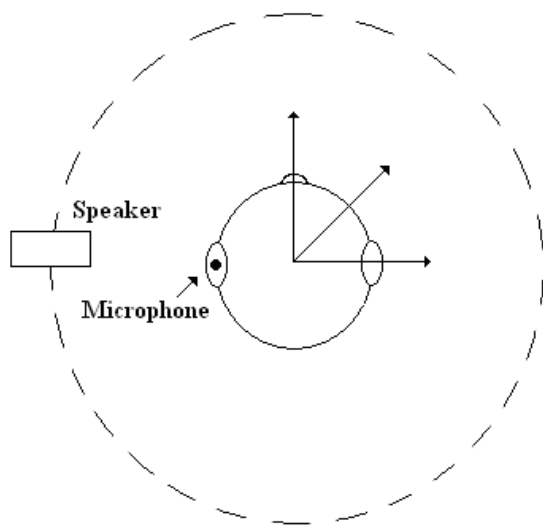


圖 1: 侵入式測量法的簡單示意圖，將微型麥克風放置在測量者或者是擬仿照頭型所製作模型的耳道之中，然後以測試者或是模型為圓心，和點音源的距離為半徑所產生的假想圓，讓點音源在圓周上移動，對各角度進行測量，由麥克風接收聲音，再彙整成 HRIR 或經 DFT 轉成 HRTF。

爲了要達成上述的效果，首先必須要讓使用者在耳機中聽到的聲音具有方向性，換句話說就是讓使用者可以判斷耳機中所聽到的聲音是來自於哪個方向。於是，我們便需要使用者的 Head Related Transfer Function (HRTF) 或 Head Related Impulse Response (HRIR)，來運算以達成這個目的。

要得到精確的 HRIR 資料，目前最準確的方法是使用侵入式測量法。主要的步驟是將微型麥克風放入被測量者的耳道內，或者是使用仿照頭型所製作的模型，然後在離測試者或是模型一定距離遠的位置放置點音源，以測試者或模型當作圓心，距離當作半徑假想出一個圓形。讓點音源在這個圓形的圓周上移動，對每個角度進行測量，再由麥克風接收 HRIR 參數資

料，如圖 1 所示[7]。

侵入式測量的方法雖然可以得到準確的 HRIR 數值，但是卻非常的耗費時間，還需要複雜精密的技術和儀器來進行測量，而且重要的是，侵入式測量的方法具有一定的危險性。

因此，如何使用較不複雜的方法，降低測量或計算的時間，且沒有安全上的顧慮，卻可得到一樣精準結果的研究，便成爲另一個研究的方向。

我們所使用的方法，是利用其他測試者已量測的 HRIR 資料，透過提出的演算法，合成每個人獨特的個人化 HRIR 資料，圖 2 是其中三個測試者在同一高度角下，所有方位角的 HRTF 和 HRIR 圖像資料。從圖像中我們可以看出，不同人的 HRTF 和 HRIR 的圖像，多少都有些許的差異性。

利用此種合成調整方法，最後可以得到精準的個人化 HRIR 資料。未來，此種方法可以應用在商業、娛樂、或是科學上，透過合成調整 HRIR，讓即使是不同人，卻都能感覺到同樣的聲音位置來源。甚至進而產生如同真實聽覺般相同的臨場感，成爲一種新型的視聽型態。

在之後的章節裡，第二節將對相關研究進行介紹，更加詳細說明的調整方法會在第三節進行說明，第四節會敘述我們實驗的過程，第五節將陳述結論。

二、背景與相關研究

虛擬音源的成立，主要是利用聽覺心理學 (Psychoacoustic) 中人對於聽音辨位的心理改變爲其理論依據。不同方位的聲音傳到耳中，大腦都會紀錄不同的頻譜與時間差。所以我們便可針對人類的耳朵和大腦感受的頻率響應特性來合成出 3D 立體音效。

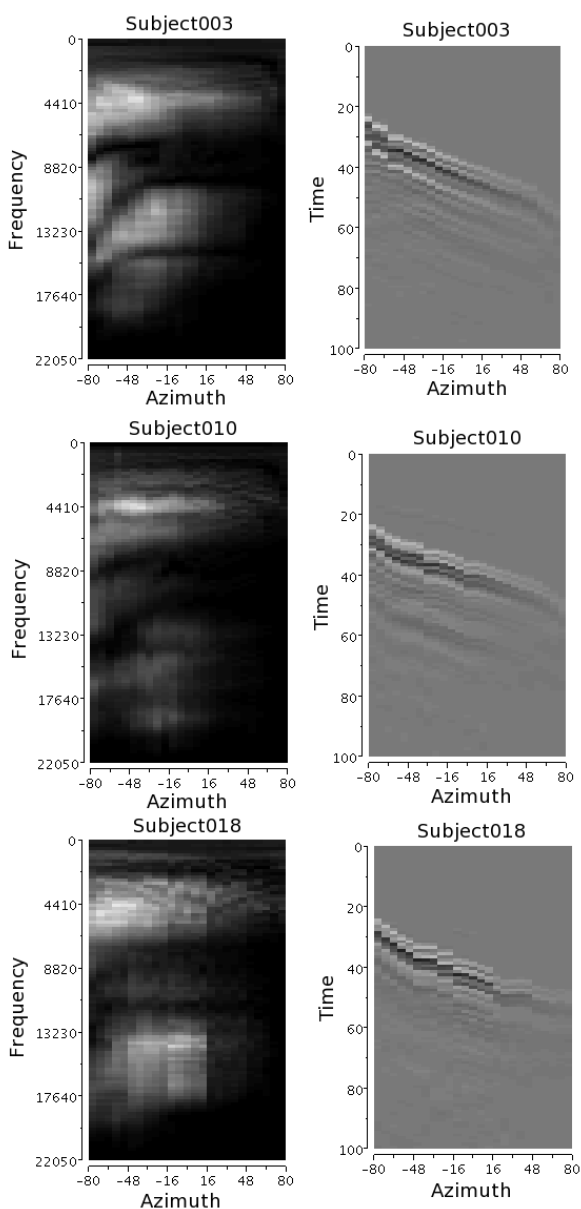


圖 2:其中三位測試者在高度角 0° ，所有的方位角 (-80° 到 80°) 中左耳部份的 HRTF 和 HRIR 圖像。左列代表的是 HRTF 圖像，右列則代表的是 HRIR 圖像。

爲了要進行音源定位的修改及調整，必須先探討兩個和音源定位有主要相關性的因素：Interaural Time Difference (ITD) 和 Interaural Intensity Difference (IID)。當聲波經由音源到達我們的雙耳，會因爲距離的不同而產生時間

差，這個時間差就是 ITD。IID 則是當聲音從某個方位傳來，兩耳所接收的聲音強度差便稱爲 IID[8]。

當音源的來源位置要傳到位置較遠的耳朵時，會受頭部遮蔽效應 (Head shadow) 影響，這遮蔽效應會使聲波如光波般以繞射的方式沿著頭部外圍進入到較遠的耳朵中。這聲波繞射的現象除了和音源的角度有關外，也和頻率有關係。一般來說，高頻的聲音較不容易繞射，所以能量的衰減也較大，傳達到兩耳的強度差也就較爲明顯。所以高頻的聲音適合使用 IID 進行定位。此外，因爲高頻波長較短，產生的相位變化較大，於是高頻聲波延遲就比低頻聲波較短，故延遲較長的低頻聲波便適合使用 ITD 來定位[1]。

另外，聲波除了透過頭部的遮蔽效應，以繞射方式進入耳朵中，也會經由肩膀或者是軀幹反射進入。故在討論音源定位的時候，肩膀和軀幹這些因素還是有討論的必要。

在文獻中，建構頭部或者是耳廓的模型通常在測量個人化 HRIR 資料已經是個常見的步驟之一。例如：建構了一個稱爲 snowman (head-and-torso) 的頭和軀幹模型，透過相關影響因素 ITD、IID 以及聲波經過身體 (肩膀、軀幹...etc) 和耳廓的反射，計算出不同角度的 HRIR[2][6]。

但是要如何建構出一個誤差非常小，幾乎完美的模型？這是非常困難的問題。每個人身體外觀的差異性完全不同，只建構一個模型並無法適用在每個人身上，如此一來得到 HRIR 參數就可能會有較大的誤差。

如果不使用建構模型，也不使用侵入式測量，是否還有其他方法來進行調整？有文獻提出了一種使用現有資料來進行調整的方法。首先測量測試者雙耳的幾何參數，然後和樣本裡

的 HRIR 資料進行比對，雙耳各自比對誤差最小的樣本資料，就將他當作測試者的資料來進行使用[9]。

但是進行幾何參數測量的話，就會有測量誤差的問題。不管是測量的精準度，還是參數進行測量的位置，這些都會影響到比對誤差的結果，而導致得到錯誤的樣本資料，造成 HRIR 產生較大的誤差。

所以雖然我們同樣選擇使用現有資料來進行調整，但是不進行幾何參數的量測，而是直接利用聆聽的方法，對樣本資料進行替換，以及使用內插法計算來調整得到我們所想要的個人化 HRIR 數值。

關於內插方法，已經有文獻提到了如何能讓內插的數值更為精準。例如運用了聲音能量位能和一種稱為 Truncation Number 的變數，內插和外推運算其他距離上的 HRIR 值。或者是使用提可諾夫規化法 (Tikhonv Regularization) 來減少對於內插上面可能會產生的一種鋸齒狀的誤差情形[7]。

三、HRTF 合成調整方法

我們提出一個不用建構模型的方法，直接使用多個已測試樣本的 HRIR 資料來進行合成調整。整個方法大略可分成三個階段，半環繞音效測試，多頻率聲音測試，以及最後的試聽測試。

(一) 半環繞音效測試

第一步是進行半環繞音效測試，主要的步驟如同圖 3 的流程圖，首先固定高度角，取出其中一個已測樣本的所有方位角 HRIR 值製作半環繞音效，讓測試者聆聽。

由於要產生半環繞音效，需要在每個方位上都有 HRIR 的值。但是樣本資料裡方位角並非連續的 (-45 到 45 度的間隔為 5 度；-45~-65 和

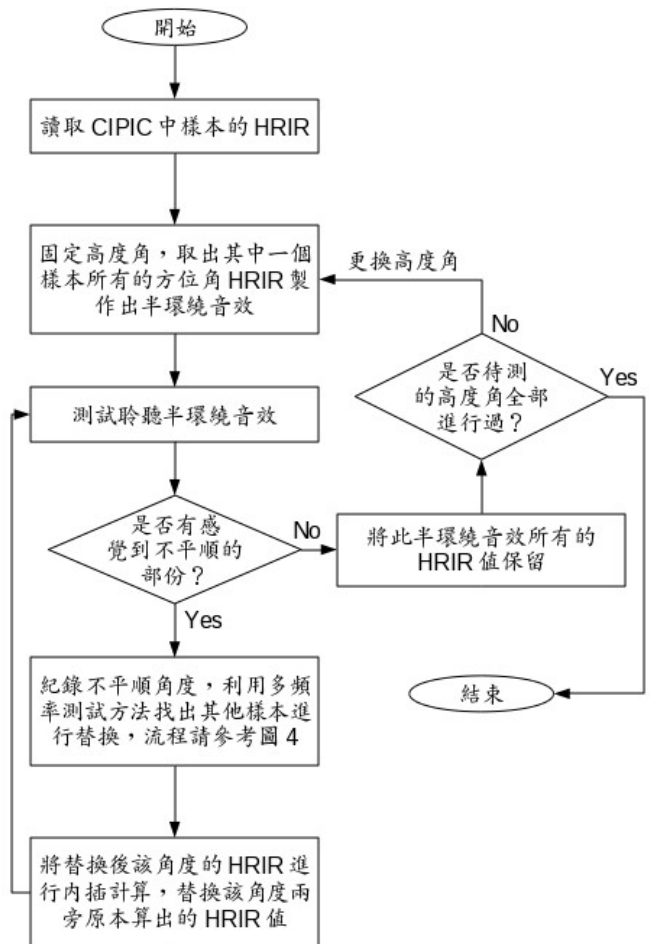


圖 3: 半環繞音效測試流程圖

45~65 度的間隔為 10 度；以及 -65~-80 和 65~80 度的間隔為 15 度。），所以在這些方位角間隔內角度的 HRIR 值，便需要使用內插法來計算出來，如此才能產生半環繞音效。關於內插的方法，我們使用前一節中所提到文獻內的方法來進行計算。

因為樣本和測試者的個人化 HRIR 之間並不相同，所以在聆聽半環繞音效時會感覺到幾個不平滑的點。這裡所謂不平滑的特徵，例如某些地方聲音感覺到的位置變化量超出預期、聲音產生了停留感或是聲音無法判別位置。一旦聆聽時發現到了這些狀況，便紀錄這些不平滑點的角度，進行下一個步驟。

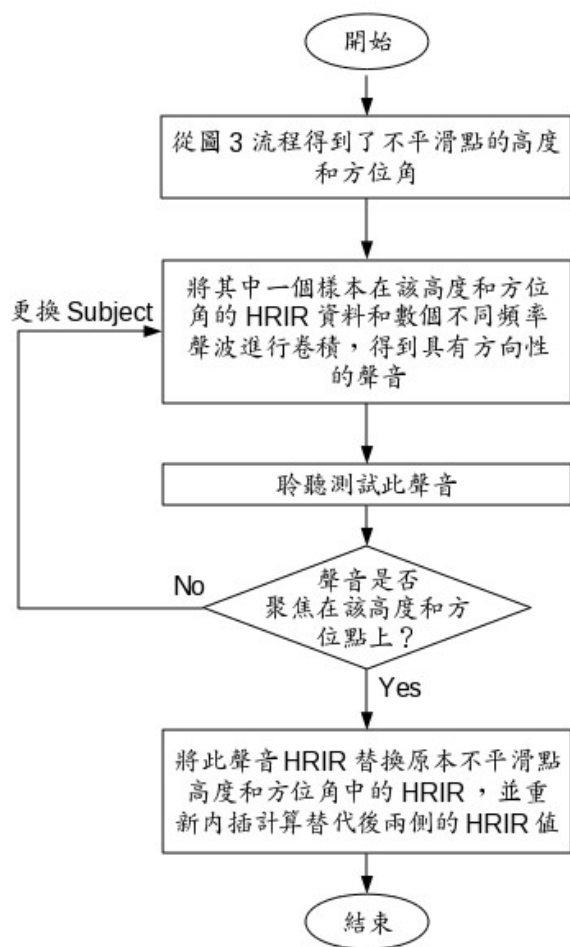


圖 4: 多頻率聲音測試流程圖

(二) 多頻率聲音測試

確認不平滑點的角度之後，接著進入第二步多頻率聲音測試，進行步驟如同圖 4 的流程圖。我們使用數個包含了不同頻率的聲波，用這些聲波來和其他樣本該角度的 HRIR 來進行卷積，製造了包含了不同的頻率，並擁有同一方向性且長度約為一秒的聲音，讓測試者聆聽並判斷聲音方向。若聲音聆聽的效果幾乎聚焦在設定的角度，便顯示了在這些不同頻率下，該點聲音都能保持在同樣位置。同時這也表示了這個樣本在該角度中的 HRIR 非常接近測試者個人獨特的 HRIR。於是該角度的 HRIR 資料，就能用來替換之前感覺到不平滑點的角度。

替換之後，重新聆聽測試這半環繞音效是否已經變的平滑。若有，則這半環繞音效便進行保留，換另一個高度角再從步驟一重新製作新的半環繞音效。若還是感覺到不平滑，則回到步驟二繼續對不平滑點的角度進行替換。

重複進行以上步驟數次，最後便得到了經過代換完成後在半環繞音效裡這些角度的 HRIR，這代表這些音效中的各個位置 HRIR 已經合成調整完畢，接近測試者獨特的 HRIR。

(三) 視聽測試

之後進行步驟三試聽測試。我們讓系統隨機挑出幾個角度的 HRIR，並使用卷積產生該角度的聲音來進行測試。這邊比較特殊的是，挑選出來進行測試的 HRIR 並不限只在半環繞音效上面，也有可能是在半環繞音效之外的位置。

最後讓測試者進行聆聽測試，比較調整過後的聲音和聆聽感覺的角度誤差，並完整記錄下來，評估這個合成調整方法所執行的調整改善成果如何。

四、系統實作

我們使用加州大學的 CIPIC (Center for Image Processing and Integrated Computing) 實驗室所量測的 HRTF 資料。CIPIC 的測試者資料一共有 46 個 (45 個實測者樣本以及一個 KEMAR 樣本資料)，除了 HRTF 和 HRIR 的資料之外，還包含了每個測試者的有關人體測量學上的身體 (軀幹，肩膀以及耳廓) 參數[3][10]。

關於實驗環境，我們使用 Scilab-5.1.1 這個版本的軟體來進程式寫作以及測試。首先我們挑選了 CIPIC 裡其中一個樣本，將高度角 0 度下所有的方位角 HRIR 值選取出，並使用頻率為 1KHz 的聲波，讓各方位角的 HRIR 和聲波進行卷積，產生了一個半環繞音效。

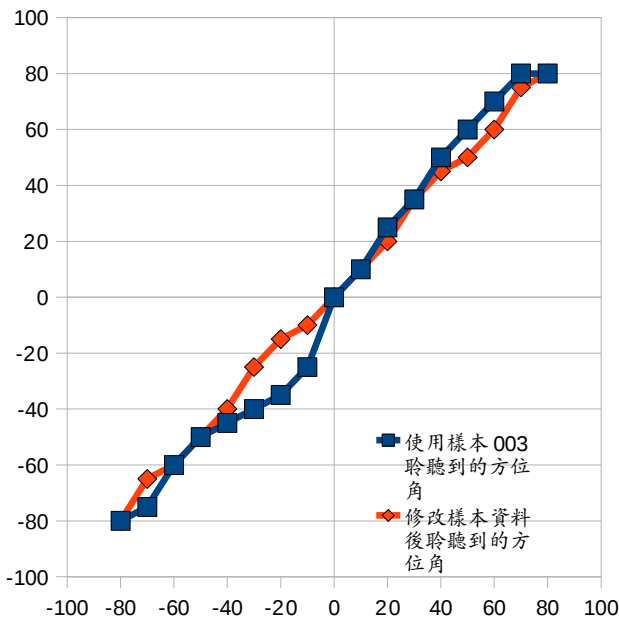


圖 5: 在高度角 0° 下，所有的方位角 (-80° 到 80°) 的相位圖。藍線為使用 CIPIC 裡其中一個樣本進行聆聽測試後所得到的相位圖，紅線為替換同一個樣本裡的 HRIR 後再進行聆聽測試所得到的相位圖。

接著，我們讓測試者去聆聽這個半環繞音效，看是否會感覺到不平滑的部份。之後，我們對於測試者所感覺到的不平滑點的角度，取出了 CIPIC 裡其他樣本但同樣角度的 HRIR 值，利用包含了數個不同頻率的聲音，讓測試者去聆聽聲音是否都聚焦在同一個點上。

圖 5 的是 CIPIC 裡其中一個樣本的原始 HRIR 資料，和同一個樣本但經過替換 HRIR 後來進行聆聽測試後的所記錄下來的相位結果。比較後，可以發現方位角在 -20 和 60 度附近，替換樣本比原始樣本進行聆聽的結果較為接近原本預設的方位角。而且在聆聽感覺上，原本使用未替換的樣本所製造出的半環繞音效，在某些方位角會感覺不平順，但在替換後所製造出的音效中卻變得平順，沒有再感覺到不平滑

的部份。因此可以確定，替代調整過後的半環繞音效上各個方位角的 HRIR 值，和測試者個人獨特的 HRIR 經過合成調整後已經變得相近。

五、結論

我們提出了一個直接利用現有的 HRTF 資料來進行個人化合成調整的方法。透過製造半環繞音效來測試並使用多頻率聲音測試代換聆聽不平滑點的角度，最後得到在半環繞音效裡所有方位角和個人化相近的 HRIR 值。

目前實驗的結果，可以確定了用半環繞音效和多頻率聲音測試尋找不平滑點的角度來進行合成調整，確實可以調整出和測試者的個人化相近的 HRIR。

相信透過此種合成調整的方法，未來我們可以快速的得到精確的個人化 HRIR 以及 HRTF，節省了麻煩的建構模型和侵入式測量所耗費的時間，也不需要昂貴的設備，更沒有安全顧慮上問題的麻煩。

參考文獻

- [1] 張淑君，“空間音效模式化與數位實現之研究”，國立成功大學，May 2002.
- [2] V. R. Algazi, R. O. Duda, D. M. Thompson (2002). “The use of head-and-torso models for improved spatial sound synthesis”, Proc. AES 113th Convention, Los Angeles, CA, preprint 5712.
- [3] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. “The CIPIC HRTF database”, Proc. IEEE WASPAA01, New Paltz, New York, October 2001.
- [4] K. J. Faller II, A. Barreto, N. Gupta, N. Rish. “Decomposition and Modeling of Head-Related Impulse Responses for Customized Spatial Audio”, WSEAS Transactions on Signal Processing, vol.1, pp. 354-361, 2005.
- [5] Alan Kraemer. “Two speakers are better than

- 5.1”, IEEE Spectrum, May, 2001.
- [6] Chi-Jang Pan. “Study of Adaptive Mechanism for Personalized HRTF”, Tatung University, July 2003.
 - [7] D. Ramani, D. N. Zotkin, N. A. Gumerov. “Interpolation and range extrapolation of HRTFs”, IEEE ICASSP, 2004.
 - [8] J. W. Strutt (Lord Rayleigh). “On our perception of sound direction”, Phil. Mag., 13:214-232, 1907.
 - [9] D. N. Zotkin, Ramani Duraiswami, Larry S. Davis, Ankur Mohan, and Vikas Raykar . “Virtual audio system customization using visual matching of ear parameters” Proc. ICPR 2002, Quebec City, Canada, August 2002.
 - [10] CIPIC HRTF Database Files, Release 1.0, August 15, 2001, available at <http://interface.cipic.ucdavis.edu/>