

多聲道錄音與訊號處理

Multi-Channel Recording and Signal Processing

組別：

組員：林秉橙、林蓁卉、王彥晴、朱恩寬

指導教授：劉奕汶

一、摘要

本專題主要探討多聲道錄音與訊號處理。多聲道錄音方面，以3D音效中”Ambisonic”技術結合數位訊號處理之推廣應用，使用市面上的360°全景麥克風 ”Zoom H3-VR” 作為實驗器材之一。為了解Ambisonic中Ambix B格式與其他常見平面單、雙、多聲道錄音格式的不同，嘗試格式之間進行數學運算轉換的方法後，以”Zoom H3-VR”錄製素材，並在Matlab程式執行轉換，輸出5.1聲道格式。

此外關於訊號處理的部分，我們以麥克風錄製單聲道音檔，經由我們實作的調頻系統，由原有頻率經過數位訊號處理轉移後產生調頻後的新音檔。另一方面，原素材與特定的頭部相關函數卷積，以數位處理方式讓單聲道音檔產生環繞音的感覺，輸出雙聲道格式。

本實驗最後將藉由邀請多位受試者填寫本專題所設計的問卷，並請受測者答覆關於實際聆聽後的感受；藉此便可驗證我們的系統能否有效建構出預期的成果。

二、前言

日新月異的科技帶來進步的技術，音訊領域的發展也擁有相當大的潛力。近年來，許多電影、遊戲等虛擬實境，都包含擬真且能夠隨著視角轉換而跟著改變的沉浸式音效。本專題實作中使用的技術有別於人們所熟知的單聲道、立體聲、5.1 等聲道導向格式，而是一種場景導向的技術，名為”Ambisonic”。

閱覽相關資料後，我們得知由於弦樂器能夠演奏出持久且相同的音色，因此對其進行訊號處理是可行的。不禁讓我們開始思考，打擊類樂器是否能藉由移頻方式調整音高？便開始了對「磬」的研究。此外我們也根據人類的頭部相關傳遞函數處理訊號，透過調整兩耳時間延遲量差、音量大小等，用原聲訊實作出聲源移動的錯覺，最終達到足以欺騙人耳的擬真程度。

二、原理分析與系統設計

2.1 原理分析

2.1.1 Ambix B format

(1) 公式分析

在此專題中使用的 ambisonic 麥克風，是4個 capsule 以四面體結構所組成，可錄製出分別指向右前(RF)、左前(LF)、右後(RB)、左後(LB)之四聲道 A-format 格式檔案。經過分析後，可將 A-format 格式轉換為目前最廣泛使用的 B-format 格式，此格式代表空間中360°的球體之全向(W)、左右(Y)、前後(X)、上下(Z) 四個方位。

其中，A-format 格式轉換至 B-format 格式的公式如下：

$$LF = 0.25 * (W + X + Y + Z) \quad (式1)$$

$$RF = 0.25 * (W + X - Y - Z) \quad (式2)$$

$$LB = 0.25 * (W - X + Y - Z) \quad (式3)$$

$$RB = 0.25 * (W - X - Y + Z) \quad (式4)$$

求解式1至式4，可得 B-format 格式轉換至 A-format 格式的公式如下：

$$W = LF + LB + RF + RB \quad (式5)$$

$$X = LF - LB + RF - RB \quad (式6)$$

$$Y = LF + LB - RF - RB \quad (式7)$$

$$Z = LF - LB - RF + RB \quad (式8)$$

2.1.2 Frequency shifting and Hilbert transform

(1) 轉移頻率(Frequency shifting)

由離散時間傅立葉轉換公式(Discrete-Time Fourier Transform)：

$$x[n]e^{j\Omega_0 n} \xleftrightarrow{DTFT} X(\Omega - \Omega_0) \quad (式9)$$

可將整段頻域訊號 $X(\Omega)$ 移動固定頻率 Ω_0 。

(2) 希爾伯特轉換(Hilbert Transform)

在訊號處理中，希爾伯特轉換可將實數訊號拓展到複數平面，並使其滿足柯西－黎曼方程式。於本專題裡則用以將頻域中的負數部分移除，同時讓原實數訊號轉為複數進行移頻。函數 $u(t)$ 的希爾伯特轉換為 $u(t)$ 與 $h(t) = \frac{1}{\pi t}$ 的卷積，可寫為：

$$H(u)(t) = \int_{-\infty}^{\infty} u(\tau)h(t - \tau)d\tau = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} d\tau \quad (\text{式10})$$

其經由傅立葉變換後的頻率響應(frequency response)為：

$$H(\omega) = (-i * \text{sgn}(\omega)) * F\{h\}(\omega) \quad (\text{式11})$$

其中

$$\text{sgn}(\omega) = \begin{cases} 1, & \text{for } \omega > 0 \\ 0, & \text{for } \omega = 0 \\ -1, & \text{for } \omega < 0 \end{cases} \quad (\text{式12})$$

既然對於一個函數 $s(t)$ 有：

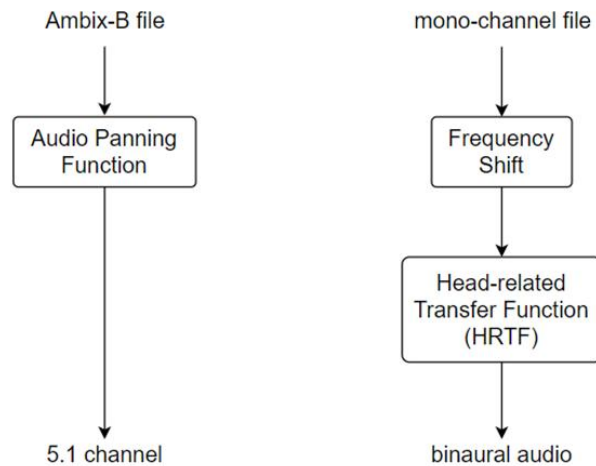
$$F\{\hat{s}\}(\omega) = H(\omega) * F\{s\}(\omega) \quad (\text{式13})$$

則希爾伯特轉換就是將 $S(\omega)$ 的負頻率成分偏移 $+90^\circ$ ，而正頻率成分偏移 -90° ，如此便可用於去除負頻率成分。

2.1.3 Head-Related Transfer Functions

頭部相關傳導函數 (Head-Related Transfer Functions, HRTF) 是一種音效定位演算法，它利用聲音傳播到雙耳的 ITD(Interaural Time Delay)及 IAD(Interaural Amplitude Difference) 設計出一組濾波器，此濾波器為 Head-Related Impulse Response (HRIR)，而 HRTF 則是 HRIR 的傅立葉轉換。

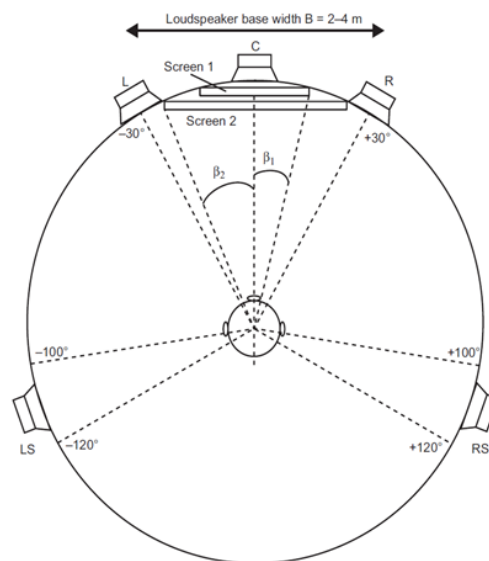
2.2 系統設計



圖一、system flow chart

2.2.1 Ambix B format to 5.1 channel panning formula

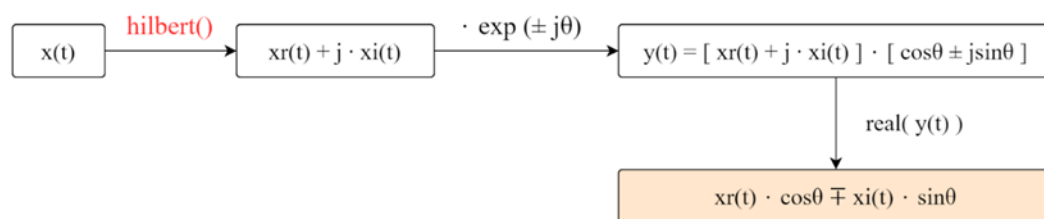
Dolby 5.1 聲道的音頻格式擴展了左右傳統立體聲通道，提供包絡 360° 的音訊，被廣泛運用於電影院、電視與遊戲之中。5.1 音頻各方向通道具不同用途，左、中、右三個前置聲道提供清晰、乾淨的對話音與精準的定位；左右雙環繞聲道則能營造出身歷其境的感覺。低頻聲道提供深沉的低音效果，由於只需要其餘通道大約十分之一的頻寬，因此被稱為“.1”聲道，圖二是常見的 5.1 聲道揚聲器擺設位置：



圖二、The placement of loudspeakers in 5.1 channel

2.2.2 Frequency shift using Fourier and Hilbert transform

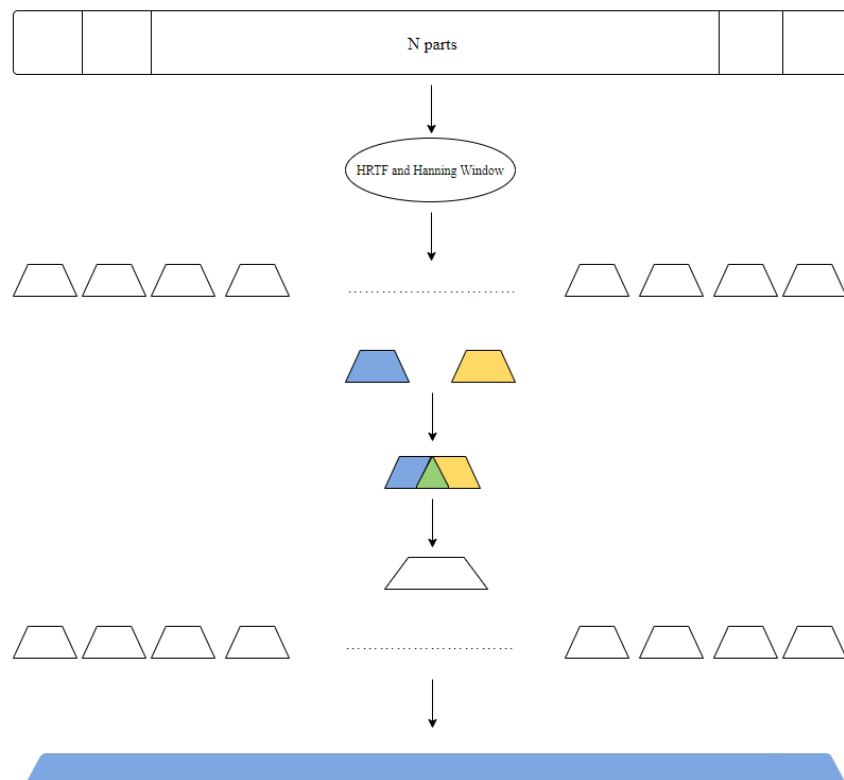
我們在移頻之前進行希爾伯特轉換，同時去除負頻率成分，並將原始純實數音檔轉為複數，在不更動正頻率成分的條件下讓左右移頻的值產生差異，使實際聆聽感受較為準確。下圖三為其流程圖。



圖三、Flow chart of frequency shifting using Hilbert transform

2.2.3 HRTF

我們輸入一段音檔，將音檔平均平分為 N 等分，然後將分割後的每段音檔與 HRIR 進行捲積。接著，我們將捲積過後的音檔利用 hanning window 來製造淡入及淡出的效果，再將每一段淡出的部分及下一段淡入的部分進行合併，即可合成出具有 3D 立體效果的音檔，如下圖四所示。



圖四、HRTF 實作示意圖

三、實驗結果

3.1 Result of Ambix B format to 5.1 channel panning formula

根據我們的計算得出下列式 14，並使用錄音檔案且利用 matlab 製作出 Ambix B format 的檔案製作出 5.1 channel 的檔案，將同一錄音檔案與市面上的聲道轉換之軟體”zoom player”做比較，可發現有人聲的部分誤差小於 1%，而其他無人聲的部分，誤差會大至 5%，推測是因為音檔是於隔音良好的錄音室內錄製，因此無人聲的部分音量極小，檔案輸入進 matlab 儲存後，便會無法儲存到完整的數值，因此才與 zoom player 轉換出的 5.1 聲道之檔案呈現較大的誤差。

$$S = 0.5(X \cdot \cos \theta + Y \cdot \sin \theta + W) \quad (\text{式 14})$$

3.2 Result of frequency shift of audio files

從上述的調頻實作方法可知，使用希爾伯特轉換(Hilbert transform)以及頻率移動(frequency shift)，可以對一沒有明顯泛音存在的音檔在保留完整功率頻譜下，讓它整段頻譜平移固定的距離，進而達成調頻的目的。如前所述，我們此種調頻方法是針對沒有明顯泛音存在的樂器，例如本次專題中所使用的素材「磬」。有明顯泛音之樂器不適用本專題中所提及的調頻方法之原因為，泛音序列為構成其音色及聆聽感受的主要成分，但 frequency shift 會使樂器中原本存在的泛音序列被打亂，因而使音檔失去原本的音色，聽起來相當不自然。

在專題的最後，我們設計了一份問卷，並邀請了 35 位受測者，為了使實驗在同一條件下進行調查，以同一副耳機進行聽力測試，此份問卷的目的聚焦在我們所實作的調頻方法是否可以成功虛擬出樂器的聲音，因此問卷中虛擬出的音檔皆是使用本專題中所述之調頻方式。而本次問卷調查共有以下兩個主要目的：

1. 分析利用調頻系統虛擬出的音檔之聲音品質與參考音準相較之音高準度。
2. 利用調頻系統虛擬出的音檔與真實音檔比較，分析虛擬音檔是否夠真實。

根據收集到的數據結果，非泛音型樂器，例如磬，其樂器本身的音色占比很重，因此當調頻的範圍愈大時，由於誤差累積，音色會越失真，故整體聆聽感受會較移頻範圍小的音檔組成的旋律差。其二為由於磬為打擊樂器，因此其敲擊瞬間的金屬音色也是音檔中很重要的一部分，且其敲擊瞬間的金屬音與磬敲擊後持續震動所產生的音，兩者音高並不相同，金屬音主要與樂器本身有關，換句話說，磬的金屬音與持續音兩者之間並沒有明確的線性關係，因此使用本專題的調頻系

統進行大範圍移頻時，敲擊瞬間的金屬音與持續音的不和諧性會增加，故整體聆聽感受會下降。

此外，問卷中的實驗一中有三個音檔，其中只有一個是真實的聲音檔，其他兩者皆為聲所移頻出的虛擬音檔，在此要請受測者判斷何者為他們認為的真實音檔。而在實驗四中，有六個音檔，同樣地，其中只有一個是真實的聲音檔，其他五個皆為聲所移頻出的虛擬音檔，再請受測者判斷何者為他們認為的真實音檔。根據收集到隻的數據結果，實驗一中成功選擇到真實音檔的比例為 37%，實驗四中成功選擇到真實音檔的比例則為 20%。從上述結果可知，使用本專題中調頻系統所製作出的虛擬音檔真實度非常高，意即我們在移動頻率外，也很好的保留了樂器本身的特色，因此受測者難以區別出何者為真實音檔，何者為虛擬音檔。

3.3 Result of HRTF

將 HRTF 系統設計完成後，我們理論上可以將音源位置平移在平面上的任一位置。在討論環繞音效時，我們有注意到人耳對於左右聲道的區別較能分辨，但對於前後聲道的分辨則較難判斷。因此本實驗目的不僅在於我們想了解 HRTF 系統的成效，也想透過設計題目讓受試者做測驗，以驗證音源距離半徑固定時，人耳對於前後聲道較難分辨的客觀事實。

本實驗我們透過調整 HRIR 函數，與音源素材進行合成，設計 3 組不同的題目給受試者評價 HRTF 系統製造的環繞音效。本實驗共蒐集 35 位受試者的作答，3 組實驗問題如下：

1. 聲源方位判定

本題使用相同的原始音檔(Nokia 手機鈴聲)，在 HRTF 系統裡，選特定方位 HRIR 函數與原始音檔合成，使原本單聲道的音檔可在特定方位播放。本題測試音檔共有 2 個，每個測試音檔均分成 2 段，來自不同方位的播放。

受試者應可透過耳機感知每段測試音檔來自不同方位。請受試者先選出這 2 個音檔出現的音源方位，再依辨識程度排序這 2 個音檔。目的為了解受試者對於空間音感的辨識程度。以下為受試者的作答結果，以雷達圖表示方位呈現。音檔 1 正確答案於第一段方位應為右前方，而第二段應為右後方；音檔 2 正確答案於第一段方位應為左後方，而第二段應為右前方。

從結果可知，音檔 1 第一段方位答對率 49%，音檔 1 第二段方位答對率 17%。音檔 2 第一段方位答對率 69%，音檔 1 第二段方位答對率 46%。透過第一個實驗，根據受試者的填答(關於兩個音檔判斷方位的難易度)，

我們驗證受試者對於音源在左右的位置播放較易辨識，當音源在前後的位置播放時，則較難以辨識區別。

2. 聲源環繞效果

本題使用三種素材，分別是磬、鋼琴和 Nokia 手機鈴聲。本題將特定方向 HRIR 函數改為全平面和素材音檔做合成，營造出環繞一圈的環繞音效，而這三題我們皆設計為順時針方向。以下為受試者的作答結果。

從結果可知，音檔 1 答對率 34%，音檔 2 答對率 40%，音檔 3 答對率 29%。透過第 2 個實驗，我們更加確定受試者對於音源在前後的位置較不易辨識，從結果可見大多數的人無法判斷正確的環繞方向。

3. 聲源環繞移動速度

本題延續上題的環繞音效，藉由分配 HRIR 函數在每個方位的時間長度，營造出聲源不只在移動，並且有速度變化的感覺。本題一樣與上題使用原始音檔素材合成。素材時間長度依序為鋼琴、磬、Nokia 手機鈴聲。設計的移動速度皆由快變慢。以下為測試結果。

從結果可知，音檔 1 答對率 57%，音檔 2 答對率 57%，音檔 3 答對率 51%。透過第 3 個實驗，我們可以驗證，當音檔時間長度越長，受試者較能判斷聲源移動速度的快慢之分。反之，如果音檔時間較短，有將近一半的受試者會無法判斷聲源移動速度的快慢。

四、結論

4.1 Ambix B format to 5.1 channel

本實驗由原始 Ambix B format 之多聲道環繞音效檔案，經我們推得的公式轉換為 5.1 聲道格式，再與由 Zoom Ambisonic Player 轉換而得的 5.1 聲道檔案做比對，其資料誤差幾乎介在 0.1% 以下，有相當高的準確性，實現並給予聽眾沉浸式 5.1 聲道的聲覺體驗。

4.2 Frequency shift

本次專題的調頻部份我們使用到兩種不同樂器，如磬與鋼琴(短按)，皆為泛音不明顯的樂器。因此，保留了完整的功率頻譜，成功的讓整段頻譜平移固定的距離。根據問卷收到的回覆，在品質、音高的整體聆聽感受也都有相當良

好的反饋，並且在真實的音檔透過我們的調頻系統後，實作出足以以假亂真的調頻音檔；亦即我們在保留原音檔音色的同時，成功地騙過了受試者的耳朵，讓他們無法分辨虛與實。

4.3 HRTF

經由問卷結果數據，驗證了當聲源距離半徑固定時，人耳確實對於前後聲道的感知，相較於左右聲道的辨別是不敏感的。此外，我們也了解到如果考量到聲源在固定的範圍進行有加速度的移動時，音檔的播放長度會影響到聆聽者對於速度變化的判斷能力。

五、心得

在這一年中，我們從對訊號處理一竅不通，到能夠實際運用手邊資源對我們錄製的音檔進行訊號處理，儘管過程走的路不算筆直，但當初發現 ambisonic 這個音訊處理技術還是令我們大吃一驚，感嘆科技發展帶來嶄新的感官體驗，也希望能透過這個研究更深入地理解這門科技。此外，關於調頻技術的研究，使我們更加理解並實現課堂上所學訊號處理相關的理論知識，讓文字不再只存在於課本中，令其跳脫框架，也讓我們的思想不再受侷限，能夠以更寬廣的視野、更多不同的角度「聽」待日常生活接觸的音樂。對於 HRTF，是一門嶄新的領域，不太了解的我們在做出結果後，對於此專題中所接觸到的訊號處理更加熟悉，收穫了經驗與成就感。我們相信這一年的經歷會扶持我們繼續成長。

六、參考資料

- [1] The ZOOM H3-VR, ZOOM.
Retrieved from <https://www.geogebra.org/graphing?lang=zh-TW> (Sep. 28, 2022)
- [2] Agnieszka Roginska and Paul Geluso, “Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio”, Oct. 2017.
- [3] Dolby Digital 5.1, Dolby Professional.
Retrieved from <https://professional.dolby.com/tv/dolby-digital/> (Sep. 28, 2022)
- [4] Sz, 2019. “VR 虛擬實境之聲音—淺談 Spatial audio、Ambisonic 等聲音處理技術,” Digilog. Retrieved from <https://digilog.tw/posts/1188> (June. 5, 2022)
- [5] V. R. Algazi, R. O. Duda and D. M. Thompson, “THE CIPIC HRTF DATABASE”, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2001.
- [6] Dimitris G. Manolakis, Vinay K. Ingle, “Applied Digital Signal Processing: Theory and Practice 1st edition”, Nov. 2011.