

# Ambisonics Microphone Recording and Playback to Multiple Channels

## 多聲道錄音與訊號處理

組別: B376

組員: 湯家昶

指導教授: 劉奕汶 教授

### 一、 摘要

本研究主要探討 Ambisonics 這種帶有空間資訊的音訊格式，利用多聲道的 Ambisonics 麥克風進行錄製，經過訊號處理，最終於多聲道的揚聲器系統進行播放。而 Ambisonics 麥克風可以依照麥克風的聲道數目，對應至球諧函數的階，區分成一階 Ambisonics(FOA)麥克風與高階 Ambisonics(HOA)麥克風。本研究中，使用一階 Ambisonics 麥克風 “Zoom H3-VR”以及高階 Ambisonics 麥克風 “Zylia ZM-1”作為實驗器材，對空間中的聲音進行錄製。此外，由於 Ambisonics 常利用球諧函數編碼處理，儲存成 B 格式，我們利用 MATLAB 設計並測試兩種解碼的方式，基礎解碼以及解碼矩陣。最終將錄音素材，經過解碼後儲存成多聲道的 WAV 檔案後，利用播放介面 “RME Fireface UCX II”進行四聲道的播放。

### 二、 簡介

近年來隨著科技進步，元宇宙或虛擬實境概念及技術日漸成形，因此在音訊方面也有相關的研究領域，期望能在空間中塑造或還原真實情境的聲音環境，讓使用者享有沉浸式的感官體驗。本研究利用 Ambisonics 這種帶有空間資訊的環繞聲音技術，以多聲道的 Ambisonics 麥克風進行錄製，儲存成 B 格式，並於 MATLAB 解碼儲存成多聲道的音訊格式，在多聲道的系統進行重放。

本研究著重探討 Ambisonics 儲存格式 B 格式，利用球諧函數的數學模型儲存時參數的調整與概念，還有利用球諧函數的特性進行解碼。由於球諧函數在錄音或是解碼方面對於空間中的參數有可調控性，因此增加了錄音時或播放音訊時的自由度，而突破杜比 5.1 或 7.1 聲道特定揚聲器擺放位置與規格的限制。

由於 Ambisonics 麥克風應用環境的自由性，我們想錄製能夠體現 Ambisonics 環繞效果的音訊格式，因此我們以二到三個樂器合奏的室內樂作為錄音素材。經過解碼與訊號處理後，最終期望能還原錄音的空間音訊，讓這二到三個樂器會有環繞聽者的感官體驗。

### 三、 背景

- **Spherical Harmonics and B-format**

球諧函數(Spherical Harmonics)是源自於在球型域解拉普拉斯方程所得出的解，因為是一組正交函數也是標準正交基底，可以利用球諧函數的線性組合表示球型域的任何的面，因此稱之為「球諧函數」。球諧函數除了能用來表示聲音在空間的波動，也常常應用在表達原子軌域模型[2]。

諧函數定義在球坐標系，位置可以利用 $\theta = (\varphi, \vartheta)$ 表示，如圖2-1可以看到球座標系與卡式座標系相對關係：

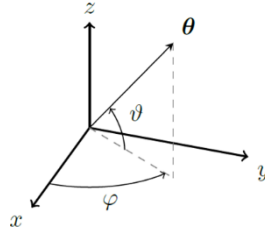


圖 2-1 卡式坐標系與球坐標系的位置關係[1]

因此，也可以表示卡式座標系的單位向量為：

$$\theta = \begin{pmatrix} \theta_x \\ \theta_y \\ \theta_z \end{pmatrix} = \begin{pmatrix} \cos \varphi \cos \vartheta \\ \sin \varphi \cos \vartheta \\ \sin \vartheta \end{pmatrix}$$

同時，也可以轉換為 $(\varphi, \vartheta)$ ：

$$\varphi = \tan^{-1} \frac{\theta_y}{\theta_x}, \quad \vartheta = \tan^{-1} \frac{\theta_z}{\sqrt{\theta_x^2 + \theta_y^2}}$$

球諧函數在數學上，分成以下幾個部分表示：標準化的部分以 $N_n^{|m|}$ 表

示、伴隨拉格讓德多項式 $P_n^{|m|}$ 、以及三角函數的部分，並表示成下式：

$$Y_n^m(\varphi, \vartheta) = N_n^{|m|} P_n^{|m|}(\sin(\vartheta)) \begin{cases} \sin|m|\varphi, & \text{for } m < 0 \\ \cos|m|\varphi, & \text{for } m \geq 0 \end{cases}$$

[2]。其中， $n$  是球諧函數的階(order)，而  $m$  是球諧函數的次數(degree)。在  $n$  階的次數  $m$  是從 $-n$  到  $n$ ，總共 $2n+1$ 個部分。可以從圖2-2看到球諧函數的圖形。

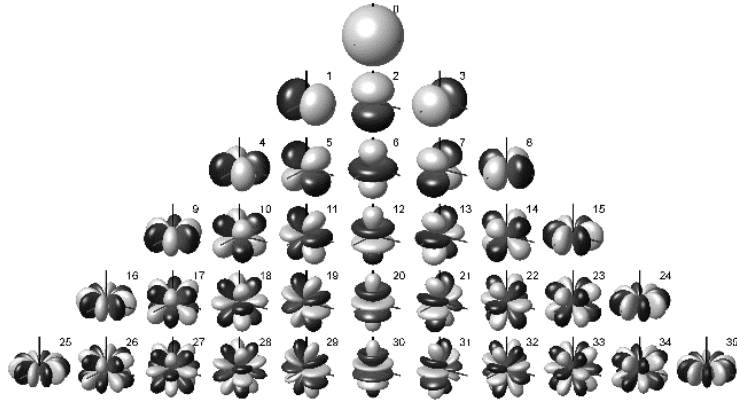


圖2-2 球諧函數0階到5階的函數圖形

(資料來源: [https://en.wikipedia.org/wiki/Spherical\\_harmonics](https://en.wikipedia.org/wiki/Spherical_harmonics))

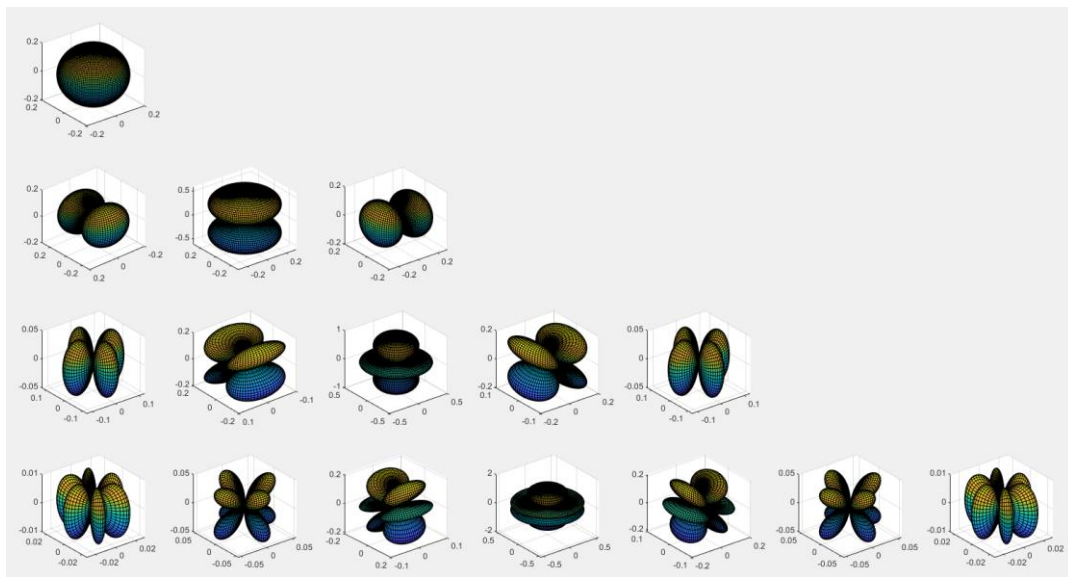


圖2-3 利用 MATLAB 繪製球諧函數0階到3階圖形

另外，標準化的部分  $N_n^{|m|}$ ，Ambisonics 常用 SN3D 和 SND 兩種標準化的方式。其中，SN3D(Schmidt semi-normalization)有0階的最大值特點，其他階不會超過0階的部分。另外，SND(full three-D normalization)則是確保編碼後音檔的每個成分有相等的能量。兩者的公式如下：

SN3D:

$$N_{m,n}^{SN3D} = \sqrt{(2 - \delta_m) \frac{(n - |m|)!}{(n + |m|)!}}, \delta_m \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0 \end{cases}$$

N3D:

$$N_{m,n}^{N3D} = N_{m,n}^{SN3D} \sqrt{2n + 1}$$

這兩者是 Ambisonics 的標準化模式，因此在編碼與解碼時須特別注意。[2][4]

B 格式是 Ambisonics 音訊的儲存格式，概念是將聲源訊號投影至球諧函數這組標準正交基底，可以表示成以下公式：

$$B_n^m = Y_n^m \cdot S$$

其中， $S$  是聲源訊號， $Y_n^m$  是球諧函數， $B_n^m$  則是 B 格式的編碼音訊。[1]

而一階 Ambisonics 的 B 格式訊號，因為一階包含0階和1階的部分，所以總共4個部分在 WAV 檔表示成  $n \times 4$  的音訊檔案。以次類推，三階包含0階、1階、2階、3階的部分，因此 WAV 檔表示成  $n \times 16$  的音訊檔案。

Ambisonics 麥克風錄音檔以球諧函數表示的階，是由錄音麥克風的數量決定，如果麥克風數量為  $N_m$ ，則  $\lfloor \sqrt{N_m} - 1 \rfloor = n$  其中  $n$  為此麥克風球諧函數的階。如此，能最有效的表示 Ambisonics 的音檔。[2]

## 四、 研究方法

### • System Design

本研究利用 Ambisonics 麥克風錄音，錄音過後經過球諧函數編碼成 B 格式的 WAV 音檔，再經過 MATLAB 設計解碼音檔並儲存成四聲道的音訊檔案，最終在四聲道揚聲器系統。如圖2-4:

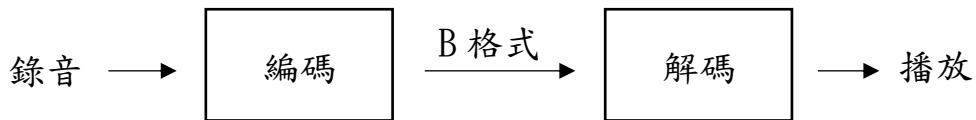


圖 2-4 系統流程圖

其中錄音利用 Ambisonics 麥克風，分別利用一階 Ambisonics 麥克風“Zoom H3-VR”和高階 Ambisonics 麥克風“Zylia ZM-1”。播放則利用“RME Fireface UCX II”播放介面在四個揚聲器播放。

### • Decoding

#### Basic Decoding

基礎解碼是利用一階的球諧函數，分別可以代表0階分量為  $W$  而1階分量為  $XYZ$ ，因此可以用以下通式解碼，並投放到位置於  $(\varphi_l, \vartheta_l)$  揚聲器  $L_l$ :

$$L_l = \frac{1}{N_l} \left( \frac{W}{\sqrt{2}} + X \cos \varphi_l \cos \vartheta_l + Y \sin \varphi_l \cos \vartheta_l + Z \sin \vartheta_l \right)$$

[1]。並且將四個揚聲器放置成正方形(左前、右前、左後、右後)，我們可以用以下公式解碼並投放到四個揚聲器:

$$L_{RF} = \frac{1}{\sqrt{2}}(W + X - Y)$$

$$L_{LF} = \frac{1}{\sqrt{2}}(W + X + Y)$$

$$L_{LB} = \frac{1}{\sqrt{2}}(W - X + Y)$$

$$L_{RB} = \frac{1}{\sqrt{2}}(W - X - Y)$$

[1]。利用 MATLAB 將 $L_{RF}$ 、 $L_{LF}$ 、 $L_{LB}$ 、 $L_{RB}$ 儲存成4個聲道的 WAV 音檔，並用 Windows 內建音樂播放軟體 “Media Player”播放介面播放音檔。

### Decoding Matrix

解碼矩陣是利用 B 格式的分量，投影到揚聲器所在位置的球諧函數。我們可以表達 $S$ 是總共 $N_l$ 的揚聲器向量， $D$ 是解碼舉證， $B$ 是 B 格式的音檔訊號向量。因此我們可以表達為以下公式：

$$S = DB$$

[1][2][3]。當我們將揚聲器放在同一平面時，可以將揚聲器發出的音訊視為平面波，而讓矩陣 $LL^T = 1$ ，且 $D = L^T$ 。其中 $L$ 矩陣定義為以下：

$$Y_l = L = \begin{bmatrix} Y_0^0(\varphi_1, \vartheta_1) & Y_0^0(\varphi_2, \vartheta_2) & \dots & Y_0^0(\varphi_{N_l}, \vartheta_{N_l}) \\ Y_{-1}^1(\varphi_1, \vartheta_1) & Y_{-1}^1(\varphi_2, \vartheta_2) & \dots & Y_{-1}^1(\varphi_{N_l}, \vartheta_{N_l}) \\ \vdots & \vdots & \ddots & \vdots \\ Y_m^n(\varphi_1, \vartheta_1) & Y_m^n(\varphi_2, \vartheta_2) & \dots & Y_m^n(\varphi_{N_l}, \vartheta_{N_l}) \end{bmatrix}$$

[1][2]。其中， $n$  是的 B 格式的階，而  $m$  是最高階  $n$  的次數。

而我們也在 MATLAB 利用以上原理設計一階 B 格式轉四個揚聲器的解碼矩陣：

$$D^T = \begin{bmatrix} Y_0^0(-45^\circ, 0^\circ) & Y_0^0(45^\circ, 0^\circ) & Y_0^0(135^\circ, 0^\circ) & Y_0^0(-135^\circ, 0^\circ) \\ Y_{-1}^1(-45^\circ, 0^\circ) & Y_{-1}^1(45^\circ, 0^\circ) & Y_{-1}^1(135^\circ, 0^\circ) & Y_{-1}^1(-135^\circ, 0^\circ) \\ Y_0^1(-45^\circ, 0^\circ) & Y_0^1(45^\circ, 0^\circ) & Y_0^1(135^\circ, 0^\circ) & Y_0^1(-135^\circ, 0^\circ) \\ Y_1^1(-45^\circ, 0^\circ) & Y_1^1(45^\circ, 0^\circ) & Y_1^1(135^\circ, 0^\circ) & Y_1^1(-135^\circ, 0^\circ) \end{bmatrix}$$

我們也設計三階 B 格式轉四個揚聲器的解碼矩陣：

$$D^T = \begin{bmatrix} Y_0^0(-45^\circ, 0^\circ) & Y_0^0(45^\circ, 0^\circ) & Y_0^0(135^\circ, 0^\circ) & Y_0^0(-135^\circ, 0^\circ) \\ Y_{-1}^1(-45^\circ, 0^\circ) & Y_{-1}^1(45^\circ, 0^\circ) & Y_{-1}^1(135^\circ, 0^\circ) & Y_{-1}^1(-135^\circ, 0^\circ) \\ \vdots & \vdots & \vdots & \vdots \\ Y_3^3(-45^\circ, 0^\circ) & Y_3^3(45^\circ, 0^\circ) & Y_3^3(135^\circ, 0^\circ) & Y_3^3(-135^\circ, 0^\circ) \end{bmatrix}$$

[1]。最終，得出4聲道的 WAV 檔案，利用2.3.1.的方式進行播放。

## 五、 研究結果

本專題執行兩次的錄音實驗，並且對兩種麥克風與兩種解碼方式的效果進行比較，以下陳述實驗的方式與過程：

- 本專題第一次實驗，利用一階 Ambisonics 麥克風錄製兩分鐘以上的音檔，以鋼琴、大提琴、小提琴圍繞著麥克風合奏錄音，如圖3-1，並儲存成 B 格式。利用基礎解碼方式解碼 B 格式音檔，並且在播放介面播放四聲道的環繞音訊。也利用解碼矩陣的方式對 B 格式的音檔解碼，並播放在相同的介面上。在個人主觀的聽感上，兩個音檔有音量上的差別，基礎解碼方式的音量較大，而解碼矩陣的方式音量較小，我認為這是由於

參考基礎解碼公式與通識相差 $N_l$ 倍，而造成音量較大的情況。另外，由於鋼琴距離麥克風較遠的距離，兩種解碼方式播放時的鋼琴聲音的方向性較不明顯。相對而言，小提琴與大提琴的聲音有較良好的方向性，而解碼矩陣呈現的方向性又較為顯著。

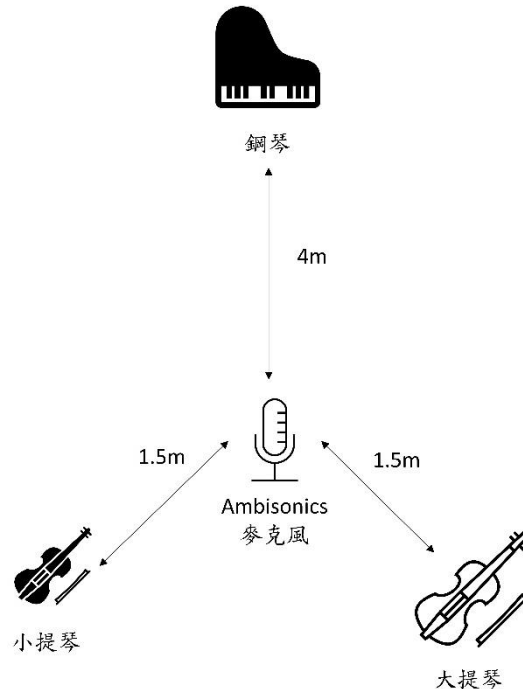


圖3-1 第一次錄音實驗樂器與麥克風配置示意圖  
鋼琴位於麥克風的正前方較遠的位置

- 本專題第二次實驗，利用一階 Ambisonics 麥克風與高階麥克風錄製五分鐘的音檔，以鋼琴和大提琴圍繞著麥克風合奏錄音，其中也嘗試兩種方位配置，如圖3-2與圖3-3，並儲存成 B 格式。相較於初次實驗，調整了鋼琴與麥克風的距離，讓鋼琴的聲音更具方向性。在個人主觀的聽感上，由於對距離進行了調整，整體錄音更佳清晰，並且比較兩種方位的音檔，無論兩種解碼方式，都能清楚聽出相對位置。另外，基礎解碼方式的音量也有較大的情況，而解碼矩陣的方式音量較小。兩種麥克風無明顯的差異。

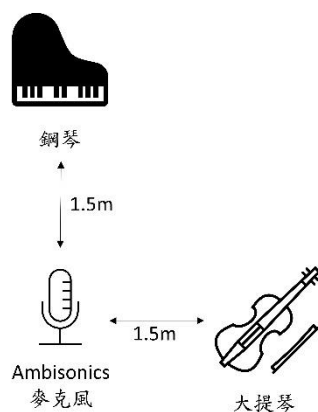


圖3-2 第二次錄音實驗樂器與麥克風配置示意圖1  
鋼琴位於麥克風的正前方，大提琴位於麥克風正右方



圖3-3 第二次錄音實驗樂器與麥克風配置示意圖2  
鋼琴位於麥克風的正前方，大提琴位於麥克風正後方

以上許多為個人的主觀感受與推論，若要更具科學的實驗結果呈現，應該招募受試者，並設計問卷進行統計上數據的研究。

## 六、 總結

本研究探討了從 Ambisonics 麥克風錄製空間音訊，到編碼並儲存為 B 格式，解碼與投放道揚聲器的方式和數學模型，並且成功利用播放介面進行播放。本專題也介紹許多近年來 Ambisonics 的學術研究，像是空間轉換與振幅平移技術，希望未來能在對於 Ambisonics 解碼的研究基礎上，測試更多相關實驗。

## 七、 參考資料

- [1] Roginska, A., & Geluso, P. (Eds.) "Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio," Routledge, 2017
- [2] Kronlachner, Matthias, and Franz Zotter. "Spatial transformations for the enhancement of Ambisonic recordings." Proceedings of the 2nd International Conference on Spatial Audio, Erlangen. 2014.
- [3] Marentakis, Georgios, Franz Zotter, and Matthias Frank. "Vector-base and ambisonic amplitude panning: A comparison using pop, classical, and contemporary spatial music." Acta Acustica united with Acustica 100.5 (2014): 945-955.
- [4] M. Kentgens and P. Jax, "Space Warping Based Dimensionality Reduction of Higher Order Ambisonics Signals," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 131-135, doi: 10.1109/ICASSP.2019.8682250.

## 八、 心得

首先，很榮幸與感謝能在劉奕汶教授的指導下，完成專題實作，過程中老師建議與教導都令我受益良多，更感謝能參加實驗室的討論會議，讓我瞭解除了研究內容外，更多的音訊相關研究。由於，對於耳機的興趣，因此在空間音訊相關的領域研究過程是樂在其中的，並且實作許多系統(像是播放介面原本是用於錄音的錄音介面)需要自行摸索，所以在過程中須自行尋找相關資料。另外，也感謝學長們，能夠在摸索學習時遇到問題，學長們在音訊處理背景知識下提出相關建議或幫助。對於身為生醫學院學士班的我，雖然修習電機系訊號處理課程，但是對音訊處理領域完全陌生，因此專題實作對我十分具有挑戰性，並且在研究時獲得成就感。