

中華民國第 65 屆中小學科學展覽會

作品說明書

高級中等學校組 工程學科(一)

佳作

052307

見光聞聲：以雷射都卜勒震動儀結合深度學習辨
認聲音的方向

學校名稱： 國立臺灣師範大學附屬高級中學

作者： 高二 王品惟 高二 吳濬丞 高二 陳巧安	指導老師： 潘冠錡 蕭煜修
---	-----------------------------

關鍵詞： 都卜勒效應、干涉、深度學習

摘要

本研究在探討如何將「雷射都卜勒震動儀」(Laser Doppler Vibrometer, 以下簡稱 LDV) 所測量出的一維訊號, 進一步轉換出二維資訊, 意即聲源的來向。本研究包含兩個部分: 以物理推導出系統模型, 與利用音檔進行深度學習以訓練AI模組, 並根據不同聲源來向進行分類。經過推導, 我們發現LDV所輸出訊號與待測物表面之關係。同時, 也利用程式模擬待測物受聲波影響後之振動模式。藉此, 我們以模擬還原出了LDV所收到的音檔, 並與實際錄製音檔比較。最終, 我們成功建立了分類準確率將近100%的AI模型, 並改善了過去的聲源定位方法。

壹、前言

一、研究動機

自從錄音技術問世以來, 氣導式麥克風一直是最普遍的聲音接收裝置, 能滿足單聲道錄製的需求; 然而, 為了實現立體聲錄製, 人們採用了麥克風陣列技術——在空間中佈置多個氣導式麥克風, 並藉由它們捕捉聲音的時間差與強度差, 以進行聲源定位。而近年有實驗表明以單個氣導式麥克風即可辨認聲音的方向[1], 其利用環境因子找出屬於每個方向的「聲紋」以達到辨認聲音方向的目的。儘管氣導式麥克風已廣泛應用於語音辨識與環境監測等領域, 但它仍承襲了氣導式麥克風的固有限制: 在真空或低壓環境下無法正常運作。

為突破此一瓶頸, 本研究嘗試將雷射都卜勒震動儀 (Laser Doppler Vibrometer, 以下簡稱 LDV) 引入聲源定位領域。LDV為一種非接觸式精密測量工具, 其原理基於光學干涉與都卜勒效應, 可直接測量物體表面的振動速度與位移, 並將振動訊號轉換成音檔。迄今為止, LDV的應用仍集中於單點的一維振動訊號分析, 鮮少延伸至多維度的聲源定位。

本計畫將結合物理模型推導與人工智慧技術, 探討將LDV所擷取的單點一維振動訊號轉換成二維聲源方向資訊的可行性。我們將利用LDV的非接觸、高靈敏度與不需介質傳播的特性, 整合深度學習以強化數據處理, 並降低環境干擾對定位精度的影響。若能成功實現, 該方法不僅具備遠距離非接觸偵測能力, 也能在真空或低壓環境下正常運作, 未來更有可能應用於太空探測等氣導式麥克風無法達成的艱難任務。

二、研究目的

- (一) 以數學公式分析LDV測量方式
- (二) 推導平面在受到聲波影響而產生的振動方式
- (三) 利用程式模擬平面受聲波影響所產生的振動
- (四) 以卷積神經網路 (Convolutional neural network, 以下簡稱CNN) 建立AI模型以分類不同角度的聲音來源音檔

三、文獻探討

麥克風陣列、精密頭部相關傳遞函數 (HRTF) 是傳統聲源定位常仰賴的方法，然而其這些方法總是帶有部署複雜度或成本過高等問題。因此，論文中[2]作者由人耳的單耳定位能力作為靈感，嘗試以單支氣導式麥克風隨機搭配散射結構辨認聲音的方向。

人耳單耳定位的原理是利用頭部的散射與過去經驗，而論文中，作者以LEGO積木作為散射結構，並在不同角度下撥放白噪音，並透過氣導式麥克風所收到的頻率響應建立「聲紋」的子空間，如下圖1.1所示。此時，只要在撥放白噪音時，將氣導式麥克風所收到的功率頻譜進行子空間投影，即可以最小投影誤差判斷聲源的方位。此方法即可省去多支麥克風或複雜的硬體設備而能達成高準確度的聲源定位。

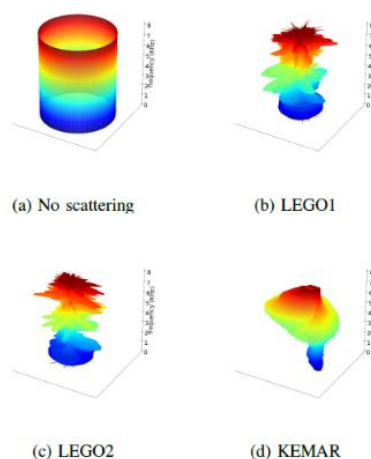


圖1.1：不同散射結構與角度下氣導式麥克風所收到的頻率響應 [3] (出處詳圖片來源1)

我們將參考上述論文對於「聲紋」的提取方法，同時嘗試是否能夠以單點雷射取代論文中的單支氣導式麥克風，透過震動的表面不同角度下的頻譜特徵結合深度學習達到聲源方向定位的效果，並嘗試提升此方法的準確率。

貳、研究設備與器材

壹、研究設備

表3.1 硬體器材(出處詳圖片來源2)

器材名稱	圖片	用途
LDV		(見貳之二)
DT9837A DAQ		將訊號由LDV進行類 比轉數位並傳送到 SpectraPLUS_DT
喇叭		播放音檔， 使物體表面震動。
分貝計		在距離聲源一公尺處 測量喇叭分貝數。
雷射測距儀		用雷射測量距離。
電子量角器		用以測量喇叭方向 和待測物體之夾角。




壓克力板		為本次實驗所使用的待測物體。
反光貼		貼在壓克力板上以利光反射回LDV。
彈簧		用於支撐磁鐵。
磁鐵		為單擺的重物； 其上做有標記，以提供tracker做為參考點。

表3.2 使用軟體 (出處詳圖片來源3)

軟體名稱	軟體照片	用途
Python		模擬聲波在平面上的震動。 (詳見參之二)
Excel		用以分析數據與繪製表格。
Matlab		對實驗數據取對數功率頻譜 (Log Power Spectrum，以下簡稱LPS)。

Tracker		用以提取磁鐵震動速度及位移。
SpectraPLUS_DT		擷取來自LDV的訊號，可以將訊號輸出成為音檔。
Cool Edit Pro 2.1		查看音檔的時頻圖與剪接音檔。
CLIO 12		用以校正喇叭音量與撥放音檔
OptoGUI		用以遠端控制LDV，調整LDV雷射光焦距

二、LDV簡述

雷射都卜勒震動儀（Laser Doppler Vibrometer, LDV）是一種非接觸式的精密測量工具，主要用於檢測物體表面的振動特性。其基本工作原理基於光學干涉與都卜勒效應，能夠測量振動的速度與位移，並由物體的振動方式輸出音檔。

本研究採用之儀器為德國Optomet的雷射都卜勒震動儀（Vector-Series, Optomet, Darmstadt, Germany）。

LDV具有以下特性：

（一）非接觸測量：LDV以雷射光束在不接觸目標物體的情況下測量其振動，可遠距測量獲得物體的聲學資訊。

（二）具高精度與高靈敏度：LDV具有極高的精度，能以微米/秒（ $\mu m/s$ ）為單位測量振動速度，因此適合需要高解析度的應用。

（三）不易受介質干擾：LDV測量以光束傳遞，不需其他介質傳遞訊號，因此不易受到介質可能引起的波反射、折射或吸收問題。

以上幾點特性使LDV在許多工程應用上佔有優勢，可應用於結構震動分析、機械震動測試或本研究欲討論的聲源定位等領域。

參、研究過程及方法

為了達成分辨角度的目的，我們將從LDV內部光路分析、物體振動方式的模擬發端，利用理論逐漸推廣到實驗，並透過程式擷取各個角度的聲紋，也就是頻譜特徵，最後再交由AI學習，以彙整「方向性」這個主題。

一、LDV光路分析

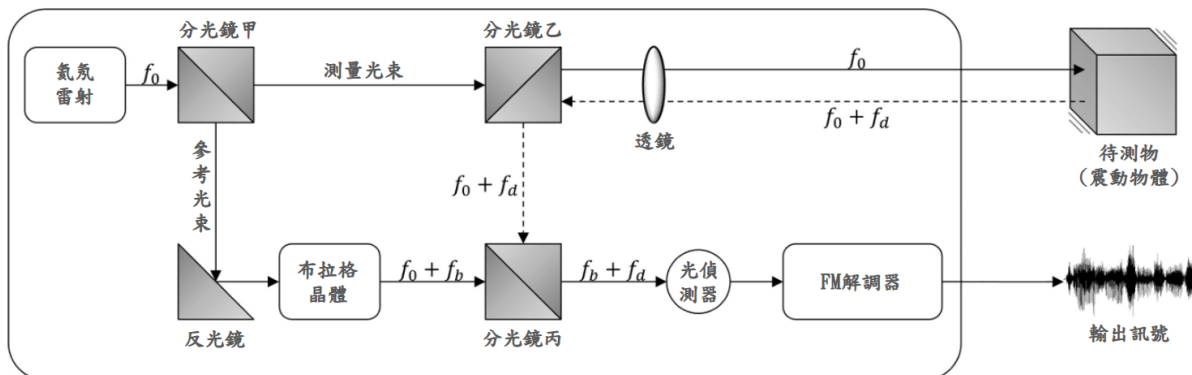


圖3.1 LDV內部光路圖[4] (出處詳圖片來源4)

如上圖3.1所示，氦氖雷射（He-Ne laser）射出的光束（波長 $\lambda_0 \approx 632.8\text{nm}$ ，頻率 $f_0 \approx 4.74 \times 10^{14}\text{Hz}$ ），經由分光鏡甲被分為測量光束（Measurement beam）與參考光束（Reference beam）。其中，測量光束會再經過分光鏡乙以及透鏡，打到待測物上，並產生都卜勒效應[5]，如下式（一）：

$$f_d(t) = \frac{2v(t) \cos \psi}{\lambda_0}$$

其中， t 為時間， $f_d(t)$ 為都卜勒頻移， $v(t)$ 為測量點表面振動速度， ψ 為測量點物體振動速度與測量光束方向之夾角。

產生都卜勒效應後，測量光束會帶有物體表面振動的頻率 $f_d(t)$ ，也就是帶有欲測量的訊息。其電場可以表示成下式（二）：

$$E_m(t) = \text{Re} \left\{ A_m e^{i \left[2\pi (f_0 + f_d(t)) t + \varphi_m \right]} \right\}$$

其中， A_m 為測量光束的振幅， φ_m 為測量光束的相位角。

而參考光束在LDV中經由反光鏡射入布拉格晶體（Bragg cell，又稱聲光調製器AOM,acousto-optic modulator）後，會產生布拉格繞射，將布拉格晶體的振動頻率 f_b 鑲嵌至參考光束中，使其頻率改變為 $f_0 + f_b$ 。其電場 $E_r(t)$ 可以表示成下式（三）：

$$E_r(t) = \text{Re} \left\{ A_r e^{i[2\pi(f_0+f_b)t+\varphi_r]} \right\}$$

其中， A_r 為參考光束的振幅， φ_r 為參考光束的相位角。

接著，測量光束與參考光束在分光鏡丙相遇後，會產生干涉作用。其干涉後電場可表示為下式（四）：

$$E_i(t) = E_m(t) + E_r(t) = \text{Re} \left\{ A_r e^{i[2\pi(f_0+f_b)t+\varphi_r]} \right\} + \text{Re} \left\{ A_m e^{i[2\pi(f_0+f_d(t))t+\varphi_m]} \right\}$$

其中， $E_i(t)$ 為干涉後的電場。

再將共同的雷射光頻率項 $e^{i2\pi f_0 t}$ 提出並以歐拉公式化簡並取其實部，可得式（五）：

$$E_i(t) = (A_r \cos \theta_b + A_m \cos \theta_d) \cos \theta_0 - (A_r \sin \theta_b + A_m \sin \theta_d) \sin \theta_0$$

其中 $\theta_0 = 2\pi f_0 t$ 、 $\theta_d = 2\pi f_d t + \varphi_m$ 、 $\theta_b = 2\pi f_b t + \varphi_r$

最後利用三角的疊合得到化簡過後的 $E_i(t)$ ，如式（六）：

$$E_i(t) = \cos(\theta_0 + \psi) \sqrt{A_r^2 + A_m^2 + 2A_r A_m \cos(\theta_b - \theta_d)}$$

我們可以得出最後干涉光的強度 I_i 如下式（七）：

$$I_i = \frac{\varepsilon_0 c}{2} E_i^2(t) = \frac{\varepsilon_0 c}{2} \cos^2(\theta_0 + \psi) [A_r^2 + A_m^2 + 2A_r A_m \cos(\theta_b - \theta_d)]$$

其中， ε_0 為真空中的電容率， c 為光速。

考慮到光偵測器（Photo Detector）所測得的能量強度主要由包絡成分決定，因此可忽略高頻快速振盪的部分 $\cos^2(\theta_0 + \psi)$ ，又已知理想的光偵測器所輸出之電壓大小與其收到的光強度 I_m 成正比，因此光偵測器所輸出的電壓大小 $V_p(t)$ 如下式（八）：

$$V_p(t) \propto I_m = \frac{\varepsilon_0 c}{2} [A_r^2 + A_m^2 + 2A_r A_m \cos(\theta_b - \theta_d)]$$

已知頻率調變（FM, frequency modulation）的原理如下，考慮有一訊號 $m(t)$ ，則經頻率調變後所輸出的訊號 $s(t)$ 如下式（九）：

$$s(t) = A \cos \left[2\pi f_c t + 2\pi \Delta f \int_0^t m(\tau) d\tau \right]$$

其中， A 為載波的振幅， f_c 為中心頻率， Δf 為頻偏。

而頻率解調變就是頻率調變的逆運算，其運算方式如下式（十）：

$$f_s(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt}$$

其中， $\varphi(t)$ 為輸入訊號的角位移， $f_s(t)$ 為輸入訊號之瞬時頻率。

最終得到的訊號 $m(t)$ 即為下式（十一）：

$$m(t) = \frac{f_s(t) - f_c(t)}{\Delta f} \propto f_s(t) - f_c(t)$$

在光偵測器接收到訊號後，會將訊號傳遞給頻率解調器（FM demodulator），而最終所輸出之電壓 $V_f(t)$ 正比於 $m(t)$ 如下式（十二）：

$$V_f(t) \propto m(t) \propto f_s(t) - f_c(t) = [f_b - f_d(t)] - f_b \propto f_d(t)$$

又已知在測量點物體振動速度與測量光束方向之夾角 φ 與雷射光波長 λ_0 皆不變的情況下， $f_d(t)$ 與測量點表面震動速度 $v(t)$ 成正比，因此最終的電壓 $V_f(t)$ 可表達如下式（十三）：

$$V_f(t) \propto f_d(t) \propto v(t)$$

最後，此訊號再傳送至DT9837A DAQ進行類比轉數位訊號（ADC, Analog-to-digital Conversion），以傳送至電腦上，得到音檔。

二、模擬

（一）原理

1. 定義座標系以及變數

如下圖，訂定LDV測量點為原點，壓克力板為 $x-y$ 平面，且喇叭之座標為 $(x_s, 0, z_s)$ ，喇叭與測量點之距離為 L ，其撥放音檔之頻率為 f 。定義夾角 θ 為 $\tan^{-1}\left(\frac{x_s}{z_s}\right)$ ，即壓克力板法線與喇叭方向之夾角。定義壓克力板在受空氣中聲波影響而振動時，其 z 方向上的位移為 $\zeta(x, y, t)$ 。

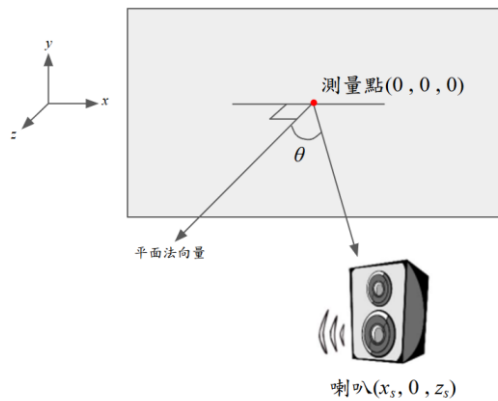


圖3.2 模擬所定義的坐標系(出處詳圖片來源5)

2.平面震動

(1) 球面波

將喇叭視作點波源，其播放音檔時發出球面波。假設聲壓 $p(x, y, t)$ 與到聲源的距離 R 的一次方成反比，且喇叭放出之音檔頻率固定，因此可設聲壓 p 如下式（十四）：

$$p(x, y, t) = \frac{A_p}{R} \times \cos(kR - 2\pi ft)$$

其中， A_p 為聲壓振幅，波數 $k = \frac{2\pi}{\lambda}$ 。

並且在距離喇叭 L 處喇叭所造成的聲壓級 SPL 滿足下式（十五）：

$$SPL = 20 \log_{10} \left(\frac{A_p}{L \cdot p_0} \right)$$

其中，基準聲壓 $p_0 = 2 \times 10^{-5} (Pa)$ 。

校正時，距離一公尺處之聲壓級為75dB，因此 A_p 滿足下式（十六）：

$$A_p = 1 \times 2 \times 10^{-5} \times 10^{\frac{75}{20}} \approx 0.11 (N/m)$$

且在壓克力板上，每個點與喇叭的距離 $R(x, y)$ 可表示為下式（十七）：

$$R(x, y) = \sqrt{(x_s - x)^2 + y^2 + z_s^2}$$

(2) 表面張力

設壓克力板上的水平表面張力 T 為定值，考慮壓克力板上的一個質點，座標為 $(x, y, 0)$ ，則在其上的垂直表面張力為 $T \nabla \zeta(x, y, t)$ 。接著考慮一個小平面，其長與寬分別為 Δx 與 Δy ，則在此小平面上 x 、 y 方向所受的表面張力合力大小 $T_x(x, y, t)$ 與 $T_y(x, y, t)$ 分別滿足下式（十八）與式（十九）：

$$T_x(x, y, t) = T \Delta x \frac{\partial \zeta(x, y, t)}{\partial x}$$
$$T_y(x, y, t) = T \Delta y \frac{\partial \zeta(x, y, t)}{\partial y}$$

因此，在此小平面所受單位面積的力 $f(x, y, t)$ 如下式（二十）：

$$f(x, y, t) = T \nabla^2 \zeta(x, y, t)$$

(3) 阻尼項

考慮到傳遞時，小平面同時會受到空氣所造成的阻力 $f_{\text{阻}}(x, y, t)$ ，且已知在低速下，阻力與速度的一次方成正比，方向與速度相反，因此設 $f_{\text{阻}}(x, y, t)$ 滿足下式（二十一）：

$$f_{\text{阻}}(x, y, t) = -\gamma \frac{\partial \zeta}{\partial t}$$

其中， γ 為阻尼係數。

(4) 牛頓第二運動定律

根據牛頓第二運動定律，壓克力板上座標為 $(x, y, 0)$ 的每個質點之運動皆應滿足下式（二十二）：

$$\begin{aligned} \mu \frac{\partial^2 \zeta(x, y, t)}{\partial t^2} &= f(x, y, t) + p(x, y, t) \times \cos(\alpha) + f_{\text{阻}}(x, y, t) \\ &= T \times \nabla^2 \zeta(x, y, t) + \frac{A_p}{R} \times \cos(kR - 2\pi ft) \times \cos(\alpha) - \gamma \frac{\partial \zeta}{\partial t} \end{aligned}$$

其中， α 為法向量與 $(x, y, 0)$ 至喇叭向量之夾角。

因此 α 滿足下式（二十三）：

$$\cos(\alpha) = \frac{z_s}{R(x, y)}$$

最終，可求得壓克力板上的每個質點皆滿足以下微分方程式式（二十四）：

$$\begin{aligned} \mu \frac{\partial^2 \zeta(x, y, t)}{\partial t^2} &= T \times \nabla^2 \zeta(x, y, t) + \frac{A_p z_s}{R^2} \times \cos(kR - 2\pi ft) - \gamma \frac{\partial \zeta}{\partial t} \\ &= T \times \nabla^2 \zeta(x, y, t) + \frac{A_p z_s \times \cos(k \sqrt{(x_s - x)^2 + y^2 + z_s^2} - 2\pi ft)}{(x_s - x)^2 + y^2 + z_s^2} - \gamma \frac{\partial \zeta}{\partial t} \end{aligned}$$

3.邊界條件：我們假設其邊界為自由端，可自由震動。

（二）程式架構

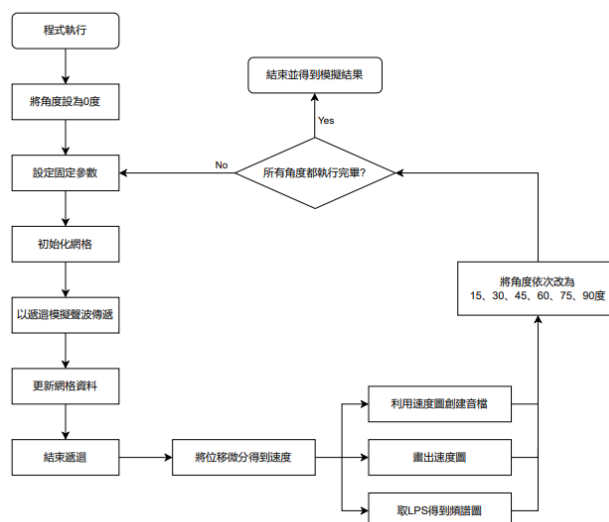


圖3.3 模擬程式架構圖 (出處詳圖片來源6)

如上圖3.3，我們使用python以波動方程的迴圈模擬出 $v(t)$ 圖，經過大迴圈處理即可得出中心點位置函數隨時間變化序列，序列經過處理後，可以獲得中心點位移與速度隨時間的變化。並且根據參之二的推論，所得音檔應與測量點的振動速度成正比，因此我們將 $v(t)$ 圖經過校正後直接轉換為音檔的波形圖，接著再將 $v(t)$ 圖進行LPS處理即可得到模擬的數據。

（三）實驗數據設置

張力 T 設為 $100000(N/m)$ ，阻尼係數 γ 設為 0.15 ，面密度 μ 設為 $4.76(kg/m^2)$ ，模擬平面大小設為 $0.13m \times 0.18m$ 。

三、實驗

（一）實驗一

1. 實驗目的

驗證參之二的推導結果，即LDV輸出的訊號正比於測量點上的速度 $v(t)$ 。

2. 實驗裝置

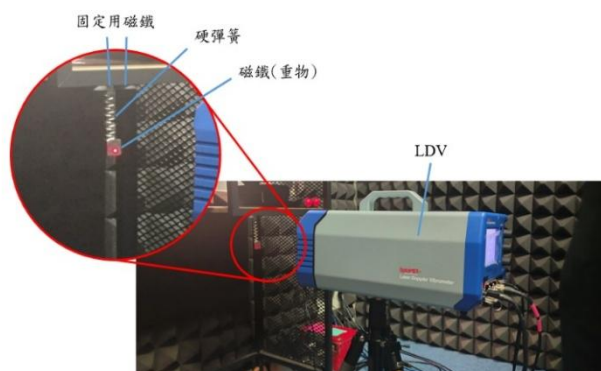


圖3.4 實驗一裝置圖(出處詳圖片來源7)

3.實驗步驟

- (1) 如上圖3.4架設裝置，將LDV對準彈簧下的磁鐵並完成校正。
- (2) 固定錄影設備，並確保磁鐵上的標記清晰可見。
- (3) 開啟錄影與SpectraPLUS_DT的錄音。
- (4) 輕撥彈簧下的磁鐵，以使其開始振盪。
- (5) 重複步驟(1)~(4)三次。
- (6) 以tracker分析磁鐵振盪的速度 $v(t)$ ，並以Cool Edit Pro 2.1開啟LDV錄製的音檔，再取開始振盪後相同秒數進行比較。

(二) 實驗二

1.實驗目的

測量不同壓克力板法線與喇叭方向之夾角 θ 與LDV測得音檔之關係。

2.實驗裝置

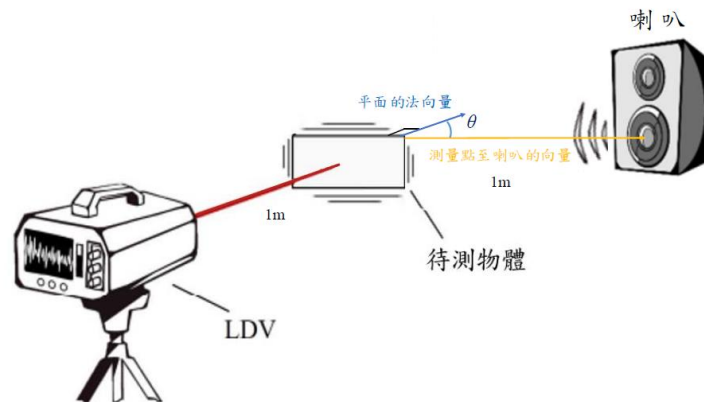


圖3.5 實驗二實驗裝置圖 (出處詳圖片來源8)

3.實驗步驟

- (1) 如上圖3.5架設裝置，依序將角度 θ 設為 0° 、 30° 、 45° 、 60° 、 90° 。
- (2) 將LDV與喇叭接至電腦，並在電腦上開啟SpectraPLUS_DT、Optogui與 CLIO。
- (3) 在距離喇叭1m處（利用雷射測距儀測量）以分貝計測量喇叭播放1kHz音檔時的分貝數，並調整喇叭播音大小，使分貝計測得75dB。
- (4) 以OptoGUI校正LDV雷射焦距。
- (5) 以CLIO播放1kHz的sine波音檔，並以SpectraPLUS_DT錄音並存檔。
- (6) 重複步驟(1)~(5)三次。

四、AI模型訓練

（一）概述

本研究接著導入深度學習的技術來幫助我們判斷聲源的方向。我們利用卷積神經網路，訓練了兩組AI模型，第一組將不同聲源方向（左、右）的音訊進行分類，第二組將同個象限內不同聲源角度（ 0° 、 15° 、 30° 、 45° 、 60° 、 75° 、 90° ）的音訊進行分類。

（二）第一組程式架構

1.實驗參數設定：

採樣率16000Hz、幀長48 點、重疊數24 點、FFT大小48 點。

2.資料與數據分配：

- (1) 預先準備好聲源方向為左邊與右邊的兩組音檔，並存入指定資料夾。
- (2) 將兩類別（left、right）資料從指定資料夾讀取。
- (3) 將兩個類別的音檔依據 4：1 之比例照標號取音檔分為訓練集與測試集。
- (4) 利用短時傅立葉轉換（STFT）將音訊轉換為頻譜圖，再取LPS作為最終特徵表示。

3.資料預處理：

- (1) 對每筆音檔進行數據特徵標準化（Standardization）。
- (2) 依照指定樣本數和時間長度將檔案擷取，確保每筆資料時間長度均一致。
- (3) 利用STFT將調整後的音訊轉換為頻譜圖。
- (4) 取對數功率頻譜LPS作為最終特徵表示。

4.頻譜圖生成與儲存：

利用 librosa 與 matplotlib 將每筆 LPS 數據繪製成圖像。

5.卷積神經網絡（CNN）模型建構與訓練：

本系統採用 Keras Functional API 設計 CNN 模型，主要結構包含：

- (1) 輸入層：根據預處理後頻譜圖的尺寸建立四維資料張量。
- (2) 多層卷積與池化：以 Conv2D、BatchNormalization 與 MaxPooling2D 層，逐層提取局部特徵。
- (3) 全域性池化層：採用 GlobalAveragePooling2D 將卷積層輸出降維。
- (4) 輸出層：以 Dense 層配合 softmax 使用，將結果映射到兩個類別。

模型編譯時使用 Adam 優化器及 categorical_crossentropy 損失函數，設定訓練次數（epochs）為 10，batch_size 為 12。訓練完成後，再以測試集進行評估，並將模型儲存以備後續應用。

6.中間層輸出抽取與 t-SNE 可視化

為了瞭解最終模型學習到的特徵，本實驗利用中間層輸出抽取的方式，分別建立各卷積層之特徵模型。

- (1) 將所得特徵向量經過標準化，並利用 t-SNE 將多維特徵降至二維，並以散佈圖呈現。
- (2) 圖中各類別以不同顏色標示，以利觀察各類之分佈情形與相互區隔效果。

7.預測結果與效能評估

- (1) 訓練結束後，模型針對測試集進行預測，逐筆比對真實類別與預測結果，並列印出各檔案之分類狀況。
- (2) 程式印出訓練集與測試集樣本數、資料類別及檔案數量，確認資料流程正確。
- (3) 根據測試集所得之準確率（Accuracy）與損失值（Loss），對模型效能進行評估，達到理論預期後再進行模型存檔。

（三）第二組程式架構

1.實驗參數設定：

採樣率16000Hz、幀長1024 點、重疊數512 點、FFT 大小1024 點。

2.資料與數據分配：

- (1) 將實驗二測得之音檔每6秒擷取為一個檔案，存進指定資料夾中。
- (2) 各類別（0°、15°、30°、45°、60°、75°、90°）資料從資料夾讀取。
- (3) 將指定資料夾中讀取之音檔取其正中間的4秒。
- (4) 依據 4：1 之比例隨機分為訓練集與測試集。

3.資料預處理：同第一組程式設定。

4.頻譜圖生成與儲存：同第一組程式設定。

5.卷積神經網絡（CNN）模型建構與訓練：同第一組程式設定。

6.中間層輸出抽取與 t-SNE 可視化：同第一組程式設定。

7.預測結果與效能評估：同第一組程式設定。

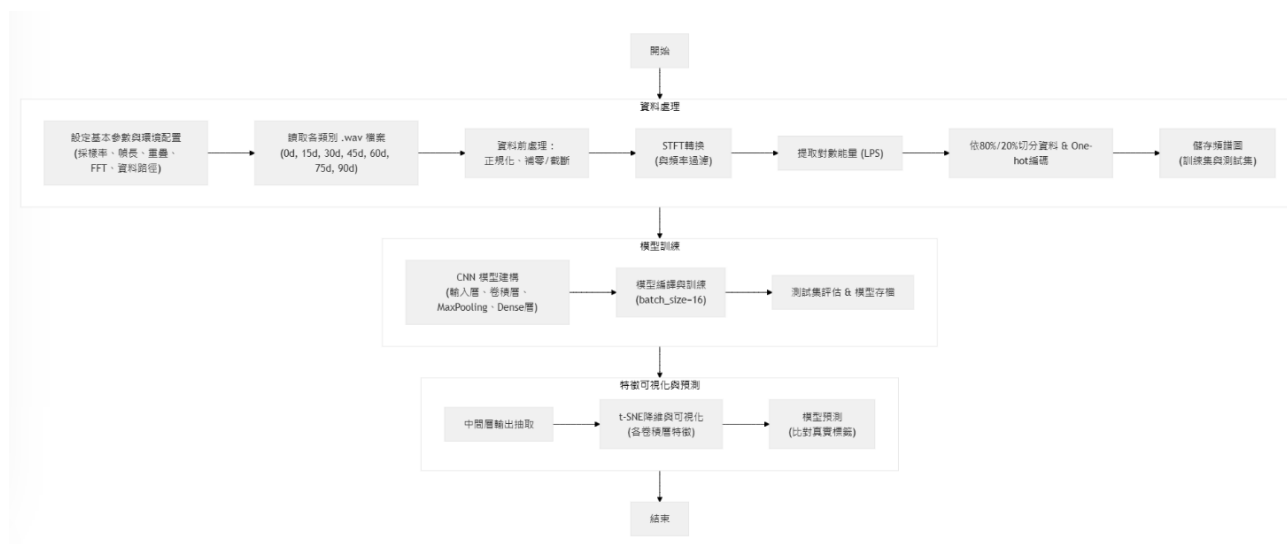


圖3.6 AI模型程式架構(出處詳圖片來源9)

以上即為本次實驗之程式說明與運作流程，各部分內容均依據實際程式碼實作，並結合視覺化及效能評估結果。整個報告內容提供了從資料預處理、特徵提取、模型訓練到中間層可視化等完整流程，為後續研究或調整提供參考。

肆、研究結果

一、實驗一結果

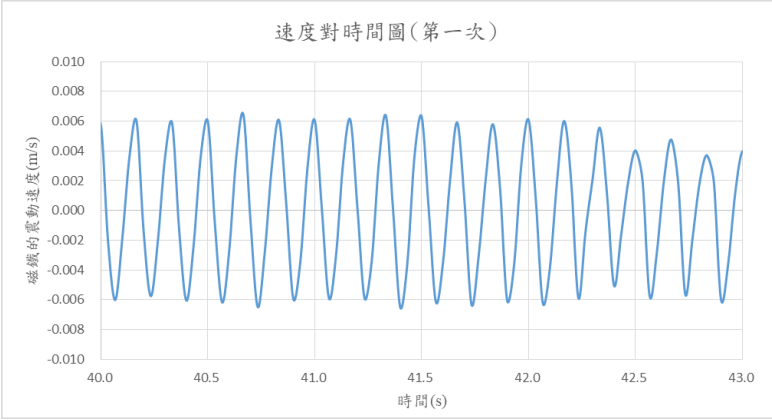
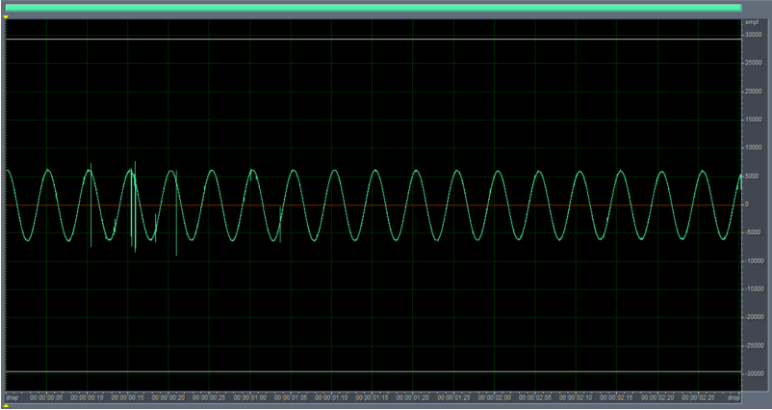
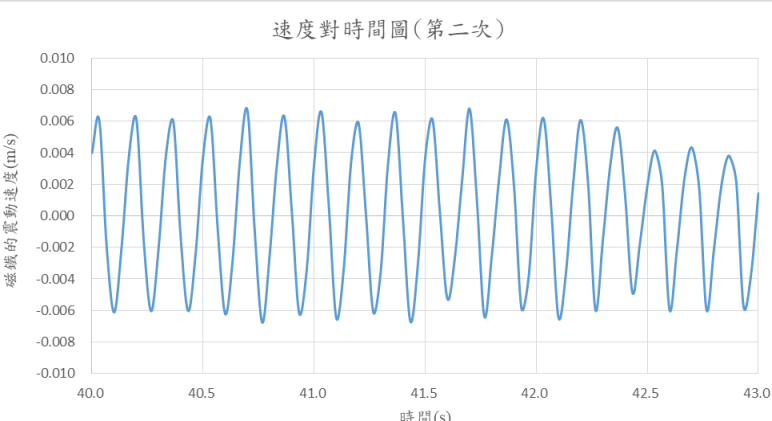
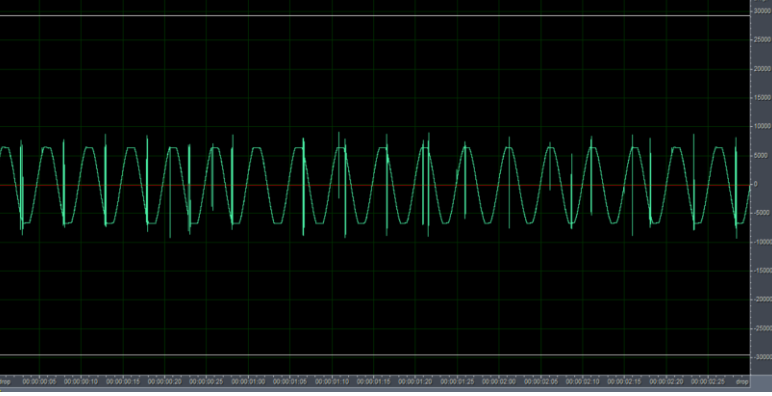
(一) tracker

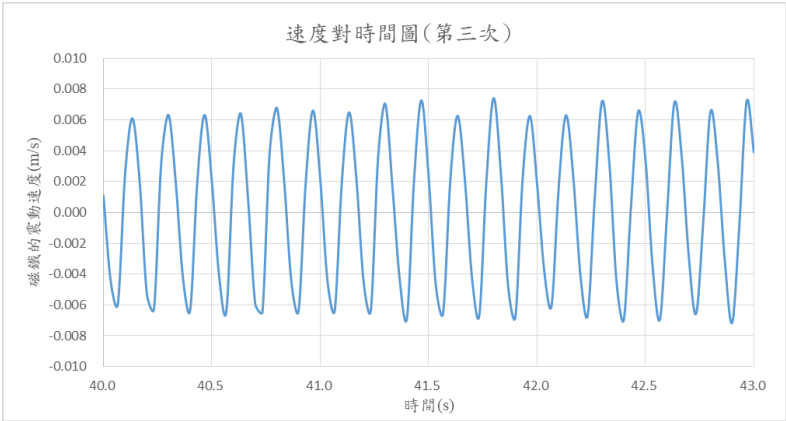
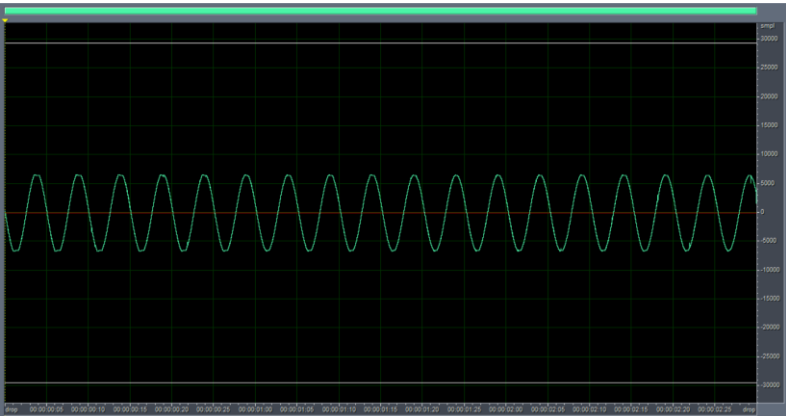
以tracker開啟錄製的影片，並以其分析磁鐵上參考點的速度 $v(t)$ ，最後將資料匯入到Excel再繪圖。

(二) 實驗結果對照圖

為了取時間相同的數據，我們取開始震動後的40-43秒的圖形進行比對（系統穩定後）。

表4.1 實驗一結果(出處詳圖片來源10)

第一次	磁鐵 $v(t)$ 圖	
	LDV 測得音檔	
第二次	磁鐵 $v(t)$ 圖	
	LDV 測得音檔	

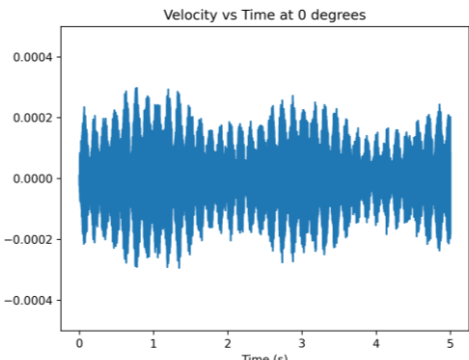
第三次	磁鐵 $v(t)$ 圖	
	LDV 測得音檔	

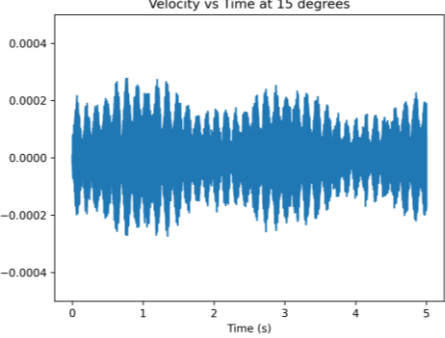
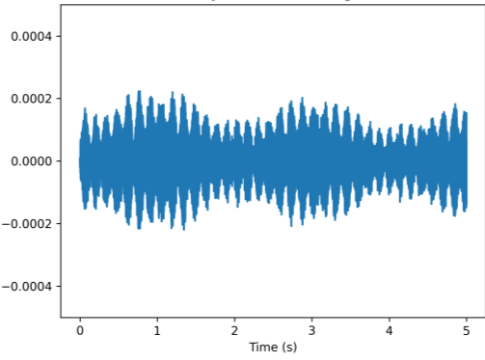
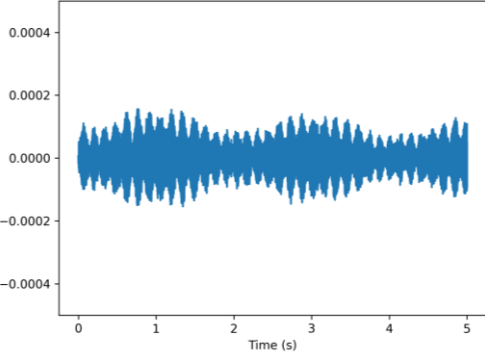
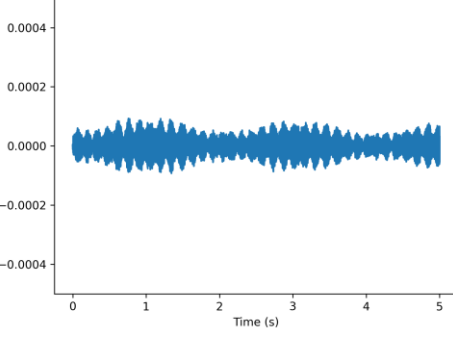
對照後可以發現LDV測得之音檔與磁鐵的 $v(t)$ 圖非常接近，因此可以支持我們推出的式（十三） $V_f(t) \propto v(t)$ 的結果。

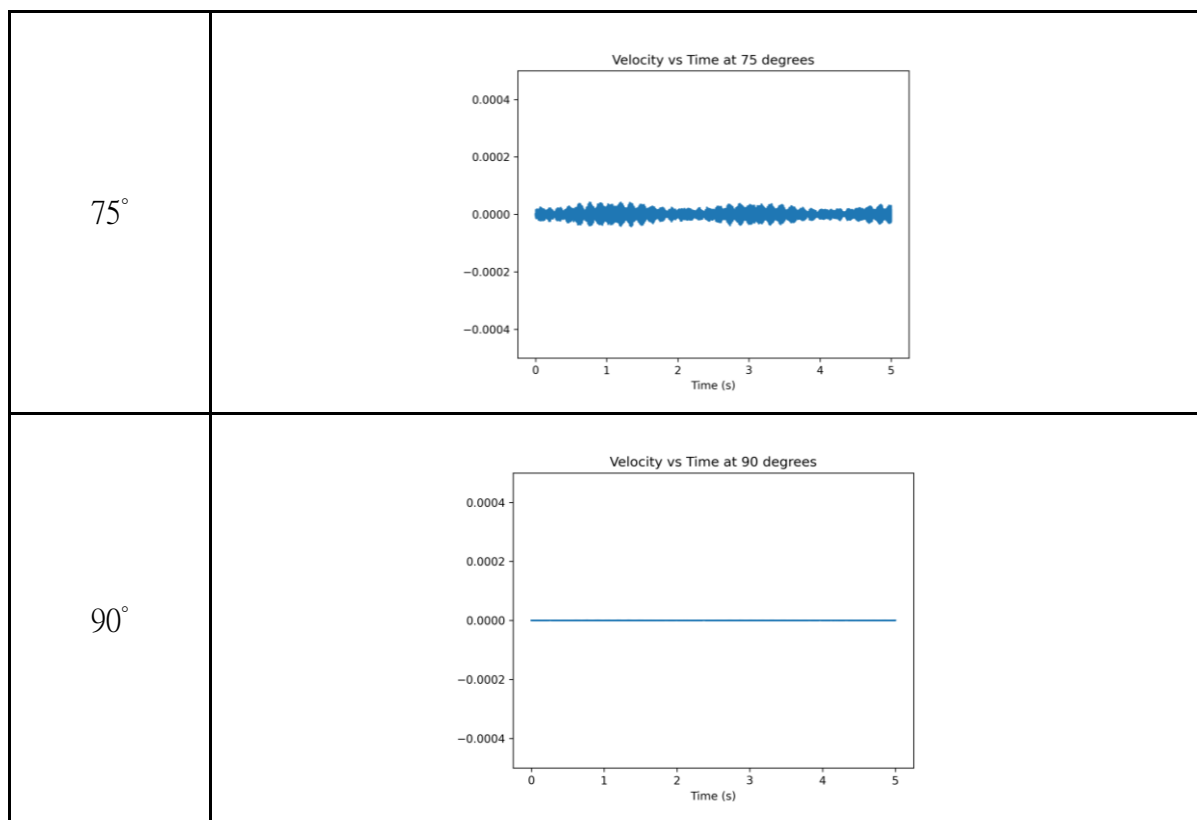
二、模擬結果

（一）模擬出的 $v(t)$ 圖

表4.2 以python程式模擬所得到的測量點速度圖(出處詳圖片來源11)

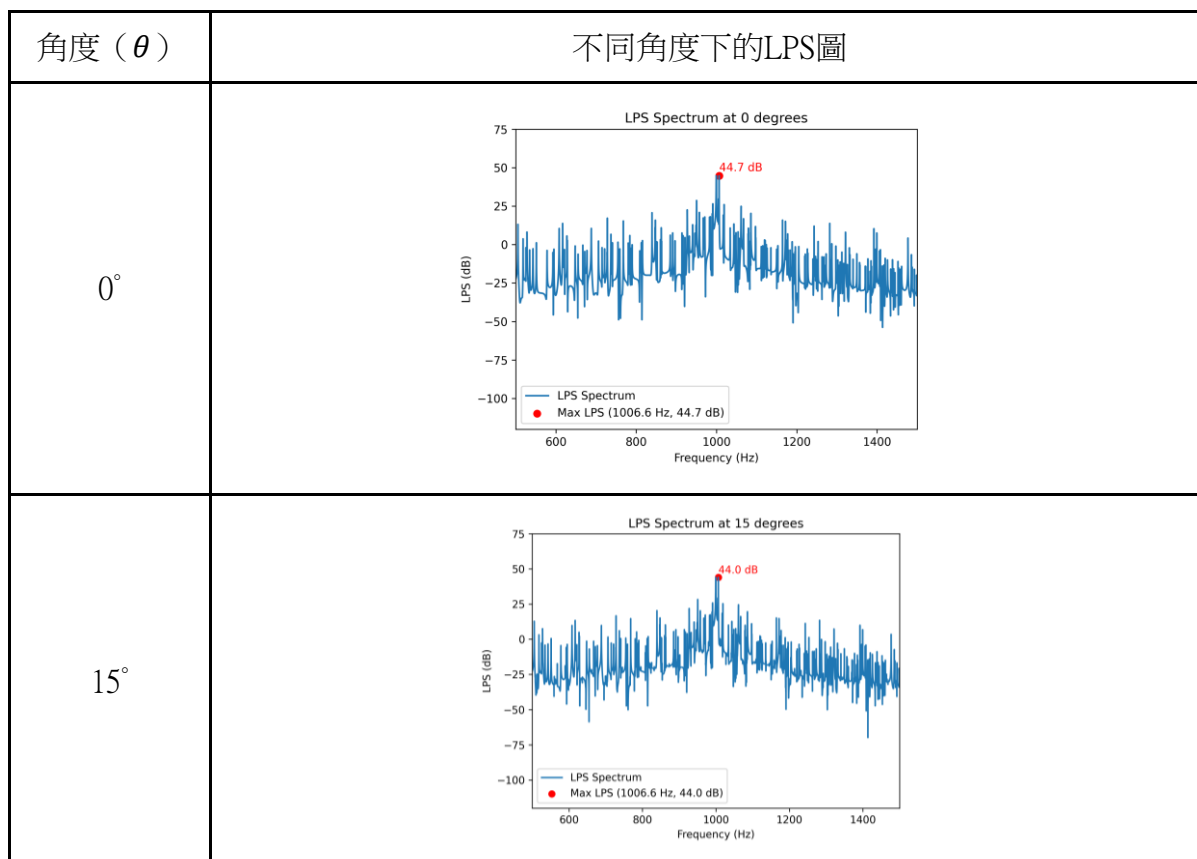
角度 (θ)	不同角度下的 $v(t)$ 圖
0°	

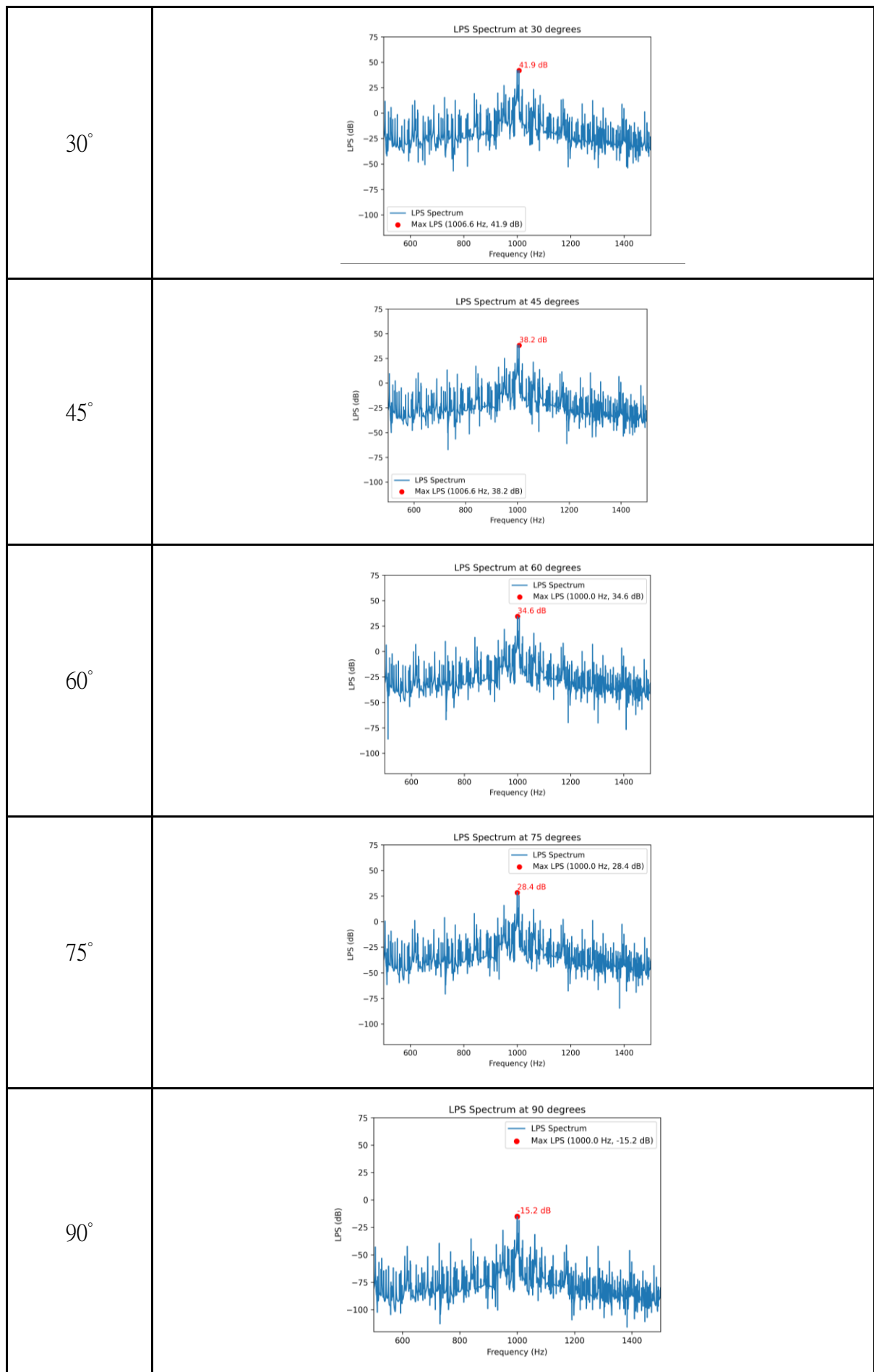
15°	 <p>Velocity vs Time at 15 degrees</p> <p>The plot shows a highly oscillatory signal with a peak-to-peak amplitude of approximately 0.0003, centered around 0.0000, over a 5-second interval.</p>
30°	 <p>Velocity vs Time at 30 degrees</p> <p>The plot shows a highly oscillatory signal with a peak-to-peak amplitude of approximately 0.0002, centered around 0.0000, over a 5-second interval.</p>
45°	 <p>Velocity vs Time at 45 degrees</p> <p>The plot shows a highly oscillatory signal with a peak-to-peak amplitude of approximately 0.0001, centered around 0.0000, over a 5-second interval.</p>
60°	 <p>Velocity vs Time at 60 degrees</p> <p>The plot shows a highly oscillatory signal with a peak-to-peak amplitude of approximately 0.00005, centered around 0.0000, over a 5-second interval.</p>



(二) 模擬出的 $v(t)$ 圖經LPS處理的圖表

表4.3 模擬後的LPS結果(出處詳圖片來源12)



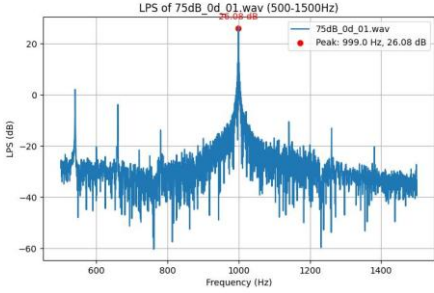
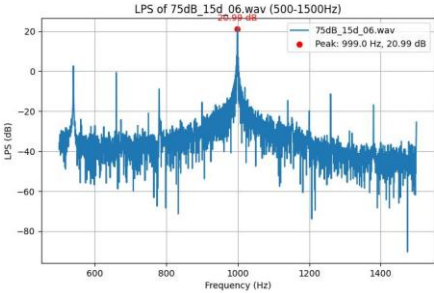
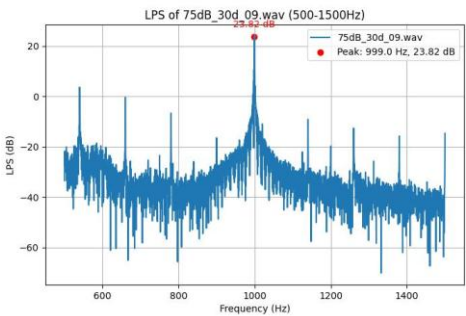
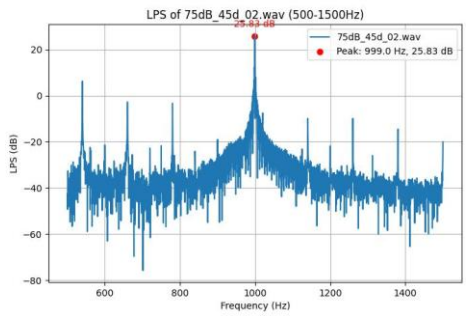


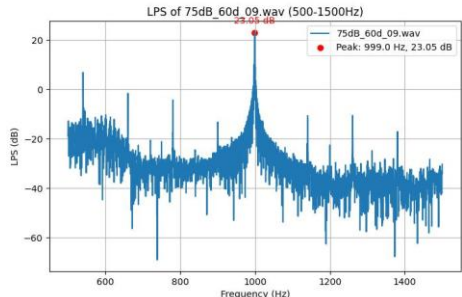
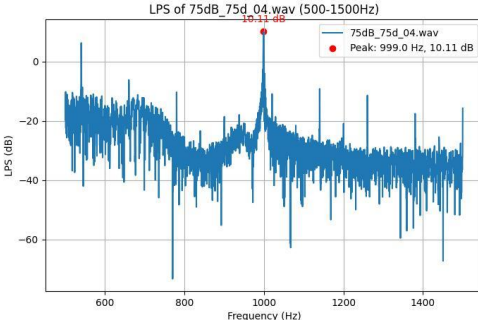
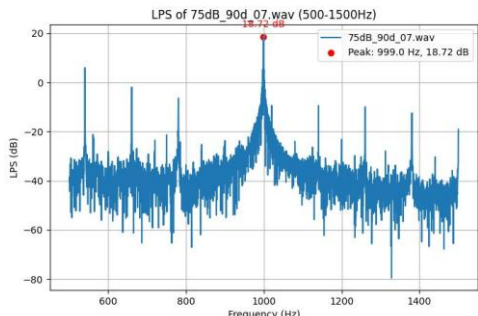
三、實驗二結果

(一) LPS處理

為了方便觀察不同角度時音檔的特性，我們同樣將每個音檔進行LPS處理。

表4.4 實驗音檔LPS的結果 (出處詳圖片來源13)

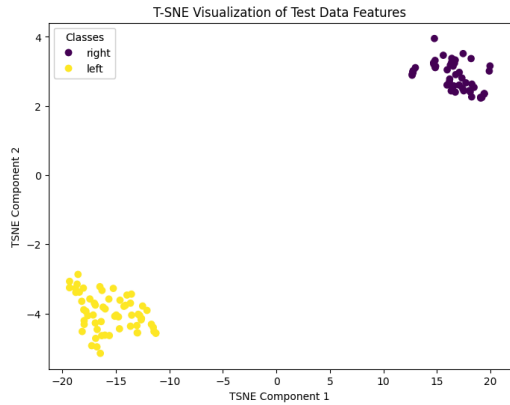
角度 (θ)	不同角度下的LPS圖
0°	
15°	
30°	
45°	

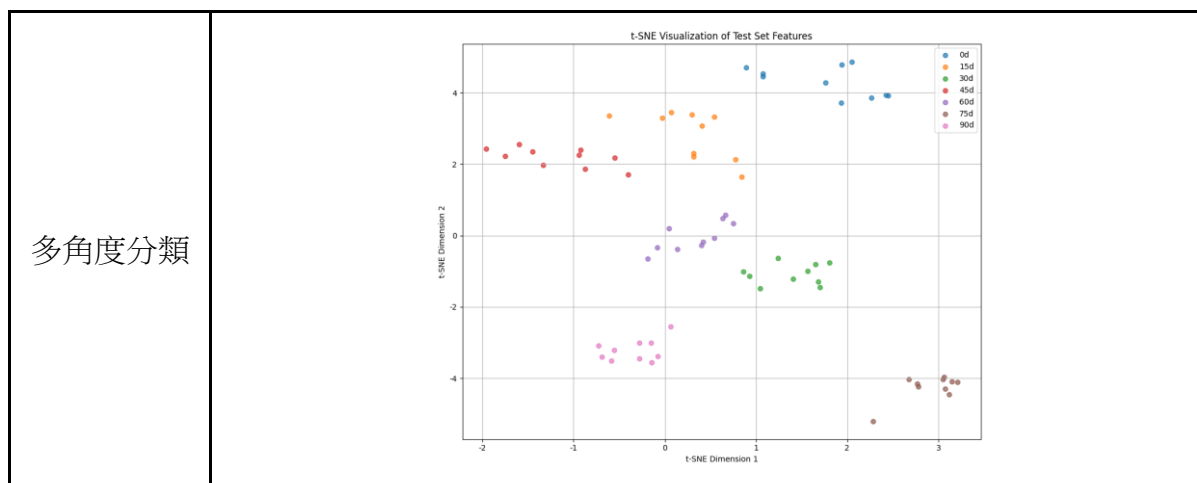
60°	
75°	
90°	

四、AI模型訓練結果

(一) CNN最終的t-SNE分析圖

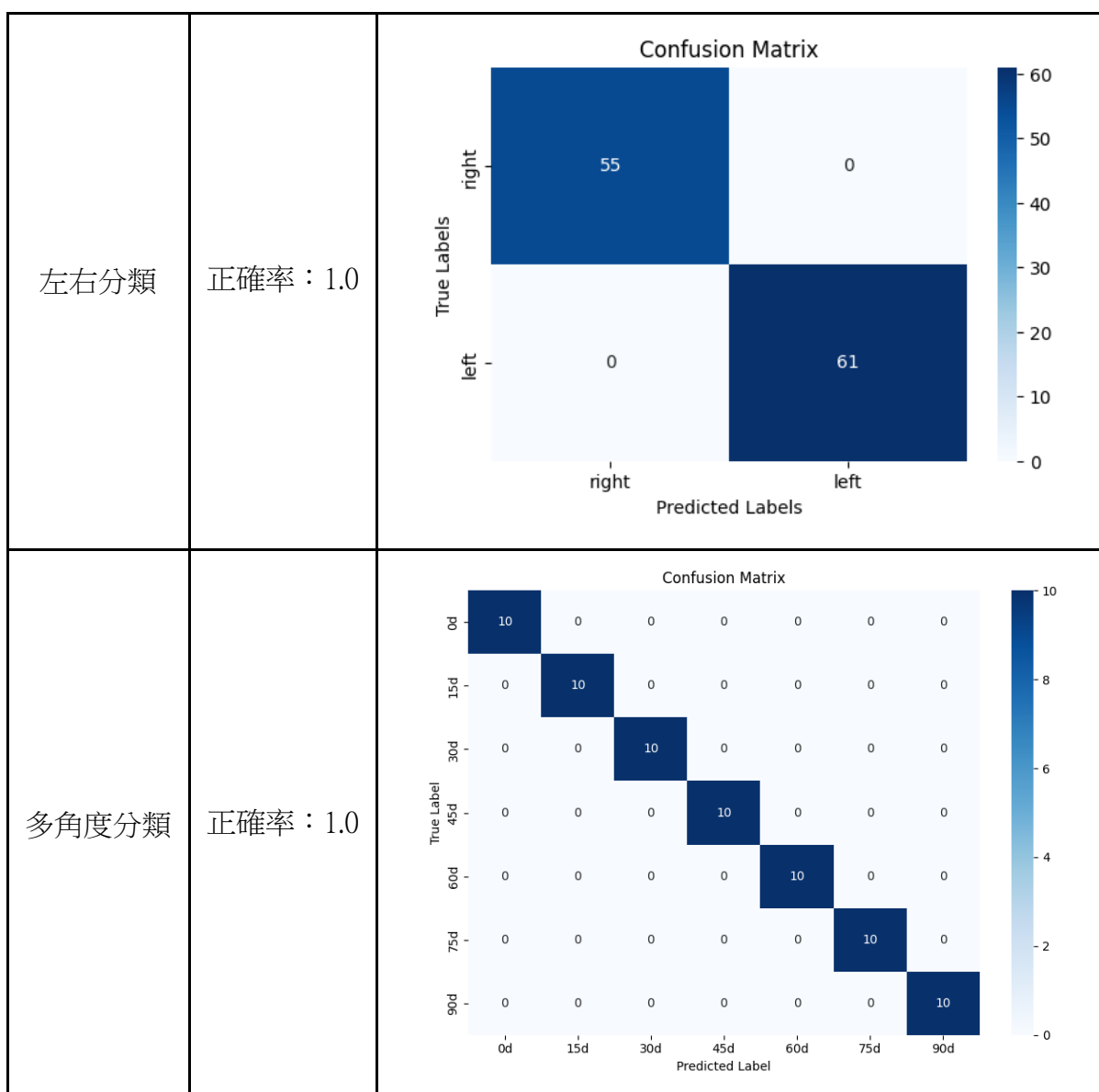
表4.5 測試集分類後的分類結果 (出處詳圖片來源14)

左右分類	
------	--



(二) 混淆矩陣

表4.6 測試集分類後的混淆矩陣 (出處詳圖片來源15)



伍、討論

一、實驗一討論

（一）結果

根據實驗結果，LDV測得的音檔波形與Tracker分析出的磁鐵速度 $v(t)$ 圖非常接近，這支持了式（十三） $V_f(t) \propto v(t)$ 的正確性。可觀察到兩者波形接近，皆可看出類似sine波，接近簡諧的週期運動。

（二）誤差分析

1.LDV的輸出最大值：LDV輸出有最大值的限制，此限制來自類比轉數位訊號時預先設置好的最大與最小值。實驗中由40秒開始取波形的用意就是為了避免超出此最大值。

2.錄影時的解析度誤差：雖然就肉眼觀察，磁鐵上做的記號非常小，但錄影時仍然占了好幾個像素，並且為了避免超過LDV的輸出最大值，我們將磁鐵的晃動幅度控制得非常小，因此這個記號大小的誤差相對而言也跟著放大，在Tracker分析出的 $v(t)$ 圖看起來也就不是完美的sine波。

3.實驗時的噪音干擾：雖然實驗在隔音室中進行，但磁鐵在晃動時難免會產生微小的噪音，因此就有可能出現表4.1中第1、2次實驗中的雜訊。

二、模擬音檔與實際錄製比較

（一）模擬結果與實際錄音的相似度

比較肆之二之（二）與肆之三之（一）模擬與實際的LPS分析圖，可以發現模擬結果與實際錄音結果接近：

- 1.皆在1000Hz出現分貝數最大值
- 2.皆有出現倍頻現象，也就是在1000Hz以外的頻率出現分貝數極大值
- 3.1000Hz的分貝數皆隨著角度增加而減少
- 4.倍頻的出現次數皆隨著角度增加而增加

（二）模擬與實際錄製音檔之LPS與角度之關係比較

1. 模擬音檔

模型中，我們假設壓克力板上的小平面受的力與小平面到喇叭方向之向量與其上法向量之夾角餘弦值 $\cos\alpha$ 正比（詳見式（二十二）），因此我們欲比較模擬音檔之LPS圖分貝數最大值與 $\cos\alpha$ 之關係，並與實際錄製之音檔做比較。

已知在測量點上， α 與 θ 相同，因此以下我們取 θ 做為 α 。

表5.1 模擬音檔1kHz分貝數與角度（餘弦值）之關係

θ	0°	15°	30°	45°	60°	75°
$\cos\theta$	1.00	0.97	0.87	0.71	0.50	0.26
分貝數 (dB)	44.2	43.5	41.4	38.2	34.6	28.4

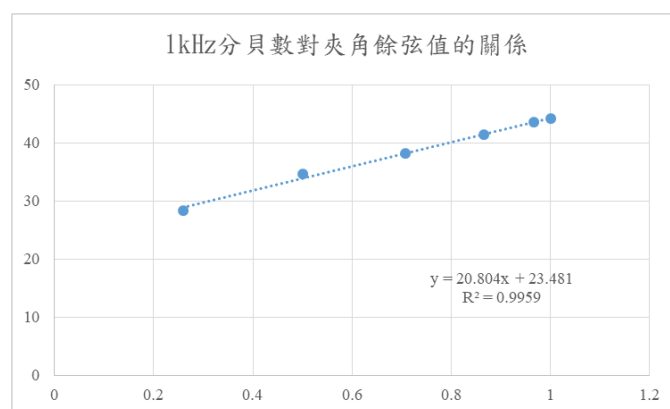


圖5.2 模擬音檔LPS圖中1kHz之分貝數對夾角餘弦值之關係(出處詳圖片來源16)

2. 實際錄製音檔

表5.3 實際錄製音檔LPS圖中，1kHz之分貝數對夾角（餘弦值）之關係

θ	0°	15°	30°	45°	60°	75°
$\cos\theta$	1.00	0.97	0.87	0.71	0.50	0.26
分貝數 (dB)	34.42	30.48	28.91	30.55	30.25	25.1

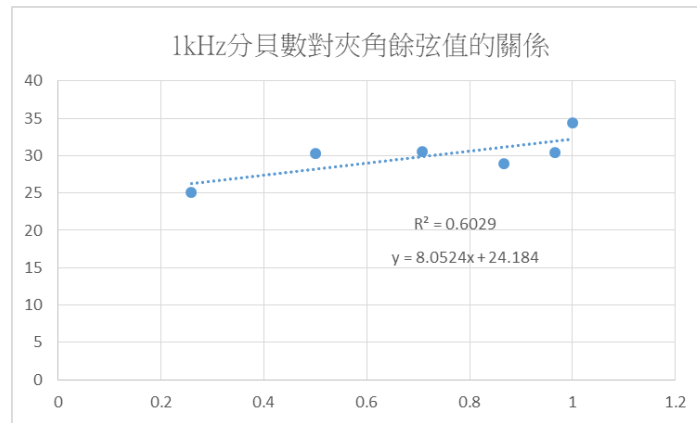


圖5.4 實際錄製音檔LPS圖中1kHz之分貝數對夾角餘弦值之關係(出處詳圖片來源17)

實際測得音檔1kHz處的分貝數對夾角餘弦值 $\cos\alpha$ 之相關係數約等於0.77，為高度正相關，因此支持我們的模擬結果。

(三) 模擬結果與實際錄製音檔之拍音現象

在上面的結果中，我們可以發現模擬音檔的波形圖有出現類似拍音的圖形，如下圖：

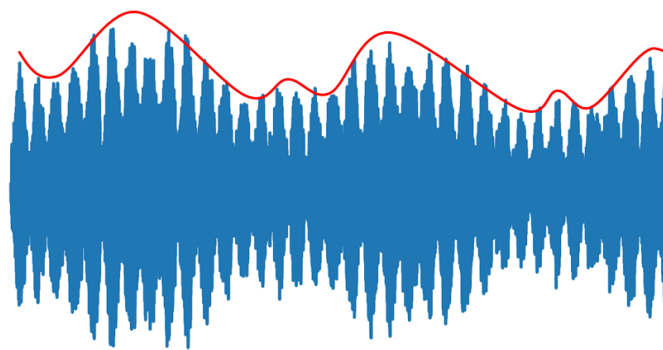


圖5.5 模擬音檔0度時的波形圖包絡線示意圖（圖中音檔長約5秒）(出處詳圖片來源18)

接著我們與實際錄製音檔的波形圖比較：

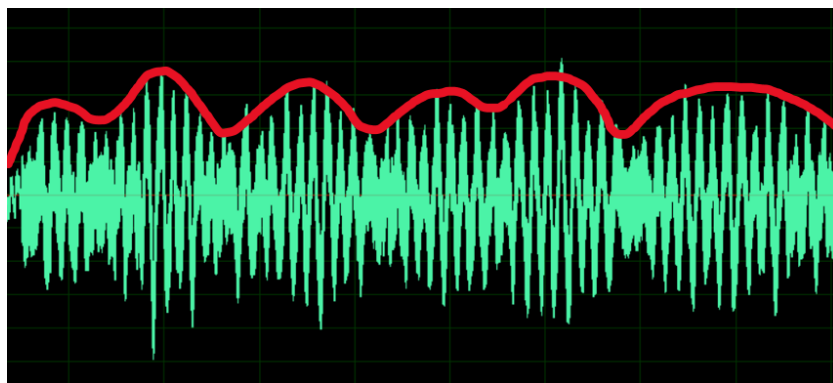


圖5.6 實際錄製音檔0度時的波形圖及包絡線示意圖（圖中音檔長約8秒）(出處詳圖片來源19)

可以發現，實際錄製音檔也有出現類似拍音的圖形，並且拍音出現的周期數量級也相近。例如：上圖5.5可以計算出5秒內含有2.5個拍音，2秒一個拍音；而圖5.6出現約3個完整的拍音，約2.6秒一個拍音。

（四）模擬結果與實際錄音的相異處出現原因

1.模型的簡化假設：模擬假設壓克力板為理想均勻的平面，忽略了厚度變化或質量不均可能造成的影響。在實際實驗中，壓克力板也有底部的支撐結構，可能會產生額外的反射與局部變形，進而影響振動模式。

2.環境因素：實驗中可能存在未考慮的環境干擾，如背景噪音、空氣流動或設備震動，這些因素會對LDV測量結果產生微小影響，模擬中則無法完全重現。

3.阻尼與非線性效應：模擬中設定的阻尼係數 γ 為固定值，未考慮隨速度變化的非線性阻尼特性。此外，空氣阻力與材料內部的能量損耗也可能導致模擬與實驗之間的差異。

四、AI模型訓練

對於我們錄製的音檔，第一組AI在分類音源左右方向時可依照音檔的特徵對此二分類進行分析和預測，而第二組AI在我們錄製的音檔中進行多分法的同象限多角度分類，也可以成功訓練出準確率將近100%的模型。以下我們將聚焦於程式架構的選擇與原因。

（一）資料預處理與特徵提取影響

頻率範圍選擇：

聚焦在特定頻率範圍有效濾除了不必要的低頻與高頻噪音，提升了模型辨識聲音方向的準確性，卻也同時可能會忽略部分頻率產生的有用特徵。

所以，在測試後發現，我們將500~1500Hz以外的部分進行歸零，而不是直接刪除，以確保觀察的準確性。

（二）模型訓練與效能分析

1.訓練曲線與模型收斂：從訓練過程中可以觀察到損失值（Loss）隨著訓練次數的增加逐漸下降，而準確率（Accuracy）則持續上升，顯示模型在學習過程中有效地擷取了數據特徵。此觀察幫助模型在接近收斂時進一步微調，提升了模型的穩定性。

2.分類表現與混淆矩陣：

(1) 如表4.5中多角度的t-SNE圖，混淆矩陣中顯示模型在大多數角度的分類表現良好，但部分接近角度（如45°與60°）有時可能出現混淆。

(2) 分類結果中，精確率（Precision）、召回率（Recall）與F1-score可以達到不錯的水準，尤其是在極端角度的分類上表現出色，證明模型具備良好的適應性與應變力。

（三）與參考論文之比較

1.論文中，判斷方向時需要播放各頻率下強度皆相同的白噪音，也就是所有的頻率同時播放，但我們的方法僅需要單頻率即可辨認聲源的方位。

2.我們利用LDV取代氣導式麥克風，透過其高靈敏度，我們僅需一個平面即可取代散射結構。

3.LDV相較於傳統氣導式麥克風，在傳導途中可以經過真空或低壓等極端環境，且依然能夠正常運作，不易受介質影響。

4.論文中的各方法準確率平均僅有約75%，而我們成功建立了分類準確率將近100%的模型。

陸、結論

本研究結合雷射都卜勒震動儀（LDV）與深度學習技術，以辨識聲音來源的方向，並取得以下主要成果：

一、推導並驗證LDV輸出的訊號：以數學模型推導得到LDV所測得之音檔波形圖即為測量點之速度圖，並且以實驗成功驗證。

二、模擬與實驗結果一致：經由數學模型推導及模擬，所得之振動數據與實際錄製音檔進行比較，結果顯示兩者在頻譜特性、倍頻現象及分貝數變化非常接近，進一步支持模型的正確性。

三、兩組AI模型的高準確率：透過卷積神經網路（CNN）的訓練，本研究建立了能有效辨識聲音來源方向的AI模型。無論是左右方向或多角度分類，模型的準確率皆接近100%，並透過t-SNE可視化與混淆矩陣進一步驗證分類效能。

四、本研究改善傳統應用氣導式麥克風的聲源定位方式，避免了中途無法經過低壓或真空等極端環境的問題，並成功將單一雷射點所收到的一維振動訊號轉換為二維的方向資訊。

柒、參考文獻資料

[1] [2] [3] 擷取並翻譯自Dalia El Badawy,Ivan Dokmanić 《Direction of Arrival with One Microphone, a few LEGOs, and Non-Negative Matrix Factorization》（發表時間：2018年12月）

<https://arxiv.org/abs/1801.03740>

[4] [5] 擷取並翻譯自趙國維《以單點光學振動感測與深度學習技術合成立體語音之研究》（發表時間：2023年8月）

<https://thesis.lib.nycu.edu.tw/items/ea34e996-3907-4d64-bae1-e521e402d1df/full>

圖片來源：

- 1.圖片擷取自Dalia El Badawy,Ivan Dokmanić 《Direction of Arrival with One Microphone, a few LEGOs, and Non-Negative Matrix Factorization》
2. LDV：圖片擷取自官網 <https://www.optomet.com/>、DT9837A DAQ：圖片擷取自官網<https://iea.com.tw/spectraplus-dt>、喇叭：圖片由作者親自拍攝、分貝計：圖片擷取自官網 <https://www.chien-leh.com.tw/tes-1350a.html>、雷射測距儀：圖片擷取自購物平台 https://m.momoshop.com.tw/goods.momo?i_code=10401426、電子量角器：圖片由作者親自拍攝、壓克力板：圖片由作者親自拍攝、反光貼：圖片由作者親自拍攝、彈簧：圖片由作者親自拍攝、磁鐵：圖片由作者親自拍攝
- 3.圖片皆擷取自軟體本身圖示
- 4.圖片擷取並翻譯自趙國維《以單點光學振動感測與深度學習技術合成立體語音之研究》
- 5.圖片由作者親自繪製
- 6.圖片由作者親自繪製
- 7.圖片由作者親自拍攝
- 8.圖片由作者親自繪製
- 9.圖片由作者親自繪製
- 10.v(t)圖利用excel繪製；音檔波形圖擷取自cool edit pro 2.1
- 11.圖片利用python模擬後繪製
- 12.圖片利用python模擬、計算後繪製
- 13.圖片利用python計算後繪製
- 14.圖片利用python執行、計算後繪製
- 15.圖片利用python執行、計算後繪製
- 16.圖片利用python計算取得數值後以excel繪製
- 17.圖片利用python計算取得數值後以excel繪製
- 18.圖片利用python模擬繪製後由作者親自繪製
- 19.圖片擷取自cool edit pro 2.1並由作者親自繪製

【評語】 052307

1. 本作品探討以雷射光學都卜勒震動儀結合深度學習辨認聲音的方向，值得鼓勵。
2. 建議應該針對以光學量測震動取代聲學量測的優勢，加強說明適當的應用場景與環境。
3. 建議應進行較深入的文獻回顧，對國內外相關研究現況分進行析，說明作品與習知技術之異同並彰顯本作品創新之處。
4. 建議對於 LDV 輸出訊號的上下限範圍、環境背景音與材料表面不均等潛在誤差源，可再進一步討論。
5. 建議若能進一步說明模型在未見過的聲音或不同距離下的泛化表現，將更具實用性。

作品海報

見 光 聞 聲：

以雷射都卜勒震動儀結合深度學習辨認聲音的方向

摘要

本研究在探討如何將「雷射都卜勒震動儀」所測量出的一維訊號，進一步辨認出二維資訊，意即聲源的來向。本研究包含兩個部分：以物理推導出系統模型，以及使用人工智能，利用音檔進行深度學習以訓練AI模組，並根據不同聲源來向進行分類。經過推導，我們發現LDV所輸出訊號與待測物表面之關係。同時，也利用電腦模擬待測物受聲波影響後之振動模式。藉此，我們以模擬還原出了LDV所收到的音檔，並嘗試確認模擬所得的音檔與實際錄製的音檔的一致性，以證明我們的推論與模擬可信。

壹、前言

動機

錄音技術從氣導式麥克風的單聲道錄音發展至雙聲道與麥克風陣列技術，透過捕捉聲波的時間差與強度差實現聲源定位。然而，在真空或低壓環境中，聲波無法傳播，導致傳統聲學定位方法失效。故本研究利用雷射都卜勒振動儀（LDV）的非接觸與高靈敏度特性，結合物理模型與人工智慧技術，嘗試將一維振動訊號轉換為二維聲源方向資訊，克服極端環境中的定位限制，並提供新的應用可能。

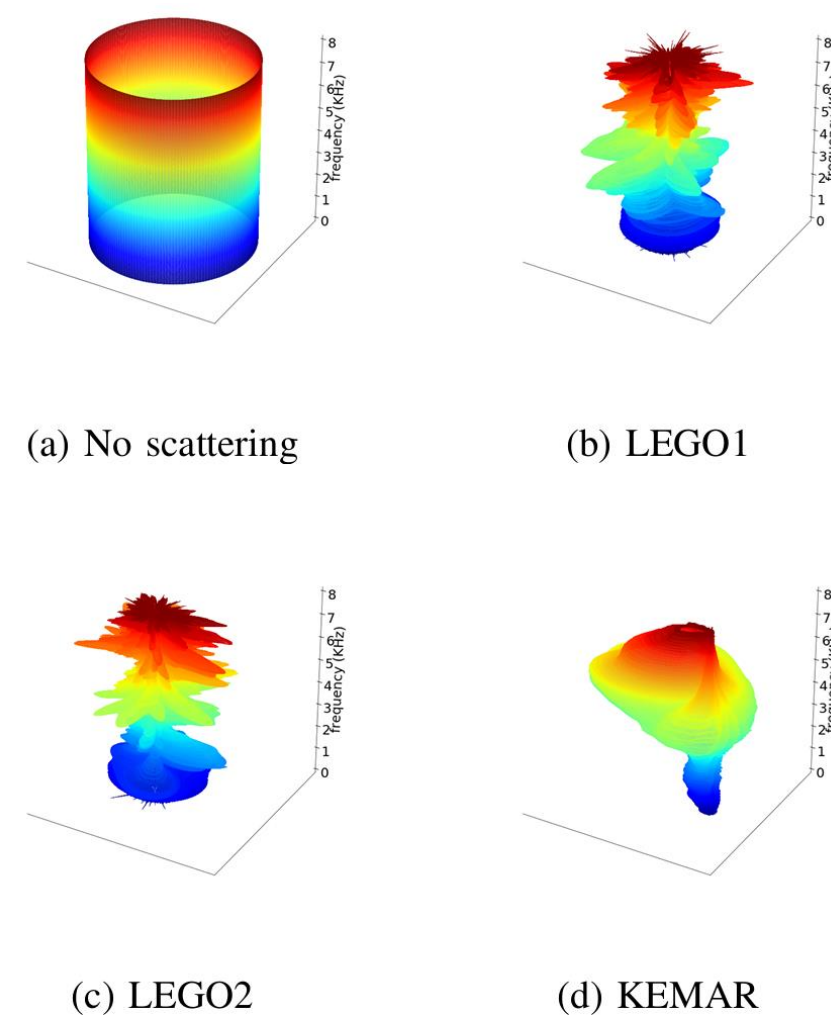


目的

- (一) 以數學公式分析LDV測量方式
- (二) 推導平面在受到聲波影響而產生的振動方式
- (三) 利用程式模擬平面受聲波影響所產生的振動，並與實際錄製音檔比較
- (四) 以卷積神經網路建立AI模型以分類不同角度的聲源方向

文獻回顧

傳統麥克風陣列與HRTF皆為聲源定位的常用工具，但有著部署複雜與成本高等問題。論文中，作者為了解決這些問題，借鑒了人耳的單耳定位能力，以單支氣導式麥克風搭配LEGO積木的散射結構，在不同角度下播放白噪音，擷取頻率響應以建立「聲紋」子空間，並且透過子空間投影與一些數學工具即可判斷聲源的方位。

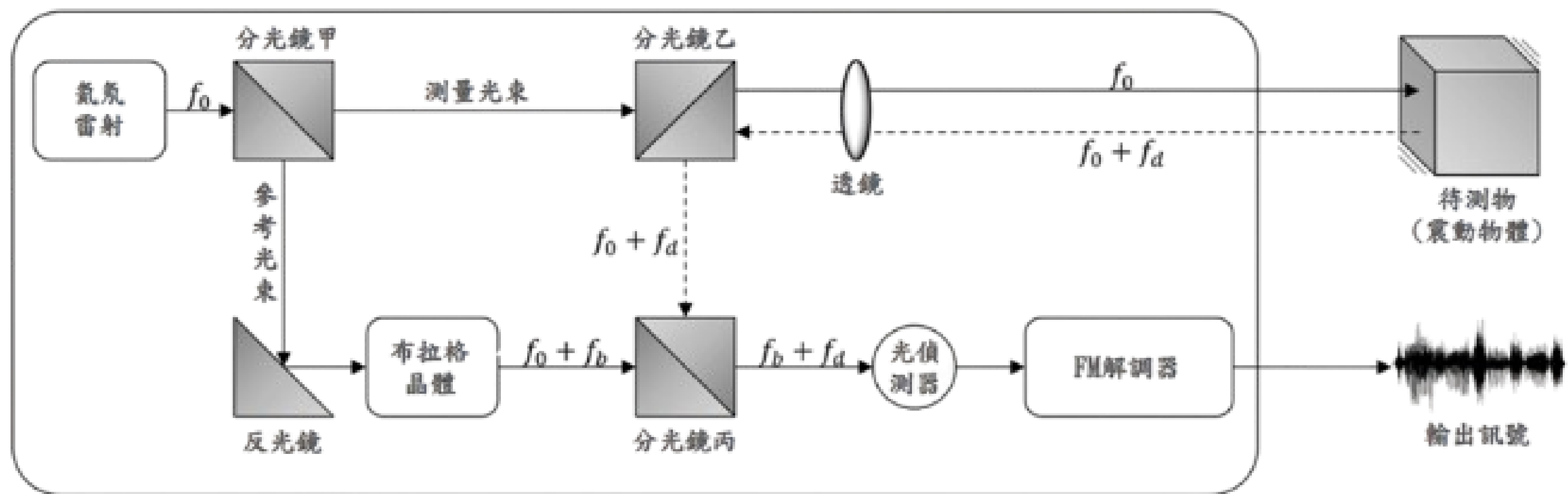


貳、研究設備與器材

LDV	DT9837A DAQ	喇叭	分貝計	雷射測距儀	電子量角器	壓克力板	反光貼	彈簧
磁鐵	Python	Excel	Matlab	Tracker	SpectraPLU S_DT	Cool Edit Pro 2.1	CLIO 12	OptoGUI

參、研究過程與方法

LDV光路分析



光打到待測物時，會產生都卜勒效應，產生頻移 $f_d(t)$ ：

$$f_d(t) = \frac{2v(t) \cos \psi}{\lambda_0}$$

測量光束與參考光束干涉後，其電場 $E_i(t)$ 為：

$$E_i(t) = E_m(t) + E_r(t) = Re\left\{A_r e^{i[2\pi(f_0+f_b)t+\varphi_r]}\right\} + Re\left\{A_m e^{i[2\pi(f_0+f_d(t))t+\varphi_m]}\right\}$$

以歐拉公式處理後得到下式：

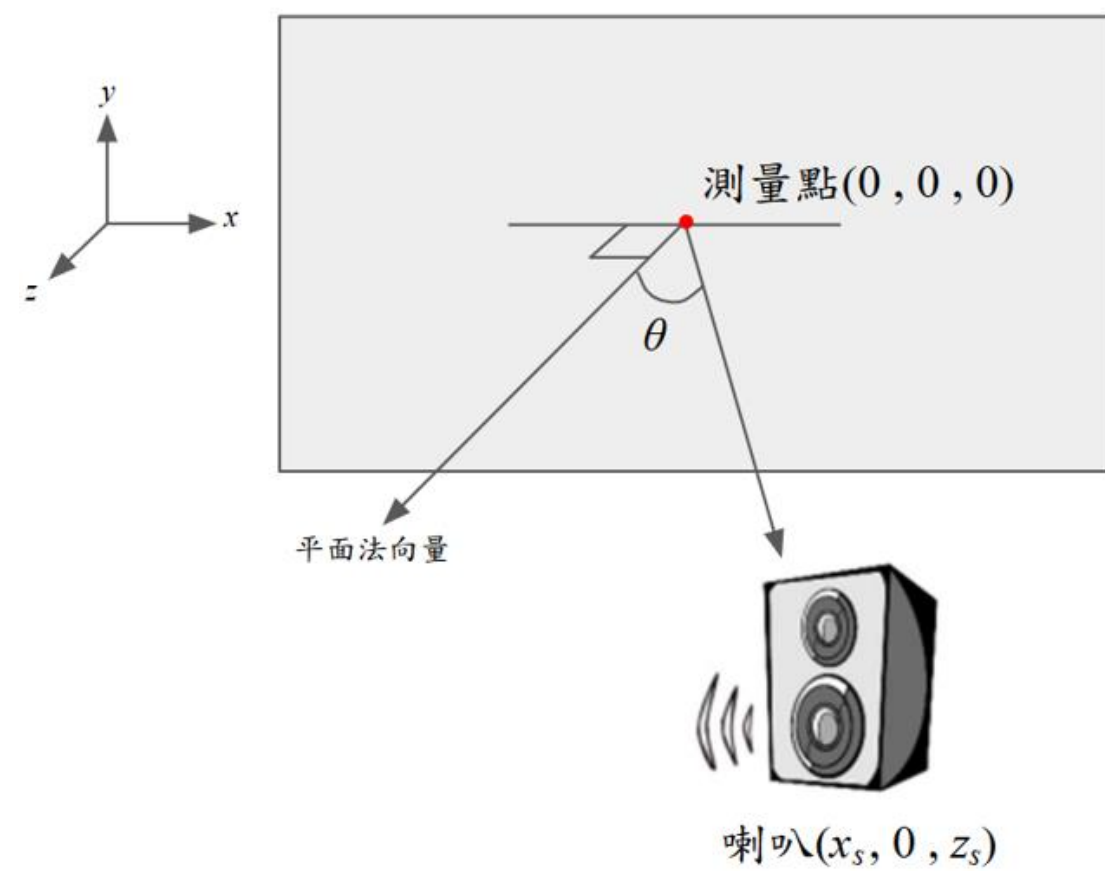
$$E_i(t) = (A_r \cos \theta_b + A_m \cos \theta_d) \cos \theta_0 - (A_r \sin \theta_b + A_m \sin \theta_d) \sin \theta_0$$

其中， $\theta_0 = 2\pi f_0 t$ 、 $\theta_d = 2\pi f_d t + \varphi_m$ 、 $\theta_b = 2\pi f_b t + \varphi_r$

利用疊合，我們可以進一步將干涉光束的電場化簡：

$$E_i(t) = \cos(\theta_0 + \psi) \sqrt{A_r^2 + A_m^2 + 2A_r A_m \cos(\theta_b - \theta_d)}$$

物體表面震動方式推導



將喇叭視作點波源，假設聲壓 $p(x, y, t)$ 如下式：

$$p(x, y, t) = \frac{A_p}{R} \times \cos(kR - 2\pi ft)$$

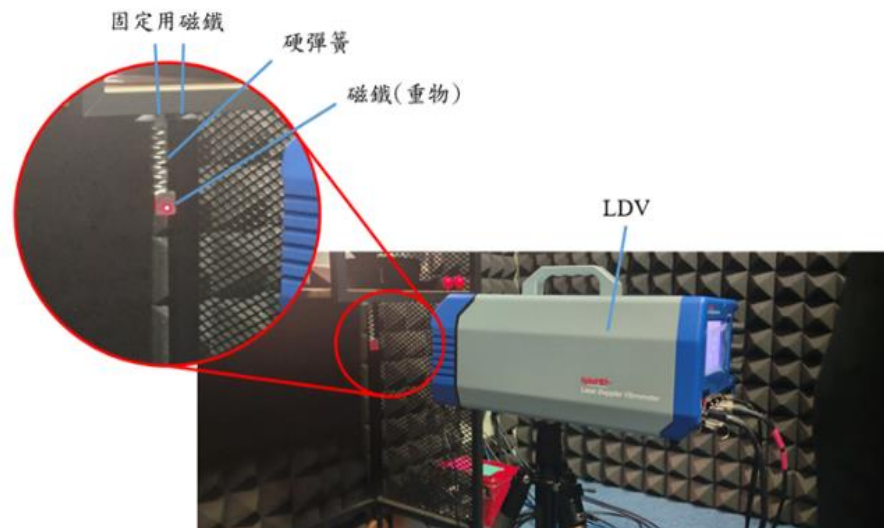
實驗

實驗一

✓目的：

驗證LDV輸出的訊號是否正比於測量點上的速度 $v(t)$

✓裝置：



✓方法：

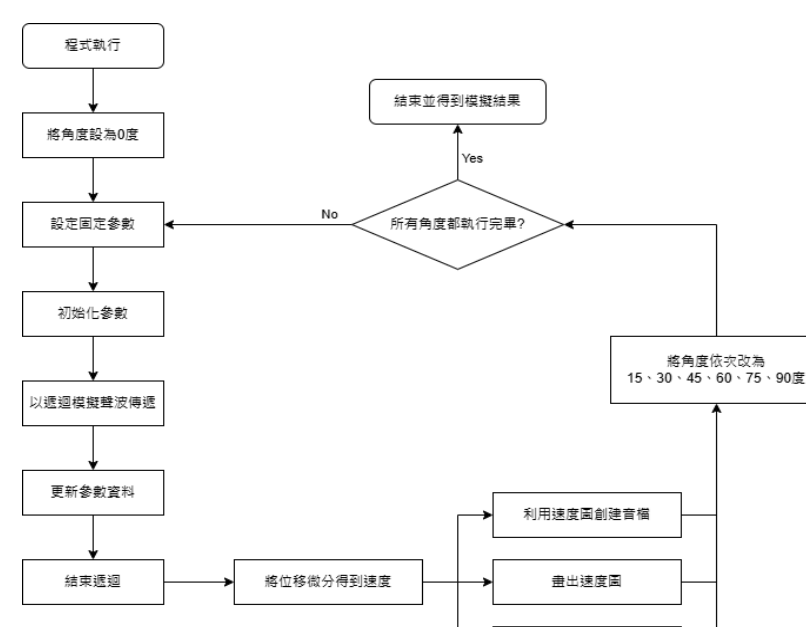
比較Tracker追蹤簡諧中磁鐵的 $v(t)$ 圖和LDV在磁鐵上所錄製到的音檔。

模擬

✓目的：

模擬出聲波打到物體後的結果

✓流程：



✓方法：

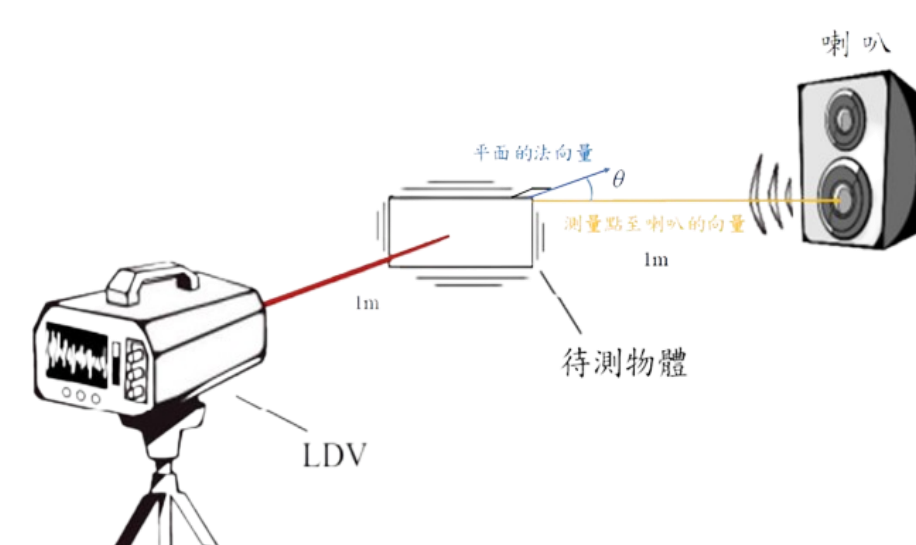
將推導出壓克力板上質點的微分方程式代入程式中，並使其輸出模擬出的音檔與速度圖。

實驗二

✓目的：

測量不同壓克力板法線與喇叭方向之夾角 θ 與LDV測得音檔之關係

✓裝置：



✓方法：

多次改變夾角 θ ，並以LDV錄音，再比較不同夾角下音檔的差別。

AI模型訓練

第一組模型

✓目的：

辨識左右聲源

✓參數設定：採樣率 16000Hz、幀長 48、重疊 24、FFT size 48

✓資料處理：

➢讀取 left/right 音檔並依 80：20 比例分割

➢利用 STFT 轉換後取 500~1500Hz 內的對數功率頻譜 (LPS)

✓CNN架構：

➢多層卷積、池化與全局平均池化

➢使用 Adam優化器，訓練 100 次 (batch_size=16)

✓特徵視覺化與評估：

➢t-SNE 降維展示中間層特徵

➢預測結果與真實標籤比對

第二組模型

✓目的：辨識同象限內不同角度（0°、15°、30°、45°、60°、75°、90°）

✓參數設定：採樣率 16000Hz、幀長 1024、重疊 512、FFT size 1024

✓資料處理：從音檔中擷取正中 4 秒，分為 7 個角度類別，80：20 比例分割

✓CNN架構：

➢多層卷積、池化與全局平均池化

➢使用 Adam 優化器，訓練 100 次 (batch_size=16)

✓特徵視覺化與評估：

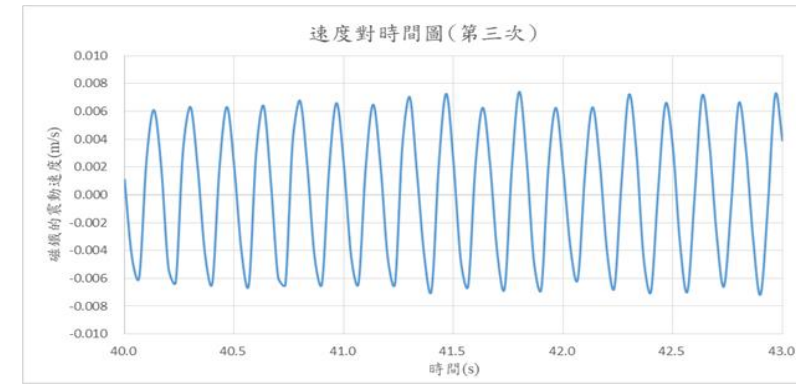
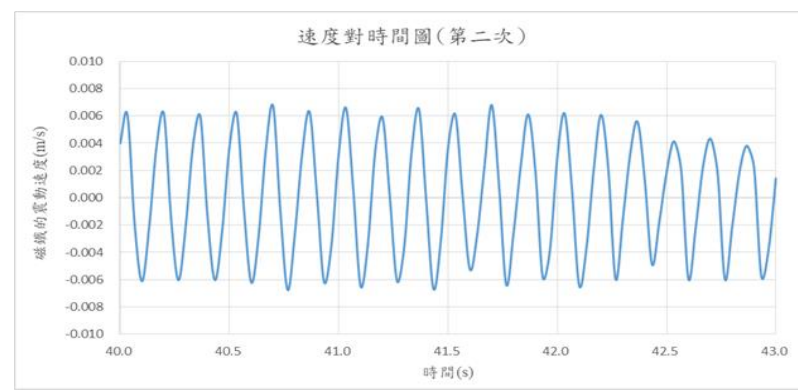
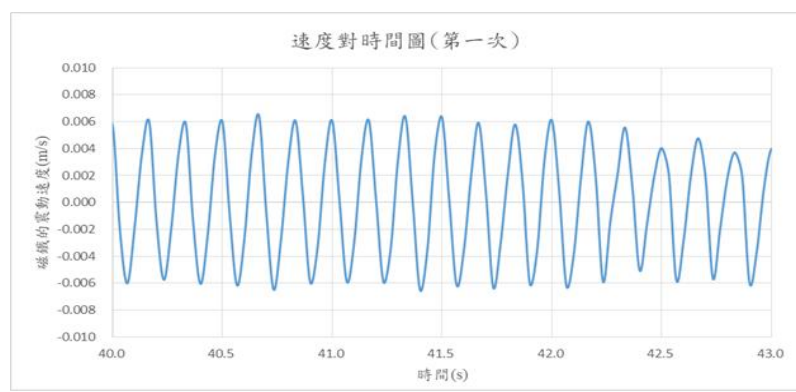
➢t-SNE 降維展示中間層特徵

➢預測結果與真實標籤比對

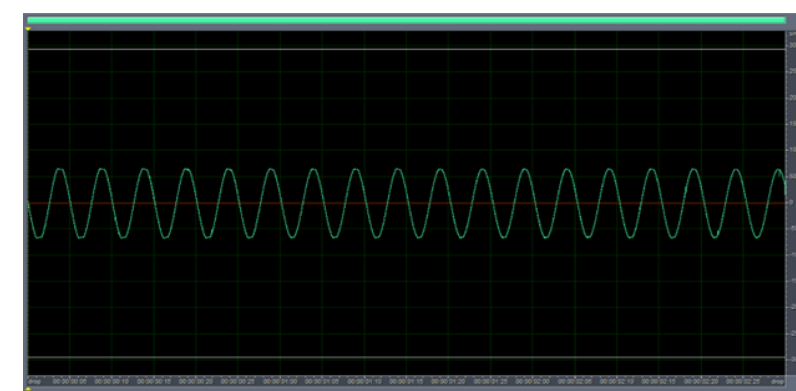
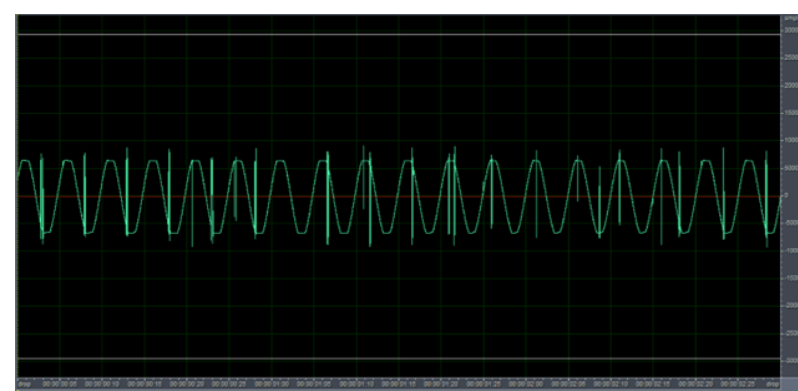
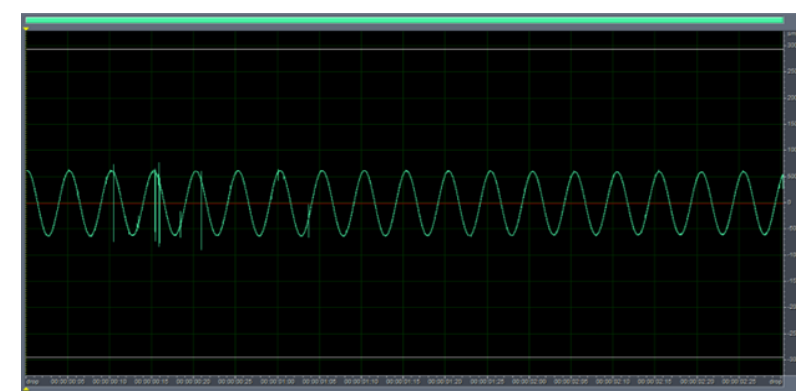
肆、結果

實驗一

對照後可以發現LDV測得之音檔與磁鐵的v(t)圖確實非常接近，因此可以支持我們推出的結果。



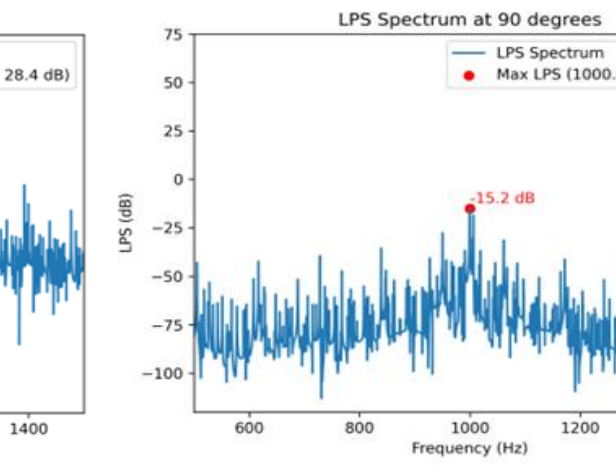
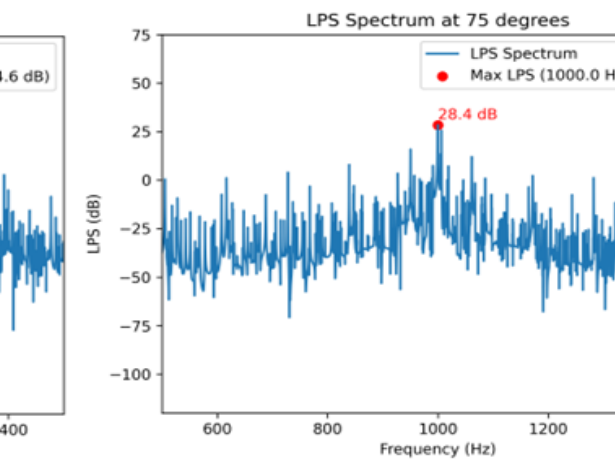
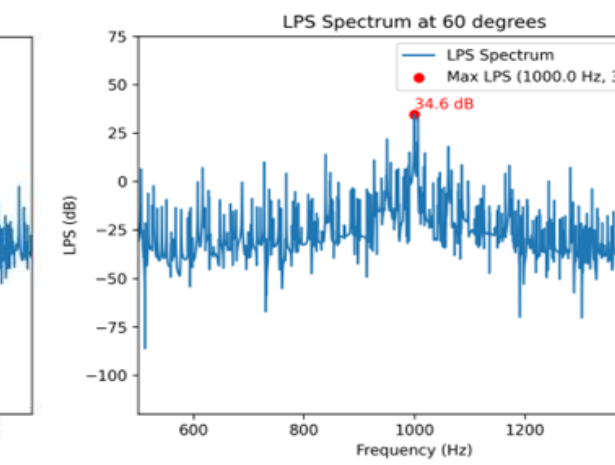
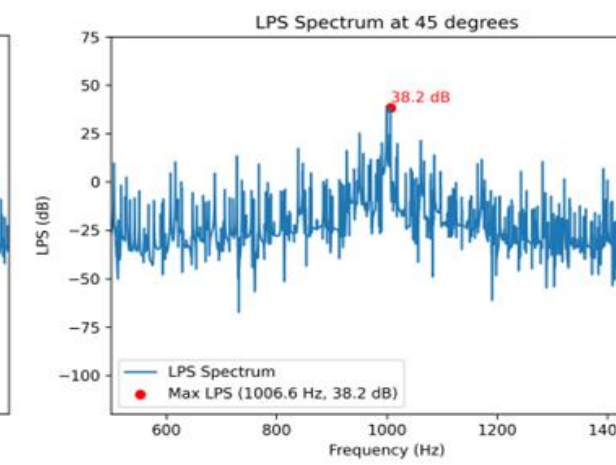
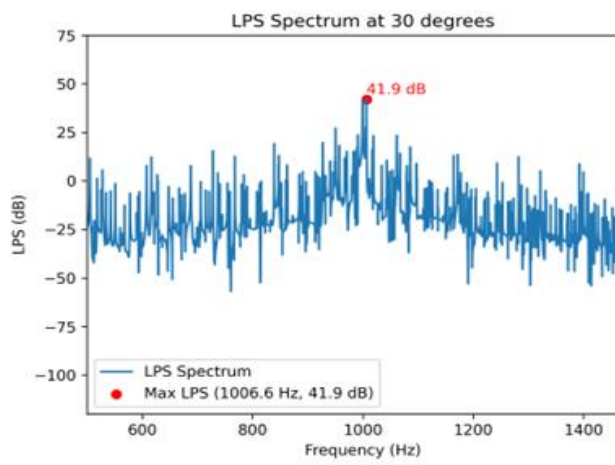
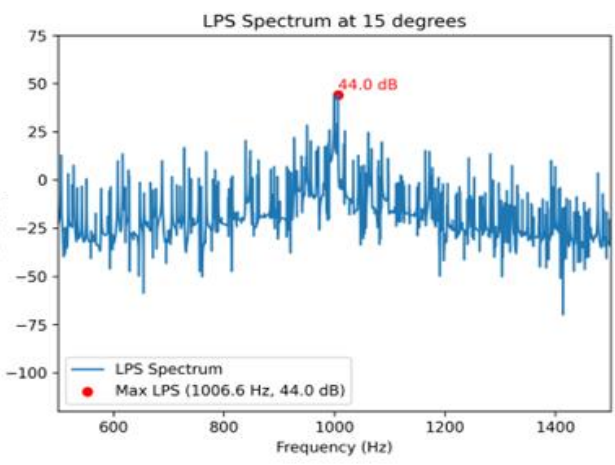
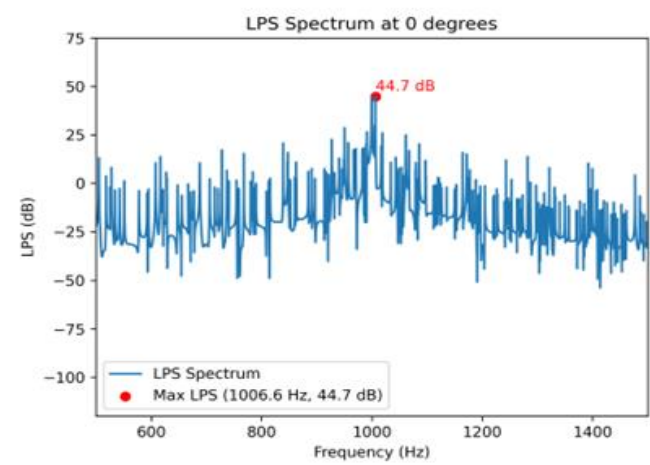
磁鐵震動速度



音檔波形

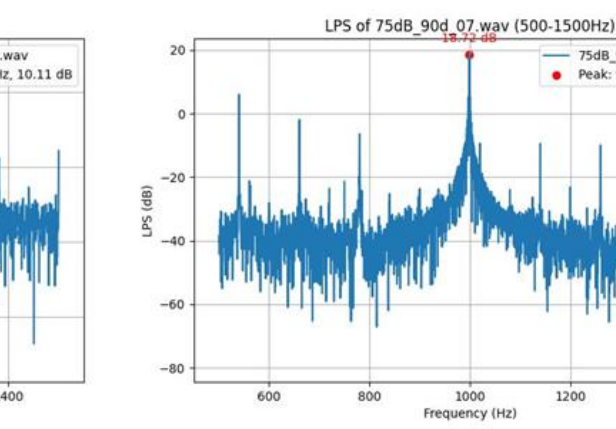
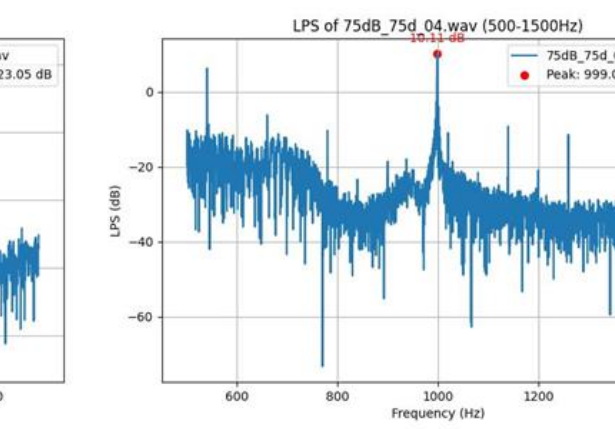
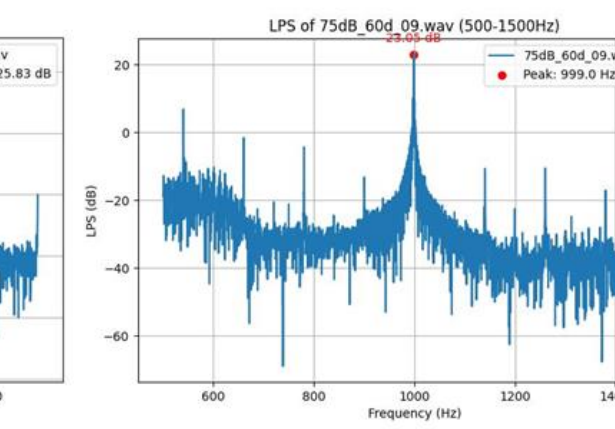
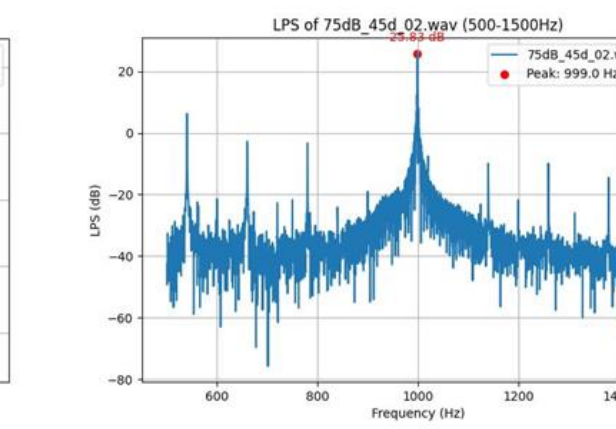
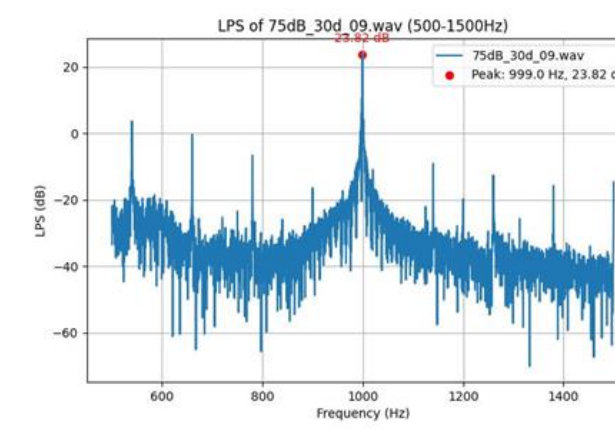
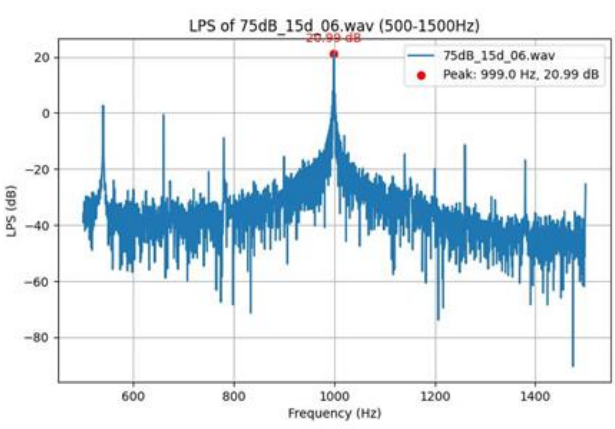
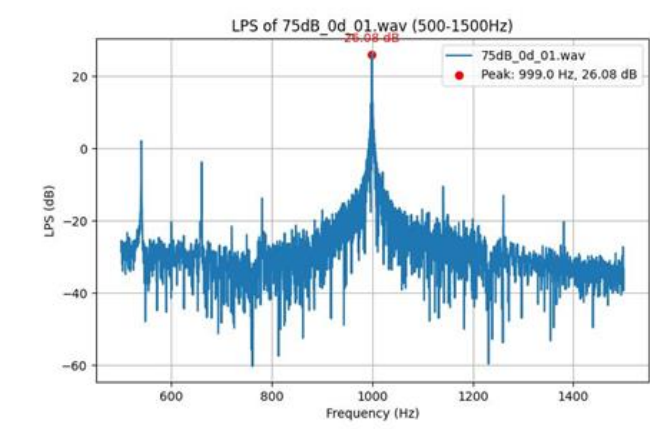
模擬

模擬音檔各角度下的LPS圖

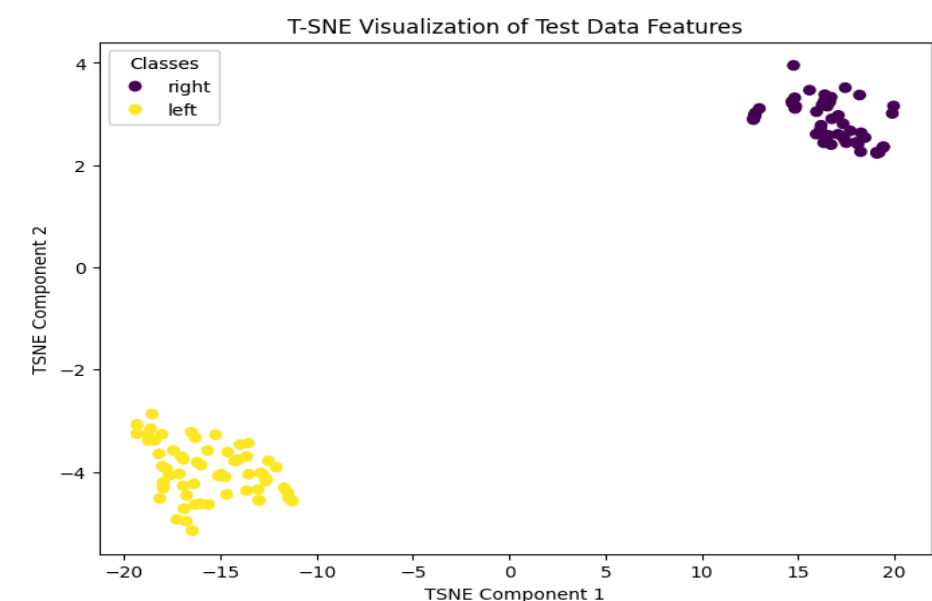


實驗二

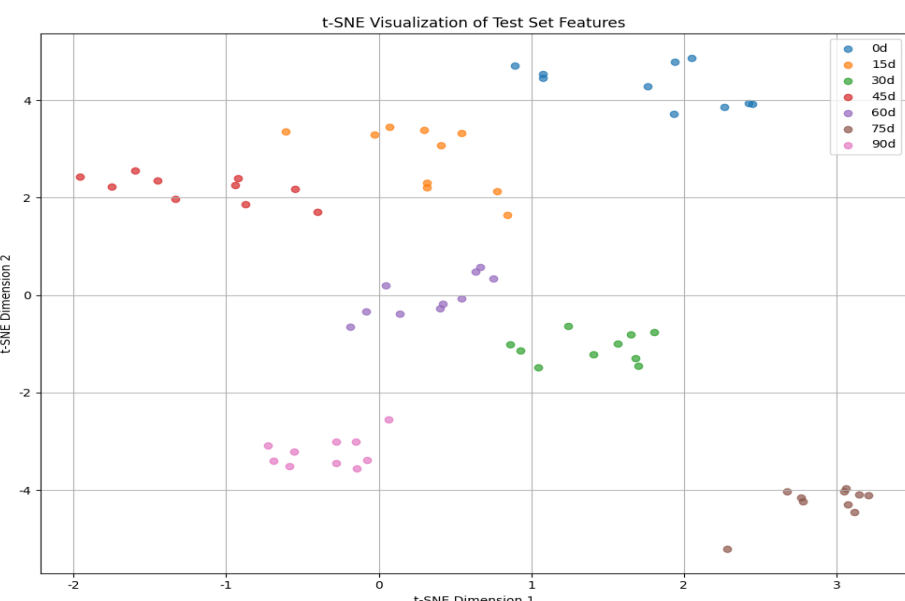
實際錄製音檔各角度下的LPS圖



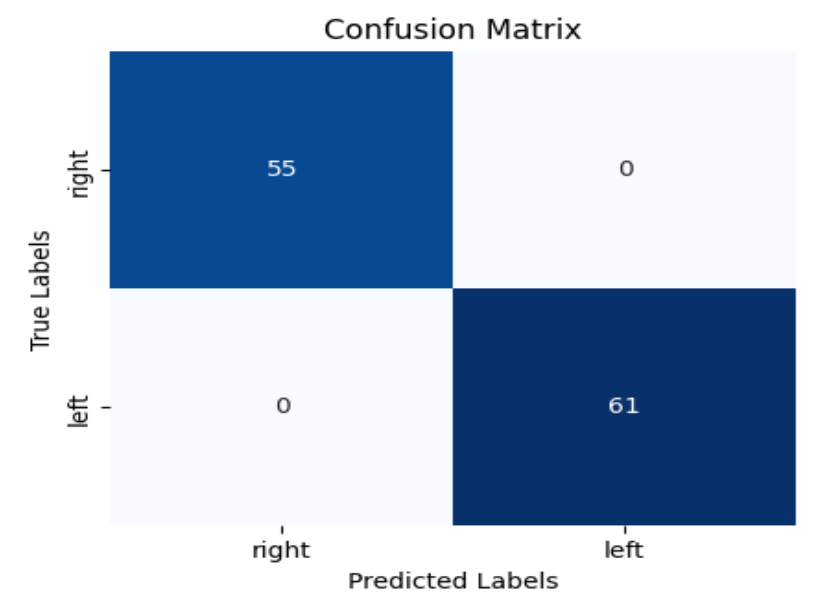
AI模型訓練



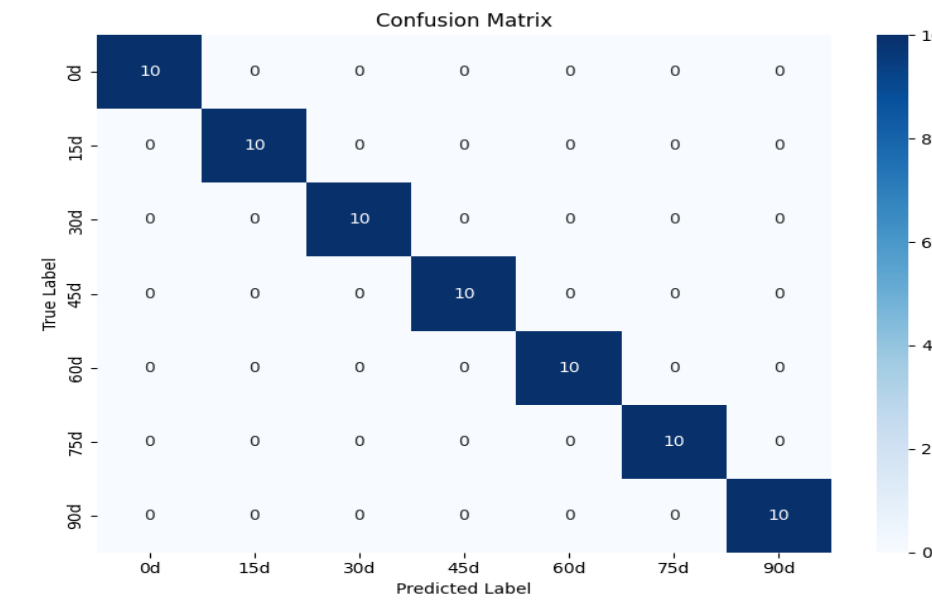
t-SNE 第一組



t-SNE 第二組



混淆矩陣 第一組



混淆矩陣 第二組

伍、討論

實驗一

一、結果

比較磁鐵振動速度與音檔波形後可以發現兩者接近，皆可看出接近正弦波的函數圖形，類似簡諧運動的速度圖。因此此結果支持LDV光路分析中推出的結果 $V_f(t) \propto v(t)$ 。

二、誤差分析

- LDV的輸出最大值
- 錄影時的解析度誤差
- 實驗時的噪音干擾

AI模型

一、結果：兩組模型皆達到100%成功率。

二、資料前處理與特徵提取影響

在測試後發現，只保留頻率範圍為[500,1500] Hz時成功率最高，即可偵測到有用特徵，亦可刪除不必要的雜訊。

三、模型訓練與效能分析

- 訓練曲線與模型收斂：損失函數訓練時逐漸下降，正確率逐漸上升，提升模型穩定性。
- 分類表現與混淆矩陣：多角度模型中，大多數角度的分類表現良好，部分接近角度偶爾出現混淆。

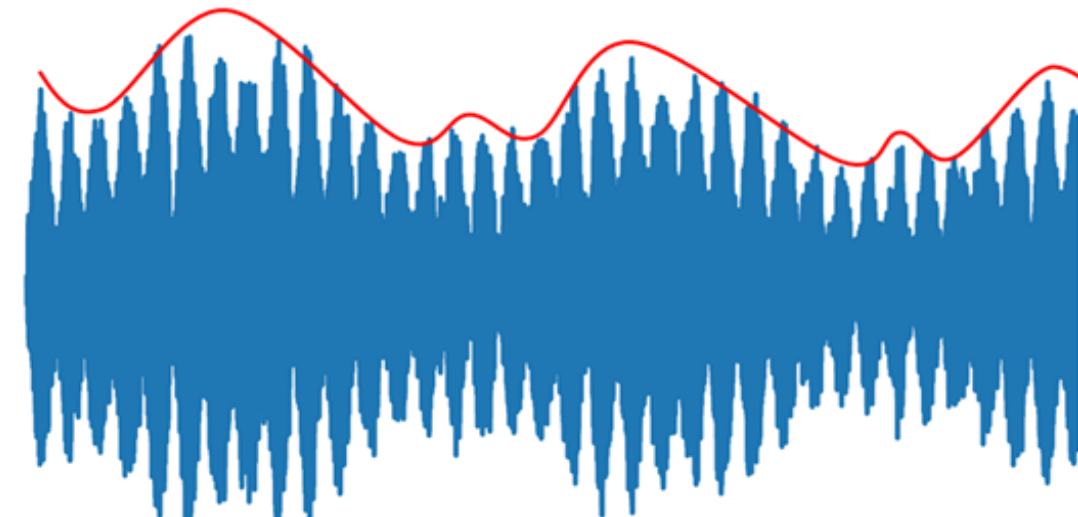
模擬與實驗二

一、模擬結果與實際錄音LPS圖的相似之處

- 皆在1000Hz出現分貝數最大值
- 皆有出現倍頻現象
- 整體圖形在角度增加時下降
- 倍頻的出現次數皆隨著角度增加而增加

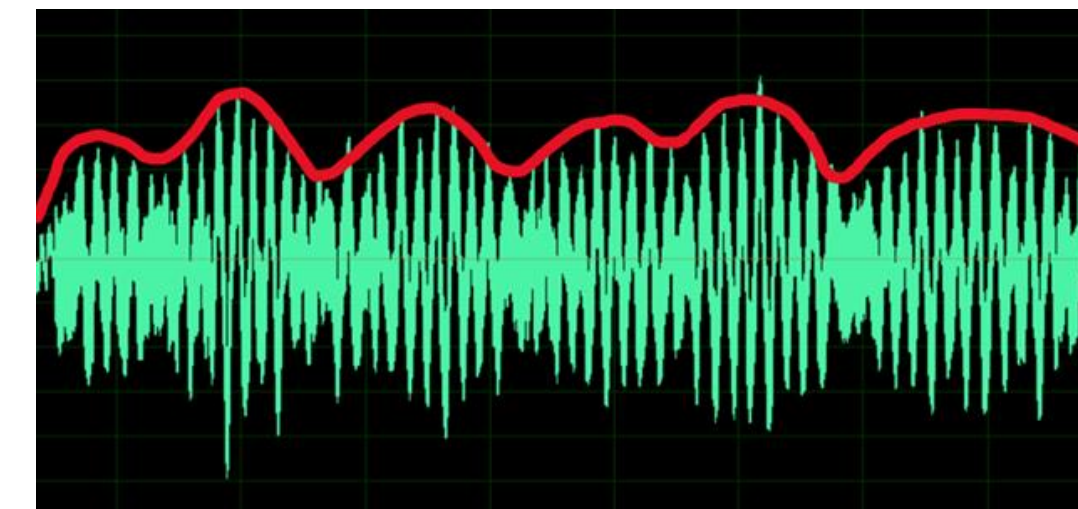
二、模擬結果與實際錄製音檔之拍音現象

實驗過程中，我們發現無論是模擬音檔或實際錄製音檔的波形圖都會出現類似拍音的現象，如右圖：



模擬音檔

兩者皆有出現拍音現象，且出現頻率相近，因此實驗二的結果支持模擬結果。



實際錄製音檔

四、模擬結果與實際錄音的相異處出現原因

- 平面的簡化假設：假設平面均勻、忽略壓克力板底部支撐
- 環境因素：背景噪音、空氣流動或設備震動

陸、結論

一、以數學模型推導得到LDV所測得之音檔為測量點之速度，並且以實驗成功驗證。

二、將模擬所得音檔與實際錄製音檔進行比較，發現波形圖非常接近，支持模型的正確性。

三、透過卷積神經網路（CNN）的訓練，本研究建立了接近準確分類聲音來源方向的AI模型。

四、本研究可以有效解決傳統麥克風定位的侷限，利用雷射提供更高準確率且適用於更廣泛範圍的聲源定位方法。