

智慧通報系統

專題成果報告書

指導導師：劉志俊 教授

組員姓名：

資工四 A 410719346 許智淵

資工四 A 410715936 黃駿瑜

資工四 A 410703507 林哲宇

資工四 A 410715318 郭駿頤

中華民國 110 年 12 月

目錄	
摘要.....	4
壹、簡介.....	5
(一) 背景.....	5
(二) 動機.....	5
(三) 目的.....	6
(四) 論文發表.....	7
貳、系統整體性描述.....	8
(一) 產品角度.....	8
(二) 產品功能.....	8
(三) 使用對象.....	8
參、系統架構與介面.....	9
(一) 系統架構圖.....	9
(二) 網頁使用者介面.....	10
肆、實驗.....	16
(一) 深度學習模型訓練.....	16
(二) 相關研究.....	22
(三) 參考文獻.....	23
(四) 開發工具.....	24
伍、結論.....	25

陸、附件.....	26
-----------	----

(一).....	26
----------	----

摘要

隨著臺灣高齡人口不斷攀升，照護措施與人力顯得不足；另外，幼童的身心發展也是大家重視的議題。我們提出一套智慧通報系統，運用情緒聲音辨識技術輔助照護員了解此二族群的情緒訴求，得以更快速地給予適當的關切。由於聲音情緒類型增加時過擬合問題嚴重影響辨識效能，本文利用希爾伯特-黃轉換的經驗模態分解將聲波分解，再根據分解後各分量訊號轉換為頻譜來訓練卷積網路來改善情緒聲音辨識效能。最終，我們使用集成式學習技術來整合各分量的預測結果。實驗結果最佳的情緒聲音測試準確率結果為 91.6%，比單純卷積網路模型的測試準確率 73.18% 大幅改善。我們希望能藉由此技術能應用在高齡人口照護和幼童監護。

關鍵詞：高齡照護、幼童照護、情緒分類、深度學習、經驗模態分解

壹、簡介

(一) 背景

在有限資源下長期照護身心障礙孩童是一件非常不容易而辛苦的工作。老師與照護員往往必須面對過於忙碌而無法兼顧到所有院童的情緒狀態的時刻。近年來人工智慧與深度學習技術發展迅速，也在醫療、工業製造、農業生產、金融科技等領域開始進行實際應用。

傳統上情緒辨識是透過臉部表情進行識別，但障礙院童，尤其是多重障礙院童往往無法單獨透過臉部表情有效進行情緒辨識。

(二) 動機

在大二參加瑪麗亞姆姆杯時，老師帶著學長和我們去[瑪麗亞基金會-台中據點]實地走訪，那是關懷弱勢孩童，像家一樣的地方，在那裡，我們看到了一些孩童們，有溝通上的困難，無法像正常人一樣說話，也無法控制自己的身體，來表達自己的想法。就像家中身邊的年長者一樣，隨著時間的過去，身體逐漸的衰弱，只要發生點小意外，當下哀號聲和呼救聲沒人發現的話，往往就產生了不可逆的影響。



(三) 目的

為了即時知道特定場所和附近有意外發生，如:獨居老人摔倒以及幼教機構虐童，來讓照護者可以更快速的了解被照護者的需求，我們提出通過人工智慧來辨識情緒聲音的系統，以此來解決現階段還無法解決的問題，並結合經驗模態分解技術來提高此系統辨識的準確率，最後，將辨識結果即時通報給相關機構以及親屬，以此降低個人意外發生後無法及時求救的機率。

(四) 論文發表

專題已發表論文：

- 1. 利用深度學習技術來辨識情緒聲音表達類型，楊俊元、郭駿頤、黃駿瑜、許智淵、劉文德、劉志俊*(通訊作者)，NCS 2019 全國計算機會議，2019.
- 2. 使用經驗模態分解與集成式深度學習模型來改善情緒聲音的辨識效能，楊俊元、郭駿頤、黃駿瑜、楊淑君、黃駿瑜、許智淵、劉文德、劉志俊*(通訊作者)，TANET 2020 臺灣學術網際網路研討會，2020.
- 3. Lin, Y. Y., Yang, J. Y., Kuo, C. Y., Huang, C. Y., Hsu, C. Y., & Liu*, C. C. (2020, December). Use Empirical Mode Decomposition and Ensemble Deep Learning to Improve the Performance of Emotional Voice Recognition. In 2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS) (pp. 1-4). IEEE.



- 1. 利用深度學習技術來辨識情緒聲音表達類型，楊俊元、郭駿頤、黃駿瑜、許智淵、劉文德、劉志俊*(通訊作者)，NCS 2019 全國計算機會議，2019
- 2. 使用經驗模態分解與集成式深度學習模型來改善情緒聲音的辨識效能，楊俊元、郭駿頤、黃駿瑜、楊淑君、黃駿瑜、許智淵、劉文德、劉志俊*(通訊作者)，TANET 2020 臺灣學術網際網路研討會，2020.
- 3. Lin, Y. Y., Yang, J. Y., Kuo, C. Y., Huang, C. Y., Hsu, C. Y., & Liu*, C. C. (2020, December). Use Empirical Mode Decomposition and Ensemble Deep Learning to Improve the Performance of Emotional Voice Recognition. In 2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS) (pp. 1-4).

貳、系統整體性描述

(一) 產品角度

隨著高齡化及照護扶養比趨勢逐漸攀升影響，社會上的潛在被照護人口也隨之攀升，然而照護的人力與資源是有限的，情緒辨識系統會對該環境持續進行監控，當發生意外時會立即通知關係人，使被通知者能第一時間做應對。

(二) 產品功能

接收該場合的情緒聲音，如笑、哭、尖叫、哀號、嘆息所發出來的情緒聲音，然後會透過此系統去分類聲音，並且判斷是哪一類的情緒，再將情緒以實體化的方式呈現出來

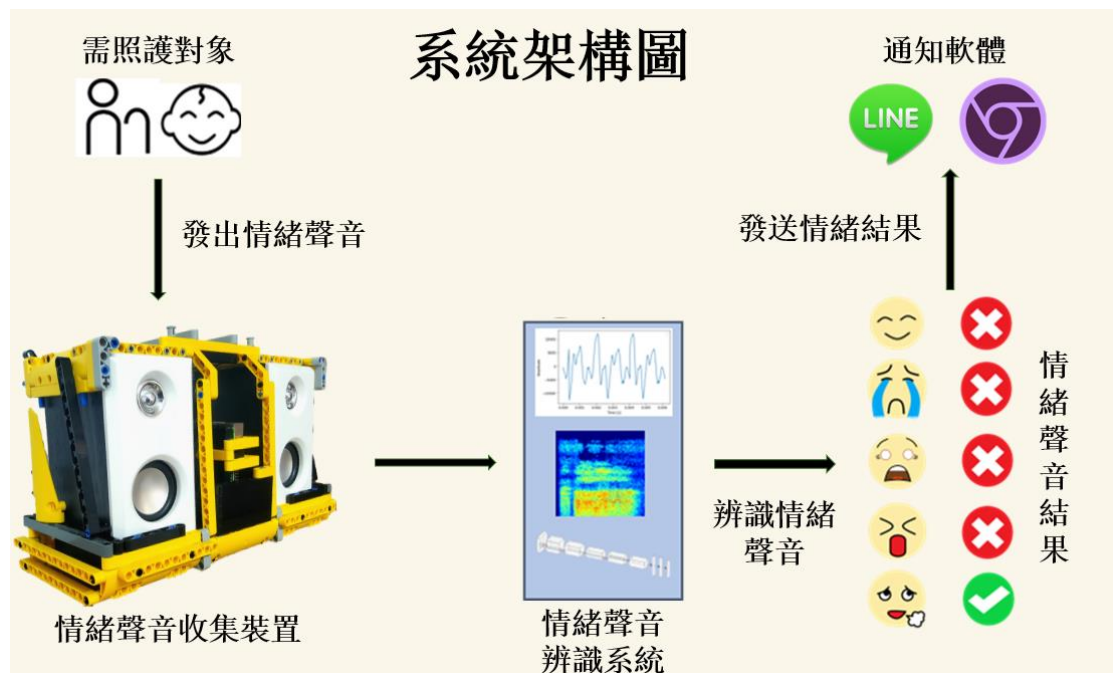
(三) 使用對象

1. 被照護者(發生意外的人、潛在發生意外的人)
2. 照護者(政府單位、照護機構、關係人)

參、系統架構與介面

(一) 系統架構圖

本研究提出之透過情緒聲音辨識來協助進行高齡人口和幼童監護系統的應用架構，首先情緒聲音蒐集裝置持續錄製監護被照護者可能發出的情緒聲音，以每五秒為一單位持續將錄製音訊傳送至情緒聲音辨識系統。此系統能將接收到的音訊進行情緒聲音辨識，當偵測到發生情緒聲音事件時，發送提醒訊息給照護員的行動裝置，照護員將按照對應情況適時給予受照護者適當關切與照護。應用架構各主要步驟說明如下。



(二) 網頁使用者介面

歡迎登入

帳號

密碼

登入

還沒有帳號嗎? 註冊


圖(一)登入畫面

輸入使用者帳號密碼登入網站



圖(二)整體畫面

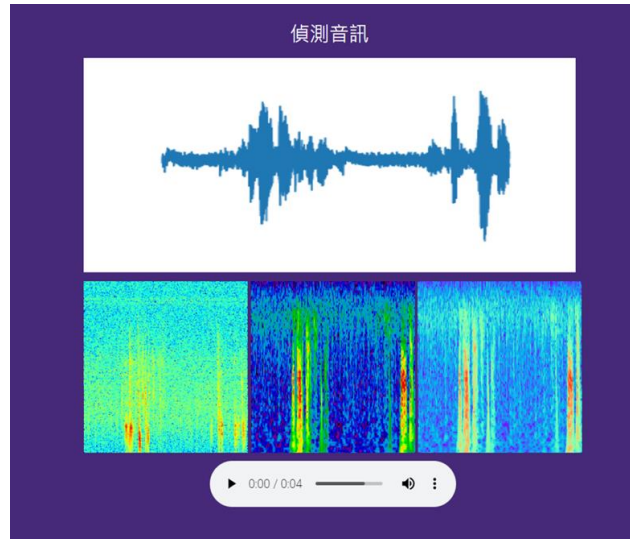
整體的排版網頁畫面。



日期：Dec. 7, 2021，時間：5:49 a.m. 偵測到Cry
日期：Dec. 7, 2021，時間：5:48 a.m. 偵測到Laugh
日期：Dec. 7, 2021，時間：5:48 a.m. 偵測到Laugh
日期：Dec. 7, 2021，時間：5:48 a.m. 偵測到Cry
日期：Dec. 7, 2021，時間：5:48 a.m. 偵測到Laugh

圖(三)網頁目錄

紀錄偵測有效情緒的發生日期、時間、情緒結果，點下任何一筆紀錄，可以查看分析報告。



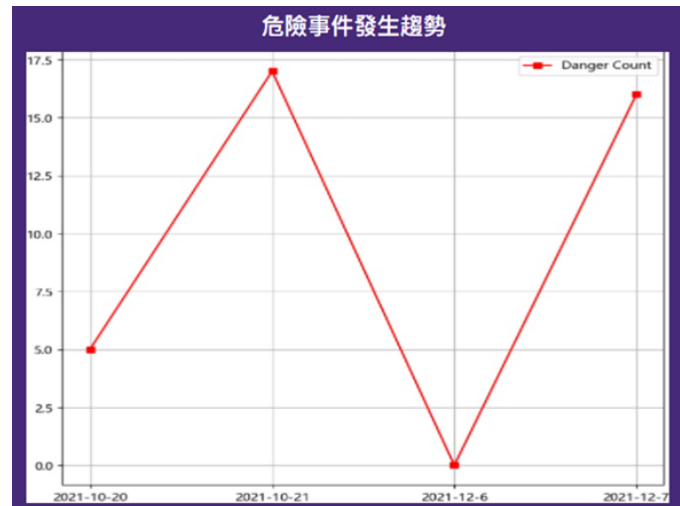
圖(四) 聲音頻譜資訊

顯示偵測到的聲音波型以及三張進行頻譜轉換後的分量圖



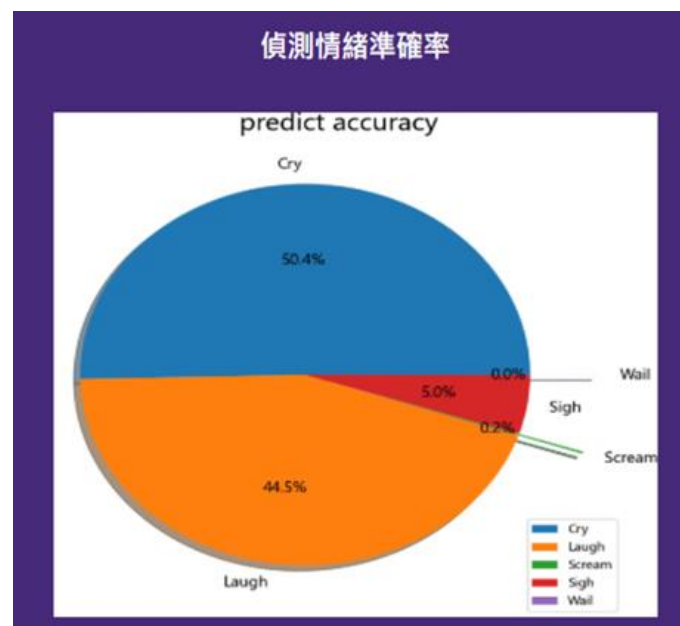
圖(五)情緒偵測結果

偵測報告為使用者最後的情緒分析結果。



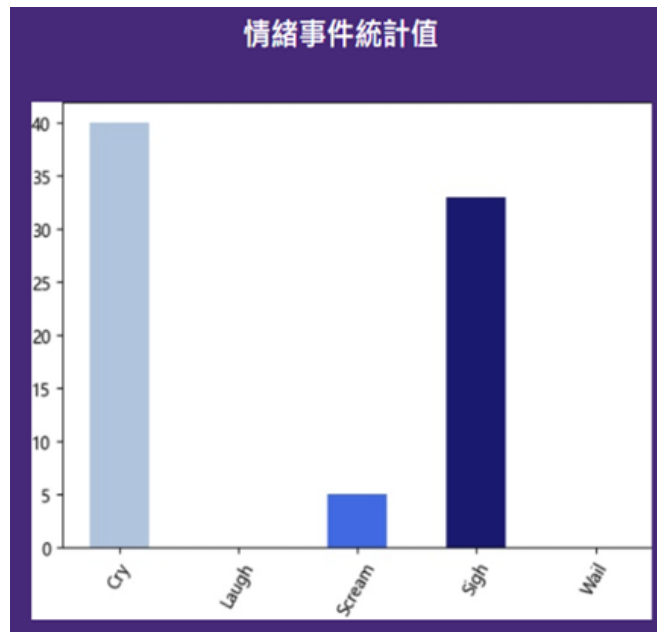
圖(六)危險趨勢統計圖

記錄在過去的五種情緒的發生機率，藉此了解使用者在這段期間的心情。



圖(七) 偵測情緒圓餅圖

統計該筆聲音經由分析過後，五種情緒發生的機率。



圖(八) 情緒事件統計值

目前已處理的每種情緒事件。



圖(九) 分析報告

音檔最後的分析報告結果，內容有機率最高的情緒結果，以及最近幾日會發生的趨勢和危險事件頻率。



圖(十) APP 使用者接收結果介面

接收當下網頁音檔所產生的最後分析報告情緒結果，並且發送到手機 APP，已告知使用者最後的分析結果。

肆、實驗

(一) 深度學習模型訓練

深度學習模型訓練

優化函數 (optimizer)	SGD, RMSprop, <u>Adagrad</u> , <u>Adadelta</u> , Adam, <u>Adamax</u> , <u>Nadam</u>
學習率 (learning rate)	0.5, 0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001
批次大小(batch size)	8, 16 , 32, 64, 128, 256
學習世代(epoch)	10, 20, 50 , 100, 150
頻譜類型(spectrum)	STFT, Mel-Scale , CQT
頻譜著色 (color map)	<u>Gist_ncar</u> , <u>Jet</u> , <u>Nipy_spectral</u> , Rainbow
丟棄率(dropout)	1, 0.8 , 0.5, 0.2, 0.1
批次正規化 (batch normalization)	Yes , No
影像解析度 (image resolution)	56x56, 112x112 , 168x168, 224x224

95.85% 的測試準確率

網路模型對辨識效能影響

網路 模型	3CN N	<u>LeNet</u>	<u>Alex Net</u>	VGG1 6	ResNet 50	網路 模型	Ince ptio n3	Incept ion4	Xcept ion	Mobil eNet	Mobile Net2
模型 大小 (MB)	6	46	250	62	101	模型 大小 (MB)	94	220	90	21	20
CNN 層數	3	3	5	13	49	CNN 層數	50	69	36	27	24
訓練 準確 率 %	97.9 7%	100.00 %	99.32 %	81.02%	96.61%	訓練 準確 率 %	85.0 3%	79.25 %	76.96 %	87.07 %	81.61%
測試 準確 率 %	83.1 6%	82.15%	87.54 %	80.81%	36.36%	測試 準確 率 %	62.6 3%	77.10 %	63.59 %	72.05 %	64.80%

優化函數對辨識效能影響

網路模型	AlexNet						
優化函數	SGD	RMSprop	<u>Adagrad</u>	<u>Adadelta</u>	Adam	<u>Adamax</u>	Nadam
訓練準確率 %	99.32%	98.31%	100.00%	99.66%	93.20%	97.96%	96.62%
測試準確率 %	87.54%	77.80%	88.22%	85.20%	85.91%	86.20%	85.19%

學習率對辨識效能影響

網路模型	AlexNet								
學習率	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
訓練準確率 %	49.66%	48.84%	74.24%	69.83%	94.22%	99.32%	98.64%	100.00 %	100.00 %
測試準確率 %	45.45%	47.47%	69.02%	68.35%	82.25%	82.83%	86.53%	87.21%	88.22%

批次大小對辨識效能影響

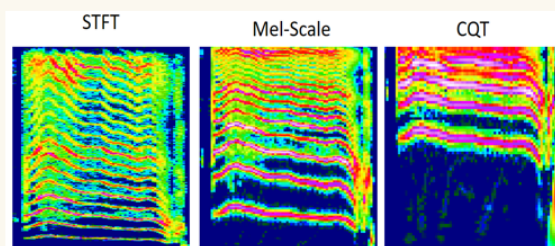
網路模型	AlexNet					
批次大小	8	16	32	64	128	256
訓練準確率 %	100.00%	99.30%	100.00%	91.89%	99.22%	40.43%
測試準確率 %	88.22%	88.58%	37.00%	53.11%	38.42%	38.28%

訓練世代對辨識效能影響

網路模型	AlexNet				
訓練世代	10	20	50	100	150
訓練準確率 %	95.45%	98.61%	99.30%	100.00%	100.00%
測試準確率 %	34.26%	35.64%	88.58%	87.54%	91.35%

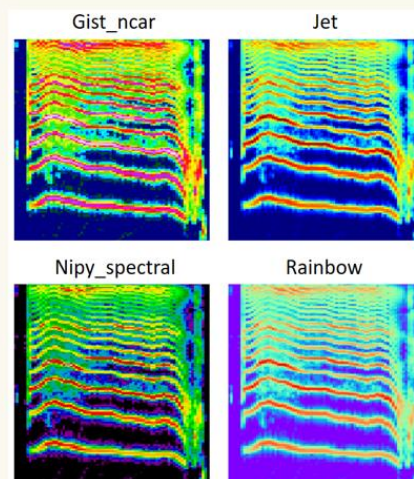
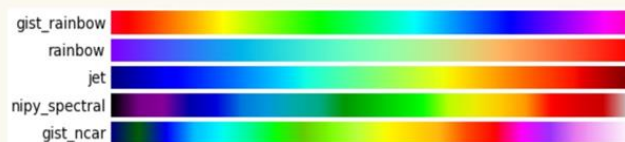
頻譜類型對辨識效能影響

網路模型	AlexNet		
學習率	STFT	Mel-Scale	CQT
訓練準確率 %	100.00%	100.00%	99.65%
測試準確率 %	91.35%	92.04%	91.35%



頻譜著色調色盤對辨識效能影響

網路模型	AlexNet			
頻譜著色	Gist_ncar	Jet	Nipy_spectral	Rainbow
訓練準確率 %	100.00%	100.00%	100.00%	100.00%
測試準確率 %	92.04%	92.04%	92.73%	93.08%

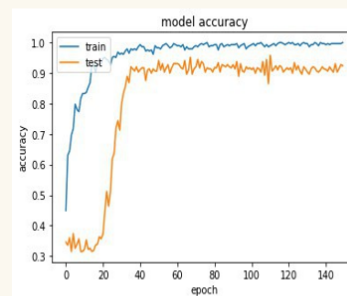


丟棄率對辨識效能影響

網路模型	AlexNet				
丟棄率	1	0.8	0.5	0.2	0.1
訓練準確率 %	100.00%	100.00%	100.00%	100.00%	100.00%
測試準確率 %	91.35%	95.85%	93.08%	93.08%	91.35%

影像解析度對辨識效能影響

網路模型	AlexNet			
影像解析度	56x56	112x112	168x168	224x224
訓練準確率 %	99.65%	100.00%	99.65%	97.87%
測試準確率 %	94.81%	95.85%	93.43%	93.42%



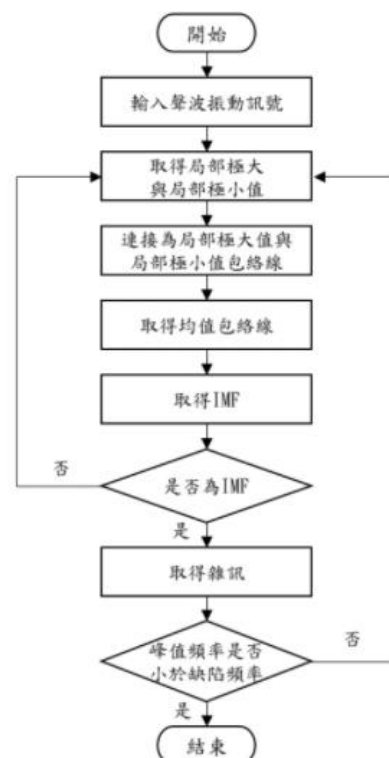
批次正規化對辨識效能影響

網路模型	AlexNet	
	無	有
批次正規化		
訓練準確率 %	100.00%	100.00%
測試準確率 %	93.43%	95.85%

(二) 相關研究

希爾伯特-黃轉換的經驗模態分解

卷積網路是一項非常強力的自動特徵選取方法，但在資料樣型非常發散或訓練資料集不足的狀況下，往往會發生過擬合(overfitting)的現象，導致模型的訓練準確度可能很高，但測試準確度不佳的情形。雖然有一些技術如網路架構最佳化、丟棄法(dropout)、批次正規化(batch normalization)、適當地進行交叉驗證等方法，可以試著來改善過擬合現象，但如果資料集本身特徵難以一般化時，往往這些技術能改善的程度有限。以本文所探討的情緒聲音分類問題而言，每個人的情緒聲音頻譜差異很大。當要辨識的情緒聲音類型增加而訓練樣本不足時，過擬合問題變得更嚴重。此時如果能透過預處理技術，來讓同類資料更有一致性，可以改善過擬合問題讓卷積網路的辨識效能提升。



(三) 參考文獻

(1) Abou-Abbas, L., Tadj, C., Gargour, C., & Montazeri, L. (2017). Expiratory and inspiratory cries detection using different signals' decomposition techniques. *Journal of voice*, 31(2), 259-e13.

(2) Kim, Jinwoo, et al. "Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition." *Building and Environment* 181 (2020): 107092.

(3) Santamaria-Granados, Luz, et al. "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)." *IEEE Access* 7 (2018): 57-67.

(四) 開發工具

軟體開發

1 . PyCharm 、 Spyder

2. 人工智慧開發

3. Keras 、 Tensorflow

硬體開發

1.樹梅派 4 代

伍、結論

1. 未來，我們希望持續蒐集更多的聲音樣本與情緒類型，提供更多樣化的情緒聲音辨識服務。
2. 以保障隱私權為前提，在照護機關或場域實際錄製聲音樣本。
3. 結合影像辨識技術，達到最佳的照護品質。

陸、附件

(一)

2019 年參加教育部舉辦的智慧跨域創作專題競賽，榮獲佳作。

2019智慧跨域創作專題競賽



本競賽可提升跨域整合實務經驗，並於校園培養軟硬體設計人才，拓展國內人才培育管道，協助廠商發掘軟體專業人才，以強化國內開源軟體研發技術能量。對於競賽獲獎團隊，將主動輔導團隊學生進入產業界服務，媒合獲獎作品，促成國內優質師生團隊邁向全球優質企業與產業接軌。本系資工四A楊俊元、資工二A郭駿頤、資工二A黃駿瑜、資工二A許智淵以「傾『聽』」此專題作品榮獲佳作，此組由資工系劉志俊老師指導。



教育部智慧創新跨域人才培育聯盟計畫
智慧感知與互動體驗跨校聯盟中心

智慧跨域創作專題競賽

第一階段 書面報告審查

第二階段 口頭簡報與實機展示 (決賽)

國立高雄科技大學前莊大樓

(高雄市三民區建國路415號)

2019年11月29日 決戰

主辦單位：教育部智慧創新跨域人才培育聯盟計畫

承辦單位：國立高雄科技大學前莊大樓

協辦單位：國立高雄科技大學前莊大樓

國立高雄科技大學前莊大樓



獎項	班級	姓名	高中畢業學校
佳作	資工四A	楊俊元	國立岡山高中
	資工二A	郭駿頤	新北市立三重高中
		黃駿瑜	私立文興高中
		許智淵	私立屏榮高中



獎 狀

靜宜大學 資訊工程學系 學生許智淵、黃駿瑜、楊俊元、郭駿頤
(隊名：天上天下唯我獨尊；指導老師：劉志俊老師)，以作品「傾
聽」參加教育部智慧創新跨域人才培育聯盟計畫智慧感知與互動體
驗跨校聯盟中心舉辦之「智慧跨域創作專題競賽」，表現優異，榮
獲佳作，特頒此狀，以資鼓勵。

教育部智慧創新跨域人才培育聯盟計畫
智慧感知與互動體驗聯盟中心

計畫主持人
國立台北科技大學
資訊工程學系

劉傳銘



中 華 民 國 1 0 8 年 1 1 月 2 9 日