

# 應用 DNN-SVM 於台灣語音情緒語料庫之跨語言語音情緒辨識

曾維翰 施學琦

國立雲林科技大學資訊管理系

M10723018@yuntech.edu.tw

## 摘要

當今人工智慧、機器學習與深度學習的發展下，有越來越多具有自動語音辨識的科技出現，但是由於語言、對話經常會因為情緒這個因素而有所變化，而過去許多研究也證明了結合情緒的語音辨識系統是能有效提升成效的，因此語音的情緒辨識是未來的一個重要研究方向，本研究使用 DNN 結合 SVM 加入瓶頸層的辨識方式來研究語音情緒辨識，並透過結合台灣語音情緒語料庫以及柏林語料庫，進行跨語言、跨語料庫的語音情緒辨識。在本研究的單一語言、混合語言、跨語言三種實驗中，國語、台語、客語均比過去的研究取得明顯更加良好的結果。

**關鍵詞：**語音情緒辨識、台灣情緒語料庫、支持向量機、深度神經網路

## 1. 前言

### 1.1 研究背景與動機

近幾年出現越來越多具有自動語音辨識 (Automatic Speech Recognition, ASR) 的科技，例如：Apple Siri、Google 語音助理、Amazon Alexa、小米的小愛同學，自動語音辨識也有越來越多的研究被提出[1]，而過去許多研究也證明了結合情緒的語音辨識系統是能有效提升成效的[2-4]，所以各種語音助理、語音程式若是能考量到對話時每句話的情緒，那麼使用者在使用上將能有更好的效果。

現在有越來越多語音情緒辨識 (Speech Emotion Recognition, SER) 的研究，但因為許多情緒語料庫大多沒有開放使用，所以大部分研究都是針對單一語料庫進行研究，因此跨語料庫之間的研究相對稀少。

[1]學者們提出一種由三種子模型組成的情緒識別系統，分別為 LLD-DNN、LLD-SVM、DNN-SVM，其中發現 DNN-SVM 相較於其他兩種具有更加良好的辨識率以及召回率。本研究受到[1]學者們研究的啟發，藉由修改該研究使用的方法，並應用在台灣情緒語料庫以及柏林情緒語料庫上進行跨語料庫的研究，從而提升使用台灣情緒語料庫的語音情緒辨識率。

### 1.2 目的

本研究目的有兩點：

(1).本研究修改由 DNN-SVM 組成的語言情緒辨識方法，對台灣語音情緒語料庫進行單一語

言、混合語言、跨語言三種語言情緒辨識，透過在 DNN 中加入瓶頸層來提高 DNN-SVM 的辨識率，並與傳統 DNN、SVM 以及加入瓶頸層前後 DNN-SVM 的結果做比較。並將修改後的 DNN-SVM 與過去使用台灣語音情緒語料庫的相關研究進行比較。

(2).本研究透過結合台灣語音情緒語料庫以及柏林語料庫，進行跨語料庫的語音情緒實驗，並與未跨語料庫的實驗比較各實驗間的結果。

## 2. 文獻探討

### 2.1 情緒語料庫

情緒的分類一直都是心理學、情緒科學和情緒研究等不同領域的熱門話題，而根據[5]針對32個語音數據庫進行的研究發現，絕大部分研究人員在建置情緒語料庫時是採用[6]所提出的六種基本情緒，憤怒(anger)、厭惡(disgust)、恐懼(fear)、幸福(happiness)、悲傷(sadness)和驚訝(surprise)，並加以變化。而根據[7]情緒語料庫的來源可以分為：演員以預先定義的情緒說話而記錄的語料庫、從現實生活中的系統(例如：呼叫中心、學習和遊戲軟體)獲得數據的語料庫。

[8]的研究指出，現在語料庫語言大多集中德語、英語，再來是希伯來語、俄語、西班牙語，其他語言的語料庫數量不是很充足，不只是中文缺乏語料庫，台灣目前的語料庫也較稀少。

而[9]的研究也指出目前現有的語料庫絕大部分都沒有開放公開使用，很少有標準語料庫可以在研究人員之間共享，這也使得大部份研究都侷限於單一語料庫，因此公開的情緒語料庫就顯得極其重要，而柏林情緒語料庫是目前研究上最多人使用的公開情緒語料庫。

柏林情緒語料庫(Berlin Database of Emotional Speech, Emo-DB)是柏林工業大學所建立的德文公開情緒語料庫，該語料庫的文本使用日常口語以及日常用語，並由五位男性以及五位女性共十位演員使用中性、憤怒、悲傷、恐懼、無聊、厭惡以及喜悅七種情緒來演繹，共產生了大約535個語音情緒樣本，壓縮後的採樣率為16kHz[10]。

台灣情緒語料庫是由國立中山大學資訊工程系的學者們於2013年仿製柏林情緒語料庫所錄製，語料庫包含七種情緒：中性、憤怒、悲傷、恐懼、無聊、厭惡以及喜悅。錄製語言採用台灣常見的國語、台語及客語三種語言[11]。

### 2.2 語音情緒辨識

語音情緒辨識 (Speech Emotion Recognition, SER)，是屬於情緒運算 (Affective Computing) 的一

部分，現有的語音情緒辨識方法可以分為兩種，第一種是先將語音轉換成頻譜圖，之後使用卷積神經網路（Convolutional Neural Networks，CNN）並結合其他分類方法進行辨識。

第二種從語音中提取語音特徵，然後根據特徵判斷情緒狀態。[12]針對語音情緒辨識進行調查與研究，該研究認為設計語音情緒辨識系統有三個方面，首先是選擇出合適的語音特徵，再來是設計合適的分類方法，最後是準備情緒語料庫及評估分類器性能。語音特徵可以分為四類：連續特徵、質性特徵、光譜特徵、能量算子特徵。而其中光譜特徵中的 MFCC 被廣泛用於語音情緒識別[13]，它在描述人耳的聽覺特徵方面具有良好的表現。

[1]學者們針對 DNN 與 SVM 的各自優點提出了一種由三種子模型組成的情緒識別模型，第一個子模型以低階語音特徵（Low Level Descriptors, LLD）為特徵，並用深度神經網路（Deep Neural Networks, DNN）作為情緒分類器，稱為 LLD-DNN。第二個子模型一樣使用 LLD，但使用支持向量機(Support vector machine, SVM)進行分類，稱為 LLD-SVM。第三個子模型是透過提取特徵以訓練 DNN，並饋入 SVM 分類器中以進行最終語音情緒辨識，然後使用投票機制將這三個子模型組合在一起，由這三個子模型個別去分類每句話，而分類後的多數結果被選為最終分類，如果三個子模型均給出不同的分類，則隨機選擇一個分類做為最終分類，該研究證明組合後的系統在準確性和召回率方面都有良好的表現，其中 DNN-SVM 相較於其他兩種方法的準確性和召回率方面都有更加良好的表現。

## 2.3 台灣情緒語料庫之相關研究

[11]的研究中以多層感知器(Multilayer Perceptron, MLP)進行單一、混合及跨語言之實驗，在單一語言實驗中各語言最佳辨識率為國語60.0%、台語48.9%、台語61.3%；混合語言實驗中各語言最佳辨識率為國語63.5%、台語53.1%與客語64.6%；跨語言實驗中各語言最佳辨識率為國語58.7%、台語48.6%與客語53.5%，該研究認為使用混合語言能提升情緒辨識率。

[14]的研究提出使用極限學習機(Extreme Learning Machines, ELM)進行單一、混合及跨語言之實驗，並結合柏林情緒語料庫進行跨語料庫的實驗。在未跨語料庫的單一語言實驗中各語言的最佳辨識率為國語65.32%、台語56.06%、台語60.00%；混合語言實驗中最佳辨識率為國語71.69%、台語61.23%與客語67.45%；跨語言實驗中最佳辨識率為國語55.89%、台語50.94%、客語50.14%。

而在跨語料庫的混合語言實驗中最佳辨識率為國語66.76%、台語59.05%與客語64.54%；跨語言實驗中最佳辨識率為國語56.18%、台語51.17%

與客語48.77%。該研究認為增加多個語料庫的訓練對情緒辨識是有利的。

## 2.4 支持向量機

支持向量機(Support vector machine, SVM)於1992年提出，是一種機器學習(machine learning)中常使用的一種監督式學習方式，被廣泛應用於統計分類和迴歸分析。目的是在特徵空間上找到一個可以區分兩筆資料的超平面，而此超平面與兩個類別的邊界越遠，分類效果就越好。SVM 具有較少的訓練複雜度以處理大量屬性的能力，許多研究也表明 SVM 即使是使用較小的訓練集，它們也可以產生良好的分類精度[15]。

## 2.5 深度神經網路

深度神經網路（Deep Neural Networks, DNN）是一種判別模型。DNN 內部具有多層的人工神經網路，DNN 內部的神經網路可以分為三類，輸入層、隱藏層與輸出層。

當一個隱藏層中的神經元數量遠少於其他隱藏層中的神經元數量時，該層稱為瓶頸層。而透過瓶頸層可以得到瓶頸特徵。過去許多研究發現使用瓶頸層來獲取瓶頸特徵是能提升績效的，[16]針對醫療臨床方面的語音輔助系統，透過使用瓶頸層提取瓶頸特徵來進行自動語音辨識，發現使用瓶頸特徵是能提升績效的。

## 2.6 openSMILE

openSMILE 是一個由德國 audEERING GmbH 開發的開放程式，主要用於從音頻信號中提取特徵以及對語音和音樂信號進行分類。openSMILE 在許多研究中被廣泛用於提取聲學特徵[17]。

## 3. 研究方法

### 3.1 實驗流程

本實驗流程如圖1所示。首先使用 openSMILE 提取語音特徵並將資料正規化，之後設定語言的排列組合並切割訓練資料以及測試資料。接下來使用 DNN 加入瓶頸層提取瓶頸特徵並使用 SVM 進行訓練以建立模型，之後利用訓練後的模型進行測試，最後依照結果來分析此方法對於台灣情緒語料庫以及跨語言、跨語料庫的成效。

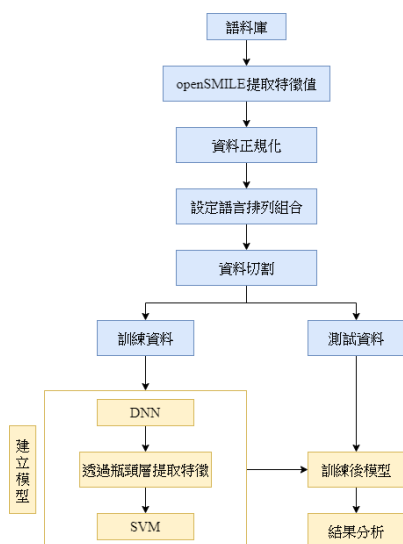


圖 1 實驗流程圖

### 3.1.1 openSMILE 提取語音特徵

本研究將使用「IS09\_emotion.conf」來提取語音特徵，該設定檔將會提取16個低階描述(Low-Level Descriptors, LLD)，該特徵檔包含 12個統計函數以及2個線性迴歸係數，因此每句話總共包含  $16 \times 2 \times 12 = 384$  個特徵參數。

### 3.1.2 資料預處理

依照[1]的研究，由於每筆資料間的數值差異非常大，所以首先使用方程式(1)將每筆特徵向量按比例縮小至0到1區間。之後將每個特徵向量使用方程式(2)。

$$\text{Scaling: } X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

$$L_1 \text{ Normalization: } \|x\|_1 = \sum_{i=1}^n |x_i| \quad (2)$$

### 3.1.3 設定語言排列

本研究將測試單一語言、混合語言、跨語言三種語言情緒實驗，表1、表2、表3的排列方法是以國語為例：

表 1 單一語言實驗排列

測試語言	國語
訓練語言	國語

表2 混合語言實驗排列

測試語言	國語
未加入 Emo-DB	
訓練語言	國語+台語
	國語+客語
	國語+台語+客語
加入 Emo-DB	
訓練語言	國語+德語
	國語+台語+德語

	國語+客語+德語
	國語+台語+客語+德語

表3 跨語言實驗排列

測試語言	國語
未加入 Emo-DB	
訓練語言	台語
	客語
	台語+客語
加入 Emo-DB	
訓練語言	德語
	台語+德語
	客語+德語
	台語+客語+德語

### 3.1.4 資料切割

測試單一語言將資料集隨機分為70%的訓練資料和30%的測試資料。混合語言是將主要語言的資料集隨機分為70%的訓練資料和30%的測試資料後，依照3.1.3的語言排列，在切割出的訓練資料中加入不同語言的全部資料作為訓練資料。而跨語言則是根據3.1.3的語言排列，使用不同的語言當作測試資料，其他語言則按排列組合做為訓練資料。訓練資料用來訓練資料建立模型，測試資料則使用訓練後的模型來進行測試。

### 3.1.5 建立模型

DNN 使用 Keras 中的 model 套件進行建置，透過 openSMILE 提取出的特徵檔為6373個，因此輸入層的節點數為6373，隱藏層節點數使用300~1500個輪流進行測試，隱藏層層數使用5層~2層進行測試，最終測試發現普通 DNN 使用兩層隱藏層並每層使用900個節點的效果最佳。

瓶頸層使用五層隱藏層分別輪流當瓶頸層以尋找出最佳的瓶頸層位置，瓶頸層的結點數須小於其他隱藏層的結點數，因此瓶頸層節點數使用100~800個輪流進行測試，測試後使用第二層隱藏層作為瓶頸層，第一層隱藏層維持使用900個神經元，而瓶頸層則經過測試後使用500個神經元，之後將瓶頸層後的隱藏層刪除提取瓶頸特徵至 SVM 進行訓練。

### 3.1.6 結果分析

將測試資料輸入訓練完成的模型中做測試後，比較不同語言、不同實驗的單一語言、混合語言、跨語言的辨識結果，並與先前的研究做比較，查看研究在各語言的辨識度是否有提升。

## 4. 實驗結果

### 4.1 單一語言情緒辨識實驗結果

單一語言情緒實驗辨識率彙整如表4所示，得到的最佳結果分別為國語情緒76.12%、台語情緒58.75%、客語情緒72.72%。最佳結果均是使用加入瓶頸層的 DNN-SVM。

表4 單一語言情緒實驗辨識率

測試/ 訓練語言	國語	台語	客語
SVM	21.21%	16.53%	15.70%
DNN	65.85%	48.03%	61.15%
DNN (瓶頸層)	69.37%	53.60%	70.24%
DNN-SVM	68.85%	53.54%	67.23%
DNN-SVM (瓶頸層)	76.12%	58.75%	70.72%

### 4.2 混合語言情緒辨識實驗結果

混合語言情緒辨識實驗各語言結果彙整如表5、表6及表7所示。在台灣情緒語料庫之混合語言情緒辨識實驗中得到的最佳結果為國語76.51%、台語62.97%、客語72.48%。在加入 Emo-DB 之跨語料庫混合語言情緒辨識實驗中，得到的最佳結果為國語62.12%、台語55.62%及客語64.46%。最佳結果均是使用加入瓶頸層的 DNN-SVM。

表5 國語混合語言情緒實驗辨識率

未加入 Emo-DB				
	國台	國客	國台客	
SVM	19.86%	19.86%	26.51%	
DNN	65.27%	65.36%	67.31%	
DNN 瓶頸層	70.11%	68.55%	71.49%	
DNN-SVM	71.78%	68.33%	75.75%	
DNN-SVM 瓶頸層	75.46%	71.37%	76.51%	
加入 Emo-DB				
	國德	國台德	國客德	國台客德
SVM	13.01%	15.60%	15.60%	18.45%
DNN	45.45%	55.13%	53.22%	58.80%
DNN 瓶頸層	47.79%	57.32%	55.16%	60.48%
DNN-SVM	54.54%	59.05%	57.37%	58.33%
DNN-SVM 瓶頸層	56.06%	61.36%	59.09%	62.12%

表6 台語混合語言情緒實驗辨識率

未加入 Emo-DB				
	台國	台客	台國客	
SVM	14.17%	14.17%	17.23%	
DNN	49.33%	44.58%	54.97%	
DNN 瓶頸層	53.55%	48.57%	55.68%	
DNN-SVM	52.84%	47.63%	55.21%	
DNN-SVM 瓶頸層	54.97%	50.94%	62.97%	
加入 Emo-DB				
	台德	台國德	台客德	台國客德
SVM	14.17%	14.17%	17.23%	14.17%
DNN	49.33%	44.58%	54.97%	49.33%
DNN 瓶頸層	53.55%	48.57%	55.68%	53.55%
DNN-SVM	52.84%	47.63%	55.21%	52.84%
DNN-SVM 瓶頸層	54.97%	50.94%	62.97%	54.97%

表7 客語混合語言情緒實驗辨識率

未加入 Emo-DB				
	客國	客台	客台國	
SVM	14.17%	14.17%	17.23%	
DNN	49.33%	44.58%	54.97%	
DNN 瓶頸層	53.55%	48.57%	55.68%	
DNN-SVM	52.84%	47.63%	55.21%	
DNN-SVM 瓶頸層	54.97%	50.94%	62.97%	
加入 Emo-DB				
	客德	客國德	客台德	客國台德
SVM	14.96%	14.96%	14.96%	18.11%
DNN	39.37%	40.49%	47.77%	48.03%
DNN 瓶頸層	43.24%	47.11%	51.96%	51.96%
DNN-SVM	40.12%	47.07%	49.32%	50.75%
DNN-SVM 瓶頸層	45.66%	51.18%	52.75%	55.62%

### 4.3 跨語言情緒實驗辨識結果

跨語言情緒辨識實驗各語言結果彙整如表8、表9及表10所示。在台灣情緒語料庫之跨語言情緒辨識實驗中得到的最佳結果為國語67.35%、台語55.68%、客語54.61%。在加入 Emo-DB 之跨語料庫跨語言情緒辨識實驗中，得到的最佳結果分別為國語60.27%、台語49.76%及客語45.38%。最佳結果均是使用加入瓶頸層的 DNN-SVM。

表8 國語跨語言情緒實驗辨識率

未加入 Emo-DB				
測試語言	台語	客語	台語+客語	
SVM	13.24%	15.29%	13.24%	
DNN	58.90%	54.79%	60.04%	
DNN 瓶頸層	59.58%	57.30%	64.15%	
DNN-SVM	60.73%	56.84%	63.21%	
DNN-SVM 瓶頸層	61.87%	58.21%	67.35%	
加入 Emo-DB				
測試語言	德語	台德	客德	台客德
SVM	13.01%	13.01%	13.01%	13.01%
DNN	13.31%	43.83%	43.71%	54.30%
DNN 瓶頸層	14.15%	44.29%	46.80%	58.27%
DNN-SVM	13.69%	41.15%	44.66%	55.99%
DNN-SVM 瓶頸層	14.84%	44.54%	47.71%	60.27%

表9 台語跨語言情緒實驗辨識率

未加入 Emo-DB			
測試語言	國語	客語	國語+客語
SVM	15.16%	16.58%	17.16%
DNN	45.75%	43.03%	50.97%
DNN 瓶頸層	50.84%	47.63%	53.21%
DNN-SVM	49.45%	44.83%	51.79%
DNN-SVM 瓶頸層	53.55%	48.57%	55.68%
加入 Emo-DB			
測試語言	德語	國德	客德
SVM	13.27%	13.27%	13.27%
DNN	14.92%	35.61%	37.91%
DNN 瓶頸層	14.92%	40.28%	40.18%
DNN-SVM	14.45%	38.03%	38.86%
DNN-SVM 瓶頸層	15.63%	41.23%	42.24%

表10 客語跨語言情緒實驗辨識率

未加入 Emo-DB			
測試語言	國語	台語	國語+台語
SVM	15.16%	16.58%	17.16%
DNN	45.75%	43.03%	50.97%
DNN 瓶頸層	50.84%	47.63%	53.21%
DNN-SVM	49.45%	44.83%	51.79%

DNN-SVM 瓶頸層	53.55%	48.57%	55.68%	
加入 Emo-DB				
測試語言	德語	國德	台德	國台德
SVM	13.27%	13.27%	13.27%	13.27%
DNN	14.92%	35.61%	37.91%	40.83%
DNN 瓶頸層	14.92%	40.28%	40.18%	44.54%
DNN-SVM	14.45%	38.03%	38.86%	47.86%
DNN-SVM 瓶頸層	15.63%	41.23%	42.24%	49.76%

#### 4.4 各語言情緒實驗結果

本研究未跨語料庫各實驗最佳結果詳細統整如表11所示，結合 Emo-DB 跨語料庫各實驗最佳結果如表12所示。

表11 未跨語料庫各情緒最佳實驗辨識率

測試語言	單一語言	混合語言	跨語言
國語	76.12% 國語	76.51% 國+台+客	67.35% 台+客
台語	58.75% 台語	62.97% 台+國+客	55.68% 國+客
客語	70.72% 客語	72.48% 客+國+台	54.61% 國+台

表12 跨語料庫各情緒最佳實驗辨識率

測試語言	混合語言	跨語言
國語	62.12% 國+台+客+德	60.27% 台+客+德
台語	55.62% 台+國+客+德	49.76% 國+客+德
客語	64.46% 客+國+台+德	45.38% 國+台+德

#### 5. 結論

本研究的研究目的有兩點，第一點透基於[1]的研究，並提出修改使其能夠應用於台灣情緒語料庫的情緒辨識，進而有效提升過去研究的情緒辨識率。

本研究的方法，用於國語情緒辨識在單一語言情緒辨識獲得76.12%的辨識率、混合語言情緒辨識獲得76.51%的辨識率、跨語言情緒辨識67.35%的辨識率；台語情緒辨識在單一語言情緒辨識獲得58.75%的辨識率、混合語言情緒辨識獲得62.97%的辨識率、跨語言情緒辨識55.68%的辨識率；客語情緒辨識在單一語言情緒辨識獲得70.72%的辨識率、混合語言情緒辨識獲得72.48%的辨識率、跨語言情緒辨識54.61%的辨識率，根據圖2、圖3、圖4的與過去研究的對比可以發現以上三種語言的

三種實驗均比過去研究的辨識率有所提升。

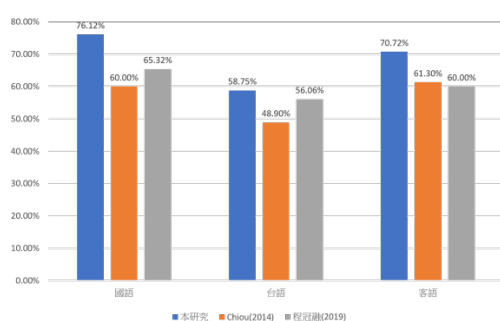


圖 2 單一語言情緒辨識比較圖

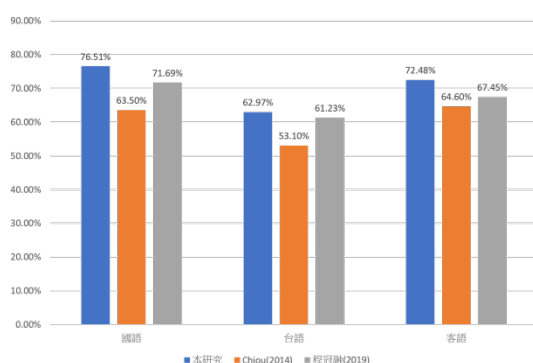


圖 3 跨語言情緒辨識比較圖

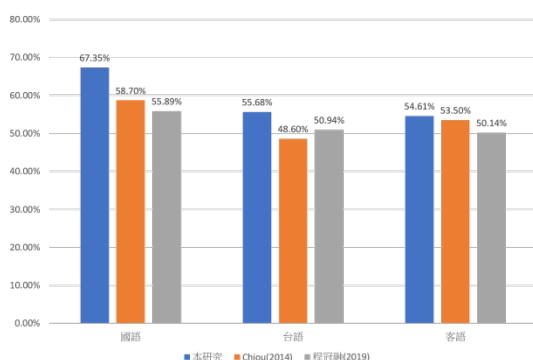


圖 4 混合語言情緒辨識比較圖

前一章節的實驗結果可以發現加入瓶頸層的 DNN-SVM 的結果會高於其他方法，因此本研究認為透過在 DNN 加入瓶頸層來提取瓶頸特徵識至 SVM 進行分類的確是可以有效提升辨識率的。

在進行混合語言情緒辨識時，發現加入不同語言資料進行混合語言訓練比使用單一語言進行訓練更能有效提升情緒辨識率，因此本研究推測若是能在台灣情緒語料庫中再加入更多的語音樣本，將可能使跨語言情緒辨識率有更進一步的提升。

本研究第二點目的是希望透過加入 Emo-DB 進行跨語料庫情緒實驗，過去研究認為使用不同語料庫的資料相互結合是能提升績效的，但本研究的實驗結果卻顯示加入 Emo-DB 後會導致各方面的辨識率下降，並且根據前一章節的實驗結果可以發現，在只使用單一語言進行跨語言實驗時，只使用德語的辨識率會大幅度低於只使用台

灣情緒語料庫內其他語言的辨識率，台灣語料庫內的國語、台語、客語均屬於亞洲語系，而 Emo-DB 使用到的德語為歐洲語系，因此本研究推測若是能使用一些同為亞洲語系的語料庫進行實驗，或許能提升跨語料庫實驗的辨識率。

## 參考文獻

- [1] Chen, R., Zhou, Y., & Qian, Y. (2017). Emotion recognition using support vector machine and deep neural network. In National Conference on Man-Machine Speech Communication (pp. 122-131). Springer, Singapore.
- [2] Hansen, J. H., & Cairns, D. A. (1995). Icarus: Source generator based real-time recognition of speech in noisy stressful and lombard effect environments. *Speech communication*, 16(4), 391-422.
- [3] Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F., & Huber, R. (2009). Detecting real life anger. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 4761-4764). IEEE.
- [4] Yildirim, S., Narayanan, S., & Potamianos, A. (2011). Detecting emotional state of a child in a conversational computer game. *Computer Speech & Language*, 25(1), 29-44.
- [5] Koolagudi, S. G., & Rao, K. S. (2012). Emotion recognition from speech: a review. *International journal of speech technology*, 15(2), 99-117.
- [6] Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4), 169-200.
- [7] Gupta, S., & Mehra, A. (2015). Speech emotion recognition using svm with thresholding fusion. *Proceedings of the 2nd International Conference on Signal Processing & Integrated Networks*, 570-574.
- [8] Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech communication*, 40(1-2), 33-60.
- [9] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572-587.
- [10] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In Ninth European Conference on Speech Communication and Technology.
- [11] Chiou, B.-C., & Chen, C.-P. (2014). An Emotional Speech Database in Taiwan: Collection and Recognition. *Proceedings of the 26th Conference on Computational Linguistics and Speech Processing*, 26, 21-30.
- [12] El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3), 572-587.
- [13] Zhang, W., & Zhao, D., & Chai, Z., & L.-T.ang., & X.Liu., & F.ong, S.Yang. (2017). Deep learning and svm-based emotion recognition from chinese speech for smart affective services *Softw. Pract. Exp.*, 47 (8), 1127-1138
- [14] 程冠融,施學琦 (2019)。應用語音辨識技術於跨語言及跨語料庫之語音情緒辨識。「TANET2019臺灣網際網路研討會」，高雄國際會議中心,245-250。
- [15] Gupta, S., & Mehra, A. (2015). Speech emotion recognition using svm with thresholding fusion. *Proceedings of the 2nd International Conference on Signal Processing & Integrated Networks*, 570-574.
- [16] Emre, Y., & Vikramjit, M., & Ganesh, S., & Horacio, F. (2019) Articulatory and bottleneck features for speaker-independent ASR of dysarthric speech, *Computer Speech & Language*, 319-334.
- [17] Eyben, F., & Wöllmer, M., & Schuller, B. (2010). Opensmile: the Munich versatile and fast open-source audio feature extractor *ACM International Conference on Multimedia*, 459-1462.