

## 中文語音測謊技術開發初探

曾春僑\*、高一瑛\*\*、曹昱\*\*\*

### 要 目

壹、前 言	參、研究方法
貳、文獻探討	一、模擬案例研究方法建構
一、聲壓分析	二、實測案例
二、與神經活動關係	三、語音資料分析方式
三、倒頻譜係數偵測	肆、結 果
四、動態語音特徵搭配各種演算法分析	一、基本語音特徵分析
五、搭配語音情感資料庫 檢測當事人壓力程度	二、模擬案件分析結果
六、聲學韻律特徵與母語 影響因子	三、實測案件分析結果
七、結合多模態偵測模式 之語音測謊	四、特徵值分布狀況
八、文獻綜合評析	伍、結論與建議
	一、誠實者與說謊者，其語 音特徵有所差異
	二、無法用單一語音特徵做 分析

DOI：10.6460/CPCP.202108\_(29).05

\* 臺灣警察專科學校副教授，中央警察大學犯罪防治研究所博士。

\*\* 法務部調查局科長，臺北醫學大學藥學研究所博士。

\*\*\* 中央研究院研究員，喬治亞理工學院電機電腦工程學系博士。

三、不同語者間，作為分類基礎之語音特徵有所不同

四、適當訪談情境為重要影響因素

五、如何建立當事人基本語音資料為後續重要基本工作

六、跨國語文使用者偵測問題

七、錄音品質影響程度尚待進一步釐清

## 摘 要

為因應國內測謊相關規範以及企業對謊言偵測需求，達到犯罪預防與偵查並重目的，非接觸式測謊技術之使用已成為必然趨勢，本研究基於此概念，開發出本土中文語音測謊雛形。受限於語音資料庫因素，本研究從個人化測謊概念進行設計，模擬試驗中，以不同時序收集27位，每位共六段誠實／說謊語音；實測案例，則由司法機關委託測謊案件中選取50件，根據測前晤談程序內容，區分為中性／相關／比對問題之語音。語音經過剪輯，保留受測者聲音後，以語音活性檢測、openSMILE語音情緒特徵分析與選取等步驟處理，後以區別分析判定各組差異性。結果顯示，僅以基本語音特徵作為說謊判別可能存在盲點；模擬案件之交叉驗證正確率可達約85%，實測案件甚至可達92%，檢視語音特徵分類狀況，可發現作為分類基礎之語音特徵大致相同，但每位語者存在細部差異，且系統選取之特徵會隨著錄音品質而有所不同，錄音品質較佳者可以選取較多特徵，但使用特徵數與正確率並未發現正相關。綜合前述結果，本研究認為，本研究使用的個人化語音分析方式，具有發展成為本土新式測謊系統之潛力，惟如何事先收集到當事人足資比對之語料則為下一步開發重點。

關鍵詞：測謊、語音、測前晤談、特徵選取、區別分析

# A Preliminary Research of Chinese Acoustic Lie Detection Technology Exploitation

Chun-Chiao Tzeng\* & Yi-Ying Kao\*\* & Yu Tsao\*\*\*

## Abstract

The polygraph test can be used for criminal investigation and prevention. Contactless lie detection has become a trend sooner rather than later in our country due to the requirement of polygraph policy and business management for lie detection. The goal of this research is to create a prototype of a native acoustic lie detection technology based on this concept. Due to the scarcity of a Chinese audio database, this study relies on the principle of personalized lie detection. In a simulation test, 27 participant sample data sets were obtained, containing truth and deception acoustic files in varied chronological sequences. In 50 genuine polygraph cases, audio files from the neutral/relevant/comparison questions during the pretest interview stage were

---

\* Associate Professor of Taiwan Police College; Ph.D., Graduate School of Crime Prevention and Corrections, Central Police University,

\*\* Section Chief of Ministry of Justice Investigation Bureau; Ph.D., College of Pharmacy, Taipei Medical University.

\*\*\* Research Fellow of Academia Sinica; Ph.D., Graduate School of Electrical and Computer Engineering, Georgia Institute of Technology.

collected to evaluate the real polygraph cases. Only the subjects' voices are preserved when the audio files are edited. These files are submitted to voice activity detection, openSMILE feature extraction and selection, and discrimination analysis during the data editing process. The results indicate that identifying the truth purely based on specific auditory features is unreliable. The accuracy rate of corss-validation is 85% in simulation cases and 92% in real cases. By examining the classification results, we can see that the acoustic characteristics used for categorization are roughly the same, although each person has subtle variances. The number of features selected varied with recording quality; more features were picked in higher recording quality files, but there is no positive correlation between the number of features selected and the accuracy rate. Based on these results, we believe that the personalized acoustic analysis methodology utilized in the study has the potential to be used for native lie detection. The subsequent research will concentrate on the subject's standard corpus collection for comparison.

**Keywords:** Lie Detection, Acoustic, Pretest Interview, Feature Seletion, Discriminat Analysis

## 壹、前言

說謊是故意誤導他人思考或認知的行為，所說內容，可以是無害或善意的謊言，也可以是導致他人權益損失的話語（DePaulo et al., 2003）。近年因為社會型態的改變，社會監督工具日趨多樣化，然隨著各種監控設備日漸發達，人類又有各種隱私權期待需求，面對他人追根究柢詢問時，更容易有說謊狀況出現，而辨別謊言的技術及方法，也隨著科學進步不斷演變。

辨別謊言方法甚多，準確率則因使用方法與收集資料多寡不同而有所差異，理論上凡與人類生理或心理訊號有關的現象，且可證明與謊言有所關聯的表徵均可作為監測標的，用於辨識說謊行為的研究主題，大致上可分為非口語行為（nonverbal behavior）、口語行為（verbal behavior）、副語言行為（paralinguistic behavior）及生理訊號（physiological reaction）等四大類。然實務上則會受限於操作難易度、人權、設備、經費、準確度要求等情況，而有不同的謊言偵測方法。這些偵測標的如呼吸、心跳、血壓、膚電、語詞、表情、聲調、肢體動作、腦波、身體各部位血流量、眼動、體溫等，然過去一百年來，使用最廣，也較可數據化的設備仍以測謊儀為大宗。

為建立測謊專業倫理，並將測謊技術、設備與教育訓練標準化，美國在1966年成立美國測謊協會（American Polygraph Association, APA），針對各種儀

器、設備、程序等作一連串規範，迄今已成為多數國家引用的標準，我國早期部分測謊技術源自日本，後來則陸續派員至美國受訓，故目前均以美國相關規範做基準，而迄今美國測謊協會主要仍以測謊儀搭配各種制式程序做規範；一般程序下，當事人在經測前晤談後，連接測謊儀，配合問題詢問步驟以收集當事人呼吸、心跳與心脈訊號，最後透過系統性分析測謊圖表以研判當事人是否說謊（National Research Council, 2003）。測謊儀的使用，雖然無法達到百分之百精確，但已有相當準確度，因此除了刑事案件偵查外，包括性侵害監控、反情報測試、任用前測試等，目前均仍以測謊儀進行（Grubin & Madsen, 2005）。

法律層面上，有學者認為測謊儀操作某種程度上仍會拘束到當事人行動，因此會有適法性問題，但實務層面又有大量測謊需求，尤其是偏向監控作為的各種預防措施，如忠誠測驗、員工監控等，需求日益增多，例如美國之員工測謊保護法（Employee Polygraph Protection Act, Public Law 100-347），雖明定原則上禁止對員工實施測謊，但對於包括公務機關（含地方政府、州政府及其附屬機構）、國防及安全體系、FBI契約承包商、私人企業已造成損失案件之調查、保全公司與醫藥安全體系之調查則屬相關除外條款，不受禁止測謊之限制（Onder & Brittan, 2009），代表在許多特殊行業，對安全之要求大過人權時，法律則允許對當事人之人權做一定程度限

縮，因此測謊技術在特殊領域仍有許多發展空間。

隨著測謊技術的發展，亦有部分人士開始傳播各種測謊反制方法，導致過去測謊儀的使用可能出現瓶頸，另一方面，隨著各種電腦科技的進步與生理檢測儀器的開發，在當事人未察覺下進行測謊的非接觸式測謊方法逐漸成為主流。如前述狀況，在我國亦有類似需求，一來刑事法令逐步限制測謊使用，二來反制測謊技術也在國內逐步傳播，三來則因半導體企業為我國重要命脈，但偶有員工將營業機密資料外洩，造成重大損失後之亡羊補牢措施又緩不濟急，在不能強制要求測謊情況下，如何透過一般訪談程序對當事人進行信用評估，亦為企業重要營運關鍵。基於此需求，非接觸式之聲紋測謊技術應運而生，但語言發音結構複雜，且牽涉到出生地、文化背景、使用習慣、生理結構差異等因素，導致發音重點有所不同。現代聲紋測謊研究多從大量語料庫內分析，找出可能的誠實或說謊特徵，但此模式在我國現有環境中有其限制，一來目前國內並未有足夠語料庫可以訓練預測模型，二來不同語文發音模式之差異亦須特別考量；故為發展適合我國使用之語音測謊雛形，本研究擬以個人化測謊為出發點，先從已知誠實與說謊語音資料中，透過各種分析工具軟體找出其差異性，並建立分析流程，後再以實際案例進行驗證，瞭解分析方法之實用性，希能在未來做到以聲紋分析協助犯罪預防、企業肅貪、性侵害監控與犯罪偵查之工作。



## 貳、文獻探討

利用聲音進行謊言分析之概念已發展許久，惟因各種技術瓶頸尚待突破，加上測謊儀已存在運作近百年，故在未能超越測謊儀準確率之狀況下，先前研究處於停滯狀態；惟近二十年來，因非接觸式測謊需求日增，加上電腦科技進步，故聲紋測謊技術研究開始蓬勃發展，與此有關之研究論文與概念列舉如下：

### 一、聲壓分析

或稱為語音壓力分析（Voice Stress Analysis, VSA），此為聲紋測謊早期之概念，主要偵測當事人說話時，輸出能量的變化。聲音訊號（Audio Signals）則泛指由人耳聽到的各種訊號，當發音體會產生震動，此震動會對空氣產生壓縮與伸張的效果，形成聲波，以每秒大約340公尺的速度在空氣中傳播，當此聲波傳遞到人耳，耳膜會感覺到一伸一壓的壓力訊號，內耳神經再將此訊號傳遞到大腦，並由大腦解析與判讀，來分辨此訊號的意義；人類發聲震動源為聲帶，其係由一連串大小不同肌肉構成，心理學上的3F概念認為，人遇到壓力時，會產生戰鬥、逃避或僵硬等三大反應，這些動作均由肌肉活動完成，又聲音之產生與肌肉活動有關，因此偵測肌肉活動相關指標，就可作為語音測謊之標的。

（一）此部分偵測的人類基本聲學特徵主要有三種，包括1.代表聲音的大小的音量（volume），此可由聲音

訊號的震幅來類比，因此又稱為能量（energy）或強度（intensity），此亦為聲紋測謊最早的偵測標的，在人類發聲器官中，通常代表肺部壓縮力量的大小，力量越大，音量越大，所以早期之聲紋測謊又常被認為語音能量偵測即屬此原因；2.代表聲音高低之音高（pitch），此可由基本頻率（fundamental frequency）來類比，為基本週期之倒數（fundamental period），此為後續開發之聲紋測謊，以及現代機器學習分析系統最基本之語音特徵成分，對應人類發聲器官，則代表聲帶震動的快慢，震動越快，音高會越高；3.代表聲音內容之音色（timbre），此可由每一個波形在一個基本週期的變化來類比，對應發聲器官，代表嘴唇和舌頭的位置和形狀，不同的位置和形狀，就會產生不同的語音內容，此部分就牽涉到各種語言與人種發音習慣之不同，這也是後續建立個人化測謊偵測模式時需要詳細分析之標的。

（二）肌肉微震動觀念（micro-muscle tremors, MMT）認為，在不同壓力狀況下，為應付隨時而來的變化，人的肌肉震動程度會有所不同，正常情況下，隨意肌的收縮正常時，MMT會出現以每秒大約10個週期小幅振盪。1972年三位美國陸軍退役軍官，從MMT理論發展語音測謊儀，其概念認為人在說謊時腎上腺素分泌增加，改變肌肉協調性，因為專注程度增加，降低了8~12Hz附近生理性震顫。Merck手冊中認為這些肌肉的生理顫動可能因為焦慮、壓力、疲勞、代謝失調或因某些藥物而變

動 (Beers & Berkow, 1999) , 而聲壓即是透過偵測聲音中各種語音傳遞時微弱的肌肉震動變化, 以推論其說謊程度。

(三)多數市面上的VSA儀器, 是偵測說話時 $f_0$ 的變化狀況, 藉由 $f_0$ 斜率變動差異判斷當事人承受壓力狀況, 學者評估市面上各種聲壓分析儀後, 認為這類儀器必須透過有經驗的測謊人員操作, 藉由會談情境的掌控與對波型觀察的正確詮釋, 才能獲得有效結果, 若僅是單純使用這類儀器是無法達到測謊之效果 (Hopkins et al., 2005) 。早期美國及以色列以此概念推出許多商業化產品, 例如Verimetrics之PSE、National Institute for Truth Verification之CVSA、Diogeres Group之Lantern、Makh-Shevet之Truster等, 均是基於此概念所開發。

(四)學者比較基於震動 (jitter) 與音高 (pitch) 強度變化兩種VSA之差異, 安排5位受測者透過紙牌吹牛遊戲, 全程錄影監視受測者是否說謊。該研究使用Praat軟體分析相關語音, 結果顯示MMT的正確率約為40.7%, 低於機率值以下, 無法有效偵測說謊; 而基於 $f_0$ 強度變化的偵測, 正確率能達70%左右, 顯示只單對某一頻率區間之分析效果並不顯著 (Liu, 2005) 。

(五)由於MMT現象通常出現於8~12Hz間, 而人在緊張時, 因為肌肉緊繃, 抑制了喉嚨對低頻聲音之產生, 這些低頻聲音人類本就無法察覺, 但透過頻譜儀檢測8~12Hz低頻聲音變化情況, 就可瞭解當事人處於緊

張或放鬆狀態。學者共安排四個實驗，第一個實驗要求受測者寫下各五種最喜歡與最不喜歡的食物、各五種最具價值與最討厭的五種特徵、五種最不想做的事情，並於測試時，刻意對這些問題以說謊回答；第二個實驗，則要求受測者在短時間內回答大量問題，蒐集其在面臨壓力情境下之回答狀況；第三個實驗，要求受測者融入特定情境，讓其對說謊的行為感到羞恥，之後再詢問特定問題細節；第四個實驗則進行實測，分析特定嫌犯之錄音檔。相關分析以Matlab軟體進行，實驗結果發現，當人類於壓力下之聲音，8~12Hz間的頻率分量會大幅度減小，但有些問題的回答並未出現預期的變化，亦即此規則並非適用於所有人之說謊反應（Cósetl & López, 2011）。

## 二、與神經活動關係

透過文獻分析，認為當事人神經活動程度會與個體發展程度有關，故 $f_0$ 的變化亦會與心電圖、呼吸、血壓、膚電（GSR）以及腦電波活動變化有關，進一步研究發現語音特徵中，性別可能是重要影響因素，尤其是女性在月經週期間對包含語音有關之事項特別敏感，因此認為進行語音相關研究時，需考慮原本性格、壓力類型、基頻位置、身體結構與性別等特性因子（Gidden et al., 2013）。

### 三、倒頻譜係數偵測

(一)梅爾倒頻譜 (Mel-Frequency Cepstrum, MFC) 為一個可用來代表短期音訊的頻譜，而梅爾倒頻譜係數 (Mel-Frequency Cepstral Coefficients, MFCC) 則是一組用來建立梅爾倒頻譜的關鍵係數，1980年代時，S.B. Davis和Paul Mermelstein提出此觀念，進一步用於語音辨識分析上。梅爾倒頻譜係數之取得步驟為訊號輸入、取音框、計算能量、凸顯高頻共振峰之預強調、將音框間左右連續性增加之漢明窗、快速傅利葉 (Fast Fourier Transform, FFT) 轉換，利用三角窗函式 (triangular overlapping window)，將頻譜對映至梅爾刻度取對數，再取離散餘弦轉換後取得。

(二)我國自2005年2月修訂性侵害犯罪防治法增訂「受保護管束之加害人經評估應接受身心治療或輔導教育者，觀護人得報經檢察官、軍事檢察官之許可，對其實施測謊。」之條文，導入測謊技術作為降低性犯罪再犯率手段之一。為評估聲紋測謊實用性，研究人員採用心理壓力評估與聲音壓力分析延伸之技術，以一個人說話時聲音的振幅，分析語音訊號是否能檢測出當事人緊張程度，再進一步判斷當事人是否說謊。研究以嘉義地區一般役男8人及在監性罪犯12人為第一及第二對照組，假釋或緩刑性罪犯8人為觀察組。參考測謊題組之修正一般問題技術 (MGQT, Modified General Question Technique)，在實施儀器測謊時，同步以錄音筆錄音，

後以MATLAB軟體偵測音頻與音調，再用儀器測謊作為效標分析聲紋測謊資料。研究發現三組中均各有2位說謊者，與之對照未說謊者，發現三個參數中只有MFCC的辨別度最高，大致上呈現一定的趨勢，比較個人差異性時，當MFCC值較低代表此人是說實話，反之MFCC值偏高代表受測者說謊；而音調輪廓PC (pitch contour, PC) 和音調輪廓數FC (frame number of pitch contour, FP) 未能充足反應說謊狀況，主要原因在於PC與FP都是取平均值，特別是此兩項參數代表音調輪廓，所以在平均過程中，會將許多特徵與特性平均掉 (林明傑等，2005)。

(三)針對分數梅爾倒譜係數 (fractional Mel cepstral coefficient, FrCC) 偵測謊言研究發現，不同的分數階層可以預測語者個性，再配合線性分析法 (LDA)、隱藏馬爾可夫模型 (HMM) 分析，發現最優分數階層效果配合FrCC模型，準確度可以大幅提升。在LDA模式下使用FrCC分別為59.9%和56.2%，使用HMM時男性和女性的準確度可以提高到71.0%和70.2%。此外引入FrCC時，某些個人準確度可以提高20%以上，使得準確率超過85% (Pan et al., 2015)。

#### 四、動態語音特徵搭配各種演算法分析

##### (一)決策樹演算法

決策樹是最簡單的機器學習算法，預測時，在樹的內部節點處用某一屬性值進行判斷，根據判斷結果決定進入哪個分支節點，直到到達葉節點處 (最終處)，得

到分類結果。清華大學透過網路招募20位學生，先請當事人填具相關問卷，後以遊戲方式蒐集受試者說謊與誠實之語音資料，錄音內容以數位訊號處理方法，透過Praat軟體擷取相關語音特徵，包括平均共振峰（average formant）、短時間聲音強度（short-time intensity）、音高（pitch）、平均靜音比例（average silence-ratio）、平均講話速率（average speaking rate）、音高／聲音強度輪廓（pitch/intensity contour）、平均音高／聲音強度變化率（average pitch/intensity changing rate）、音高／聲音強度範圍（pitch/intensity range）、平均音高／聲音強度下降率（average pitch/intensity dropping rate）、時長（duration）等24個特徵值，搭配用Python 機器學習套件Scikitlearn決策樹與迴歸樹演算法分析，最後以決策樹學習的訓練模型，對一段未知的語音辨別真偽；結果顯示第一至第三共振峰、平均基頻、基頻中位數、最小聲音強度值、平均聲音強度值與聲音強度中位數等特徵被選取次數相對較多，其中共振峰部分，研究者搭配訪談結果，認為是當事人說謊時講話可能含糊不清所致，而共振峰又和發音時的口腔形狀關係密切；利用聚合式階層分群法將受試者行為向量分群後發現，可將同類別受測者區分為3~4個族群；個人化分析中，當分為三群時，平均辨識率介於0.57~0.73間，分為四群時，平均辨識率則介於0.52~0.73間。整體模型部分，特徵權重分配後之辨識率為0.56（吳乙彤，2018）。

## (二)非線性演算法

中央大學實驗中，先給受試者問題清單，請受試者填上正確答案後，再請受試者隨機對所填答案進行誠實或說謊回答，就其回答與實際作答內容做誠實與說謊之檔案分類，每筆錄音資料約4秒鐘，取樣頻率為48,000取樣點／秒，每筆僅取2.5秒錄音，並重複三次以增加訓練量，再將所擷取語音參數轉換成梅爾倒頻譜係數，總共收集20個人，共200筆資料，以回聲狀態網路（Echo State Network, ESN）為架構，隨機抽取其中15個人錄音資料作為訓練資料集，其餘5人之錄音資料作為測試資料集，共隨機抽樣10次。並將其結果與受試者填寫於問題列表清單之答案做比較，作為其判斷正確性依據，結果顯示其正確率為65%（劉丁瑋，2017）。

## (三)雙向長短期神經記憶

與單層類神經網路相比，深度類神經網路具有更好的分類與回歸能力，此模型是由多層的連接層所構成，訓練時是以倒傳遞方式對各層模型參數進行調整。詳細的流程如下述：得到初始權重後，由輸入層開始，透過前傳遞逐步求得各層的輸出，直到最後的輸出層後，基於與正確答案的差距算出誤差函數，再依此誤差值往回調整各層參數；鑑於說謊者相關線索除在語音中外，也可能受詢問者對話方式影響，因此學者以BLSTM-DNN為架構設計一套包含一般聲學特徵與雙向對話變化特徵的分析系統，其中雙向長短期記憶（Bidirectional Long



Short-Term Memory, BLSTM) 技術常被用來提取整句話的特徵，深度學習神經網路 (Deep Neural Networks, DNN) 則是透過模仿生物神經系統的數學模型，進行不同階層與架構的多次運算和訓練，找出最佳化、最有效的語音學習模型。實測顯示此系統對於繁體中文對話之謊言偵測率達74.71%，故學者認為除語音特徵外，若再加入雙方對話過程如語調、語速等特徵共同分析，對於提升準確度有所幫助 (Chou et al., 2019)。

## 五、搭配語音情感資料庫檢測當事人壓力程度

(一)分析語音情緒時，需要先對基本情感項目做分類，目前語音情感資料庫，主要均從Ekman於1972年所提出之項目再延伸，當時提出之參考項目包括憤怒 (anger)、厭惡 (disgust)、恐懼 (fear)、快樂 (happiness)、悲傷 (sadness)、驚訝 (surprise)，在進行語料收集時一般會再加入中性 (neutral) 特徵，而語音情緒表現會因地域、語種、年齡不同而有所差異，因此使用時須考量資料庫合適性，目前常用的幾個情感資料庫，如Belfast英語情感資料庫係由40位錄音者用五種情感傾向進行演講後建立；柏林Emo-DB情感資料庫由10位演員以回憶自身真實經歷或體驗進行情緒醞釀後，表達出七種情感；FAU AIBO兒童德語情感資料庫請51名10-13歲兒童，透過與索尼公司的AIBO機器狗進行自然互動，從而進行情感資料的收集；CASIA漢語情感資

料庫則由兩男兩女透過演講得到的六種情感<sup>1</sup>；ACCorpus系列漢語情感資料庫由50位錄音者各自表演五種情感；NNIME中文情緒互動多模態語料庫則請44位演員透過戲劇表演呈現情緒，錄製了將近11小時的聲音、影像與心電圖資料，請共49個標記人員進行六類情緒與PAD標記而得<sup>2</sup>。

(二)日本以柏林情感數據庫(EmoDB)、慶應大學日語情感語音數據庫(KeioESD)、賴爾森語音歌曲視聽數據庫(RAVDESS)為訓練與測試標的，透過偵測語詞中之平均能量、強度、與梅爾頻率倒譜係數搭配機器學習檢測語音中的壓力，發現使用神經網絡可獲得最佳結果，壓力檢測準確度分別為97.98%(EmoDB)，95.83%(KeioESD)和89.16%(RAVDESS)，顯示壓力、情緒變化與語音間存在一定程度關係(Tomba et al., 2018)。

## 六、聲學韻律特徵與母語影響因子

(一)哥倫比亞大學提出的欺騙檢測系統是以聲學韻

---

<sup>1</sup> 由中國科學院自動化研究所錄製，包括4位錄音人(2男2女)在純淨錄音環境下(信噪比約為35db)分別在五類不同情感下(高興、悲哀、生氣、驚嚇、中性)，對500句文本進行的演繹所得，共有960句。

<sup>2</sup> 國立清華大學——國立台灣藝術大學中文情緒互動多模態語料庫(NTHU-NTUA Chinese Interactive Multimodal Emotion Corpus, NNIME)為一個指定情境而無語句腳本的自發性情緒語料庫，內部包含許多不同如笑聲、哭聲、氣音等自然情感對話所具有的非口語聲音片段。

律當作特徵參數，實驗結果證實聲學韻律是判斷欺騙與否的重要特徵，為釐清不同文化、性別之間因素是否影響檢測系統準確性，故在收集語料時，也一併記錄受測者人格因素、性別、種族和信心等級，最後將這些因素及語音訊號綜合判斷真實／說謊語音的正確性，結果發現，將語音訊號與說話人性格、性別有關參數相結合後，可得到最佳的辨識度（Levitan et al., 2015）；為瞭解何種語音特徵造成語音辨識謊言時之跨文化差異，以及是否需要考慮性別與母語間之特性，因此進一步以個人化欺騙檢測系統作分析，實驗結果證實，在個人化系統下，聲韻特徵在分類欺騙性言語方面非常有效（Levitan et al., 2018）。

（二）母音共振峰分析：此係由前述MMT理論延伸之研究，且為少數針對國語注音符號之母音進行分析，研究分為baseline和deception兩階段，在baseline階段，測試者詢問受測者一些簡單問題，如早餐種類，目的為蒐集心情平靜且真誠無隱瞞的聲音，類似測謊之中性問題概念；deception階段要求受測者選出10張不喜歡的圖片，再假裝很喜歡這些圖片，並說出喜歡的原因，營造當事人說謊情境。全程採用中文提問與答覆，再將錄音檔切割出所對應的聲音區塊加以分析。錄音檔用Praat軟體將16個母音ㄚ ㄛ ㄜ ㄝ ㄝ ㄞ ㄟ ㄠ ㄡ ㄢ ㄣ ㄤ ㄨ ㄩ ㄚ ㄛ ㄜ ㄝ ㄞ ㄟ ㄠ ㄡ ㄢ ㄣ ㄤ ㄨ ㄩ 進行時頻分析（JTFA）轉換以獲得個別母音的不同共振峰 $f_0$ 、 $f_1$ 與 $f_2$ ，並請一位執業14年的心理諮商師重新聽二階段的

聲音，採五分法來判讀表裡不一的程度。該研究對象共有30人，結果顯示，以 $f_1, f_2$ 之座標所圍成之向量面積，在30人中有25人呈現 $Dec\_area > Bas\_area$ ，且統計檢定 $p = .01$ ，代表此向量面積可當成識別表裡不一的指標，亦即說謊與否之判斷；在不考量樣本數量條件下，發現隨著學歷下降， $p$  value有逐漸上升的趨勢，統計檢定顯示，隨著學歷越高，對說謊特徵掌握度就越好（陳世烽，2017）。

### 七、結合多模態偵測模式之語音測謊

（一）結合動作影像（3D-CNN）、語詞、語音（openSMILE）、微表情的測謊系統，並以多模態（multimodal）與單峰模型（unimodal）做預測，發現MLPC之收斂速度較快，在少量樣本（121件）下發展的模型，具有九成以上準確度（Krishnamurthy et al., 2018）。

（二）馬里蘭大學提出了一種使用錄影檔案中可用信息進行欺騙檢測的系統。在視覺方面，所架構的系統結合常用動作辨識以及人臉微表情辨識，可有效預測錄影檔案中的欺騙行為。在語音部分，該系統提出以MFCC為特徵參數的辨識技術，結合視覺及語音辨識，可獲得比人類自身更好的判斷結果（Wu et al., 2018）。

### 八、文獻綜合評析

（一）測謊需搭配良好會談與題目建構：相關文獻

中，部分研究並不支持聲壓測謊之有效性，這兩部分主要差異在於，多數認為聲壓可運用於測謊之實驗，多由測謊人員所操作，故某一程度上，顯示良好的對談與問題設計，並配合測謊相關／比對問題的設計概念衍生出的晤談方式，可能是影響聲壓能否用於測謊的重要因子。根據國內清華大學的研究，發現語音特徵之說謊辨識率最高為73%，在使用機器學習權重最適當分配情況下，此辨識率尚有向上改善空間，檢視其研究內容，雖有透過特徵選取與機器學習等分類步驟，但並未使用儀器測謊之測前晤談步驟，且選用特徵數較少，故推測訪談設計與特徵數量選取，可能為語音測謊成功重要關鍵因素。

(二)訓練資料集大小亦可能為重要因素：檢視前述研究，部分研究之訓練集時間過短，如回答是／否，或是僅錄製一段話，且部分研究因為取樣時間過短，因此用重複取樣，或將同一段語詞重複擷取多次方式處理，在取樣數太少下，推測亦為正確率未能提升原因之一。

(三)個人化差異是重要考量因素：根據聲韻特徵與跨種族、文化關聯性分析發現，除聲學韻律是判斷欺騙與否的重要特徵外，配合個人化系統分析下，更能凸顯聲韻特徵之效用，主要原因在於聲韻特徵與其所屬種族、人格因素、性別與個人自信心有關，因此將語音訊號與說話人性格、性別有關資訊相結合，轉化為個人語音特徵，更能提升辨識效率。而根據清華大學之研究亦

發現，整體訓練模型測試結果，辨識率僅為0.56，然個人化模型辨識結果，可以達到0.73，亦顯示建立個人化模型是重要關鍵因素之一。

(四)大量運用各種學習理論、語料庫與分析模型：近幾年趨勢為將各種學習理論建構之模型大量運用於語音分析上，而學習資料來源之一為各種語音資料庫，除此之外，在機器學習技術廣泛使用後，就算可擷取之語音特徵數量較少，亦會套用各種機器學習模型進行分析，此亦已經成為大量普遍使用之技術。

(五)綜上所述，目前趨勢為使用各種語料庫，找出整體族群說謊特徵，而相關實驗結果又發現個人文化、種族背景等為重要考量因子，進一步研究這些分析模式發現，在我國狀況下仍有部分窒礙難行處，舉例如下，1.我國目前僅陸續建立語音情緒資料庫，尚未建立各種說謊資料庫，而中文語音情緒與說謊連結程度，尚未有深入研究，2.國外多數實驗並未以認知訪談為基礎進行或經測前會談後再詢問與設計題目，此與測謊實際作業有所不同，且分析時所收集語音長度不足。在此前提下，為發展適合我國使用之語音測謊雛形，因此本研究排除語音資料庫運用模式，改從個人化比對需求出發，搭配記憶增強認知訪談模式，希望能建構出本土語音測謊雛形。

## 參、研究方法

研究方法區分為模擬實驗建構、實際案例收集比對，與各種語音資料分析方法建構三大區塊。

### 一、模擬案例研究方法建構

仿照測謊測前晤談步驟，與後續儀器測試、數據分析中之比對問題／相關問題設計意旨，本研究需收集受測者誠實與說謊情況下的各種訊號以資比對，故以誠實、說謊兩類題組為主軸設計

(一)誠實內容：係針對受測者原本親身經歷，僅需稍微回憶即可自由敘述的內容作詢問，亦即較輕易提取腦中記憶之資料。

1.順時間序列自述今日生活，2分鐘。

2.逆時間序列自述今日生活，2分鐘。

3.針對前面提到的重點，由訪談者逐一去詢問更詳細內容，約4分鐘。

(二)說謊內容：係針對受測者未親身經歷，需事先思考建構的虛擬內容，例如殺人、偷竊軍火等內容作詢問，屬較須思考，認知負荷較多之資料。

1.順時間序列自述虛擬事件經過，2分鐘。

2.逆時間序列自述虛擬事件經過，2分鐘。

3.針對前面提到的重點，由訪談者逐一去詢問更詳細內容，約4分鐘。

(三)每位受測者之蒐集資料內容共計六段，包括誠

實三段，說謊三段，誠實檔案部分可作為比較之基礎。為便利後續分析，相關語音以Audacity錄音軟體錄製。

(四)將每一案件基本資料建檔，包括案類、涉案刑度、當事人基本背景資料（如年齡、性別、教育程度）、涉案屬性（如關係人、加害人等）編碼後，逐一輸入SPSS（Statistical Product and Service Solutions）軟體或其他適用分析系統之資料庫建檔。

## 二、實測案例

選取調查局實際測謊案件中測前晤談的內容，依詢問內容分別選取不同問題之對話，包括比對／相關／中性問題之討論內容，擷取音檔後，交叉比對準確率。此部分因資料來源為實際測試案例，考量到實際偵審過程中，可能因為程序正義、證物採驗、緩起訴、當事人和解等不同狀況，導致測謊結果與法院審判結果有所差異，故本研究誠實與說謊之判定基礎，以實務與學術機構中，持有美國測謊學會認可訓練機構完訓證書之人員共4人，分別根據案情內容與圖譜反應據以判定，4人針對案情內容與圖譜均達成說謊或誠實一致結論之案件方進行分析。

## 三、語音資料分析方式

模擬案件部分，主要比對說謊／誠實陳述內容兩者語音情緒特徵之差異；實測案件部分為比較相關／比對／中性問題間之差異。



(一)語音活性檢測 ( Voice activity detection, VAD)：包括降噪，訊號區塊提取特徵，與使用分類器對區塊進行分類，瞭解是否為語音訊號。

(二)語音特徵初步分析：此部分針對音高 (pitch) 與語速 (speaking rate) 部分作分析。音高部分，假設語音訊號在音框內是穩定的，先以短時距分析 (short-term analysis) 將音訊切割成多個音框；在一個特定音框內，可以觀察到音框的音高，此音高是由基本頻率 (fundamental frequency) 來計算，亦即基本週期 (fundamental period) 的倒數，聲音的基本頻率越高，代表音高越高，反之則音高越低。語速主要是計算在固定時間中說話的字數，本研究使用語音活性檢測法，將一段話語音部分標記後，依照整段話的長度算出說話的語速，由於語音活性檢測是以能量計算，故本研究亦以語音能量大小判斷是否有語音存在，能量超過預定閾值時標記為Voiced，未超過則標記成Unvoiced，藉由計算所有Voiced的部分除以所有的語音長度，即可以算出語速。

(三)語者情緒特徵萃取：由於建構語音情緒辨識系統往往需要考量到資料庫、情緒類別、特徵擷取、分類器架構、評量指標等因素，因此本研究以現有之openSMILE軟體將每個音檔做情緒萃取，未另外建置語音情緒系統。

(四)特徵選取：由於openSMILE抽取的特徵過多，

許多屬於次級且較無意義特徵，故使用特徵選取技術減少不必要的情緒特徵值，以增加處理語音的速度，特徵選取以嵌入法（Embedded）進行，先以機器學習與模型進行訓練，得到各個特徵權值係數，再根據係數從大到小排列選擇特徵。

（五）差異性比較：將每個字的情緒特徵值視為任一座標，然後利用線性判別分析工具（Linear discriminant analysis）來判定分析不同字元間是否有差異性。

圖1

測試環境建構





依據。

(一)音高特徵分析：為瞭解音高特徵是否全面具有代表性，因此隨機選取數個對象進行音高特徵檢測，結果發現音高特徵不一定可作為說謊與否判定基準，下面列出其中一例無法以音高特徵做判定基準之分析結果。

1.音高特徵圖：物理意義而言，音高代表聲帶震動的快慢，震動越快，音高會越高。通常人類在情緒變化時，音高也可能有明顯變化，因此分析音高特徵與說謊是否有關，此目的亦在檢測早期部分聲壓研究報告中，認為說謊與聲壓存在一定關聯性之理論是否一定成立，或是是否僅存在於某些族群。音高係以基本頻率（fundamental frequency）來計算，當基本頻率越高，代表音高越高，反之則音高越低。針對語音活性檢測以及音高在各種不同特性的語音訊號上做觀察，結果如圖3～6所示，依序為語音波形圖、頻譜圖、語音活性檢測及音高於四句語音訊號的結果，此四句話分別來自誠實（順時鐘敘述）、誠實（逆時鐘敘述）、說謊（順時鐘敘述）、說謊（逆時鐘敘述）四種狀況。從圖中可知，基於單句話分析，無法直接看出語速及音高對於說話偵測有明顯效果。

圖3

誠實（順時鐘敘述）之波形圖、頻譜圖、語音活性檢測及音高結果圖

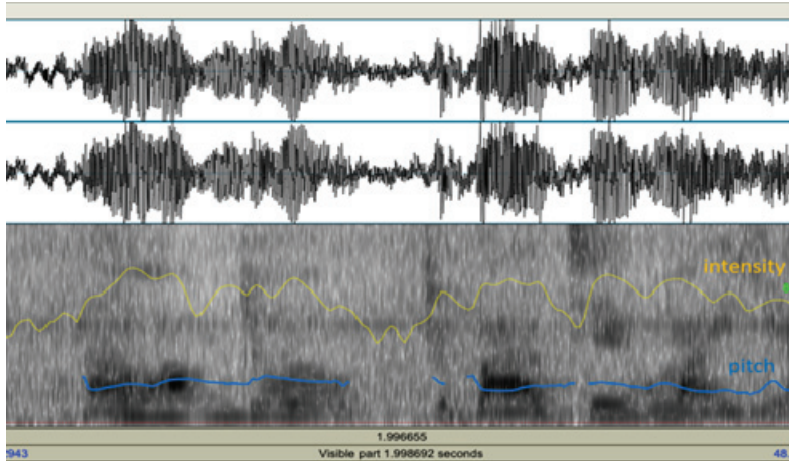


圖4

誠實（逆時鐘敘述）之波形圖、頻譜圖、語音活性檢測及音高結果圖

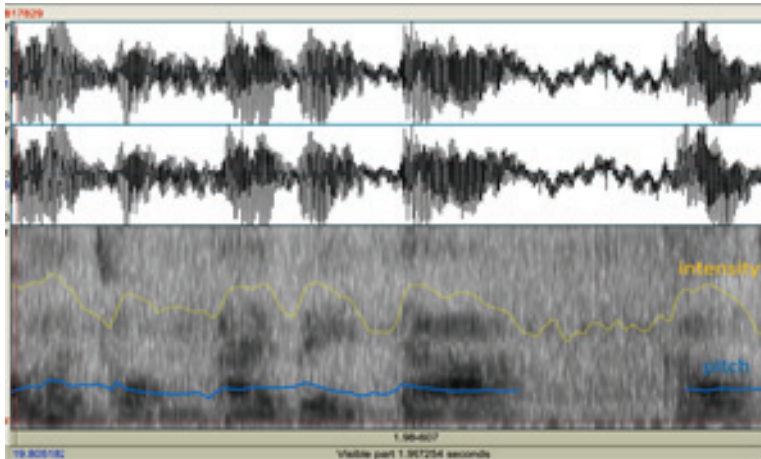


圖5  
說謊（順時鐘敘述）之波形圖、頻譜圖、語音活性檢測及音高結果圖

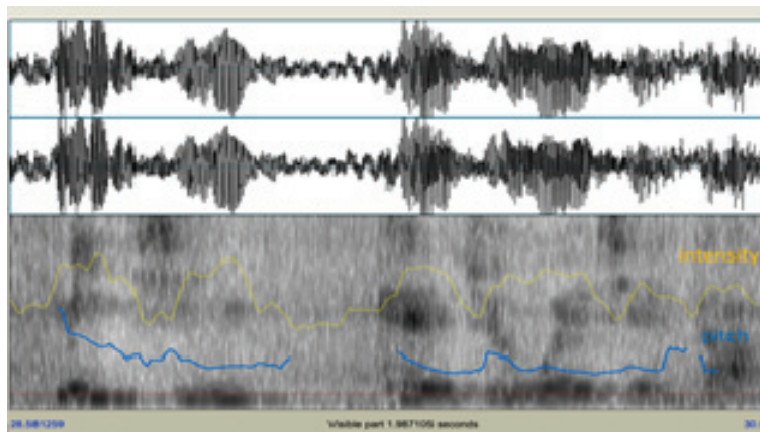
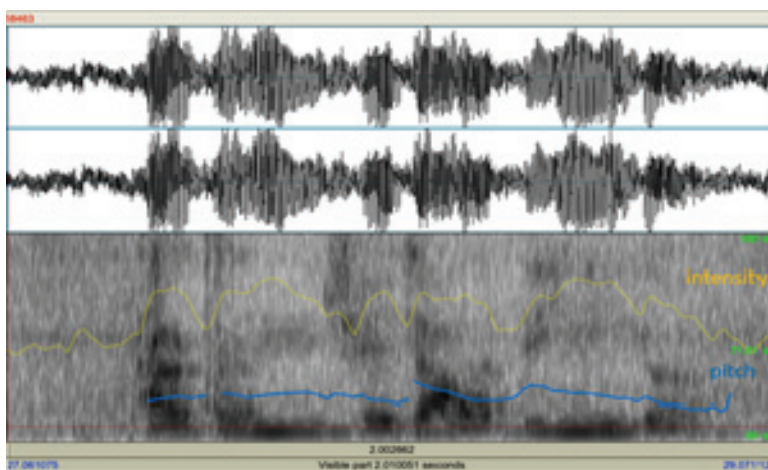


圖6  
說謊（逆時鐘敘述）之波形圖、頻譜圖、語音活性檢測及音高結果圖



2.音高統計值：相較於誠實（順時鐘敘述）語音，說謊（順時鐘敘述）時的音高平均值較高（182.967Hz vs. 175.110Hz）；而相較於誠實（逆時鐘敘述），說謊（逆時鐘敘述）時的音高平均值反而較低（160.285Hz vs. 170.581Hz），基於實驗結果可以發現此受試者在說謊與誠實狀態的音高平均值並無明顯區別。由於音高的變化反映在標準差的數值上，進一步分析音高變化標準差可發現，相較於誠實（順時鐘敘述）狀態，受測者在說謊（順時鐘敘述）時的音高標準差較高（51.599Hz vs. 44.948Hz）；另相較於誠實（逆時鐘敘述）狀態，說謊（逆時鐘敘述）時的音高反而較低（22.087Hz vs. 46.143 Hz），基於實驗結果可以發現此測試者在說謊跟誠實的音高標準差並無明顯分別。實驗結果顯示，基於音高特徵對該語者是否說謊的鑑別度不大。

表1

四種狀況下的音高分析結果，包括1.誠實（順時鐘敘述），2.誠實（逆時鐘敘述），4.說謊（順時鐘敘述），5.說謊（逆時鐘敘述）

	1誠實， 順向	2誠實， 逆向	4說謊， 順向	5說謊， 逆向
平均數 (Hz)	175.110	170.581	182.967	160.285
標準差 (Hz)	44.948	46.143	51.599	22.087
中位數 (Hz)	162.967	106.187	168.734	158.840

(二)語速特徵統計分析：隨機選取一人進行語速分析，相較於誠實（順時鐘敘述）狀態，該受測者於說謊（順時鐘敘述）時的語速較高（2.928 word/sec vs. 3.035 word/sec）；而相較於誠實（逆時鐘敘述）狀態，該受測者於說謊（逆時鐘敘述）時的語速也較高（2.744 word/sec vs. 2.942 word/sec），基於實驗結果可發現此測試者在說謊時（不論是順時鐘或是逆時鐘），語速都有所提高，代表語速可能可以當作說謊與否判斷特徵之一。

表2

四種狀況下的語速分析結果，包括1.誠實（順時鐘敘述），2.誠實（逆時鐘敘述），4.說謊（順時鐘敘述），5.說謊（逆時鐘敘述）

	1誠實， 順向	2誠實， 逆向	4說謊， 順向	5說謊， 逆向
語音總時數 (sec)	86.397	66.338	74.801	54.047
總字數	253	182	227	159
每字平均秒數 (不含沉默時間)	.315	.309	.319	.301
沉默總時數	6.807	10.088	2.395	6.168
每字平均秒數 (含沉默時間)	2.928	2.744	3.035	2.942

## 二、模擬案件分析結果

考量個人說話方式存在差異，且前述初步測試結果顯示，無法僅以單一特徵比對當事人是否說謊，故將所有音檔依據前述實驗方法進行分析。模擬案件中共收集27位受測者在不同狀況下錄製的語音片段，每段約2~5



分鐘；為使比較基礎相同，故挑選同時序之語音片段進行相互比較，在表3中，1-4、2-5、3-6分別代表誠實敘述語音群（第一群）與說謊語音群（第二群）配對比較結果；另為比較同一談話性質語音是否沒有差異，因此隨機將同一段語音分成兩群（第一／二群）進行比較。理想狀況為不同語段（誠實／說謊）具高預測率，代表誠實／說謊的確存在不同語音特徵；而同一語段具低預測率，代表同一語段內，語音特徵相同，歧異度低。而除檢驗預測正確率數值外，另以leave one out交叉分析法檢驗預測準確率，相關結果敘述如下：

（一）選取特徵數：不同語段下，不同個案與不同語段，系統選取特徵項目與數量有所不同，平均為60～85個特徵項目，此結果代表系統選取作為說謊判斷之發音特徵，會因受測者與談話內容而所不同。

（二）同一說話時序，不同陳述內容（說謊或誠實）之分類與預測結果：此部分共有三組，分別比較同一說話時序（正向／逆向／隨機），不同性質（誠實／說謊）之結果，由預測率可知，誠實原始分類正確率介於80.62～90.48%間，說謊原始分類正確率介於73.84～86.91%間；誠實之交叉驗證預測正確率介於79.15～85.98%間，說謊之交叉驗證預測正確率介於75.63～80.73%間。

（三）同一說話時序，相同陳述內容之分類與預測結果：此部分共有六組，分別比較同一時序下不同語段分

類與預測結果，由預測率可知，誠實之原始分類正確率介於76.37～90.48%間，說謊之原始分類正確率介於71.82～86.91%間；誠實之交叉驗證預測正確率介於63.5～68.41%間，說謊之交叉驗證預測正確率介於53.59～56.96%間。

(四)綜合上述結果，代表在原始分類中，說謊／誠實語段之預測尚未見到明顯差異，但在交叉驗證中，不同談話性質間可見明顯差異，相同談話性質間未見明顯差異，代表本研究使用之方法，能有效區分誠實／說謊談話內容，且相同談話性質間歧異度低，故為一良好分類預測方法。

表3  
模擬案件語音分析結果

比對內容	平均特徵數	第一群原始分類正確率 (%) <sup>c</sup>	第二群原始分類正確率 (%) <sup>c</sup>	第一群交叉驗證預測正確率 (%) <sup>c</sup>	第二群交叉驗證預測正確率 (%) <sup>c</sup>
1-4 <sup>a</sup>	69.12	90.48	86.91	<b>85.98</b>	<b>80.73</b>
2-5 <sup>a</sup>	59.12	83.73	80.63	<b>85.35</b>	<b>80.09</b>
3-6 <sup>a</sup>	85.44	80.62	73.84	<b>79.15</b>	<b>75.63</b>
1-1 <sup>b</sup>	69.12	80.78	75.57	66.11	56.96
2-2 <sup>b</sup>	69.12	76.37	71.82	64.75	53.59
3-3 <sup>b</sup>	85.44	82.20	76.55	63.50	57.27
4-4 <sup>b</sup>	59.12	83.51	79.17	67.55	56.53
5-5 <sup>b</sup>	59.12	76.96	72.65	68.41	55.03
6-6 <sup>b</sup>	85.44	90.48	86.91	64.12	56.58

a：分別代表1與4，2與5，3與6段之比對。

b：同一人，同段語音隨機切成兩群後比對。

c：若為不同人，第一／二群分別代表誠實／說謊；若為同一人，則第一／二群則由同一語段隨機分派。

### 三、實測案件分析結果

本研究共收集50件實測案例進行分析，考量這類案件之測前會談內容主要包括中性／比對／相關三種問題，因此擷取實測案件中之這三類問題之會談內容作為分析標的，並以原始測謊報告最終判定之說謊／誠實結果進行分組，以計算其準確度。

(一)選取特徵數：不同組合下選取語音特徵數不同，平均在30～46個特徵數間，且誠實組所選取特徵較說謊組多。

(二)比對／相關問題之分類與預測結果：此部分以誠實／說謊組做區分，在比對／相關問題中，說謊組之原始分類正確率分別為97.92%與99.01%，交叉驗證之預測正確率分別為87.01%與86.31%；誠實組之原始分類正確率分別為95.62%與98.52%，交叉驗證之預測正確率分別為81.36%與84.07%。而僅以單一問題種類做分析之結果，亦即將比對或相關問題隨機分成兩組，原始分類正確率介於80.76～86.03%間；交叉驗證之預測正確率則介於47.96～65.39%，明顯低於比對／相關問題之組合，顯示上述分類法，不論在誠實或說謊群組，均能將比對／相關問題正確區分。

(三)中性／相關問題之分類與預測結果：此部分亦以誠實／說謊組做區分，在中性／相關問題中，說謊組之原始分類正確率分別為98.58%與98.2%，交叉驗證之預測正確率分別為92.7%與86.32%；誠實組之原始分類

正確率分別為96.9%與97.5%，交叉驗證之預測正確率分別為88.95%與87.02%。僅以單一問題種類做分析結果，亦即將中性或相關問題隨機分成兩組，原始分類正確率介於77.38~88.65%間；交叉驗證之預測正確率則介於42.07~64.02%，亦明顯低於比對／相關問題之組合。這些結果代表上述分類法，不論在誠實或說謊群組，均能將中性／相關問題有效區分。

(四)綜合上述結果，代表使用中性／相關問題，或比對／相關問題之組合，均能作為說謊與否之判定指標。

表4

實測案件語音分析結果

比對內容	平均特徵數	第一群原始分類正確率(%)	第二群原始分類正確率(%)	第一群交叉驗證預測正確率(%)	第二群交叉驗證預測正確率(%)
DI/C-R <sup>a</sup>	36.47	97.92	99.01	<b>87.01</b>	<b>86.31</b>
C <sup>a</sup>	36.47	86.03	85.12	65.39	50.93
R <sup>a</sup>	36.47	85.32	84.13	61.91	48.78
NDI/C-R <sup>b</sup>	45.87	95.62	98.52	<b>81.36</b>	<b>84.07</b>
C <sup>b</sup>	45.87	83.13	79.95	61.20	51.44
R <sup>b</sup>	45.87	85.38	80.76	61.55	47.96
DI/N-R <sup>c</sup>	29.78	98.58	98.20	<b>92.70</b>	<b>86.32</b>
N <sup>c</sup>	29.78	82.18	82.15	57.92	42.07
R <sup>c</sup>	29.78	82.23	85.61	62.67	49.52
NDI/N-R <sup>d</sup>	40.63	96.90	97.50	<b>88.95</b>	<b>87.02</b>
N <sup>d</sup>	40.63	88.65	83.36	64.02	54.58
R <sup>d</sup>	40.63	83.21	77.38	60.49	48.40

a：說謊者（DI），比對問題C／相關問題R。

b：誠實者（NDI），比對問題C／相關問題R。

c：說謊者（DI），中性問題N／相關問題R。

d：誠實者（NDI），中性問題N／相關問題R。



## 一、誠實者與說謊者，其語音特徵有所差異

不論在模擬或實案測試中，均可發現利用本研究開發的語音情緒特徵分析法，可將誠實者與說謊者有效區分，代表陳述可靠度不同時，所展現的語音特徵亦有所差異，故可在此基礎上，建立更多分析模型，提升預測準確度。

## 二、無法用單一語音特徵做分析

早期林明傑等人性侵害監控語音測謊研究中，顯示基頻作為說謊偵測標的準確率較高，而音高等未發現一致性；本研究隨機挑選數人做語音特徵檢測，亦發現無法僅以音高作為說謊判斷依據，如同儀器測謊需要呼吸、心跳、膚電等多項指標綜合判定一樣，在語音測謊領域，亦無法僅用某一語音特徵進行判斷。本研究綜合各種語音情緒特徵做權重調整後，發現每個人可資區分誠實／說謊之語音情緒特徵不盡相同，因此過去使用少數特徵作為判斷基礎之聲壓測謊儀，就算再配合各種晤談步驟施行，準確率亦可能無向上大幅增加空間。

## 三、不同語者間，作為分類基礎之語音特徵有所不同

逐一檢視特徵選取步驟後，每位語者使用之誠實／說謊語音特徵可發現，用來作為誠實／說謊分類基礎之語音特徵均不相同，亦即使用之特徵會隨著語者而有所變動，雖多數語者具有共同語音特徵，但細部特徵則無

法見到一致性，因此未來中文聲紋測謊發展，究係要從族群中尋找共同說謊／誠實特徵，或是要以共同特徵為基礎，再參酌個人特徵後進行分析，亦為未來研究與評估方向。

#### 四、適當訪談情境為重要影響因素

清華大學研究中，個人化比對辨識率可達73%，而本研究之模擬試驗，提升至86%，實案測試則為92%，推測原因除套用語音情緒特徵軟體以選取較多特徵外，另一個原因可能與情境塑造與標準化訪談模式建立有關。在模擬試驗中，扣除筆者提問後，雖僅收錄誠實／說謊各約7分鐘語料，惟實驗者對於每位受試者於收錄前均花費近一小時說明測試過程、目的與原理，並讓其在不受打擾房間有足夠時間醞釀與塑造說謊情境，且誠實／說謊內容係參照警專日常生活而得，可以明確知道是否說謊，語料內容相當具代表性；而實測案例，係直接從測前晤談中擷取語音，故所收集之比對／相關／中性問題語音長度更充足，讓電腦有較足夠訓練集建模，這些因素推測均與本實驗正確率能提升有關。

#### 五、如何建立當事人基本語音資料為後續重要基本工作

本研究係從已知之誠實／說謊語料，找出分類方法，這些語料收集，在模擬部分，係在可控制環境下達成，且參酌陳述內容，可以明確知道當事人陳述真偽，

而實案部分，除以訪談技術收集外，並透過生理反應、案情資料、專家評估等步驟重複驗證，故所得語料內容真偽判定亦具代表性。但未來實際運用時，如何事先收集當事人誠實／說謊語音資料，或是在會談時，即能透過訪談獲得足資作為訓練集之語音特徵，將是一大挑戰。目前我國測謊實務中，測前晤談大約耗時一小時左右，在此時間內，除須考量正常晤談步驟外，未來又須在當事人未察覺下，獲得足夠且精確之誠實／說謊語音資料，因此若未有經驗豐富晤談者協助，將會減弱本研究開發方法之成效。而聲紋測謊最後目的，則是希望擺脫過去測謊測前晤談架構，開發出透過一般談話方式即能完成測謊之系統，為能透過一般談話方式收集到足夠語音資料，建議仍需以半結構式訪談方式進行，這些半結構式訪談題目設計與驗證，將會是語音測謊技術另一重要議題。

## 六、跨國語文使用者偵測問題

本研究係將中文語音進行分類，然伴隨著我國國際化程度日益增加，未來訪談過程將會遇到夾雜不同語言之談話內容，面對混合有不同語言之語音資料，從訪談內容設計、語音活性檢測、特徵選取、分析模型建立等可能與單純中文語料有所不同，面對這些狀況，亦為未來研究重點之一。



## 七、錄音品質影響程度尚待進一步釐清

本研究之模擬案件語音資料係以獨立麥克風收集，錄音品質較佳；實測案例受限於過往測謊室錄影設備建構目的，僅用於監控與記錄測試過程是否有違反程序正義情形，並非以收錄高品質語音分析作為出發點，故錄音品質較差。本研究一如預期結果，錄音品質較佳者，系統自動選取語音特徵較多，錄音品質不佳者，選取特徵數較少。但交叉驗證預測準確率卻顯示，實測案件雖特徵數較少，但準確率反而較高，其中模擬案件之交叉驗證預測準確率介於75.63～85.98%間，實測案件之相關／比對問題組則介於81.36～87.01%，而相關／中性問題組則介於86.37～97.2%間，整體均高於模擬案件之準確率。呈現的整體趨勢為挑選特徵數較多，反而準確率偏低，未出現正相關情況，確切原因，究係為特徵數因素，或實測案例中當事人可能有較重認知負荷與心理壓力，導致語音差異較明顯；抑或實測案例係完全依據儀器測謊測前晤談步驟操作，故可獲得較長時間語料進行建模，所以有高正確率，仍待進一步分析才能確認。

## 參考文獻

### 一、中文文獻

- 吳乙彤（2018）。基於語音特徵判斷語句內容真實性（未出版碩士論文）。清華大學。
- 林明傑、李璟林、蔡景宏、黃敏偉（2010）。我國男性假釋及緩刑之性罪犯接受聲紋測謊相關參數之研究。《刑事科學》，68，1-14。
- 陳世烽（2017）。人類內在感受與外在情緒表達有落差時的語音特徵：影像刺激條件下分析母音共振峰之特性（未出版碩士論文）。交通大學。
- 劉丁瑋（2017）。人工智慧於雨量預測以及測謊之應用（未出版碩士論文）。中央大學。

### 二、英文文獻

- Beers, M. H., & Berkow, R. (1999). *The merck manual of diagnosis and therapy*. Hoboken.
- Chou, H. C., Liu, Y. W., & Lee, C. C. (2019, November). Joint learning of conversational temporal dynamics and acoustic features for speech deception detection in dialog games. In 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China.
- Cósetl, R. C., & López, J. D. B. (2011, February). Voice stress detection: A method for stress analysis detecting fluctuations on Lippold microtremor spectrum using FFT. In CONIELECOMP 2011, 21st International Conference on Electrical Communications and Computers, San Andres Cholula, Mexico.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L.,

- Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 74-118. <https://doi.org/10.1037/0033-2909.129.1.74>
- Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: A review. *Journal of voice*, 27(3), 390-e21. <https://doi.org/10.1016/j.jvoice.2012.12.010>
- Grubin, D., & Madsen, L. (2005). Lie detection and the polygraph: A historical review. *The Journal of Forensic Psychiatry & Psychology*, 16(2), 357-369. <https://doi.org/10.1080/14789940412331337353>
- Hopkins, C. S., Ratley, R. J., Benincasa, D. S., & Grieco, J. J. (2005, January). Evaluation of voice stress analysis technology. In proceedings of the 38th annual Hawaii international conference on system sciences, Big Island, HI, USA.
- Krishnamurthy, G., Majumder, N., Poria, S., & Cambria, E. (2018). *A deep learning approach for multimodal deception detection*. arXiv preprint arXiv:1803.00344.
- Levitan, S. I., An, G., Wang, M., Mendels, G., Hirschberg, J., Levine, M., & Rosenberg, A. (2015, November). Cross-cultural production and detection of deception from speech. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, Seattle Washington USA.
- Levitan, S. I., Maredia, A., & Hirschberg, J. (2018, September). Acoustic-Prosodic Indicators of Deception and Trust in Interview Dialogues. In Interspeech, Hyderabad, India.
- Liu, X. (2005). *Voice stress analysis: Detection of deception*. (Unpublished master's thesis). University of Sheffield.
- National Research Council (2003). *The polygraph and lie detection*. National Academies Press.

- Onder, A., & Brittan, M. (2009). Recent case law under the employee polygraph protection act: A practical review. *Privacy & Data Security Law Journal*, 4(6), 483-503.
- Pan, X., Zhao, H., & Zhou, Y. (2015). The application of fractional Mel cepstral coefficient in deceptive speech detection. *PeerJ*, 3, e1194. <https://doi.org/10.7717/peerj.1194>
- Tomba, K., Dumoulin, J., Mugellini, E., Abou Khaled, O., & Hawila, S. (2018, July). Stress Detection Through Speech Analysis. In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018), Porto, Portugal.
- Wu, Z., Singh, B., Davis, L., & Subrahmanian, V. (2018, April). Deception detection in videos. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA.