應用遺傳演算法指認嫌犯話音之研究

許芳誠講師
淡水工商管理學院資訊管理系
台北縣淡水鎮眞理街三十二號

陳稼興副教授
國立中央大學資訊管理系
桃園縣中壢市五權里三十八號

摘要

人類天生具備極佳的語音辨識能力，對認識或不認識的人所說過的話音，很自然地能保留足夠的印象，當再度聽到同一人所發出的聲音，都可以容易的指認出來。本文的主要目的是，提供一個結合「遺傳演算法」與「語音合成技術」的「話音重現」觀念架構。希望借助此架構所發展出的系統，能幫助犯罪被害人或目擊者容易的找到嫌犯的聲音，做爲警方偵查案件時的參考線索。全文分五章，第一章爲前言，第二章簡要陳述語音學、語音合成技術，第三章主要說明遺傳演算法的基本概念，第四章提出「話音再現」的理念及觀念架構，第五章爲結論。

關鍵字：遺傳演算法，語音合成技術，犯罪偵查。

1. 前言

傳統犯罪偵察程序中，辦案人員在搜證時，大多不考慮採集嫌犯聲音作爲證據。由於語音學(phonetics)長期發展的基礎，以及語音辨識科技的成就，近年來，警政機構已經可以運用電話錄音採集嫌犯的聲音，並提取聲譜中的一些特徵作爲"聲紋"，和嫌犯的聲音比對，協助案件偵破。受限於嫌犯聲音採集的困難度較高，上述的運用並不普遍。儘管如此，借助聲音協助犯罪案件的偵破總算有了起步。本文嚐試從另一個角度提出運用聲音協助犯罪案件偵破的方法。

被害人或現場目擊者通常只能提供犯案者臉部或其它的外觀特徵給警方參考，但一些蓄意犯罪的人通常會利用蒙面或化妝等技術，使得被害人或目擊者無法提供有效的資訊給警方，警方因此也失去一些辦案線索。事實上，被害人或目擊者除了目睹現場實況外，通常也可能聽到嫌犯的談話聲音。因此，若警方能提供方便及有效的方法協助目擊者提供嫌犯的聲音資訊，相信將有助於案件的偵破。

人類身體的發音構造遺傳自上一代，且每一個人的發音方式以及說話的語氣、節奏等大多也是遺傳或學習自父母或其他人。基於上述的認知，可以想像每一個人說話的許多特徵中，應有絕大部分是從許多會說話者的特徵中複製而得。這種情形和大自然的演化(natural evolution)有許多相似之處。

因此，若可找到或設計出能編碼人類話語的「染色體」，便有可能模仿大自然演化的程序，產生出和任何人的說話方式相當近似的「話音」。換言之，我們可以借用大自然演化的機制，模擬出任何人講話的聲音，包括說話者話音的音長、音色、音高、音量、語句中的停頓等。全文將分別從語音學、語音合成及遺傳演算法的發展情形，藉以說明「話音再現」的可行性。

本文的主要目的是提出一個結合語音合成(synthesis)技術及遺傳演算法(genetic algorithms)的「話音再現」觀念架構。運用此架構所發展出來的系統將可幫助目擊者「重現」犯案者在犯罪現場發出的談話聲音。爲方便說明，全文共分成五章，除第一章的前言外，第二章將簡要陳述語音學、語音合成技術，第三章主要說明遺傳演算法的基本概念，以及應用遺傳演算法解決問題時的做法，第四章提出「話音再現」的理念及觀念架構，第五章爲結論。

2. 語音學與語音合成

人類研究語音已有上千年前的歷史，如中國的「音韻學」。傳統以來，研究者都著重於語音的定性分析。本世紀初由於科學的進步，研究人員才開始運用生理實驗，並借助醫學儀器進行有關發音部位、發音氣流、和聲帶動作等研究。二次大戰後各種聲學儀器(例如，示波器、X光照像機等)被廣範應用於語音研究，使得研究人員對聲音的物理性質有更進一步的瞭解。上述的發展，不但豐富了語音學的研究範圍，也提升了語音學的實用性。

60年代後期開始運用電腦進行語音分析，此舉不但促成了電腦語音合成與語音識別研究的興起，同時也建立了語音學家與電腦科學研究人員攜手合作的管道。事實上，本世紀的語音學在「音位學(phonemics)」、「音響學(acousics)」、「實用語音學(practical phonetics)」、「實驗語音學(experiemental phonetics)」和「比較語音學(comparative phonetics)」等領域所累積的研究成果，都已成爲近代電腦語音合成與語音識別研究人員的重要參考。

時至今日，由於人們要求電腦除了能發出單個音素、音節外，還要發出連續的語流，一方面也希望電腦能聽懂不同的人所說的話，甚至要求電腦能進一步和人交談。基於上述相同的理由，語音合成和識別的研究人員已開始向語音學家提出更多有關語音現象的要求。瑞典家語音學家方特，在第十屆國際語

音學會的主題報告中提到:「我們首先需要的是第五代的語言科學家，而不是第五代的計算機」[2]。

從上述語音學的發展過程看，由於語音學研究的實用性越來越受到重視，研究人員在則被要求持續地提供更多有關語音學的研究新知，這種趨勢將可能導致語音合成與識別研究的良性進展。就語音合成的複雜度而言，發展中文語音發音系統時，一個相當有利的因素就是中國話(Mandarin Chinese)是單音節(mono-syllable)的語音。近年來學者在中國話語音合成的研究方面也有一些成果[3,4,14]。

運用遺傳演算法「再現話音」的其中二個重要關鍵是：設計「數位染色體」的編碼方法，以及將最後找到的染色體(最佳解)轉換成語音。前者需借助於語音學的研究成果，後者則有賴於語音合成技術的支援。目前語音學及語音合成的發展成果，應足以建立「話音再現」可行性的基礎。

3. 遺傳演算法

長期以來，生物學家一直對物種的演化理論(evolutionary theory)有高度的興趣，儘管學界對隱藏在演化背後的驅動機制 (mechanisms)並不全然瞭解，但透過對自然演化過程的觀察，人類尚能掌握演化當中的一些特質。包括：演化的發生主要是染色體(chromosomes)的運作所引起的；之所以有新的生命物種，部分原因隱藏於染色體的解碼過程中。同樣的，對於染色體編碼及解碼的細節，生物學家迄今也未能全然瞭解，但以下所列的幾個一般些特徵則被普遍接受[8]。

1.演化的程序是建立在染色體的運作上，而不是建立在生物(染色體所編碼成的)本身上。

2.天擇(natural selection)的過程會導致一些編碼結構較好(較能適應環境)的染色體較可能被複製(reproduce)到下一代，反之則反。

3.不同的複製程序會導致不同的演化。突變(mutation)程序可能使得下一代的染色體和上一代的染色體截然不同；而重組(recombination)程序則藉由組合上一代(雙親)的染色體要素(material)，產生和上一代略有差異的下一代染色體。

4.生物的演化並無記憶。換言之，演化論的觀點中沒有經驗法則，沒有任何指引，演化靠的是嘗試錯誤(物競天擇)。

1970年代初期， Holland[12] 受到上述「自然演化」理論的激勵，嘗試擷取演化的精神，並適度結合計算機演算法，希望能找到一種可以「模仿大自然解決複雜問題的方式」，隨後終於在1975年提出「遺傳演算法」。運用遺傳演算法解決問題基本上有六個主要的步驟[8]。

1.由電腦產生一群「數位染色體」，做爲母代族群(population)。

2.根據評估函數，評估族群中的每個染色體的「適應力」。

3.挑選「適應力」佳的染色體兩兩配對，並運用基因重組、突變等演算，產生新的染色體。

4.淘汰母代族群中「適應力」較差的染色體，爲新的染色體預留空間。

5.評估新染色體的「適應力」，並將「適應力」較佳的新染色體加入母代族群中存活下來的染色體堆裏，形成新的母代族群。。

6.如果演化的時間已截止或找到滿意的解答，則停止演化並傳回「適應力」最佳的染色體(最佳解)，否則反覆步驟3至步驟6。

根據上面的演算步驟，運用遺傳演算法解決問題的系統中，基本上需要包含五個主要的組件[9,15]：

1.設計一種能表示問題「可能解(potential solutions to the problem)」的染色體表示方式。亦即，需建立「數位染色體」的表示法。常見的表示方式是將「可能解」編碼成二進位有限長度的位元串(bit string)。

2.訂定一套能隨機產生母代族群(initial population)之方法。母代族群爲是由許多「數位染色體」(可能解)組成的。

3.定義評估函數(evaluation function)。評估函數扮演類似大自然一樣的角色，用以爲每一個可能解(染色體)打分數，此分數用以代表每個染色體的「適應力(fitness)」。

4.決定採用何種「遺傳運算子(交配或突變等)」。遺傳複製期間就是因爲這些運算子改變了子代族群染色體的組成。

5.決定遺傳演算法所需的各個參數值(族群的大小、交配發生率、突變發生率等)。

自Holland發表遺傳演算法至今，有關遺傳演算法的研究及應用報告數量，每年皆以極快的速度增加，從Alander[6]截至1993年爲止所搜集的2521篇有關遺傳演算法的文獻中，顯示1975至1985年間僅有百餘篇文章發表，但1986至1990年間則有六百多篇，1991至1993年短短三年間，所發表的文章更高達一千六百餘篇。

1995年的「第六屆國際遺傳演算法研討會」論文集[13]中所收錄的82篇論文中，有高達23篇文章是有關遺傳演算法的應用，從此也可以看見遺傳演算法的實用性。

4.「話音指認系統」之觀念架構

4.1「話音再現」之理念

由於人類天生具備極佳的影像辨識能力，警方經常利用素描專家配合目擊者的描述，取得嫌犯的面貌素描。但常受限於素描專家培養不易、目擊者無法完整記憶嫌犯面貌的細部特徵，或目擊者與素描專家之認知差異等因素，導致素描工作需耗時相當長，甚至導致素描誤差太大等缺失。目前雖然也有利用「面貌影像拼湊機器」協助合成嫌犯面貌之作法，但王朝煌[1]認爲「由於人類不斷的進化及人類求新奇的本性，面貌特徵也漸漸變化，面貌影像拼湊機器內存之底片須常更新方敷時用。再者面貌影像拼湊機器之製作不易且造價昂貴，無法大量使用」。

Holland 在解釋他的遺傳演算法觀念時，最喜歡以警方繪相師的例子作爲說明[16]。而 Caldwell and Johnston [7]運用遺傳演算法所發展出來的系統，可以不斷地產生新面貌，讓目擊者指認是否是嫌犯相似。目擊者不需要事先描述嫌犯的面部特徵，只要運用天生具備的面貌識別能力，爲系統每次產生的新面貌評分，評定每張面孔究竟和嫌犯有鄉幾分相似，系統便會根據目擊者的評分，透過交配、突變等演算，產生經過「物競天擇」後的新一代面貌。原則上，目擊者只要不斷地爲最新一代的面貌評分，經歷幾代後便可以找到近似嫌犯的面貌圖像。這種系統如果能成功的推出，就可以避免前述方法的一些缺失。事實上，在NMSU(New Maxico State University)的努力下，運用遺傳演算法發展

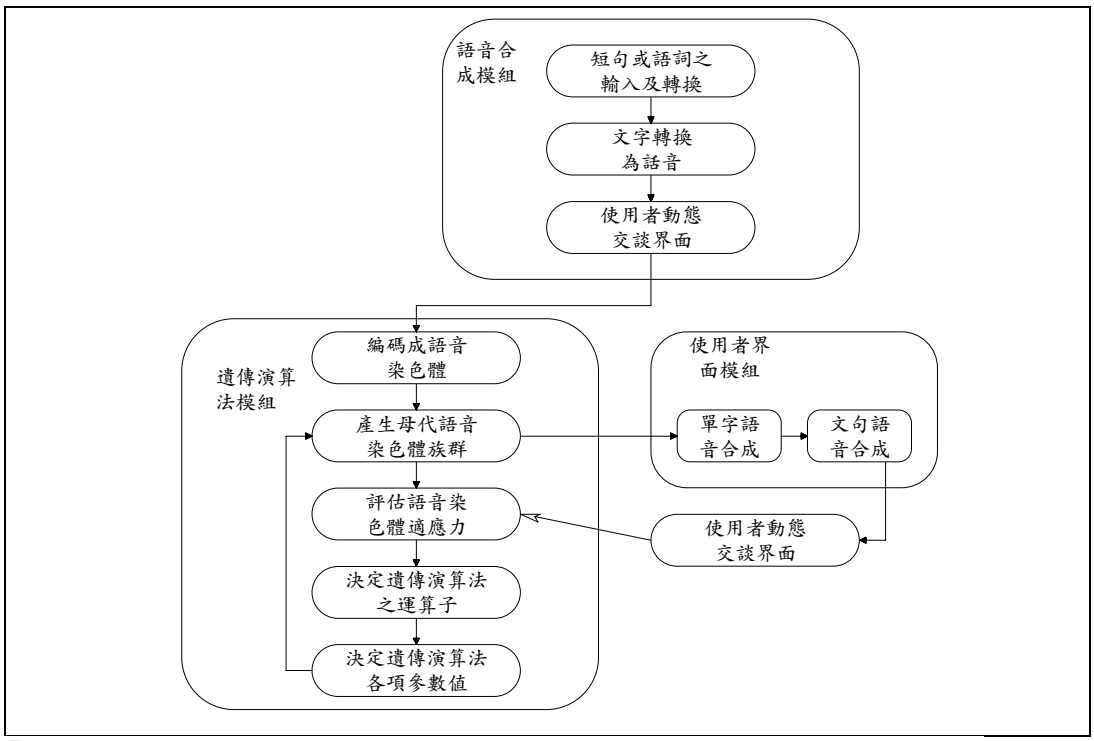
的面貌指認系統已有相當好的進展，除了取得專利外，並已很接近商業化的程度[10]。

人類除了天生具備極佳的影像識別能力外，同時也具備極佳的語音辨識能力，我們一生中認識許許多多的人，聽過幾乎無限多的聲音。而我們通常不必經由眼睛的幫助，就可以輕易地從說話者的聲音中，辨認出是那一位熟人在說話，不論已有多久沒聽見說話人的聲音。

對於不認識的人所發出的聲音，在經歷一段時間後，雖然我們可能已經完全無法回憶該聲音的一些訊息，但當同樣的聲音再度出現時，我們卻依然可以肯定確實曾經聽過該聲音。如果我們能應用遺傳演算法並結合語音合成的技術，發展出一套可以演化出人類談話聲音的系統，做爲被害人或目擊者指認嫌犯話音時的工具，相信對於提供偵查線索應很有助益。

4.2「話音再現」系統觀念架構

本文所架構之「話音再現」系統，內含如圖一所示之三個主要模組：模組一爲「使用者界面模組」，模組二爲「遺傳演算法模組」，模組三爲「語音合成模組」。三個模組之間的關係如圖所示。系統運作過程大多由「使用者交談界面」控制，當使用者確認已找到欲「再現之語音」或演化時間已截止，都會結束所有模組的運作。系統實作時，可根據本觀念架構設計更明確之流程。爲方便說明底下關於各模組的描述皆以中文語音爲例。



圖一「話音再現」系統觀念架構

模組一：「使用者界面模組」。主要有以下三個組件：

1.短句或語詞輸入及轉換。讓使用者方便輸入想要「再現」之短句或語詞(例如，「站起來」、「不要走開」等)，並將之轉換成基本語音合成單位(例如，若使用四聲基頻模式，則「站起來」將轉換成「ㄓ」「ㄢˋ」、「ㄑ」「ㄧˇ」、「ㄌ」「ㄞˊ」)，同時也記錄下該短句之長度(例如，「站起來」長度爲 3)。最後將之編成對應的碼(依一定規則編碼)。

2.文字轉換成話音(text-to-speech)。(1)處理「同字異音」(例如，「音樂」和「快樂」中的「樂」，前者發音爲「ㄩ」「ㄝˋ」，後者發音爲「ㄌ」「ㄜˋ」)。(2)處理前後字相互間對發音影響(例如，「土生土長」中的後一個「土」，受到同爲第三聲的「長」的影響，應發音爲「ㄊ」「ㄨˊ」)。

3.交談界面。系統將演化出的新一群「染色體(可能解)」以視窗圖像(icon)方式，呈現在終端機螢幕上，由使用者按鈕聽音。接著輸入對每個「圖像(染色體)」所代表的聲音的評分。最接近真正聲音的染色體給予最高分，就遺傳的觀點而言，代表該「染色體」「適應力」最佳。

模組二：「遺傳演算法模組」。基本上有五個組件：

1.將使用者輸入的一段短句或語詞編碼成「語音染色體」。亦即將之編碼成二進位有限長度的位元串(bit string)。由於短句或語詞的長度並不固定，因此染色體的長度將隨短句或語詞的長度隨機改變。染色體可以設計成如圖二所示之格式。

2.一套能隨機產生母代「語音染色體」族群之方法。母代族群是由許多編碼成符合圖二格式之「語音染色體」組成的。

3.評估「語音染色體」之「適應力」。過去以來，評估染色體「適應力」經常運用評估函數，但運用遺傳演算法進行多變量搜尋時，一個創新的作法是，透過與使用者進行動態交談方式，評估染色體「適應力」[7]。

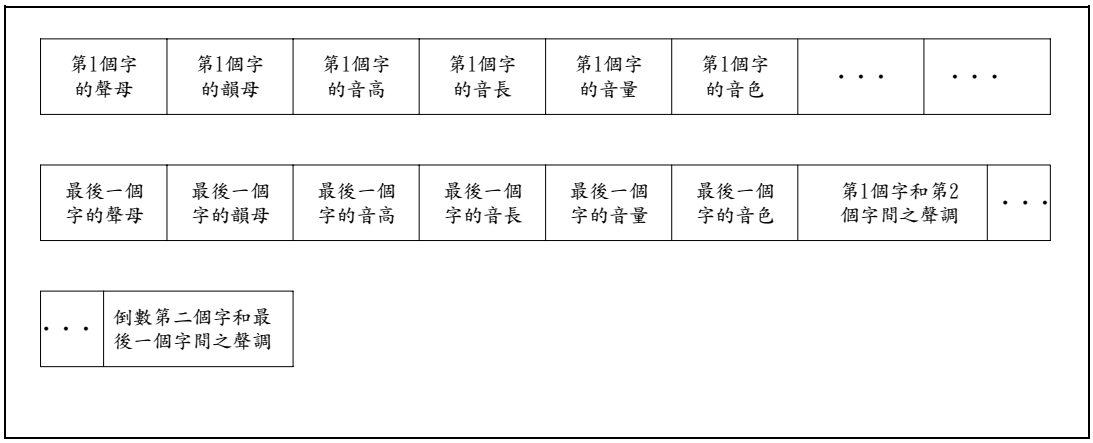
4.定義淘汰、配對、突變等遺傳運算子，並進行基因重組。遺傳複製期間，便是運用這些運算子，產生新的染色體。

5.決定遺傳演算法所需的各個參數值(族群的大小、交配發生率、突變發生率等)。理想的作法是，利用一個中間模組，持續記錄各參數值的比例，及評估值的趨近速度，動態提供最佳的參數供系統使用。

模組三：「語音合成模組」。主要含有二個組件：

1.單字合成次模組。可根據染色體中的聲母、韻母，對應出基本的語音波型，再利用染色體中的音長、音高、音量、及音色等資訊，配合模組中的單字合成規則，合成不同音質的單字語音。

2.字詞合成次模組。可根據染色體中每個單字間的聲調資訊，調整詞句間的連音聲調。



圖二 語音染色體之編碼格式

5. 結論

本文所提出之「話音指認系統」觀念架構，最大的特色是借用大自然的演化概念，運用遺傳演算法的運算，並讓使用者依曾經聽過的聲音印象，導引語音的演化趨勢，使最終得以「重現」使用者曾聽過的聲音。而語音「重現」過程中，系統從頭至尾都不必要求使用者描述曾經聽過的語音之特徵，只需要對系統持續產生的語音評分。換言之，只要告訴系統所聽到的聲音像不像就可以了。

日後依本架構所實作的系統，除了文中所述得以做爲犯罪受害人或目擊者提供嫌犯說話的話音外，也可以供廣告、戲劇業者尋找心目中理想的聲音，或讓使用者尋找自己遺失或遺忘的聲音等等。

由於聲音染色體的編碼長度相當長，且長度將隨詞句的長度增加而增加，使得找尋空間非常巨大。爲避免「話音重現」耗時過長，可考慮在觀念架構的「使用者模組」中，加入「模糊(fuzzy)語意次模組」，接受使用者輸入語義

模糊的語音描述(例如，聲音沙啞、語調急促等)，使得原本依隨機方式產生的第一代染色體族群，變成使用者可回憶的「較佳化」第一代染色體族群。

另外，根據「實用語音學」的理論，大自然或發聲儀器所產生的聲音比起人體的發音器官所能發出的聲音，或聽覺器官的所能感受到的聲音多到數不清的地步[5]。因此若能嚐試在「語音合成模組」中加入「專家系統次模組」，藉以過濾人類器官不可能發出的聲音，或聽覺器官所無法感受到的聲音，相信對避免「話音重現」耗時過長，應有相當大的助益。

若目擊者人數不止一人，則讓多位目擊者個別「尋找」各自印象中的染色體，然後再將每個人的所尋得的染色體，當成母代染色體族群，繼續往下演化，可能可以增加「重現話音」的真實性。

本文僅爲一觀念性架構，希望能提供拓展語音應用範圍的原型架構，另一方面也可供研究語音學的學者，在發展語音理論時的參考。文中對系統實作時可能遭遇到的問題並沒有深入探討，有關實作的問題可留待未來作進一步之研究。

6. 參考文獻

1.王朝煌，面貌影像拼湊編輯系統之研究，第六屆國際資訊管理學術研討會論文集，1995年5月，頁155-160 。

2.石鋒，語音學探微，北京大學出版社，1990年12月。

3.許志旭，素片編集方式的語音合成之研究，淡江大學資訊工程研究所碩士論文，1993年6月。

4.熊明德，中文語音規則合成之研究，淡江大學資訊工程研究所碩士論文，1991年6月。

5.謝雲飛，語音學大綱，台灣學生書局，1990年11月。

6.Alander, J. T. An Indexd Bibliography of Genetic Algorithms:Years 1957-1993. Dept. of Information Technology and Production Economics, University of Vaasa, Report Series No.94-1, Feb, 1994.

7.Caldwell, C. and Johnston, V. S. Tracking a Criminal Suspect through "Face-Space" with a Genetic Algorithm. Proceedings of the Fourth International Conference on Genetic Algorithms. Morgan Kaufmann, San Mateo, California, 1991, pp. 416-421.

8.Davis, L., Ed. Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York, 1991.

9.Davis, L., Ed. Genetic Algorithms and Simulated Annealing. Morgan Kaufmann Publishers, Inc., Los Altos, California, 1987.

10.Goldberg, D. E. Genetic and Evolutionary Algorithms Come of Age. Communications of the ACM, March, Vol.37, No.3, 1994, pp. 113-119.

11.Goldberg, D. E. Genetic Algorithms in search, Optimization, and Machine Learning. Addison-Wesley, Reading, Mass, 1989.

12.Holland, J. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, 1975.

13.Eshelman , L. J., Ed. Proceedings of the Sixth International Conference on Genetic Algorithms. Morgan Kaufmann, San Francisco, California, July, 1995.

14.Lee, L. S., Tseng, C.Y., and Ouh-Young M. The Synthesis Rules in Chinese Text-to-Speech System. IEEE Trans. on ASSP, Vol.37, No.9, 1989, pp.1309-1320.

15.Michalewicz, Z. Genetic Algorithms+Data Structures=Evolution Programs. Springer-Verlag, Berlin Heidelberg, 1992.

16.Woldrop, M. M. Complexity:The Emerging Science at the Edge of Order and Chaos. 天下文化出版社中譯本,1994 年,頁229.