

## 小波分析在語者識別之初步探討

黃雪珠 王昭男

台灣大學工程科學及海洋工程研究所

### 摘要

本文將以小波分析法針對電話按鍵音、樂器音、人的聲音進行辨識。電話按鍵音和樂器音為規律的訊號，所以可以得到不錯的辨識結果，而人的聲音變化很大，本文將著重在語者識別的部分。語者識別(Speaker identification)主要是藉由分析人的聲音達到識別語者身分的目的[1]，一般分為字相依(text - dependent)和字獨立(text-independent)兩種分析模式，本文所要探討的為字獨立的辨識。

影響人的聲音變化有許多因素，例如環境、年紀、性別等，因此特徵參數的抽取要能顯示出一個人獨有的聲音特質。本文結合了多解析度的小波分析[2]和常用的梅爾倒頻譜參數(Mel frequency cepstrum coefficients, MFCC)作為特徵演算，並藉由向量量化(Vector Quantization, VQ)的方式將特徵向量進行編碼，建立語者的資料庫，而在辨識時，將輸入的訊號和資料庫做 VQ 誤差比對，找出最小的偏差值，定為辨識的結果[3,4]。

**關鍵字：**小波分析，梅爾倒頻譜參數，向量量化，識別

### 一、前言

近年來語者識別(Speaker Identification)被廣泛的研究，語者識別系統可以藉由人的聲音查核身分，並進而利用在語音撥號、語音郵件、或是機密文件的安全掌控等領域。其重點在於人的聲音特徵抽取和比對，本文利用小波和常用的梅爾倒頻譜參數結合做聲音特徵分析，探討利用向量量化的方式做識別的效果。

小波理論在科學及工程領域被廣泛運用，不管在聲音或影像的處理上都有其優點，本文首先針對單純的訊號：電話音和樂器音做分析，以了解小波在時間頻率域分析的優點，接著針對本文重點：語者識別做分析。本文之第二章將先介紹小波基本理論；第三章介紹電話音和樂器音的基本概念，以及利用小波分析所得的結果，第四章將討論較複雜的訊號—人的聲音，也就是將收集到的訊號，藉由小波工具，擷取有關語者的聲音特徵向量，並且利用向量量化的分類法(classification)達到語者識別的目的。

### 二、小波簡介

小波變換於時間-頻率域上，是一種多重解析度分析(multi-resolution analysis)。所謂的蓋博轉換(Gabor Transform)其實是利用高斯視窗(Gaussian window)做傅立葉轉換，其利用高斯視窗的短時間-頻率特性，將時間-頻率域上的能量集中以達到分析訊號的目的，如果將

其中的頻率平移以階度(scale)壓縮或擴展(dilation)所取代，所推演出的時間梯度分布即為積分小波轉換(Integral Wavelet Transform)[5]。

$$W_{\psi} f(b, a) = \frac{1}{a} \int_{-\infty}^{\infty} f(t) \overline{\psi_{b,a}(t)} dt \quad (2.1)$$

其中  $\psi_{b,a}(t) = \psi\left(\frac{t-b}{a}\right)$  ,  $a > 0$

而  $\psi(t)$  (基本小波) 為一實數或複數之帶通函數， $b$  為時間的平移(translation)變數， $a$  為擴展(dilation)變數， $a$  值越大是為低頻，反之為高頻。

傅立葉積分轉為傅立葉轉換是將時間和頻率取離散，而小波積分化為小波轉換的方式不同，是藉由將  $a$  和  $b$  兩個值取離散(discrete)，將  $a$  取為  $2^{-s}$ ， $b$  為  $k2^{-s}$  的形式，

$$W_{\psi} f(k2^{-s}, 2^{-s}) = 2^{s/2} \int_{-\infty}^{\infty} f(t) \psi(2^s t - k) dt \approx 2^{s/2} \sum_n f(n) \psi(2^s n - k)$$

經由小波轉換的時間-頻率域，低頻範圍的頻率解析度好，高頻範圍的時間解析度好，這項特徵和人耳聽覺的分辨能力相仿[6]。

### 三、電話按鍵音與樂器音分析

本章首先利用小波分析電話按鍵音和樂器音，以了解其對頻率分解的機制。

電話按鍵音和樂器音是屬於較單純的訊號，其聲音是由一個基頻或數個泛頻組合而成，可以藉由小波對訊號分析具頻率多解析度的特性加以分析。

電話系統是採用雙音多頻(Dual Tone Multi Frequency,DTMF)的原理，每一個按鍵皆利用兩個音(tone)來表示，其特點是抗干擾能力強。

因為電話按鍵音是由單純的雙音所組成的，利用小波將所有的電話按鍵音做分解，如此便可以將組成的頻率找出，而每一組小波係數都代表其頻率的成分，因此可以視為特徵值，經由訓練的資料庫和測試訊號做比對，找出最小的誤差值以判定辨識結果。

電話訊號分析得出的辨識率相當良好，可以百分之百的辨識出待測訊號。

樂器的訊號變化較電話音多且複雜[7]，所以無法藉由小波分解直接找出訊號的組成頻率，因此採用將分解後的小波係數作能量的運算，計算出每一階小波係數的總能量值作為特徵值，將測試訊號與其做誤差比對以判斷結果。

樂器辨識之樂器音訊號是利用 Creative keyboard-AudioHQ 程式直接製造，針對每種樂器的 C4 音程做分析，所選用的樂器有 Piano(鋼琴)、Violin(小提琴)、Cello(大提琴)、Trumpet(喇叭、小號)、Oboe(雙簧管)、Koto((日本)十三弦古箏)等六種樂器，但是實際上樂器在發音時會有一些環境的干擾音，所以要加上雜訊以逼近實際的狀況，加入原訊號比例 3% 的雜訊結果在表 3.1。

表 3.1 樂器辨識(noise=3%)

Musical instrument	Piano	Violin	Cello	Trumpet	Oboe	Koto	Average
Recognition Rate(%) (Correct/Total)* 100	91.67	100	100	91.67	100	91.67	95.835

藉由上述分析得知，規律的訊號如電話音和樂器音等，有固定組成頻率，因此有相當高的辨識率。

#### 四、語者識別

在語者識別方面，本文採取閉集(close-set)識別的方式，亦即假定待識別語者在集合內。依據識別方式的不同，可以分為字相關(Text-Dependent)和字獨立(Text-Independent)兩種，字相關為系統辨識文字是已知的，而字獨立是系統以不同的文字作辨識，本文以字獨立的分析為研究主軸。

一般而言，人耳聽力範圍為 20Hz~20kHz，而口語發聲頻率範圍在 100Hz~10kHz 範圍內。本文採用中華民國語言學學會的國語語音資料庫：MAT-160，作為語音訊號的來源，其錄製了 160 個人的聲音，79 位女性和 81 位男性，分別說不一樣的文字內容，本次實驗主要針對 12 個人的聲音作分析，利用不同的語音內容辨識其身分。訓練的語料經過前處理步驟(沉默段去除、正規化及取音框(frames))後，將訊號作小波轉換，並將係數作梅爾倒頻譜參數特徵的抽取，收集到的所有語者的資料經過上述特徵擷取的步驟後，並利用向量量化的方式將大量的特徵向量編碼，建立其資料庫，而測試語句與語者的編碼本作誤差比對便可以識別語者身份，其擷取特徵流程如圖 4.1 所示。

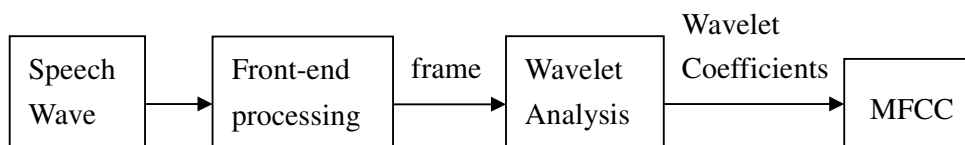


圖 4.1 擷取聲音特徵流程圖

對於語者識別的結果，建立資料庫時的語音越完整，辨識率越高，如圖 4.2 所示，

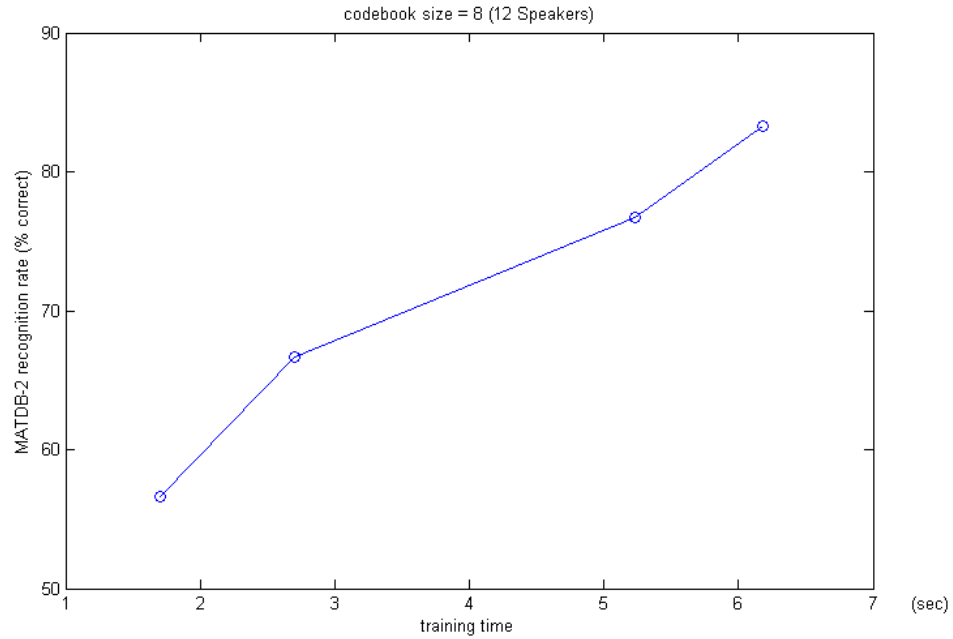


圖 4.2 訓練時間對辨識率的影響

因為小波在低頻時對頻率較敏感，因此分解的階數對辨識率也有一定的影響，如圖 4.3 所示。

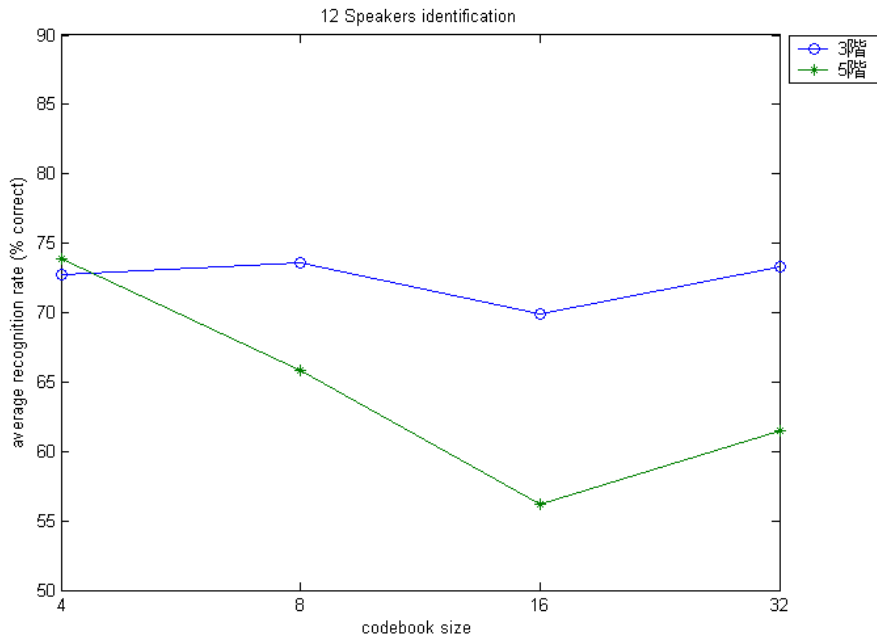


圖 4.3 小波階數對辨識率的影響

經由上述分析可以了解，資料庫訓練的時間越長，代表人的聲音變化越足夠，以及分解的階數越多對辨識的效果並不一定有效，因此要如何在這些參數中做取捨將是影響結果相當重要的因素。

## 伍、結論

利用小波對時間-頻率的多解析度特性，可以幫助分析訊號，對於較單純的訊號可以達到預期中的效果，但是對於複雜的聲音，卻有許多影響因素要考慮，因此，對於特徵參數的抽取、比對等尚有許多可以發展的空間，例如對於小波族的選取、訊號的錄製以及特徵參數的選擇等。

## 六、參考文獻

1. Joseph P. Campbell and JR, "Speaker Recognition :A Tutorial" Proc. Of the IEEE, vol.85, no.9, Sept 1997, pp. 1437-1462
2. G. Strang, T. Nguyen, "Wavelets and Filter Banks", Wellesley-Cambridge Press, New York, 1996
3. Evgeny Karpov, "real time speaker identification", University of Joensuu, Department of Computer Science, Master's thesis, 2003
4. J. R. Deller, J. H. L. Hansen, J. G. Proakis, "Discrete Time Processing of Speech Signals", Piscataway(N. J. ), IEEE Press, 2000
5. Jaideva C. Goswami and Andrew K. Chen, "Fundamentals of Wavelets, Theory, Algorithm, and Applications", New York : John Wiley & Sons, 1999
6. 潘敏俊, 姜雲生, 「聲紋信號之時頻域分析鑑識」, 中華民國音響學會第十二屆學術研討會論文集, 第 137-144 頁(1999)。
7. 謝寧, 「音樂的科學原理」, 徐氏基金會, 第 9-15 頁(1961)。