# 語者辨識
# Speaker Recognition

陳慶瀚
機器智慧與自動化技術(MIAT)實驗室
義守大學電機系
pierre@isu.edu.tw
2005年12月20日

# 語者辨識課題

語者辨認系統可分為兩個課題：

- 語音特徵擷取（Feature Extraction）
- 分類器（Classifier）

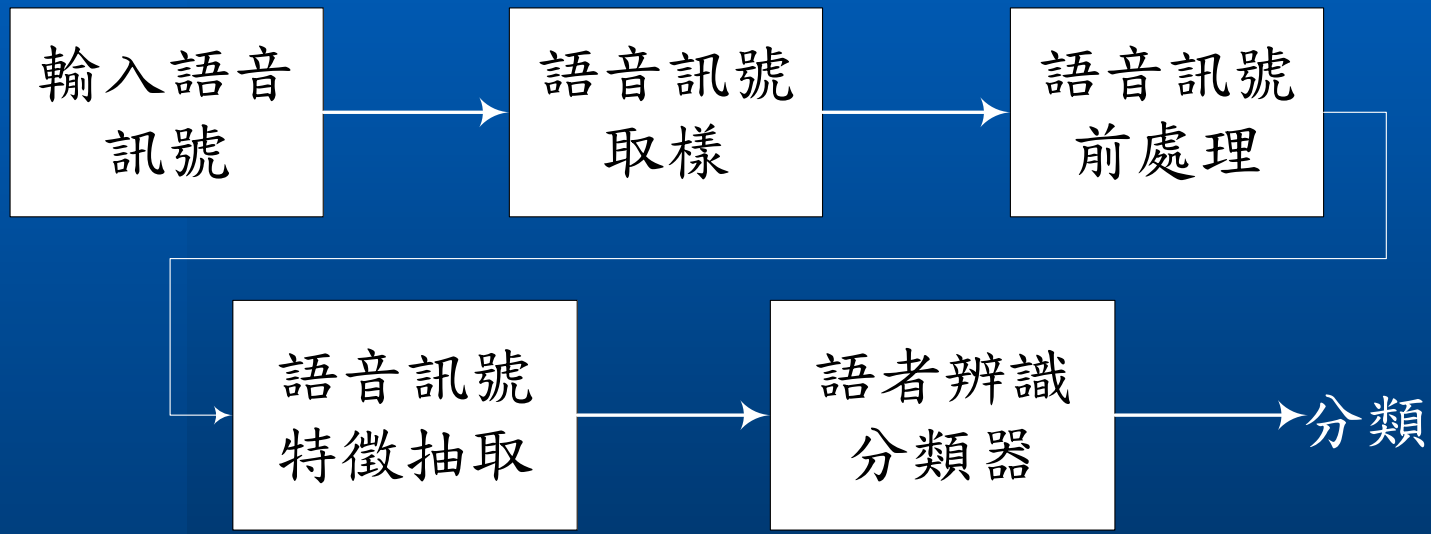
重要的研究方向：

- 提昇辨識性能
- 簡化計算複雜度
- 硬體架構

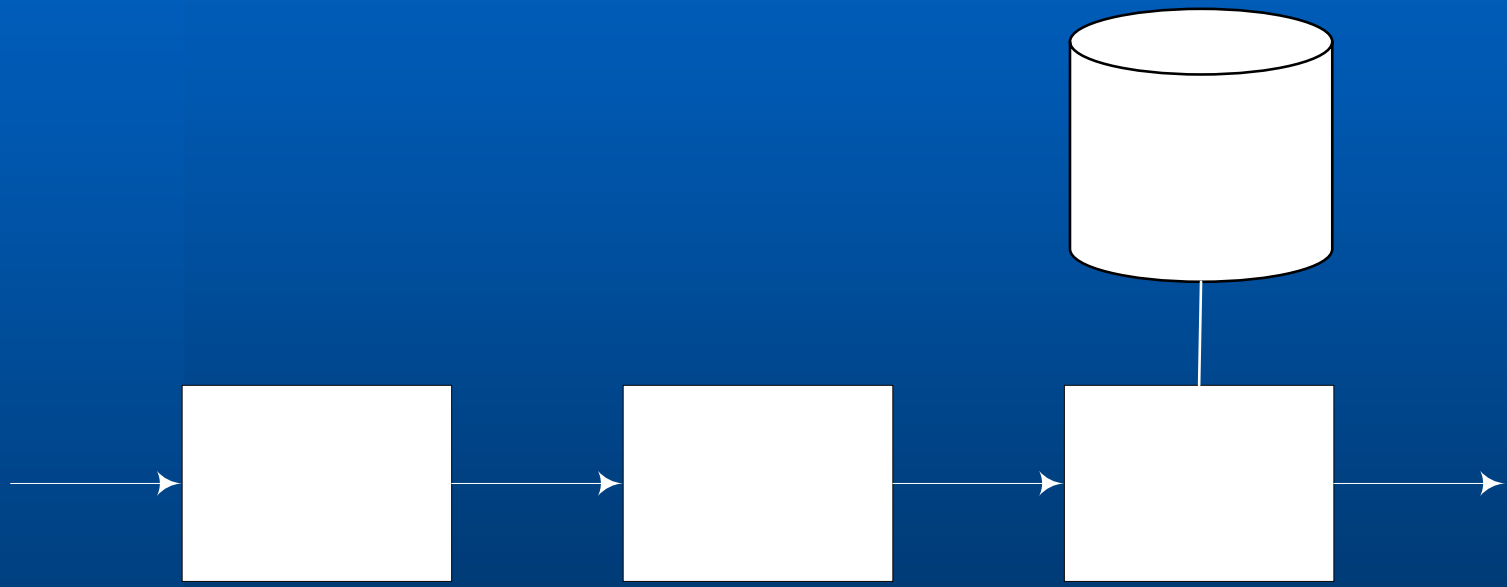
# 語者辨識模式

- 根據使用者的聲音訊號輸入，辨認使用者身份
- 辨識模式
  - 語者識別（Speaker Identification）
  - 語者驗證（Speaker Verification）
- 語音輸入模式
  - 文字相關（Text-Dependent）模式
  - 文字不相關（Text-Independent）模式

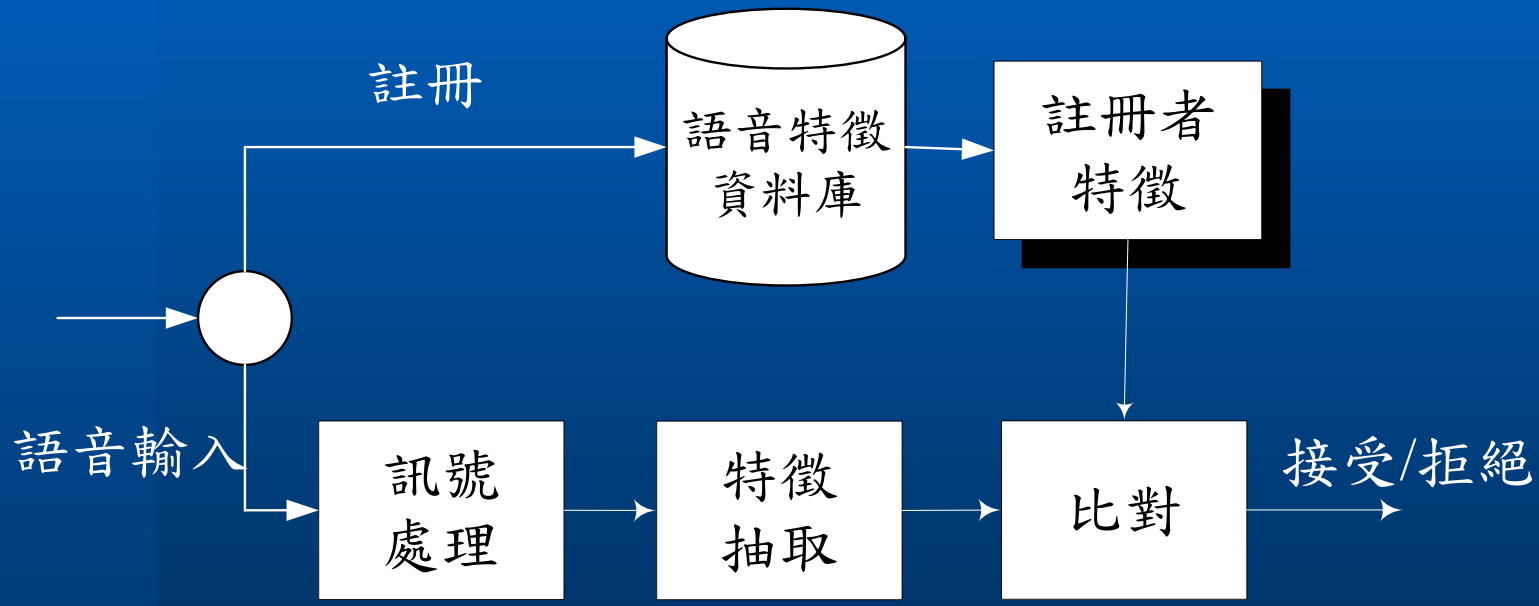
# 語者辨識流程



# 語者識別模式



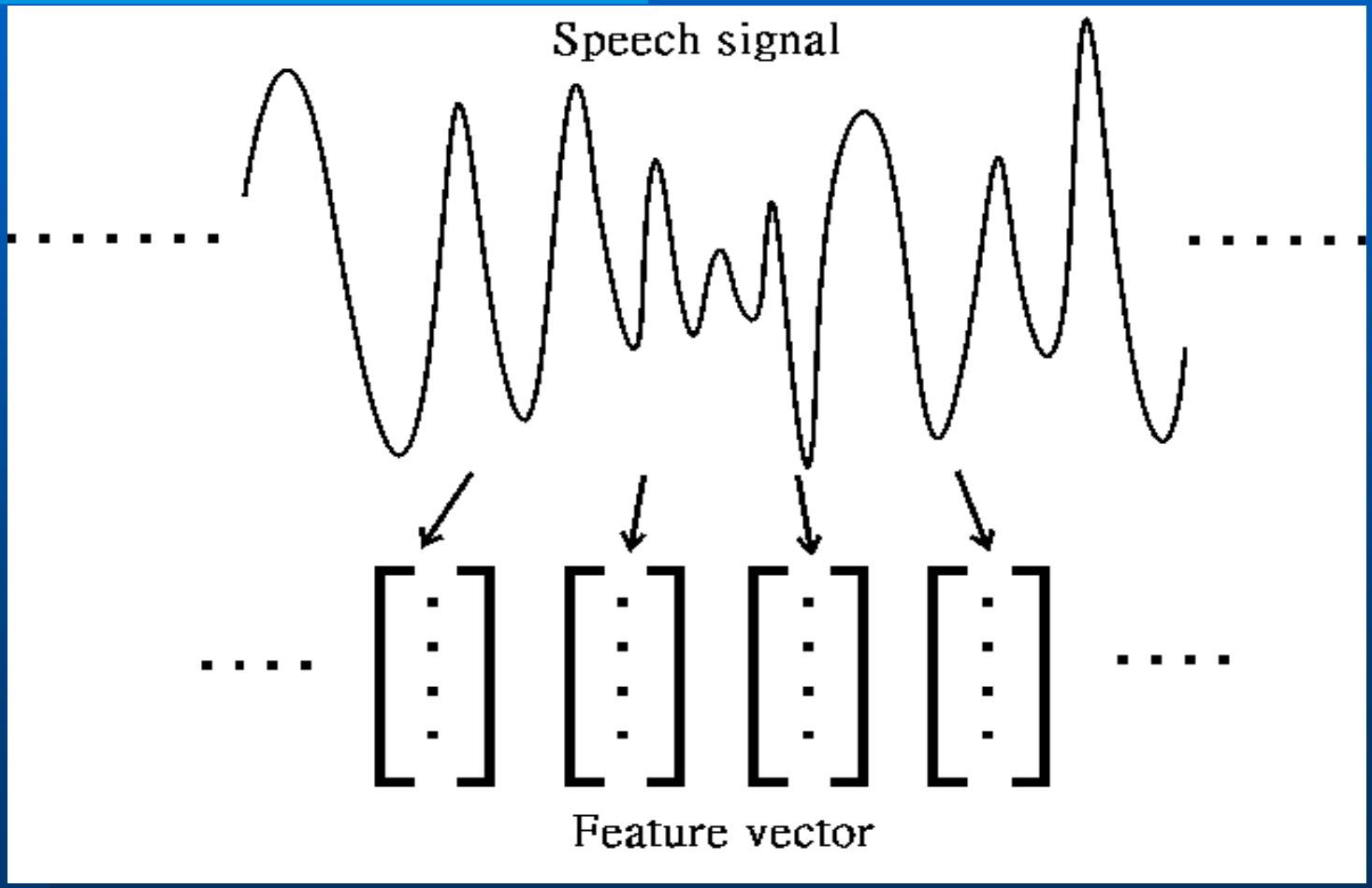
# 語者驗證(比對)模式



# 語音訊號處理及特徵抽取

- 語音訊號前處理
- 語音特徵抽取
  - 線性預測編碼導出的倒頻譜參數（Short-Time Cepstral Coefficient - LPCC）
  - 梅爾刻度式倒頻譜參數（Mel-Frequency Cepstral Coefficient - MFCC）

# 語音訊號之音框(Frame)



# 語音訊號前處理

- 漢明窗（Hamming Window）

$$W(n)=\begin{cases}0.54-0.46\times\cos(\dfrac{2\pi n}{N-1}) & 0\le n\le N-1\\ 0 & otherwise\end{cases}$$

- 正規化（Normalization）
- 預強調（Pre-Emphasis）

$$h(z)=1-az^{-1}$$

# 語音訊號特徵—LPC

1.求相關係數：

$$R(k)=\sum_{n=-\infty}^{\infty}x(n)x(n+k) \qquad k=0,\cdots P+1$$

2. 求線性預測編碼（LPC）係數：

$$\alpha_i^{(P)} \quad 1\le i\le P$$

Levinsion-Durbin演算法

$$E^{(0)}=R(0)$$

$$for\ \ i=1:P$$

$$k_i=R(i)-\sum_{j=1}^{i-1}\alpha_j^{(i-1)}R(i-j)$$

$$\alpha_{\mathrm{i}}^{(\mathrm{i})}=k_i \Big/ E^{(i-1)}$$

$$for\ \ j=1:i-1$$

$$\alpha_{\mathrm{j}}^{(\mathrm{i})}=\alpha_j^{(i-1)}-k_i\alpha_{i-j}^{(i-1)}$$

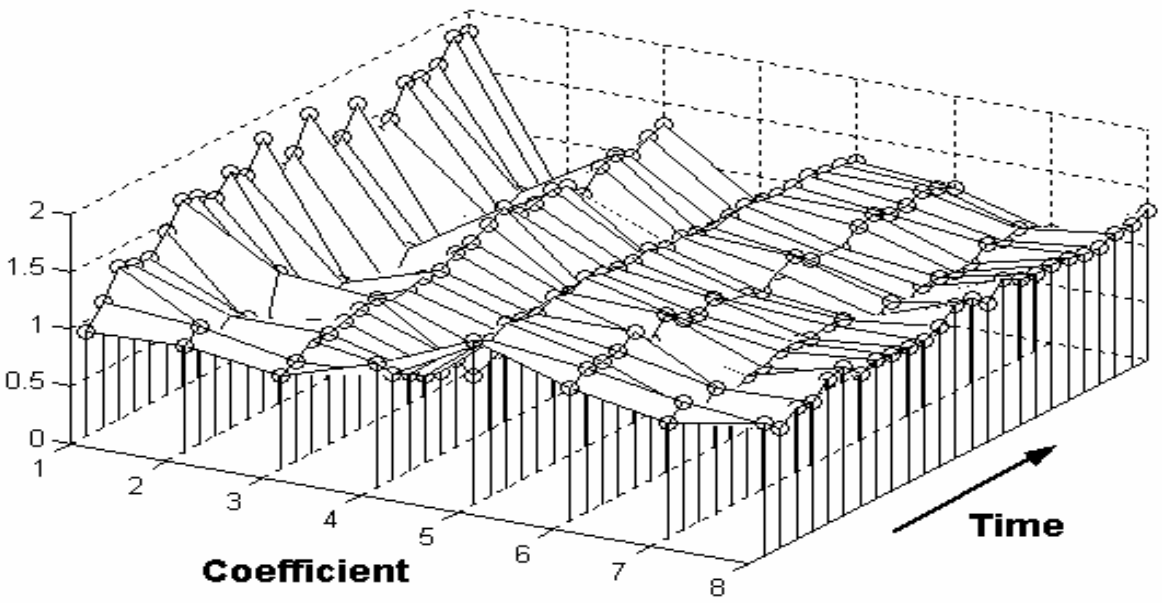
end

$$\mathrm{E}^{(\mathrm{i})}=(1-k_i^2)E^{(i-1)}$$

*end*

# 特徵抽取—LPCC

$$c_m = a_m + \sum_{k=1}^{m-1} (\frac{k}{m}) c_k a_{m-k} \qquad 1 \le m \le P$$



# LPCC特徵向量

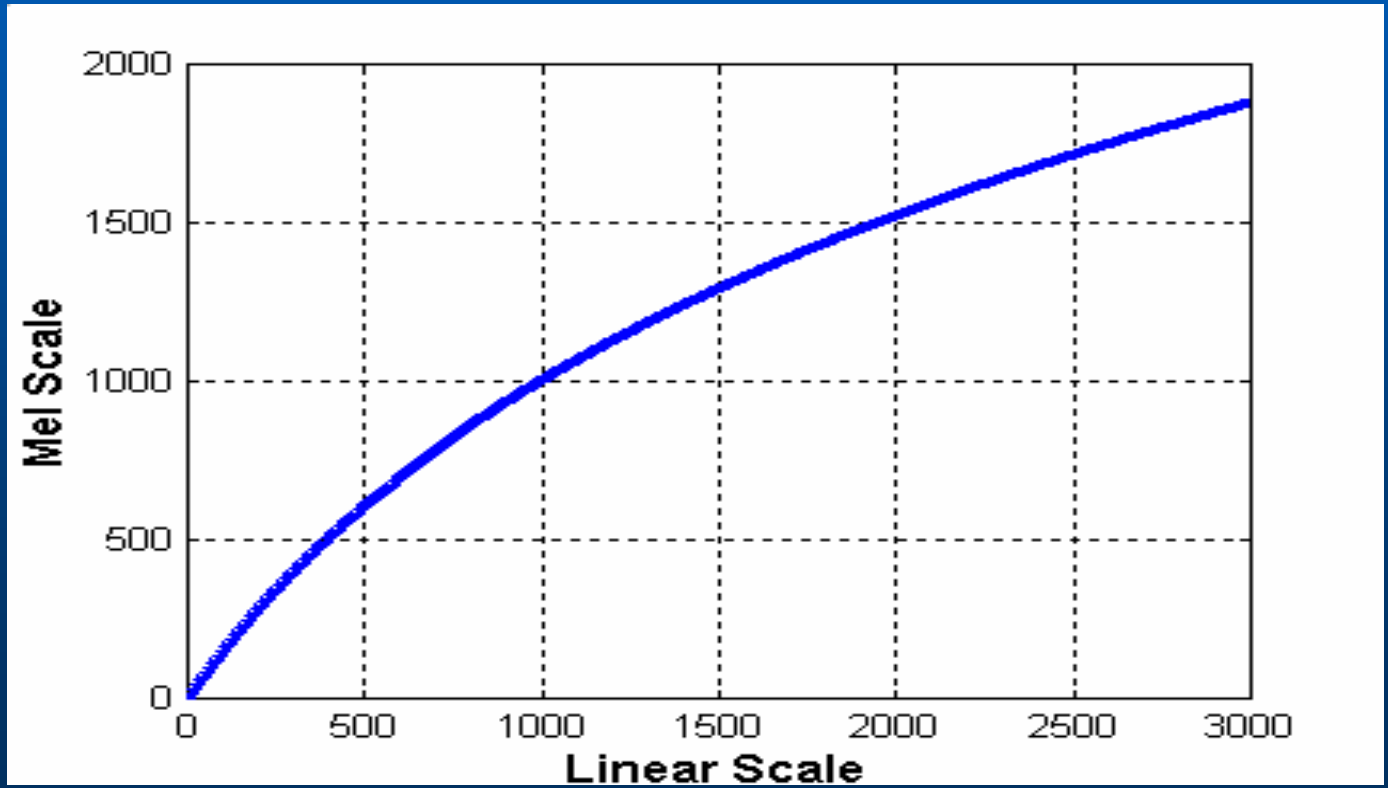
- LPCC比LPC的強健性（Robustness）及可信賴度（Reliability）都要來的高。

- 缺點：將聲音在每段頻率上的特性視為線性，這與實際上人類的聽覺反應並不一致。

# 特徵抽取—MFCC

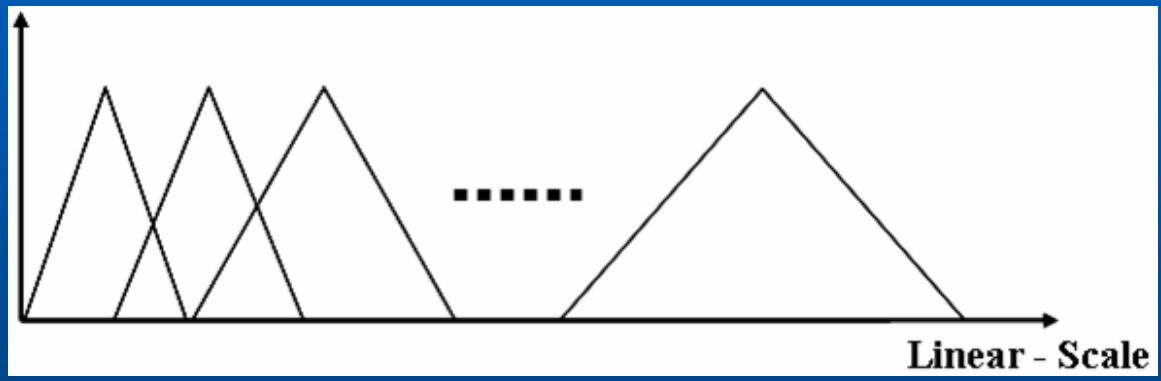
Mel Scale：

$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700})$$

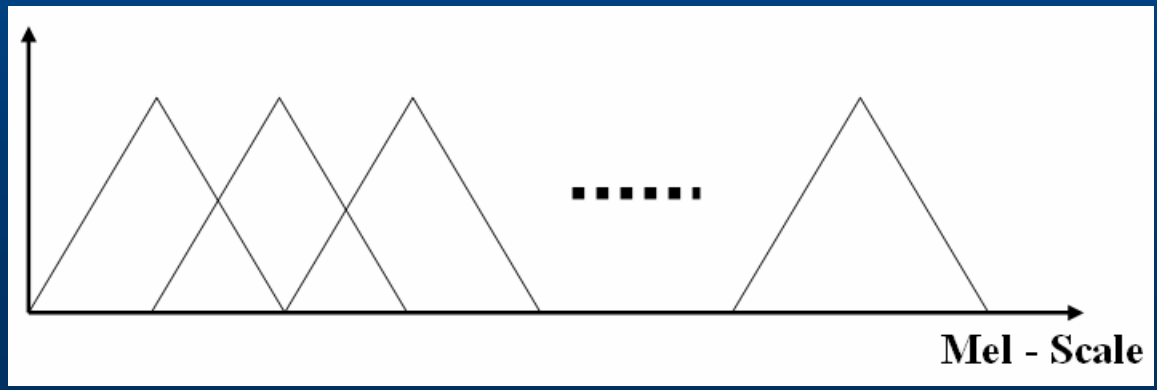


# Linear–Scale和Mel–Scale

**Linear - Scale**




**Mel – Scale**



# 三角帶通濾波器

三角帶通濾波器的函數值：

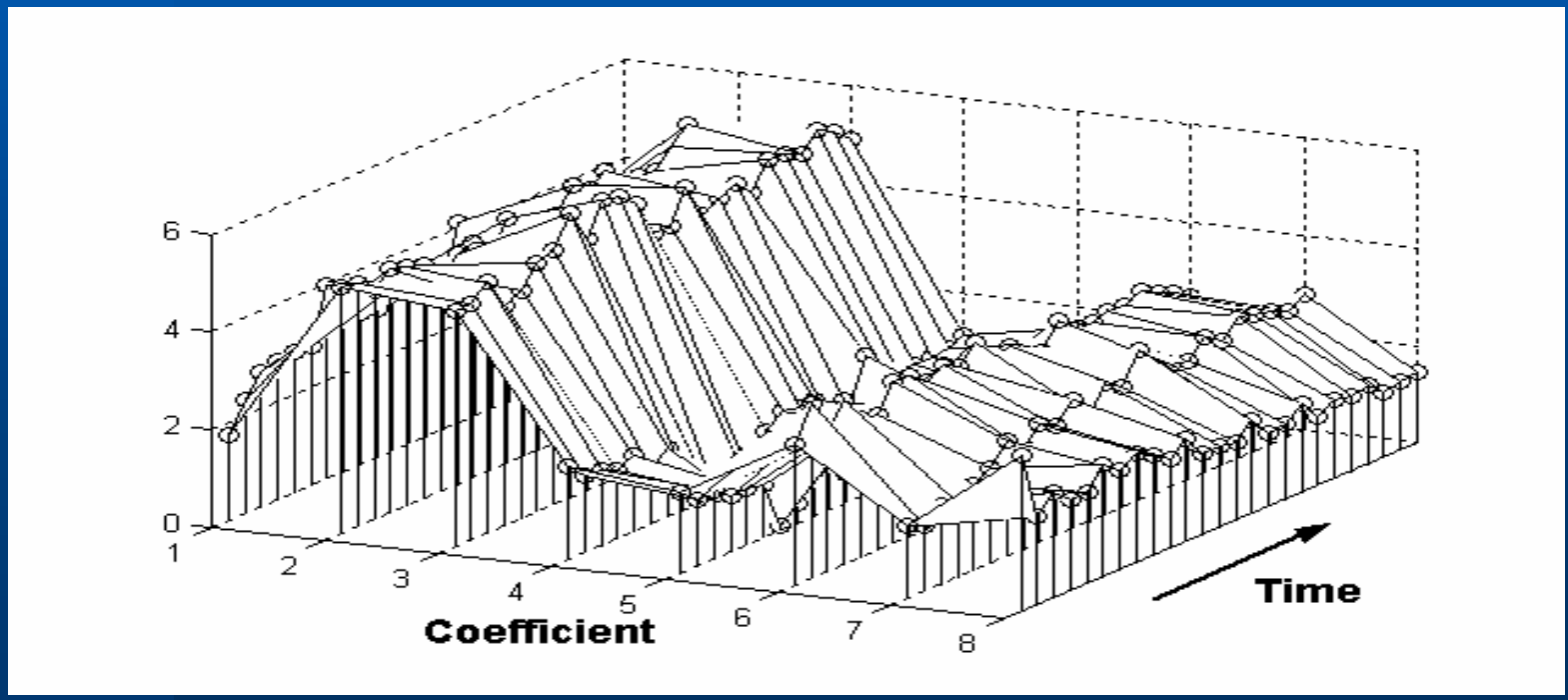
$$U_m(n)=\begin{cases}1-\dfrac{|b_m-n|}{\Delta m} & if \quad |b_m-n|\le \Delta m \\ 0 & if \quad |b_m-n|>\Delta m\end{cases}$$

帶通濾波器之輸出：

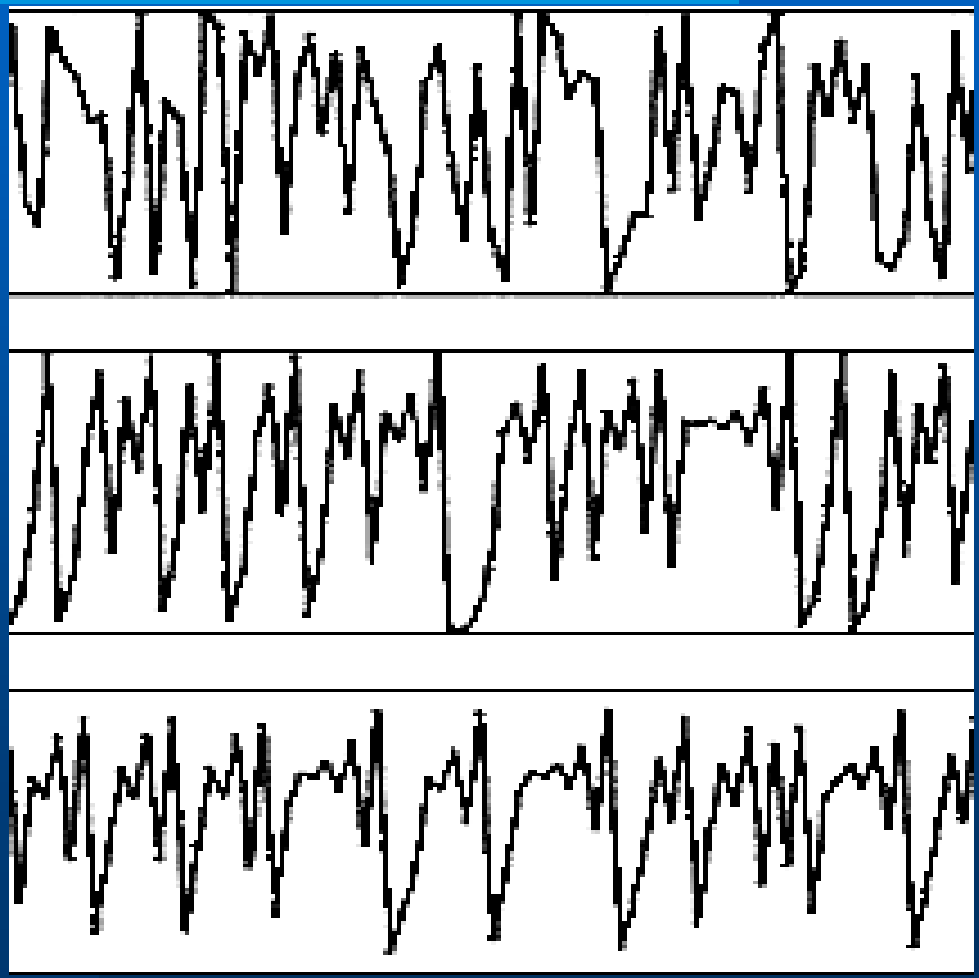
$$Y(m)=\sum_{n=b_m-\Delta m}^{b_m+\Delta m} X(n)U_m(n-b_m)$$

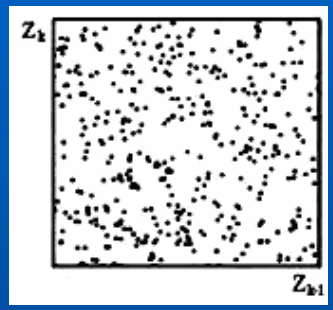
# MFCC

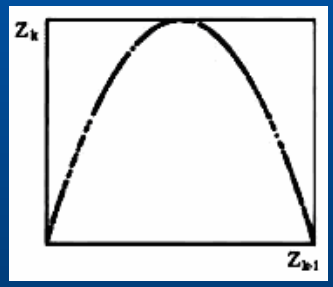
$$C(k) = \sum_{m=1}^{M} \log\{|Y(m)|\} \cdot \cos\{k(m - \frac{1}{2})\frac{\pi}{M}\} \quad , \quad k = 1,2,\cdots,P$$

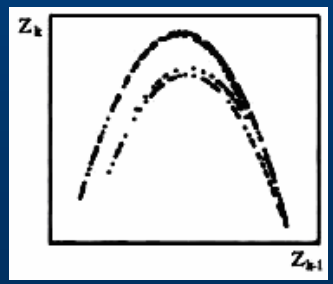


# 語音與混沌時間序列





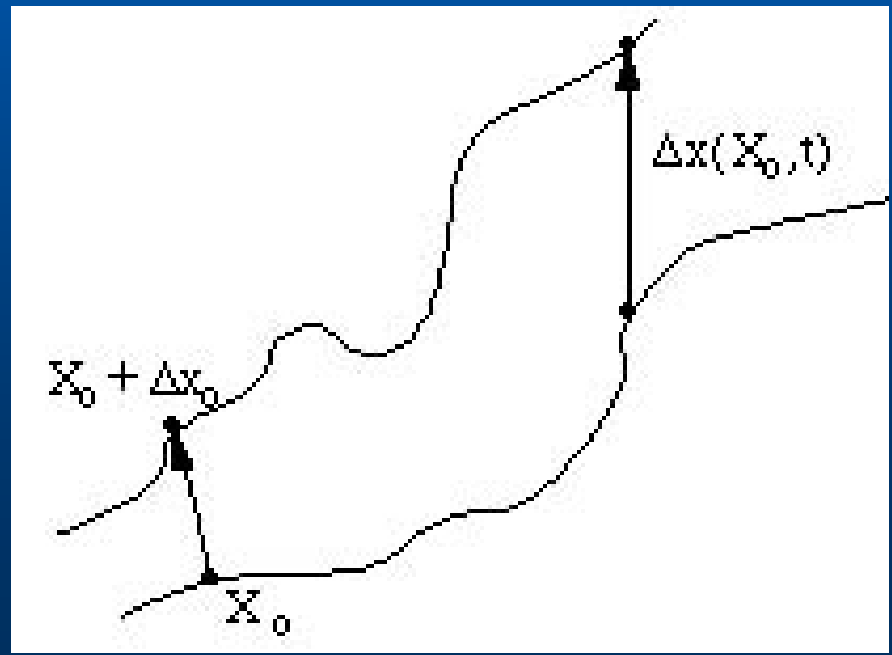



# 里雅普諾夫指數 (Liapunov exponents)

Benettin等人於1976年提出一個計算里雅普諾夫指數的方法：將動態系統特定時間狀態作為初始狀態，兩個軌跡相臨點間的距離為$\Delta x_0(X_0)$，經過時間t後兩點距離變化為$\Delta x(X_0,t)$

- 里雅普諾夫指數λ：

$$\lambda = \lim_{\substack{t\to\infty \\ |\Delta x_0|\to 0}} \frac{1}{t} \ln \left| \frac{\Delta x(X_0,t)}{\Delta x_0} \right|$$



# 里雅普諾夫指數演算法 (Liapunov exponents)

- 對一組離散的語音訊號 $x_0, x_1, x_2 \ldots\ldots x_n$ 中任一 $x_i$ $(0 \leqq i \leqq n)$ 搜尋一個與其值最接近的 $x_j$ $(0 \leqq j \leqq n)$ 。

  求出初始距離 ： $d_{ij} = \left| x_i - x_j \right|$

  觀察經過時間t後的距離 ： $dt_{ij} = \left| x_{i+t} - x_{j+t} \right|$

- $x_i$ 的里雅普諾夫指數：

$$\lambda_i = \frac{1}{t} \ln \left| \frac{dt_{ij}}{d_{ij}} \right|$$

# PCA語音特徵擷取

Mean Value：

$$m_v = \frac{1}{M}\sum_{k=1}^{M} X_k$$

Covariance Matrix：

$$E = (X - M)(X - M)'$$

PCA：

$$V = \left\{ \begin{bmatrix} \\ \\ \end{bmatrix}, \begin{bmatrix} \\ \\ \end{bmatrix}, \cdots, \begin{bmatrix} \\ \\ \end{bmatrix} \right\}$$

# 碎形語音特徵擷取

Iterated Function System(IFS)

$$a_i x_0 + e_i = x_{i-1}$$

$$a_i x_N + e_i = x_i$$

$$c_i x_0 + d_i F_0 + F_i = F_{i-1}$$

$$c_i x_N + d_i F_N + f_i = F_i$$

$a_i$、$c_i$、$d_i$、$e_i$和$f_i$是IFS係數

# 碎形語音特徵擷取

$$x_i^* = \begin{bmatrix} c_i & d_i & f_i \end{bmatrix}$$

$$A^* = \begin{bmatrix} x_1^* & x_2^* & \cdots & x_N^* \end{bmatrix}^T$$

$$m_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}^*$$

$$S_j^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_{ij}^* - m_j\right)^2$$

# 小波語音特徵擷取

1. 計算每一層小波轉換後的語音訊號係數的平均能量：

$$v_i = \frac{1}{n_i}\sum_{j=1}^{n_i} w_{i,j}^2 \quad i = 1,2,\cdots,N+1$$

2.將每一層$i$平均能量當小波轉換的語音訊號特徵值：

$$V = \{v_1, v_2, \cdots, v_i\}^t$$

# 期末報告專題

- Ear Biometrics
- Lips Biometrics(鄧)
- Hand Geometry Biometrics(邱)
- Palmprint Biometrics
- Gait  Biometrics
- Handwriting(Signature) Biometrics(黃)
- Keystroke Biometrics(謝昇憲)
- Voice Biometrics(陳俊任)
- Multimodal  Biometrics(朱家德)