

VI. Brief Introduction for Acoustics

[參考資料]

- 王小川，“語音訊號處理”，全華出版，台北，民國94年。
- T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principle and Practice*, Pearson Education Taiwan, Taipei, 2005.
- L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- P. Filippi, *Acoustics : Basic Physics, Theory, and Methods*, Academic Press, San Diego, 1999.

◎ 6-A 聲音的相關常識

206

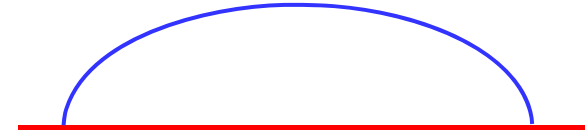
人耳可以辨識頻率：20Hz ~ 20000Hz

說話：150~2000Hz

電話系統頻域：小於 3500Hz

電腦音效卡取樣頻率：44100Hz (最新技術可達192K)

(一般用 22050Hz, 11025Hz 即可)



> 20000Hz: 超音波 (ultrasound)

< 20Hz: 次聲波 (infrasound)

波長較長 -> 傳播距離較遠，但容易散射

波長較短 -> 衰減較快，但傳播方向較接近直線

- 一般聲音檔格式：
 - (1) 取樣頻率 22050Hz
 - (2) 單聲道或雙聲道
 - (3) 每筆資料用8個bit來表示

- 電腦中沒有經過任何壓縮的聲音檔： *.wav

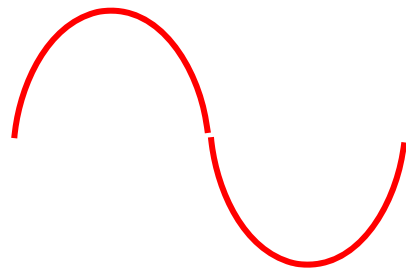
Q: What is the data size of a song without compression?

- 數位電話取樣頻率：8000Hz

聲音在空氣中傳播速度：每秒 340 公尺 (15°C 時)

所以，人類對 3000Hz 左右頻率的聲音最敏感

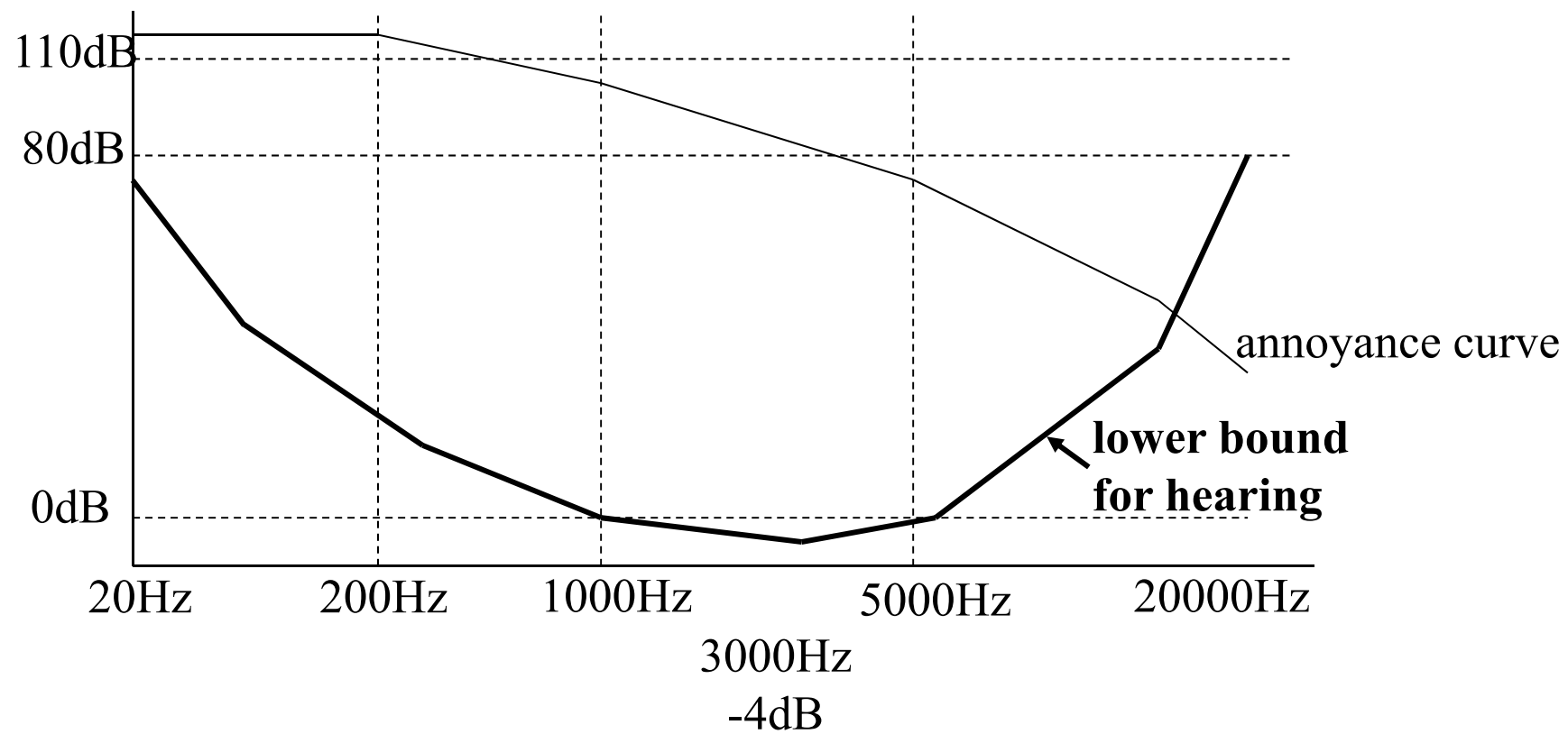
(一般人，耳翼到鼓膜之間的距離：2.7公分)



附：(1) 每增加 1°C，聲音的速度增加 0.6 m/sec

(2) 聲音在水中的傳播速度是 1500 m/sec

在鋁棒中的傳播速度是 5000 m/sec



- dB: 分貝 $10\log_{10}(P/C)$ ，其中P為音強(正比於振幅的平方)；C為0dB時的音強

每增加 10dB，音強增加10倍；每增加3dB，音強增加2倍；
所幸，內耳的振動不會正比於聲壓

- 人對於頻率的分辨能力，是由頻率的「比」決定

對人類而言，300Hz 和 400 Hz 之間的差別，與 3000Hz 和 4000 Hz 之間的差別是相同的

◎ 6-B Music Signal

電子琴 Do 的頻率：低音 Do: 131.32 Hz
 中音 Do: 261.63 Hz
 高音 Do: 523.26 Hz
 更高音 Do: 1046.52 Hz,

音樂每增加八度音，頻率變為 2 倍

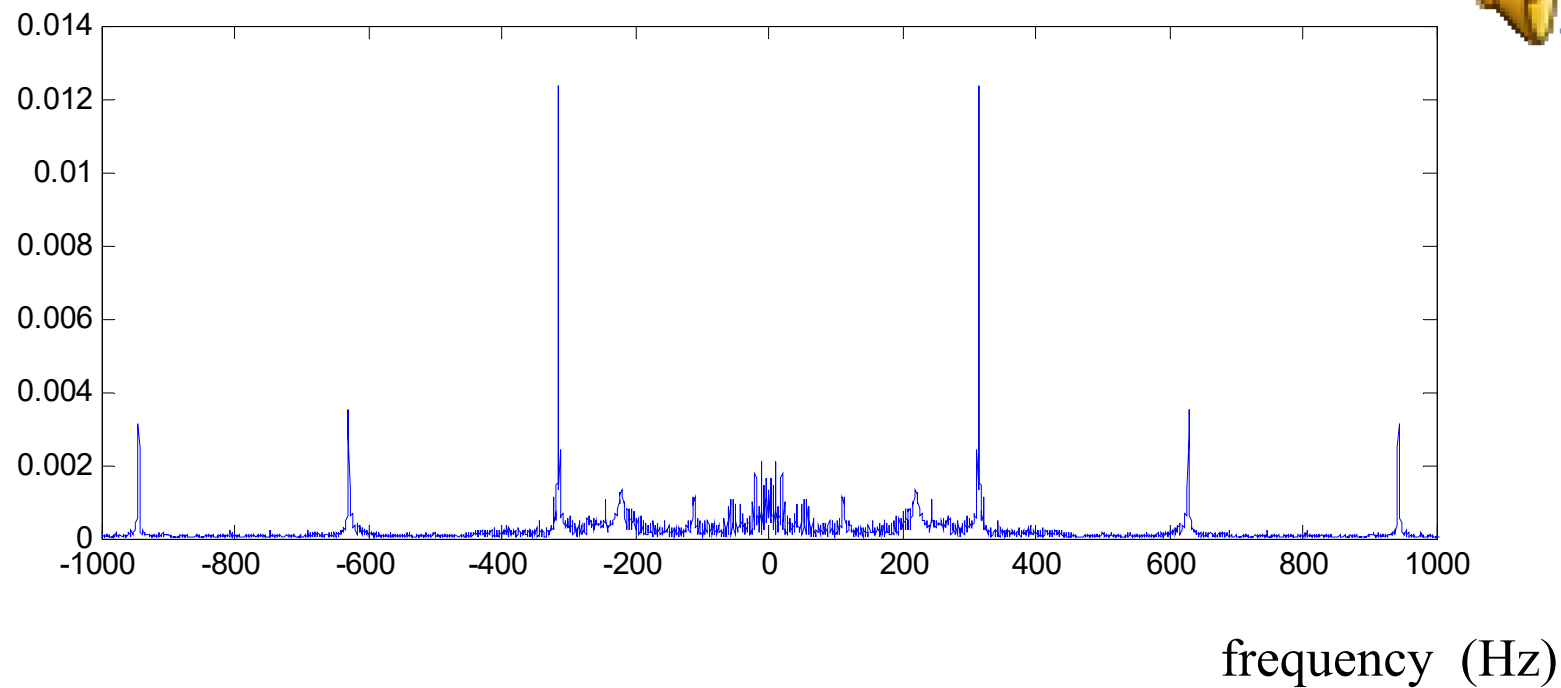
每一音階有 12 個半音

增加一個半音，頻率增加 $2^{1/12}$ 倍 (1.0595 倍)

| | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | Do | 升Do | Re | 升Re | Mi | Fa | 升Fa | So | 升So | La | 升La | Si |
| Hz | 262 | 277 | 294 | 311 | 330 | 349 | 370 | 392 | 415 | 440 | 466 | 494 |

音樂通常會出現「和弦」(chord) 的現象

除了基頻 f_0 Hz 之外，也會出現 $2f_0$ Hz, $3f_0$ Hz, $4f_0$ Hz, 的頻率



為什麼會產生和弦？

以共振的觀點：

聲音信號是一個 periodic signal，但是不一定是 sinusoid

◎ 6-C 語音處理的工作

- (1) 語音編碼 (Speech Coding)
- (2) 語音合成 (Speech Synthesis)
- (3) 語音增強 (Speech Enhancement)
 前三項目前基本上已經很成功
- (4) 語音辨認 (Speech Recognition)
 音素 → 音節 → 詞 → 句 → 整段話
 目前已有很高的辨識率
- (5) 說話人辨認 (Speaker Recognition)
- (6) 其他：語意，語言，情緒

◎ 6-D 語音的辨認

音素 → 音節 → 詞 → 句 → 整段話
音素：相當於一個音標

(1) Spectrum Analysis

Time-Frequency Analysis

(2) Cepstrum

(3) Correlation for Words

◎ 6-E 子音和母音

ㄅ ㄆ ㄇ ㄏ ㄏㄨㄥ ㄏㄨㄚ ㄏㄨㄛ ㄏㄨㄣ ㄏㄨㄥ ㄏㄨㄥ
 ㄩ ㄩㄛ ㄩㄛㄛ ㄩㄛㄛㄛ ㄩㄛㄛㄛㄛ ㄩㄛㄛㄛㄛㄛ ㄩㄛㄛㄛㄛㄛㄛ

母音：ㄩ ㄩㄛ ㄩㄛㄛ ㄩㄛㄛㄛ ㄩㄛㄛㄛㄛ ㄩㄛㄛㄛㄛㄛ ㄩㄛㄛㄛㄛㄛㄛ

單母音：a, e, i, o, u ㄩ ㄩㄛ ㄩㄛㄛ ㄩㄛㄛㄛ ㄩㄛㄛㄛㄛ

雙母音：ㄩㄛㄛㄛ ㄩㄛㄛㄛㄛ

母音+濁音：ㄩㄛㄛㄛㄛ ㄩㄛㄛㄛㄛㄛ

子音：ㄅ ㄆ ㄇ ㄏ ㄏㄨㄥ ㄏㄨㄚ ㄏㄨㄛ ㄏㄨㄣ ㄏㄨㄥ ㄏㄨㄥ

| | | | | | | | | | | | | | | |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ㄅ | ㄆ | ㄇ | ㄈ | ㄉ | ㄊ | ㄋ | ㄌ | ㄍ | ㄎ | ㄏ | ㄐ | ㄑ | ㄒ |
| 漢語拼音 | b | p | m | f | d | t | n | l | g | k | h | j | q | x |
| 通用拼音 | b | p | m | f | d | t | n | l | g | k | h | j | c | s |

| | | | | | | | | | | | | | | |
|------|----|----|----|---|---|---|---|---|---|---|---|----|----|----|
| | ㄗ | ㄘ | ㄙ | ㄨ | ㄩ | ㄑ | ㄒ | ㄚ | ㄛ | ㄜ | ㄝ | ㄞ | ㄟ | ㄠ |
| 漢語拼音 | zh | ch | sh | r | z | c | s | a | o | e | e | ai | ei | ao |
| 通用拼音 | jh | ch | sh | r | z | c | s | a | o | e | e | ai | ei | ao |

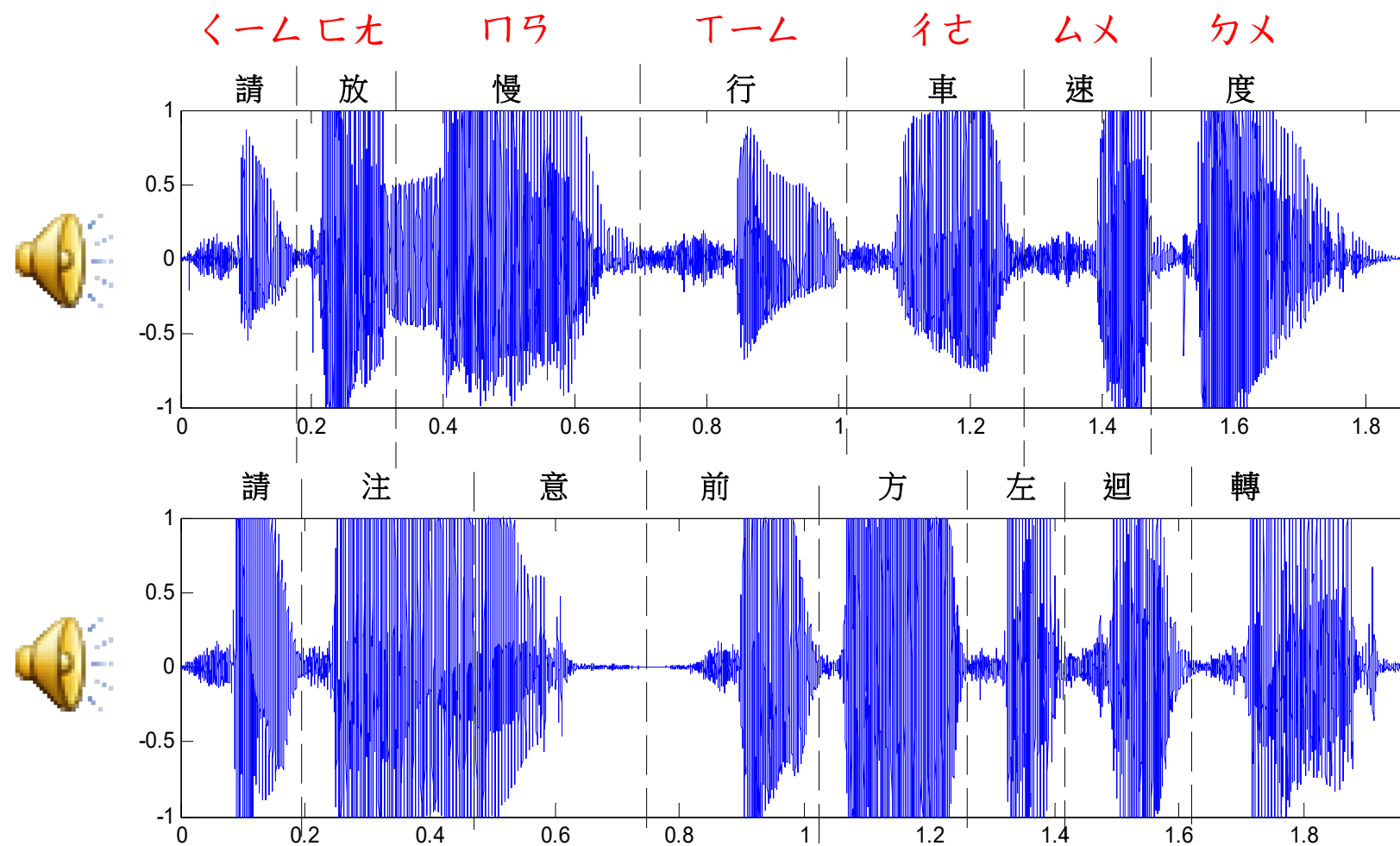
| | | | | | | | | | |
|------|----|----|----|-----|-----|----|------|------|--------|
| | ㄨ | ㄢ | ㄣ | ㄤ | ㄥ | ㄦ | ㄨ | ㄩ | ㄩ |
| 漢語拼音 | ou | an | en | ang | eng | er | i, y | u, w | yu, iu |
| 通用拼音 | ou | an | en | ang | eng | er | i, y | u, w | yu, iu |

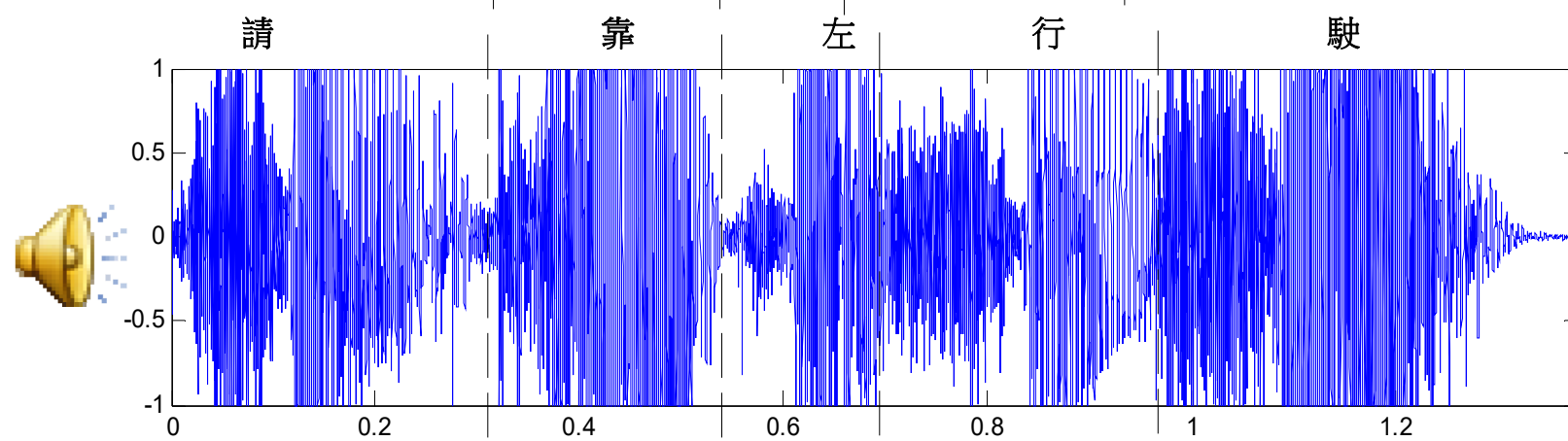
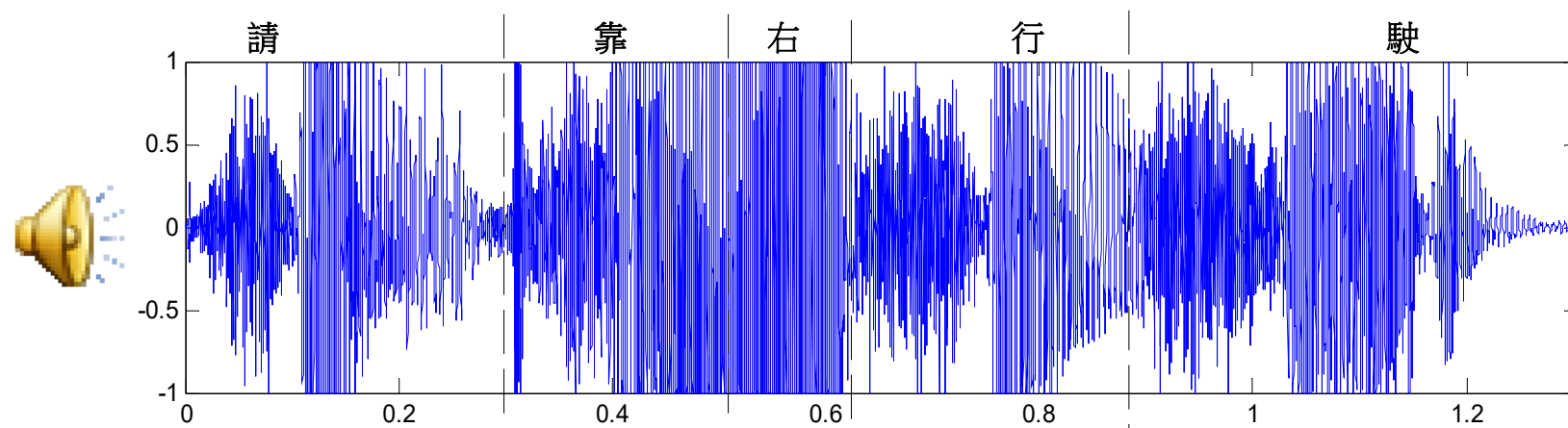
母音：依唇型而定

子音：在口腔，鼻腔中某些部位將氣流暫時堵住後放開

子音的能量小，頻率偏高，時間較短，出現在母音前

母音的能量大，頻率偏低，時間較長，出現在子音後或獨立出現





發音模型 (線性非時變近似)

$$X(z) = R(z)H(z)G(z)E_p(z)$$

$R(z)$: 嘴唇模型, $H(z)$: 口腔模型, $G(z)$: 聲帶模型

$E_p(z)$: 輸入(假設為週期脈衝)

音量和 $E_p(z)$, $G(z)$ 有關

子音和 $H(z)$, $R(z)$ 有關

母音和 $R(z)$ 有關

- 分析一個聲音信號的頻譜：

用 **Windowed Fourier Transform**

或稱作 **Short-Time Fourier Transform**

- Fourier transform

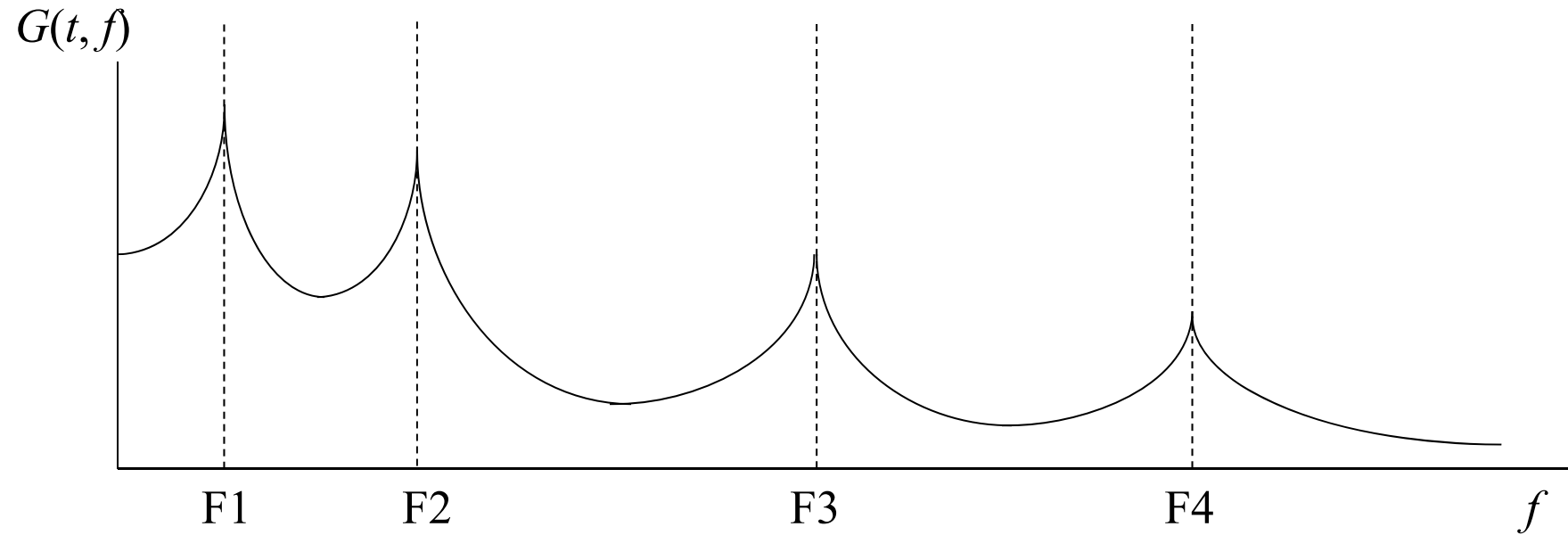
$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt$$

Windowed Fourier transform

$$G(f) = \int_{t_0-B}^{t_0+B} g(t) e^{-j2\pi ft} dt \quad \text{強調 } t = t_0 \text{ 附近的區域}$$

或
$$G(t, f) = \int_{-\infty}^{\infty} w(t-\tau) g(\tau) e^{-j2\pi f\tau} d\tau$$

典型的聲音頻譜 (不考慮倍頻)：



頻譜上，大部分的地方都不等於0。

出現幾個 peaks 值

可以依據 peaks 的位置來辨別母音

母音 peaks 處的頻率 (Hz) (不考慮倍頻)：

| | 男聲 | | | 女聲 | | |
|---|-----|------|------|------|------|------|
| | F1 | F2 | F3 | F1 | F2 | F3 |
| ㄚ | 900 | 1200 | 2900 | 1100 | 1350 | 3100 |
| ㄛ | 560 | 800 | 3000 | 730 | 1100 | 3200 |
| ㄜ | 560 | 1090 | 3000 | 790 | 1250 | 3100 |
| ㄝ | 500 | 2100 | 3100 | 600 | 2400 | 3300 |
| ㄟ | 310 | 2300 | 3300 | 360 | 3000 | 3500 |
| ㄞ | 370 | 540 | 3400 | 460 | 820 | 3700 |
| ㄡ | 300 | 2100 | 3400 | 350 | 2600 | 3200 |
| ㄩ | 580 | 1500 | 3200 | 760 | 1700 | 3200 |

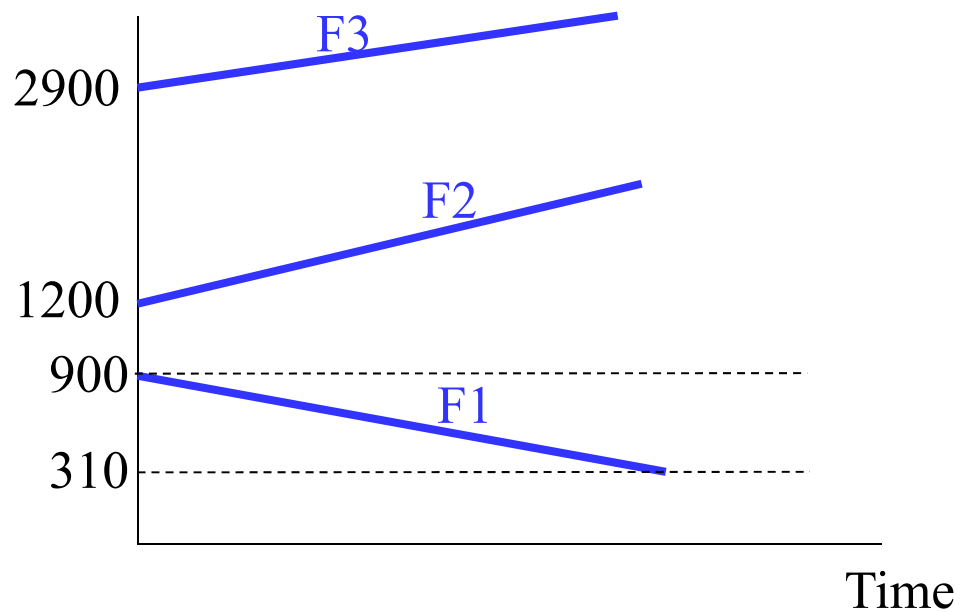
原則上：(1) 嘴唇的大小，決定F1

(2) 舌面的高低，決定 F2 – F1

- 雙母音：
 ㄚ (ai) ㄟ (ei) ㄠ (ao) ㄡ (ou)

頻譜隨時間而改變，一開始始像第一個母音，後變得像另一個母音

ㄚ 的頻譜的 peaks 位置



◎ 6-F 語意學的角色

以「語意學」或「機率」來補足語音辨識的不足

例如：經過判定，一個聲音可能是

ㄅ一 ㄇ ㄅ ㄅ一 ㄇ ㄅ

ㄅ一 ㄆ ㄅ ㄅ一 ㄆ ㄅ

這個聲音是「必然」的機率比較大。

ㄅㄆ ㄅㄆ ㄅㄆ ㄅㄆ

可能是「伯伯」，也可能是「婆婆」，看上下文

儲存詞庫

- 當前主流的語音辨識技術：

Mel-Frequency Cepstrum + 語意分析 + Machine Learning (人工智慧的一種)

附錄七之一：線性代數觀念補充

(1) \mathbf{x} 和 \mathbf{y} 兩個向量的內積可表示成 $\langle \mathbf{x} | \mathbf{y} \rangle$

(2) 兩個互相正交(orthogonal)或垂直(perpendicular)的向量，其內積為0。
可表示成： $\langle \mathbf{x} | \mathbf{y} \rangle = 0$ 或 $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

(3) 令 S 為內積空間 V 的一組正交集合(set)且由非零向量構成，

$$\text{其中 } \mathbf{x} = \sum_{\mathbf{y} \in S} a_{\mathbf{y}} \mathbf{y}, \quad a_{\mathbf{y}} = \frac{\langle \mathbf{x} | \mathbf{y} \rangle}{\langle \mathbf{y} | \mathbf{y} \rangle}$$

如果 S 是由一組正規集合(orthonormal set)構成，那麼 $a_{\mathbf{y}} = \langle \mathbf{x} | \mathbf{y} \rangle$

(4) Gram-Schmidt algorithm: 對於內積空間 V 的任意一組基底 $\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \rangle$ ，我們可以透過這演算法找到一組正交基底 $\langle \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n \rangle$

$$\mathbf{y}_j = \mathbf{x}_j - \sum_{i=1}^{j-1} \frac{\langle \mathbf{x}_j | \mathbf{y}_i \rangle}{\langle \mathbf{y}_i | \mathbf{y}_i \rangle} \mathbf{y}_i \quad \text{for each } j = 2, \dots, n$$

幾何意義: 把 \mathbf{x}_j 在 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{j-1}$ 上面的分向量全都從向量 \mathbf{x}_j 身上扣掉之後，剩下的向量 \mathbf{y}_j 自然就會跟 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{j-1}$ 垂直。

(5) Solving $\mathbf{Ax} = \mathbf{b}$ but $\text{size}(\mathbf{A}) = m \times n$ and $\mathbf{b} \in F^m$, $m > n$

Interpolation Theorem (插值定理)

1. For any inner-product function of F^m , there exists a vector \mathbf{z} that minimizes

$$\|\mathbf{Az} - \mathbf{b}\| \quad \text{where } \mathbf{z} \in F^n$$

2. If $\text{rank}(\mathbf{A}) = n$, then $\mathbf{z} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$ is the unique minimizer of $\|\mathbf{Az} - \mathbf{b}\|$

附錄七之二：PCA and SVD

PCA (principle component analysis) 是資料分析和影像處理當中常用到的數學方法，用來分析資料的「主要成分」或是影像中物體的「主軸」。

它其實和各位同學在高中和大一線代所學的回歸線 (regressive line) 很類似。回歸線是用一條一維 (one-dimensional) 的直線來近似二維 (two-dimensional) 的資料，而 PCA 則是用 M -dimensional data 來近似 N -dimensional data，其中 M 小於等於 N

在講解 PCA 之前，先介紹什麼是 SVD (singular value decomposition)

我們在大一的時候，都已經學到該如何對於 $N \times N$ 的矩陣做 eigenvector-eigenvalue decomposition

那麼.....

當一個矩陣的 size 為 $M \times N$ ，且 M 和 N 不相等時，我們該如何對它來做 eigenvector-eigenvalue decomposition?

SVD 的流程：

假設 \mathbf{A} 是一個 $M \times N$ 的矩陣。

(Step 1) 計算

$$\mathbf{B} = \mathbf{A}^H \mathbf{A} \quad \mathbf{C} = \mathbf{A} \mathbf{A}^H$$

注意， \mathbf{B} 是 $N \times N$ 的矩陣，而 \mathbf{C} 是 $M \times M$ 的矩陣。上標H代表 Hermitian matrix，相當於做共軛轉置。

(Step 2) 接著，對 \mathbf{B} 和 \mathbf{C} 做 eigenvector-eigenvalue decomposition

$$\mathbf{B} = \mathbf{V} \mathbf{D} \mathbf{V}^{-1} \quad \mathbf{C} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1}$$

其中 \mathbf{V} 的每一個 column 是 \mathbf{B} 的 eigenvector (with normalization)， \mathbf{U} 的每一個 column 是 \mathbf{C} 的 eigenvector (with normalization)， $\mathbf{\Lambda}$ 和 \mathbf{D} 都是對角矩陣， $\mathbf{\Lambda}$ 和 \mathbf{D} 對角線上的 entries 是 \mathbf{B} 和 \mathbf{C} 的 eigenvalues。並假設 eigenvectors 根據 eigenvalues 的大小排序 (由大到小)

Note: 值得注意的是，由於 $\mathbf{B} = \mathbf{B}^H$ 且 $\mathbf{C} = \mathbf{C}^H$ ，所以 \mathbf{B} 和 \mathbf{C} 的 eigenvectors 皆各自形成一個 orthogonal set。經過適當的 normalization 使得 \mathbf{U} 和 \mathbf{V} 的 column 自己和自己的內積為 1 之後， $\mathbf{U}^{-1} = \mathbf{U}^H$ 和 $\mathbf{V}^{-1} = \mathbf{V}^H$ 將滿足。因此， \mathbf{B} 和 \mathbf{C} 可以表示成

$$\mathbf{B} = \mathbf{V} \mathbf{D} \mathbf{V}^H \quad \mathbf{C} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$$

注意， \mathbf{V} 和 \mathbf{U} 是 unitary matrix

(Step 3) 計算

$$\mathbf{S}_1 = \mathbf{U}^H \mathbf{A} \mathbf{V} \quad \mathbf{S} = |\mathbf{S}_1| \quad \text{取絕對值}$$

\mathbf{S} 是一個 $M \times N$ 的矩陣，只有在 $\mathbf{S}[n, n]$ ($n = 1, 2, \dots, \min(M, N)$) 的地方不為 0

(Step 4) 若 $\mathbf{S}_1[n, n] < 0$ ，改變 \mathbf{U} 第 n 個 column 的正負號

即完成 SVD

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^H$$

\mathbf{A} 也可以表示為

$$\mathbf{A} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \lambda_k \mathbf{u}_k \mathbf{v}_k^T$$

其中 $\lambda_n = \mathbf{S}[n, n]$, $k = \min(M, N)$

註：Matlab 有內建的 `svd` 指令可以計算 SVD

從 SVD 到 PCA (principle component analysis , 主成份分析)

$$\mathbf{A} = \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \lambda_k \mathbf{u}_k \mathbf{v}_k^T \quad k = \min(M, N)$$

若 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k$

$\lambda_1 \mathbf{u}_1 \mathbf{v}_1^T$ 是 A 矩陣的最主要的成份

$\lambda_2 \mathbf{u}_2 \mathbf{v}_2^T$ 是 A 矩陣的第二主要的成份

⋮

$\lambda_k \mathbf{u}_k \mathbf{v}_k^T$ 是 A 矩陣的最不重要的成份

若為了壓縮或是去除雜訊的考量，可以選擇 $h < k$ ，使得 \mathbf{A} 可以近似成

$$\mathbf{A} \cong \lambda_1 \mathbf{u}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \lambda_h \mathbf{u}_h \mathbf{v}_h^T$$

PCA 的流程

假設現在有 M 筆資料，每一筆資料為 N dimension

$$\begin{aligned}\mathbf{g}_1 &= [f_{1,1} \ f_{1,2} \ \cdots \ f_{1,N}] \\ \mathbf{g}_2 &= [f_{2,1} \ f_{2,2} \ \cdots \ f_{2,N}] \\ &\vdots \\ \mathbf{g}_M &= [f_{M,1} \ f_{M,2} \ \cdots \ f_{M,N}]\end{aligned}$$

(Step 1) 扣掉平均值，形成新的 data

$$\mathbf{d}_m = [e_{m,1} \ e_{m,2} \ \cdots \ e_{m,N}] \quad m = 1, 2, \dots, M$$

$$\text{其中 } e_{m,n} = f_{m,n} - \tilde{f}_n, \quad \tilde{f}_n = \frac{1}{M} \sum_{m=1}^M f_{m,n}$$

(Step 2) 形成 $M \times N$ 的矩陣 \mathbf{A}

$$\mathbf{A} \text{ 的第 } m \text{ 個 row 為 } d_m, \quad m = 1, 2, \dots, M$$

(Step 3) 對 A 做 SVD 分解

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}} \\ &= \lambda_1 \mathbf{u}_1 \mathbf{v}_1^{\mathbf{T}} + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^{\mathbf{T}} + \dots + \lambda_k \mathbf{u}_k \mathbf{v}_k^{\mathbf{T}} \quad k = \min(M, N) \\ \lambda_1 &\geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_k \end{aligned}$$

(Step 4) 將 A 近似成

$$\mathbf{A} \cong \lambda_1 \mathbf{u}_1 \mathbf{v}_1^{\mathbf{T}} + \lambda_2 \mathbf{u}_2 \mathbf{v}_2^{\mathbf{T}} + \dots + \lambda_h \mathbf{u}_h \mathbf{v}_h^{\mathbf{T}}$$

則每一筆資料可以近似為

$$g_m \cong \lambda_1 u_1[m] \mathbf{v}_1^{\mathbf{T}} + \lambda_2 u_2[m] \mathbf{v}_2^{\mathbf{T}} + \dots + \lambda_h u_h[m] \mathbf{v}_h^{\mathbf{T}} + [\tilde{f}_1 \quad \tilde{f}_2 \quad \dots \quad \tilde{f}_N]$$

除了平均值 $[\tilde{f}_1 \quad \tilde{f}_2 \quad \dots \quad \tilde{f}_N]$ 之外

$\mathbf{v}_1^{\mathbf{T}}$ 是資料的最主要成分， $\mathbf{v}_2^{\mathbf{T}}$ 是資料的次主要成分， $\mathbf{v}_3^{\mathbf{T}}$ 是資料的第三主要成分，以此類推

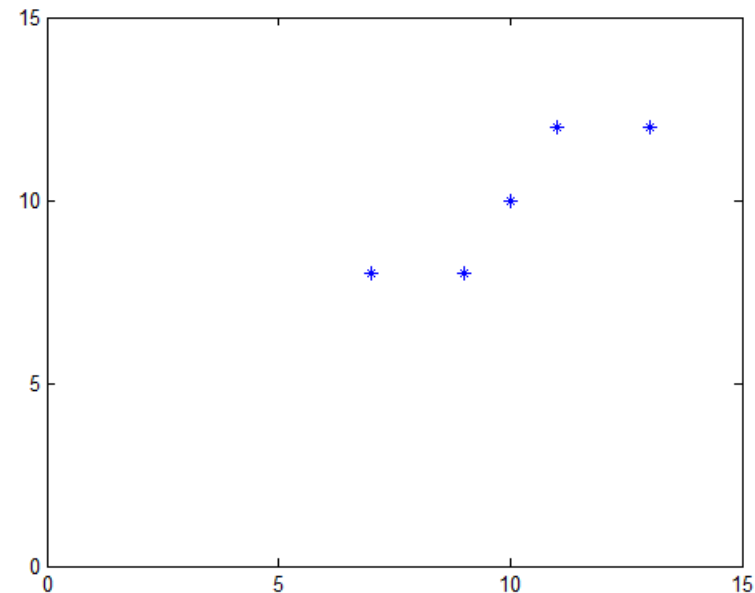
Example of PCA

假設在一個二維的空間中，有5個點，座標分別是

(7,8), (9,8), (10, 10), (11,12), (13,12)

$M = 5, N = 2$

試求這五個點的 PCA (即回歸線)



(Step 1) 將這五個座標點減去平均值 (10, 10)

(-3, -2), (-1, -2), (0, 0), (1, 2), (3, 2)

(Step 2) 形成 5x2 的 matrix

$$\mathbf{A} = \begin{bmatrix} -3 & -2 \\ -1 & -2 \\ 0 & 0 \\ 1 & 2 \\ 3 & 2 \end{bmatrix}$$

(Step 3) 計算 SVD

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H$$

$$\mathbf{U} = \begin{bmatrix} -0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} 5.8416 & 0 \\ 0 & 1.3695 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 0.7497 & -0.6618 \\ 0.6618 & 0.7497 \end{bmatrix}$$

$$\mathbf{A} = 5.8416 \begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} + 1.3695 \begin{bmatrix} 0.3549 \\ -0.6116 \\ 0 \\ 0.6116 \\ -0.3549 \end{bmatrix} \begin{bmatrix} -0.6618 & 0.7497 \end{bmatrix}$$

主成分
次要成分

(Step 4) 得到主成分 $[0.7497 \quad 0.6618]$

這五個座標點可以近似成

$$5.8416 \cdot u_m [0.7497 \quad 0.6618] + [10 \quad 10] \quad m = 1, 2, \dots, 5$$

$$u_1 = -0.6116, \quad u_2 = -0.3549, \quad u_3 = 0, \quad u_4 = 0.3549, \quad u_5 = 0.6116$$

回歸線

$$[10 \quad 10] + c [0.7497 \quad 0.6618]$$

$$c \in (-\infty, \infty)$$

