

行政院國家科學委員會專題研究計畫 成果報告

餘弦轉換統計模型在音訊浮水印之研究

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-009-121-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學電信工程學系(所)

計畫主持人：張文輝

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 12 月 8 日

餘弦轉換統計模型在音訊浮水印之研究
Statistical Analysis of DCT Coefficients
for Audio Watermarking

計畫編號：NSC-92-2213-E-009-121

執行期限：92. 8. 1-93. 7. 31

主持人：張文輝 交通大學電信工程學系所 教授

一、中文摘要

(關鍵詞：數位浮水印、離散餘弦轉換、展頻調變、聲響心理分析)

音訊編碼可以有效運用傳輸頻寬與儲存容量，唯仍須配合數位浮水印的智財權保護以對抗非法複製和線上傳播。理想的浮水印系統設計應兼顧其隱密性與強韌度，前者要求嵌入的認證訊號不能被盜版者察覺其存在，而後者則強調非法的惡意攻擊不致於破壞浮水印的完整性。此計劃應用離散餘弦轉換配合展頻調變原理來建構浮水印的嵌入函數，其關鍵在於設定一能涵蓋音訊轉換係數特性的機率模型，同時依據統計預估原理設計一最佳的浮水印檢測機制，進而提昇檢測器效能及其快速實現的能力。至於浮水印嵌入函數的設計，則是透過人耳聽覺的聲響心理分析，計算一組離散餘弦頻域上的遮蔽臨界曲線，以取得浮水印透明度及強韌度間的最佳平衡。

英文摘要

(Keywords: digital watermarking, discrete cosine transform, spread spectrum modulation, psycho-acoustic model analysis)

Illegal misuses necessitate the development of digital watermarking techniques that allow for copyright protection of audio. In this project, we develop an additive spread spectrum watermarking system in the discrete cosine transform (DCT) domain. The embedded watermark must be imperceptible by the user and should be robust to various signal manipulations. The imperceptibility constraint can be attained by taking into account the frequency masking properties of the human auditory system. Closely related to the robustness is the detection process, which consists in testing whether a given data was watermarked with a certain key. The detection problem can be formulated as a binary hypothesis test with the aim at maximizing the probability of detection conditioned to a given probability of false alarm. We propose a statistical approach to model the DCT coefficients of the audio and then use it as a basis for the application of statistical decision theory to the design of efficient detection structures.

二、計劃緣由與目的

近幾年來網際網路的廣泛應用，配合商業化的多媒體資料壓縮技術，使得數位音樂與影像可以無失真地複製，非法拷貝和線上傳播已嚴重威脅到原創作者的生存空間。為了有效保障其智慧財產權，數位浮水印技術逐漸被應用於保護數位化的多媒體資訊[1]，而利用展頻調變技術將可認證用的訊息隱藏嵌入原始影音訊號頻域的數位浮水印機制則最為常見[2, 3]。在數位浮水印的技術中，影音訊號扮演著展頻通訊中寬頻雜訊的角色，而認證用的浮水印則代表窄頻訊號，經過處理嵌入浮水印的影音檔即可放心的傳播流通。當著作權的確認有爭議時，即可藉由檢測解碼隱含於影音訊號中的浮水印來保障原創作者的正常性。

理想的浮水印系統設計應兼顧其隱密性與強韌度，前者要求嵌入的認證訊號不能被盜版者察覺其存在，而後者則強調非法的惡意攻擊不致於破壞浮水印的完整性。目前已知的相關研究集中在影像浮水印[4, 5]及其相關式檢測[2, 3]，主要是基於頻域係數呈高斯機率分佈的假設而推導其檢測演算法，但這不符合影像頻域係數的實際量測結果[6]，因此我們深信浮水印的檢測機制仍有許多改善的空間。相對而言，音訊浮水印技術的開發仍在起步階段[2, 3]，而音訊在頻域係數的機率分佈模型也未見分析整理。有鑑於此，我們將推導一種廣義的高斯機率函數，以期能有效地描述不同演奏內容音樂的頻域分佈變化。我們鎖定在離散餘弦轉換的頻域進行相關的訊號處理，主要是考慮到該轉換已廣泛地應用於多項音訊壓縮的國際標準格式。更值得深入探討的是，人耳對音樂與人眼對影像的感知分析存在明顯的差異，具體反應在所對應的離散餘弦係數的機率分佈，進而影響其最佳浮水印檢測的快速實現演算法。至於音訊浮水印的隱密需求，我們將考慮人耳聽覺在響度與頻率解析度存在非線性響應的事實，計算其遮蔽臨界曲線再據以調整各個離散餘弦轉換係數的加權比重。

三、研究方法與成果

本計劃旨在探討音訊展頻浮水印技術，並配合離散餘弦轉換係數的機率分佈、人耳聽覺模型分析及最大相似度演算法的推導，設計一最佳的音訊浮水印檢測機制。針對各項研究進行步驟說明如下：

(1) 加成性展頻浮水印機制：

主要是仿造數位通訊系統中對抗干擾的直序展頻通訊技術，把原始音訊當成通道雜訊，並運用虛擬亂數序列調變一組隱藏資訊製成浮水印，再嵌入音訊以保護其智慧財產權。數位浮水印系統可用圖一模型表示，訊號處理流程敘述如下：以 $N_f = 256$ 點做音框切割，再經離散餘弦轉換得其係數向量 $\mathbf{x} = [x_0, x_1, \dots, x_{N_f-1}]^T$ ，配合秘密金鑰 K 及隱藏位元向量 $\mathbf{b} = [b_1, b_2, \dots, b_N]^T$ ，經函數 f 產生認證用的浮水印訊號 \mathbf{w} ，再嵌入 \mathbf{x} 而製成加印(watermarked)音訊 \mathbf{y} 。浮水印的嵌入流程細節如圖二所示，將音框的 N_f 個轉換係數索引劃分成互不相關的 N 個子集合 $\{S_j\}_{j=1}^N$ ，滿足 $S_j \cap S_i = \emptyset, \forall i \neq j$ ，則 S_j 代表第 j 個訊息位元 b_j 將嵌入的轉換係數之索引集合。為不被人耳察覺音訊內含嵌入訊息，我們執行聲響心理模型分析，計算個別係數 x_k 的加權比重值 α_k ，再與用秘密金鑰 K 產生的虛擬亂數 s_k 相乘，製成音訊浮水印訊號 w_k 。明確地說 $\mathbf{w} = \mathbf{P}(K, \mathbf{x}) \cdot \mathbf{b}$ ，其元素 $w_k = b_j \cdot \alpha_k \cdot s_k, \forall k \in S_j$ ，而 $\mathbf{P}(K, \mathbf{x})$ 代表 $N_f \times N$ 的矩陣，其元素 p_{kj} 滿足：

$$p_{kj} = \begin{cases} \alpha_k \cdot s_k, & \text{if } k \in S_j \\ 0, & \text{otherwise} \end{cases}$$

由此得知，隱藏訊息的 N 個位元 \mathbf{b} 被展頻調變成 N_f 個維度的碼字 \mathbf{w} ，進而涵蓋整段的離散餘弦域係數，嵌入後的加印音訊為 $y_k = x_k + w_k$ 。

(2) 音訊轉換係數的機率分佈：

有鑒於音訊頻域機率分佈的未知性，我們先建構其離散餘弦轉換係數之廣義高斯分佈函數機率模型，其定義如下：

$$f_c(x) = \frac{\beta \cdot C}{2\Gamma(1/C)} \cdot e^{-\beta|x|^C}, \quad \beta = \frac{1}{\sigma} \left[\frac{\Gamma(3/C)}{\Gamma(1/C)} \right]^{1/2}$$

我們利用離散餘弦域係數的平均絕對值及變異數來估算 c 值如下：

$$E[|x|] = \int_{-\infty}^{\infty} |x| \cdot p(x) dx = \frac{1}{\beta} \frac{\Gamma(2/C)}{\Gamma(1/C)}$$

$$\sigma^2 = \text{Var}[x] = \int_{-\infty}^{\infty} x^2 \cdot p(x) dx = \frac{1}{\beta^2} \frac{\Gamma(3/C)}{\Gamma(1/C)}$$

則第 k 個轉換係數 x_k 所對應的機率模型參數值 c_k 計算公式為：

$$\frac{E[|x_k|]}{\sigma_k} = \frac{\Gamma(2/C_k)}{\sqrt{\Gamma(1/C_k)} \cdot \Gamma(3/C_k)} = F(C_k)$$

$$\Rightarrow C_k = F^{-1} \left\{ \frac{E[|x_k|]}{\sigma_k} \right\}, \forall k \in \{0, 1, \dots, N_f - 1\}$$

由於離散餘弦域上每個係數的分佈皆不盡然相同，據此模擬之廣義高斯分佈函數在不同位置上所求出之最佳 C 值，也存在著明顯的差異。透過以上統計分析，就是要求出每個係數所對應的最適化機率分佈函數。

(3) 人耳聽覺模型分析：

透過人耳聽覺模型的分析，計算出遮蔽臨界曲線並據以調整各轉換係數的加權比重，將有助於浮水印的透明度及強韌度間取得最佳平衡。根據傳統的聲響心理學理，我們計算出一組傅利葉轉換頻域上的遮蔽臨界曲線，其計算流程包括有臨界頻帶濾波、等響度預處理，以及主觀響度單位轉換。該曲線就是浮水印能嵌入音訊內而不被察覺所允許的最大頻域值，而尚待解決的課題是如何將其於離散餘弦轉換軸上展開。我們的解決方案如下：

$$DFT: x_{DFT}[k] = \sqrt{\frac{1}{N_f}} \cdot \sum_{n=0}^{N_f-1} q[n] \cdot e^{-j2\pi kn/N_f}$$

$$DCT: x_k = \sqrt{\frac{2}{N_f}} \cdot \gamma[k] \cdot \sum_{n=0}^{N_f-1} q[n] \cdot \cos\left(\frac{\pi \cdot k(2n+1)}{2N_f}\right)$$

$$\gamma[k] = \begin{cases} 1/\sqrt{2} & k=0 \\ 1 & k=1, 2, \dots, N_f-1 \end{cases}$$

為發展兩轉換間關係式，首先我們考慮一段訊號序列 $q[n], 0 \leq n \leq N_f - 1$ ，將其表示為一個行向量 $\mathbf{Q} = [q[0], q[1], \dots, q[N_f - 1]]^T$ ，將離散傅利葉轉換向量 $\mathbf{X}_{DFT} = [x_{DFT}[0], x_{DFT}[1], \dots, x_{DFT}[N_f - 1]]^T$ 表示為 $\mathbf{X}_{DFT} = \mathbf{A}\mathbf{Q}$ ，其中

$$\mathbf{A} = \frac{1}{\sqrt{N_f}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-j2\pi(0)/N_f} & e^{-j2\pi(1)/N_f} & \dots & e^{-j2\pi(N_f-1)/N_f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi(N_f-1)/N_f} & e^{-j2\pi(N_f-2)/N_f} & \dots & e^{-j2\pi(N_f-0)/N_f} \end{bmatrix}$$

餘弦轉換向量亦可轉換為 $\mathbf{X}_{DCT} = \mathbf{B}\mathbf{Q}$ 其中

$$B = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \dots & \frac{1}{\sqrt{2}} \\ \cos\left(\frac{\pi(1)(1)}{2N_f}\right) & \cos\left(\frac{\pi(1)(3)}{2N_f}\right) & \dots & \cos\left(\frac{\pi(1)(2N_f-1)}{2N_f}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \cos\left(\frac{\pi(N_f-1)(1)}{2N_f}\right) & \cos\left(\frac{\pi(N_f-1)(3)}{2N_f}\right) & \dots & \cos\left(\frac{\pi(N_f-1)(N_f-1)}{2N_f}\right) \end{bmatrix}$$

$X_{DCT} = [x_{DCT}[0], x_{DCT}[1], \dots, x_{DCT}[N_f-1]]^T$ ，其間互轉關係為 $X_{DCT} = BQ = BA^{-1}X_{DFT}$ ，因此我們可以把 DFT 頻域求得之遮蔽曲線，先標準化後，再將其轉換到 DCT 域，如此即可求出浮水印的加權比重 α_k ，亦即為容許浮水印嵌入的最大變動值。

(4) 浮水印檢測演算法：

浮水印檢測器的設計旨在正確判別待測試的音訊檔內是否已嵌入認證用的浮水印訊號，所以此檢測問題可視為兩事件的二元假說測試，其中事件 H_1 代表音訊經過測試含有金鑰 K 產生的浮水印， H_0 則代表音訊經過測試不包含金鑰 K 產生的浮水印。

$$H_1: z = x + w$$

$$H_0: z = x$$

我們定義誤警機率 $P_r = P\{d(z, K) = H_1 | H_0\}$ ，及確認機率 $P_D = P\{d(z, K) = H_1 | H_1\}$ ，作為量測浮水印檢測效能的指標，而最佳化檢測器的設計旨在降低誤警機率且提高確認機率。利用先前估算轉換係數的分佈函數 $f_x(\mathbf{x})$ ，再透過奈曼-皮爾生法則(Neyman-Pearson test)推導出最佳的決策公式為：

$$\Lambda(z) = \frac{P_{obs}(z|H_1)}{P_{obs}(z|H_0)} = \frac{f_z(z|H_1, K)}{f_z(z|H_0)}$$

$$= \sum_b \frac{p(b) \cdot f_z(z - Pb)}{f_z(z)} \frac{H_1}{H_0} \eta$$

其中 η 為決策臨界值。假設無訊號處理失真 $y=z$ ，進一步將音訊頻域係數的廣義高斯分佈機率代入函數 $I(y) = \ln \Lambda(y)$ 中，即可推導出浮水印檢測的快速實現演算法。若考慮 $N=1$ 且 $b_1=1$ 的特殊情況，定義 $\hat{k} = (k \bmod N_f)$ ， K 為音訊取樣點的總數目，則函數 $I(y)$ 可簡化為

$$I(y) = \sum_{k=1}^K \beta_k^{C_k} \cdot [|y_k|^{C_k} - |y_k - \alpha_k s_k|^{C_k}]$$

進一步分析發現不同兩事件的相似函數呈現高斯機率分佈，其變異數相同而平均值互為正負。更明確地說，

$$I(y | H_1) \sim G(m_1, \sigma_1^2)$$

$$I(y | H_0) \sim G(m_0, \sigma_0^2) = G(-m_1, \sigma_1^2)$$

其中 $m_1 = E\{I(y) | H_1\}$

$$= \frac{1}{2} \sum_{k=1}^K \beta_k^{C_k} \cdot (|x_k + \alpha_k|^{C_k} + |x_k - \alpha_k|^{C_k}) - \sum_{k=1}^K \beta_k^{C_k} |x_k|^{C_k}$$

$\sigma_1^2 = Var\{I(y) | H_1\}$

$$= \frac{1}{4} \sum_{k=1}^K \beta_k^{2C_k} \cdot (|x_k + \alpha_k|^{C_k} - |x_k - \alpha_k|^{C_k})^2$$

而確認機率 P_D 和誤警機率 P_r 則為：

$$P_D = Q\left(\frac{\eta - m_1}{\sigma_1}\right), \quad P_r = Q\left(\frac{\eta + m_1}{\sigma_1}\right)$$

其中 $Q(v) = \frac{1}{\sqrt{2\pi}} \int_v^\infty e^{-t^2/2} dt$ 。若我們定義 $SNR_1 = \frac{m_1^2}{\sigma_1^2}$

且設定誤警機率為 P_r ，則其對應值為

$$v = Q^{-1}(P_r) = \frac{\eta + m_1}{\sigma_1} \text{。而確認機率則表示為：}$$

$$P_D = Q\left(\frac{\eta - m_1}{\sigma_1}\right) = Q\left(\frac{\eta + m_1}{\sigma_1} - 2\frac{m_1}{\sigma_1}\right)$$

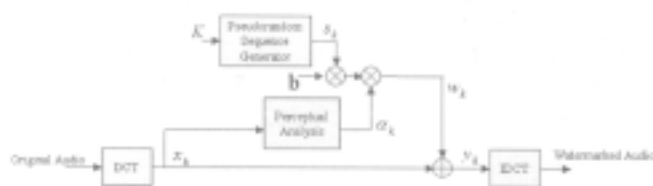
$$= Q\left(Q^{-1}(P_r) - 2\sqrt{SNR_1}\right)$$

四、結論與討論

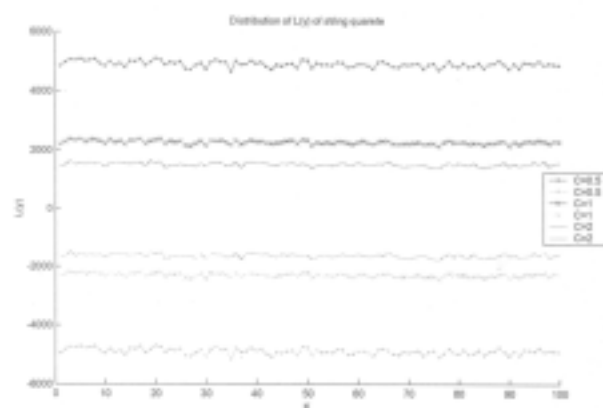
我們取一百組擬亂序列分別針對三種不同的音樂(鋼琴、長笛、弦樂四重奏)做浮水印之嵌入，再對加印音訊及原始音訊做浮水印的偵測。C 值為 0.5、1、2 的廣義高斯機率模型可各別得到一百組之 $I(y)$ 值，其 m_1 及 σ_1^2 所求得之 SNR_1 示於表一。我們的結論是訊雜比 SNR_1 之實驗值與理論值具有一致性，且弦樂四重奏在 $C=0.5$ 時檢測效能為最佳。弦樂四重奏音樂檢測所得之 $I(y)$ 值分布情形如圖三所示，加印音訊及原始音訊所得之 $I(y)$ 值分布有明顯差距，可據以準確地判斷音訊有否嵌入浮水印。表二列出 $C=0.5$ 時，檢測器對三種音樂檢測之加印及原始音訊所求得 $I(y)$ 的平均值與變異數，其實驗值與理論值大致相同。



圖一：浮水印系統



圖二：浮水印的嵌入流程



圖三：在弦樂四重奏音訊中，對一百組不同之秘密金鑰之實驗中，針對不同的C值(C=0.5、1、2)之廣義高斯分布所得到的對數相似函數 $I(y)$ 的分布。

SNR_i	C=1/2		Laplace(C=1)		Gaussian(C=2)	
	empirical	Theoretical	empirical	Theoretical	empirical	Theoretical
Audio						
Piano	19.45	21.08	20.35	22.51	19.65	20.10
Flute	35.16	38.00	34.62	37.83	33.50	25.42
String quartet	34.05	33.42	32.02	32.85	28.44	27.63

表一：浮水印檢測器訊雜比(單位 dB)之實驗值與理論值

C=1/2	Empirical				Theoretical	
	m_1	σ_1^2	m_0	σ_0^2	m_1	σ_1^2
Audio						
Piano	709.28	5706.6	-727.90	5708.1	717.95	4015.7
Flute	10397	32967	-10399	32967	10394	17130
String quartet	4903.8	9463	-4898.0	9463	4880.3	10845.0

表二：在 H_1 和 H_0 的條件下，對C=0.5的廣義高斯分布之對數相似函數 $I(y)$ 的實驗值與理論值。

參考文獻

- [1] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," Proceedings of the IEEE, Vol. 86, Jun 1998, pp. 1064-1087.
- [2] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shanon, "Secure spread spectrum watermarking for multimedia," IEEE Transactions on Image Processing, Vol. 6, Dec 1997, pp. 1673-1687.
- [3] M. D. Swanson, B. Zhu, A. H. Tewfik and L. Boney, "Robust audio watermarking using perceptual masking," Signal Processing, Vol. 66, Nov 1997, pp. 337-355.
- [4] M. Kutter, and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," IEEE Transactions on Image Processing, Vol. 11, Jan 2002, pp. 16-25.
- [5] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," IEEE Transactions on Image Processing, Vol. 9, Jan 2000, pp. 55-68.
- [6] K. A. Birney, and T. R. Fischer, "On the modeling of DCT and subband image data for compression," IEEE Transactions on Image Processing, Vol. 4, Feb 1995, pp. 186-193.