

# 基於線性預測編碼技術之頻寬延展方法應用於高品質音訊解碼

雷曉方<sup>1</sup> 吳子懿<sup>2</sup>

## 摘要

感知音訊編碼已是現今普及的數位音訊處理技術，然而在低位元率、高壓縮率的要求條件下，往往是捨棄音訊高頻部份，因而造成解碼端音訊品質的劣化。近年來，頻寬延展是音訊編碼系統發展的一種新技術，根據人耳對於低頻跟高頻音訊的敏感度不同，利用少量的資訊來重建失去的高頻頻譜，使得低位元率的音訊可以保存住全頻寬，提升傳統感知型音訊編碼器的品質。本論文，提出編碼端利用線性預測編碼演算法來記錄高頻頻譜的特徵，然後在解碼端可以還原回近似原音訊之高頻頻譜的方法，並結合傳統感知型音訊解碼器，彌補其高頻頻帶處理的缺陷，最後達成數位音訊編碼系統的新目標—低位元率和高品質音質要求。

關鍵字：音訊編碼、頻寬延展、頻帶複製、線性預測編碼

## High-Quality Audio Bandwidth Extension Method Based on LPC

### Abstract

Perceptual Audio Coding is very popular for digital audio signal processing systems. The requirements of low bit rate and high compression efficiency by this technology cause the distortion in high frequency components of audio signals and the degradation of the audio quality. Bandwidth extension is a new technology in the recent years. It designs a processing method different from traditional perceptual audio coding. According to the different sensitivity to different frequency band in human auditory system, it uses a little information to reconstruct the lost high frequency components and makes low bit rate audio streams without losing audio quality. This paper proposes an algorithm based on linear prediction coding (LPC) to record original high frequency characteristics in the encoder and reconstruct based on high frequency components in the audio decoder. The proposed algorithm can be combined with traditional perceptual audio decoder to overcome the distortion for high frequency to achieve the goals of low bit rate with high sound quality of new digital audio coding systems.

Keyword : Audio coding, Bandwidth extension, Band replication, Linear prediction coding

### 一、緒論

隨著電腦與網路科技的發展，數位化成為了高科技產業的趨勢，其所包含的多媒體應用也不例外。在多媒體應用中，龐大的音訊資料在傳輸(transmission)與儲存(storage)的過程中受到網路頻寬與記憶空間的限制。需要將原始音訊資料量加以縮減，因此音訊編碼技術應運而生。多數音訊編碼技術為了增進編碼效率，能有效壓縮資料量，大多以降低位元率為主要目標，然而在追求此目標下，難免對原始音訊的音質有所犧牲。

---

<sup>1</sup> 雷曉方，成功大學電機工程系助理教授

<sup>2</sup> 吳子懿，成功大學電機工程系研究生

在 2000 年，許多研究者想開發一種新的音頻壓縮技術並計畫放入 MPEG-4 的音訊編碼標準中。新技術主要的訴求是“高音質”與“低位元率”。2001 年，頻帶複製技術(Spectral band replication, SBR)首先在 MPEG-4 的架構下標準化，它是一種頻寬延展技術，訊號頻寬的重建主要來自於接收端的低頻部分。今天的數位廣播收音機系統(Digital Radio Mondiale System)與 XM 衛星廣播系統(XM Satellite Radio System)都採用了 SBR 的技術來編碼，且被標準化在 ISO/IEC 14496-3:2001 [1-3]。

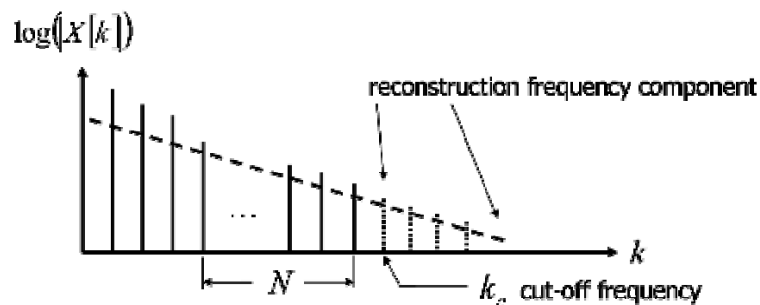
雖然 SBR 在一個新的數位音訊編碼領域中展現出成果，但仍然有些缺點，諸如系統的複雜度增加，編碼延遲、感知型編碼器與頻譜延展的匹配問題 [4]，所以持續有許多的研究者投身於頻寬延展技術的研究。本論文，提出編碼端利用線性預測編碼演算法來記錄高頻頻譜的特徵，然後在解碼端可以還原回近似原音訊之高頻頻譜的方法，並結合傳統感知型音訊解碼器，彌補其高頻頻帶處理的缺陷，最後達成數位音訊編碼系統的理想目標，能符合低位元率和高音質的要求。

## 二、頻寬延展演算法

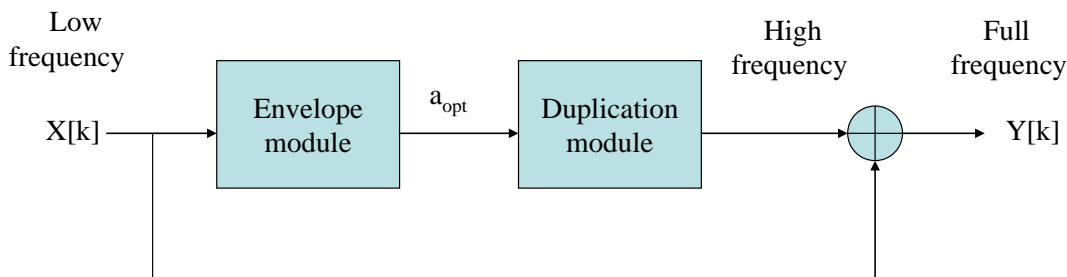
人耳聽覺系統是以非線性關鍵頻帶(critical Bands)來分解音頻信號 [5]，人耳對於不同頻率的訊號成分有不同的敏感度，在鑑別分析率上低頻比高頻較為敏感(Bark1 的頻寬是 0~100Hz 以及 Bark24 的頻寬是 12~15.5kHz，關鍵頻帶的測量尺度為 Bark，Bark1 為低頻頻帶，Bark24 為高頻頻帶)，而且，遮蔽效應(masking effect)也有相同的情況，聲音在低頻比高頻更容易超出遮蔽門檻(masking threshold)。藉由人類耳朵去辨別高音頻的原始訊號與近似訊號之間的差異是困難的，根據這些現象，低頻訊號成分對於人類聽覺系統而言比高頻來的重要多了。所以多數音訊編碼演算法皆是以編碼低頻的信號為優先，在有限的位元率限制下，捨棄掉高頻成分的編碼，因而造成音訊品質的劣化，所以如何藉助高頻重建技術，將解碼的音訊品質提升是一值得研究的課題。

### 2.1 高頻重建演算法(High-Frequency Reconstruction, HFR)

圖一所示，HFR [6] 利用線性外插(linear extrapolation)方式對低頻信號頻譜以對數座標幅度(magnitude with logarithm scale)來求得高頻部份的頻譜（試著用線性外插去推估高頻包絡線(envelope)）。



圖一、HFR 以線性外插來推估信號的高頻頻譜



圖二、HFR 系統方塊圖

圖二所示為 HFR 的系統方塊圖，此圖中的輸入信號是由傳統感知型解碼器還原信號的低頻成份，然後計算低頻的頻譜分佈，再利用線性外插與迴歸方法來推估信號的高頻頻譜成份，然後產生出相對於此高頻頻譜成份的時域信號與低頻信號合成出全頻寬的音頻信號。

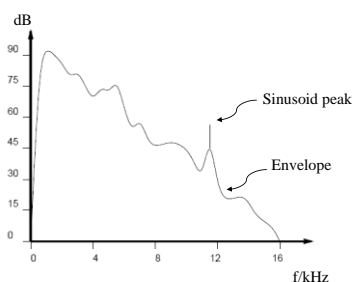
## 2.2 Plus V 演算法

Plus V 的概念是記錄原始高頻訊號的頻譜包絡線的平均能量以及有意義的正弦曲線 (significant sinusoid)。在圖三裡，我們可以看到 Plus V 所要注意的包絡線，還有正弦曲線峰值 (peak)。

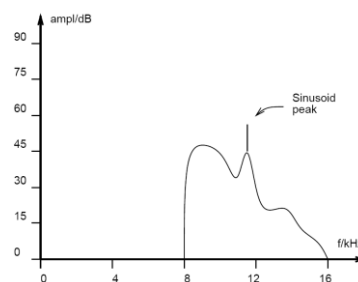
Plus V 演算法 [7] 的步驟先將音頻信號以濾波器分成低頻與高頻成份的信號，低頻的信號中，因高頻成份已被去除，所以可以降低信號取樣率 (例如：原信號取樣率為 44.1kHz，若低頻濾波的截止頻率在 8kHz 以下，故取樣率可降為 22kHz，仍可滿足取樣率處理要求)，經過降低取樣率的信號，由傳統感知型編碼器來作音訊壓縮 (如 Lame MP3 編碼)，即可達到低位元率的效果。高頻成份的信號如圖四所示，因為耳朵的鑑別率不高，所以只要將高頻成份的信號以白噪音與正弦單音信號合成近似，根據此概念 Plus V 將信號的高頻成份分為 8 個頻帶，然後計算每個頻帶的能量分佈如圖五所示，若某頻帶有特別突顯的正弦單音信號，即將此正弦信號的頻率與對應信號峰值記錄下來，如圖六所示，在 Plus V 演算法所記錄的高頻頻譜資訊，可以在解碼端將高頻成分的信號合成出來。

與 HFR 演算法的比較，HFR 在編碼端不需要記錄高頻信號的資料，所以 HFR 是完全在解碼端以後處理器(post-processor)達成， Plus V 需要紀錄的原始高頻訊號成分的粗略資訊(不同頻帶的噪音能量值與正弦單音信號的頻率與峰值)。所以 Plus V 在編碼端是前處理器(pre-processor)而解碼端則為後處理器來實現。

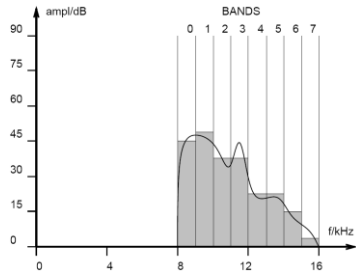
圖七 (a) 和圖七 (b) 顯示了 Plus V 的編碼與解碼器架構。原始的音頻訊號輸入 Plus V 的編碼器，首先會算出正弦曲線的部分，接下來在移除掉正弦曲線後計算包絡線的能量  $e_i$ 。在解碼器中的操作則相反，先用包絡線的能量去重建粗略的包絡線，再加入正弦訊號，最後則將重建的高頻訊號成分與使用感知型解碼器的低頻訊號成分結合，產生全頻帶的音頻訊號。



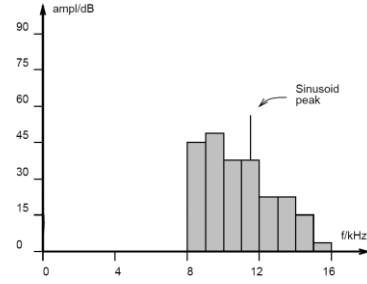
圖三、原始音頻訊號



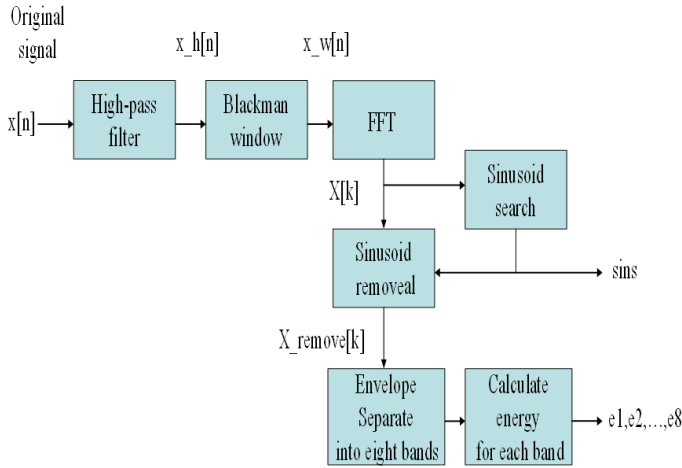
圖四、含有正弦單音成分的高頻頻譜信號



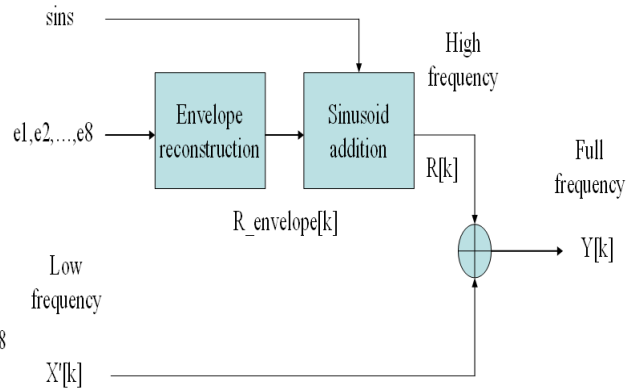
圖五、Plus V 演算法中移除正弦單音成分與計算高頻頻帶能量分布



圖六、Plus V 演算法還原後的高頻頻譜



圖七 (a)、Plus V 編碼器的系統方塊圖



圖七 (b)、Plus V 解碼器的系統方塊圖

### 2.3 線性預測編碼(proposed LPC)演算法

線性預測編碼技術可從信號的序列中抽取少量的特徵值，這些值可用以描述信號的頻域特性，所以 LPC 作為頻譜預估已有許多應用 [8]，從先前討論中，高頻重建需要對高頻頻譜做粗略近似，並且不需要太過複雜的計算。LPC 使用一組有限係數能預估出訊號的包絡線形狀，此外 LPC 也不需要很多的位元來儲存係數，而仍保有可辨識音頻的品質，這也適合用少量的紀錄資訊來幫助感知型編碼達到低位元率而高頻音質有所提昇 [9]。

若在時域中來處理信號，線性預測的觀念為利用先前的 N 個取樣點來估測現在的取樣點，其現在取樣點可表示成先前的取樣點的線性組合。

$$\hat{s}[n] = -\sum_{i=1}^N a_i s[n-i] \quad (1)$$

$$e[n] = s[n] - \hat{s}[n] = s[n] + \sum_{i=1}^N a_i s[n-i]$$

$\hat{S}[n]$  為預估現在取樣點， $e[n]$  為真實取樣點與預測值的差， $a_i$  為預估係數。換言之，線性預測是建立一個預估模型 (prediction model)。藉由其中的預估係數  $a_i$ ，可推測出信號的波形變化。為了求得較準確的預估係數，可利用最小的均方值預測誤差 (找出一組線性預估係數  $a_i$ ，使得  $e[n]$  的均方值為最小)，根據此觀念，利用數學中統計隨機程序的推導可得出。

$$R_s a = -r_s \quad (2)$$

$$\text{where } R_s = \begin{pmatrix} R_s[0] & R_s[1] & \cdots & R_s[N-1] \\ R_s[1] & R_s[0] & \cdots & R_s[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_s[N-1] & R_s[N-2] & \cdots & R_s[0] \end{pmatrix}$$

$$a = [a_1 \ a_2 \ \cdots \ a_N]^T$$

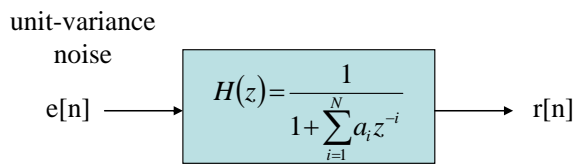
$$r_s = [R_s[1] \ R_s[2] \ \cdots \ R_s[N]]^T$$

$R_s[i]$  是  $s[i]$  的自相關函數矩陣。

$$R_s[i] = E\{s(n)s(n-i)\}$$

$E\{\cdot\}$  為平均值計算

我們採用 Levinson-Durbin 遞迴演算法 [10] 去求解方程式(2)。在計算  $a_i$  的同時，可得  $e[n]$  的均方值與線性預估模型的增益值有關，在解碼器中可以 IIR 濾波方式合成還原高頻端的信號。

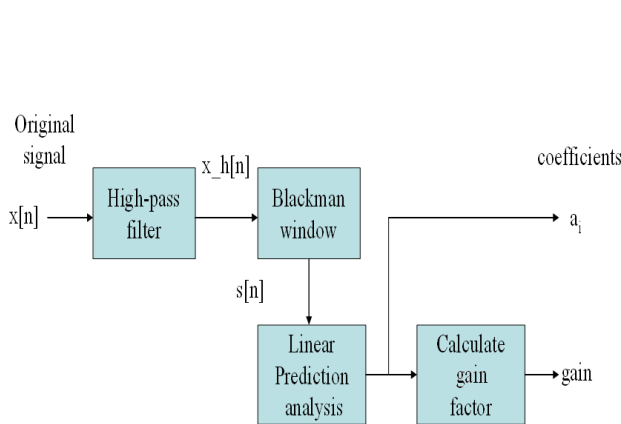


圖八、IIR 濾波器還原高頻信號

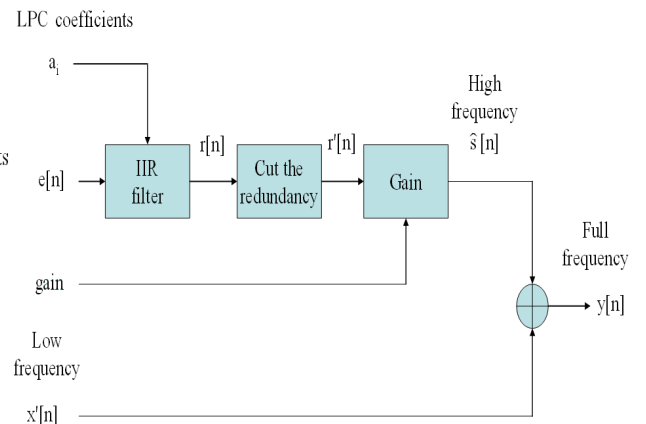
$$H(z) = \frac{R(z)}{E(z)} = \frac{1}{1 + \sum_{i=1}^N a_i z^{-i}} \quad (3)$$

$$r[n] = e[n] - a_1 r[n-1] - a_2 r[n-2] - \cdots - a_N r[n-N]$$

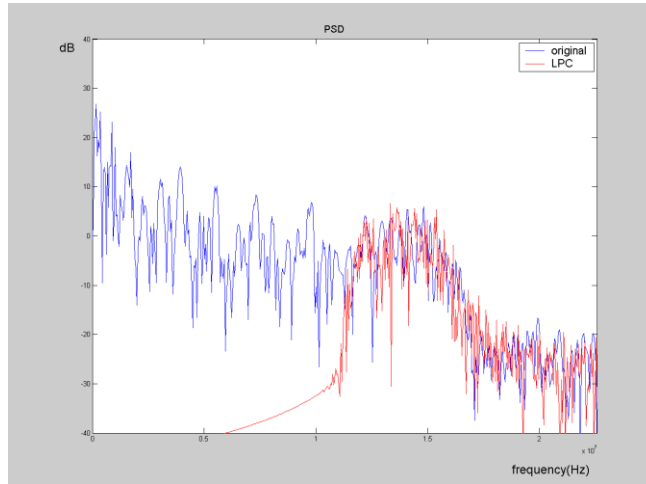
圖九所示為以線性預測編碼技術實現頻寬延展的編解碼器方塊圖架構。圖九 (a) 原始信號經過一高頻濾波器，分離出高頻信號，加以 windowing 運算，然後將以線性預測分析計算線性預估模型的增益值，然後將線性預估係數及增益值萃取出來作傳輸，達到以少量資訊為原來代表信號中高頻的資訊，圖九 (b) 所示為解碼端，利用 IIR filtering 方式，以單位變異量為 1 的白噪音當成激發源，此濾波器的輸出乘以增益值，還原高頻的信號，再與傳統解碼器中解回來的低頻信號相加，即可恢復全頻寬的音頻信號。



圖九 (a)、proposed LPC 編碼器



圖九 (b)、proposed LPC 解碼器



圖十、原始訊號與 proposed LPC 還原回來高頻信號的頻譜

圖十所示，為原始信號與利用線性預測推估演算法所還原的高頻信號的頻譜圖，藍線全頻寬的頻譜為原始信號，紅線僅有高頻部份的頻譜是以 proposed LPC 演算法所還原的高頻成份信號，由頻譜的比較可知，線性預測演算法所還原的頻譜與原始信號頻譜相當近似。

### 三、實驗結果與分析

本論文利用一些客觀品質量測數據來比較頻寬延展方法的性能結果。實驗所採用 30 種不同的音訊樣本作測試，這些樣本包括：古典音樂、流行音樂、樂器演奏、聲樂演唱等。

首先我們以音譜圖和計算頻譜距離的參數來比較不同的頻寬延展演算法，比較的演算法包括：HFR 演算法、Plus V 演算法及我們所提出以線性預測 proposed LPC 的演算法。

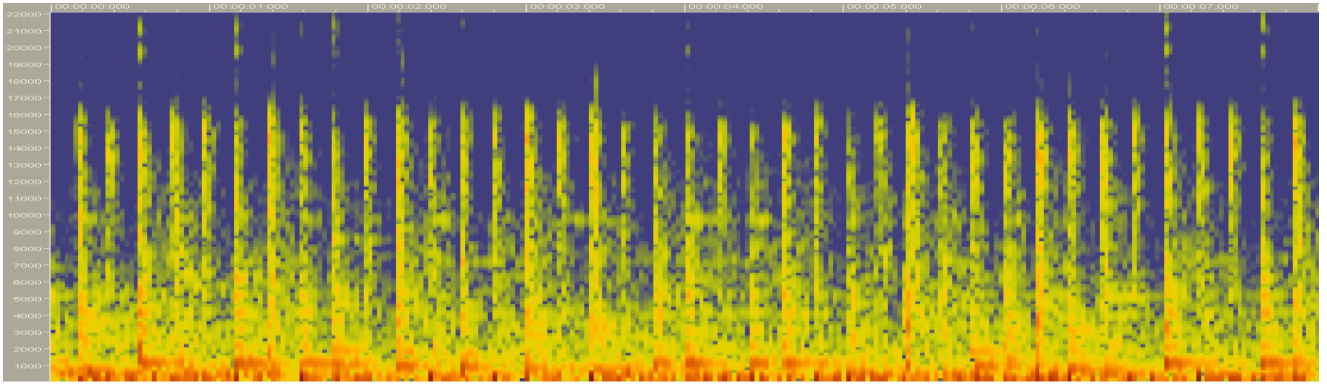
我們使用 Nero 波型編輯器來顯示各種頻寬延展的音譜圖。我們使用的感知型音訊編碼是 MP3 Lame v3.96.1，而位元率是 64kbps。這實驗的樣本是三個十秒長的流行歌曲。我們用 COSH 長度 [11] 來計算頻譜長度，如下方程式所示：

$$COSH = \int_{-\pi}^{\pi} \left[ \frac{1}{2} \left( \frac{Y(w)}{X(w)} + \frac{X(w)}{Y(w)} \right) - 1 \right] \frac{dw}{2\pi} \quad (4)$$

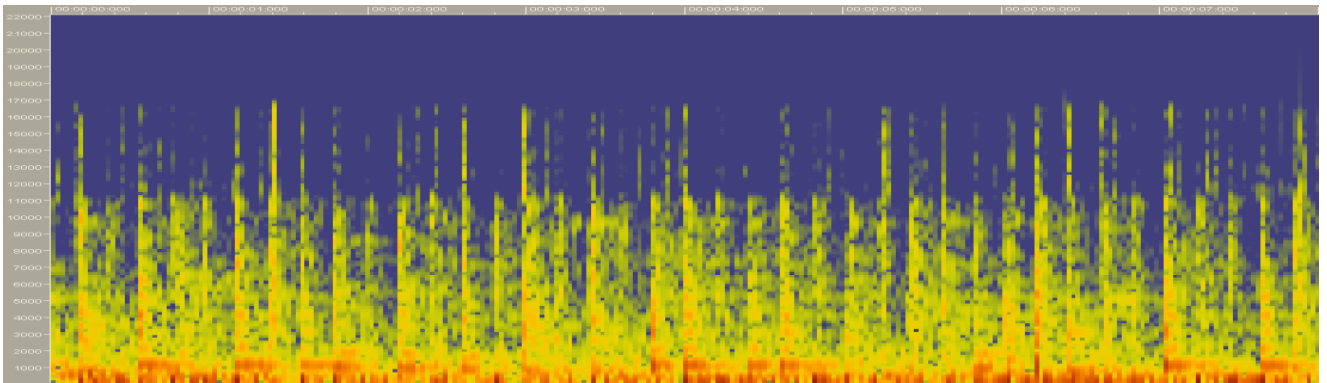
原始訊號以  $x[n]$  表示， $y[n]$  為失真訊號。COSH 長度(COSH distant)是單位長度失真(distant distortion)的對稱變化(symmetric version)。因此，如果 COSH 長度較小，頻譜變化則較少。COSH 長度愈大時頻譜失真將會愈差。

圖十一所示為音譜圖的結果，(a) 為原始音訊的音譜圖 (b) (c) (d) 分別為 HFR、Plus V 以及 proposed LPC 的音譜圖，與 (a) 比較，顯然 (b) 所示 HFR 方法的高頻部份與原始訊號有較多的差異，失真程度比較大。Proposed LPC 較接近原始信號的頻譜，顯示 proposed LPC 高頻還原的效果較佳，表一顯示為頻譜距離參數的量測。HFR 方法的結果有較大的 COSH 值，proposed LPC 方法的 COSH 平均值最小，這與音譜圖的結果比較相呼應，顯示 proposed LPC 法有較好的高頻重建頻譜，所以 proposed LPC 有較好高頻還原的效果。

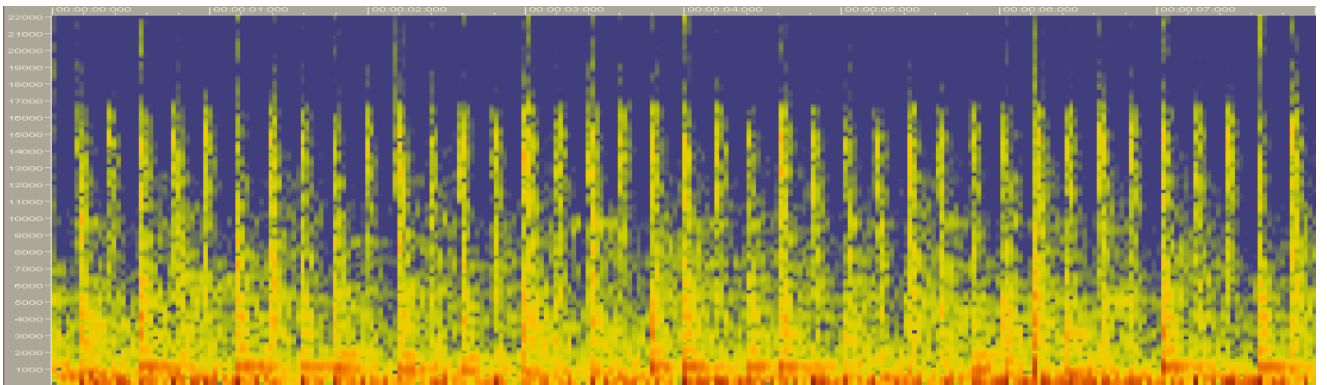




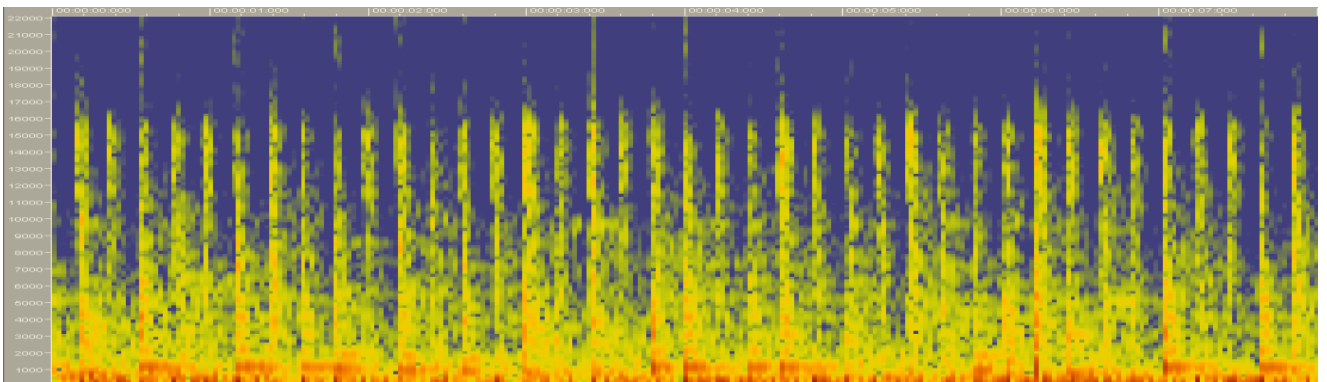
圖十一 (a)、原始信號的音譜圖



圖十一 (b)、HFR 演算法的音譜圖(COSH=117.4979)



圖十一 (c)、Plus V 演算法的音譜圖(COSH=29.8137)



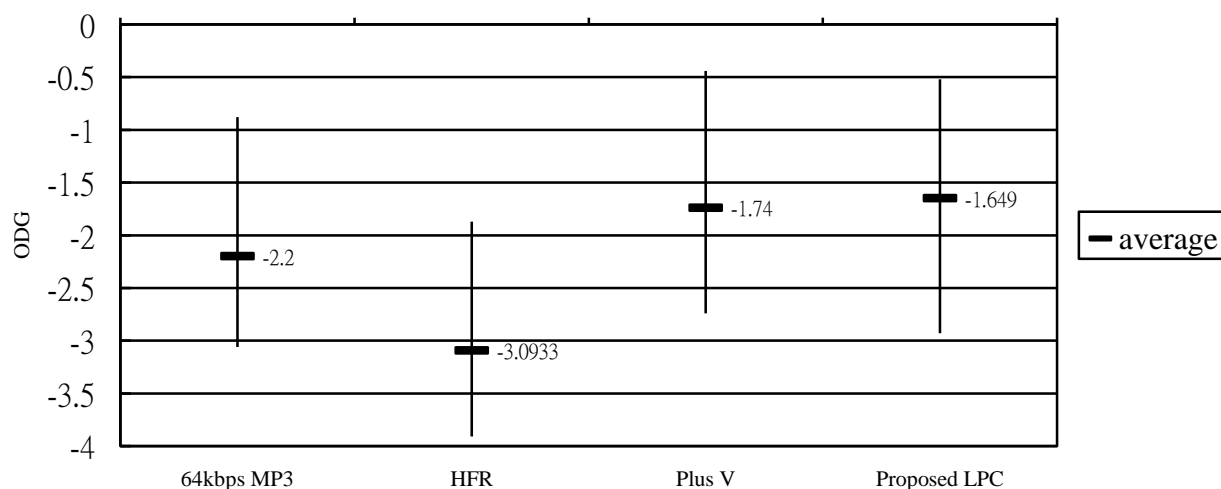
圖十一 (d)、proposed LPC 演算法的音譜圖(COSH=18.7476)

表一、頻譜展延演算法的 COSH 長度值

Sample	HFR	Plus V	Proposed LPC
1	117.4979	29.8137	18.7476
2	129.4460	37.0981	25.5393
3	203.6958	19.9227	13.9725
平均	150.2132	28.9448	19.4198

音訊品質感知評估 [12] 乃是一套演算法來客觀量化音訊的品質，所根據的觀念是將一參考信號樣本與經過編碼器輸出信號樣本平行輸入一感知聽覺模型來處理。此感知模型系統的輸出乃模擬音訊經過人類聽覺系統所有的內部代表形態。由此輸出端的內部代表形態產生可聽性差值(audible difference)，然後再由大腦認知模型(模擬大腦聽覺處理，感知聲音信號的過程)來計算出一量化的數值稱作客觀差評分(Objective Difference Grade, ODG)，其數值範圍為 0 至-4，0 代表經過編碼的信號與原參考信號幾乎完全一樣，聽不出來有任何失真的感覺，數值愈負，代表失真愈容易被聽出來，若 ODG 值到達-4，表示此編碼器所編碼的音訊失真程度已經非常惱人，品質到達不能接受的程度。

圖十二所示為比較沒有使用及使用各種頻寬延展的 64kbps MP3 PEAQ 的結果，實驗的樣本為三十首流行歌曲，每個樣本都是 10 秒長，加以平均後的 ODG 值，比較此結果，顯示本論文所提的線性預測演算法的 ODG 值優於其他兩種演算法。



圖十二、頻寬展延演算法的 ODG 值

表二、頻寬展延演算法所需增加的位元率

	HFR	Plus V	Proposed
附加位元率 (kbps)	0	6.31	4.47

表二所示為頻寬展延演算法所需增加的位元率，我們所提 LPC 的方法若以線性預測係數為 12 階為例，由表中結果可看出，本論文所引用線性預測的演算法，所需增加的位元率比 Plus V 的演算法要低。



#### 四、結論

本論文提出利用線性預測編碼演算法的頻寬延展技術來還原高頻的信號，與其他寬頻展延技術比較(HFR、Plus V 的演算法)，能結合傳統感知型音訊編解碼系統，在低位元率與音訊品質上有較佳的表現。

#### 五、參考文獻

1. P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication," in IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, Nov. , 2002.
2. M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th AES Convention, Munich, May 2002.
3. ISO/IEC JTC1/SC29/WG11/N5203 MPEG, "Bandwidth extensions," Audio, Oct. 2002.
4. Anibal J. S. Ferreira and Deepen Sinha, "Accurate Spectral Replacement," AES 118th Convention, Barcelona, Spain, May, 2005.
5. Zwicker, E. and Fastl, H., "Psychoacoustics : Facts and Models," Springer-Verlag, second ed., 1999.
6. Chi-Min Liu, Wen-Chieh Lee, and Han-Wen Hsu, "High Frequency Reconstruction for Band-limited Audio Signals," Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September, 2003.
7. Tapani Ritoniemi, "Plus V specification," <http://www.vlsi.fi/plusv/plusv.shtml>.
8. Florian Keiler, Daniel Arfib, and Udo Z'olzer, "Efficient Linear Prediction for Digital Audio Effects," Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December, 2000
9. Arttu Laaksonen, "Bandwidth Extension in High-Quality Audio Coding", Master Thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, May, 2005
10. L.R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.
11. Hamza Ozer, Ismail Avcibas, Bulent Sankur, Nasir Memon, "Steganalysis of Audio Based on Audio Quality Metrics," Proceedings of SPIE-IS&T Electronic Imaging, SPIE vol. 5020,pp.55-66, 2003.
12. ITU-R Recommendation BS. 1387, Method for Objective Measurement of Perceived Audio Quality 1998.