行政院國家科學委員會專題研究計畫成果報告

高壓縮比語音編碼技術之研究

Research of Low Bit-Rate Speech Coding Technology

計畫編號: NSC 87-2213-E-002-019
執行期限: 86年8月1日至87年7月31日
主持人: 闕志達　　國立臺灣大學電機學院電機系
電子信箱: chiueh@cc.ee.ntu.edu.tw

一、中文摘要

在本計劃中，針對音訊訊號的遮罩效應 (masking effect) **進行研究, 提出一個新的後遮罩** (forward masking effect) **的模型，應用在音訊壓縮音質的改善上。這個模型利用了人耳音訊系統中接收感應器與神經刺激的效應，這些效應通常在心理聲音學中後遮罩的原因。其中，人耳中的非線性效應我們以一個非線性電路的差分方程式來建立模型。我們將這個模型加入** MPEG Layer III **音訊壓縮架構當中的遮罩效應，建立在時間頻率空間中的遮罩曲面。加入這個模型我們可以在相同壓縮比下得到比較好的音訊音質。在我們的實驗中，主觀與客觀的音質測試顯示我們可以比** MPEG Layer III **的音訊壓縮減少** 12%**到** 25%**所需的位元數。**

關鍵詞: **音訊壓縮、遮罩效應、音質量測**

二、英文摘要

This paper presents a new forward masking model for perceptual audio coding. This model exploits adaptation of the peripheral sensory and neural elements in the auditory system, which is often deemed as the cause of forward masking. Nonlinearity of the ear is modeled by a nonlinear analog circuit with difference equations. We incorporate this model in the MPEG Layer III audio coding scheme and construct a masking plane in the frequency-time space. With some extra computations, the new audio coding scheme can improve the sound quality of the decoded audio signals. In our experiments, subjective and objective sound quality measurements show that, to achieve the same reconstructed sound quality, the new scheme requires 12% to 23% less bits than the original MPEG Layer III scheme.

Keywords: audiocoding, masking effect, sound quality measure.

三、計劃緣由與目的

音訊訊號壓縮在許多系統中都有廣泛的應用，包括多媒體訊號壓縮、高品質音訊壓縮和高品質音訊儲存等。由於頻寬的限制，想要在固定的位元速率 (bit-rate)**來壓縮，能達到最小感官損失為重要的一個目標。在感官式音訊壓縮演算法中，將量化誤差值保持在恰可觀測扭曲（**JND just notice distortion**）之下，可大幅減低所需的位元數。在一般的感官式壓縮演算法中，只使用了頻率象限的遮罩（**frequency masking**），少數演算法有提出時間象限的遮罩（**temporal masking**），但缺乏模型的建立以預測遮罩值。**

在後遮罩的模型中常以電阻電容電路來建立人耳音訊系統中神經元素

（neural element）的激發現象。這是因為電阻電容電路是時間相關的系統，對於時間相關的音聲現象可以有效的建立模型。因此我們以一個電路來做後遮罩模型的建立，用以預測後遮罩的值。合併預測的後遮罩的值與頻率軸上的遮罩值可以建立更精確的遮罩預測。在有限的位元數下，可以得到最佳化的位元分配（bit allocation）因此可以進一步改善音訊的品質。

四、研究方法與成果

4.1 後遮罩模型的建立

後遮罩效應與前面發生的訊號有很大的相關性，先前訊號的能量會有部份殘留在 basilar membrane 中，因此造成了後遮罩效應。我們因此提出一個音訊壓縮的架構，用來計算位元分配的遮罩量，不但與目前是窗框內的訊號能量分佈有關，而且和前面一個是窗框內的訊號能量有關，如圖(一)所

$$M(t,f)=\max\{M_f(t,f),M(t-\Delta t,f)\exp^{\frac{-\Delta t}{z(f)\cdot N}}\}$$

示，我們可以用以下式子來計算出來:

其中 $M(t,f)$ 為用來計算位元分配的遮罩值，$M_f(t,f)$ 為使用頻率遮罩效應所計算出來的遮罩值，Δt 為視窗框間的時間差距，$z(f)$ 為在不同決定頻帶(critical band)內最大的時間衰減常數，N 為在心理聲音學中的總和音量(total loudness)。

總和音量 N 與後遮罩的效應大小有很大的關係，N 被限定在 1 與 0 之間，當 N 等於 1 的時候，basilar membrane 內的能量達到飽和，每一個決定頻帶有最大的時間衰減常數。當 N 等於 0 時，前面一個視窗框內的訊號沒有留下足夠產生後遮罩的能量，因此沒有後遮罩的效應。我們取頻率遮罩與後遮罩的最大值為計算位元分配的依據。因此我們可以從時間頻率空間中決定遮罩的效果為是頻率遮罩效果主導或是後遮罩效果主導。

4.2 後遮罩效應的電路模型

為了模擬人耳模型中非線性的效應，我們使用圖（二）中的電路來預估後遮罩效應的程度。每一個決定頻帶由一個圖（二）的電路來模擬，輸入 Ns 為每一個頻帶上的特定音量(Specific Loudness)，輸出為經由後遮罩效應處理過後的音量值，總和音量 N 即為 No 的總和。

在心理聲音學中,特定音量可以從 basilar membrane 內的激發能量(Excitation Energy)來算出，如下列式子可得:

$$N_s=0.08(E_T/E_0)^{0.23}[(0.5+0.5*E/E_T)^{0.23}-1]$$

其中 E 為激發能量,從分佈函數(spreading function)與頻譜能量做迴旋積分所獲得。這個式子將感官外的物理能量值轉換成腦內的感官聲音大小的量測值。

為了將類比電路模型應用在數位音訊壓縮演算法中，我們將此一電阻電容電路的微分方程式以一差分方程式來近似如下:

Ron 為二極體的通路電阻

若 $I_n<0.0$

$$I_n=0.0$$

否則

$$I_n=\frac{N_s-N_o^*}{R_{on}}$$

$$N_o^*>V^*$$

$$N_o=\frac{I_n+C2\cdot V^*/(\Delta t+C2\cdot R2)+C1\cdot N_o^*/\Delta t}{1/R1+C2/(\Delta t+C2\cdot R2)+C1/\Delta t}$$

若

否則

$$V=\frac{N_o^*\cdot\Delta t+V^*\cdot C2\cdot R2}{\Delta t+C2\cdot R2}$$

$$V = N_o = \frac{I_n + (C1 + C2) \cdot V^* / \Delta t}{1/R1 + (C1 + C2)/\Delta t}$$

其中 V^* 與 N_o^* 分別為前一個視窗框差距時間 Δt 內對應電壓值。在這個電阻電容電路中，遮罩訊號越長，透過二極體與電阻儲存在電容內的電荷也就越多，因此以較長時間脈波輸入充電電荷也越多，輸出的特定電壓值 N_o 也就越大，此時時間衰減常數也就越大。相反地，以較短時間脈波輸入充電電荷越少，輸出的特定電壓值 N_o 就會越小，此時時間衰減常數也就越小。 二極體在這個電路中的功用是用來模擬在 basilar membrane 內能量飽和的情況，二極體會限制輸入的電荷，達到限制能量輸入的目的，此時時間衰減常數為最大值。電路中的電阻電容值必須經由特殊設計以達到在量測現象中在 200ms 以上後遮罩飽和的情形。

4.3 計算量的評估

表(一)列出在心理音訊模型計算中的算數複雜度。前八項為原有頻率遮罩效應的運算，後三項後遮罩效應模型加入後的運算。為由表中可知，加入的後遮罩模型計算量遠小於頻率遮罩所需要的計算量， 值得注意的是這些並不包含修改式數位餘弦轉換(MDCT)、位元分配演算法(Bit allocation)、量化器(Scalar Quantization)以及赫夫曼編碼(Huffman coding)。如果這些都考慮進去的話，所需額外計算量的比例會更小。

4.4 主觀式與客觀式的音質評量

我們收集了十個 CD 品質的單音道音樂， 以取樣頻率 44.1 千赫十六位元的數位音訊訊號為壓縮的原音。這十個樂音的種類詳列在表(二)中。這十個樂音分別以 MPEG Layer III 壓縮標準來做壓縮解壓縮，一組沒有加入後遮罩模型，另一組有加入後遮罩模型，壓縮時的位元速率分別使用 MPEG Layer III 中所定義的 80、64、56、48、40 和 32kbps 來做壓縮解壓縮, 依照解壓縮過後的聲音訊號來做聲音品質的評量。

在主觀式聲音音質評量方面，我們採用最常用的平均意見分數(Mean Opinion Score)的方式，以十個人來做主觀聽音評分。在聽音時，所有的解壓縮樂音是在安靜室內透過耳機隨機播放，每聽一段音樂即給一個從 5 到 1 的聲音品質評量分數。

在客觀式聲音音質評量方面，我們採用感官式聲音品質量測(Perceptual Audio Quality Measure PAQM)，這種方式將壓縮架構的輸入及輸出從物理量值轉換成腦內感官量值來做比較，這種轉換的方式可以將聲音品質的好壞用一個量化的量值來做評量分數，並且能夠預測主觀式聲音音質量測的結果，有節省主觀式量測成本的好處。此外 PAQM 可以對於每一個短時間聲音訊號作量測，因此聲音品質隨著時間的變化也可以精確地預測出來。

4.6 聲音品質結果

圖（三）為一個編號一的樂音，以 48kbps 的位元速率壓縮的 PAQM 量測結果，實線沒有加後遮罩效應模型，虛線則有。在 PAQM 的量測值中是以噪音干擾值（noise disturbance）為分數，值越高代表聲音品質越差，值越低代表聲音品質越佳。圖（三）顯示的結果在加入後遮罩模型汁後的結果有比較好的聲音品質。圖（四）欲說明後遮罩效應的效果，圖（四）(a)為原音，圖（四）(b)為經由加入後遮罩效應模型之後，遮罩效應的兩種遮罩的分佈。深色的區域為頻率遮罩效應主導的區域，白色區域為後遮罩效應主導的區域，從這個分佈圖可以知道後遮罩效應在樂音變化的過程中佔了

很重要的比例，因此加入此效應可改善聲音的品質。

圖（五）為主觀式音質評量的平均結果，從中可知隨著位元速率的降低，有加入後遮罩效應模型的壓縮架構，聲音品質的遞減不至於像沒加入此模型時那麼快。圖（六）為 PAQM 量測的平均結果，隨著位元速率的降低，沒加入後遮罩效應的壓縮架構其噪音干擾值很快就升高至-0.4，加入後遮罩效應模型的壓縮架構仍可維持在-1.2。

從圖表統計的結果我們可以知道加入後遮罩效應對於固定位元速率下的壓縮，可以獲得較佳的聲音品質。

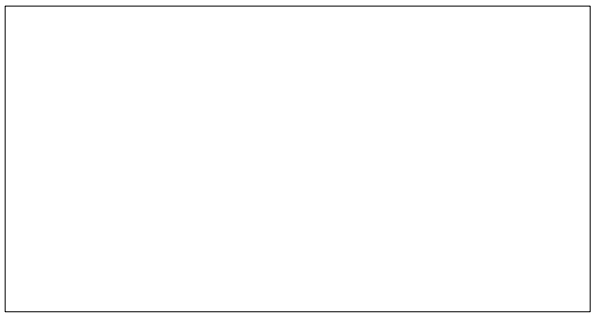
五、結論與討論

本計畫對於感官式音訊壓縮加入了後遮罩的效應，由於遮罩效應的模型建立更加完備，位元分配的結果更趨於最佳化，因此在固定位元速率下的壓縮架構下，可以進一步的提昇原有標準壓縮方式的聲音品質。

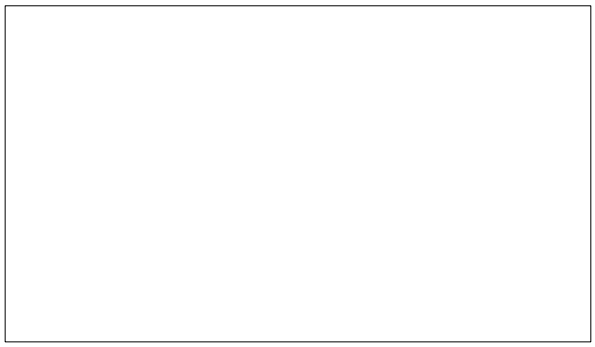
六、參考文獻

1. E. Zwicker and H. Fastl, "Psychoacoustics – Facts and Models," Springer-Verlag, 1990.
2. K. Brandenburg, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," J. of Audio Eng. Soc., vol. 42, no. 10, Oct. 1994, pp.780-792.
3. E. Zwicker "Dependence of post-masking on masker duration and its relation to temporal effects in loudness" J. Acoust. Soc. Of Amer. Vol.71(4), Apr. 1982, pp. 950-962.
4. J. G. Beerends and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on Psychoacoustic Sound Representation", J. of Audio Eng. Soc. , vol. 40, no. 12, Dec. 1992, pp.963-978.

七、圖表



圖（一）加入後遮罩效應的壓縮架構



圖（二）後遮罩效應電路模型

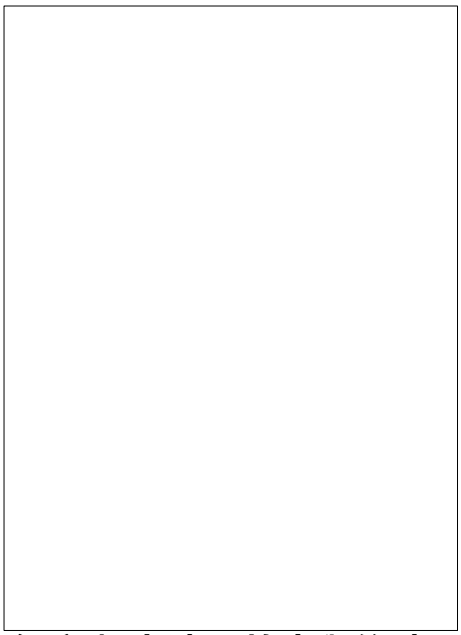
功能	複雜度	運算	N
FFT	Nlog(N)	* + cos sin	1024
Unpredict	N	* + sqrt cos sin	512
Grouping	N	* +	512
Spreading	N^2	* +	63
Tonality	N	* + log comp.	63
Pre-echo	N	Comp.	63
Entropy	N	* + log comp.	63
Threshold	N	* + log exp	63
Loudness	N	* + / power	63
Diff. Equ.	N	* + /	63
masking	N	Comp. + exp	63

表（一）計算複雜度分析

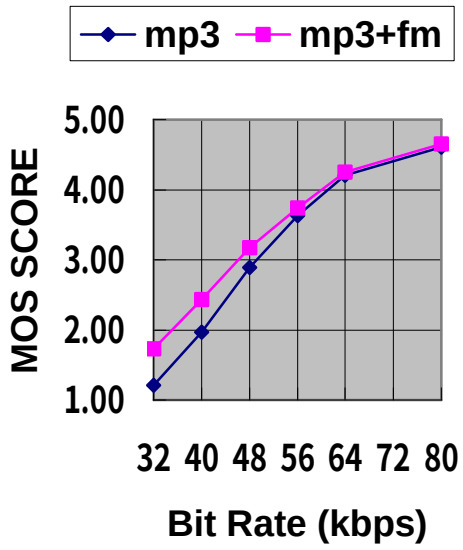
No.	樂音總類
1	薩克斯風
2	男聲歌聲
3	女生歌聲、鼓、大提琴
4	電吉他、小提琴
5	小提琴、鋼琴
6	交響樂
7	鋼琴獨奏
8	歌聲和聲
9	排笛
10	低音吉他

表（二）壓縮樂音的種類表

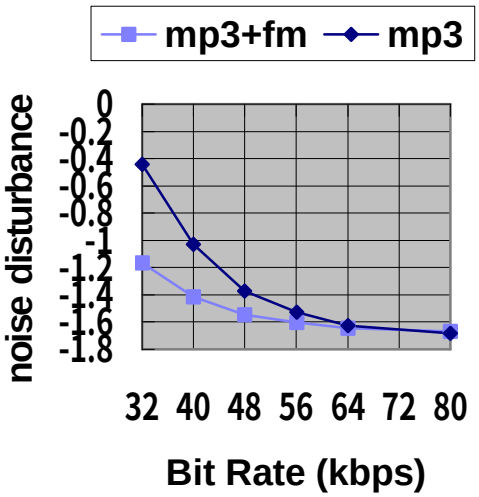
圖（三）噪音干擾對時間變化圖



圖（四）頻率遮罩效應與後遮罩效應分佈圖



圖（五）MOS vs. 位元速率



圖（六）噪音干擾值 vs.位元速率