

AMR-WB+之 ACELP 碼簿快速搜尋之研究

王木良*

樹德科技大學資訊工程系 副教授
e-mail: mulwang@mail.stu.edu.tw*

李孟達

高苑科技大學電子工程系 講師
e-mail: mtlee@cc.kyu.edu.tw

摘要

3GPP 針對低位元率多媒體服務制定“延伸型寬頻調適性多位元率”(AMR-WB+)作為音訊編碼器之標準, AMR-WB+對混合音訊內容提供良好之音訊壓縮品質。AMR-WB+是基於複合式“代數型碼激式線性預估/轉換編碼激發訊號”(ACELP/TCX)模型之編解碼器,對不同音訊內容採用不同之編碼核心以提高編碼效能。本論文針對 AMR-WB+音訊編解碼器之 ACELP 編碼核心,研究快速演算法以降低固定碼簿搜尋之計算複雜度。透過脈衝位置之預測機制可以降低碼簿搜尋之次數,經由實驗證實,本文所提出之演算法結合 3GPP AMR-WB+編碼器之深先搜尋法可進一步降低 ACELP 碼簿搜尋之計算複雜度約 50%,而聽覺上之品質傷害幾乎無法察覺。¹

關鍵詞: 延伸型寬頻調適性多位元率、複合式編碼、代數型碼激式線性預估、深先搜尋法、固定碼簿搜尋。

Abstract

The 3GPP “Extended Adaptive Multi-rate Wideband (AMR-WB+)” has been selected as recommended audio codecs for low bitrate multimedia service. AMR-WB+ showed good performance for speech and mixed content. AMR-WB+ codec is based on a hybrid ACELP/TCX model consisting of time and transform domains coding selection. For different content, the better coding core is selected to improve the coding efficiency. In this paper, the fast codebook search scheme of ACELP core of AMR-WB+ has been investigated. A fast codebook search scheme has been proposed to reduce the computational complexity of fixed codebook search. The number of search loop has been reduced by prediction mechanism of candidate pulse position. The experimental results demonstrate that the proposed scheme incorporated with the Deep First Tree Search (DFTS) scheme of 3GPP AMR-WB+ ACELP core enables a reduction of over 50 percent of the

computational load of ACELP codebook search without suffering any subjective quality degradations.

Keywords: Extended adaptive Multi-rate Wideband (AMR-WB+), Algebraic Code Excited Linear Predict (ACELP), Depth-First Tree Search (DFTS), Fixed Codebook Search.

1. 前言

近年來由於無線網路及行動裝置技術之進步,透過可攜式裝置使用者可以隨時隨地取得創新的多媒體服務。新一代的 3G 手機提供更豐富多元的應用功能,各種行動多媒體服務之應用,例如音訊及影音內容之串流、下載及上傳逐漸普及。3GPP 訂定了第三代行動通訊網路新的多媒體服務內容,而與這些服務相關之音訊內容有:音樂、語音及語音混合著音樂之音訊型式 (speech between music, speech over music),例如運動廣播、新聞及電影。如何在低位元率之下,對各種型態之音訊作有效率之編碼,對上述服務之成功與否影響甚大。在 2005 年,3GPP 選定“延伸型寬頻調適性多位元率”(Extended Adaptive Multi-rate Wideband, AMR-WB+)作為多媒體服務之推薦音訊編解碼標準[1-3,7], AMR-WB+在低位元率對語音及混合音訊型式具備良好之編碼效能。

AMR-WB+是基於複合式“代數型碼激式線性預估/轉換編碼激發訊號”(ACELP/TCX)模型之編解碼器[6],具備時域及頻域編碼模式選擇。AMR-WB+與現行固定位元率語音編碼器之差異在於其搜尋脈衝數目較多;此外,固定碼簿搜尋每一“次音框”(subframe)執行一次,由於寬頻語音編碼器之取樣頻率較高,所以單位時間內要執行碼簿搜尋之次數亦隨之增加,因而造成其 ACELP 固定碼簿搜尋的計算量很高。由文獻得知,AMR-WB+編碼在 ACELP 之計算複雜度比 TCX 要高,附圖一所示為 AMR-WB+編碼器之計算複雜度分析之實驗結果,若能再進一步降低 ACELP 之計算量,將可有效降低編碼器之計算複雜度。因此本論文針對 AMR-WB+之 ACELP 編碼核心研究碼簿快速搜尋演算法,並評估快速搜尋法在各種

¹本研究由國科會贊助,計畫編號 NSC 96-2221-E-244-012-

不同型式之音訊輸入信號情況，對輸出信號品質之影響。

2. AMR-WB+編碼演算法

AMR-WB+編碼器支援多種不同的輸入信號取樣頻率，為簡化編碼器之設計，輸入信號先轉換成 12.8kHz 的內部取樣率 F_s 。每 2048 個取樣點之輸入訊號，分割成低頻帶(LF)及高頻帶(HF)，每一頻帶各有 1024 取樣點，稱之為“超音框”(superframe)，每一超音框可再分割成 4 個長度為 256 個取樣點之音框(frame)。附圖二所示為 AMR-WB+之編碼器之架構圖，對信號之低頻帶(LF)部份，利用 ACELP 及 TCX 編碼核心作編碼，然後依合成信號之訊雜比(Signal to Noise Ratio, SNR)選擇最合適之編碼核心作為 LF 信號之編碼，由於語音與音樂信號特性之差異，若是輸入音訊為語音類型之信號，編碼器傾向於採用 ACELP 技術；反之，對非語音類型之信號則傾向採用 TCX 技術；對高頻帶(HF)部份之信號則採用頻寬延伸技術(BandWidth Extension, BWE)完成高頻帶信號之編碼[5]。

AMR-WB+之 ACELP 編碼器，使用合成分析法(Analysis by Synthesis, AbS)架構，對每一音框信號編碼其線性預估係數(LPC)、基週(pitch)與固定碼簿之碼字，而固定碼簿與調適性碼簿之增益則一併作向量量化。固定碼簿使用 ACELP 型態之碼簿架構以簡化碼簿的搜尋計算量，每一音框信號再分割成 4 個長度為 64 點之次音框，對每一次音框進行固定碼簿最佳碼字之搜尋，AMR-WB+之 ACELP 碼簿搜尋採用深先搜尋法(Depth-First Tree Search, DFTS)之搜尋法則，以進一步減少編碼過程的計算量。

3. ACELP 固定碼簿之搜尋之研究

3.1 ACELP 碼簿搜尋法

ACELP 之固定碼簿的脈衝可能位置採用分軌(Track)方式，附表一所示為 AMR-WB+所使用的固定碼簿結構。執行碼簿搜尋時，必需在每一分軌，搜尋最佳脈衝位置以及該脈衝正 / 負極性，而每一分軌搜尋之最佳脈衝位置個數與所設定之位元率有關，附表二所示為 AMR-WB+在不同位元率下，每一次音框之碼字非零值之脈衝數。

AMR-WB+內部取樣率比一般 G.723.1 /

G.729 系列語音編碼器之 8 kHz 取樣率高，另一方面，其位元率也比一般語音編碼器之位元率高，使得每一次音框之碼字脈衝數亦較多，因此其固定碼簿搜尋之計算複雜度相對於目前普遍使用之語音編碼器高。

ACELP 編碼器之固定碼簿搜尋原則為：尋找最佳脈衝位置、極性和其增益值(G)使其合成語音與標的訊號(Target Signal, \mathbf{r})的均方誤差(E_ξ)為最小，即符合最小誤差準則：

$$E_\xi = \|\mathbf{r} - \mathbf{G}\mathbf{H}\mathbf{v}_\xi\|^2, \quad (1)$$

上式中， \mathbf{v}_ξ 為代數碼字之碼字，其在碼簿的索引值為 ξ 。在方程式(1)中，自相關矩陣 $\mathbf{H}=\mathbf{h}^T\mathbf{h}$ 為一個下三角拓撲旋積矩陣(Low Triangular Toeplitz Convolution Matrix)，而 \mathbf{h} 為聽覺加權(perceptual weighting)後之 LPC 合成濾波器之脈衝響應。依理論推導可知，最佳的碼字會使得方程式(2)中的 τ_ξ 得到最大值：

$$\tau_\xi = \frac{(\mathbf{r}^T \mathbf{H} \mathbf{v}_\xi)^2}{\mathbf{v}_\xi^T \mathbf{H}^T \mathbf{H} \mathbf{v}_\xi} = \frac{C_\xi^2}{\varepsilon_\xi} = \frac{(\mathbf{d}^T \mathbf{v}_\xi)^2}{\mathbf{v}_\xi^T \Phi \mathbf{v}_\xi}, \quad (2)$$

其中， $\mathbf{d}=\mathbf{H}^T\mathbf{r}$ 為標的(Target)訊號 \mathbf{r} 與脈衝響應 \mathbf{h} 的相關函數， $\Phi=\mathbf{H}^T\mathbf{H}$ 為共變異矩陣(Covariance Matrix)。 \mathbf{d} 與 Φ 於代數碼簿搜尋之前先計算求得，因此每一次音框只需計算一次。對 AMR-WB+之 ACELP 模式，此矩陣 Φ 之大小為 64×64 ，因為語音信號之高頻成份能量較小，因此以線性預估方法所得到的 LPC 合成濾波器，在統計上具備低通濾波器之特性，因此脈衝響應 \mathbf{h} 之數值會隨時間有逐漸遞減之趨勢，使得共變異矩陣 Φ 在偏離主對角線較遠之數值會偏低而可以忽略不計，參考論文[10]之方法，我們對共變異矩陣計算作簡化之計算，定義符號 $\phi_r(k)$ 如下：

$$\phi_r(k) = \Phi(64 - k - r, 64 - k), \quad (3)$$

其中矩陣 Φ 主對角線之 $r=0$ ，與主對角線平行之對角線編號依序為 $r=1, 2, \dots, 63$ ，在編號 r 之對角斜線上，右下至左上之 k 值依序為 $k=0, 1, 2, \dots, (63-r)$ 。利用脈衝響應 \mathbf{h} 之遞減趨勢，當 $|\phi_r(0)|$ 小於臨界值 η 時，該對角線上之 $\phi_r(k)$ 直接設為零，此外當式(4)之條件滿足時，直接設定後續 $\phi_r(k)$ 之值為 $\phi_r(K)$ ， $k=K+1, K+2, \dots, (63-r)$ 。

$$|\phi_r(K+1) - \phi_r(K)| \leq \alpha \quad (4)$$

3.2 非零碼字脈衝位置之預測機制

本文對 ACELP 編碼核心之固定碼簿搜尋

作簡化之研究，利用固定碼簿編碼過程中的相關函數 \mathbf{d} 所提供的資訊，來預先篩選較可能出現編碼脈衝位置的區域，以減少固定碼字中脈衝之候選位置，因而降低固定碼簿搜尋的計算量。因為代數碼字之碼字有 N 個非零之脈衝位置 $\{p_0, p_1, \dots, p_{N-1}\}$ ，脈衝的正/負極性與該位置之函數 \mathbf{d} 之極性相關性極高，因此最佳的碼字傾向於使方程式(5)中的數值為最大；

$$C_{\xi}^i = \sum_{i=0}^{N-1} |d(p_i)| \quad (5)$$

經由式(5)得知，非零脈衝位置較可能落在函數 \mathbf{d} 絕對振幅較大之位置附近，因此我們可以藉此預選可能最佳位置的組合，以簡少所需之搜尋組合，進而降低計算量。當然會使得式(5)為最大之碼字未必是最佳碼字，因為不同碼字亦會使得式(2)的分母數值有所不同，有必要將每一分軌預選位置之鄰近位置再進一步搜尋，以得到最佳之固定碼簿碼字。在脈衝位置預選時，先對每一分軌上之 16 個預選位置，依位置分成 M 區，找出 S 個具有較大之函數 \mathbf{d} 值之區域，然後將篩選過 S 個區域中之脈衝位置，以 DFTS 法進一步搜尋出所需之最佳脈衝位置。

4. ACELP 固定碼簿快速搜尋法之效能評估

本實驗針對低複雜度 ACELP 碼簿搜尋法作效能評估，在實驗中我們比較三種方法，一者為 AMR-WB+編碼標準中所採用之深先搜尋法(DTFS)，另一者為參考文獻[4]中所提之 CP 法結合 DTFS 演算法之搜尋法(簡稱為 CP 法)，再者為本文所提之候選脈衝位置預測結合 DFTS 演算法與共變異矩陣計算簡化法(以下簡稱為 FCP 法)。首先對共變異矩陣計算簡化之參數作評估，因為 AMR-WB+之編碼信號包括音樂、語音及語音混合著音樂之音訊型式，編碼器在壓縮信號前並未對信號型式作正確之判別，因此未必只有語音信號才會採用 ACELP 編碼核心，所以本實驗之測試信號樣本採用各種型式之信號，語音信號採用 16kHz 取樣寬頻語音信號之聲音樣本，該聲音樣本取自於 TIMIT 之英語聲音樣本[9]，男聲與女聲音長分別為 593 及 610 個超音框。音樂與混合信號樣本取自 VOICEAGE 公司網頁[12]中之純音樂樣本(music)與混合聲音樣本(Speech over music)。

由於 ACELP 在編碼器之作用為低頻信號

編碼，為評估簡化演算法對聲音品質之影響，我們不對寬頻解碼器之輸出信號作品質量測，而是將低頻編碼核心之輸出合成信號與原始寬頻語音信號經過 AMR-WB+編碼器所得到之低頻帶信號(LF)作比較，計算每一音框之分段訊號雜訊比(Segmental SNR, SEGSNR)，藉此評估在不同碼簿搜尋法下之合成信號品質差異。

由於音樂與語音信號頻譜特性之差異，第一個實驗先透過測試選取 FCP 法中共變異矩陣計算簡化之參數。利用 TIMIT 語音信號樣本與音樂信號樣本(classical)，統計在不同參數下所需計算之矩陣元素 $\phi_r(k)$ 之比例及對信號品質之影響，計算比例之實驗數據如附圖三所示，其中臨界值 η 選擇 0.01，在該圖中水平軸為式(4)中之臨界值 α 之值，由圖三可看出音樂信號因為其頻譜不如語音信號具備明顯之低通特性，所需之計算比率高於語音信號，由於編碼器未對輸入信號模式做正確判別，所以選擇之臨界值參數值必需同時適用於各種不同之信號類型，為避免簡化計算對聲音品質影響太大，在以下實驗中我們選擇 α 之值為 0.007，在圖三，音樂信號與語音信號之計算比例各約為 49%與 39%。

由於 ACELP 被設計作為語音類型信號之編碼技術，因此在第二個實驗只針對寬頻語音信號樣本，實驗時將 AMR-WB+程式之編碼模式強制設定成 ACELP 模式。在此實驗採用 DFTS、CP 與 FCP 三種不同碼簿搜尋法，並量測合成音訊品質，實驗結果如附表三與圖四所示，其中 CP 與 FCP 法之區域分割數 M 與篩選區域數 S 之值分別為 8 與 4，因此可將搜尋之迴圈數降為原始標準之一半，相對的降低其碼字搜尋之計算量約 50%。與原始編碼標準之 DFTS 法的合成語音相較之下，本文所提之 FCP 法之分段訊號雜訊比平均約下降 0.036dB，而 CP 法下降約 0.167dB，由圖四得知 FCP 法之信號 SNR 量測值接近於原始編碼標準之方法且優於 CP 法。

在第三個實驗我們將各種不同型式之音訊輸入 AMR-WB+ 編碼器作編碼，此時不再將編碼模式強制設定成 ACELP 模式，因此每一音框可能為 ACELP 或 TCX 編碼，為避免 BWE 技術影響品質之評估，我們仍將低頻帶之合成訊號與原始信號之 LF 部份作比較，聲音樣本包含兩組語音樣本及純音樂樣本與混合聲音樣本。根據實驗統計得知，男聲與女聲音樣本採用 ACELP 編碼模式之音框數比率分別為

68.6%與 63.6%；而音樂與混合信號樣本採用 ACELP 編碼模式音框數比率分別為 0.8%與 13.3%。我們將 AMR-WB+編碼標準中所採用之深先搜尋法與本文所提之 FCP 法之低頻合成輸出信號作品質量測結果如附表四所示，由實驗結果得知，其信號品質之 SNR 量測數據下降低於 0.01 dB，若透過主觀聆聽比較解碼器之輸出合成語音，幾乎無法分辨信號品質之差異。

5. 結論

本文針對 3GPP AMR-WB+音訊編解碼器之 ACELP 編碼核心，研究固定碼簿搜尋之快速演算法以降低其碼字搜尋之計算複雜度，透過候選碼字脈衝位置之預先篩選機制與共變異矩陣之簡化計算，降低 ACELP 碼簿搜尋之計算量。利用不同之聲音樣本，我們測試快速演算法對各種不同音訊信號型式之信號品質之影響，經由實驗證實，所提之演算法結合 3GPP AMR-WB+之深先搜尋法，可再進一步降低其 ACELP 碼簿搜尋之計算複雜度約 50%，經由聆聽實驗證實，幾乎無法察覺原始方法與快速搜尋法兩種情形下輸出語音品質之差異。

參考文獻

[1] 3GPP, TS 26.190 : "Adaptive multi-rate wideband speech transcoding".

[2] 3GPP TS 26.290 : "Extended AMR Wideband codec; Transcoding functions".

[3] 3GPP TS 26.304: "ANSI-C code for the Floating-point; Extended AMR Wideband codec".

[4] Chen F. K., Yang J. F. and Yan Y. L., "Candidate scheme for fast ACELP search," *IEE Proceedings - Vision, Image and Signal Processing*, Vol. 149, No. 1, pp.10-16, 2002.

[5] Ekstrand Per, "Bandwidth Extension of Audio Signal by Spectral Band Replication," *Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium, 2002.

[6] Lefebvre R., Salami R., Laflamme C., and Adoul J. P., "High quality coding of wideband audio signals using Transform Coded excitation (TCX)," *Proc. ICASSP'94*, pp. 1-193-196, 1994.

[7] Makinen, J.; Bessette, B.; Bruhn, S.; Ojala, P.; Salami, R.; Taleb, A., "AMR-WB+: a new audio coding standard for 3rd generation mobile audio services", *Proc. ICASSP'05*, pp.ii/1109 - ii/1112, 2005.

[8] Salami, R.; Lefebvre, R.; Lakaniemi, A.; Kontola, K.; Bruhn, S.; Taleb, A., "Extended AMR-WB for high-quality audio on mobile devices," *IEEE Communications Magazine*, Vol. 44, No.5, pp.90 - 97, May 2006.

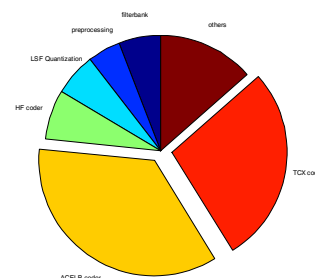
[9] TIMIT, DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (CD-ROM), *NIST Speech Disc 1-1.1*, USA. 1990.

[10] Tsai S. M. and Yang J. F., "Efficient algebraic code-excited linear predictive codebook search" *IEE Proceedings - Vision, Image, and Signal Processing*, Vol. 153, No. 6, pp. 761-768, 2006.

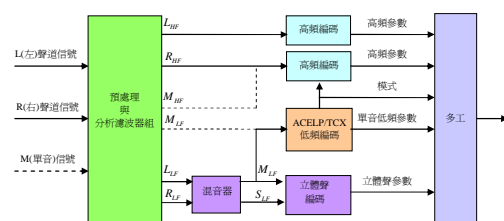
[11] Wang M. L. and Yang J. F., "Generalized Candidate Scheme of Stochastic Codebook Search for Scalable Coders", *IEE Proceedings: Vision, Image and Signal Processing*, Vol. 151, no. 5, pp. 443-452, 2004.

[12] WWW homepage, <http://www.voiceage.com>.

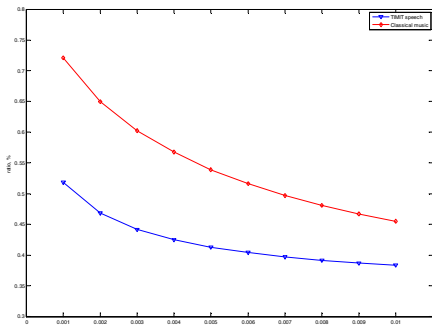
圖表



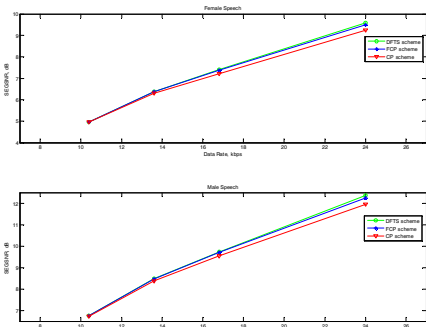
圖一：編碼器計算複雜度分析 (16.8kbps, 單音)



圖二：AMR-WB+之編碼器架構



圖三：不同臨界值下
需計算之變異矩陣元素之比率



圖四：各種搜尋法之合成語音品質比較

表一：AMR-WB+的固定碼簿結構

分軌	可能脈衝位置
T1	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
T2	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
T3	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
T4	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

表二：ACELP 碼字之非零值脈衝數

位元率 (kbps)	脈衝數
10.4	4
12	6
13.6	8
15.2	10
16.8	12
19.2	16
20.8	18
24	24

表三：各種搜尋法之合成語音品質比較表
(SEGSNR, dB)

		位元率 搜尋法			
		10.4 kbps	13.6 kbps	16.8 kbps	24 kbps
男 語 音	DFTS 法	6.756	8.490	9.741	12.370
	CP 法	6.733	8.386	9.553	11.960
	FCP 法	6.752	8.475	9.728	12.252
女 語 音	DFTS 法	4.973	6.388	7.407	9.580
	CP 法	4.962	6.300	7.227	9.246
	FCP 法	4.963	6.382	7.373	9.491

表四：各種音訊信號在不同演算法下
之品質比較表 (SEGSNR, dB)

		位元率 搜尋法		
		10.4 kbps	13.6 kbps	24 kbps
男 語 音	DFTS 法	7.140	8.806	12.592
	FCP 法	7.139	8.793	12.502
女 語 音	DFTS 法	11.301	7.969	6.520
	FCP 法	11.237	7.969	5.515
音 樂	DFTS 法	9.638	11.583	16.168
	FCP 法	9.638	11.584	16.163
語 音 混 合 音 樂	DFTS 法	7.413	8.859	12.601
	FCP 法	7.409	8.839	12.582