

# 行政院國家科學委員會專題研究計畫 期中進度報告

## 無線串流多媒體之新生代可調式聲視訊編碼技術研究(1/3)

計畫類別：個別型計畫

計畫編號：NSC94-2213-E-009-144-

執行期間：94年08月01日至95年07月31日

執行單位：國立交通大學電子工程學系及電子研究所

計畫主持人：杭學鳴

計畫參與人員：陳繼大，林鴻志，黃育彰，韓志岡，鄭凱庭

報告類型：精簡報告

報告附件：國際合作計畫研究心得報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 5 月 29 日

# 行政院國家科學委員會專題研究計畫成果報告

## 無線串流多媒體之新生代可調式聲視訊編碼技術研究(1/3)

### Emerging Scalable Audio/Video Coding Technologies for Wireless Streaming Multimedia (1/3)

計畫編號: NSC 94-2213-E-009-144

執行期限: 94 年 8 月 1 日至 95 年 7 月 31 日

主持人: 杭學鳴 國立交通大學電子工程學系 教授

計畫參與人員: 陳繼大, 林鴻志, 黃育彰, 韓志岡, 鄭凱庭

交通大學電子工程學系 研究生

#### 中文摘要

多媒體服務 (multimedia service) 咸信為寬頻無線網路的最重要推手。在無線寬頻網路上傳輸即時多媒體資訊, 面臨許多挑戰, 例如封包遺失 (packet loss), 網路與接收機非同質性 (heterogeneity) 等特性。因此, 新一代的通用音訊編碼 (universal audio coding) 與可調式視訊編碼法 (scalable video coding) 目前正熱烈發展中, 以降低接收機複雜度與通道穩定度等問題。

本計畫之主要目標在採用通用音訊編碼與可調式視訊編碼法技術, 來研究與製作寬頻無線網路環境中的串流音訊與視訊系統。本計畫全程為三年, 本報告將前十個月的成果做一摘要敘述。經過模擬實驗經驗, 我們的通用音訊編碼方案結合 AAC 與 LPC 之技術。LPC 需從新設計以符合現有之音訊壓縮方式。另外, 心理聲學模式亦需改善, 以適用於低位元率之壓縮架構。針對低位元率壓縮, MPEG-4 version 3 訂定了 HE-AAC。本年度已對其演算法做加速, 將之實現在 DSP 上。加速的項目包括縮小 Transient Detector 之偵測範圍, 簡化 Window Grouping 步驟, 降低心理聲學模式計算量等。在不影響聲音品質下, 經由加速後, HE-AAC 之速度提升了 1 倍。我們也開始研究 MPEG 可調式視訊編解碼 (Scalable Video Coding) 標準, 目的在設計低複雜度編解碼器。為達到空間、時間與品質可調性, 現有的可調式視訊參考軟體非常複雜, 運行緩慢。我們經由模擬與演算法分析, 已找出應加速之模組, 正進行加速的設計中。最後一項為我們與美國伊利諾大學香檳分校, 胡文美 (W-M. Hwu) 教授之團隊進行的國際合作計畫。我們協助簡化 MPEG-4 視訊演算法, 設計低複雜度架構, 以適合軟體/硬體共同設計平台。上述這些課題我們在未來兩年將持續發展, 並做更詳盡的報告。

**關鍵詞:** MPEG, DSP, 音訊壓縮, 語音壓縮, AAC, HE-AAC, SBR, 心理聲學模式, 軟體/硬體共同設計, 可調式視訊編解碼。

#### Abstract

Multimedia service is believed to be the driving force for developing wide-band wireless communication systems. However, there are a number of challenges in delivering real-time multimedia over wideband wireless networks such as packet loss and heterogeneous mobile receivers. New technologies for audio and video representations such as universal audio coding and scalable video coding are thus under development.

The goal of this project is to study, simulate and implement streaming audio/video algorithms based on the universal audio coding and scalable video coding concepts. The entire duration of

this project is 3 years. We will outline the intermediate results for the first 10 months in this report. The universal audio codec in our design is a combination of AAC and LPC (Linear Predictive Coding). The LPC component is redesigned and tuned to fit into this structure. The key is to modify and include an adequate psychoacoustic model into this system. HE-AAC is a very efficient audio coding algorithm specified in the MPEG-4 version 3. We successfully reduce its computation and memory requirements in DSP implementation. We simplify the transition detection procedure, window grouping process, psychoacoustic model and other modules. At the end, we achieve over 50% complexity reduction with little subjective quality loss. Another topic we started is the on-going MPEG scalable video coding (SVC) standard. After carefully studying its algorithm and reference software, we found that SVC achieves spatial, temporal, and SNR scalability at the price of a tremendously high computational complexity. Modules are identified for future complexity reduction. The last topic is an international collaboration project with Prof. W. Hwu at University of Illinois at Urbana-Champaign (UIUC) on efficient video codec implementation. The target is to port MPEG-4 and other video algorithms to a software/hardware co-design platform. Preliminary results show significant speed-up when a well-designed C program is adopted as the starting point in the porting process. Further research results on all the above topics are to be reported in the next two years.

**Keyword :** MPEG, DSP, Audio/Speech Compression, AAC, HE-AAC, SBR, Psychoacoustic model, Software/Hardware Co-design, Scalable Video Coding, SVC

## A. Part 1: Universal Audio and Speech Coding

### A.1 研究目的與背景

近年來由於手機頻寬的逐漸增加，直接在手機上傳送多媒體的需求逐漸的增加。因此，直接經由有限的手機頻寬，傳送音訊訊號成為逐漸提升的需求。然而，音訊編碼(audio coding)及語音編碼(speech coding)其資料來源雖然皆為音波，由於應用領域的不同，一直以來皆無法使用一個技術來同時達成這兩種壓縮。

為了達到能在手機上傳送多媒體，必須考慮三種因素。最重要的因素是壓縮率。在狹小的頻寬下，壓縮率的需求是非常緊迫的。在需要即時傳輸音訊的情況下，傳輸所能提供的頻寬僅 14~24k bps。相較於音訊編碼所需的 48k bps，位元率少了許多，這使得一般的音訊編碼無法達到有效的品質要求。此外，傳送多媒體時有時候也會有夾雜音樂，甚至於同時含有音訊與語音訊號的傳輸需求。在這種情況下，手動切換是不可行的；即使使用自動切換，也必須良好的切換方法與兩種編碼之間轉換機制的配合。這代表著不可能單純的以同時放兩個 codec 的方式來實現。最後，硬體的複雜度對手機來說是很重要的因素。由於音訊及語音編碼的不同，一隻多媒體的手機需要同時包含兩種不同的 codec。這代表著一隻手機可能要為這多出來的 codec 而增加額外的成本，甚至於導致成本過高。

由於上述三種因素，一個能同時壓縮音訊與語音之 codec 有其必要性。這個 codec 最好是不需要模式(mode)的轉換。本計畫擬從 MPEG 現有標準出發以發展 universal coding 演算法，使其達到同時達到以相似之品質壓縮音訊與語音，並能在極低之位元率下壓縮音訊訊號。

### A.2 文獻探討

MPEG-4 是 ISO/IEC MPEG Committee 所訂定的標準，第一版的 MPEG-4 於 1998 年完成，並於 1999 年 12 月正式出版。相較於 MPEG-2 AAC (Advanced Audio Coding)，MPEG-4 加入了許多新的工具(tool)來增加壓縮效率。其中 TwinVQ 就是在此時加入，與純量量化 (Scalar quantization)和 Huffman 編碼部分互相替換的編碼 (亦即如果採用了 TwinVQ，就無

法使用純量量化和 Huffman 編碼)。

TwinVQ 其全名為 Transform-Domain Weighted Interleave Vector Quantization。相較於 MPEG-4 AAC，TwinVQ 在低位元率(bitrate)時，壓縮效率比 MPEG-4 AAC 要好。也因此，MPEG-4 GA 的最低位元率可下降到每個聲道 6 kbps。由於 TwinVQ 採用向量量化的關係，其複雜度比 MPEG-4 AAC 高且與 MPEG-2 相容性較低。TwinVQ 大致上可以分為 4 個部分，如 Fig A-1 所示。輸入訊號將先經過三階段展平(flattening)，直到第四部分才正式進入向量量化部分。

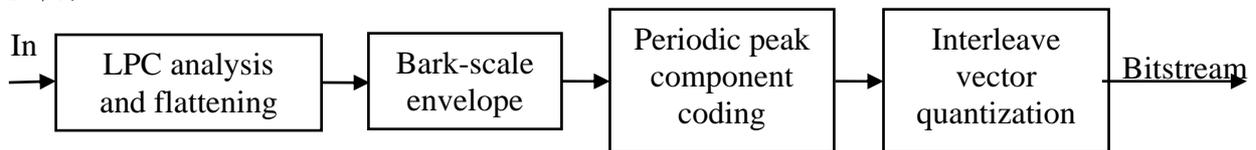


Fig A-1 TwinVQ 架構圖

仔細觀察 TwinVQ，可以發現 LPC 及 pitch 的分析都是語音編碼常見而在 MPEG-4 AAC 所沒有的部分。實際上，TwinVQ 確實有較佳之語音編碼效能。然而，即使是 TwinVQ，其語音壓縮效率仍然與傳統語音編碼相距甚遠。

### A.3 研究方法

由於 TwinVQ 與語音編碼的相似性比較高，亦有尚佳之語音編碼效能。因此選擇 TwinVQ 作為研究起始點，改善其效能。

#### A.3.1 TwinVQ 之分析與其缺點

根據觀察 TwinVQ 的設計架構中缺少音訊壓縮中的一個重要技術，那就是心理聲學模型 (Psychoacoustic Model)。由於來自於語音壓縮的技術，TwinVQ 在做向量量化(vector quantization)時，僅使用簡單的 weighting function。然而這種方式使得量化後之訊號並非與心理聲學模型配合的很好，這代表著 TwinVQ 將浪費位元來壓縮人耳聽不到的音樂。

然而在實際工作時卻發現，在 TwinVQ 的架構下，心理聲學模型難以置入。主要的原因在於 TwinVQ 為了增加向量量化之效率，必須採用 interleaving 方式將所有的 MDCT 係數打散，這導致無法在向量量化時，計算雜訊遮蔽比(Noise to Masking Ratio)。因此我們捨棄向量量化，改為原本 AAC 之純量量化(Scalar quantization)，重新設計整個 codec 之架構。

#### A.3.2 Universal audio and speech codec 之設計架構

我們以 AAC 為基礎，並將 TwinVQ 之 LPC 加入以期改善音訊壓縮之效能。Fig A-2 是目前發展之架構，其中 LPC 來自於 TwinVQ 之想法，以 frequency domain 的方式來實現 LPC，以期 LPC 能相容於目前 AAC 之架構。Other tools 則為 MPEG-4 之其他 tools，這些 tools 目前尚未打開。

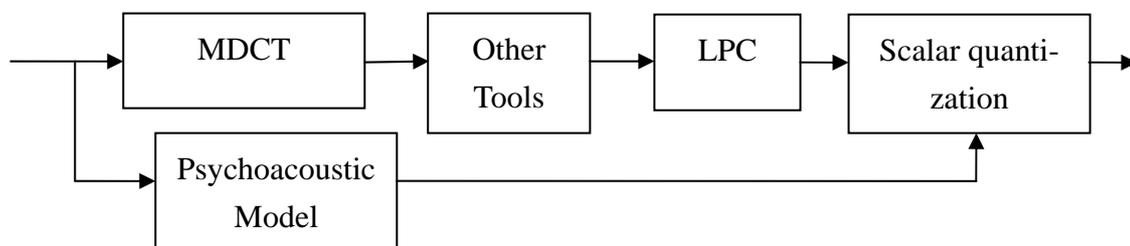


Fig A-2 A.3.2 Universal audio and speech codec 之設計架構

#### A.3.3 低位元率之調整

經由實驗的結果發現，MPEG-4 之 reference software 無法在低位元率下運做，一旦位元率低於 40kbps 時，程式將會產生錯誤。根據觀察後發現發生的原因為缺乏適當的頻寬限制

調整與量化之 Distortion loop 之中斷條件設計不良。由於頻寬之調整僅需根據實驗來給予固定之數值，又其與量化有很大之關係。如能使用少量之位元來壓縮係數即可以使用較大之頻寬，因此，調整 Distortion loop 之中斷條件為首要之工作。

原本的 reference software 之中斷條件為當雜訊遮蔽比(Noise to Masking Ratio)低於一個動態調整之比值時，即跳出此迴圈。然而這個條件卻有可能在位元不夠時，此比值過高導致無法跳出迴圈。

新加入的中斷條件有兩個。以程式的觀點來說，由於 AAC 其 scale factor 最大為 60，因此以當相鄰兩個 scale factor 皆為 60 作為第一個中斷條件，因為，此時增加 scale factor 不會有任何好處。另一個中斷條件為當每個 band 皆已 amplified 時即跳出中斷迴圈。

## A.4 結果與討論

為了比較 codec 之效能，我們使用 ITU-R BS. 1387[3]作為客觀評量(Objective Evaluation)的方法。Fig A-3為實現後之結果與原先 AAC 之比較，其中前 11 個訊號為音訊，剩下的則為語音。由比較圖可以發現，實現的架構其音訊壓縮效率在某些訊號下有會比較高，整個平均來看此架構之效能比 AAC 還好。但在語音壓縮部分仍須改善。

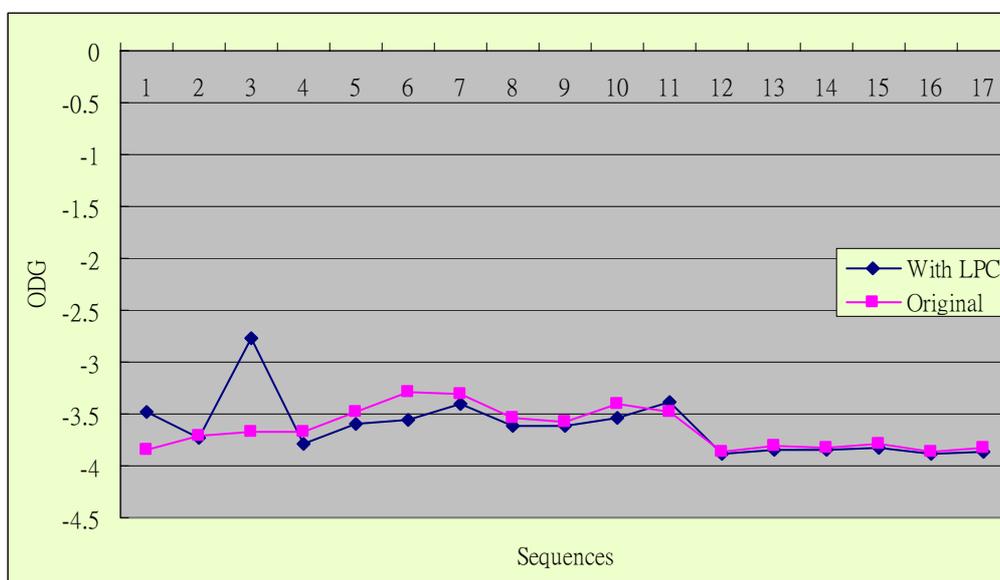


Fig A-3 64kbps 比較圖

經由實驗觀察，可能原因在於目前 LPC 之方法並不夠好，產生之 LPC 係數無法滿足某些訊號。為此，我們擬重新設計 LPC 之架構，以期能改善語音壓縮之效率。另外一個可能的因素是，對語音壓縮而言以 1024 sample 為一個 frame 實在是太長，必須對 AAC 之架構做更動才行。

對於未來之改善方面，除了 LPC 之改善外，目前 AAC 使用心理聲學模型方式偏向於高品質之壓縮。當位元率很低的時候，不可能將雜訊控制在遮蔽(masking)下。此外，將量化雜訊控制在與遮蔽平行的方式亦難達成。因此，低位元率時，必須調整心理聲學模型的使用方式或者心理聲學模型本身。

## B. MPEG-4 High Efficiency Advanced Audio coding

### B.1 HE-AAC 背景

MPEG-4 HE-AAC [4]技術是結合了 AAC [1]與 spectral band replication (SBR 頻帶複製技術)而產生的一種新的 audio coding，SBR 技術是德國 Coding Techniques 組織在 2001 年 12 月的

MPEG 會議中提出，其主要的概念是利用低頻與高頻的高度相關性，在傳輸的過程中只需要少量的 bits (side information) 來壓縮高頻訊號，解碼端再根據低頻的訊號與少量的高頻資訊還原高頻訊號，其壓縮效率比較 AAC 可以提升至少 30%，改善了在某些應用中 (手機傳輸系統)，傳輸頻寬不足的限制，使在低位元率的壓縮中依然有不錯 audio 品質的。

## B.2 HE-AAC Encoder 介紹

HE-AAC encoder 中，SBR encoder 與 AAC encoder 平行的處理高頻與低頻訊號。Fig B-1 為 SBR encoder 的架構圖，Input samples 先經由 quadrature mirror filter (QMF) bank，將訊號分成 64 個 subband signals，這些 subband signals 再經由 envelope estimator 和 various detectors 把高頻訊號的資訊分析出來。從 SBR encoder 的架構圖可以看到，transient detector、frame splitter、frame generator，用來產生 time/frequency grid，而 additional control parameters module 是用來偵測當解碼端把低頻訊號複製到高频時，可能會損失一些 tonality、sinusoidal 的成分，additional control parameters module 可以把這些損失的訊號還原。最後把 AAC encoder 與 SBR 編碼後的 bitstream 結合在一起，傳送到解碼端。

本計畫使用德州儀器 C6416T 這顆 DSP 作為實現平台，同時把 HE-AAC 編碼器實現在 Sundance 所提供的 DSP 平台上，並且針對需大量運算的部分做加速與演算法的改善，期望能加速並且降低消耗 DSP 資源。

## B.3 加速方法

我們將 HE-AAC 程式碼改寫並且放在 DSP 平台上，利用 DSP 系統所提供的 profile 功能，分析每個 module 的執行時間，結果如 Fig B-2。較耗時部分是下面討論加速的重點。

### B.3.1 Transient Detector Acceleration

從 decoder 解碼出來的 audio 訊號分析，很明顯的發現高於 16kHz 以上的訊號都不會被還原，在 HE-AAC 的定義中，複製的頻帶之前就會先有砍掉高頻的動作，因此在偵測 transient 我們也作類似的動作，分析出 T/F grid 時，高於 16kHz 以上的訊號不要分析，這樣一來可獲得相當程度的加速。

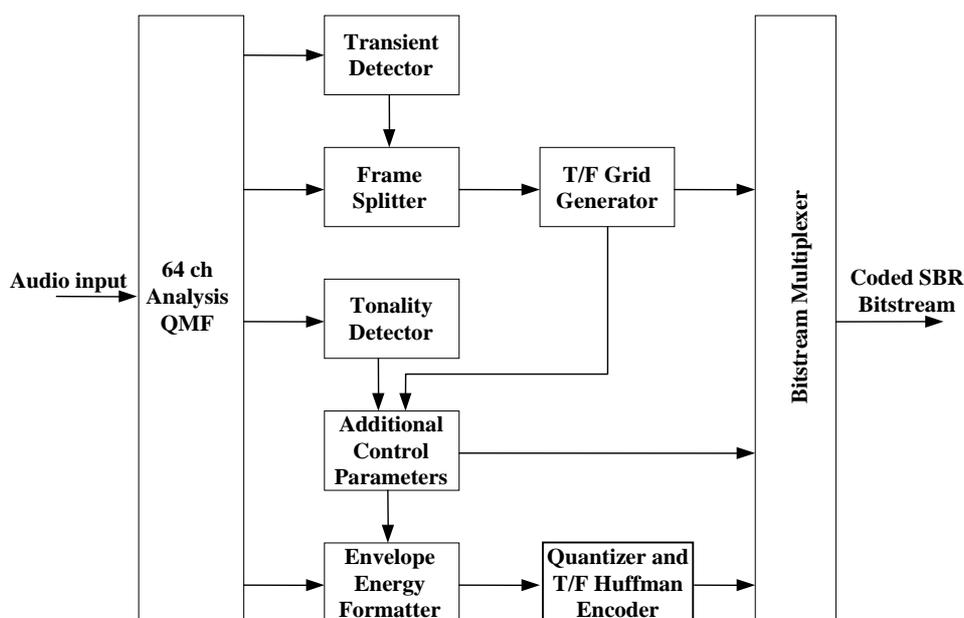


Fig B-1 SBR encoder

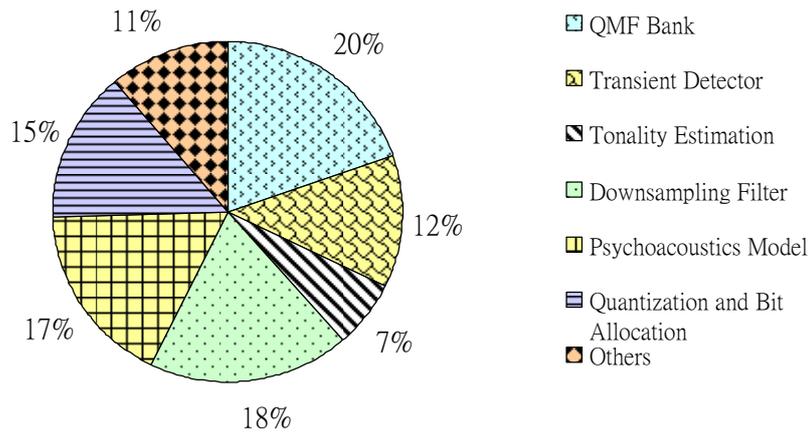


Fig B-2 HE-AAC 複雜度分析

### B.3.2 Simplified Long/Short Window Detection

程式中，判別用 Long or Short window 做編碼是在時間軸上做判斷，將一個 frame 中的 1024 個 samples 分成八個 groups，用能量的多寡去偵測出哪個 group 有 transient 產生，偵測到 transient 後，此 group 就判定為 short window。做法主要是先將 input samples 先通過一個 high pass filter，然後再用能量去分析 transient。此處需要通過 high pass filter 主要的原因是通常 transient 會產生的地方都在高低頻變化劇烈處。High pass filter 是想要把低頻的訊號做一過濾，使低頻訊號在判斷 transient 不會造成影響。

但此處的 high pass filter 需要與 input 做 convolution，將會花費相當多的時間，因此將這 filter 拿掉，可以改善運算時間，而簡化過後的 long/short window detection 並未對音訊品質造成太大的影響，主因是 TNS filter 也會針對 pre-echo 效應做處理，補償了此處 long/short window detection 可能的損失。

### B.3.3 Low Computation Psychoacoustic Model [6]

在 PAM 這邊的加速主要是把算 spreading function 所需要用到的參數都建成 table 的形式，用查表的方式來大幅降低運算 spreading function 的時間。Spreading function 所用到的參數與取樣頻率和位元率有關，因此根據不同取樣頻率與位元率可以建出不同的 table，用 memory 換取時間來達成加速。

### B.3.4 Simplified TNS Filter

TNS filter 主要是用一個 12 階係數的 linear predictive coding 所算出來的 prediction gain 與一 TNS threshold 比較看是否有大於此 threshold，若大於的話則做 TNS filter，並且傳送 filter 後的 residual。此處需要算 12 階的 LPC 非常耗時，因此檢查 TNS filter 需要使用到的次數，實驗的結果顯示啟動 TNS filter 的次數其實是非常少的。因此我們的作法是先用一個 6 階的 LPC 去算 prediction gain，6 階的比起 12 階來說可以降低運算量，而 6 階 LPC 會造成的錯誤幾乎小到只有 1% 以內。因此這個作法是可以降低複雜度，而且不會對降低音訊品質。

### B.3.5 Quantization Acceleration [7]

參照前人的作法，將 R-D control(two loop)的架構改成只需要 single loop distortion control，大幅降低 iteration 的數目，達到加速的效果。

### B.3.6 Window Grouping

MPEG 參考程式中會針對 short window 做 group 的動作，但是當分完 group 後再去做 quantization 及 bit allocation 時，並沒有特地去判斷哪些 group 需要分到比較多 bits。原作法是從第一個 group 開始做 quantization 與 bit allocation，因此可能即使做了 grouping，但是在 bit allocation 時，bits 分配仍會不平均，造成某個 group 有 transient 但有可能分配到較少的

bits，而造成失真的現象。另一個因素是分 group 後造成 side information 增加，encode 訊號的 bits 減少，使音訊品質變差。因此我們的作法是只用一個 group (原本的程式中都分成四個 group)，這樣一來在 quantization 時的 iteration 次數也可以減少，降低複雜度。最後結果發現只用一個 group 的音訊品質比分成四個 group 的品質來的好，這是令人意外的結果。

## B.4 結果

最後的結果可加速一倍以上。在音訊品質方面，用 EQUAL[3]測音訊品質大概會降低一點，但是人耳並無法聽出品質的差異。

## C. Scalar Video Coding

多媒體(Multimedia)影音技術在近年來被廣泛地應用在各種產品上，小至手持式之畫面接收器，大至高解析度接收機。在此一異質的接收與傳輸環境下，各種接收機擁有不同的傳輸頻寬、運算能力與解析度。

為了能夠讓編碼系統所編碼出的位元流(Bitstream)能夠讓各種不同頻寬大小的播放系統使用，MPEG 標準會議發展出了一個可調式的視訊編碼系統。其目的是希望能以單一位元流來滿足各種接收機的傳輸需求。目前，納入標準的整體架構是德國 HHI 提出以 MPEG-4 14496-10 精進視訊壓縮標準做為延伸的技術。預計於今年(2006年)年底會制訂完成。這一新壓縮標準是以 H.264 為基本層架構，可同時提供空間上(Spatial)、時間上(Temporal)、和畫質精細(Quality)可調的特性。

### C.1 文獻探討

#### MPEG-4 Part 10 Amd.1 可調視訊編解碼(Scalable Video Coding)標準技術

Fig C-1說明現階段可調視頻編碼標準的整體架構。在空間的可調性上(Spatial Scalability)目前的方法採用和過去 MPEG-4 和 MPEG-2 相似的概念，也就是利用降頻(Decimation)方式先對輸入視訊產生不同的畫面解析度。而為了提升壓縮率，目前的方法將低解析度層的動態補償預估後的殘餘訊號經過插補(Interpolation)用來預估高解析度層的殘餘訊號。此外，不同解析層的動態向量(Motion Vector)也採用類似的預估來改善壓縮率。而為了在每一個解析度上同時提供時間的可調性(Temporal Scalability)，在時間軸上的動態補償預估採用動態補償時間軸濾波(Motion Compensated Temporal Filtering, MCTF)。透過階梯結構(Lifting Structure)的轉換，動態補償時間軸濾波可進一步簡化成多層次的雙向預估結構(Hierarchical B Frame Prediction)，可容易達成時間上的可調性，同時也可用現有的 MPEG-4 14496-10 精進視訊壓縮標準來實現。另一方面，在畫質可調上(Quality Scalability)，預估過後的殘留值使用類似 MPEG-4 精細可調的位元層編碼(FGS Bit-plane Coding)做壓縮，使得每個解析度層都有一定的畫質可調範圍。

### C.2 可調視訊編解碼標準運算複雜度

可調視訊標準在編碼時需要大量的運算、記憶體搬移與額外的運算延遲。因此，對於即時的應用帶來極大的挑戰。以下我們針對每一種可調性來說明所需的運算和類型：

#### ● 時間可調性(Temporal Scalability)

其主要的運算包括了多層次的雙向預估和動態補償時間軸濾波。其中的運算包含了：

- 整數加減法用以選擇適合的動態向量
- 記憶體搬移並根據多層次的雙向預估和動態補償時間軸濾波結構做動態補償
- 額外的記憶體用以暫存在同一個GOP裡的多張畫面

#### ● 空間可調性(Spatial Scalability)

其主要的運算包括了 FIR 濾波用以完成層次間的插補(Interpolation)和決定層次間的估測模式。其中的運算包含了：

- 整數加減法與乘法用來做濾波運算

- 整數加減法用以決定層次間的估測模式
- 畫質可調性(Quality Scalability)
  - 其主要的運算包括了多重量化與熵編碼(Entropy Coding)，其中包含了:
    - 整數除法用來做多重量化
    - 單位元式的運算(Bit-wise Operation)與條件判斷
    - 額外的記憶體用以暫存DCT係數

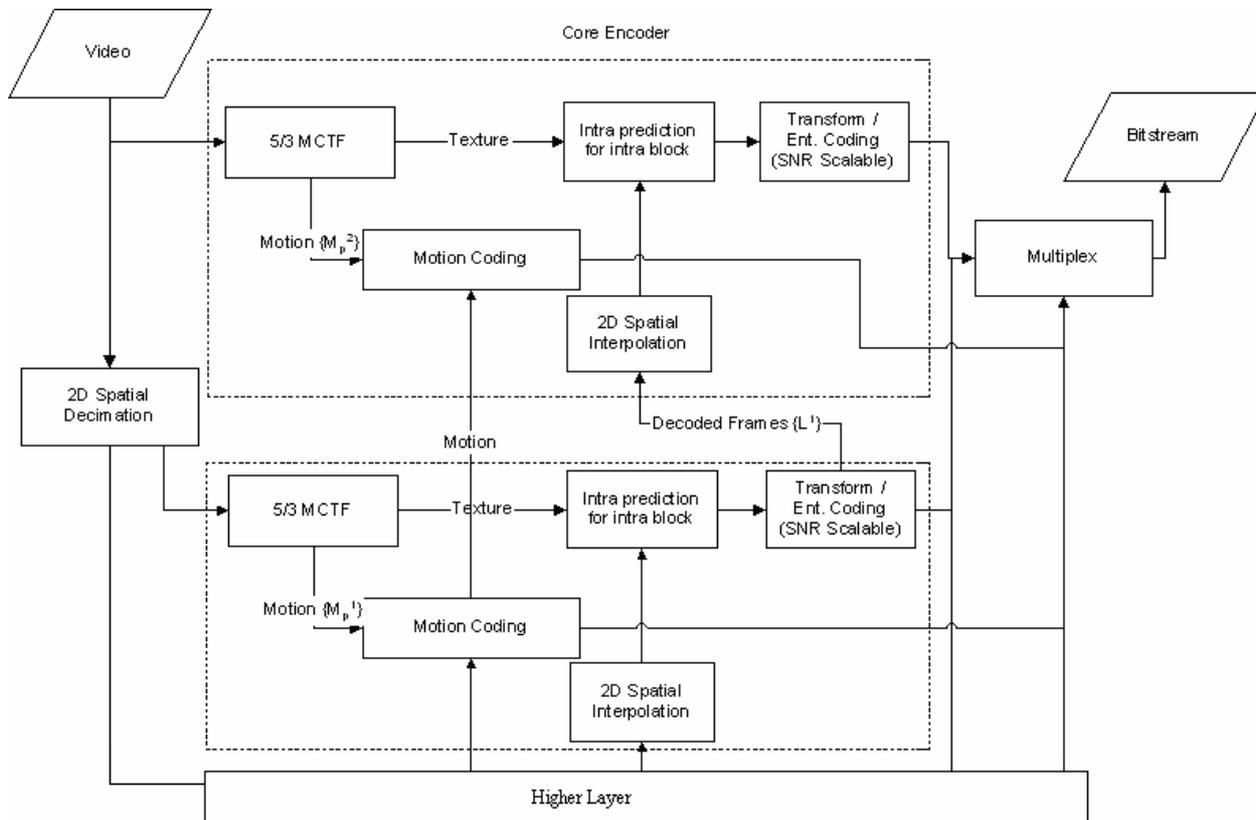


Fig C-1 可調視訊編碼架構

### C.3 結果與討論

實地使用可調視訊編碼參考軟體經驗，其速度極慢。因為使用大量的運算、甚高的記憶體頻寬與額外的記憶體，若要提高編解碼速度，則必須針對上述之運算模組作加速。

交大團隊(含其他老師)過去兩年積極參與 MPEG 可調視訊標準會議，提出數項提案。此外，我們撰寫專文描述 MPEG 可調視訊標準，見 Publication [5]。

### D. 針對低功率架構之 MPEG-4 視訊演算法之研究

MPEG-4 為一混合式(hybrid)編碼，由許多編碼工具組合而成，包括形狀編碼(shape coding)、移動估算 (motion estimation) 與補償、影像紋理編碼 (texture coding)、抗錯性、和可調性(scalability)。而如手機，數位相機等，MPEG-4 始停留在非常低階的應用，其主要原因在於 MPEG-4 編解碼器的設計困難，尤其許多高複雜度的運算造成消耗大量的計算能量，對於強調低功率(low-power)的產品來說，實現 MPEG-4 低功率編解碼器是一大挑戰。

因此我們以低功率需求之 MPEG-4 多媒體處理器為目標，並且進行國際交流，與美國伊利諾大學香檳分校胡文美教授之平行架構團隊合作，參與實作。結合我方團隊對於多媒體視訊之經驗對方之處理器設計能力，達到 MPEG-4 低功率處理需求。

## D.1 研究方法

主要進行有兩大方面：快速演算法與適用於硬體之軟體改進

### D.1.1 快速演算法

在快速演算法方面，由於原始的 MPEG-4 參考軟體的運算複雜度過高，除了會造成硬體實現的困難外，亦會造成編解碼過程之使用功率過高，因此我們利用快速移動估算法來降低移動向量搜尋複雜度，並且利用整數點餘弦運算來降低浮點運算造成的高運算量。

### D.1.2 適用於硬體之軟體改進

在與胡文美教授的合作上，所使用的平台為 Xilinx ML310 平台，此平台為利用 Power PC CPU 為硬體核心，並且利用 Linux 作為 embedded OS，可用來發展多媒體運算之運用。

我們將經過快速演算法改善過之 MPEG-4 軟體，再利用由胡文美教授實驗室所研發出的 IMPACT 分析軟體，可將程式切割成適用於軟體執行的部份(compiled in C)以及適用於硬體的部份(in HDL)，如Fig D-1；如此一來即可將運算的過程在考慮功率的分配下，分為軟硬體進行平行運算，可充分將 Xilinx ML310 之軟硬體資源進行利用，加速 MPEG-4 軟體的運算效率。

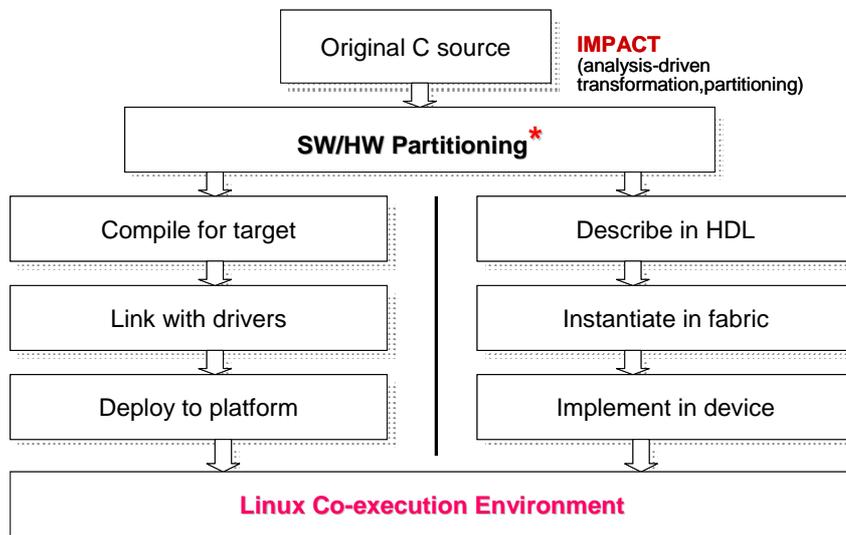


Fig D-1 利用 Xilinx ML310 之系統設計流程

## D.2 國際合作

由於主持人與研究生學期中得教學上課，已規劃於 2006 年 7 月初，計畫主持人和兩位博士生前往伊利諾大學，進行實質交流合作一個月。國際合作心得報告因此擬於八月初繳交。

## E. 參考文獻

- [1] ISO/IEC 14496-3, 1999, *Information technology – Coding of audio-visual objects – Part 3: Audio*
- [2] T. Moriya, A. Jin, N. Iwakami, and T. Mori, “Design of an MPEG-4 general audio coder for improving speech quality,” *IEEE Workshop on Speech Coding*, pp.139–141, Sept. 2000.
- [3] “*Method for objective measurements of perceived audio quality*”, ITU-R Draft Revision to Recommendation BS.1387, Dec. 1998.
- [4] ISO/IEC JTC1/SC29/WG11, “Draft ISO/IEC 14496-3:2001/Amd.2:2004 (Audio 3<sup>rd</sup> Edition)”, ISO/IEC JTC1/SC29/WG11 N7027, Jan. 2005.
- [5] Per Ekstrand, “*Bandwidth Extension of Audio Signals by Spectral Band Replication*”, IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002),

Leuven Belgium, Nov 15, 2002.

- [6] S. W. Huang, T. H. Tsai, L. G. Chen, "A low complexity design of psychoacoustic model for MPEG-2/4 advanced audio coding", IEEE Trans. on Consumer Electronics, Nov., 2004.
- [7] C. Y. Lee and et al., "A fast audio bit allocation technique based on a linear R-D model", IEEE Trans. on Consumer Electronics, Vol. 48, pp. 662-670, Aug. 2002
- [8] "Joint Draft 5: Scalable Video Coding," ITU-T and ISO/IEC JTC1, JVT-R201, Jan. 2006.
- [9] "Joint Scalable Video Model JSVM-5," ITU-T and ISO/IEC JTC1, JVT-R202, Jan. 2006.

## F. 計畫成果自評

多媒體服務是寬頻無線網路的最重要應用之一。本計畫重點在先進音、視訊標準編碼的演算法加速與實作，也開發設計適合在無線網路上傳送串流多媒體的音訊編碼法設計。本專題研究將承繼我們過去的經驗與前人的成果，進一步設計發展解決問題方式。所發展出的技術、經驗及成品極具實用價值，可促進國內工業研發技術開發。

參與工作人員(研究生)在學理上習得音訊與視訊編碼技術與國際標準。針對寬頻無線網路，設計開發可調式編碼等演算法，成員得到此課題研究與開發產品的經驗與知識。畢業後進入產業，直接有助於產業界開發新產品，提昇我國工業技術能力，達到人才培育之目的。

綜合評估：研究內容與原計畫進度與內容大致相符，已達成學術研究創新與人才培育之預目標。整體成效良好。研究成果頗具學術與應用價值，承繼之前的同一系列研究項目，已發表期刊論文一篇，兩篇學術會議論文，及專書一章以及碩士學位論文一冊如下表。並積極參與國際 MPEG 標準會議，將我人研發成果推展到國際舞台。

### Publications: (含前兩年同一系列計畫產出)

- [1] Yu-Chang Huang 黃育彰, *MPEG-4 High Efficient AAC Codec Acceleration and Implementation on TI DSP*, MS Thesis, NCTU, (to be finished in June 2005).
- [2] C.-H. Yang and H.-M. Hang, "Cascaded trellis-based rate-distortion control algorithm for MPEG-4 Advanced Audio Coding," *IEEE Transactions on Speech and Audio Processing*, Vol 14, pp.998-1007, May 2006.
- [3] C.-Y. Chen and H.-M. Hang, "DSP Implementation of AMR Speech Coding and Reed-Solomon Decoder in IEEE 802.16a," *2005 Workshop on Consumer Electronics and Signal Processing*, Nov. 17 – 18, Yunlin, Taiwan, 2005. (最佳學生論文獎)
- [4] Y.-M. Wang and H.-M. Hang, "MPEG-4 AAC Codec Acceleration and DSP Implementation," *2005 Workshop on Consumer Electronics and Signal Processing*, Nov. 17 – 18, Yunlin, Taiwan, 2005.
- [5] W.-H. Peng, C.-Y. Tsai, T. Chiang, and H.-M. Hang, "Advances of MPEG Scalable Video Coding Standard," book chapter in *Intelligent Multimedia Data Hiding*, to be published, 2006.

### 本研究群發表之相關 MPEG 標準文件(2005)

- 1. W.-H. Peng, Tihao Chiang and H.-M. Hang, "ISO/IEC JTC1/SC29/WG11 M11914: Adding Selective Enhancement Functionality to Scalable Video Coding," April 2005 (72nd, Busan, Korea)