

第二章

語音/視訊技術

■CODEC ■Video CPE

撰稿/通訊系統組

一、CODEC

(一)數位語音與音訊編解碼器(Speech and Audio Codec)

1.技術簡介與定義

雖然語音與音訊都是聲音，但是它們卻是有分別的。人類可以發聲的範圍通常稱為語音；而人類可以聽覺的範圍通常稱為音訊。一般而言，音訊的範圍比語音大。語音之基礎頻寬為 4KHz、音訊的基礎頻寬為 22.05KHz。現代通訊常將類比的語音訊號變成數位的語音訊號來傳送；近幾年來，音訊訊號也被轉化成數位訊號，在媒介上傳送或儲存。因而「語音與音訊編解碼(Speech and Audio Coding)」，可以將它定義成：將類比的語音訊號或音訊訊號經一定程序轉化為一組數位符號的過程以及其反向過程。「語音與音訊編解碼器」則是實行此一功能之器具。

2.發展現況與應用

語音為人類溝通的主要方式，在沒有發明文字以前語音就是人類社會中彼此聯繫的媒體。因而語音通訊的發展也相當早。從貝爾發明電話起，人類就逐步實現長距離通話的夢想。從那時開始，電話的功能隨著電子、電機技術的發展愈來愈精緻、愈來愈便宜。同時，隨著通訊技術的發展，也使電話

網路愈來愈普及、愈來愈快速。

近代，通訊更往數位化的方向發展，使得語音編解碼技術的需求也就愈趨明顯。除此之外，音訊隨著朝向數位化發展，而成爲多媒體通訊的重要一環。這種發展，乃肇因於語音與音訊數位化的優點，包括：

- (1)計算性：可以作各種計算式的處理。
- (2)壓縮性：可以節省通訊與儲存成本。
- (3)結合性：可與其他媒體整合成一體。
- (4)存取性：使隨意或指定存取更快速。
- (5)傳真性：多次拷貝可以達到零失真。
- (6)交換性：使訊號的交換更快速方便。

因而，語音與音訊編解碼的研究也就成爲現代的通訊技術的重要課題。

早期發展的編碼技術的發展以「語音編解碼」爲主。而「語音編解碼」技術以 PCM (Pulse Code Modulation)爲濫觴，它也是大部份「語音、音訊編解碼」技術的發展基礎。它是以簡單的取樣(sampling)與量化(quantization)將語音訊號以數位資料來代表。最常見的位元率爲每秒 64 仟位元。由於它的簡單，所以被廣泛地採用。CCITT(International Telephone and Telegraph Consultative Committee)即將其制訂爲標準，稱爲 CCITT G.711，乃爲「語音編解碼」標準之始。

從 64 Kbps PCM 開始，各種新「語音編解碼」技術就不斷地出現。發展的趨勢是往更低的位元率推進。在 32Kbps 層的代表爲 CCITT 的 G.721。它爲一典型的 ADPCM (Adaptive Differential PCM)，乃是應用語音訊號相鄰取樣(samples)的相似性作適應性預測將原來的 PCM 資料壓縮兩倍。另外，CCITT 也訂了 G.723、G.726 標準，它們仍爲 ADPCM，但它提供了 40Kbps, 32Kbps, 24Kbps, 16Kbps 等位元率的選擇。

中位元率的語音編解碼以 16Kbps 爲代表，有 CCITT 之 G.727 及 G.728。它們用的是 backward-CELP 技術。這一系列的標準重視編解碼的延遲時間。因爲，在通訊的應用上延遲時間過長往往會帶來迴音干擾等等問題。尤其在 DCME(Digital Circuit-Multiplication Equipment)、ISDN(Integrated Service Digital Network)、Packetized speech、Cordless telephones、Speech for videophone 等應用，低延遲變成語音編碼的必要條件。因而，低延遲的語音編解碼技術

也成爲研究發展的重點之一。

而 8Kbps 位元率的「語音編解碼」多半以 CELP (Code Excited Linear Prediction)爲主要運算法則。它是以 LPC (Linear Prediction Coding)爲骨幹加上 VQ(Vector quantization)技術混合而成的新技術。北美的 CTIA (Cellular Technology Industry Association)即提議將之訂爲數位式行動電話之標準。

隨著影音通訊之需，ITU-T 乃制訂一套試用於電話線上作數位語音通訊之新標準 G.723.1。G.723.1 是一個雙位元率的語音編解碼標準。它可以視影像傳輸的需求不同而使用不同的位元率。它有兩種位元率，分別爲 6.3Kbps 位元率與 5.3Kbps 位元率。所使用的編解碼技術分別爲 MPCELP(Multiple Pulse Code Excited Linear Prediction)及 ACELP(Algebraic Pulse Code Excited Linear Prediction)。它的音質在通訊應用可達到不錯之通話品質而它所產生之延遲則控制在 37.5ms 左右，算是一種很不錯之新技術，目前被廣泛應用於 PSTN 視訊電話及網路電話產品上。

4.8 Kbps 的「語音編解碼」以 NSA 的 Federal Standard 1016 爲代表。它仍是應用 CELP 技術，初始是要應用於軍方的保密電話，但後來不僅用於軍中，有些商用產品也採用相類似的方法。例如，AT&T 在 1990 開始生產 4.8Kbps 之 CELP 語音編碼器。

位元率低到 2.4 Kbps 的語音編解碼器多半是用 LPC vocoder 技術，語音品質多半只強調語意之傳達，在音色方面較無法顧及，通話雙方可能無法辨認彼此的個人聲音特色。而再低位元率的語音編解碼技術也有不少研究投入，但由於位元率低，所能呈現的語音品質則還不適用於一般語音通訊，這類技術範疇多半被分類至「語音合成」的領域。表 x-x-x-1 列出了目前國際上主要的語音編碼標準比較，其中 MOS (Mean Opinion Score)爲語音編碼品質評估之泛用標準依據。評估方式乃是由一群受過訓練之專業聽眾依照 5 分法評分以取其平均值，而評分方式有兩種：(A)依照話質的好壞：excellent (5), good (4), fair (3), poor (2), or bad (1)；(B) 依照語音損壞之可察覺程度：imperceptible (5), barely perceptible but not annoying (4), perceptible and annoying (3), annoying but not objectionable (2), or very annoying and objectionable (1)。

表x-x-x-1 Speech Coder Standard

發展年代	Bit Rates (kbps)	Description	MOS
1972	64	PCM (for PSTN)	4.4
1976	2.4	LPC-10 (U.S. Federal Standard 1015)	2.7
1984	32	G.721 ADPCM (for PSTN)	4.1
1990	4.15	INMARSAT (satellite)	≈3.2
1991	13	GSM (European cellular)	3.6
1991	4.8	CELP (U.S. Federal Standard 1015)	3.2
1992	16	G.728 (low-delay CELP)	4.0
1992	8	VCELP (NA cellular)	3.5
1993	1-8	QCELP (North America CDMA)	≈3.4
1993	6.8	VSELP (Japanese cellular)	≈3.3
1995	8	G.729 (new toll quality)	≈4.2
1995	6.3	G.723.1 (in H.323 and H.324)	3.98
1995	5-6	Half-rate GSM	≈3.4
1996	2.4	New low-rate Federal Standard	≈3.3

另外，除了位元率的降低外，「語音編解碼」考慮的頻寬也會隨需求而有不同。一般最常用的頻寬為 4KHz。4KHz 對一般用戶已算足夠，因為語音訊號大部份能量均集中在 4KHz 以下。但也有人採用 3.3KHz 以節省頻帶成本，也有增加頻寬以提高語音品質者，如 CCITT G.722 標準即將頻寬訂為 7KHz，位元率為 64Kbps，而這已跨越語音的範圍而進入音訊編解碼的領域。因此，業界即漸感音訊編解碼的需求，而競相投入音訊編解碼的研究。

為此，ISO 也招集了一群影音編解碼的專家討論制訂一套國際標準。通常稱為 MPEG (Moving Pictures Expert Group)，MPEG 除了視訊外還包含音訊的編解碼規範。MPEG1 於西元 1992 年底定案，規範了三層的音訊編解碼標準，目前市面上已出現相關的產品。MPEG1 之後，又有 MPEG2 的制訂，於西元 1994 年 11 月定案。音訊方面，在 MPEG1 上新增取樣率的選擇、位元率的選擇、新的位元分配表等。從西元 1995 年開始，MPEG 的專家們則開始投入制訂 MPEG4，也見到音訊編解碼技術的蓬勃發展。

而在美國杜比公司則發展出 AC-1、AC-2、AC-3 等音訊編解技術，其中 AC-3 更成為美國國家標準。在未來的 DVD 應用，AC-3 可能成為音訊解碼

之主要標準，因而格外引人注意。目前主要之商用音訊編碼標準比較請參考表 x-x-x-2。

表x-x-x-2 Comparison of Commercially Available Audio Coding Systems

	Bit Rates	品質	複雜度	主要應用	年代
MPEG-1 Layer 1	32-448 kbps total	Transparent @ 192 kbps/channel, as per (ISO 1991c)	Low decoder/encoder	DCC	1991
MPEG-1 Layer	32-384 kbps total	Transparent @ 128 kbps/channel, as per (ITU 1994)	Low decoder	DAB, CD-I DVD	1991
MPEG-1 Layer 3	32-320 kbps total	Transparent @ 96 kbps/channel, as per (ITU 1994)	low decoder	ISDN, satellite radio systems, internet audio	1993
Dolby AC-2	128-192 kbps total	Transparent @ 128 kbps/channel, as per (ITU 1994)	low encoder/decoder	Point to point, cable	1989
Dolby AC-3	32-640 kbps/channel	Transparent @ 384 kbps/5.1 channel, as per (ITU 1995)	low decoder	Point to multipoint, HDTV, cable, SD-DVD	1991
SONY ATRAC	140 kbps /channel		low encoder/decoder	MD	1992
AT&T PAC MPEG-AAC (NBC)	64 kbps/channel	Transparent	low decoder		1997

3. 未來發展趨勢

近年來的語音、音訊編解碼有三個大方向值得注意：(1)多媒體化；(2)網路化；(3)民生化。多媒體化使語音、音訊與影像、視訊相結合。比如，視訊會議系統或影音通訊應用。因為視訊與語音、音訊並存，所以許多標準設計是一起考慮的。代表性的是 MPEG1、MPEG2、MPEG4、H.320、H.324、H323 等標準，不但包含有視訊標準，也搭配有各自之語音或音訊標準。

而在網路熱潮之下，語音編解碼在網路上的應用也受到相當的關注，就連電信單位對此應用也投注密切的觀察。目前已出現網路電話的產品，使得

消費者能以相當便宜的網路費用打長途電話，這對傳統電話市場帶來相當大的衝擊，但目前仍有許多問題待克服。其中語音編解碼技術亦是重要的關鍵，現今所見的網路電話所用的語音編解碼技術是百家爭鳴的狀況，產品品質亦差異甚大。因此，如何設計一套適合網路通訊環境的語音編解碼器，是一項新的技術考驗，也有許多專家開始著手商討一套標準，不久的將來可能會有更明朗的結果。

因為媒體數位化的普及，使得數位語音與音訊之應用更形擴展，而其產品價格也不斷降低。使語音與音訊的產品走向大眾，這些相關的技術也就更深入生活成為民生產品中之重要技術。

4.結語

「語音與音訊編解碼」技術的發展是多樣化的。早期的「語音編解碼」發展多只注意位元率與語音品質。但漸漸地「延遲時間」、「容錯力」、「系統耗電量」、「頻寬」、「非語音訊號接受力」等等因素也被列入設計考慮重點。根據不同的應用，決定不同的設計重點。比方說，傳真機送出的訊號也要透過「語音編解碼器」時，該「語音編解碼器」就需要有能力接受非語音訊號。又如，用於個人通訊時可能「系統耗電量」就要考慮。此時，系統就不宜過於複雜以減低硬體需求。而「音訊編解碼」也正蓬勃發展，漸漸地進入實用領域。網路與多媒體的熱潮也在在影響此一技術的需求。可以預期的是，「語音與音訊編解碼」在通訊上將成為重要的技術，其未來的發展價值也將更形提昇，而民生化之趨勢也將使這些技術更有經濟價值。

(二)數位視訊編解碼器(Digital Video Codec)

1.技術簡介與定義

電視是本世紀的重大發明，讓人類可以見到遠方的表演。但是，人們一直都有個理想，希望傳送影像能像打電話一樣方便，不需要隨著電視台的節目表作息。這個理想在下一個世紀很快就能辦到。重要的關鍵之一，就在於數位視訊編解碼器。它的功能是將數位視訊資料轉化成容易儲存與傳送的型式。因而，「數位視訊編解碼器」的定義為一種將數位視訊資訊精簡成特

定的格式、並反解該特定格式成數位視訊的器具。

2.發展現況與應用

數位視訊編解碼的應用可以分成兩類，一為單向地儲存與傳送視訊視號；另一則為雙向的視訊通訊。例如，如電話一樣方便的視訊電話可以讓地球兩端的人不需要往返交通就能面對面溝通。或者，偏遠的醫院可以利用視訊傳輸技術與遠方的名醫進行即時的會診。這些都人們嚮往以久的應用，如今都可能實現，數位視訊編解碼技術則是這些應用的重要基礎。因為，要使每一個終端用戶都能享受到這種服務並不是一件簡單的工程，傳送傳統每秒上億位元的視訊資料並不是現有陳舊的通訊網路所能提供的，而重新架構更寬頻的通訊網路不但成本至巨，而且也曠日廢時，於是視訊壓縮便成為解決頻道資源有限的有效辦法。所謂視訊壓縮即將視訊中冗餘的部分及不可區分的部分除去的信號處理技術，視訊壓縮的處理過程可稱為「視訊編碼」，其反向過程則稱為「視訊解碼」。視訊編解碼器基本架構如圖 x-x-x-1 所示，不同的編解碼技術則應用不同的建構元件(transform, quantization, etc.)

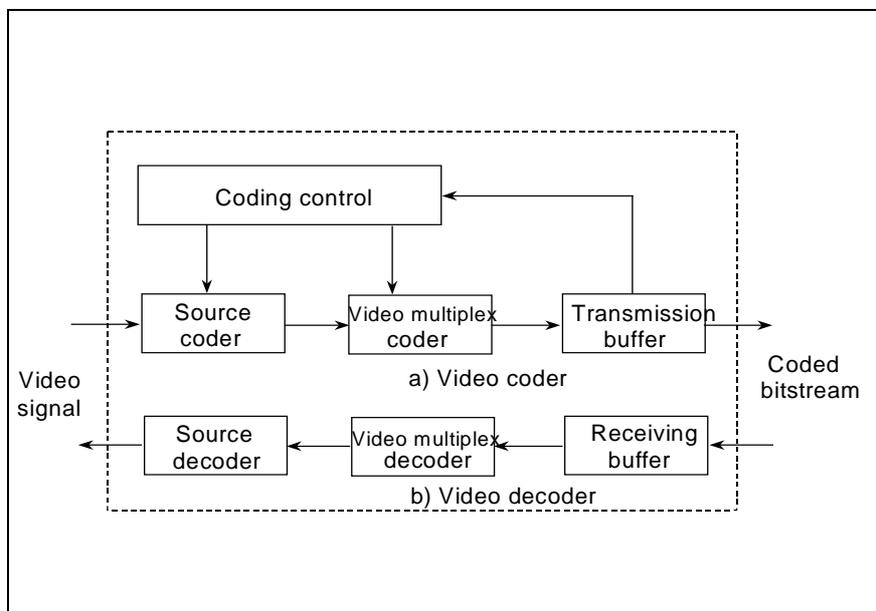


圖 x-x-x-1 視訊編解碼器基本架構

視訊編解碼技術發展已有一段歷史，從早期的單張靜態影像編碼發展至連續畫面的動態影像編碼，乃稱為視訊編碼。也有商品已經問市，例如，用於 CD-ROM 的 MPEG1 視訊編碼。而為了產品相容的需求，各國廠商紛紛投入標準產品之研究。國際標準組織如 ITU-T 及 ISO 也不遺餘力地推動國際共通之標準，並組成了若干工作小組，針對不同的應用和需求而研制各種共同的視訊編解碼標準。

為了提供娛樂產品應用，尤其是在多媒體方面的視訊需求，ISO 制訂了 MPEG1 標準。MPEG1 是針對 CD-ROM 的應用，而其位元率(bit rate)上限為 1.5 Mb/s，畫面達 768x576 像素。基本上，它的編碼演算法和 H.261 類似，其主要的差異在於 MPEG1 允許雙向的動向補償 (Bi-directional Motion Compensation) 並將每一幅畫面依編碼方式區分為 I、P、B 三類 (依次為 Intra、Predictive、Bi-directional)。I 畫面使用幅內編碼方式，P 畫面利用先前解碼過的 I 或 P 畫面作前向(forward)的動向補償，B 畫面穿插於 I 與 I、I 與 P 或 P 與 P 畫面之間，其編碼方式為雙向的動向補償。同時 MPEG1 將一個 I 畫面及若干個 P、B 畫面組合成一個畫面組(Group of Picture，簡稱 GOP)以防止因傳輸造成的錯誤傳播至下一個畫面組。

MPEG-1 基本上並無法滿足在廣播、HDTV 或有線電視等方面的需求。這些應用不僅要求更高的畫質和解析度，同時需要能透過各種網路做不同的服務。於是，ISO 所召集的專家又於 1994 年底完成制訂了 MPEG2 標準。與 MPEG1 比較，MPEG2 在各項特徵上都做了極大的擴充。在視訊格式它允許了 4:2:2 及 4:4:4 的彩訊格式，同時也增加了 interlaced 的輸入訊號。在解碼演算法方面它提出了 Field DCT、16x8 Motion Compensation、Field Prediction 及 Dual Prime 移動向量等新的處理方式。在解析度方面，MPEG2 定義了各種 profile 及 level，使用者可針對不同的應用而選擇不同的解析度。同時在 MPEG2 洋蔥式(onion ring)的文法架構下，定義了 spatial、temporal 及 SNR 等 Scalability 方式。使得解碼器在同一個位元流中，可根據解碼器的能力和網路週邊的容量等關係，解出低至 MPEG1，高至超過 HDTV 的解析度及畫質的不同視訊。

在雙向視訊通訊應用如視訊電話及視訊會議方面，則主要由 ITU-T 主導標準之制訂，ITU-T 首先於 1990 年底推出 H.261 標準，這是針對 ISDN

上應用的視訊會議系統而設計的。又被稱為「 $p \times 64$ codec ($p = 1 \sim 30$)」，因為其傳輸速率為 ISDN 中 B 通道的整數倍，而一個 B 通道有 64kb/s 的容量。H.261 所使用的解碼演算法為 Predictive Transform Coding，即對連續畫面先作畫面間的動向補償(Motion Compensation)，再對所餘的差值作離散餘弦轉換(Discrete Cosine Transform)，轉換後的頻譜則再加以量化，並以可變長度方式允予編碼。因其應用的傳輸速率不高，無法提供娛樂用途的高畫質畫面，所以以視訊會議(Videoconference)為其主要應用。

ITU-T 更進一步於 1996 年完成 H.263 標準。現在，H.263 標準受到相當地重視，原因是它可以在傳統的電話網路上傳送視訊訊號。基本上，H.263 的架構與 H.261 很相似。但因為視訊編解碼技術的進步，在應用了半像素(half-pel)運動估測技術及四種新的編碼選項(Un-restricted Motion Vector Mode, Advanced Prediction Mode, Syntax-based Arithmetic Coding Mode, 及 PB-frame Mode)，使其比 H.261 的效能高出許多，在相同的壓縮比，於 <64kbps 之應用可提昇 3~4dB 之畫面品質。乃使其在標準制訂會議中脫穎而出。再經各國的專家集思廣益，使它成為新一代視訊電話的標準。

而 ITU-T 的專家們更再接再厲地對 H.263 系列標準進行後續之改進及增訂工作。目前已接近完成了 H.263+ (前稱 H.263 version 2) 的標準，其主要是在原有的 H.263 架構上加入了 12 項新的編碼選項(連同原 H.263 之 4 項，共有 16 個編碼選項)以提高原有的壓縮效率及畫面品質。此外正制訂中的標準尚有預計於 2000 年完成之 H.263++ 及預計於 2002 年完成之 H.26L。其中 H.263++ 可和 H.263/H.263+ 前溯相容(backward compatible)；而 H.26L 則為新一代之標準，並不要求和 H.263 系列標準具有相容性，很多新技術可能和 MPEG-4 共用。

目前最令人矚目則為 ISO 進行之新一代標準 MPEG-4 之制訂。和原有編碼標準相較，在 MPEG-4 加入了相當多新的觀念和技術：如物件導向式編碼及處理(例如個別物件之存取、操作及判別等)，及提供了編碼工具之觀念以支援不同功能，也增加因應傳輸錯誤所造成之資料損毀之錯誤修補功能，並且允許虛擬物件和真實影像物件的混合編碼(Synthetic & Natural Hybrid Coding; SNHC)。這些新技術使得 MPEG-4 除了可以達成高效率之壓縮外，更增加了內容導向之互動式操作功能，其物件觀念請參見圖 x-x-x-2。MPEG-4

預定於 1999 年初完成第一版標準之制訂，其標準將允許和 MPEG-2 及 H.263 之前溯相容。

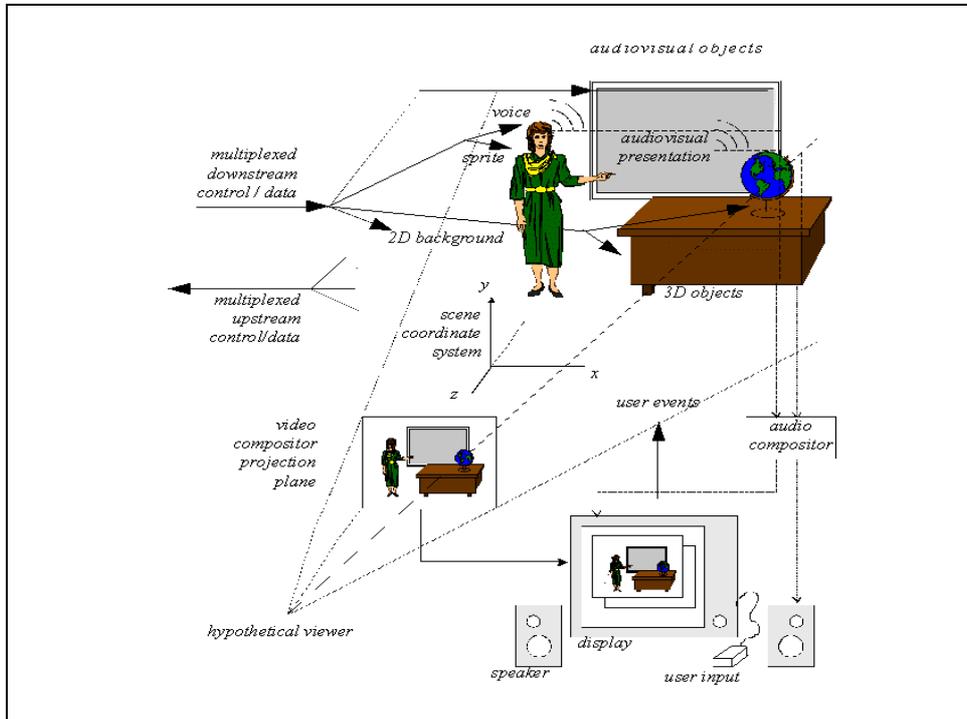


圖 x-x-x-2 MPEG-4 系統構成

3. 未來發展趨勢

往後的視訊編解碼器的發展將以兩系列的標準為主軸，並針對不同的應用發展出特有的視訊編解碼器，值得注意的有三種發展方向：

在 DVD (Digital Video Disc)、DVB (Digital Video Broadcast) 的應用將朝向更高畫質的畫面發展。這樣的應用以 MPEG2 的高層的 profile 及 level 為依據。但這類應用需要很寬的通訊頻道支持，也需要電影業者的配合，才能形成氣候。

在一般電話網路上之視訊電話的需求將愈來愈大，這種應用已受到許多廠商的關注。在 ITU-T 有 H.263 標準，在 ISO 有 MPEG4 的標準。這兩種標準都希望能在很窄的頻道上傳送視訊與音訊，以利大眾化影音通訊的需求。目前，H.263、H.263+ 已完成，而 MPEG4 也將完成，廠商也漸漸地推出產品，

可預期將來的商業潛力相當大，值得密切注意。

隨著網路通訊之發展，視訊傳送也漸漸成為網路應用的重要課題，於是 ISO 與 ITU-T 也召募專家投入下一代視訊編解碼之研究，包括 MPEG7 與 H26L，可稱之為「先進型視訊編碼(Advanced Video Coding)」。當然 AVC 可以用來更進一步改進視訊通訊之品質，亦可為網路應用提供一種更高效之視訊傳送與擷取方法。而如何於視訊編碼技術中結合網路通訊考量以提供網路視訊服務為不管是學術界或產業界均相當重視的課題，也是一個重要的商機所在。

表x-x-x-3 Comparison of Video Coding Standards

Standards Organization	Video Coding Standard	Typical Range of Bit Rates	Typical Applications
ITU-T	H.261	p×64 kbits/s, p=1...30	ISDN Video Phone
ISO	IS 11172-2 MPEG-1 Video	1.2 Mbits/s	CD-ROM
ISO	IS 13818-2 MPEG-2 Video	4-80 Mbits/s	SDTV, HDTV
ITU-T	H.262		
ITU-T	H.263	64 kbits/s or below	PSTN Video Phone
ISO	CD 14496-2 MPEG-4 Video	24-1024 kbits/s	
ITU-T	H.263 Version 2	< 64 kbits/s	PSTN Video Phone
ITU-T	H.26L	< 64 kbits/s	-

4. 結語

數位視訊訊號的傳輸受限於傳輸通道的容量不足與視訊資料的龐大。解決的方法除了增加傳輸通道的頻寬，如更新銅線網路為寬頻的光纖網路及採用更先進的通訊技術。無論從短、中、長期來看，視訊編解碼技術都是不可

或缺的關鍵技術，它的進步牽動著視訊通訊的發展與消費者的購買意願。

視訊編解碼技術的進步降低視訊通訊的成本，而促進消費者的購買。消費者的需求又刺激視訊編解碼技術的推陳出新。如此發展，必會帶引通訊進入「有聲有色」的時代，國內廠商應緊隨潮流趨勢，提昇自己的技術層次，才能在適當的時機切入市場掌握先機。

(本小節作者為通訊系統組林嘉文課長)