

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

-----o0o-----



BÀI TẬP LỚN

Môn: XỬ LÝ DỮ LIỆU ĐA PHƯƠNG TIỆN

Đề tài: Mã hoá dải tần và các ứng dụng

Giáo viên hướng dẫn: PGS. TS. Nguyễn Thị Hoàng Lan

Sinh viên thực hiện: Đoàn Hải Đăng 20070735

Trương Minh Đức 20070884

Trần Trung Hiếu 20071129

Đặng Thu Thủy 20072815

Nguyễn Thị Thanh Tú 20073378

Lớp: Truyền thông Mạng K52

Mục lục

Mở đầu	2
I. Tổng quan về mã hoá dải tần.....	3
1. Tư tưởng chung.....	3
2. Quy trình mã hoá SBC	4
3. Một số bộ lọc sử dụng trong SBC	5
Quadrature Mirror Filter (QMF).....	5
Bộ lọc Johnston.....	6
II. Mã hoá dải tần qua phép biến đổi.....	7
1. Biến đổi DCT:.....	8
2. Biến đổi Wavelet:.....	8
III. Ứng dụng mã hoá dải tần trong nén video.....	10
IV. Ứng dụng mã hoá dải tần trong nén audio	13
V. Mã hoá dải tần trong nén audio: Chuẩn MPEG.....	23
1. Giới thiệu chung về chuẩn MPEG.....	23
2. Phân tích kỹ thuật SBC trong nén audio chuẩn MPEG	23
2.1. MPEG-1	23
2.2. MPEG-2	27
2.3. MPEG-4	31
Kết luận.....	33
Tài liệu tham khảo	34

Mở đầu

Truyền thông đa phương tiện là một phương thức truyền đạt thông tin, nội dung thông qua các dữ liệu đa phương tiện như hình ảnh, âm thanh, video,... Gắn liền với sự ra đời và phát triển của vô tuyến, radio, truyền hình và sau này là mạng internet, truyền thông đa phương tiện ngày càng phát triển, mang lại dịch vụ truyền thông chất lượng cao thể hiện ngay bởi chất lượng âm thanh, hình ảnh.

Truyền thông tín hiệu tương tự có đặc trưng là dữ liệu nhiều, đòi hỏi băng thông rất lớn để đảm bảo chất lượng âm thanh hình ảnh. Mặt khác, việc xử lý dữ liệu đối với tín hiệu tương tự là có thể nhưng rất phức tạp và tốn kém. Chính vì thế, truyền thông tín hiệu số ra đời đã giải quyết được các vấn đề của truyền thông tương tự, và đang trở thành hình thức truyền thông đa phương tiện phổ biến hiện nay. Truyền thông tín hiệu số cho phép truyền dữ liệu với cùng chất lượng nhưng lượng dữ liệu thấp hơn, mặt khác, cho phép nén dữ liệu để giảm băng thông yêu cầu, từ đó giảm chi phí truyền thông một cách đáng kể.

Với đặc thù của tín hiệu số, nhiều phương pháp nén ra đời với mục đích giảm kích thước dữ liệu, giảm băng thông và giảm chi phí truyền thông. Các phương pháp nén được xây dựng trên cơ sở lý thuyết xác suất, lý thuyết truyền tin và xử lý tín hiệu số, được hoàn thiện và ứng dụng trong các chuẩn nén ảnh, audio, video. Mặt khác, đây cũng chính là động lực thúc đẩy các nghiên cứu trong xử lý tín hiệu số, xử lý đa phân giải nhằm đưa ra các giải pháp mã hoá nén ngày một hiệu quả hơn. Một trong số những phương pháp mã hoá có nền tảng lâu đời nhất là mã hoá dải tần, hay mã hoá băng con (Subband coding – SBC), hiện đang được ứng dụng trong nén MPEG.

Vì lý do đó, nhóm chúng em chọn đề tài về mã hoá dải tần và ứng dụng của nó trong nén audio và video. Bài tập lớn này sẽ trình bày khái quát về mã hoá SBC, mã hoá bằng phép biến đổi, ứng dụng của SBC trong nén audio và video, sau cùng là nghiên cứu ứng dụng có SBC trong một chuẩn nén cụ thể, đó là chuẩn MPEG, đang được sử dụng rộng rãi trong truyền thông đa phương tiện.

I. Tổng quan về mã hoá dải tần

1. Tư tưởng chung

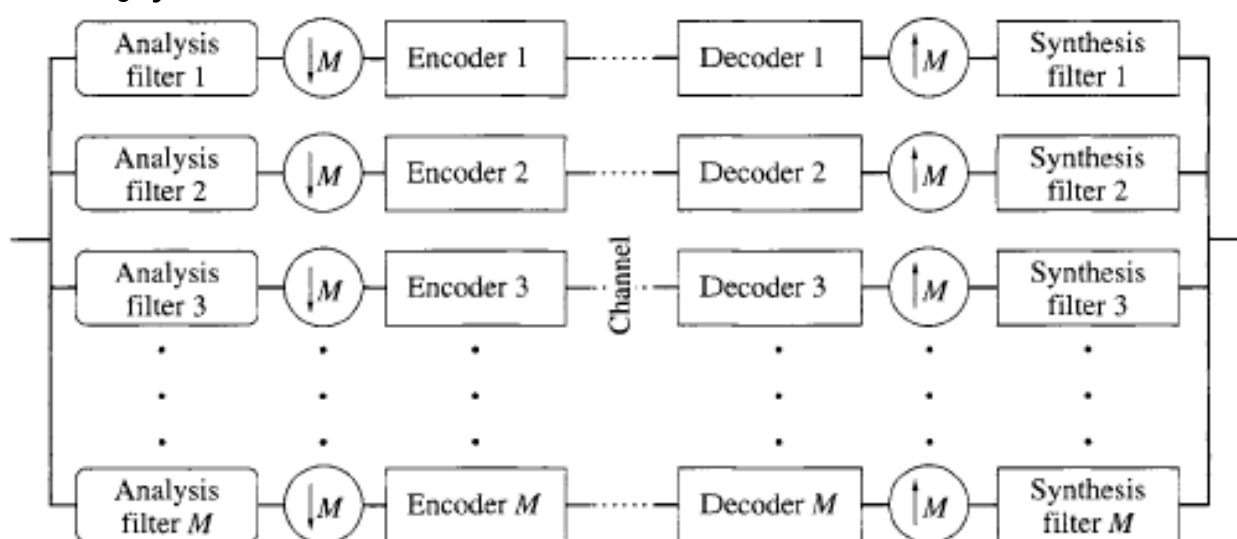
Có nhiều nguyên nhân và hướng tiếp cận dẫn đến việc sử dụng phương pháp mã hoá dải tần (Subband coding- SBC). Hiện nay có nhiều phương pháp mã hoá nén (lượng tử hoá và mã hoá) tín hiệu âm thanh, hình ảnh, dựa trên phân bố của tín hiệu theo một dạng nào đấy. Chẳng hạn như đối với tín hiệu có biên độ phân cụm, người ta hay sử dụng lượng tử hoá vector; đối với tín hiệu có biên độ chênh lệch nhau không nhiều, người ta dùng lượng tử hoá vi sai DPCM. Tuy nhiên dạng phân bố tín hiệu là hoàn toàn ngẫu nhiên, do đó khó có thể chọn được một phương pháp mã hoá nén tối ưu cho toàn bộ tín hiệu. Do đó, người ta đưa ra ý tưởng chia tín hiệu ra làm nhiều phần nhỏ, mã hoá mỗi phần theo một cách khác nhau để đạt được hiệu quả nén cao nhất.

Mặt khác, đối với âm thanh, thực nghiệm cho thấy tai người có một số hiệu ứng đối với âm thanh như hiệu ứng che tần số, che thời gian,... Do đó có thể tiết kiệm dữ liệu bằng cách loại đi các thành phần tín hiệu bị che, chỉ mã hoá những phần nghe thấy được. Hiệu ứng che ở tai người là khác nhau trên từng thành phần tần số, vì thế ta cần chia tín hiệu âm thanh thành các thành phần tần số con (dải băng con) rồi xác định các thành phần bị che và không bị che của tín hiệu trong từng dải tần số để xử lý.

Đó là tư tưởng cơ bản của mã hoá SBC. Mã hoá SBC chia tín hiệu gốc thành các thành phần tín hiệu thuộc các dải tần (sub-band) để xử lý và mã hoá riêng biệt từng thành phần sau đó gửi đi. Việc nhận tín hiệu, giải mã và khôi phục tín hiệu ban đầu được thực hiện theo trình tự ngược lại. Các dải băng con không chồng chập lên nhau (mà thực chất là liền kề nhau), do đó tín hiệu thu được ở các dải tần được giải tương quan, nhờ đó có thể mã hoá riêng biệt mà vẫn khôi phục được thành tín hiệu ban đầu.

Mã hoá SBC được ứng dụng rộng rãi trong nén dữ liệu audio, video, đặc biệt là trong chuẩn nén dữ liệu MPEG.

2. Quy trình mã hoá SBC



Mã hoá SBC gồm các bước sau:

- Lọc phân tích: tín hiệu được đưa qua các bộ lọc nhằm chia tín hiệu thành các thành phần tín hiệu ứng với các dải tần không chồng nhau. Đáp ứng xung của các bộ lọc sau bằng bộ lọc trước nhân với hàm số cos (để dịch đáp ứng đi một khoảng)
- Tín hiệu của các dải tần được giảm tần số lấy mẫu đi M lần (M là số dải tần chia tín hiệu). Việc giảm tần số lấy mẫu làm phổ rộng ra, trải hết trục tần số. Số mẫu lúc này cũng giảm theo. Giảm tần số lấy mẫu được thực hiện dựa theo định lý lấy mẫu Nyquist và hệ quả của định lý Nyquist. Tỷ lệ giảm tần số lấy mẫu phụ thuộc vào tỷ lệ giữa dải thông của bộ lọc với dải tần tín hiệu vào.
- Thực hiện lượng tử hoá và mã hoá riêng trên từng dải tần số. Lựa chọn phương pháp lượng tử hoá và mã hoá, thực hiện và phân phối bit dựa trên tính chất của dòng tín hiệu sau bộ lọc. Đây là bước mang lại hiệu quả nén cho toàn bộ quá trình mã hoá SBC.

Phân phối bit là việc phân chia số bit hiện có cho mã hoá từng dải, dựa theo tín hiệu của từng dải do trong từng dải, lượng thông tin là khác nhau. Phân phối bit ảnh hưởng lớn đến khôi phục dữ liệu sau quá trình gửi nhận dữ liệu.

Giải mã, tổng hợp SBC thực hiện theo trình tự ngược lại:

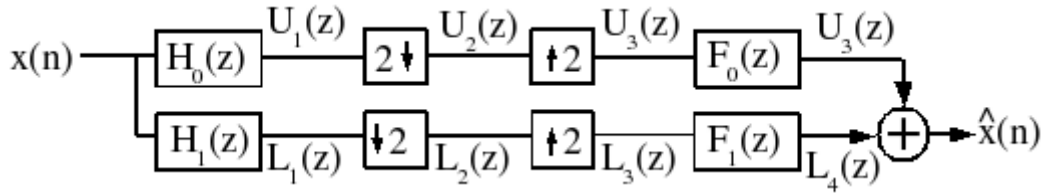
- Tín hiệu nhận được được phân kênh về các bộ xử lý với các dải tần tương ứng
- Giải mã tín hiệu, sử dụng các hệ số của quá trình lượng tử hoá và mã hoá ở trước.
- Tăng tần số lấy mẫu lên M lần
- Đi qua các bộ lọc để tín hiệu có xung giống như lúc sau khi đi qua bộ lọc phần mã hoá
- Ghép các tín hiệu lại thành tín hiệu ban đầu

3. Một số bộ lọc sử dụng trong SBC

Người ta thường sử dụng một số bộ lọc khác nhau cho băng lọc để lọc tần số thành các dải tần. Một số bộ lọc thường dùng có QMF, bộ lọc Johnson, bộ lọc Smith-Barnwell

3.1. Quadrature Mirror Filter (QMF)

Bộ lọc QMF thường được sử dụng để chia tín hiệu thành 2 dải hoặc 2^n dải có độ rộng bằng nhau. Sơ đồ mã hoá SBC sử dụng bộ lọc QMF như sau:



Trong đó $H_0(z)$ là bộ lọc thông thấp, $H_1(z)$ là bộ lọc thông cao. $F_0(z)$ và $F_1(z)$ là 2 bộ lọc tổng hợp để khôi phục tín hiệu ban đầu.

Hai bộ lọc phân tích của QMF thoả mãn điều kiện ràng buộc đối xứng:

$$H_0(z) = H_1(-z)$$

Từ đó, ta có tín hiệu thu được sau mỗi bộ lọc:

$$U_1(z) = H_0(z).X(z)$$

$$L_1(z) = H_1(z).X(z)$$

Tín hiệu thu được sau khi giảm tần số lấy mẫu:

$$U_2(z) = \frac{1}{2} \left(U_1(z^{1/2}) + U_1(-z^{1/2}) \right)$$

Thay U_1 và L_1 ở trên vào, ta thu được:

$$U_2(z) = \frac{1}{2} \left(H_0(z^{1/2}).X(z^{1/2}) + H_0(-z^{1/2}).X(-z^{1/2}) \right)$$

Tín hiệu sau khi qua bộ nội suy như sau:

$$U_3(z) = U_2(z^2) = \frac{1}{2} (X_k(z) + X_k(-z)) = \frac{1}{2} (H_k(z).X_k(z) + H_k(-z).X_k(-z))$$

Tín hiệu sau khi qua bộ lọc tổng hợp được khôi phục lại như sau:

$$\hat{X}(z) = F_0(z).U_3(z) + F_1(z).L_3(z)$$

$$\begin{aligned} \hat{X}(z) &= \frac{1}{2} (F_0(z).H_0(z) + F_1(z).H_1(z)).X(z) \\ &\quad + \frac{1}{2} (F_0(z).H_0(-z) + F_1(z).H_1(-z)).X(-z) \end{aligned}$$

Thành phần thứ 2 ứng với $X(-z)$ là thành phần nhiễu do chồng phổ. Để khử thành phần này, ta đặt hệ số của thành phần đó bằng 0, giải ra ta được điều kiện:

$$\begin{cases} F_0(z) = H_1(-z) \\ F_1(z) = -H_0(-z) \end{cases}$$

Người ta chứng minh được bộ lọc QMF để có thể tái tạo lại tín hiệu ban đầu sau quá trình lọc và giảm tần số lấy mẫu, phải có dạng như sau:

$$H_0(z) = c_0 z^{-2n_0} + c_1 z^{-(2n_1+1)}$$

$$H_1(z) = c_0 z^{-2n_0} - c_1 z^{-(2n_1+1)}$$

Một ví dụ của bộ lọc QMF là bộ lọc Haar với $c_0 = c_1 = 1, n_0 = n_1 = 0$:

$$H_0(z) = 1 + z^{-1}$$

$$H_1(z) = 1 - z^{-1}$$

3.2. Bộ lọc Johnston

Xét bộ lọc QMF thỏa mãn điều kiện triệt tiêu nhiễu do chồng phổ, khi này hàm truyền đạt của toàn hệ sẽ là:

$$H(z) = \frac{1}{2}(H_0(z)F_0(z) + H_1(z)F_1(z)) = \frac{1}{2}(H_0^2(z) - H_0^2(-z))$$

Hàm truyền đạt của hệ lúc này có thể dẫn đến sai lệch về pha và biên độ của tín hiệu so với tín hiệu gốc. Nếu $H_0(z)$ là FIR với pha là tuyến tính thì $H(z)$ cũng có pha tuyến tính, khi đó sai lệch pha sẽ bị triệt tiêu. Để đạt được điều này, $H_k(z)$ phải đối xứng trên miền thời gian. Mặt khác, để giảm sai lệch về biên độ, người ta chứng minh được điều kiện:

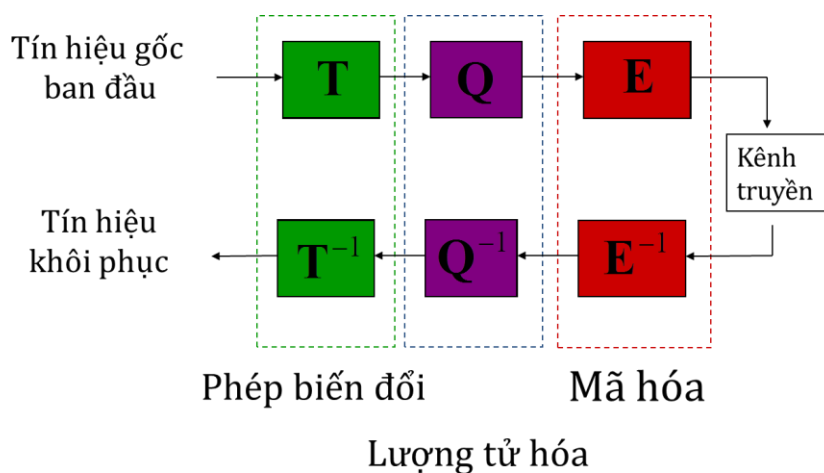
$$|H_0^2(e^{j\omega})|^2 + |H_1^2(e^{j\omega})|^2 \approx 1$$

Johnston năm 1980 đã đưa ra phương pháp thiết kế bộ lọc giảm thiểu sai số biên độ sử dụng các hàm tối ưu nhiều biến, điều kiện bình phương biên độ đáp ứng tần số của 2 bộ lọc gần xấp xỉ 1 trên toàn bộ trục tần số. Bộ lọc Johnston với 8 xung như sau:

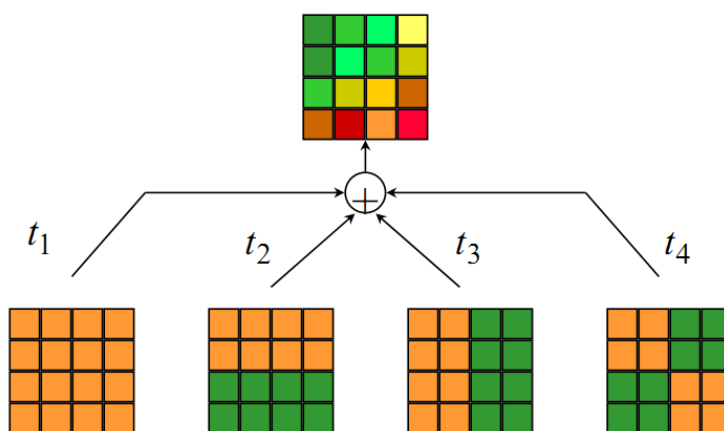
h0, h7	0.00938715
h1, h6	0.06942827
h2, h5	-0.07065183
h3, h4	0.48998080

II. Mã hoá dải tần qua phép biến đổi

Quá trình nén sử dụng phép biến đổi được thực hiện theo trình tự như sơ đồ sau:



Trong sơ đồ trên, phép biến đổi đóng một vai trò rất quan trọng, nhờ đó, tín hiệu ban đầu được biểu diễn hiệu quả hơn. Sau phép biến đổi, các tham số thu được đòi hỏi ít bit hơn để biểu diễn, do đó, giảm dung lượng truyền và lưu trữ. Các phép biến đổi thực hiện việc biểu diễn tín hiệu dưới dạng tổ hợp tuyến tính của các thành phần tín hiệu cơ bản khác.



Các phép biến đổi được sử dụng hiện nay có thể kể đến như:

- Dự đoán tuyến tính LPC
- Subband
- DCT
- DWT

1. Biến đổi DCT:

Ta xét tín hiệu $s(n)$ 1 chiều được biểu diễn dưới dạng 1 ma trận cột. Bằng phép biến đổi DCT, s có thể được biểu diễn dưới dạng tổ hợp tuyến tính của N vector cơ sở u .

$$s = \begin{bmatrix} s_0 \\ s_1 \\ \dots \\ s_{N-1} \end{bmatrix} \quad u = \begin{bmatrix} u_{k,0} \\ u_{k,1} \\ \dots \\ u_{k,N-1} \end{bmatrix} \quad t_k = (u_k, s) = \sum_{n=0}^{N-1} u_{k,n}^* s_n$$
$$s = t_0 u_0 + t_1 u_1 + \dots + t_{N-1} u_{N-1}$$

Công thức của phép biến đổi như sau:

$$\begin{cases} X[m] = \sqrt{\frac{2}{M}} K_m \sum_{n=0}^{M-1} x[n] \cos \left[\frac{(2n+1)m\pi}{2M} \right] \\ x[n] = \sqrt{\frac{2}{M}} K_n \sum_{m=0}^{M-1} X[m] \cos \left[\frac{(2m+1)n\pi}{2M} \right] \end{cases}$$
$$m, n = 0, 1, \dots, M-1$$
$$K_i = \begin{cases} \frac{1}{\sqrt{2}}, & i = 0 \\ 1, & i \neq 0 \end{cases}$$

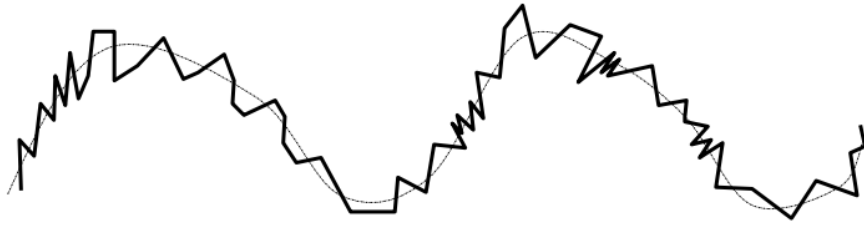
Như vậy, phép biến đổi DCT phân tích một tín hiệu thành tổ hợp tuyến tính của các thành phần tín hiệu cơ bản. Sau khi biến đổi DCT, ta thu được các hệ số tương ứng với mỗi thành phần tần số cơ bản. Từ công thức trên, ta thấy các thành phần đó chính là các thành phần tần số.

Phép biến đổi DCT cho ra hệ số của các thành phần tần số cơ bản, trong khi đó Subband coding cho ra tín hiệu của các thành phần tần số khác nhau. Trong quy trình nén JPEG, ảnh được chia làm các khối 8×8 , sau đó biến đổi DCT, lượng tử hoá và mã hoá. Về mặt hiệu quả tính toán, việc biến đổi DCT trên từng khối tương đương với việc nhân chập toàn bộ ảnh với các thành phần DCT cơ bản rồi giảm tần số lấy mẫu bằng kích thước của thành phần cơ bản. Nói cách khác, mã hoá dùng phép biến đổi DCT cũng được coi là một phương pháp SBC.

2. Biến đổi Wavelet:

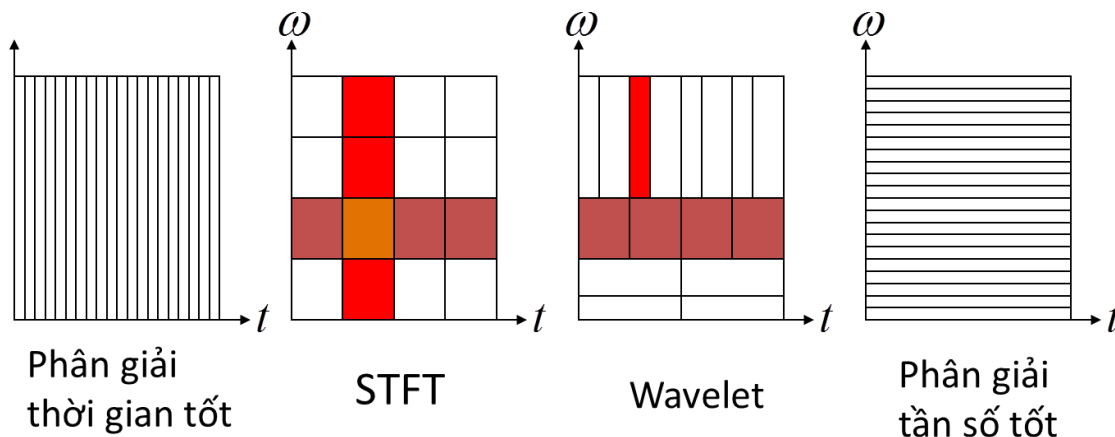
Bất kì tín hiệu nào cũng có thể được xấp xỉ bởi 2 thành phần:

- Phần thô (thành phần tần số thấp)
- Phần chi tiết (thành phần tần số cao)



Tại một thời điểm, tín hiệu có thể được biểu diễn dưới dạng tổng của thành phần tín hiệu thô và nhiều thành phần tín hiệu chi tiết. Tuy nhiên, phép biến đổi Fourier thường dùng có một hạn chế là làm mất đi hoàn toàn thông tin về miền thời gian. Sử dụng biến đổi Fourier, ta không thể biết được tại một thời điểm sẽ xuất hiện những thành phần tần số nào. Để khắc phục nhược điểm này, các nhà khoa học sử dụng biến đổi Fourier thời gian ngắn (STFT). Theo đó, tín hiệu được chia thành các khoảng nhỏ và được biến đổi Fourier trong từng khoảng đó. Để thực hiện kỹ thuật này cần chọn một hàm cửa sổ w sao cho độ dài của cửa sổ đúng bằng các khoảng tín hiệu phân chia. Tuy nhiên phương pháp này lại có một nhược điểm là nếu hàm cửa sổ càng hẹp thì độ phân giải thời gian càng tốt nhưng phân giải tần số càng kém, và ngược lại.

Để khắc phục cả 2 phương pháp trên, biến đổi Wavelet ra đời. Biến đổi Wavelet được thực hiện như sau: tín hiệu được nhân với hàm Wavelet (tương tự như nhân với hàm cửa sổ trong biến đổi STFT), rồi thực hiện biến đổi riêng rẽ cho các khoảng tín hiệu khác nhau trong miền thời gian tại các tần số khác nhau. Cách tiếp cận như vậy còn được gọi là: phân tích đa phân giải.



Phép biến đổi Wavelet rời rạc sử dụng công thức:

$$f(x) = \sum_k c_{j_0}(k) \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_j(k) \psi_{j,k}(x)$$

$$c_{j_0}(k) = \langle f(x), \tilde{\phi}_{j_0,k}(x) \rangle = \int f(x) \tilde{\phi}_{j_0,k}(x) dx$$

$$d_j(k) = \langle f(x), \tilde{\psi}_{j,k}(x) \rangle = \int f(x) \tilde{\psi}_{j,k}(x) dx$$

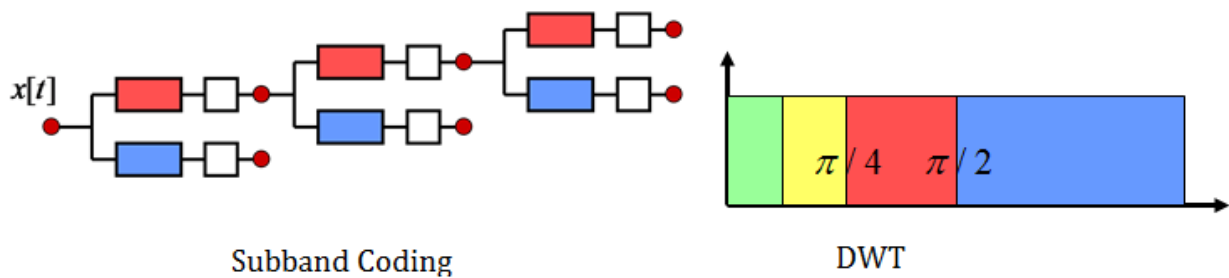
Trong đó:

$c_{j_0}(k)$ là hệ số xấp xỉ (hệ số co giãn)

j_0 là điểm bất kì bắt đầu co giãn

$d_j(k)$ là hệ số chi tiết hay hệ số wavelet.

Việc tính toán các hệ số của phép biến đổi Wavelet có thể được thực hiện nhờ Subband coding:



Từ đó ta có thể thấy mã hoá sử dụng DWT tương đương với biến đổi SBC, có thể thực hiện được bằng SBC.

III. Ứng dụng mã hoá dải tần trong nén video

Video là một chuỗi các frame ảnh (gọi là khung hình), xuất hiện liên tiếp tạo cảm thụ chuyển động theo thời gian. Quan hệ thời gian giữa các frame biểu diễn ảnh động.

Nén video sử dụng 2 kỹ thuật cơ bản, đó là Interframe và Intraframe :

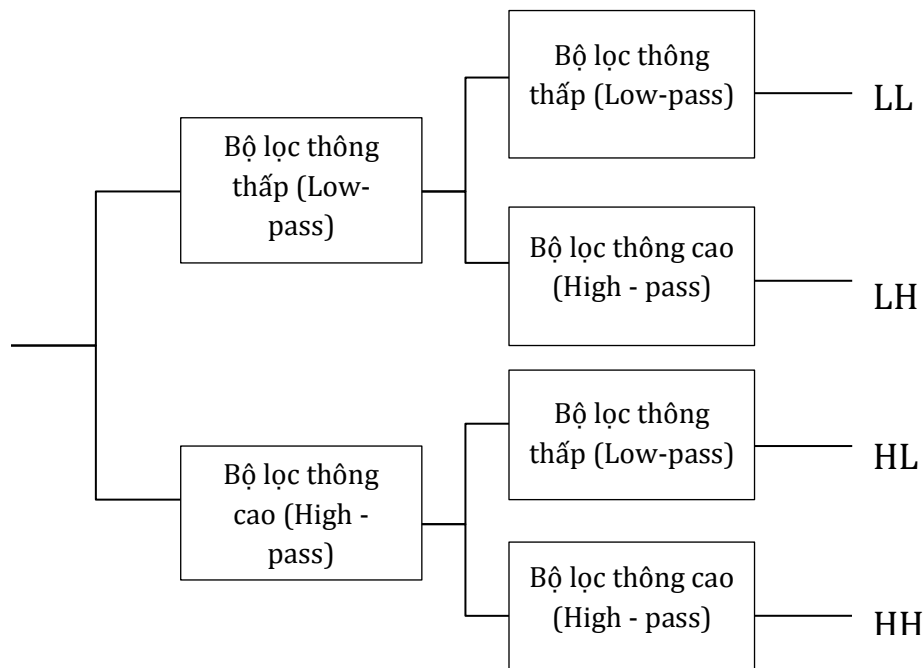
- Mã hóa Intraframe là kỹ thuật nén ảnh tĩnh giảm độ dư thừa về không gian
- Mã hóa Interframe là kỹ thuật đánh giá chuyển động và mã hóa để giảm độ dư thừa về thời gian

Đối với mã hóa Intraframe, mỗi frame (Khung hình) trong video được nén như ảnh tĩnh JPEG với điều kiện đảm bảo tốc độ khung hình. Vì vậy, Subband coding áp dụng trong nén video bản chất chính là Subband coding áp dụng đối với ảnh gốc.

Tư tưởng cơ bản của mã hóa SBC là chia tín hiệu gốc thành các thành phần tín hiệu thuộc các dải tần để xử lý và mã hóa riêng biệt từng thành phần nhằm làm tăng hiệu quả nén. Tư tưởng này được áp dụng trong nén ảnh như sau :

Dữ liệu ảnh được cho đi qua một băng lọc (filter bank) được gọi là băng lọc phân tích nhằm chia tín hiệu thành các thành phần ứng với các dải tần không chồng nhau.

Do dữ liệu ảnh bao gồm 2 thành phần riêng biệt : thành phần thô chứa phần lớn dữ liệu ảnh thường nằm ở thành phần tần số thấp và thành phần biên chứa thông tin về đường nét ảnh thường nằm ở thành phần tần số cao. Vì vậy trong nén ảnh thông thường, người ta sử dụng một bộ lọc thông thấp và một bộ lọc thông cao để tiến hành chia dải tần. Do sau khi đi qua hai bộ lọc, dữ liệu ảnh được chia làm 2 dải tần nên theo nguyên tắc chung của SBC, mỗi dải tần được giảm tần số lấy mẫu đi 2 lần.



Giả sử ảnh lúc đầu có kích thước $N \times N$. Trước hết chúng ta áp dụng bộ lọc theo chiều ngang cho từng dòng của dữ liệu ảnh với một bộ lọc thông thấp và một bộ lọc thông cao. Sau đó giảm tần số lấy mẫu ở đầu ra của 2 bộ lọc đi 2 lần, ta được 2 mới có kích thước $(N \times N)/2$. Tiếp theo đó, chúng ta áp dụng 2 bộ lọc thông thấp và thông cao theo chiều dọc của 2 ảnh mới. Ở đầu ra của 2 bộ lọc tiếp tục giảm tần số lấy mẫu đi 2 lần, ta thu được 4 ảnh có kích thước $N/2 \times N/2$.

Sau khi đi qua băng lọc, ta thu được 4 ảnh với ký hiệu như sau :

LL	LH
HL	HH

Trong đó :

- LL : Thành phần sau khi đi qua 2 bộ lọc thông thấp, chứa phần lớn dữ liệu ảnh gốc. Đây chính là thành phần thô của ảnh.
- LH : Thành phần đi qua bộ lọc thông thấp rồi đến bộ lọc thông cao
- HL : Thành phần đi qua bộ lọc thông cao rồi đến bộ lọc thông thấp, chứa phần lớn dữ liệu về biên của ảnh gốc
- HH : Thành phần sau khi đi qua 2 bộ lọc thông cao, chứa rất ít dữ liệu ảnh gốc

Sau khi thu được 4 thành phần, do 2 thành phần LH và HH chứa ít dữ liệu ảnh gốc trong khi thành phần LL và HL chứa phần lớn dữ liệu ảnh gốc và biên của nó nên ở các bước sau đó, ta có thể loại bỏ 2 thành phần LH và HH, chỉ xử lý các thành phần LL và HL.

Thành phần LL, chứa phần lớn dữ liệu ảnh gốc. Do thành phần này chứa những dữ liệu ảnh đã đi qua 2 bộ lọc thông thấp nên độ chênh lệch giữa các mẫu tín hiệu trong thành phần này không cao, vì vậy thành phần LL thường được mã hóa bằng phương pháp mã hóa dự đoán DPCM. Trong khi đó thành phần HL được lượng tử hóa bằng phương pháp lượng tử hóa vô hướng (Scalar Quantization). Đây là giai đoạn quyết định hiệu quả nén cho toàn bộ quá trình.

Nén ảnh sử dụng SBC đem lại hiệu quả nén cao do dữ liệu ảnh được phân ra làm các dải tần và chỉ lựa chọn một số dải tần chứa nhiều thông tin trong các dải tần đó để áp dụng các phương pháp mã hóa và lượng tử hóa phù hợp.

Từ hiệu quả của áp dụng SBC trong nén ảnh, người ta cũng áp dụng kỹ thuật này mã hóa Intraframe. Do dữ liệu của từng frame ảnh được nén với tỉ số nén cao hơn nên tổng thể một khung hình cũng được nén với hiệu quả tốt hơn các biện pháp nén ảnh thông thường.

IV. Ứng dụng mã hoá dải tần trong nén audio

1. Tổng quan

Âm thanh có bản chất là tín hiệu ; dựa trên dải tần của tín hiệu âm thanh được chia ra làm hai loại :

- Âm thanh dải tần cơ sở : âm thanh tiếng nói thoại có dải tần số từ 300Hz – 4KHz
- Âm thanh dải rộng : tiếng nói trình diễn, hát, âm nhạc... có dải tần số từ 100Hz – 20KHz

Phương pháp mã hóa các dải tần subband coding được sử dụng chủ yếu trong mã hóa âm thanh dải rộng do phương pháp này phù hợp với mô hình cảm thụ âm thanh hệ thống thính giác của con người. Các thông số cơ bản của âm thanh dải rộng

- Giới hạn dải phổ tín hiệu : 20KHz
- Tần số lấy mẫu $F_s = 44.1 \text{ KHz}$ (hoặc 48KHz)
- Lượng tử hóa giá trị : mã hóa bởi 16 bits, 20 bits
- Tốc độ stereo cần thiết 176Kbyte/s

2. Mô hình cảm thụ âm thanh hệ thống thính giác của con người

2.1. Phân chia dải tần

Tai con người có thể nghe được tần số trong dải từ 20Hz – 20KHz. Tuy nhiên tai không tiếp nhận tất cả các tần số của tín hiệu tiếng nói như nhau. Tai tiếp nhận các âm thanh có tần số nhỏ dễ dàng hơn các âm thanh có tần số lớn. Ví dụ tai ta có thể nghe thấy các âm thanh với tần số 1KHz có biên độ lớn hơn 5 dB trong khi đó với một âm có tần số 14KHz chỉ khi biên độ của âm thanh này khoảng 40dB tai ta mới có thể nghe thấy được. Do sự cảm nhận khác nhau của tai con người với các tần số âm thanh. Tai có thể được coi như 1 băng lọc với 25 dải dải tần có tần số từ 20Hz – 20KHz. Các dải tần số này không phân bố đều (độ rộng của các dải là khác nhau) và không liên tục (các dải này có phần chồng lên nhau). Tai chúng ta không thể phân biệt được các âm thanh ở trong cùng 1 dải khi chúng xuất hiện đồng thời. Mỗi dải này được gọi là 1 critical band (dải tần giới hạn). Độ rộng của mỗi dải tần :

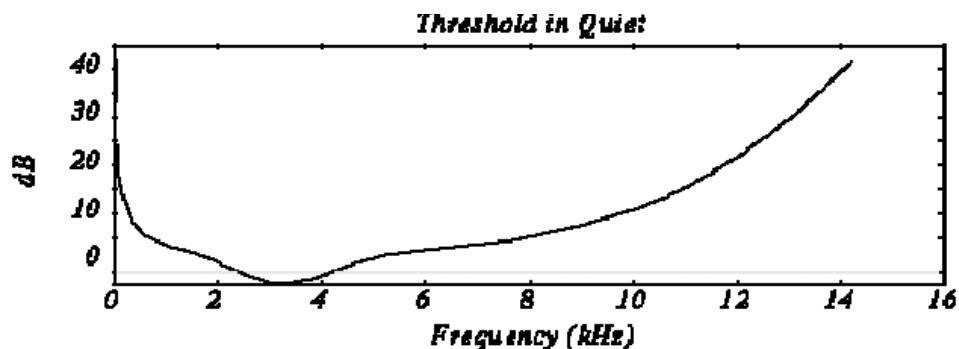
- 100 Hz với các dải tần có tần số nhỏ hơn 500Hz
- Tăng dần từ 500 tới 5000Hz với các dải tần còn lại

2.2. Các hiệu ứng che

Hiệu ứng che là hiện tượng khi 1 hoặc 1 số âm thanh tuy tồn tại trong không gian (vẫn nằm trong dải tần số mà con người cảm nhận được) nhưng tai ta không nghe thấy do 1 số lý do nhất định (biên độ nhỏ hoặc bị âm thanh khác lấn át) hiện tượng này được gọi là che hoàn toàn. Một trường hợp khác của hiệu ứng che là che một phần khi đó âm thanh bị che không biến mất hoàn toàn mà vẫn có thể cảm nhận được tuy nhiên với một biên độ thấp hơn (âm thanh trở nên bé hơn). Các giá trị của của hiệu ứng che chủ yếu được xác định dựa trên thực nghiệm và có giá trị không giống nhau đối với từng người. Có 3 hiệu ứng che là :

- Ngưỡng nghe (Threshold in Quiet)
- Che về mặt tần số (Frequency masking)
- Che về mặt thời gian (Temporal masking)

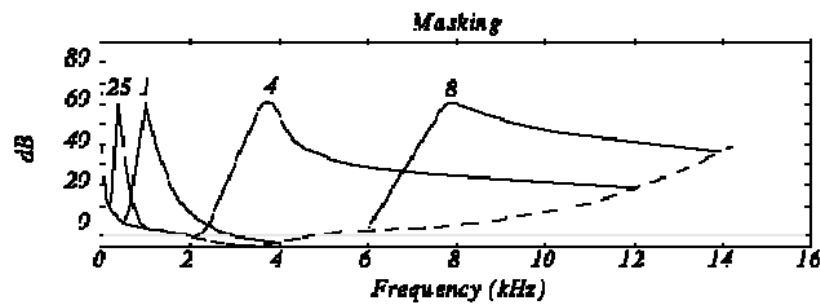
Ngưỡng nghe : Tai con người có cảm nhận không giống nhau đối với các dải tần số do đó với 1 số tần số ta có thể nghe rất rõ ràng chỉ với âm lượng nhỏ (biên độ tín hiệu nhỏ) nhưng với 1 số tần số âm thanh ta chỉ có thể nghe thấy khi chúng có biên độ khá lớn.



Như ta thấy tai nhạy cảm nhất với vùng tần số từ 2 – 4KHz

Frequency Masking : Là hiện tượng khi ta có 2 hay nhiều tín hiệu âm thanh đồng thời cùng xuất hiện với tần số và biên độ khác nhau nhưng ta chỉ nghe thấy một hoặc 1 số âm thanh còn phần còn lại thì không nghe được (Các âm thanh này đều có biên độ ở trên ngưỡng nghe). Ta nói âm thanh không nghe được bị âm thanh nghe được che đi.

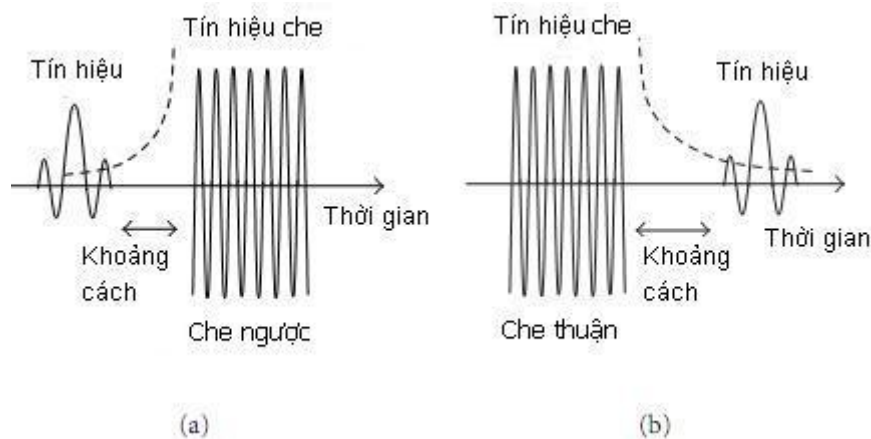
Thực nghiệm với tín hiệu âm thanh cho ta kết quả sau :



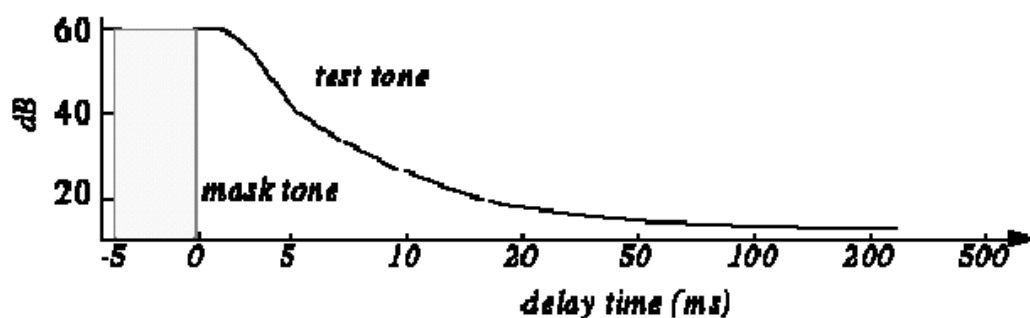
Qua biểu đồ trên ta thấy hiện tượng che khiến ngưỡng nghe được bị nâng lên cao hơn so với bình thường. Vì thế tất cả các tín hiệu ở dưới ngưỡng che cũng có thể được loại bỏ trong quá trình mã hóa.

Temporal masking : Với hiện tượng che về mặt tần số hai âm thanh cùng xuất hiện đồng thời thì với hiện tượng che về mặt thời gian hai âm thanh che và bị che có thể xuất hiện ở các thời điểm khác nhau. Hiện tượng che về mặt thời gian có thể chia ra làm hai loại :

- Che thuận : Tín hiệu che xuất hiện trước tín hiệu bị che. Với hiện tượng che thuận khi âm thanh che có cường độ lớn xuất hiện thì sau đó một lúc tai ta mới có thể nghe được âm thanh khác. Nếu âm thanh bị che xuất hiện trong khoảng thời gian này thì ta không thể nghe thấy được. Đây là dạng thường thấy của hiện tượng che về mặt thời gian. Hiện tượng này xuất hiện khi khoảng cách giữa 2 tín hiệu che và bị che là nhỏ hơn 200ms
- Che nghịch : Tín hiệu che xuất hiện sau tín hiệu bị che. Đây là hiện tượng khi âm thanh che sẽ che mất phần cuối của âm thanh trước đó được phát ra. Hiện tượng này ít xảy ra hơn và ngược chỉ có thể xảy ra khi mức độ của tín hiệu che cao hơn tương đối nhiều so với mức độ của tín hiệu âm thanh bị che phủ và khoảng cách thời gian giữa hai tín hiệu này là nhỏ hơn 25ms . Mặc dù tín hiệu âm thanh tới tai trước nhưng do não bộ lại xử lý tín hiệu che (có độ ồn cao) trước nên hiện tượng che ngược xuất hiện.



Một ví dụ cho hiện tượng che thuận với âm thanh che có tần số 1 kHz và biên độ 60dB. Âm thanh sử dụng để thí nghiệm(test tone) có tần số 1.1KHz. Giảm dần biên độ của test tone từ 40 dB xuống ta được kết quả sau



Bên cạnh đó hiện tượng che cũng xảy ra khi tín hiệu che tới một bên tai và tín hiệu bị che tới bên tai còn lại. Khi cường độ của tín hiệu che tăng đến một ngưỡng nào đó, nó có thể che phủ được tín hiệu âm thanh ở tai bên kia. Hiện tượng này gọi là hiện tượng che chênh lệch tuy nhiên hiện tượng này có hiệu quả thấp hơn nhiều so với việc cả hai tín hiệu tới cùng một bên tai (tín hiệu che cần phải có biên độ lớn hơn nhiều so với tín hiệu bị che). Trong trường hợp che chênh lệch các tín hiệu ở tần số cao lại có tác dụng che phủ lớn hơn nhiều so với các tín hiệu ở tần số thấp.

3. Subband coding trong mã hóa audio

3.1. Ý tưởng cơ bản

SBC trong mã hóa audio dựa trên mô hình cảm thụ âm thanh của con người. Như đã trình bày ở trên tai con người có độ nhạy cảm khác nhau với các tần số khác nhau. Do đó khi ta nghe một tín hiệu âm thanh với dải tần số rộng tai ta không thể nghe thấy các thành phần tín hiệu tại một số tần số nhất định với năng lượng nhỏ (nhỏ hơn ngưỡng nghe) và một số thành phần tín hiệu có năng lượng nhỏ có tần số gần với một thành phần tín hiệu khác có năng lượng lớn (Frequency masking và Temporal masking).

Ý tưởng cơ bản của SBC là tiết kiệm băng thông của tín hiệu bằng cách loại bỏ thông tin về các thành phần tần số bị che hoặc có năng lượng nhỏ mà tai không nghe được. Tín hiệu sau thực hiện quá trình sẽ khác với tín hiệu ban đầu nhưng tai không thể phân biệt được sự khác nhau giữa tín hiệu ban đầu và tín hiệu sau khi tiến hành đi qua SBC.

3.2. Các bước thực hiện SBC

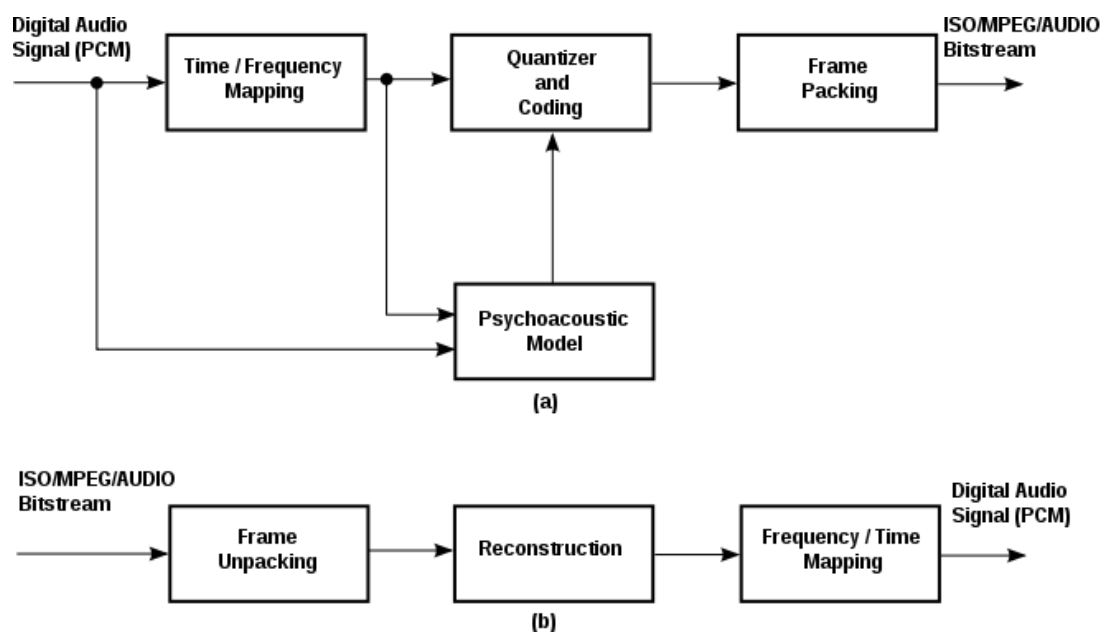
Phân chia một tín hiệu thành các dải tần riêng biệt sử dụng băng lọc như đã trình bày trong các phần trước. Sau khi thực hiện xong việc phân chia, ta có thể xử lý từng dải tần số đó một cách riêng rẽ. Các tín hiệu trong mỗi dải tần này được tiến hành phân tích năng lượng và tính toán các hiệu ứng che trong mỗi dải tần bởi tín hiệu trong các dải tần khác (Frequency masking) và tín hiệu trong các thời điểm khác (Temporal masking).

Sau khi tiến hành phân tích ta thu được năng lượng của các tín hiệu trong từng dải tần và một ngưỡng che (masking threshold) tương ứng cho từng dải tần. Dựa vào năng lượng và ngưỡng che của từng dải tần ta tiến hành lượng tử hóa và mã hóa một cách phù hợp :

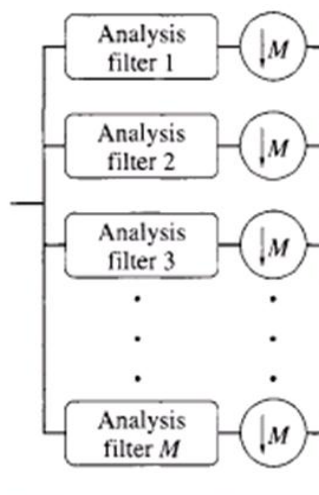
- Với các dải tần có các tín hiệu với mức năng lượng nhỏ hơn ngưỡng che tính toán được ở bước trên ta bỏ qua và không tiến hành lượng tử hóa lại và mã hóa.
- Các tín hiệu có năng lượng lớn hơn ngưỡng che được lượng tử hóa với bước lượng tử phụ thuộc vào ngưỡng che. Từ đó ta tính ra số bit cần sử dụng cho mỗi mẫu tín hiệu trong từng dải tần sao cho việc méo về mặt cảm thụ được giảm tối đa (Giữ cho nhiều lượng tử hóa nằm dưới ngưỡng che và không thể nhận ra bởi tai).
- Tín hiệu có cường độ càng cao thì con người sẽ càng cảm nhận được nó một cách rõ ràng, vì vậy nó sẽ được mã hóa bởi nhiều bit hơn so với những tín hiệu có cường độ thấp.

3.3. Sơ đồ thực hiện SBC

SBC có sơ đồ thực hiện chung như sau :



+ Khối Time/Frequency Mapping : Khối có đầu vào là tín hiệu gốc. Được sử dụng để chia tín hiệu đầu vào thành các subband. Khối này thường là băng lọc với nhiều bộ lọc với đáp ứng tần số khác nhau.



Sơ đồ trên thể hiện một băng lọc tổng quát với M bộ lọc phân tích. Sau khi đi qua băng lọc này ta thu được M dải tần tương ứng với từng bộ lọc. Nếu các dải tần này có độ rộng đều nhau thì với tín hiệu trên mỗi dải tần mà thu được ta có thể giảm tần số lấy mẫu đi M lần (Down Sampling $\downarrow M$). Ta có thể giảm như vậy vì theo định lý Nyquist mở rộng tần số lấy mẫu yêu cầu cho một tín hiệu :

$$F_s \geq 2(f_{max} - f_{min})$$

Hay nói cách khác tần số lấy mẫu của một tín hiệu cần phải lớn hơn hoặc bằng 2 lần độ rộng dải tần số của tín hiệu đó.

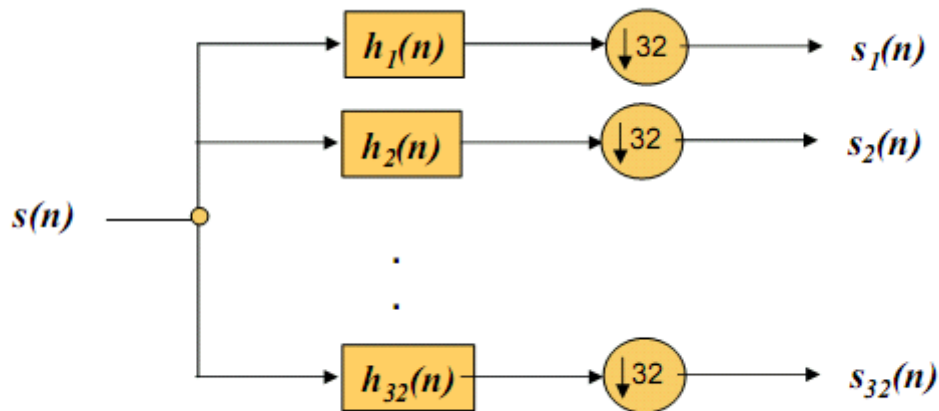
Có hai loại băng lọc để chia tín hiệu ban đầu ra thành các dải tần dựa vào độ rộng của các dải tần sau khi chia. Ta có thể chia tín hiệu ban đầu thành các dải tần có độ rộng đều nhau. Thường thì với loại băng lọc này tín hiệu ban đầu sẽ được đưa vào các bộ lọc của băng lọc cùng một lúc. Loại băng lọc này có thể áp dụng được cho mọi trường hợp do đó nó thường được áp dụng cho trường hợp ta không biết trước phân bố năng lượng của tín hiệu đầu vào.

Ta cũng có thể sử dụng băng lọc tạo ra các dải tần có độ rộng không đều nhau. Loại băng lọc này thường sử dụng một dãy các bộ lọc có cấu trúc cây. Độ rộng dải tần và mức độ down sampling của từng nhánh phụ thuộc vào độ sâu của nhánh trên cây. Khác với kiểu băng lọc trên với loại băng lọc này tín hiệu được đưa qua nhiều tầng bộ lọc với tín hiệu đầu ra của bộ lọc tầng trên là đầu vào của bộ lọc tầng dưới. Băng lọc này thường áp dụng cho các tín hiệu mà ta biết trước phân bố năng lượng. Do vậy loại băng lọc này có thể áp dụng khá tốt cho tín hiệu audio số vì thường thì hầu hết năng lượng của audio số đều nằm ở các tần số thấp hơn nhiều so với tần số lấy mẫu.

Tuy nhiên khác với tín hiệu ảnh với các miền tần số cao và thấp được phân chia khá rõ ràng với miền tần số cao là các cạnh và đường nét, miền tần số thấp là các vùng ảnh và sau khi đi qua băng lọc ta loại bỏ phần đi (Low-High và High-High) thì tín hiệu âm thanh không thể phân chia rõ ràng như vậy. Các âm thanh ở cả miền tần số thấp và cao với tín hiệu âm thanh đều có thể bị loại bỏ nếu có mức năng lượng nhỏ hơn ngưỡng che. Các âm thanh có tần số cao hoàn toàn có thể là nội dung của tín hiệu nếu có năng lượng lớn.

Số lượng dải tần được chia ra từ tín hiệu gốc và việc chia đều hay không đều phụ thuộc vào từng chuẩn mà ta sử dụng. Thường thì số dải tần được chia ra là lũy thừa của 2 để dễ tiến hành tính toán ; chia ra càng nhiều dải tần và chia các dải tần không đều nhau thì hiệu quả nén càng cao tuy nhiên chi phí tính toán sẽ tăng lên. Ví dụ với hệ thống mã hóa Dolby AC-2 ta chia ra làm 256 dải tần ; với ISO(MPEG) Layer 1,2 và MUSICAM là 32 dải tần với độ rộng đều nhau; với ISO Layer 3 là 576 subbands và độ rộng các dải tần không đều.

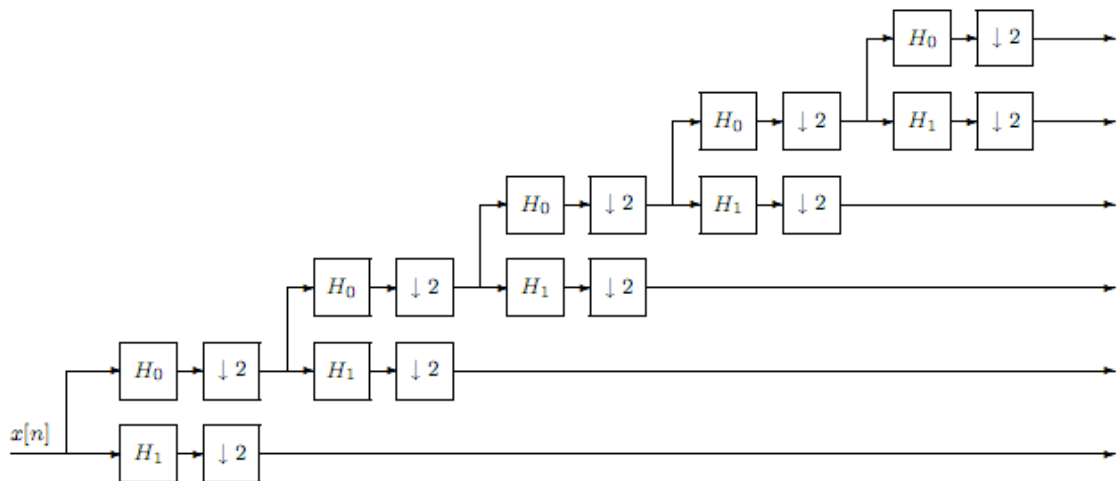
Một ví dụ về 2 loại băng lọc :



Sơ đồ trên thể hiện một băng lọc với 32 subbands có độ rộng bằng nhau sau khi cho tín hiệu gốc đi qua băng lọc này. Với $s(n)$ là tín hiệu đầu vào s_1 tới s_{32} là các subband tương ứng sau khi đi qua băng lọc. Tín hiệu trên mỗi subband có tần số lấy mẫu giảm 32 lần so với tín hiệu gốc $s(n)$. Với đáp ứng của từng bộ lọc :

$$h_{\text{sub}(k)}(n) = h(n) \cos \left[\frac{(2k+1)\pi n}{2M} + \varphi(k) \right] \text{ với } M = 32 ; k = 0, 1, \dots, 31,$$

Trong đó $h(n)$ là bộ lọc thông thấp



Sơ đồ trên thể hiện một băng lọc với 6 subband đầu ra sau khi cho tín hiệu gốc $x(n)$ qua băng lọc. Subband đầu tiên có độ rộng dải tần bằng $1/2$ độ rộng dải tần của tín hiệu đầu và có tần số lấy mẫu giảm 2 lần. Subband thứ 2 có độ rộng dải tần bằng $1/4$ và tần số lấy mẫu giảm 4 lần. Tương tự tới subband số 6 có độ rộng dải tần giảm đi 32 lần và tần số lấy mẫu cũng giảm 32 lần so với tín hiệu gốc. Với H_0 là bộ lọc thông thấp và H_1 là bộ lọc thông cao.

+ Khối Psychoacoustic Model : khối này có đầu vào là tín hiệu gốc và tín hiệu sau khi đã được chia thành các subbands. Khối được sử dụng để tính toán các hiệu ứng che của tín hiệu đưa vào dựa vào mô hình cảm thụ âm thanh của con người.

Tín hiệu được đưa vào khối này được tính toán để tìm ra các thành phần tần số có năng lượng nhỏ dưới ngưỡng nghe và các thành phần tần số bị che bởi các thành phần tần số trên các dải tần khác. Từ đó tính ra ngưỡng che cho mỗi dải tần và đưa tới khối lượng tử hóa để tiến hành lượng tử.

Độ rộng về thời gian của khối tín hiệu đưa vào khối Psychoacoustic Model để tiến hành phân tích càng lớn ta thu được hiệu quả nén càng cao (Temporal masking sẽ được tính toán tốt hơn). Masking threshold tính được chính xác hơn lượng thông tin không cần mã hóa sẽ cao hơn. Tuy nhiên việc này sẽ làm tăng độ trễ của quá trình mã hóa.

+ Khối lượng tử hóa và mã hóa (Quantizer and Coding) : Khối có đầu vào là các subband được chia sau khi tín hiệu gốc đi qua băng lọc và các hiệu ứng che được tính toán từ khối Psychoacoustic Model. Nhiệm vụ của khối này là tiến hành lượng tử hóa tín hiệu trên các subbands một cách phù hợp sao cho số bit sử dụng cho mỗi subband là ít nhất (tất nhiên vẫn cần đảm bảo chất lượng tín hiệu) và sau đó mã hóa tín hiệu trên các subband với các phương pháp mã hóa như Huffman, mã hóa đại số ...

Ta có sai số lượng tử với trường hợp lượng tử đều và bước lượng tử là Q gần như xấp xỉ với phân bố đều trong khoảng $(-Q/2, Q/2)$ với phương sai là $Q^2/12$ (Đây là nhiễu gây ra do lượng tử). Giả sử tín hiệu gốc có phân phối đều trên khoảng B . Với R bits/mẫu ta sẽ có 2^R mức lượng tử có thể sử dụng. Mối liên hệ giữa bước lượng tử Q và R : $Q = B/(2^R)$

Nhiều do lượng tử hóa sẽ giảm đi 6 dB với việc ta sử dụng thêm một bit cho việc lượng tử :

$$\begin{aligned} \text{SNR} &= 10 \log_{10} \frac{\sigma_f^2}{\sigma_q^2} \\ \sigma_q^2 &= \frac{Q^2}{12} = (B/2^R)^2/12 = B^2/12 * 2^{-2R} \\ \text{SNR}(R+1) - \text{SNR}(R) &= 10 \log_{10} \frac{\sigma_q^2(R)}{\sigma_q^2(R+1)} \\ &= 10 \log_{10} \frac{2^{-2(R)}}{2^{-2(R+1)}} = 10 \log_{10} 2^2 = 6 \text{ dB} \end{aligned}$$

Như ta đã biết mỗi subband có ngưỡng che riêng . Khối lượng tử hóa làm nhiệm vụ lượng tử tín hiệu của mỗi subband với đủ số bit để duy trì khoảng cách giữa nhiễu do lượng tử và ngưỡng che của mỗi subband (Giữ cho nhiễu lượng tử nhỏ hơn ngưỡng che của subband để các thành phần nhiễu này không thể nghe thấy được bởi tai con người). Như vậy ta có thể giảm được số bit cần để lượng tử hóa mỗi mẫu tín hiệu của subband tùy thuộc vào ngưỡng che của subband tương ứng nhờ vậy ta có thể tăng được hiệu quả nén dữ liệu (Việc giảm mỗi bit làm tăng nhiễu lượng tử lên 6 dB ta cần giữ cho giá trị nhiễu này nhỏ hơn ngưỡng che).

Bên cạnh đó, ta cũng thấy được một trong những giới hạn của mã hóa tín hiệu âm thanh là khi nhiễu do lượng tử hóa vượt quá ngưỡng che và trở thành một phần của tín hiệu âm thanh mà tai có thể nghe được. Ta luôn phải cố gắng giảm số bit sử dụng để lượng tử hóa giá trị của mẫu sao cho giá trị của nhiễu do lượng tử xấp xỉ với giá trị của ngưỡng che nhưng luôn phải nhỏ hơn giá trị của ngưỡng che.

Ví dụ : Ta có 16 trong 32 dải tần như sau

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level (db)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

Với dải tần số 8 có giá trị biên độ 60dB giả sử rằng dải tần này gây ra hiệu ứng che cho 2 dải tần bên cạnh nó : 12 dB với dải tần số 7 và 15 dB với dải tần số 9. Ta thấy giá trị biên độ của dải tần số 7 là 10 dB < 12 dB của ngưỡng che. Do đó ta bỏ qua dải tần số 7 mà không cần mã hóa.

Giá trị biên độ của dải tần số 9 là 35 dB > 15 dB của ngưỡng che do đó ta tiến hành mã hóa và gửi đi. Giả sử tín hiệu ban đầu đưa vào được biểu diễn với 8 bits/mẫu Với giá trị ngưỡng che là 15 dB ta có thể giảm số bit sử dụng cho 1 mẫu đi 2 bit (Sai số lượng tử sẽ tăng lên 12 dB tuy nhiên vẫn nhỏ hơn ngưỡng che do vậy kết quả vẫn có thể chấp nhận được).

Sau khi tiến hành lượng tử hóa với số bit tương ứng cho từng subband ta tiến hành encode cho từng subband bằng các phương pháp mã hóa đã nêu.

+ Khối Frame Packing : Dữ liệu sau khi ta lượng tử và mã hóa được đưa vào khối này để đóng gói thành các frame và chuyển đi trong hệ thống mạng.

+ Các khối trong quá trình giải mã tại bên nhận : Tại bên nhận ta làm ngược lại với quá trình mã hóa tuy nhiên việc giải mã với SBC đơn giản hơn việc mã hóa vì ta không cần tới khối Psychoacoustic Model. Ta unpacking frame, giải mã và đưa tín hiệu của các subband vào băng lọc để tổng hợp lại tín hiệu ban đầu.

V. Mã hoá dải tần trong nén audio: Chuẩn MPEG

1. Giới thiệu chung về chuẩn MPEG

MPEG là viết tắt của Motion Picture Experts Group. Đây là một hệ thống tiêu chuẩn cho nén video, audio và và đặc tả hệ thống được phát triển bởi tổ chức cùng tên – Hội Phim Ảnh Thế Giới.

Các chuẩn MPEG được sử dụng trong nén tín hiệu video và audio hiện nay gồm có:

- MPEG-1: Gồm chuẩn cho nén video và audio cho việc lưu trữ trên đĩa CD-ROM
- MPEG-2: Chuẩn này cho chất lượng video và audio tốt hơn với tỉ lệ nén cao, được sử dụng trong truyền hình kỹ thuật số vệ tinh, truyền hình cáp và định dạng đĩa DVD. Sau MPEG-2 có sự phát triển của MPEG-3. Tuy nhiên, do chuẩn này có nhiều điểm tương đồng với MPEG-2 nên đã bị xóa bỏ và chuẩn MPEG-2 tiếp tục được phát triển.
- MPEG-4: Phát triển từ MPEG-1 và nâng cao chất lượng hơn và giảm bit rate, phù hợp với internet audio video streaming.

2. Phân tích kỹ thuật SBC trong nén audio chuẩn MPEG

2.1. MPEG-1

a. Giới thiệu chung về MPEG-1

Chuẩn MPEG-1 là tiêu chuẩn cho nén tín hiệu có mất mát thông tin với tốc độ tổng cộng cho cả video và audio là 1.5Mbps. Trong đó, tín hiệu audio được nén từ 1.536 Mbps xuống còn 0.3Mbps.

Một số định dạng file sử dụng chuẩn MPEG-1 để nén tín hiệu là *.mpg, *.mpeg, *.mp1, *.mp2, *.mp3, *.m1v, *.m1a, *.m2a, *.mpa, *.mpv.

Âm thanh trong chuẩn MPEG-1 được truyền trên 2 kênh Left và Right (tín hiệu stereo). Tại mỗi kênh, tốc độ lấy mẫu là 48K mẫu/s và mỗi mẫu được lượng tử hóa bởi 16 bits. Như vậy, tốc độ bit cần thiết cho riêng tín hiệu audio là 1.536 Mbps. Với khả năng nén xuống còn 0.3 Mps, có thể thấy tỉ lệ nén đạt tới 5 lần.

b. MPEG-1 Audio Layers

Tín hiệu âm thanh trong MPEG-1 được chia làm 3 thành 3 loại tương ứng với 3 layer:

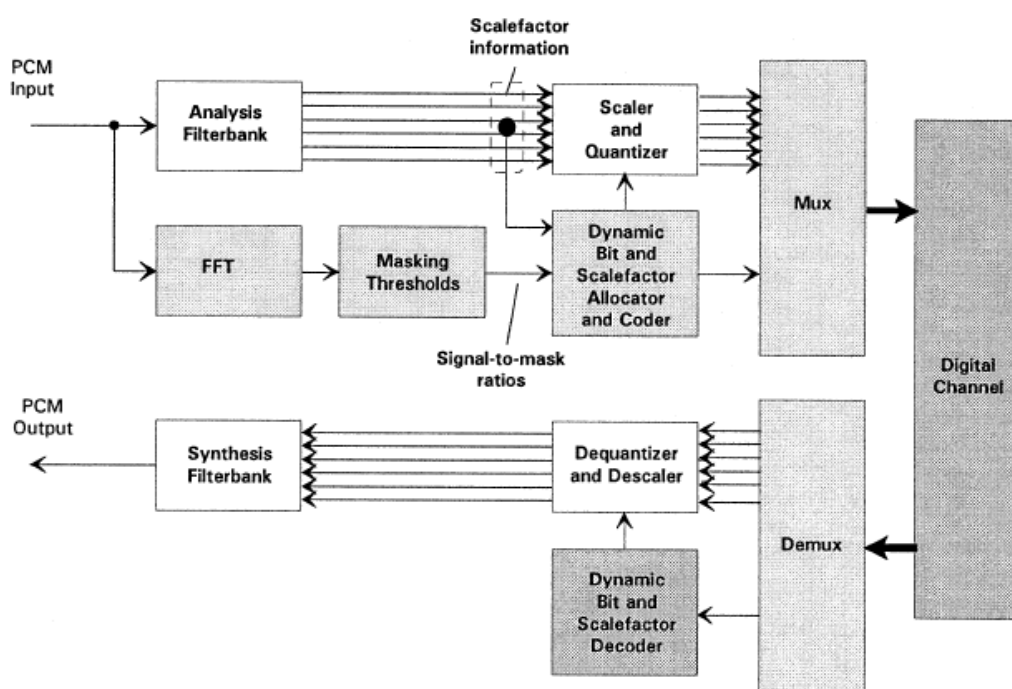
- Layer I
- Layer II
- Layer III

Các layer khác nhau ở độ phức tạp tính toán và hiệu năng. Layer III cần sự tính toán phức tạp hơn Layer I và II nhưng lại có hiệu quả nén tại tốc độ bit rate thấp cao hơn. Bên cạnh đó, các kỹ thuật nén cũng được tiếp cận theo các cách khác nhau để đạt được những hiệu quả nhất định tại từng layer. Phần tiếp theo sẽ trình bày chi tiết kỹ thuật SBC trong việc nén tín hiệu audio tại các layer.

Layer I và II

Layer I và II có nhiều đặc điểm giống nhau trong việc mã hóa và giải mã. Layer II có hiệu năng cao hơn một chút do khối lượng tử hóa có xử lý tốt hơn khối lượng tử hóa của Layer I.

Sơ đồ mã hóa và giải mã:



Băng lọc - Filterbank

Layer I & II chia tín hiệu audio thành 32 băng con khác nhau sử dụng kỹ thuật DCT (Discrete Cosine Transform – Biến đổi cosine rời rạc). Cách chia này được dựa trên mô hình cảm thụ âm thanh của con người: tai nghe được coi như một băng lọc gồm 25 dải băng con. Tuy nhiên để phù hợp hơn với các tính toán kỹ thuật số, tín hiệu được chia thành $2^5 = 32$ dải băng con thay vì 25.

Lượng tử hóa

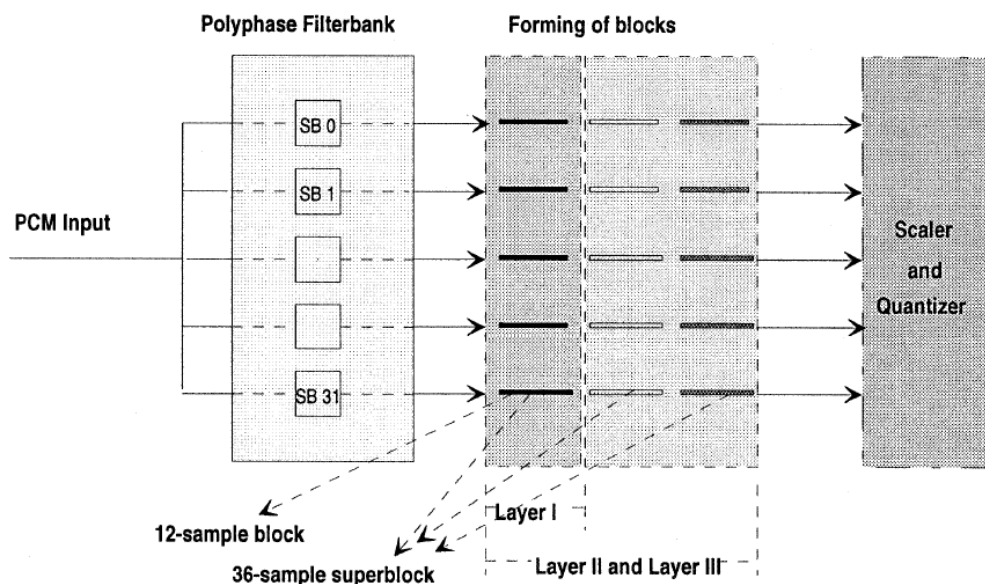
Số lượng bit dùng để lượng tử hóa cho từng băng con được quyết định dựa trên kỹ thuật phân bố bit động (dynamic bit allocation). Kỹ thuật này cũng dựa trên mô hình cảm thụ âm thanh của con người để đưa ra những tính toán phù hợp. Như đã trình bày ở phần trên, mô hình này đưa ra 3 ngưỡng che trong việc cảm thụ của

con người. Trong MPEG-1 Layer I và II, mức độ che (masking level) tại mỗi băng tần chỉ được tính toán dựa trên hiệu ứng che về mặt tần số (frequency masking).

Các mẫu sẽ được xử lý theo từng block gồm 12 mẫu. Kỹ thuật này được gọi là block companding. Tại Layer I, mỗi một khối sẽ được lượng tử hóa bằng 1 bit. Như vậy với 32 băng con sẽ có $32 \times 12 = 384$ mẫu được xử lý. Với tần số lấy mẫu là 48 KHz, 12 mẫu sẽ tương ứng với 8ms.

Tại Layer II, thay vì xử lý từng khối 12 mẫu, tín hiệu sẽ được xử lý theo superblock gồm 3 block liên tiếp, có độ dài là $8 \times 3 = 24$ ms. Mỗi superblock sẽ được lượng tử hóa bởi 1 bit. Như vậy tương ứng với 32 băng con sẽ có $32 \times 12 \times 3 = 1152$ mẫu được xử lý.

Tại Layer III, khối lượng tử hóa cũng thực hiện việc xử lý với các superblock.

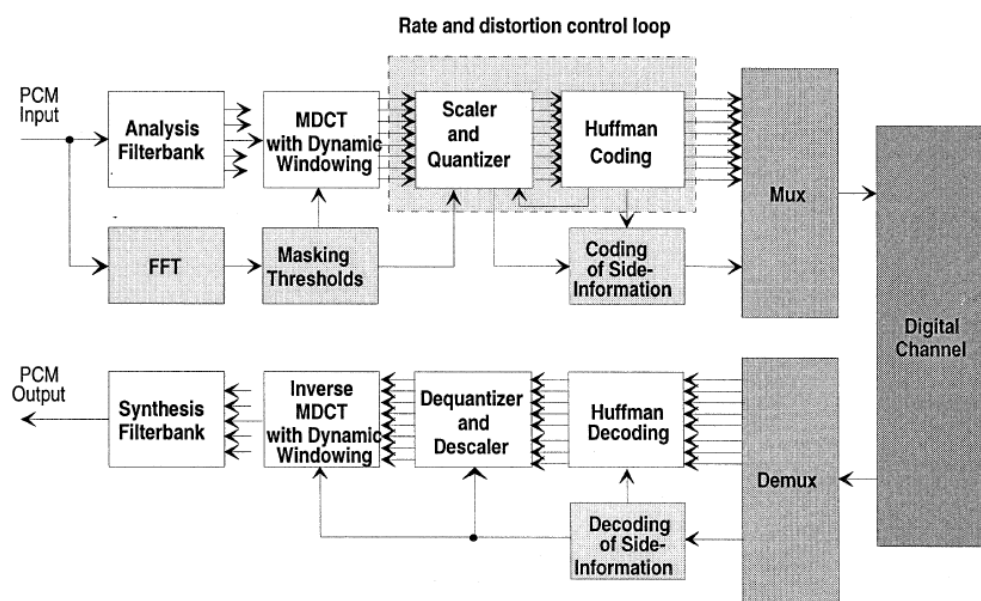


Mã hóa không được sử dụng trong Layer I và Layer II.

Layer III

Layer III của MPEG được biết đến với một tên gọi khác phổ biến hơn là MP3. Chuẩn này có thêm nhiều tính năng hơn so với Layer I và Layer II do đó có cho hiệu quả nén tốt hơn và cũng được sử dụng rộng rãi hơn.

Sơ đồ mã hóa và giải mã:



Bảng lọc – Switched Hybrid Filterbank

Để đạt độ phân giải tần số cao hơn, nghĩa là tính toán chính xác hơn với các critical band trong mô hình cảm thụ tiếng nói của con người, MPEG-1 Layer III tiếp tục chia nhỏ 32 băng tần con bằng khối DCT cải tiến (modified DCT – MDCT) 6-point hoặc 18-point với mức độ chồng lấp (overlap) giữa các băng tần là 50%. Như vậy số lượng băng tần con được chia ra tối đa là $32 \times 18 = 576$ băng con. Khối 18-point sẽ cho độ phân giải về mặt tần số tốt hơn và được sử dụng rộng rãi hơn trong khi khối 6-point cho độ phân giải về mặt thời gian tốt hơn. Do đó khối 6-point chỉ thường được sử dụng trong trường hợp âm thanh có tiếng vọng. Về mặt tín hiệu, tiếng vọng được coi là một số các bit có sự xuất hiện cao một cách tức thời.

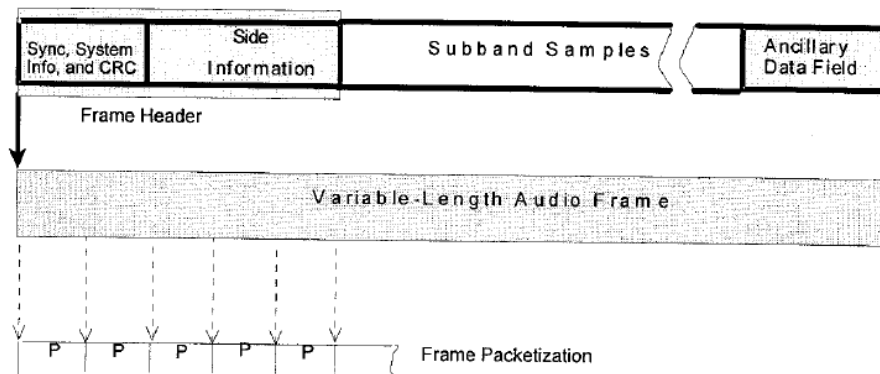
Lượng tử hóa và mã hóa

Các mẫu tại các băng con sau khi đi qua khối MDCT sẽ được lượng tử hóa khác nhau, do đó, sẽ sinh ra lỗi lượng tử nhỏ hơn. Như đã trình bày ở phần trên, 3 superblock tương ứng với 36 mẫu sẽ được lượng tử hóa bằng 1 bit.

Sau khi lượng tử hóa, các băng con sẽ được xử lý thêm bước mã hóa để tăng hiệu quả của việc nén. Giải thuật được sử dụng là Huffman.

Đóng gói gói tin

Dữ liệu sau khi được lượng tử hóa và mã hóa sẽ được đóng gói thành các frame. Cấu trúc của các frame như sau:



- **Frame Header:**
 - 12 bit đồng bộ (Sync)
 - 20 bit thông tin hệ thống (System Info)
 - 16-bit check lỗi CRC
 - Side Information: Thông tin phụ về số bit lượng tử
- Thông tin chính: các mẫu của các băng tần con, tương ứng với từng Layer: 32x12 cho Layer I (8ms) và 32x36 cho Layer II và III (24ms).
- Thông tin bổ sung (Ancillary Data Field): Các thông tin bổ sung, phụ thuộc tính chất của từng trường hợp sử dụng.

Trong MPEG-1 Layer I và II các frame được coi là “tự trị” (autonomous) – mỗi frame đã tự chứa các thông tin cần thiết để giải mã. Do vậy các frame có thể được giải mã một cách độc lập và được ghép lại với nhau nhờ các điểm đánh dấu. Độ dài của các frame là khác nhau do:

- Độ dài của trường thông tin chính khác nhau, phụ thuộc vào bit rate và tần số lấy mẫu
- Độ dài trường side information khác nhau
- Độ dài trường thông tin bổ sung khác nhau

2.2. MPEG-2

a. Giới thiệu chung về MPEG-2

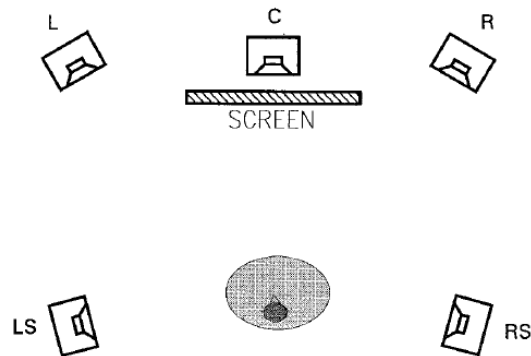
Chuẩn MPEG-2 cho chất lượng nén tốt hơn MPEG-1 và được sử dụng trong truyền hình kỹ thuật số vệ tinh, truyền hình cáp và định dạng đĩa DVD. Chuẩn MPEG-2 đưa ra 2 kỹ thuật mã hóa cải tiến dựa trên MPEG-1 là MPEG-2 BC có khả

năng tương thích ngược với MPEG-1 và MPEG-2 AAC không tương thích ngược với MPEG-1.

Tín hiệu âm thanh của chuẩn MPEG-2 thường được truyền trên nhiều kênh. Mô hình phổ biến nhất là mô hình 5.1 kênh với cấu hình 3/2. Cả 2 kỹ thuật nói trên đều hỗ trợ tốt việc mã hóa đa kênh (multichannel encoding).

Một số cấu hình đa kênh được sử dụng phổ biến:

- 1 channel 1/0-configuration: centre (mono)
- 2 channels 2/0-configuration: left, right (stereophonic)
- 3 channels 3/0-configuration: left, right, centre
- 4 channels: 3/1-configuration left, right, centre, mono-surround
- 5 channels: 3/2-configuration: left, right, centre, surround left, surround right



Mô hình thực tế cho Cấu hình 3/2

b. MPEG-2 BC và AAC

MPEG-2 BC

Chuẩn MPEG-2 Backward Compatible có sự tương thích ngược với MPEG-1 có nghĩa là một thiết bị giải mã thuộc chuẩn MPEG-1 có thể tái tạo lại âm thanh chuẩn MPEG-2 dưới cấu hình 2/0.

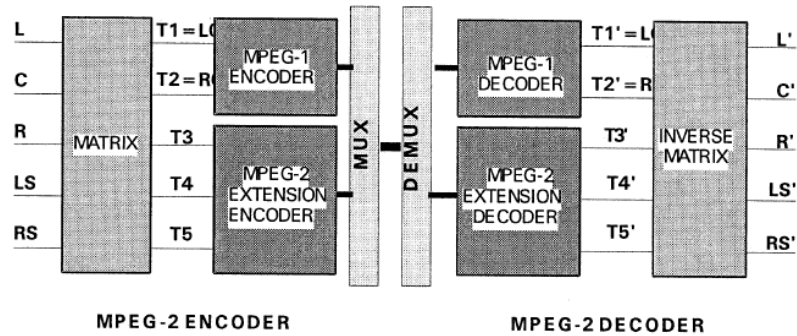
Trong quá trình mã hóa, tín hiệu trên 5 kênh truyền của MPEG-2 sẽ được đưa về 2 kênh truyền LO và RO tương ứng với kênh left và right của MPEG-1 dựa trên công thức sau:

$$LO = \frac{1}{1+\sqrt{2}}(L + \sqrt{2}C + \sqrt{2}LS)$$

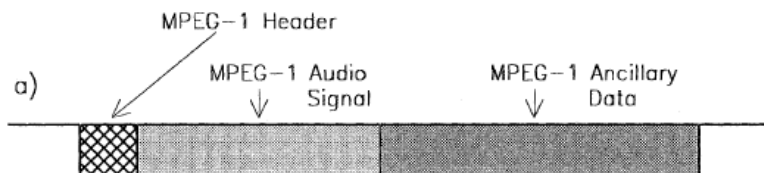
$$RO = \frac{1}{1+\sqrt{2}}(R + \sqrt{2}C + \sqrt{2}RS)$$

Ngoài ra, trong nhiều trường hợp, để đơn giản hơn, có thể coi $LO = L$ và $RO = R$.

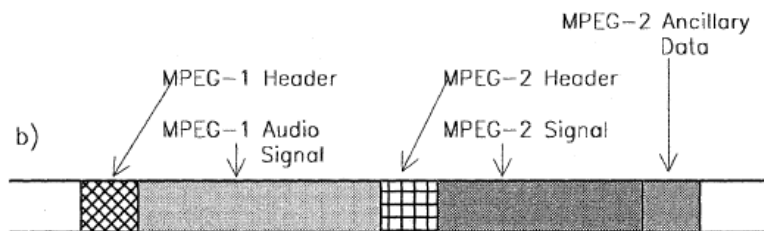
Tín hiệu trên LO và RO sẽ được truyền đi trên kênh T1 và T2. Các kênh T3, T4 và T5 sẽ tạo thành các kênh tín hiệu mở rộng có tác dụng tái tạo lại âm thanh đúng cấu hình 3/2 như trên. Trong quá trình mã hóa, sự dư thừa giữa các kênh và các ngưỡng che vẫn được tính toán đúng với nguyên tắc của phương pháp SBC cho từng kênh.



Tín hiệu trên các kênh LO và RO sẽ được mã hóa SBC theo chuẩn MPEG-1 và gửi đi dưới định dạng gói tin của chuẩn này trên kênh T1 và T2. Khi đó, các thông tin bổ sung chính là thông tin về các kênh tín hiệu mở rộng T3, T4 và T5. Với bộ giải mã của chuẩn MPEG-1, các trường thông tin này sẽ bị bỏ qua và tín hiệu âm thanh vẫn được khôi phục bình thường (tuy nhiên chất lượng sẽ kém hơn do chỉ còn là Cấu hình 2/0). Với bộ giải mã của chuẩn MPEG-2, tín hiệu âm thanh sẽ được khôi phục lại đúng với tín hiệu ban đầu. Có thể thấy rõ sự tương thích ngược này thông qua cấu trúc gói tin của MPEG-1 và MPEG-2.



MPEG-1 audio frame

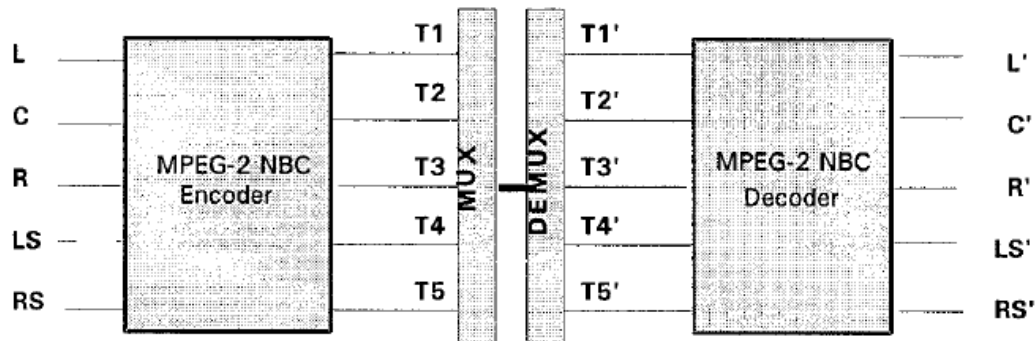


MPEG-2 audio frame

Việc phân chia 5 kênh truyền thành 2 nhóm MPEG-1 Encoder và MPEG-2 Extension Encoder như trên được gọi là matrixing. Khi giải mã, 2 nhóm encoder lại được chia về 5 kênh truyền – kỹ thuật này được gọi là dematrixing. Đây chính là sự giải thích cho tính tương thích ngược giữa chuẩn MPEG-2 BC và MPEG-1.

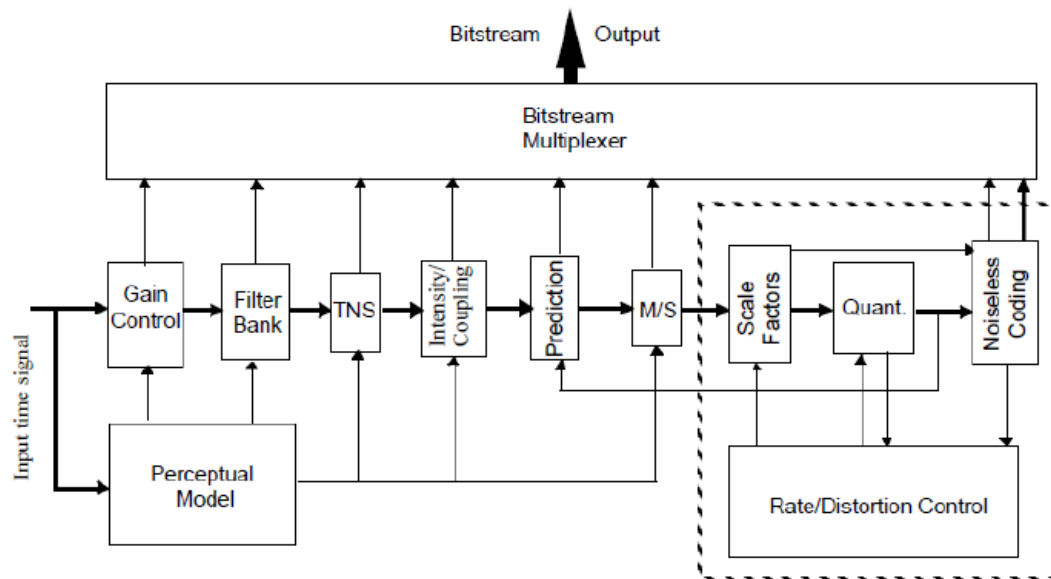
MPEG-2 AAC

Chuẩn MPEG-2 Advanced Audio Coding không hỗ trợ tính tương thích ngược với chuẩn MPEG-1. Chính vì vậy kỹ thuật matrixing và dematrixing không còn xuất hiện trong MPEG-2 AAC.



Chuẩn MPEG-2 AAC được cải tiến bằng cách module hóa các khối chức năng và phát triển thêm nhiều công cụ mã hóa, giúp cải thiện chất lượng audio ở bit rate thấp:

- **Optional Preprocessing:** Khối tiền xử lý
- **Filterbank:** MPEG-2AAC cũng sử dụng kỹ thuật MDCT với kích thước cửa sổ lớn hơn và cho tạo ra 1024 băng con giúp tăng độ phân giải tần số tốt hơn MP3 với 576 băng con.
- **TNS – Temporal Noise Shaping:** Dựa vào các hệ số của khối MDCT, kỹ thuật này cho phép tạo dạng nhiễu tạm thời, hỗ trợ việc cải thiện chất lượng tiếng nói ở bit rate thấp.
- **Psychoacoustic Modeling:** Vẫn dựa trên mô hình cảm thụ âm thanh của con người để sinh ra các ngưỡng che cho từng băng con.
- **Quantization and coding:** MPEG-2 AAC cũng sử dụng mã hóa Huffman sau khi lượng tử hóa
- **Optional Temporal Prediction:** Khối dự đoán này hỗ trợ tăng tỉ lệ nén bằng cách dự đoán các mẫu tín hiệu để định hướng khối lượng tử hóa tập trung vào những mẫu tín hiệu cần quan tâm



Bên cạnh việc module hóa các khối chức năng, để hỗ trợ nhiều mục đích sử dụng khác nhau, MPEG-2 AAC còn cho phép lựa chọn các profile khác nhau. Ba profile đã được chuẩn hóa là:

- Profile chính
- Profile với độ phức tạp thấp
- Profile với tốc độ lấy mẫu đa dạng

Với sự cải tiến về của các module, chuẩn MPEG-2 AAC tạo ra tín hiệu âm thanh với chất lượng tốt tại bit rate từ 320 đến 384 Kbps cho cả 5 kênh và được ứng dụng rộng rãi cho các tín hiệu yêu cầu chất lượng cao.

2.3. MPEG-4

a. Giới thiệu chung về MPEG-4

Chuẩn MPEG-4 hướng tới việc tương thích với nhiều ứng dụng đa phương tiện và truyền tải trên Internet với tỉ lệ nén cao. Chuẩn MPEG-4 được tập hợp từ rất nhiều phần khác nhau với nhiều chức năng đa dạng cho mỗi phần. Với tín hiệu audio, chuẩn MPEG-4 sử dụng Part 3 để chuẩn hóa các kỹ thuật nén hỗ trợ nhiều loại âm thanh từ nhiều loại dải tần như tiếng nói, tiếng nói tổng hợp, âm thanh tổng hợp... Hiện tại đã có những công cụ sử dụng chuẩn MPEG-4 Part 3 để hỗ trợ mã hóa âm thanh đơn âm (chỉ sử dụng một kênh truyền) với bit rate chỉ từ 2 tới 64 kbps.

b. MPEG-4 Part 3


Đối với tín hiệu audio, chuẩn MPEG-4 xây dựng 3 bộ mã hóa chính (core coders) là:

- Mã hóa dựa trên thông số cho tiếng nói với bit rate thấp

- Mã hóa phân tích – tổng hợp (Analysis-by-Synthesis) cho âm thanh với bit rate từ khoảng 6 đến 16 kbps
- Mã hóa sub-band/transform cho âm thanh với bit rate cao (MPEG4 AAC)

Các bộ mã hóa được xây dựng từ rất nhiều kỹ thuật mã hóa khác nhau và rất đa dạng như mã hóa có mất mát thông tin (HVXC, CELP), mã hóa AAC, TwinVQ, BSAC, mã hóa không mất mát thông tin (MPEG-4 SLS, Audio Lossless Coding, MPEG-4 DST)...

Với tính chất như vậy, chuẩn MPEG-4 được sử dụng rất rộng rãi trong các thiết bị đòi hỏi chất lượng cao đối với tín hiệu audio. Với từng đối tượng âm thanh, MPEG-4 Part 3 xây dựng các profile khác nhau để đạt được hiệu quả nén tốt nhất. Các profile được xây dựng từ nhiều thông số, trong đó, 2 thông số quan trọng nhất là tốc độ lấy mẫu và số lượng kênh truyền:

MPEG-4 Audio Profiles	
Audio Profile 	Audio Object Types
AAC Profile	AAC LC
High Efficiency AAC Profile	AAC LC, SBR
HE-AAC v2 Profile	AAC LC, SBR, PS
Main Audio Profile	AAC Main, AAC LC, AAC SSR, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI, Main synthesis
Scalable Audio Profile	AAC LC, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI
Speech Audio Profile	CELP, HVXC, TTSI
Synthetic Audio Profile	TTSI, Main synthesis
High Quality Audio Profile	AAC LC, AAC LTP, AAC Scalable, CELP, ER AAC LC, ER AAC LTP, ER AAC Scalable, ER CELP
Low Delay Audio Profile	CELP, HVXC, TTSI, ER AAC LD, ER CELP, ER HVXC
Natural Audio Profile	AAC Main, AAC LC, AAC SSR, AAC LTP, AAC Scalable, TwinVQ, CELP, HVXC, TTSI, ER AAC LC, ER AAC LTP, ER AAC Scalable, ER TwinVQ, ER BSAC, ER AAC LD, ER CELP, ER HVXC, ER HILN, ER Parametric
Mobile Audio Internetworking Profile	ER AAC LC, ER AAC Scalable, ER TwinVQ, ER BSAC, ER AAC LD
HD-AAC Profile	AAC LC, SLS
ALS Simple Profile	ALS

Hiện tại chuẩn MPEG-4 vẫn đang được nghiên cứu và tiếp tục phát triển.

Kết luận

Mã hoá dải tần là kết quả của nghiên cứu trong lĩnh vực xử lý tín hiệu số, nhằm phân tách tín hiệu thành các thành phần tín hiệu trên các dải tần và từ đó mã hoá riêng biệt nhằm đạt hiệu quả nén cao với từng thành phần. Phương pháp này có sự tương đồng về mặt toán học với các phép biến đổi DCT, DWT đang được dùng trong nén ảnh JPEG và JPEG 2000. Nói cách khác, cho dù diễn giải khác nhau, thứ tự thực hiện khác nhau, tuy nhiên chúng đều có đặc điểm chung là giải tương quan tín hiệu bằng cách phân tách thành các thành phần cơ bản và từ đó có phương pháp lượng tử hoá, mã hoá phù hợp nhằm đạt hiệu quả nén cao nhất.

Mã hoá dải tần được ứng dụng trong nén audio và video, cùng với các kết quả nghiên cứu về độ cảm thụ của tai người là nền tảng cho chuẩn nén MPEG. Nhóm đã nghiên cứu về ứng dụng của mã hoá dải tần trong nén audio và video, đồng thời nghiên cứu ứng dụng cụ thể của SBC trong chuẩn MPEG.

Do thời gian nghiên cứu có hạn cũng như khuôn khổ của tiểu luận không cho phép, có rất nhiều vấn đề khác liên quan tới SBC chưa được trình bày trong tiểu luận này. Nhóm rất mong nhận được sự góp ý và giúp đỡ của cô để sau này có thể tiếp tục nghiên cứu về vấn đề này.

Tài liệu tham khảo

1. **Stephen J. Solari**, "Digital video and audio compression", McGraw-Hill, 1997.
2. **Martin Vetterli, Jelena Kovacevic**, "Wavelet and Subband coding", Prentice Hall, 1995
3. **Jason R. VandeKieft**, "Computational improvements to linear convolution with multirate filtering methods", "Chapter 3. Multirate Filter banks"
Bản online:
<http://mue.music.miami.edu/thesis/jvandekieft/jvchapter3.htm>
4. **Eero Simoncelli, Edward H. Andelson**, "Subband Coding", "Chapter 4. Subband transforms", Kluwer Academic Press, 1990.
5. **Wikipedia**, "Subband Coding",
http://en.wikipedia.org/wiki/Sub-band_coding
6. **Aria Nosratinia, Geoffrey Davis, Zixiang Xiong**, "Subband image compression"
Bản online:
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.1554>
7. **Jennie Abraham**, "Subband Coding"
Bản online:
[www-ee.uta.edu/dip/Courses/EE5351/Subband_Coding%20\(1\).ppt](http://www-ee.uta.edu/dip/Courses/EE5351/Subband_Coding%20(1).ppt)
8. **Peter Noll**, "MPEG Digital Audio Coding Standards", CRC Press LLC, 2000
Bản online:
<http://erdos.csie.ncnu.edu.tw/~hychen/multimedia/mpeg%20audio%20coding.pdf>
9. **Wikipedia**, "Moving Picture Experts Group"
http://en.wikipedia.org/wiki/Moving_Picture_Experts_Group
10. **Yao Wang**, "Audio Coding", 2004
Bản online:
http://eeweb.poly.edu/~yao/EE3414/audio_coding.pdf
11. **Wikipedia**, "MPEG-4 Part 3"
http://en.wikipedia.org/wiki/MPEG-4_Part_3
12. **Yao Wang**, "Audio coding", Slide, Polytechnic University , Brooklyn
13. **Hung-Quoc Lai, Steven Tjoa**, "Filter Bank Design and Subband Coding for Digital Audio"
14. **Otolith**, "Subband Coding"
<http://www.otolith.com/otolith/olt/sbc.html>