BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

HUỲNH TRỌNG NGUYÊN

NGHIÊN CỨU CÁC KỸ THUẬT NÉN TÍN HIỆU AUDIO TRONG TRUYỀN HÌNH SỐ

Chuyên nghành : KỸ THUẬT ĐIỆN TỬ
Mã số : 60.52.70

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2011

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **TS. Phạm Văn Tuấn**

Phản biện 1: **TS. Ngô Văn Sỹ**

Phản biện 2: **TS. Nguyễn Hoàng Cẩm**

Luận văn được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 03 tháng 12 năm 2011

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại Học Đà Nẵng
- Trung tâm Học liệu, Đại Học Đà Nẵng.

MỞ ĐẦU

1. Cơ sở nghiên cứu của luận văn

Tín hiệu audio số PCM được sử dụng trong truyền hình, truyền thông đa phương tiện cũng như trong nhiều ứng dụng khác. Các dòng số này có tốc độ bít rất cao, không thể ghép với dòng video số

2. Mục đích nghiên cứu

Luận văn tập trung nghiên cứu các nội dung sau:

- Nghiên cứu nguyên lý về các kỹ thuật nén âm thanh
- Nghiên cứu các kỹ thuật nén âm thanh MPEG và AAC
- Đánh giá hiệu quả các kỹ thuật nén âm thanh MPEG và AAC

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu

- Nghiên cứu kỹ thuật nén âm thanh chuẩn MPEG3 và AAC

3.2. Phạm vi nghiên cứu

- Tìm hiệu tổng quan về kỹ thuật nén
- Nghiên cứu các thuật toán nén
- Đánh giá chất lượng các kỹ thuật nén

4. Phương pháp nghiên cứu

- Nghiên cứu lý thuyết các kỹ thuật nén
- Xây dựng các File âm thanh
- Thực hiện chương trình nén
- Đánh giá

5. Ý nghĩa khoa học của đề tài

- Hổ trợ cho việc sử dụng các công nghệ truyền tải dữ liệu truyền hình số
- Phát triển sử dụng chương trình nén âm thanh

6. Cấu trúc của luận văn

Luận văn được chia làm 4 chương. Phần mở đầu luận văn trình bày tóm tắt mục đích nghiên cứu, đối tượng nghiên cứu, phạm vi nghiên cứu, phương pháp nghiên cứu và ý nghĩa khoa học đề tài.

Chương 1: CÁC CHUẨN NÉN TÍN HIỆU AUDIO TRONG TRUYỀN HÌNH SỐ

Chương 2: CÁC KỸ THUẬT MÃ HÓA ÂM THANH.

Chương 3: KỸ THUẬT MÃ HÓA ÂM THANH MPEG-3 VÀ MPEG-2 AAC,

Chương 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

CHƯƠNG 1: CÁC CHUẨN NÉN TÍN HIỆU AUDIO TRONG TRUYỀN HÌNH SỐ

1.1. Một số chuẩn truyền hình số hiện nay trên thế giới

1.1.1. Chuẩn ATSC

Hệ thống ATSC có cấu trúc dạng lớp. Mỗi lớp ATSC có thể tương thích với các ứng dụng khác cùng lớp. ATSC sử dụng dạng thức gói MPEG-2 cho cả Video, Audio và dữ liệu phụ.

1.1.2. Chuẩn DVB

Chuẩn DVB được sử dụng ở Châu Âu, truyền tải Video số MPEG-2 qua cáp, vệ tinh và phát truyền hình mặt đất.

Chuẩn DVB có một số đặc điểm như sau:

- Mã hoá Audio tiêu chuẩn MPEG-2. MPEG-2-AAC
- Mã hoá Video chuẩn MPEG-2.

DVB gồm một loạt các tiêu chuẩn. Trong đó cơ bản là:

- DVB - S: Hệ thống truyền tải qua vệ tinh. Bề rộng băng thông mỗi bộ phát đáp từ 11 đến 12 G hz.

Hệ thống DVB - S sử dụng phương pháp điếu chế QPSK (Quadratue Phase - Shift Keying), mỗi sóng mang cho một bộ phát đáp. Tốc độ bit truyền tải tối đa khoảng 38,1Mbps.

- DVB - C: Hệ thống cung cấp tín hiệu truyền hình số qua mạng cáp. Tốc độ bit lớp truyền tải MPEG-2 tối đa là 38,1 Mbps.
- DVB - T: Hệ thống truyền hình mặt đất với các kênh 8MHz. Tốc độ bit tối đa 24 Mbps. Sử dụng phương pháp điều chế RF mới đó là COFDM.

1.2. Các chuẩn nén âm thanh trong truyền hình số

1.2.1. Chuẩn mã hóa âm thanh MPEG

MPEG-1 là thuật toán nén âm thanh tiêu chuẩn quốc tế đầu tiên

cho nén âm thanh kỹ thuật số có độ trung thực cao. Chuẩn nén âm thanh MPEG-1 với tốc độ lấy mẫu 32, 44.1, 48 kHz. Tốc độ bít có thể hỗ trợ một hoặc hai kênh âm thanh và được xác định trong khoảng từ 32 đến 224 kbps cho mỗi kênh. Chuẩn mã hóa âm thanh MPEG được phân chia thành 3 lớp.

1.2.2. Chuẩn nén âm thanh AC3

Dolby Digital và công nghệ mã hóa âm thanh AC-3 được sửng dụng rộng rãi và trở thành không thể thiếu được trong các máy giải trí gia đình, DVD và HDTV. Dolby Digital trở nên không thể thiếu được trên đĩa phim DVD-Video và được thấy thường xuyên trên đĩa DVD-Audio. Hầu hết các đầu thu AV Preamp có khả năng giải mã Dolby Digital. Ưu điểm chính của Dolby Digital chính.

1.2.3. Chuẩn nén âm thanh AAC

Để tiến đến công nghệ mã hóa âm thanh vượt qua MP3, AC3, một nỗ lực được thực hiện để tạo ra một mã mới âm thanh có chất lượng không thể phân biệt tại tốc độ bít 64 kbps cho mono đó là chuẩn âm thanh MPEG-2 AAC. Về mặt kỹ thuật, định dạng AAC được tiêu chuẩn hóa vào năm 1997, được xây dựng trên một cấu trúc tương tự như MP3 và do đó giữ lại hầu hết các tính năng thiết kế của nó. Nhưng không giống như các lớp MPEG trước đây, AAC sử dụng một cách tiếp cận kiểu mô-đun (xem hình 1.2).

CHƯƠNG 2: KỸ THUẬT MÃ HÓA ÂM THANH

2.1. Tổng quan về mã hóa âm thanh

Mã hóa tín hiệu audio được thực hiện dựa trên cơ sở mô hình tâm lý thính giác, sự cảm nhận về âm thanh của hệ thống thính giác con người, sự hạn chế về mặt cảm nhận và hiện tượng che lấp các thành phần tín hiệu âm thanh.

2.1.1. Đặc tính sinh lý về sự cảm nhận âm thanh

Hệ thống thính giác của con người (Human Auditory System-HAS) như một dãy các bộ lọc thông dải.

2.1.2. Sự che lấp tín hiệu âm thanh

- Che lấp tần số
- Sự che phủ thời gian

2.2. Các kỹ thuật mã hóa âm thanh

Sơ đồ của bộ mã hóa như hình 2.3, do đặt tính của hóc tai như bộ lọc thông dãi, sự cảm nhận âm thanh của hệ thống thính giác của con người phụ thuộc vào độ phân giải tần số. Do vậy tín hiệu vào sẽ được chia thành các băng con (subband).

2.2.1. Kỹ thuật xử lý băng con (Subband)

Do thuộc tính che tần số của hệ thống thính giác trong miền tần số, việc dùng băng con (subband) hoặc biến đổi bộ lọc dãi (transform filter bank) là rất hiệu quả trong phân tích cảm quan về sự cảm nhận âm thanh của con người.

2.2.2. Kỹ thuật chia các băng con

Trong công nghệ nén audio, sử dụng một số loại bộ lọc băng con ví dụ PQMF (Polyphase Quadrature Mirror Filter). Bộ lọc này có độ chồng phổ thấp và thường được sử dụng cho các mẫu gần kề về mặt thời gian.

2.2.3. Kỹ thuật phân phối bít

Sử dụng thuật toán biến đổi Fourier nhanh (Fast Fourier Transform-FFT) được thực hiện để xác định nội dung tần số và năng lượng của tín hiệu vào. Từ ngưỡng nghe được và đặc tính che phủ tần số của HAS, người ta tính toán được đường cong che lấp như minh họa trên hình 2.4.

2.2.4. Lượng tử hóa

Quá trình lượng tử hoá các băng tần con trong phổ tín hiệu audio là một quá trình không đồng bộ. Tức là mỗi băng con được lượng tử với một bước lượng tử khác nhau phù hợp với mức năng lượng cũng như mức độ che lấp của băng tần. Bước lượng tử được xác định nhờ bộ phân phối bít.

2.2.5. Ghép kênh dữ liệu

Các khối (hay còn gọi là các nhóm) 12 mẫu dữ liệu từ đầu ra bộ lượng tử hoá được ghép kênh cùng với tham số xếp loại tương ứng của chúng và thông tin phân phối bít để hình thành nên khung dữ liệu audio trong dòng bít mã hoá.

2.2.6. Công nghệ giảm tốc độ nguồn dữ liệu audio số

Công nghệ mã hoá nguồn được sử dụng để loại bỏ đi sự dư thừa trong tín hiệu audio (khi giá trị vi sai mẫu - mẫu sắp sỉ gần giá trị 0), còn công nghệ che lấp dựa trên mô hình tâm lý thính giác của con người

CHƯƠNG 3: KỸ THUẬT MÃ HÓA ÂM THANH MP3 VÀ AAC

3.1. Kỹ thuật mã hóa âm thanh MP3

Định dạng MP3 được sử dụng để mã hóa âm thanh sử dụng kỹ thuật nén tổn hao. Dựa chủ yếu vào mô hình cảm quan. Loại bỏ một số Tần số âm thanh không được nghe theo hệ thống thính giác của con người.

3.2. Lịch sử phát triển tiêu chuẩn MP3

3.3. Thuật toán mã hóa MP3

Thuật toán nen MPEG gồm các bước sau:

- Đầu tiên tín hiệu âm thanh được chia thành các thành phần nhỏ hơn gọi là khung.
- Bước thứ hai biến đổi FFT 1024 điểm trên một mẫu và áp dụng mô hình cảm quan. Sử dụng mặt nạ và ngưỡng để loại bỏ các dữ liệu là không nghe được theo hiệu ứng tâm lý thính giác.
- Bước thứ ba định lượng và mã hóa mỗi mẫu của băng con (subband) bằng cách tính toán hệ số cần thiết đại diện cho tỷ lệ (SNR). Xem xét đầu ra các mẫu từ bộ lọc và tỷ số SMRs từ mô hình cảm quan (psychoacoustic) để điều chỉnh việc phân bổ tỷ lệ bit theo yêu cầu mặt nạ.
- Giai đoạn cuối cùng bao gồm các định dạng dòng bít (bitstream). Lượng tử hóa kết quả đầu ra từ bộ lọc, phân bổ nhiễu và các thông tin yêu cầu được thu thập sau đó mã hóa và định dạng.

Thông số kỹ thuật khác cho các thuật toán như sau:

- Tỷ lệ bit từ 8 kbps đến 320 kbps. Tỷ lệ bit đề cập đến số lượng dữ liệu (bit) được lưu trữ cho tất cả âm thanh sau. Tỷ lệ bit tiêu chuẩn là 128 kbps.

- Tỷ lệ lấy mẫu là 32 kHz, kHz 44.1, 48 kHz. Tỷ lệ lấy mẫu liên quan đến tần số mà tín hiệu được lưu trữ. Tỷ lệ lấy mẫu mặc định tiêu chuẩn là 44,1 kHz.
- Dòng bit được mã hóa với một tốc độ bit không đổi (CBR) hoặc với một biến thay đổi (VBR)
- Chế độ hỗ trợ sẽ là mono, dual channel, stereo and joint stereo.

3.3.1. Bộ lọc thời gian – tần số

Bộ lọc phân tích các băng con là một bộ lọc đa pha. Được thiết kế từ các bộ lọc dãi bao gồm toàn bộ dải tần số âm thanh. Được sử dụng để phân chia các tín hiệu PCM đầu vào với tần số lấy mẫu fs thành các băng con (subbands).

3.3.1.1. Lọc thông cao

Tiêu chuẩn MP3 sử dụng một bộ lọc thông cao. Cho phép tần số trên tần số cắt nhất định đi qua và không cho phép những tần số thấp hơn đi qua. Việc áp dụng loại bộ lọc này tránh được yêu cầu tốc độ bit cao không cần thiết cho các băng con thấp làm tăng chất lượng âm thanh tổng thể.

3.3.1.2. Bộ lọc phân tích băng con

Giàn lọc phân tích các băng con cơ bản là một bộ lọc đa pha. Được thiết kế từ các bộ lọc dãi bao gồm toàn bộ dải tần số âm thanh. Được sử dụng để phân chia các tín hiệu PCM đầu vào với tần số lấy mẫu fs trong băng con (subbands). Kết quả sẽ có 32 subbands bằng nhau với tần số lấy mẫu fs/32.

3.3.1.3. Bộ lọc đa pha

Các bộ lọc đa pha được sử dụng trong MP3, nguyên mẫu từ biến đổi cosin của bộ lọc thông thấp với bộ lọc thông dãi song song M

kênh. Được gọi là QMF (Quadrature Mirror Filter) bộ lọc gương tứ cầu. Với M chạy từ 0 đến 31. Ưu điểm của bộ lọc là:

- Thiết kế đơn giản từ bộ lọc một bộ lọc FIR có đáp ứng xung hữu hạn
- Các kênh đều có pha tuyến tính

3.3.1.4.Thực hiện thuật toán

Thực hiện phân tích băng con từ các thuật toán nén MP3 gồm các bước sau đây:

- Đầu vào 32 mẫu âm thanh Wi với i = 0 đến 31.
- Xây dựng một vector đầu vào X gồm 512 phần tử

$$Xi = X_{i-32} \ \ cho\ i = 511 \text{ xuống } 32$$

32 mẫu âm thanh tại các vị trí từ 0 đến 31, gần nhất tại vị trí 0, và 32 phần tử cũ nhất được chuyển ra.

$$Xi = W_{31-i} \ \ chor\ i = 31 \text{ xuống } 0$$

- Cữa sổ vector X bằng vector C. Với C là các hệ số được tìm thấy trong Bảng 3.1,3.2, 3.3, 3.4, 3.5, 3.6, 3.7 và 3.8

$$Zi = Ci * Xi \ ; \ chor\ i = 0 \text{ đến } 511$$

- Tính 64 giá trị của Yi bởi công thức sau:

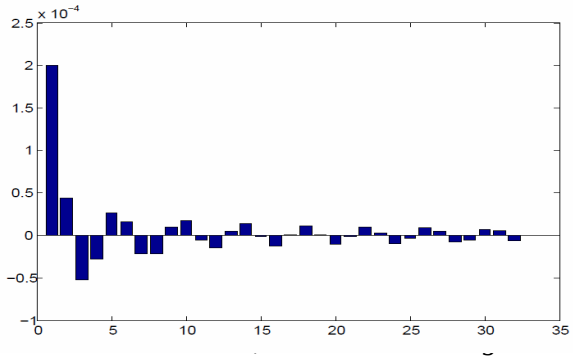
$$Y_i = \sum_{j=0}^{7} z_i + 64j \ \ \textbf{cho}\ \text{i = 0 đến 63}$$

Tính toán 32 mẫu subband ma trận Si

$$S_i = \sum_{k=0}^{63} M_{i,k} + Y_k \qquad \textbf{cho}\ \text{i = 0 đến 31}$$

Các hệ số ma trận M có thể được tính theo công thức sau đây:

$$M_{i,k} = \cos\left[\frac{(2i+1)(k-16)\pi}{64}\right] \quad \textbf{cho} \;\; i = 0:31,\ k=0:63$$



3.3.2. Mô hình cảm quan

3.3.2.1. Ngưỡng nghe tuyệt đối

Ngưỡng nghe tuyệt đối là đại lượng biểu thị về mức năng lượng âm thanh cần thiết có thể nghe được trong môi trường yên lặng.

Giá trị ngưỡng được thể hiên là dB SPL (Sound Pressure Level) và được đặc trưng bởi hàm tuyến tính sau:

$$T_q(f) = 3{,}64\left(\frac{f}{100}\right)^{-0,8} - 6{,}5e^{-0,6\left(\frac{f}{1000}-3{,}3\right)^2} + 10^3\left(\frac{f}{1000}\right)^4 \; dB$$

3.3.2.2. Băng tới hạn (band tới hạn)

Hóc tai của con người được xem như bộ lọc thông dãy với độ rộng băng thông không đồng đều và được dùng theo công thức sau:

$$Z(f) = 13 \arctan(.00076f) + 3{,}5 \arctan\left[\left(\frac{f}{7500}\right)^2\right] \text{Bard} \qquad (3.8)$$

Tần số và độ rộng băng tần của các băng tới hạn được cung cấp tại bảng 3.9.

3.3.2.3. Thực hiện thuật toán

3.3.2.4. Biến đổi FFT

Mẫu âm thanh đến, s (n), được chuẩn hóa [8] theo chiều dài FFT là N, và số bit cho mỗi mẫu là b, sử dụng theo phương trình:

$$x(n) = \frac{s(n)}{N\left(2^{b-1}\right)}$$

Ngưỡng mặt nạ ước tính từ của mật độ phổ công suất, P (k) được tính bằng phép phân tích FFT 1024-điểm

h (n) là một cửa sổ Hann tính từ:

$$P(k) = PN + 10\log\left|\sum_{n=0}^{N-1} h(n)x(n)e^{-j\frac{2\pi kn}{N}}\right|^2 dB \text{ với } 0 \le k \le N/2$$

Và PN năng lượng ở mức tham chiếu 96 dB SPL.

- Các cửa sổ Hann phải trùng khớp với các mẫu băng con của khung.
- Kích thước cửa sổ tùy thuộc vào tần số lấy mẫu fs được liệt kê trong Bảng 3.10

3.3.2.5. Xác định SPL (SPL determination)

L_{SB} mức độ áp lực âm thanh trong subband n được tính bằng:

$$L_{SB}(n) = \max[P(k),\ 20\ \log(SCF_{max}(n)\ * 32768) — 10]\ (dB)$$

P (k) là mức áp suất dòng âm thanh với chỉ số k của FFT và biên độ tối đa trong phạm vi tần số tương ứng với subband n. SCFmax(n) biểu hiện tối đa ba hệ số chia tỷ lệ băng con thứ n trong một khung.

Mức -10 db là hiệu của đỉnh và RMS (root-mean-square). $\boldsymbol{L_{SB}(n)}$ được tính cho mỗi băng con thứ n.

3.3.2.6. Ngưỡng yên lặng

Các ngưỡng yên lặng $T_q(k)$, hoặc ngưỡng nghe tuyệt đối theo phương trình: 3.13 được tính trong bảng 3.11, 3.12, 3.13, 3.14 và 3.15.

$$T_q(f) = 3{,}64\left(\frac{f}{100}\right)^{-0,8} - 6{,}5e^{-0,6\left(\frac{f}{1000}-3{,}3\right)^2} + 10^3\left(\frac{f}{1000}\right)^4 \; dB$$

3.3.2.7. Thành phần âm và không âm

Tính toán ngưỡng mặt nạ toàn phần để lấy được các thành phần âm và không âm từ phổ của biến đổi FFT.

Bắt đầu với việc xác định biến cục bộ cực đại, sau đó trích các thành phần âm và tính toán cường độ của các thành phần không âm trong một băng thông của các băng tần tới hạn. Biên của các băng tần tới hạn được đưa ra trong Bảng 3.16, 3.17.

Để xác định biến cục bộ tối đa của một thành phần âm trong dải tần số df xung quanh biến cục bộ tối đa được đưa ra bởi bảng 3.18.

Để xác định dãy các vạch phổ của P (k) là âm hay không âm, có ba cách sau đây được thực hiện:

- Ghi nhãn của biến cục bộ cực đại

Một dòng phổ, X (k), được dán nhãn nếu:

$$P(k) > P(k - 1) \text{ và } \quad P(k) > P(k + 1):$$

- Lập Danh sách các thành phần âm và tính toán mức độ áp suất âm thanh

  Biến cục bộ tối đa là có trong các thành phần âm nếu

  $$P(k) - P(k + j) = 7 \; dB \qquad j$$ là lựa chọn theo Bảng 3.19.

Nếu P (k) được tìm thấy là một thành phần âm, thì các thông số sau đây được liệt kê:

- Chỉ số số k của dòng phổ.
- Mức áp suất âm thanh

$$P_{TM}(k) = P(k-1) + P(k) + P(k+1)\ \text{dB}$$

- Tonal flag.

Xác lập, tất cả các đường phổ trong dải tần số kiểm tra là 8 dB.

- Lập các thành phần không âm và tính toán năng lượng phổ các thành phần không âm (nhiễu) từ các dòng phổ còn lại. Để tính toán các thành phần không âm từ các vạch phổ P (k), các băng tần tới hạn z (k) bằng cách sử dụng Bảng 3.8.

3.3.2.8. Giảm các thành phần mặt nạ

Số lượng maskers được xem xét để giảm khi:

- Các thành phần Âm $P_{TM}(k)$ hoặc không âm $P_{MN}(k)$ được xem xét để tính toán ngưỡng mặt nạ nếu:

$$P_{TM}(k) >= T_Q(k)$$

Hoặc $$P_{MN}(k) >= T_Q(k)$$

$T_Q(k)$ là ngưỡng tuyệt đối tại tần số k. Những giá trị này được đưa ra trong Bảng 3.11, 3.12, 3.13, 3.14 và 3.15

- Hai hoặc nhiều thành phần âm trong một khoảng cách ít hơn 0,5 Bark. Có năng lượng cao nhất thì được lưu giữ, và các thành phần nhỏ hơn từ danh sách của các thành phần âm thì được loại bỏ. Một cửa sổ trượt trong các băng tần tới hạn sẽ được sử dụng với chiều rộng là 0,5 Bark.

3.3.2.9.Tính toán các ngưỡng mặt nạ

Trong N/2 mẫu miền tần số, k là số mẫu, i là ngưỡng mặt nạ toàn phần. Các mẫu sử dụng được thể hiện trong bảng 3.11, 3.12, 3.13, 3.14 và 3.15

Số lượng mẫu, i, trong lĩnh vực tần số lấy mẫu các băng con là khác nhau tùy thuộc tỷ lệ lấy mẫu (Bảng 3.20).

Chỉ số i chỉ thành phần âm và không âm trong miền trong tần số lấy mẫu các băng con gần nhất với tần số dòng phổ gốc P (k). Chỉ số này được đưa ra trong bảng 3.11, 3.12, 3.13, 3.14 và 3.15.

Các ngưỡng mặt nạ riêng lẽ của các thành phần âm và không âm được đưa ra bởi biểu thức sau đây:

$T_{TM}[z(j), z(i)] = P_{TM}[z(j)] + AV_{TM}[z(j)] + VF[z(j), z(i)]$ (dB)

$T_{NM}[z(j), z(i)] = P_{NM}[z(j)] + AV_{NM}[z(j)] + VF[z(j), z(i)]$ (dB)

Mặt nạ âm được cho bởi:

$$AV_{TM} = -1.525 - 0.275z(j) - 4.5(dB)$$

Mặt nạ không âm được cho bởi:

$$AV_{NM} = -1.525 - 0.175z(j) - 0.5(dB)$$

Chức năng che VF khác nhau, cao, thấp phụ thuộc vào khoảng cách dz = z (i) +z (k) để che.

Tỷ lệ các băng tần tới hạn z (j) và z (i) như trong các bảng 3.11, 3.12, 3.13, 3.14 và 3.15. Các chức năng tạo mặt nạ, mặt nạ âm và không âm là như nhau, và được cho bởi phương trình (3.19).

Trong các biểu thức P [z (j)] là mức áp suất âm thanh của các thành phần mặt nạ thứ j dB. Nếu dz <-3 Bark, hoặc dz ≥ 8 Bark, thì P [z (j)] không còn được coi mặt nạ và (T_{TM} và T_{MN} được thiết lập -8 dB).

4.3.2.10. Ngưỡng mặt nạ toàn phần

Ngưỡng T_g(i) mặt nạ toàn phần (Eq. 3.19) tại mẫu thứ i, tần số trên và dưới của ngưỡng mặt nạ riêng lẽ j của thành phần âm và không âm, và ngưỡng yên lặng T_q(i) được đưa ra trong bảng 3.11, 3.12,

3.13, 3.14 và 3.15. Ngưỡng các mặt nạ toàn phần được tính, bằng cách tổng hợp các năng lượng tương ứng với ngưỡng mặt nạ âm và không âm với ngưỡng yên lặng.

$$T_g(i) = 10\log\left(10^{(0.1T_q(i))} + \sum_{l=1}^{L} 10^{(0.1T_{TM}(i,l))} + \sum_{m=1}^{M} 10^{(0.1T_{NM}(i,m))}\right)(dB)$$

Tổng số mặt nạ âm được cho bởi l, và tổng số của mặt nạ không âm được cho bởi m. Đối với i, phạm vi của j có thể được giảm xuống chỉ còn những thành phần mặt nạ trong vòng -8 <=i<= +3 Bark. Bên ngoài của phạm vi này T_{TM} và T_{NM} là -8 dB.

3.3.2.11. Ngưỡng Mặt nạ tối thiểu

Mặt nạ mức tối thiểu T_{min} (n) trong subband n được xác định [4] như sau:

$$T_{min}(n) = min[Tg(i)]\ (dB)$$

Tg(i) là tần số của mẫu thứ i trong subband n. Tg (i) như trong bảng 3.11, 3.12, 3.13, 3.14 và 3.15

3.3.2.12. Tính toán SMR

SMR được tính cho mỗi subband n (Eq. 3.21).

$$SMR_{SB}(n) = L_{SB}(n) - T_{min}(n)\ (dB)$$

3.3.3. Biến đổi MDCT (Modified Discrete Cosine Transform)

Các mẫu âm thanh Đầu ra từ các bộ lọc P-QMF là không đạt được hoàn hảo. Biến đổi MDCT sẽ làm hạn chế các nguồn tín hiệu biến dạng cho giai đoạn lượng tử hóa. MDCT là trường hợp đặc biệt của biến đổi DCT với L = 2M. Các hệ số từ bộ lọc phân tích MDCT được cho bởi:

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos[\frac{\pi}{4M}(2n + M + 1)(2k + 1)]$$

và hệ số tổng hợp là:

$$g_k(n) = h_k(n)\ (2M\text{-}1\text{-}n)$$

3.3.3.1. Biến đổi MDCT thuận và nghịch

Biến đổi MDCT Thuận theo phương trình (3.24)

3.3.3.2. Cửa sổ biến đổi

Cửa sổ được chọn cho biến đổi MDCT là:

$$W(n)=\sin\left[(n+\frac{1}{2})\frac{\pi}{2M}\right]$$

3.3.3.3. Cách tính

3.3.4. Lượng tử hóa

Trong kỹ thuật mã hóa MP3, thực hiện lượng tử hóa và mã hóa các thành phần phổ với yêu cầu nhiễu lượng tử hóa thấp hơn ngưỡng mặt nạ. Các giá trị lượng tử hóa được mã hóa Huffman với bảng mã thay đổi đối với những dải tần số khác nhau, để thích nghi tốt hơn với tín hiệu.

3.3.4.1.Thiết lập giá trị trung bình bít chứa

Số lượng trung bình của các bit cho granule (1/2 mẫu) được tính toán từ các kích thước khung. Ví dụ tốc độ bit 64 kbps tần số lấy mẫu là 48.000 Hz, thì số bít có trong hồ chứa được tính:

$$\frac{(64000 * 0.024 bit / frame}{2 granules / frame} = 768 bit / granule$$

Phần đầu lấy 32 bit, phần thông tin lấy 17 byte (136 bit) chế độ kênh đơn, số bít trung bình cho các dữ liệu chính cho 1 granule đưa ra bởi:

$$mean_bit = 768 bit / granule - \frac{(32+136 bit / frame}{(2 granules / frame)} = 684 bit / granul$$

3.3.4.2. Thiết lấp các thông số vòng lặp

3.3.4.3. Kiểm soát bít chứa

Bit được lưu vào hồ chứa, khi ít hơn so với các bit trung bình

(mean_bits) thì được sử dụng để mã một hạt chứa bít (granule) trong 1/2 frame. Nếu các bit được lưu đủ cho một khung thì kết thúc việc tăng số bít lên. Cách tổ chức phân phối các dòng bít (bitstream) được minh họa như hình 3.4.

3.3.4.4. Hệ số tỷ lệ thang lượng tử

Scfsi chứa các thông tin, (được phân nhóm trong các dãy hệ số scfsi), thông tin tại chỉ số đầu tiên cũng được sử dụng cho các thông tin trong chỉ số thứ hai. Do đó hệ số bit khi đạt được hệ số được sử dụng để mã hóa Huffman không được truyền. Việc xác định sử dụng các thông tin của scfsi, cho mỗi hạt phải được lưu trữ theo các kiểu:

- Kiểu khối

Kiểu tính tổng năng lượng trong granule như phương trình (3.40)

Năng lượng phổ theo dãi các hệ số như phương trình (3.41)

- Sai số chấp nhận cho hệ số dãi như phương trình (3.42)
- Các giá trị phổ khác 0
- Không có granule nào chứa trong khối ngắn
- Trị tuyệt đối $\left| en_tot_0 - en_tot_1 \right| < en_tot_{krit}$

$$\sum_{all\ \ scale\ factor\ \ band} \left| en\ (cb)_0 - en\ (cb)_1 \right| < en_dif_{krit}$$

Khi:

$$\sum_{all\ \ cb's\ \ in\ \ scfsi_band} \left| en(cb)_0 - en(cb)_0 \right| < en(scfsi_band)_{krit}$$

$$\sum_{all\ \ cb's\ \ in\ \ scfsi_band} \left| xm(cb)_0 - xm(cb)_0 \right| < xm(scfsi_band)_{krit}$$

Các hằng số theo chỉ số (krit) được chọn các giá trị sao cho scfsi không bị cấm trong trường hợp năng lượng phổ bị méo dạng. Các giá trị đó được chọn là:

- en_ tot = 10
- en _dif = 100
- *en*(scfsi band) = 10, cho mỗi scfsi band
- *xm*(scfsi band) = 10, cho mỗi scfsi band

3.3.4.5. Vòng lặp

Trong kỹ thuật mã hóa MPEG-3 Các dữ liệu được lượng tử hóa và mã hóa trong hai vòng lặp lồng nhau. Được gọi là vòng lặp trong và vòng lặp ngoài

3.3.4.6. Vòng lặp ngoài

Vòng lặp bên ngoài có chức năng kiểm soát nhiễu (distortion control loop) lượng tử hóa. Hiệu chỉnh hệ số tỷ lệ để giảm nhiễu lượng tử hóa.

3.3.4.7.Vòng lặp trong

Chức năng vòng lặp bên trong là kiểm soát tốc độ bít lượng tử. Lựa chọn bảng, phân vùng các giá trị và lựa chọn kích thước bước lượng tử hóa.

3.3.5. Mã hóa Huffman

Lượng tử hóa hoàn toàn giá trị vector phổ được tính như sau:

$$ix(i) = \left[\left(\frac{|xr(i)|}{4\sqrt{2^{qquant + quan\tan f}}} \right)^{\frac{3}{4}} - 0.0946 \right] \qquad (3.44)$$

3.4 . Kỹ thuật mã hóa AAC [17]

3.4.1. Mô hình cảm quan

Trong mã hóa AAC. Các mô hình cảm quan được sử dụng để tính toán năng lượng biến dạng tối đa được che bởi năng lượng tín hiệu. Năng lượng này được gọi là ngưỡng.

3.4.1.1. Chức năng phân bố

Sử dụng các biến tạm thời tmp để tính các giá trị tín hiệu được phân bố trong dãi

3.4.1.2. Các bước tính toán ngưỡng

3.4.2. Quá trình xử lý kiểm soát độ lợi trong mã hóa

Các hệ thống kiểm soát được bao gồm một giàn lọc PQF (Polyphase Quadrature Filter),

3.4.2.1. PQF(Polyphase Quadrature Filter)

3.4.2.2. Kiểm soát phát hiện

Kiểm soát phát hiện với chức năng thay đổi dữ liệu về số lượng, vị trí, độ lớn. Phát hiện các điểm thay đổi được thực hiện trong nửa thứ hai của các khu vực cửa sổ MDCT và trong khu vực không chồng phổ

3.4.3. Xử lý chuyển đổi giàn lọc và khối

3.4.3.1. Chuyển đổi cửa sổ khối

Nhằm đáp ứng về độ phân giải thời gian- tần số của giàn lọc với đặc điểm của tín hiệu đầu vào được thực hiện bằng cách dịch chuyển giữa các biến đổi cửa sổ có chiều dài đầu vào là 2048 hoặc 256 mẫu.

3.4.3.2. Biến đổi MDCT

3.4.4. Mã hóa dự đoán

3.4.4.1. Xử lý mã hóa dự đoán

Dự đoán được sử dụng để giảm sự dư thừa nhằm cải thiện yêu cầu về tốc độ bít. Thông tin dự đoán được truyền theo 2 bước:

Bước đầu tiên xác định hệ số chia tỷ lệ thang đo cho mỗi dãi tần, xác định có mã hóa hay không mã hóa.

Bước thứ 2 kiểm soát thông tin dự đoán tính giá trị cho dự đoán

3.4.4.2. Xây dựng lại các thành phần phổ lượng tử hóa

Việc xây dựng lại giá trị các thành phần phổ lượng tử hóa của tín hiệu đầu vào như dự đoán là tùy thuộc vào giá trị của bit dự đoán prediction_used.

➢ Nếu bit prediction_used được thiết lập 1, các lỗi dự đoán lượng tử hóa được tái tạo lại dữ liệu và được truyền đi, và được cộng thêm vào x_{est} (n).

$$x_{rec}=x_{est}+e_q(n) \quad (3.59)$$

➢ Nếu bit prediction_used không được thiết lập 0, thì giá trị lượng tử hóa của các thành phần phổ giống hệt giá trị tái tạo trực tiếp từ các dữ liệu được truyền.

3.4.5. Định hình nhiễu tạm thời (TNS)

TNS được sử dụng để kiểm soát tạm thời nhiễu lượng tử hóa trong mỗi cửa sổ của biến đổi. Điều này được thực hiện nhờ áp dụng quá trình lọc một phần phổ dữ liệu vào.

3.4.6. Lượng tử hóa

3.4.6.1. Thiết lập lại của tất cả các biến lặp

➢ Scalefactor [sb] được thiết lập 0 cho tất cả các giá trị của sb

➢ Tính toán giá trị start_common_scalefac cho lượng tử hóa để tất cả các giá trị khi biến đổi MDCT có thể được lượng tử hóa và mã hóa trong bitstream

3.4.6.2. Điều khiển bít chứa(Bit Reservoir Control)

Bit được lưu trong bit chứa khi ít hơn số lượng bít trung bình mean_bits được sử dụng để mã hóa một khung.

Số bit tối đa của các bit có sẵn cho một khung là tổng của mean_bits và các bit được lưu trong bit chứa.

3.4.6.3. Lượng tử hóa hệ số MDCT (Quantization of MDCT Coefficients)

3.4.6.4. Vòng lặp ngoài

3.4.6.5. Thực hiện gọi vòng lặp trong

Đối với mỗi vòng, vòng lặp bên ngoài được gọi là vòng kiểm soát nhiễu (distortion control loop), vòng lặp bên trong được gọi là vòng điều khiển tốc độ (rate control loop). Các thông số hệ số chia tỷ lệ

thang đo trong miền tần số được áp dụng cho các giá trị hệ số chia tỷ lệ thực tế trong dãi. Kết quả khi gọi vòng lặp trong thực hiện, số lượng của các bit thực tế sử dụng lượng tử hóa là x_quant (i), thì xác lập giá trị common_scalefac mới.

3.4.6.6. Giảm hệ số chia tỷ lệ thang đo

Tất cả các giá trị phổ trong các dãi hệ số chia tỷ lệ thang đo có một giá trị thay đổi vượt quá sự thay đổi cho phép (xmin (sb)), thì xác lập lại hệ số chia tỷ lệ mới.

3.4.6.7. Vòng lặp trong

Các vòng lặp bên trong tính toán lượng tử hóa thực tế các dữ liệu của miền tần số.

3.4.7. Mã hóa Huffman

Mã hóa Huffman được sử dụng để biểu diễn cho n-bộ dữ liệu của hệ số lượng tử với mã Huffman được rút ra từ một trong 11 codebooks. Các hệ số phổ trong n-bộ dữ liệu được sắp xếp theo thứ tự (thấp đến cao) và kích thước n-bộ dữ liệu là hai hoặc bốn hệ số. Được thể hiện trong bảng 3.47.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

4.1. Phương pháp đánh giá tỷ số SNR và nghe thử

Để đánh giá chất lượng mã hóa, về cơ bản có ba phương pháp đó là:

- Nghe thử (Đánh giá chủ quan)
- Đánh giá khách quan
- Nghe chủ quan.

4.1.1. Đánh giá chủ quan sử dụng mô hình MOS(Mean Opinion Score)

Dùng chức năng Recoring thu 5 file âm thanh có cùng tần số lấy mẫu là 44100, 48000 ở dạng *.WAV . Chất lượng âm thanh sau khi thu được phân tích tỷ số SNR.

Sau đó tiến hành nén 5 file âm thanh này với các chuẩn mã hóa MPEG-3 và MPEG-2 AAC bằng chương trình thực nghiệm trên Mathlab và trên C++.

Nhờ mười người nghe, kiểm tra và đánh giá hộ. Kết quả từng người sẽ được tính trung bình và lập hồ sơ đánh giá.

4.1.1.1: Thực hiện kiểm tra tỷ số SNR

Các file âm thanh sau khi được ghi, *dùng* đoạn chương trình trong mathlab kiểm tra tỷ lệ SNR. Sau đó phân loại theo tỷ số SNR từ thấp đến cao. Mã hóa các file âm thanh này được với hai tốc độ bít khác nhau là 64kbps và 128kbps với tần số lấy mẫu là 44.1KHz và 48KHz .

4.1.1.2. Đánh giá kết quả SNR cho 2 loại tốc độ bít và Fs

Đánh giá kết quả cho các file được thực hiện mỗi file 5 lần, và tính trung bình cho mỗi file. Kết quả như các bảng: 3.48, 3.49, 3.50, 3.51

4.1.1.3. Nghe kiểm tra

Kết quả nghe thử được thể hiện trong các bảng 3.49 đến 3.52 trong đó “tốt” nghĩa là file nén gần với file gốc, khá là còn ồn so gốc, trung bình khá còn méo và ồn so với gốc, trung bình là còn méo, ồn và đứt đoạn

4.1.1.4. Đánh giá tỷ lệ mã hóa

4.1.1.5.Nhận xét

- Qua đánh giá thực nghiệm về tỷ lệ SNR, nhận thấy rằng tỷ số SNR thay đổi rất lớn ở tốc độ bít thấp đói với kỹ thuật mã hóa MP3. Tỷ số SNR với chuẩn mã hóa MP3 là tăng lên so với

file gốc, điều này chứng tỏ với tốc độ bít thấp kỹ thuật mã hóa MP3 là không bằng AAC.

- Qua kết quả nghe thử cho thấy:
  - Tốc độ bít càng cao và tỷ lệ nén càng thấp thì file nén gần như là file gốc ở cả 2 chuẩn MP3 và AAC.
  - Tỷ lệ SNR là không thay đổi nhiều trong quá trình mã hóa kể cả 2 chuẩn.

KẾT LUẬN VÀ KIẾN NGHỊ

Luận văn đã trình bầy hai kỹ thuật mã hóa tín hiệu Audio, các nguyên lý nén và một số khái niệm quan trọng trong lĩnh vực mã hóa âm thanh. Luận văn cũng đã tập trung trình bầy các nguyên lý mã hóa âm thanh theo mô hình cảm quan, nén có tổn hao điển hình dựa vào hiệu ứng tâm lý nghe của tai, đang là các kỹ thuật cốt lõi của các chuẩn nén *MPEG layer 1,2,3 và MPEG-2 AAC.* Cả hai chuẩn mã hóa MP3 và MPEG-2 AAC đều có thể nén tín hiệu audio với chất lượng gần chất lượng của CD. Trong hai chuẩn trên, MP3 ít phức tạp hơn AAC, AAC cung cấp chất lượng tốt hơn MP3 với cùng tần số lấy mẫu và tỷ lệ nén.

Mục đích của luận văn là nghiên cứu các kỷ thuật mã hóa âm thanh của hai chuẩn MPEG-3 và MPEG-2 AAC, làm kiến thức cơ bản cho nghiên cứu tiếp các kỹ thuật mã hóa âm thanh của các chuẩn khác như MPEG-4, và đáp ứng nhu cầu cần thiết trong công việc và lựa chọn công nghệ.