

THỬ NGHIỆM ỨNG DỤNG KỸ THUẬT MÃ HÓA NÉN TÍN HIỆU ÂM THANH TẠI ĐÀI TIẾNG NÓI VIỆT NAM

Nguyễn Thanh Phong¹, Hoàng Lê Uyên Thực²

¹ Đài Tiếng nói Việt Nam thường trú Cần Thơ, 102 Lý Tự Trọng, Q. Ninh Kiều, Cần Thơ

² Trường Đại học Bách khoa, Đại học Đà Nẵng, 54 Nguyễn Lương Bằng, Q. Liên Chiểu, Đà Nẵng

nguyenthanhphong@vov.org.vn, hluthuc@dut.udn.vn

TÓM TẮT: Tại Đài Tiếng nói Việt Nam (TNVN), các kênh âm thanh số hiện nay đang được mã hóa theo chuẩn không nén. Do bản chất mà những tệp âm thanh dạng không nén có dung lượng cực lớn, vì vậy yêu cầu dung lượng ổ cứng lưu trữ các chương trình phát thanh rất lớn và việc thao tác trên các đoạn âm thanh rất khó khăn. Nghiên cứu được thực hiện nhằm kiểm tra việc áp dụng thử nghiệm các chuẩn nén âm thanh tại Đài TNVN, chuẩn bị cho việc triển khai mã hóa nén âm thanh tại Đài theo kế hoạch. Hai chuẩn nén âm thanh số có tổn hao chất lượng cao được chọn để đánh giá hiệu năng và chất lượng là MP3 (Moving Picture Experts Group 1 - Layer 3) và AAC (Moving Picture Experts Group 2 - Advanced Audio Coding). Lần lượt các chuẩn MP3 và AAC được áp dụng cho 18 tệp âm thanh trích chọn từ các chương trình khác nhau gần đây của Đài với tổng dung lượng là 5546MB (tương đương thời lượng hơn 8 giờ), sau đó được đánh giá dựa trên các tiêu chí về hiệu năng bao gồm thời gian thực thi và tỷ lệ nén. Tiêu chí chất lượng nén được đánh giá bằng phương pháp chủ quan dựa vào nghe thử. Thử nghiệm nghe thử được tiến hành trên 24 đoạn âm thanh ngắn nhiều thể loại, mỗi đoạn kéo dài trung bình là 20 giây, với số lượng tình nguyện viên tham gia là 35 người. Kết quả nghiên cứu cho thấy tính khả thi của việc áp dụng các chuẩn nén âm thanh mới tại Đài TNVN, đặc biệt giúp xác định tập giá trị các tham số phù hợp nhất đối với từng chuẩn nén nhằm tiết kiệm dung lượng lưu trữ dữ liệu và giảm thời gian thao tác trên tín hiệu trong khi vẫn đảm bảo chất lượng dịch vụ phát thanh.

Từ khóa: Mã hóa nén âm thanh, mã hóa MP3 (Moving Picture Experts Group 1 – Layer 3), mã hóa AAC (Moving Picture Experts Group 2 – Advanced Audio Coding), đánh giá chủ quan, nghe thử.

I. GIỚI THIỆU

Trong vài thập niên gần đây, ngành công nghiệp đa phương tiện đã có những bước tiến ngoạn mục, chịu tác động từ sự phát triển của hạ tầng mạng máy tính và internet theo hướng dịch chuyển từ mạng chuyên mạch kênh sang mạng chuyên mạch gói và chuyên mạch IP (Internet Protocol). Tiêu biểu như loại hình truyền tiếng nói qua mạng internet VoIP (Voice over IP) thay thế cho điện thoại truyền thống; truyền hình và phát thanh tương tự đang dần được thay thế bằng truyền hình và phát thanh kỹ thuật số; các đầu thu âm thanh số di động, truyền hình theo yêu cầu, v.v. ngày càng gia tăng nhanh chóng về số lượng [1]. Với những ưu điểm nổi bật về chất lượng sao chép, độ phân giải, khả năng miễn nhiễm đối với nhiễu, tính tương tác, khả năng khai thác dịch vụ gia tăng, v.v., ngành đa phương tiện phát triển theo xu hướng số hóa dần dần và hoàn toàn, kéo theo những ảnh hưởng mạnh mẽ đến sự phát triển của các ngành liên quan như phát thanh, truyền hình, game, phim ảnh, tiếp thị, v.v. [2].

Đối với loại dữ liệu âm thanh, việc số hóa trong giai đoạn đầu tiên được tiến hành theo định dạng wave với phần đuôi mở rộng là *.wav [3]. Wave sử dụng kỹ thuật điều xung mã PCM (Pulse Code Modulation) đơn giản để chuyển đổi dữ liệu từ tương tự sang số theo ba bước cơ bản là lấy mẫu, lượng tử hóa và mã hóa. Trước tiên, tín hiệu tương tự được rời rạc hóa theo thời gian, tạo thành chuỗi mẫu rời rạc. Chuỗi mẫu rời rạc này có biên độ được điều chế theo tín hiệu tương tự và có chu kỳ tuân theo định lý lấy mẫu Nyquist [4]. Sau đó, chuỗi mẫu được lượng tử hóa với các bước lượng tử đều hoặc không đều, tạo thành chuỗi mẫu có biên độ được làm tròn theo mức lượng tử gần nhất. Cuối cùng các mẫu lượng tử hóa được mã hóa để chuyển đổi thành các từ mã số nhị phân [4].

Định dạng wave dựa trên mã hóa dạng sóng nên có thể đạt được chất lượng gần với chất lượng âm thanh tương tự. Tuy nhiên, định dạng wave có nhược điểm rất lớn là dung lượng tín hiệu tỷ lệ với chất lượng tín hiệu, nghĩa là tín hiệu wave muốn có chất lượng càng cao thì tần số lấy mẫu và số bit mã hóa phải càng cao, dẫn đến dung lượng càng lớn. Dung lượng tín hiệu lớn dẫn đến không gian lưu trữ tín hiệu tăng và thời gian truyền dẫn tín hiệu lớn. Ngoài ra dung lượng lớn còn kéo theo hậu quả là băng thông đường truyền yêu cầu phải lớn, mà băng thông lại là tài nguyên hàng đầu cần thiết để có thể truyền thành công dòng dữ liệu từ đầu cuối đến đầu cuối. Ngày nay, ngay cả khi hạ tầng mạng viễn thông đã phát triển đến mức độ *băng rộng (broadband)* thì vẫn không thể đủ để truyền tải lượng dữ liệu đa phương tiện quá lớn bao gồm tiếng nói, âm thanh, hình ảnh, video, v.v. [5].

Vi thể cho nên vấn đề quan trọng đặt ra đối với lĩnh vực nghiên cứu về dữ liệu đa phương tiện nói chung và âm thanh nói riêng là phải tìm cách mã hóa dữ liệu âm thanh số nhằm nén xuống một dung lượng phù hợp mà vẫn đảm bảo chất lượng yêu cầu. Nói cách khác, chúng ta phải giải quyết mâu thuẫn giữa tỷ lệ nén dữ liệu và chất lượng âm thanh, sao cho vẫn đảm bảo yêu cầu về tiêu chuẩn phát thanh.

Tại Đài Tiếng nói Việt Nam (TNVN) hiện nay, hoạt động mã hóa dữ liệu âm thanh thực sự vẫn chưa tuân theo một bộ tiêu chuẩn quốc gia thống nhất. Dữ liệu âm thanh vẫn còn đang tồn tại song song hai hình thức tương tự và số, trong đó dữ liệu số đang sử dụng định dạng wave. Không nằm ngoài xu thế phát triển chung của phát thanh toàn cầu, hạ tầng kỹ thuật của Đài TNVN đã có kế hoạch hướng đến số hóa hoàn toàn [6]. Để đạt được mục đích này, một trong

các bài toán cấp thiết được đặt ra là phải chọn được chuẩn mã hóa nén âm thanh cùng bộ tham số nén phù hợp với tình hình Việt Nam.

Từ những phân tích trên đây, câu hỏi nghiên cứu được đặt ra là xác định điểm tối ưu (còn gọi là sweet pot) - là điểm đảm bảo sự cân đối tốt nhất giữa tỷ lệ nén và chất lượng âm thanh để có thể ứng dụng vào phát thanh số tại Đài TNVN trong tương lai. Đơn vị được lựa chọn thử nghiệm là Cơ quan thường trú của Đài TNVN khu vực *Đồng bằng sông Cửu Long (ĐBSCL)* trụ sở ở thành phố Cần Thơ. Đây là chi nhánh của Đài TNVN được thành lập từ năm 1998, chuyên theo dõi và thông tin về các vấn đề chính trị, kinh tế, xã hội, văn hóa văn nghệ, an ninh quốc phòng trên địa bàn 13 tỉnh thành ở ĐBSCL và đặc biệt có phát sóng chương trình tiếng Khmer phục vụ đồng bào Khmer Nam Bộ.

Hai tiêu chuẩn nén được lựa chọn cho nghiên cứu là MP3 (*Moving Picture Experts Group 1 – Layer 3*) và AAC (*Moving Picture Experts Group 2 – Advanced Audio Coding*), bởi lý do đây là hai trong số các chuẩn nén âm thanh số được tổ chức ISO/IEC ban hành rộng rãi và đã được sử dụng phổ biến trên thế giới. Chuẩn MP3 thường được ứng dụng để truyền tín hiệu âm thanh qua internet và lưu trữ tín hiệu âm thanh trong các thiết bị nghe nhạc bỏ túi. Chuẩn AAC là chuẩn nén tiếp theo MP3, đang được sử dụng trong các thiết bị của Apple, Sony, cửa hàng trực tuyến iTunes, v.v. hoặc là trong chuẩn phát thanh số DAB (*Digital Audio Broadcasting*) [1].

Từ vấn đề nghiên cứu đặt ra, nội dung nghiên cứu sẽ bao gồm các bước chính sau:

- Xây dựng bộ cơ sở dữ liệu gồm các đoạn âm thanh thử nghiệm dạng wave, được trích chọn đa dạng từ các chương trình khác nhau của Đài TNVN ĐBSCL,
- Thực hiện nén các đoạn âm thanh thử nghiệm theo chuẩn nén MP3 và AAC, sau đó giải nén về lại dạng wave,
- Thực hiện thí nghiệm đánh giá các chuẩn nén về hiệu năng,
- Thực hiện thí nghiệm đánh giá các chuẩn nén về chất lượng,
- Phân tích kết quả và đề xuất.

Phần tiếp theo của bài báo được cấu trúc như sau: mục II tìm hiểu nguyên lý mã hóa nén tín hiệu âm thanh số, mục III lần lượt trình bày về các thí nghiệm đánh giá hiệu năng và đánh giá chất lượng các chuẩn nén âm thanh và cuối cùng là kết luận ở mục IV.

II. MÃ HÓA NÉN ÂM THANH SỐ

Về lý thuyết, chúng ta có thể mã hóa âm thanh bằng cách bất chước mã hóa nén tiếng nói: thay vì mã hóa dạng sóng của tín hiệu như kiểu wave, ta tìm cách mô hình hóa cơ chế tạo ra tiếng nói và mã hóa các thông số quan trọng nhất của mô hình, phía giải mã sử dụng các thông số này để kích thích các bộ tạo hàm số và bộ lọc để tổng hợp lại tiếng nói. Tuy nhiên trong thực tế, cách tiếp cận này khó áp dụng hiệu quả cho nén âm thanh. Nguyên nhân là do âm thanh là tín hiệu tạo nên từ nhiều nguồn khác nhau gồm âm thanh của các nhạc cụ như bộ gõ, bộ hơi, bộ dây (do gảy/kéo); âm thanh do các tiếng động khác từ sinh hoạt, sản xuất; âm thanh từ tự nhiên, v.v. nên việc mô hình hóa riêng cho từng nguồn âm thanh là quá phức tạp và không phù hợp.

Vì vậy, đối với nén âm thanh, thay vì tiếp cận dựa trên nguồn phát ra âm thanh, chúng ta chuyển hướng tiếp cận sang dựa vào cơ chế thu nhận âm thanh, tức là dựa vào hệ thống thính giác ở người. Mã hóa theo hướng này được gọi là mã hóa nén cảm quan, là loại mã hóa dựa trên lợi dụng những đặc điểm tâm lý cảm nhận âm thanh ở người.

A. Hệ thống thính giác thu nhận âm thanh ở người

Hệ thống thính giác ở người bao gồm tai và não, thu nhận âm thanh theo hai cơ chế là sinh lý và tâm lý [7].

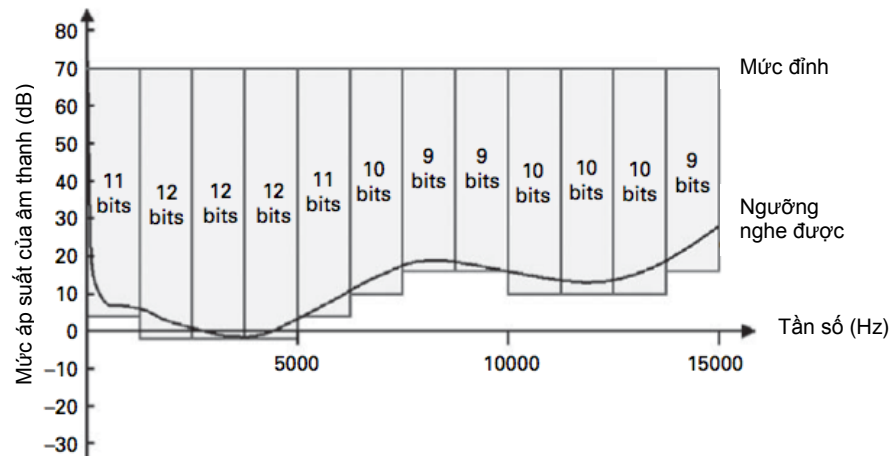
Về mặt sinh lý, tai bao gồm tai ngoài, tai giữa và tai trong. Mỗi bộ phận ở tai đảm nhiệm những chức năng khác nhau: tai ngoài tiếp nhận âm thanh rồi truyền qua ống tai và đưa vào màng nhĩ; tai giữa biến đổi sóng áp suất âm thanh thành chuyển động cơ học tác động lên các xương nhỏ; tai trong chứa ốc tai nằm trong một màng cơ sở, ốc tai chuyển đổi chuyển động cơ học ở xương tai giữa thành chuyển động ở màng cơ sở và kích thích các tế bào thần kinh thính giác ở đây rồi chuyển tín hiệu điện lên não.

Về mặt tâm lý, sự cảm nhận âm thanh ở người liên quan đến một số hiệu ứng tâm lý đặc biệt. Mã hóa nén lợi dụng điều này để nén âm thanh ở những vùng mà tai người cảm nhận âm thanh kém.

Nghiên cứu [7] cho biết một số hiệu ứng tâm lý cảm nhận âm thanh bao gồm:

- Tai cảm nhận âm thanh phụ thuộc vào mức áp lực (cường độ) âm thanh. Mã hóa nén có thể lợi dụng điều này để lượng tử hóa với kích thước bước lượng tử hóa không đồng đều trong toàn bộ dải động của tín hiệu, nghĩa là mã hóa tín hiệu ở vùng biên độ lớn với kích thước bước lớn (số bit mã hóa nhỏ) và ngược lại.
- Trong dải nghe 20Hz-20kHz, độ nhạy của tai khác nhau đối với các thành phần tần số khác nhau. Mã hóa nén lợi dụng điều này để phân chia dải 20Hz-20kHz thành các băng con rồi mã hóa với số bit khác nhau cho mỗi băng con, dẫn đến số bit trung bình giảm xuống như nêu ra trên Hình 1.
- Khi nghe đồng thời hai âm thanh mạnh yếu khác nhau với tần số gần nhau, âm mạnh hơn làm biến dạng ngưỡng nghe tuyệt đối và làm cho tai không nghe được âm nhỏ hơn. Hiệu ứng này gọi là *che tần số*

(*frequency masking*). Tương tự như vậy, nếu âm yếu hơn được phát ra ngay trước hoặc ngay sau âm mạnh hơn thì cũng bị “che”. Hiệu ứng này gọi là *che thời gian (temporal masking)*. Mã hóa nén tiến hành tạo thêm các tín hiệu “che” thời gian và “che” tần số để làm tăng ngưỡng nghe, từ đó giảm số bit mã hóa trong các băng con khác nhau mà vẫn đảm bảo lỗi lượng tử hóa (lúc này bị tăng lên) vẫn ở dưới mức ngưỡng nghe.

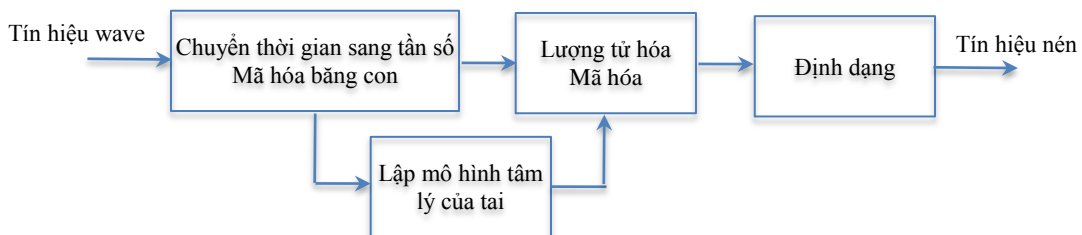


Hình 1. Phân chia băng con và mã hóa với số bit thay đổi

B. Nguyên lý mã hóa nén âm thanh số

Như đã giới thiệu ở Mục I, nghiên cứu này tập trung vào hai chuẩn nén âm thanh số phổ biến MP3 và AAC do tổ chức MPEG thiết kế và được ISO/IEC phê duyệt ban hành rộng rãi. Đây là hai chuẩn nén âm thanh cảm quan chất lượng cao dựa vào mô hình tâm lý của tai người. Về nguyên lý, mã hóa nén cảm quan thực hiện chuyển đổi tín hiệu số định dạng wave thành tín hiệu nén qua các bước xử lý như minh họa trên Hình 2. Cụ thể các bước xử lý như sau [8]:

- Chuyển tín hiệu wave sang miền tần số, sau đó phân chia thành các băng con qua một giàn lọc băng con. Việc làm này là nhằm lợi dụng đặc điểm độ nhạy của tai thay đổi đối với các thành phần tần số khác nhau như trên Hình 1.
- Lập mô hình tâm lý cảm nhận âm thanh của tai bằng cách phân tích tín hiệu vào. Đây là khâu quyết định chất lượng mã hóa nén. Mô hình này giúp xác định tỷ lệ tín hiệu và ngưỡng nghe của mỗi băng con.
- Lượng tử hóa các thành phần phổ sao cho thỏa mãn yêu cầu nhiều lượng tử hóa thấp hơn ngưỡng nghe được trong mỗi băng con.
- Mã hóa các giá trị lượng tử hóa dựa vào mã entropy với bảng mã thay đổi đối với các dải tần khác nhau để đạt hiệu quả nén tốt hơn.
- Cuối cùng là định dạng dòng dữ liệu theo những chuẩn định dạng quy định.



Hình 2. Sơ đồ nguyên lý mã hóa nén âm thanh

Dựa vào nguyên lý này, các chuẩn nén khác nhau có những biến thể khác nhau nhằm cải thiện tỷ lệ nén và chất lượng nén. Phần sau trình bày cụ thể hơn về các biến thể áp dụng cho MP3 và AAC:

1. Mã hóa MP3

MP3 thuộc nhóm mã hóa nén âm thanh cảm quan MPEG-1. MPEG-1 hoạt động ở ba chế độ khác nhau gọi là *lớp (layer)*, với mức độ phức tạp và hiệu quả tăng dần từ lớp 1 đến lớp 3. MP3 là nhóm MPEG-1 lớp 3. Đây là nhóm MPEG-1 phức tạp nhất, cung cấp chất lượng âm thanh gần với chất lượng CD với tỷ lệ nén âm thanh lên đến 12 lần [1]. Bộ mã hóa nén MP3 từ tín hiệu wave có một số đặc điểm như sau:

- Trước tiên tín hiệu wave được chia thành 32 băng con bằng giàn lọc, sau đó đưa đến *bộ biến đổi cosin rời rạc MDCT (Modified Discrete Cosine Transform)* để tiếp tục phân chia thành 576 băng con nhằm đạt độ phân giải tốt hơn trong miền tần số.

- Để ước lượng ngưỡng nghe tốt hơn, trước khi lập mô hình tâm lý cảm nhận âm thanh của tai, bộ mã hóa MP3 tiến hành ánh xạ từ miền thời gian sang miền tần số bằng *phép biến đổi Fourier nhanh FFT (Fast Fourier Transform)* 1024 điểm, để giúp phân giải tần số tốt hơn.
- Việc lượng tử hóa các thành phần phổ được thực hiện dựa vào lượng tử hóa không đều. Hệ số nén được chọn trước khi lượng tử hóa bằng cách dùng hai vòng lặp lồng vào nhau: vòng lặp trong hiệu chỉnh tăng dần kích thước bước lượng tử hóa, nhằm giảm dần số mức lượng tử hóa dẫn đến giảm số bit, vòng lặp ngoài hiệu chỉnh tăng số mức lượng tử hóa dẫn đến đảm bảo chất lượng mã hóa.
- Việc mã hóa các giá trị lượng tử hóa được thực hiện dựa vào mã Huffman. Đối với mỗi thành phần phổ khác nhau, bảng mã Huffman tối ưu khác nhau được chọn ra từ trong số nhiều bảng mã Huffman để thích nghi tốt hơn với đặc trưng thống kê của âm thanh trong từng thành phần phổ.
- Dòng bit MP3 được định dạng theo từng khung, đầu khung là header bao gồm từ mã đồng bộ, tốc độ bit, tần số lấy mẫu, lớp, mode mã hóa (mono, stereo,...).

2. Mã hóa AAC

AAC là chuẩn mã hóa thuộc nhóm MPEG-2. AAC có kiến trúc tương tự như MP3 nhưng khác với MP3 ở chỗ AAC dùng phương pháp modul hóa và phát triển thêm nhiều công cụ mã hóa mới, giúp cải thiện chất lượng âm thanh. Sau đây là một số công cụ phát triển thêm trong AAC so với MP3 [1]:

- AAC tăng kích thước giàn lọc trong MP3 từ 576 lên 1024 nhằm tăng độ phân giải tần số so với MP3.
- Bổ sung bộ tạo nhiễu trong miền thời gian dựa vào công nghệ *TNS (Temporal Noise Shaping)* nhằm thay đổi ngưỡng nghe để giảm số bit mã hóa.
- Bổ sung bộ dự đoán để tăng tỷ lệ nén bằng cách hướng cho bộ lượng tử hóa tập trung vào các mẫu tín hiệu đáng quan tâm.
- Dữ liệu trong AAC được định dạng thành từng khung như MP3, nhưng từ mã đồng bộ và tham số mã hóa thay đổi tùy theo từng ứng dụng cụ thể.

III. ĐÁNH GIÁ CÁC CHUẨN NÉN

Việc đánh giá các chuẩn nén dựa vào hai tiêu chuẩn chính là hiệu năng và chất lượng, được tiến hành trên bộ cơ sở dữ liệu xây dựng từ các chương trình phát thanh thực tế của Đài TNVN khu vực ĐBSCL.

A. Xây dựng cơ sở dữ liệu

Cơ sở dữ liệu xây dựng để đánh giá các chuẩn nén bao gồm hai tập dữ liệu:

- Tập dữ liệu thứ nhất bao gồm 18 tệp âm thanh như mô tả một phần trong Bảng 1. Các tệp âm thanh này được trích ra từ các chương trình khác nhau của Đài; xuất phát từ đa dạng nguồn gốc khác nhau bao gồm tiếng nói, âm nhạc các loại và hỗn hợp tiếng nói và âm nhạc; được phát trên các sóng điều chế khác nhau với các tần số khác nhau. Tổng dung lượng tập thứ nhất là 5.546MB, tương đương thời lượng là hơn 8 giờ (650MB/giờ). Tập thứ nhất được dùng để thử nghiệm các chuẩn nén với tỷ lệ nén thay đổi nhằm đánh giá hiệu năng của các chuẩn nén và xác định mối liên quan giữa thời gian thực thi và tỷ lệ nén.
- Tập dữ liệu thứ hai bao gồm 3 đoạn âm thanh các loại định dạng wave gốc và 24 đoạn nén/giải nén âm thanh gốc với các tỷ lệ nén khác nhau. Chuẩn nén và các tham số nén đối với tập thứ hai được nêu ra trong Bảng 2. Tập này được dùng để đánh giá chất lượng nén dựa vào phương pháp đánh giá chủ quan là nghe thử. Mỗi đoạn âm thanh có thời lượng giới hạn là 20 giây. Việc hạn chế thời gian này là nhằm đảm bảo khả năng tập trung cao của người tham gia thí nghiệm.

Toàn bộ cơ sở dữ liệu được lưu ở địa chỉ sau:

<https://drive.google.com/drive/folders/1FyX47938ErWQHxFUIKDBmEfZupxDFqs>

Bảng 1. Cơ sở dữ liệu âm thanh – Một phần tập dữ liệu thứ nhất

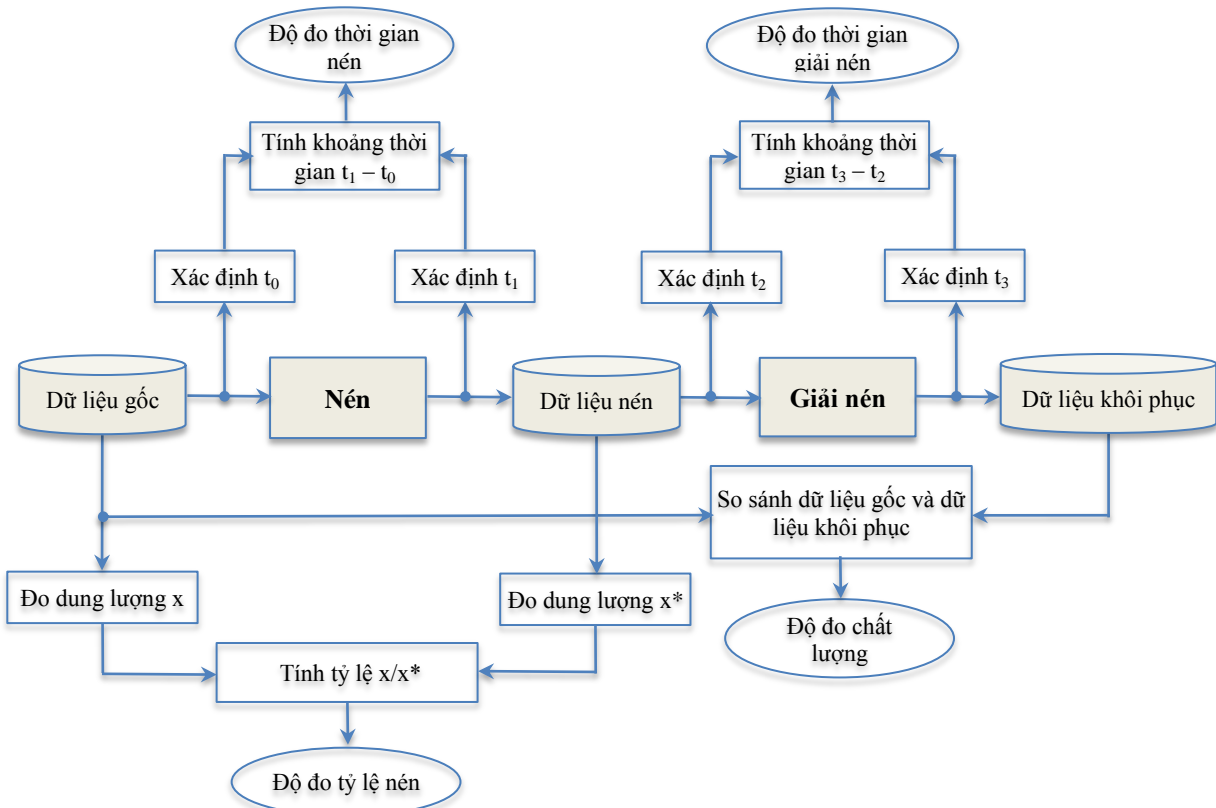
Tên tệp *.wav	Chương trình	Kênh	Phát sóng	Dung lượng	Thể loại âm thanh
01.wav	Chương trình thời sự hệ 1	VOV1	AM 711kHz	244MB	Tiếng nói
03.wav	Đọc truyện dài kỳ	VOV2	AM 558kHz	326MB	Tiếng nói
08.wav	Chiến sĩ miền Tây	VOV4	AM 873kHz	304MB	Tiếng nói
10.wav	Dân ca các miền lời cổ	VOV3	AM 783kHz	302MB	Nhạc dân ca
11.wav	Thính phòng giao hưởng	VOV3	AM 783kHz	380MB	Nhạc cổ điển
12.wav	Câu chuyện VPOP	VOV3	AM 783kHz	290MB	Nhạc POP
14.wav	Dân ca cổ nhạc	VOV4	AM 873kHz	604MB	Ca cổ
15.wav	Ca nhạc theo yêu cầu	VOV4	AM 873kHz	302MB	Nhạc trẻ
18.wav	Quà tặng âm nhạc	VOV Giao thông ĐBSCL	FM 90MHz	309MB	Tiếng nói + nhạc

Bảng 2. Cơ sở dữ liệu âm thanh - Tập dữ liệu thứ hai

Tốc độ nén	Tỷ lệ nén	Chuẩn nén	
		MP3	AAC
32kbps	44:1	3 đoạn	3 đoạn
64kbps	22:1	3 đoạn	3 đoạn
128kbps	11:1	3 đoạn	3 đoạn
192kbps	7:1	x	3 đoạn
256kbps	5,5:1	3 đoạn	x

B. Quy trình thí nghiệm đánh giá các chuẩn nén

Để đánh giá các chuẩn nén, quy trình thí nghiệm được thiết kế như trong Hình 2, với các khối hình chữ nhật thể hiện các bước xử lý/đo đạc, các khối hình trụ tròn thể hiện dữ liệu âm thanh và các ô hình bầu dục là kết quả các độ đo.



Hình 3. Quy trình thí nghiệm đánh giá các chuẩn nén

1. Thí nghiệm đánh giá hiệu năng nén

Ở thí nghiệm đánh giá hiệu năng nén, đối với mỗi tệp âm thanh rút ra từ tập dữ liệu thứ nhất, ta áp dụng lần lượt các chuẩn MP3 và AAC để nén dữ liệu, sau đó thực hiện giải nén để thu được dữ liệu khôi phục ở dạng wave.

Để đo thời gian nén, ta tiến hành ghi lại thời điểm trước khi nén t_0 và sau khi nén t_1 , từ đó xác định được độ đo thời gian nén $t_1 - t_0$. Tiến hành tương tự cho quá trình giải nén để xác định độ đo thời gian giải nén $t_3 - t_2$.

Để đo tỷ lệ nén, tiến hành xác định dung lượng tệp âm thanh trước khi nén x và sau khi nén x^* , từ đó xác định tỷ lệ nén là tỷ số giữa hai dung lượng x/x^* .

2. Thí nghiệm đánh giá chất lượng nén

Chất lượng nén được đánh giá dựa vào việc so sánh dữ liệu gốc với dữ liệu khôi phục. Việc so sánh này được ghi trong khối “So sánh dữ liệu nén và dữ liệu khôi phục” trên Hình 2. Để đánh giá chất lượng nén có thể dùng phương pháp chủ quan hoặc khách quan. Đánh giá khách quan là phương pháp dựa vào tỷ số tín hiệu trên nhiễu SNR (Signal to Noise Ratio). Tuy nhiên mã hóa cảm quan cố ý tạo thêm nhiễu che trong miền thời gian và miền tần số để thay đổi ngưỡng nghe nhằm giảm số bit mã hóa, nên có thể dẫn đến SNR thấp. Vì vậy việc đánh giá dựa vào SNR có thể dẫn đến kết quả không phù hợp.

Cho đến nay thì đánh giá chủ quan bằng cách nghe thử vẫn là phương pháp đơn giản và hiệu quả để đánh giá chất lượng của các thuật toán mã hóa âm thanh khác nhau. Tổ chức ITU-R (International Telecommunications Union,

Radiocommunications sector) cùng với nhóm MPEG phụ trách âm thanh đã đề xuất một loạt các quy tắc nghiêm ngặt để đánh giá chất lượng nén bằng cách nghe thử [9].

Trong nghiên cứu này, với nỗ lực tối đa, thí nghiệm nghe thử được tiến hành theo cách thức và các điều kiện gần giống với các quy định về đánh giá chủ quan [9]. Thí nghiệm được thực hiện trên tập dữ liệu thứ hai như mô tả trong Bảng 2. Các bước chuẩn bị và các quy tắc cho thí nghiệm bao gồm:

- Tổ chức dữ liệu thành các thư mục chứa hai tệp âm thanh, trong đó có một tệp dạng wave gốc và một tệp dữ liệu khôi phục ở dạng wave. Việc dùng chung định dạng wave nhằm giúp cho người nghe dùng chung một phần mềm xuất âm thanh ra loa, nhằm đảm bảo sự công bằng trong khi đánh giá chất lượng.
- Để tránh quán tính của người nghe là cho điểm chất lượng tệp nén thấp hơn tệp gốc, bố trí một số thư mục chứa hai tệp âm thanh đều là tệp gốc dạng wave.
- Trong khi thí nghiệm, để tránh sự phân tâm, yêu cầu chỉ có một người tham gia thí nghiệm ở trong phòng cùng với nhân viên hỗ trợ kỹ thuật khi có sự cố.
- Trong khi thí nghiệm, tắt tất cả các chương trình đang chạy trên máy tính, tắt điện thoại di động nhằm tránh ảnh hưởng đến âm thanh nghe thử (nếu có).
- Mọi người tham gia đều nghe thử âm thanh trên cùng một máy tính, dùng cùng loại tai nghe. Mặc dù điều này sẽ khiến kéo dài thời gian thí nghiệm rất nhiều (do không thể tiến hành thí nghiệm đồng thời cho nhiều người) nhưng giúp đảm bảo sự công bằng trong khi đánh giá chất lượng.
- Người tham gia được yêu cầu lắng nghe thật tập trung từng thư mục gồm một tệp âm thanh gốc và một tệp kiểm tra (có thể nén hoặc gốc), sau đó cho điểm chất lượng theo thang điểm: 1- chất lượng không chấp nhận được (âm thanh kiểm tra khác nhiều so với âm thanh gốc), 2- chất lượng chấp nhận được (âm thanh kiểm tra khác một ít so với âm thanh gốc), 3- chất lượng tốt (âm thanh kiểm tra gần giống hoặc không phân biệt được so với âm thanh gốc).
- Người nghe có thể nghe lại nhiều lần, có thể nghỉ giải lao tùy ý trong khi làm thí nghiệm nhưng không được mang theo thức ăn vào phòng thí nghiệm.

C. Kết quả đánh giá hiệu năng

Như đã nêu trên, hiệu năng của chuẩn nén được đánh giá dựa vào các tiêu chí về tỷ lệ nén x/x^* và thời gian thực thi. Ở đây thời gian thực thi bao gồm cả thời gian nén $t_1 - t_0$ và thời gian giải nén $t_3 - t_2$.

- Trước tiên, với mỗi tệp âm thanh trong cơ sở dữ liệu thứ nhất, tiến hành thí nghiệm theo quy trình như ở Hình 2 nhằm đo thời gian thực thi và tỷ lệ nén, sử dụng lần lượt các chuẩn nén MP3 và AAC với tốc độ cao nhất là 256kbps và tần số lấy mẫu 44.1kHz. Như vậy, tổng số lần thực thi nén/giải nén là 18 lần tương ứng với 18 tệp âm thanh. Kết quả thu được là tỷ lệ nén trung bình đối với MP3 là 4.46:1, độ lệch chuẩn là 0.14:1 và đối với AAC là 4.66:1 với độ lệch chuẩn là 0.40:1. Độ lệch chuẩn khá thấp trong khoảng thời gian dài chứng tỏ sự ổn định của chuẩn nén khá tốt.
- Chọn ra hai tệp âm thanh trong tập dữ liệu thứ nhất gồm một tệp tiếng nói (08.wav) và một tệp âm nhạc (15.wav). Thực hiện thí nghiệm như vừa nêu trên, sử dụng tần số lấy mẫu là 44,1kHz, với tốc độ thay đổi lần lượt 64kbps, 96kbps, 128kbps và 256kbps đối với MP3 và 64kbps, 96kbps, 128kbps và 192kbps đối với AAC. Kết quả thí nghiệm được thể hiện ở Bảng 3. Từ đây suy ra thời gian thực thi nén của hai chuẩn nén chênh lệch không đáng kể, thời gian thực thi trung bình khoảng 70 giây cho 300MB dữ liệu, chưa đến nửa giây cho 1MB dữ liệu. Như vậy khả năng nén/giải nén thời gian thực là hoàn toàn khả thi. Ngoài ra, có vẻ như thời gian thực thi nén không liên quan với tỷ lệ nén: trong thí nghiệm này, với chuẩn MP3 thì thời gian thực thi tốt nhất cho nén tiếng nói tương ứng với tỷ lệ nén là 11:1 nhưng đối với âm nhạc lại là 22:1, với chuẩn AAC thì thời gian thực thi tốt nhất tương ứng với tỷ lệ nén là 14:1.
- Tiếp tục thử nghiệm ảnh hưởng của tần số lấy mẫu đến thời gian nén thì nhận thấy rằng: muốn giảm thời gian nén, có thể xem xét giảm tần số lấy mẫu. Thí nghiệm trên tệp 08.wav cho thấy khi tần số lấy mẫu giảm một nửa (nghĩa là giảm từ 44,1kHz xuống còn 22,05kHz) thì thời gian thực thi trung bình giảm còn 38 giây cho 300MB dữ liệu (đối với tần số 44,1kHz thì thời gian này là hơn 1 phút 10 giây).

D. Kết quả đánh giá chất lượng

Như đã nêu trên, thí nghiệm đánh giá chất lượng được thực hiện bằng phương pháp nghe thử. Để có kết quả toàn diện hơn, thí nghiệm được tiến hành với hai nhóm tình nguyện viên khác nhau gồm nhóm không chuyên và nhóm chuyên gia về âm thanh.

Cụ thể là nhóm tình nguyện viên thứ nhất gồm 25 sinh viên Trường Cao đẳng Quốc tế TP. Hồ Chí Minh. Thí nghiệm diễn ra tại một căn phòng cách âm tốt đặt tại Trường ở địa chỉ 460 D Kinh Dương Vương, phường An Lạc, quận Bình Tân, TP.HCM. Nhóm tình nguyện viên thứ hai là 10 chuyên viên của Đài TNVN khu vực ĐBSCL. Thí nghiệm diễn ra ở phòng thu âm chuyên dụng của Đài đặt tại 102 Lý Tự Trọng, quận Ninh Kiều, TP. Cần Thơ. Kết quả thu được từ hai nhóm tình nguyện viên được thể hiện cụ thể trong các Bảng 4 và 5.

Bảng 3. Kết quả đánh giá hiệu năng các chuẩn nén

File *.wav	Dung lượng gốc	MP3			AAC		
		Dung lượng nén	Tỷ lệ nén	Thời gian thực thi	Dung lượng nén	Tỷ lệ nén	Thời gian thực thi
08 (tiếng nói)	304MB	13,8MB	22:1	1'10"	14,4MB	20,9:1	1'11"
		20,7MB	14,6:1	1'10"	21,3MB	14,1:1	1'10"
		27,6MB	11:1	1'09"	28,2MB	11:1	1'12"
		55,3MB	5,5:1	1'10"	42,1MB	7,17:1	1'16"
15 (âm nhạc)	302MB	13,7MB	22:1	1'08"	14,3MB	22:1	1'07"
		20,6MB	14,6:1	1'12"	21,2MB	14,6:1	1'07"
		27,4MB	11:1	1'12"	28,1MB	11:1	1'10"
		54,9MB	5,5:1	1'13"	41,9MB	7,2:1	1'13"

Bảng 4. Điểm đánh giá chất lượng âm thanh nén MP3

Loại dữ liệu	Tỷ lệ nén 44:1			Tỷ lệ nén 22:1			Tỷ lệ nén 11:1			Tỷ lệ nén 5,5:1		
	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB
1	1,72	1,00	1,36	2,32	1,90	2,11	2,56	2,50	2,53	2,84	3,00	2,92
2	1,68	1,60	1,64	1,68	1,90	1,79	1,96	2,40	2,18	2,32	2,30	2,31
3	1,24	1,00	1,12	2,00	2,10	2,05	2,40	2,40	2,4	2,68	2,70	2,69
Tổng	1,37			1,98			2,37			2,64		

Bảng 5. Điểm đánh giá chất lượng âm thanh nén AAC

Loại dữ liệu	Tỷ lệ nén 44:1			Tỷ lệ nén 22:1			Tỷ lệ nén 11:1			Tỷ lệ nén 7:1		
	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB	Nhóm 1	Nhóm 2	TB
1	1,08	1,10	1,09	2,20	2,00	2,10	2,72	2,80	2,76	2,64	2,70	2,67
2	1,80	1,90	1,85	2,32	2,10	2,21	2,56	2,60	2,58	2,64	2,60	2,62
3	1,08	1,30	1,19	2,28	2,30	2,59	2,60	2,60	2,60	2,76	2,90	
Tổng	1,38			2,30			2,65			2,83		

Trong Bảng 4 và 5, cột điểm "TB" chính là trung bình cộng của hai điểm của hai nhóm tình nguyện viên thứ nhất và thứ hai và điểm "Tổng" là điểm trung bình của tất cả các tệp âm thanh thí nghiệm. Theo lý thuyết thì việc tính điểm "TB" là trung bình cộng điểm của Nhóm 1 và Nhóm 2 là không hợp lý, do số lượng thành viên của nhóm 2 chỉ bằng 40% số thành viên nhóm 1. Tuy nhiên ở đây bỏ qua điều này là vì một số lý do sau đây:

- Điều kiện thí nghiệm của nhóm 2 tốt hơn so với nhóm 1 do phòng thí nghiệm của nhóm 2 là phòng thu âm chuyên dụng. Do đó kết quả đánh giá của nhóm 2 có khả năng chính xác hơn.
- Nhóm 2 gồm các chuyên viên chuyên về kỹ thuật âm thanh của Đài TNVN nên có kinh nghiệm tốt hơn trong việc thẩm định chất lượng âm thanh. Điều này thể hiện rõ qua kết quả thí nghiệm: ví dụ như nhóm 2 cho điểm tuyệt đối đối với các tệp âm thanh gốc được trộn vào dữ liệu thí nghiệm (trung bình nhóm 1 cho 2.8 điểm, nhóm 2 cho 3 điểm).

Quan sát Bảng 4 và Bảng 5 có thể có một số nhận xét như sau:

- Dựa vào điểm tổng nhận thấy: với cùng một tỷ lệ nén, chuẩn AAC luôn đạt được chất lượng cao hơn MP3 cho tất cả các loại dữ liệu. Ta cũng nhận thấy chất lượng AAC ứng với tỷ lệ nén 22:1 tương đương với nén MP3 tỷ lệ 11:1 và AAC ứng với tỷ lệ 11:1 có chất lượng tương đương với MP3 tỷ lệ 5.5:1.
- Nén AAC đạt chất lượng tốt hơn nhiều so với MP3 ở cùng tỷ lệ nén và thời gian thực thi AAC nhìn chung không chênh lệch nhiều so với MP3 (theo Bảng 3). Do vậy AAC được đề nghị nên sử dụng hơn so với MP3 trong phát thanh.
- Theo thang điểm đánh giá chất lượng sử dụng trong thí nghiệm thì điểm từ 2 trở lên là đảm bảo sai khác giữa tín hiệu khôi phục và tín hiệu gốc không đáng kể. Như vậy, với chuẩn nén AAC, tỷ lệ nén 22:1 là có thể chấp nhận được nhưng với chuẩn nén MP3 thì tỷ lệ này phải là 11:1.
- Chất lượng tín hiệu nén tăng khi hạ sinh tỷ lệ nén. Điều này đặc biệt thấy rõ khi giảm tỷ lệ nén từ 4:1 xuống 22:1 ở cả AAC và MP3. Tuy nhiên khi tiếp tục giảm tỷ lệ nén hơn nữa thì chất lượng gia tăng không đáng kể. Chẳng hạn như với AAC khi giảm tỷ lệ nén từ 44:1 xuống 22:1 thì chất lượng tăng thêm 30% so với điểm tối đa, nhưng khi giảm tỷ lệ nén từ 22:1 xuống 11:1 thì chất lượng chỉ tăng 12%, khi giảm tỷ lệ nén từ 11:1 xuống 7:1 thì chất lượng tăng ít hơn, chỉ tăng 6%. Từ đây ta có thể xác định được điểm sweet-pot đối với AAC là tỷ lệ nén 11:1, là điểm cân đối tốt giữa tỷ lệ nén và chất lượng. Tương tự, điểm sweet-pot cho MP3 là 5,5:1.

Tóm lại, qua thử nghiệm, một số ý kiến sau đây được đề xuất cho việc ứng dụng vào Đài TNVN giai đoạn số hóa hoàn toàn:

- Nên áp dụng chuẩn nén AAC vào nén âm thanh số.
- Nên áp dụng tỷ lệ nén AAC là 11:1 vào nén âm thanh số để có chất lượng tốt. Trong trường hợp cần tiết kiệm hơn nữa dung lượng lưu trữ, thời gian truyền, băng thông truyền thì có thể xem xét nén AAC ở tỷ lệ 22:1.

IV. KẾT LUẬN

Tóm lại, bài báo đã thực hiện đánh giá hai chuẩn nén âm thanh số chất lượng cao phổ biến là MP3 và AAC, áp dụng thử nghiệm trên một số chương trình phát thanh của Đài TNVN thường trú Cần Thơ. Quá trình đánh giá bao gồm đánh giá hiệu năng dựa vào tiêu chí thời gian thực thi và tỷ lệ nén và đánh giá chất lượng dựa vào nghe thử. Nghiên cứu được thực hiện trên cơ sở dữ liệu xây dựng bằng việc trích chọn phong phú về thể loại âm thanh, đảm bảo dung lượng dữ liệu đủ lớn từ nhiều chương trình khác nhau của Đài. Từ các thí nghiệm, bài báo đã rút ra một số đề xuất như sau: (1) chuẩn nén âm thanh nên được áp dụng là AAC và (2) tỷ lệ nén nên được áp dụng là từ 11:1 đến 22:1. Trong trường hợp dùng MP3 thì tỷ lệ nén ưu tiên nên chọn là từ 5.5:1 đến 11:1. Đây là những kết quả ban đầu để tiếp tục phát triển các nghiên cứu thử nghiệm trong tương lai với số lượng dữ liệu lớn hơn, số lượng người tham gia nhiều hơn và trên quy mô rộng rãi hơn, nhằm xác định chính xác nhất chuẩn nén và tập tham số phù hợp cho việc số hóa hoàn toàn dịch vụ phát thanh ở Việt Nam như kế hoạch [10].

V. LỜI CẢM ƠN

Chúng tôi trân trọng cảm ơn các chuyên viên công tác tại Đài Tiếng nói Việt Nam thường trú Cần Thơ, đặc biệt là Cán bộ quản lý phòng Kỹ Thuật đã nhiệt tình hỗ trợ xây dựng cơ sở dữ liệu và cơ sở vật chất cho nghiên cứu. Chúng tôi gửi lời cảm ơn sâu sắc đến các chuyên viên tại Đài và sinh viên Trường Cao đẳng Quốc tế Thành phố Hồ Chí Minh đã nhiệt tình tham gia vào thí nghiệm đánh giá chất lượng của các chuẩn nén.

TÀI LIỆU THAM KHẢO

- [1] J-N Hwang, “Multimedia Networking: From Theory to Practice”, Cambridge University, 2009.
- [2] E. Lisa, “An Introduction to the Digital Media Industry”, INKinspire non-profit Organization, 2018.
- [3] “Wave PCM soundfile format”, URL: <https://ccrma.stanford.edu/courses/422/projects/WaveFormat/>
- [4] Bernard Sklar, “Digital Communications: Fundamentals and Applications”, Prentice Hall, 2nd edition, 2017.
- [5] J. D. Gibson, “Multimedia Communications: Directions and Innovations, Communication, Networking and Multimedia Series”, Academic Press, 2000.
- [6] Bộ Thông tin và Truyền thông, “Mã hóa các đối tượng hình ảnh âm thanh- Tương tác giữa âm thanh và các hệ thống”, Tài liệu tiêu chuẩn kỹ thuật, 2016.
- [7] T. Painter and A. Spanias, “Perceptual coding of digital audio,” Proc. IEEE, no. 88(4), pp. 451–515, 2000.
- [8] P. Nolls, “MPEG digital audio coding”, IEEE Signal Processing Magazine, no. 14(5), pp. 59–81, 1997.
- [9] Stephen Bunting, “A subjective comparison of MPEG-4 AAC codecs”, 4B Technical Project 2004.
- [10] Đài Tiếng nói Việt Nam, “Nghiên cứu và ứng dụng công nghệ phát thanh số tại Việt Nam”, Đề tài KHCN cấp Nhà nước mã số KC 01.17, 2005.

TESTING RESULTS OF THE APPLICATION OF DIGITAL AUDIO TECHNIQUES TO THE VOICE OF VIETNAM NATIONAL STATION

Nguyen Thanh Phong, Hoang Le Uyen Thuc

ABSTRACT: At the Voice Of Vietnam national station (VOV), the digital audio channels are currently encoded by uncompressed audio codings in wave format. Naturally, the size of wave format-based audio files is extremely large, so the capacity to store the radio programs is very high and the manipulation on audio file is very complex. This study aims to experimentally validate the application of audio compression standards to The VOV, preparing for the deployment of audio compression coding as planned. Two high-quality lossy audio compression techniques including MP3 (Moving Picture Experts Group 1- Layer 3) and AAC (Moving Picture Experts Group 2- Advanced Audio Coding) are evaluated for performance and quality. MP3 and AAC are sequentially applied to 18 audio files extracted from different instant VOV programs during over 8 hours, and then are measured based on performance metrics including processing time and compression ratio. The audio quality is evaluated subjectively based on listening tests. Listening tests are implemented on 24 multi-genre audio files with a participation of 35 volunteers. The experimental results show the feasibility of the application of MP3 and AAC techniques to The VOV. Based on the experimental results, the most appropriate parameter values of new audio techniques are also determined in order to save the storage capacity and the signal manipulation time while ensuring the required quality of audio service.