



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0051257

(51)^{2020.01} G10L 19/008

(13) B

(21) 1-2021-05445

(22) 10/02/2020

(86) PCT/EP2020/053264 10/02/2020

(87) WO2020/177981 10/09/2020

(30) 19305254.5 05/03/2019 EP

(45) 25/09/2025 450

(43) 25/02/2022 407A

(73) ORANGE (FR)

111, quai du Président Roosevelt, 92130 Issy-les-Moulineaux, France

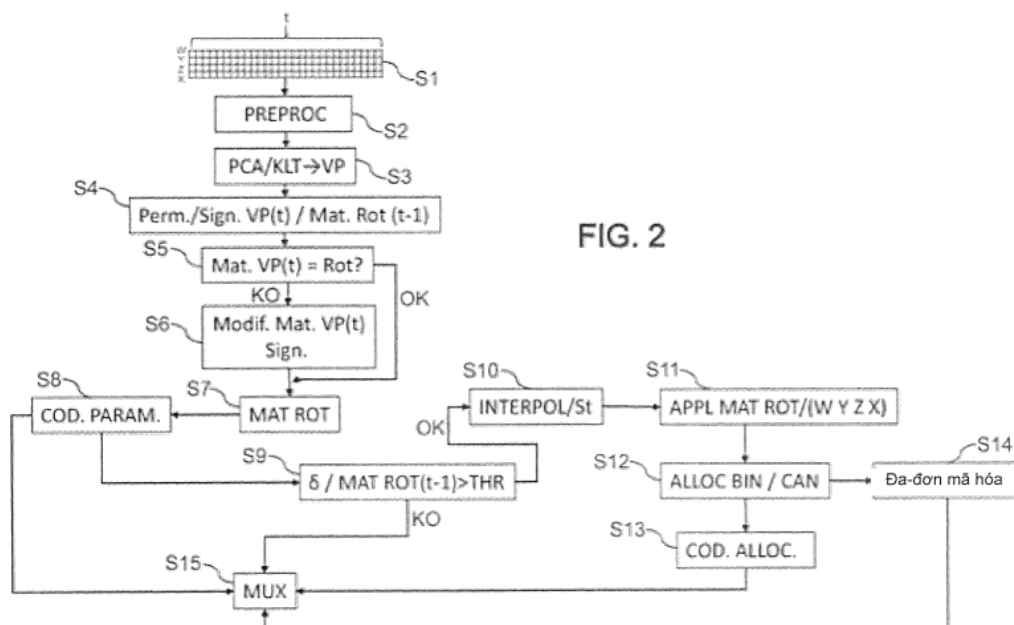
(72) RAGOT, Stéphane (FR); MAHE, Pierre (FR).

(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP MÃ HÓA ĐỐI VỚI VIỆC NÉN CÁC TÍN HIỆU AUDIO, THIẾT BỊ MÃ HOÁ VÀ THIẾT BỊ GIẢI MÃ

(21) 1-2021-05445

(57) Sáng chế đề cập đến phương pháp mã hóa đối với việc nén các tín hiệu audio, thiết bị mã hóa và thiết bị giải mã. Phương pháp mã hóa đối với việc nén các tín hiệu audio tạo ra, theo thời gian, sự liên tiếp của các khung mẫu, ở mỗi trong số N kênh của biểu diễn ambisonic của bậc cao hơn 0, phương pháp này bao gồm các bước: tạo ra, dựa trên các kênh và đối với khung hiện thời, ma trận hiệp biến liên kênh, và tìm kiếm (S3) cho các vector đặc trưng của ma trận hiệp biến nêu trên với việc xem xét để thu ma trận của các vector đặc trưng, kiểm tra (S5) ma trận của các vector đặc trưng để xác minh rằng nó biểu diễn phép quay trong không gian N chiều, và nếu không, hiệu chỉnh (S6) ma trận của các vector đặc trưng cho đến khi ma trận quay thu được, đối với khung hiện thời, và áp dụng ma trận quay nêu trên (S7) cho các tín hiệu của N kênh trước khi mã hóa kênh riêng biệt các tín hiệu nêu trên.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc mã hóa/giải mã dữ liệu audio theo không gian, cụ thể là trong ngữ cảnh ambiophonic (sau đây được gọi là "ambisonic").

Tình trạng kỹ thuật của sáng chế

Các bộ mã hóa/bộ giải mã (sau đây được gọi là "bộ mã hóa-giải mã") hiện đang được sử dụng trong điện thoại di động là mono (một kênh tín hiệu đơn lẻ để tái tạo trên một loa đơn lẻ). Bộ mã hóa-giải mã 3GPP EVS (cho "Dịch vụ thoại nâng cao") giúp nó có thể cung cấp chất lượng "Super-HD" (còn được gọi là giọng nói "độ phân giải cao +" hoặc HD +) với dải tần audio dải siêu rộng (super-wideband, SWB) cho các tín hiệu được lấy mẫu tại 32 hoặc 48 kHz hoặc toàn dải (full-band, FB) cho các tín hiệu được lấy mẫu ở 48 kHz; băng thông audio từ 14,4 đến 16 kHz ở chế độ SWB (9,6 đến 128 kbps) và 20 kHz ở chế độ FB (16,4 đến 128 kbps).

Sự cải tiến tiếp theo về chất lượng trong các dịch vụ đàm thoại do các nhà khai thác cung cấp nên bao gồm các dịch vụ nhập vai, sử dụng các thiết bị đầu cuối như điện thoại thông minh, ví dụ, được trang bị một số micrô hoặc thiết bị cho hội nghị audio không gian hóa hoặc hội nghị video kiểu truyền hình từ xa, hoặc thậm chí các công cụ để chia sẻ nội dung "trực tiếp", với kết xuất audio 3D theo không gian, nhập vai hơn nhiều so với tái tạo âm thanh nổi 2D đơn giản. Với việc nghe nội dung trên điện thoại di động bằng tai nghe audio ngày càng phổ biến và sự xuất hiện của thiết bị audio tiên tiến (phụ kiện như micrô 3D, hỗ trợ giọng nói với anten âm thanh, tai nghe thực tế ảo, v.v.) và các công cụ cụ thể (ví dụ, để sản xuất nội dung video 360°), việc ghi lại và kết xuất các cảnh âm thanh theo không gian hiện đã đủ phổ biến để mang lại trải nghiệm giao tiếp nhập vai.

Để đạt được điều này, tiêu chuẩn 3GPP tương lai "TVAS" (đối với "các dịch vụ giọng nói và audio nhập vai") đề xuất việc mở rộng bộ mã hóa-giải mã EVS để bao gồm nhập vai, bằng cách chấp nhận, dưới dạng các định dạng đầu vào cho bộ

mã hóa-giải mã, ít nhất các định dạng audio theo không gian được liệt kê sau đây (và các kết hợp của chúng):

- Định dạng đa kênh kiểu âm thanh nổi ("dựa trên kênh"), 5.1, trong đó mỗi kênh cấp cho loa (ví dụ L và R trong âm thanh nổi, hoặc L, R, Ls, Rs và C trong 5.1)

- Định dạng dựa trên đối tượng trong đó các đối tượng audio được mô tả là tín hiệu audio (thường là đơn) được liên kết với siêu dữ liệu mô tả các thuộc tính của đối tượng này (vị trí trong không gian, chiều rộng theo không gian của nguồn, v.v.), và

- Định dạng ambisonic ("dựa trên khung cảnh") mô tả trường âm thanh ở điểm đã cho, thường được thu bởi micro hình cầu hoặc được tổng hợp trong miền của sóng hài hình cầu.

Sau đây, ta thường quan tâm đến việc mã hóa âm thanh trong định dạng ambisonic, theo một phương án làm ví dụ (ít nhất một số khía cạnh được thể hiện liên quan đến sáng chế sau đây cũng có thể được áp dụng cho các định dạng ngoài ambisonic).

Ambisonics là phương pháp ghi ("mã hóa" theo nghĩa âm thanh) âm thanh theo không gian, và hệ thống tái tạo ("giải mã" theo nghĩa âm thanh). Micro ambisonic (bậc thứ nhất) bao gồm ít nhất bốn vỏ bọc (điển hình là loại cardioid hoặc sub-cardioid) được bố trí trên lưới hình cầu, ví dụ, các đỉnh của khối tứ diện đều. Các kênh audio được liên kết với vỏ bọc này được gọi là "Định dạng A". Định dạng này được chuyển đổi thành "Định dạng B", trong đó trường âm thanh được chia thành bốn thành phần (sóng hài hình cầu) được biểu thị W, X, Y, Z, tương ứng với bốn micro ảo trùng nhau. Thành phần W tương ứng với việc thu đẳng hướng của trường âm thanh, trong khi các thành phần X, Y, và Z, có tính định hướng hơn, so với các gradient áp lực được định hướng trong ba chiều không gian. Hệ thống ambisonic là hệ thống linh hoạt theo nghĩa là việc ghi hoặc tái tạo là riêng biệt và tách rời. Hệ thống này cho phép việc giải mã (theo nghĩa âm thanh) trong cấu hình loa bất kỳ (ví dụ, âm thanh vòm loại 5.1, hai tai hoặc periphonic loại 7.1.4 (có chiều cao). Tất nhiên, phương thức ambisonic có thể được tổng quát hóa cho nhiều

hơn bốn kênh ở định dạng B và sự biểu diễn tổng quát hóa này được gọi là "HOA" (đối với "Ambisonics bậc cao"). Thực tế rằng âm thanh được phá vỡ thành nhiều sóng hài hình cầu hơn để tăng cường độ chính xác theo không gian của việc tái tạo khi kết xuất trên các loa phóng thanh.

Tín hiệu ambisonic bậc N bao gồm các thành phần $(N+1)^2$, và tại bậc thứ nhất (nếu $N=1$), ta thấy bốn thành phần của ambisonic ban đầu được gọi chung là FOA (đối với ambisonic bậc thứ nhất). Cũng được gọi là biến "phẳng" ambisonic phá vỡ âm thanh được xác định trong mặt phẳng thường là mặt phẳng nằm ngang. Trong trường hợp này, số lượng các kênh là $2N+1$ kênh. Ambisonic bậc thứ nhất (4 kênh: W, X, Y, Z) và ambisonic phẳng bậc thứ nhất (3 kênh: W, X, Y) sau đây được gọi không phân biệt là "ambisonics" để dễ đọc, việc xử lý này được thể hiện là áp dụng được một cách độc lập cho dù loại này có phẳng hay không. Tuy nhiên, nếu trong ngữ cảnh nhất định cần phải phân biệt, các thuật ngữ "ambisonic bậc thứ nhất" và "ambisonic phẳng bậc thứ nhất" được sử dụng. Lưu ý rằng có thể suy ra từ định dạng B bậc thứ nhất, tín hiệu âm thanh nổi (2 kênh) tương ứng với việc thu âm thanh nổi trùng hợp của cặp chéo loại Blumlein ($X+Y$ và $X-Y$) hoặc phía giữa (kết hợp W và X ở giữa và lấy Y ở phía bên).

Sau đây, tín hiệu ở định dạng B của thứ tự định trước được gọi là "âm thanh ambisonic". Theo một số biến thể, âm thanh ambisonic có thể được xác định trong định dạng khác như định dạng A hoặc các kênh được kết hợp trước bằng ma trận cố định (duy trì số lượng các kênh hoặc giảm số lượng các kênh với trường hợp có 3 hoặc 2 kênh), như sẽ được thấy sau đây.

Các tín hiệu cần được xử lý bởi bộ mã hóa/bộ giải mã được thể hiện dưới dạng sự liên tiếp của các khối của các mẫu âm thanh được gọi là "các khung" hoặc "các khung con" sau đây.

Ngoài ra, sau đây, các ký hiệu toán học tuân theo quy ước này:

- Vector: **u** (chữ thường, in đậm)
- Ma trận: **A** (viết hoa, in đậm)

Phương thức đơn giản nhất để mã hóa tín hiệu âm thanh nổi hoặc tín hiệu ambisonic để sử dụng bộ mã hóa đơn và áp dụng nó song song với tất cả các kênh, khả thi với cấp phát bit khác nhau phụ thuộc vào các kênh. Phương thức này ở đây được gọi là “đa-đơn” (mặc dù thực tế ta có thể tổng quát hóa phương thức đa-đơn hoặc sử dụng một số trường hợp song song của cùng bộ mã hóa-giải mã lỗi).

Phương án như vậy được thể hiện trên FIG.1. Tín hiệu đầu vào được chia thành các kênh (đơn) trong khối 100. Các kênh này được mã hóa riêng biệt trong các khối 120 đến 122 theo cấp phát định trước. Dòng bit của chúng được dồn kênh (khối 130) và sau khi truyền và/hoặc lưu trữ nó được giải dồn kênh (khối 140) để áp dụng việc giải mã cho mỗi trong số các kênh (các khối 150 đến 152) được tái kết hợp (khối 160).

Chất lượng được kết hợp thay đổi theo mã hóa đơn được sử dụng, và nói chung nó thỏa mãn chỉ ở tốc độ bit rất cao, ví dụ với tốc độ bit là ít nhất 48 kbps trên mỗi kênh đơn để mã hóa EVS. Do đó đối với bậc thứ nhất ta thu được tốc độ bit tối thiểu là $4 \times 48 = 192$ kbps.

Các giải pháp được đề xuất hiện tại cho các bộ mã hóa-giải mã phức tạp hơn, đối với sự không gian hóa ambisonic cụ thể, không thỏa mãn, cụ thể về mặt độ phức tạp, độ trễ, tính sử dụng hiệu quả của tốc độ bit, để đảm bảo sự giải tương quan hiệu quả giữa các kênh ambisonic.

Ví dụ, bộ mã hóa-giải mã MPEG-H đối với âm thanh ambisonic sử dụng hoạt động xếp chồng-bổ sung mà thêm độ trễ và độ phức tạp, cũng như phép nội suy tuyến tính đối với các vectơ có hướng dưới mức tối ưu và gây ra các khiếm khuyết. Vấn đề cơ bản với bộ mã hóa-giải mã này là nó thực thi sự phân tích thành các thành phần chiếm ưu thế và môi trường xung quanh vì các thành phần chiếm ưu thế có nghĩa là khác biệt về mặt cảm giác với môi trường xung quanh, nhưng việc phân tích này không được xác định hoàn toàn. Bộ mã hóa MPEG-H gặp phải vấn đề về sự không tương ứng giữa các hướng của các thành phần chính từ một khung đến khung khác: thứ tự của các thành phần (các tín hiệu) có thể được hoán đổi theo các hướng liên quan. Đó là lý do tại sao bộ mã hóa-giải mã MPEG-H sử dụng kỹ thuật so khớp và xếp chồng-bổ sung để giải quyết vấn đề này.

Hơn nữa, có thể sử dụng các phương thức mã hóa tần số (trong miền FFT hoặc MDCT) thay vì mã hóa theo thời gian như trong bộ mã hóa-giải mã MPEG-H, nhưng việc xử lý tín hiệu trong miền tần số (các dải phụ) yêu cầu việc truyền dữ liệu đến bộ giải mã bởi dải phụ, do đó làm tăng tốc độ bit cần thiết cho việc truyền này.

Bản chất kỹ thuật của sáng chế

Sáng chế cải thiện trường hợp này.

Để đạt được mục đích này, sáng chế đề xuất phương pháp mã hóa đối với việc nén các tín hiệu audio tạo ra, theo thời gian, sự liên tiếp của các khung mẫu, ở mỗi trong số N kênh trong biểu diễn ambisonic của bậc lớn hơn 0, phương pháp này bao gồm các bước:

- tạo ra, dựa trên các kênh và đối với khung hiện thời, ma trận hiệp biến liên kênh, và tìm kiếm các vector đặc trưng của ma trận hiệp biến nêu trên với việc xem xét tới việc thu ma trận của các vector đặc trưng,
- kiểm tra ma trận của các vector đặc trưng để xác minh rằng nó biểu diễn phép quay trong không gian N chiều, và nếu không, hiệu chỉnh ma trận của các vector đặc trưng cho đến khi ma trận quay thu được, đối với khung hiện thời, và
- áp dụng ma trận quay nêu trên cho các tín hiệu của N kênh trước khi mã hóa kênh riêng biệt của các tín hiệu nêu trên.

Sáng chế do đó có thể cải thiện sự giải tương quan giữa N kênh mà tiếp theo cần được mã hóa riêng biệt. Sự mã hóa riêng biệt này sau đây cũng được gọi là "đa-đơn mã hóa".

Theo một phương án, phương pháp này có thể còn bao gồm bước:

- mã hóa các tham số được lấy từ ma trận quay nhằm các mục đích truyền qua mạng.

Các tham số này thường có thể là bộ bốn và/hoặc góc quay và/hoặc các giá trị góc Euler như sẽ được thấy sau đây, hoặc nói cách khác chỉ là các phần tử của ma trận này, chẳng hạn.

Theo một phương án, phương pháp này có thể còn bao gồm các bước:

- so sánh ma trận của các vector đặc trưng mà thu được đối với khung hiện thời, với ma trận quay thu được đối với khung ở trước khung hiện thời, và
- hoán vị các cột của ma trận của các vector đặc trưng của khung hiện thời để đảm bảo tính thống nhất với ma trận quay của khung trước.

Phương án như vậy có thể giúp phương pháp này duy trì sự đồng nhất tổng thể và cụ thể để tránh các tiếng kêu lách tách từ một khung đến khung khác, trong việc tái tạo audio.

Tuy nhiên, các biến đổi nhất định được thực thi để thu các vector đặc trưng từ ma trận hiệp biến (như "PCA/KLT" được thấy sau đây) có khả năng đảo ngược hướng của các vector đặc trưng nhất định, và sau đó thích hợp là ở cùng một thời điểm xác minh tính thống nhất của trục, sau đó tính nhất quán theo hướng trên trục này, đối với mỗi vector đặc trưng của ma trận của khung hiện thời. Để đạt được mục đích này, theo một phương án, khi phép hoán vị nêu trên của các cột đã làm cho nó có thể đảm bảo tính nhất quán của các trục của các vector, phương pháp này còn bao gồm các bước:

- xác minh, cho mỗi vector đặc trưng của khung hiện thời, tính nhất quán theo hướng với vector cột của vị trí tương ứng trong ma trận quay của khung trước, và
- trong trường hợp không có tính nhất quán, đảo dấu của các phần tử của vector đặc trưng này trong ma trận của các vector đặc trưng của khung hiện thời.

Thông thường, với phép hoán vị giữa các cột của ma trận của các vector đặc trưng để đảo dấu của định thức của ma trận của các vector đặc trưng và định thức của ma trận quay bằng 1,

ta có thể ước tính định thức của ma trận của các vector đặc trưng, và nếu nó bằng -1, sau đó ta có thể đảo các dấu của các phần tử của cột được chọn của ma trận của các vector đặc trưng để định thức bằng 1, và do đó tạo ra ma trận quay.

Theo một phương án, phương pháp này có thể còn bao gồm các bước:

- ước tính chênh lệch giữa ma trận quay thu được đối với khung hiện thời và ma trận quay thu được đối với khung ở trước khung hiện thời,

- dựa trên chênh lệch được ước tính, xác định liệu ít nhất một phép nội suy có cần được thực hiện giữa ma trận quay của khung hiện thời và ma trận quay của khung trước hay không.

Sau đó phép nội suy như vậy có thể làm nhẵn ("trung bình lũy tiến") các ma trận quay được áp dụng lần lượt cho khung trước và khung hiện thời và do đó làm suy giảm hiệu ứng tiếng lách tách nghe được từ một khung đến khung khác trong khi phát lại.

Theo phương án thực hiện như vậy:

- dựa trên chênh lệch được ước tính, số lượng các phép nội suy cần được thực hiện giữa ma trận quay của khung hiện thời và ma trận quay của khung trước được xác định,

- khung hiện thời được chia thành số lượng các khung con tương ứng với số lượng các phép nội suy cần được thực hiện, và

- ít nhất số lượng các phép nội suy này có thể được mã hóa với việc xem xét tới việc truyền qua mạng nêu trên.

Theo một phương án, biểu diễn ambisonic là bậc thứ nhất và số lượng N kênh là bốn, và ma trận quay của khung hiện thời được biểu diễn bằng hai bộ bốn.

Theo phương án này và trong trường hợp của phép nội suy, mỗi phép nội suy cho khung con hiện thời là phép nội suy tuyến tính hình cầu (hoặc "SLERP"), được tiến hành là hàm của phép nội suy của khung con ở trước khung con hiện thời và dựa trên các bộ bốn của khung con trước.

Ví dụ, phép nội suy tuyến tính hình cầu của khung con hiện thời có thể được tiến hành để thu các bộ bốn của khung con hiện thời, như sau:

$$Q_{L,interp}(\alpha) = Q_{L,t-1} \frac{\sin(1-\alpha)\Omega_L}{\sin \Omega_L} + Q_{L,t} \frac{\sin \alpha\Omega_L}{\sin \Omega_L}$$

$$Q_{R,interp}(\alpha) = Q_{R,t-1} \frac{\sin(1-\alpha)\Omega_R}{\sin \Omega_R} + Q_{R,t} \frac{\sin \alpha\Omega_R}{\sin \Omega_R}$$

trong đó:

$Q_{L,t-1}$ là một trong số các bộ bốn của khung con trước t-1,

$Q_{R,t-1}$ là bộ bốn khác của khung con trước t-1,

$Q_{L,t}^{\wedge}$ là một trong số các bộ bốn của khung con hiện thời t,

$Q_{R,t}^{\wedge}$ là bộ bốn khác của khung con hiện thời t,

$$\Omega_L = \text{Arccos} (Q_{L,t-1} \cdot Q_{L,t}); \Omega_R = \text{Arccos} (Q_{R,t-1} \cdot Q_{R,t})$$

và α tương ứng với hệ số nội suy.

Theo một phương án, việc tìm kiếm các vectơ đặc trưng được tiến hành bằng cách phân tích thành phần chính (hoặc "PCA") hoặc bằng biến đổi Karhunen-Loève (hoặc "KLT"), trong miền thời gian. Tất nhiên, các phương án khác có thể được xem xét (phân tích giá trị suy biến hoặc các giá trị khác).

Theo một phương án, phương pháp này bao gồm bước trước đó dự báo ngân sách cấp phát bit trên mỗi kênh ambisonic, bao gồm:

- cho mỗi kênh ambisonic, ước tính năng lượng âm thanh hiện tại trong kênh,
- chọn, trong bộ nhớ, điểm chất lượng định trước, dựa trên kênh ambisonic này và đối với tốc độ bit hiện thời trong mạng,
- ước tính trọng số cần được áp dụng cho việc cấp phát bit cho kênh này, bằng cách nhân điểm số được chọn với năng lượng được ước tính.

Phương án này sau đó giúp nó có thể quản lý cấp phát tối ưu của các bit cần được chỉ định cho mỗi kênh cần được mã hóa. Điều này có lợi và chính nó và có thể có khả năng là đối tượng của việc bảo vệ riêng biệt.

Sáng chế cũng đề cập đến phương pháp giải mã các tín hiệu audio tạo ra, theo thời gian, sự liên tiếp của các khung mẫu, ở mỗi trong số N kênh trong biểu diễn ambisonic của bậc lớn hơn 0, phương pháp này bao gồm các bước:

- nhận, đối với khung hiện thời, ngoài các tín hiệu của N kênh của khung hiện thời này, các tham số của ma trận quay,

- cấu trúc ma trận quay ngược từ các tham số nêu trên,
- áp dụng ma trận quay ngược nêu trên cho các tín hiệu từ N kênh nhận được, trước khi giải mã kênh riêng biệt các tín hiệu nêu trên.

Phương án như vậy cũng khiến nó có thể cải thiện, trong việc giải mã, sự giải tương quan giữa N kênh.

Sáng chế cũng đề cập đến thiết bị mã hóa bao gồm mạch xử lý để thực thi phương pháp mã hóa được thể hiện ở trên.

Sáng chế cũng đề cập đến thiết bị giải mã bao gồm mạch xử lý để thực thi phương pháp giải mã nêu trên.

Sáng chế cũng đề cập đến chương trình máy tính bao gồm các lệnh để thực thi phương pháp nêu trên, khi các lệnh này được thực thi bởi bộ xử lý của mạch xử lý.

Sáng chế cũng đề cập đến phương tiện nhớ không tạm thời lưu trữ các lệnh của chương trình máy tính này.

Mô tả vắn tắt các hình vẽ

Các dấu hiệu và các ưu điểm khác của sáng chế sẽ trở nên hiển nhiên từ việc đọc các phương án làm ví dụ được thể hiện trong phần mô tả chi tiết sau đây, và từ việc xem xét các hình vẽ kèm theo trong đó:

- FIG.1 minh họa đa-đơn mã hóa (giải pháp kỹ thuật trước),
- FIG.2 minh họa sự liên tiếp của các bước chính của phương pháp làm ví dụ theo ý nghĩa của sáng chế,
- FIG.3 thể hiện cấu trúc chung của ví dụ về bộ mã hóa theo sáng chế,
- FIG.4 thể hiện các chi tiết về việc phân tích và biến đổi PCA/KLT được thực hiện bởi khối 310 của bộ mã hóa của FIG.3,
- FIG.5 thể hiện ví dụ của bộ giải mã theo sáng chế,
- FIG.6 thể hiện việc giải mã và tổng hợp PCA/KLT mà ngược lại với FIG.4, trong việc giải mã,

- FIG.7 minh họa các phương án làm ví dụ của cấu trúc của bộ mã hóa và bộ giải mã trong ý nghĩa của sáng chế.

Mô tả chi tiết sáng chế

Sáng chế nhằm cho phép việc mã hóa được tối ưu hóa bằng:

- ma trận thời gian tương thích (cụ thể với việc biến đổi tương thích thu được bằng PCA/KLT ("PCA" chỉ định phân tích thành phần chính và "KLT" chỉ định biến đổi Karhunen-Loève),

- tốt hơn là được theo sau bằng đa-đơn mã hóa.

Ma trận tương thích cho phép sự phân tích hiệu quả hơn thành các kênh so với ma trận cố định. Ma trận theo sáng chế có lợi khiến nó có thể giải tương quan các kênh trước khi đa-đơn mã hóa, để nhiều mã hóa được đưa vào bằng cách mã hóa mỗi trong số các kênh làm biến dạng hình ảnh không gian ít nhất có thể về tổng thể khi các kênh này được tái kết hợp để tái cấu trúc tín hiệu ambisonic trong việc giải mã.

Ngoài ra, sáng chế có thể đảm bảo sự tương thích nhẹ của các tham số ma trận để tránh các thành phần lạ kiểu “nhấp” vào mép của khung hoặc các dao động quá nhanh trong hình ảnh không gian, hoặc ngay cả các thành phần lạ mã hóa do các biến thể quá mạnh (ví dụ được liên kết với sự hoán vị không đúng lúc của các nguồn audio giữa các kênh) trong các kênh riêng biệt khác nhau do ma trận sau đó được mã hóa bởi các trường hợp khác nhau của bộ mã hóa-giải mã đơn. Đa-đơn mã hóa được thể hiện sau đây tốt hơn là với việc cấp phát bit biến đổi giữa các kênh (sau ma trận hóa tương thích), nhưng trong một số trường hợp đa biến của bộ mã hóa-giải mã lỗi âm thanh nổi hoặc trường hợp khác có thể được sử dụng.

Để tạo thuận lợi cho việc hiểu sáng chế, các khái niệm giải thích nhất định liên quan đến các phép quay n chiều và các phân tích loại PCA/KLT hoặc SVD ("SVD" biểu thị phân tích giá trị suy biến) được tóm lại sau đây.

Các phép quay và "các bộ bốn"

Các tín hiệu được biểu diễn bởi các khối liên tiếp của các mẫu audio, các khối này được gọi là "các khung con" sau đây.

Sáng chế sử dụng sự biểu diễn của các phép quay n chiều với các tham số thích hợp đối với sự lượng tử hóa trên mỗi khung và cụ thể phép nội suy hiệu quả bởi khung con. Các biểu diễn của các phép quay được sử dụng 2, 3, và 4 chiều được xác định sau đây.

Phép quay (xung quanh gốc) là phép biến đổi của không gian n chiều mà thay đổi từ một vectơ đến vectơ khác, sao cho:

- Độ lớn của vectơ được bảo toàn
- Tích chéo của các vectơ xác định hệ tọa độ trực giao trước khi phép quay được bảo toàn sau phép quay (không có phản xạ).

Ma trận $\mathbf{M}$ có kích thước $n \times n$ là ma trận quay nếu và chỉ nếu $\mathbf{M}^T \cdot \mathbf{M} = \mathbf{I}_n$ trong đó $\mathbf{I}_n$ chỉ định ma trận đơn vị có kích thước $n \times n$ (tức là $\mathbf{M}$ là ma trận unita, $\mathbf{M}^T$ chỉ định chuyển vị của $\mathbf{M}$) và định thức của nó là $+1$.

Các biểu diễn được sử dụng theo sáng chế tương đương với sự biểu diễn bởi ma trận quay:

Trong hai chiều (trong mặt phẳng 2D) ($n=2$): Ta sử dụng góc của phép quay là sự biểu diễn, như sau.

Cho góc của phép quay θ ta suy ra ma trận quay:

$$\mathbf{M}_2(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Cho ma trận quay, ta có thể tính góc θ bằng cách quan sát rằng vết của ma trận là $2\cos \theta$. Lưu ý rằng cũng có thể ước tính θ trực tiếp từ ma trận hiệp biến trước khi áp dụng phân tích thành phần chính (principal component decomposition - PCA) và phân tích giá trị đặc trưng (eigenvalue decomposition, EVD) được thể hiện sau đây.

Phép nội suy giữa hai phép quay của các góc tương ứng θ_1 và θ_2 có thể được thực hiện bằng phép nội suy tuyến tính giữa θ_1 và θ_2 , tính đến giới hạn đường ngắn nhất trên đường tròn đơn vị giữa hai góc này.

Trong không gian ba chiều (3D) ($n=3$): các góc Euler và các bộ bốn được sử dụng là phép biểu diễn. Theo một số biến thể, phép biểu diễn trục-góc cũng có thể được sử dụng, không được nêu ở đây.

Ma trận quay có kích thước 3×3 có thể được phá vỡ thành tích của 3 phép quay cơ sở của góc θ dọc theo các trục x , y , hoặc z .

$$\begin{aligned} M_{3,x}(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \\ M_{3,y}(\theta) &= \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \\ M_{3,z}(\theta) &= \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Phụ thuộc vào các kết hợp của trục, các góc nêu trên là các góc Euler hoặc Cardan.

Tuy nhiên, sự biểu diễn khác của các phép quay 3D được cho bởi các bộ bốn. Các bộ bốn là việc tổng quát hóa của các phép biểu diễn bởi các số phức có bốn thành phần ở dạng số $\mathbf{q} = a + bi + cj + dk$ trong đó $i^2 = j^2 = k^2 = ijk = -1$.

Phần thực a được gọi là vô hướng và ba phần ảo (b, c, d) tạo ra vector 3D. Chuẩn của bộ bốn là $|\mathbf{q}| = \sqrt{a^2 + b^2 + c^2 + d^2}$. Các bộ bốn đơn vị (của chuẩn 1) biểu diễn các phép quay – tuy nhiên, sự biểu diễn này không phải là duy nhất; do đó, nếu $\mathbf{q}$ biểu diễn phép quay, $-\mathbf{q}$ biểu diễn cùng một phép quay.

Cho bộ bốn đơn vị $\mathbf{q} = a + bi + cj + dk$ (với $a^2 + b^2 + c^2 + d^2 = 1$), ma trận quay kết hợp là:

$$M_{3,quat}(\mathbf{q}) = \begin{pmatrix} a^2 - b^2 - c^2 - d^2 & 2(bc - ad) & 2(ac + bd) \\ 2(ad + bc) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(ab + cd) & a^2 - b^2 - c^2 + d^2 \end{pmatrix}$$

Các góc Euler không cho phép nội suy chính xác các phép quay 3D; để thực hiện điều đó, thay vì ta sử dụng các bộ bốn hoặc phép biểu diễn trục-góc. Phương

pháp nội suy SLERP ("phép nội suy tuyến tính hình cầu") thay cho việc nội suy theo công thức:

$$\text{slerp}(\mathbf{q}_1, \mathbf{q}_2, \alpha) = \frac{\sin(1 - \alpha)\Omega}{\sin \Omega} \mathbf{q}_1 + \frac{\sin \alpha\Omega}{\sin \Omega} \mathbf{q}_2$$

trong đó $0 \leq \alpha \leq 1$ là hệ số nội suy đi từ $\mathbf{q}_1$ đến $\mathbf{q}_2$ và Ω là góc giữa hai bộ bốn:

$$\Omega = \arccos(\mathbf{q}_1 \cdot \mathbf{q}_2)$$

trong đó $\mathbf{q}_1 \cdot \mathbf{q}_2$ biểu thị tích chấm giữa hai bộ bốn (giống với tích chấm giữa hai vectơ 4 chiều).

Lượng này để nội suy bằng cách đi theo vòng tròn lớn đối với quả cầu 4D có tốc độ góc không đổi là hàm của α . Người ta phải đảm bảo rằng đường ngắn nhất được sử dụng để nội suy bằng cách thay đổi dấu của một trong số các bộ bốn khi $\mathbf{q}_1 \cdot \mathbf{q}_2 < 0$. Lưu ý rằng các phương pháp khác đối với phép nội suy bộ bốn có thể được sử dụng (phép nội suy tuyến tính được chuẩn hóa hoặc nlerp, splines, v.v.).

Lưu ý rằng cũng có thể nội suy các phép quay 3D bằng phép biểu diễn trục-góc; trong trường hợp này, góc được nội suy như trong trường hợp 2D, và trục có thể được nội suy ví dụ bằng phương pháp SLERP (trong 3D) trong khi đảm bảo rằng đường ngắn nhất được lấy trên quả cầu đơn vị 3D và tính đến thực tế rằng sự biểu diễn được cho bởi trục $\mathbf{r}$ và góc θ tương đương với được cho bởi trục của hướng đối diện $-\mathbf{r}$ và góc $2\pi - \theta$.

Ở chiều thứ 4 ($n=4$), phép quay có thể được tham số hóa bởi 6 góc ($n(n-1)/2$) và ta thể hiện rằng phép nhân của hai ma trận có kích thước 4x4 được gọi là bộ bốn ($\mathbf{Q}_1$) và phản bộ bốn ($\mathbf{Q}_2^*$) được liên kết với các bộ bốn $\mathbf{q}_1 = a + bi + cj + dk$ và $\mathbf{q}_2 = w + xi + yj + zk$ cho ma trận quay có kích thước 4x4.

Có thể thấy cặp bộ bốn kết hợp $(\mathbf{q}_1, \mathbf{q}_2)$ và bộ bốn kết hợp và các ma trận phản bộ bốn sao cho:

$$\mathbf{Q}_1 = \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix}$$

và

$$Q_2^* = \begin{pmatrix} w & -x & -y & -z \\ x & w & -z & y \\ y & z & w & -x \\ z & -y & x & w \end{pmatrix}$$

Tích của chúng cho ma trận kích thước 4x4:

$$M_{4,quat}(q_1, q_2) = Q_1 Q_2^*$$

và có thể xác minh rằng ma trận này thỏa mãn các tính chất của ma trận quay (ma trận unita và định thức bằng 1).

Ngược lại, ma trận quay 4x4 đã cho, ma trận này có thể được phân tích thành tích của các ma trận ở dạng $Q_1 Q_2^*$ ví dụ với phương pháp được biết là "sự thừa số hóa của Cayley". Điều này liên quan đến việc tính toán ma trận trung gian được gọi là "biến đổi tứ giác" (hoặc ma trận kết hợp) và suy ra các bộ bốn từ điều này với một số không xác định đối với dấu của hai bộ bốn (có thể được loại bỏ bởi giới hạn "đường ngắn nhất" bổ sung được nêu sau đây).

Phân tích giá trị suy biến (hoặc "SVD")

Phân tích giá trị suy biến (Singular value decomposition, SVD) bao gồm phân tích ma trận thực A có kích thước $m \times n$ ở dạng:

$$A = U \Sigma V^T$$

trong đó U là ma trận unita ($U^T U = I_m$) có kích thước $m \times m$, Σ là ma trận chéo hình chữ nhật có kích thước $m \times n$ có các hệ số thực và dương $\sigma_i \geq 0$ ($i = 1 \dots p$ trong đó $p = \min(m, n)$), V là ma trận unita ($V^T V = I_n$) có kích thước $n \times n$, và V^T là chuyển vị của V . Các hệ số σ_i theo đường chéo của Σ là các giá trị suy biến của ma trận A . Theo quy ước, nói chung chúng được liệt kê theo thứ tự giảm dần, và trong trường hợp này ma trận chéo Σ được liên kết với A là duy nhất.

Hạng r của A được cho bởi số lượng của các hệ số khác không σ_i . Do đó ta có thể viết lại phân tích giá trị suy biến dưới dạng:

$$A = [U_r \tilde{U}_r] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ \tilde{V}_r^T \end{bmatrix}$$

trong đó $\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$ là các vector suy biến ở bên trái (hoặc các vector đầu ra) là $\mathbf{A}$, $\mathbf{\Sigma}_r = \text{diag}(\sigma_1, \dots, \sigma_r)$, và $\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ là các vector suy biến ở bên phải (hoặc các vector đầu vào) của $\mathbf{A}$. Công thức ma trận này cũng có thể được viết lại là:

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Nếu tổng bị giới hạn với chỉ số $i < r$, ta thu được ma trận “được lọc” mà biểu diễn chỉ thông tin “chiếm ưu thế”.

Ta cũng có thể viết:

$$\mathbf{A} \mathbf{v}_i = \sigma_i \mathbf{u}_i$$

thể hiện rằng ma trận $\mathbf{A}$ biến đổi $\mathbf{v}_i$ thành $\sigma_i \mathbf{u}_i$.

SVD của $\mathbf{A}$ có mối tương quan với việc phân tích giá trị đặc trưng của $\mathbf{A}^T \mathbf{A}$ và $\mathbf{A} \mathbf{A}^T$ vì:

$$\mathbf{A}^T \mathbf{A} = \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma}) \mathbf{V}^T$$

$$\mathbf{A} \mathbf{A}^T = \mathbf{U} (\mathbf{\Sigma} \mathbf{\Sigma}^T) \mathbf{U}^T$$

Các giá trị đặc trưng của $\mathbf{\Sigma}^T \mathbf{\Sigma}$ và $\mathbf{\Sigma} \mathbf{\Sigma}^T$ là $\sigma_1^2, \dots, \sigma_r^2$. Các cột của $\mathbf{U}$ là các vector đặc trưng là $\mathbf{A} \mathbf{A}^T$, trong khi các cột của $\mathbf{V}$ là các vector đặc trưng của $\mathbf{A}^T \mathbf{A}$.

SVD có thể được giải thích về mặt hình học: hình ảnh của quả cầu theo chiều n bởi ma trận $\mathbf{A}$ là, theo chiều m , siêu elip có các trục chính theo các hướng $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ và có độ dài $\sigma_1, \dots, \sigma_m$.

Biến đổi Karhunen-Loève (hoặc "KLT")

Biến đổi Karhunen-Loève (Karhunen-Loève transform, KLT) của vector ngẫu nhiên $\mathbf{x}$ được chỉnh tâm ở 0 và của ma trận hiệp biến $\mathbf{R}_{\mathbf{xx}} = E[\mathbf{x} \mathbf{x}^T]$ được xác định bởi:

$$\mathbf{y} = \mathbf{V}^T \mathbf{x}$$

trong đó V là ma trận của các vector đặc trưng (với quy ước là các vector đặc trưng là các vector cột) thu được bằng cách phân tích R_{xx} thành các giá trị đặc trưng

$$R_{xx} = V\Lambda V^T$$

trong đó $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ là ma trận chéo mà có các hệ số là các giá trị riêng. Ma trận $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ chứa các vector đặc trưng (các cột) của R_{xx} , sao cho

$$R_{xx}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

Ta có thể thấy KLT là sự thay đổi cơ sở, vì tích $V^T \mathbf{x}$ biểu diễn vector $\mathbf{x}$ trong cơ sở được cho bởi các vector đặc trưng.

Phép biến đổi ngược được cho bởi:

$$\mathbf{x} = V \mathbf{y}$$

KLT khiến nó có thể giải tương quan các thành phần của $\mathbf{x}$; các biến đổi của vector được biến đổi $\mathbf{y}$ là các giá trị đặc trưng của R_{xx}

Phân tích thành phần chính (hoặc "PCA")

Phân tích thành phần chính (Principal Component Analysis, PCA) là kỹ thuật giảm chiều mà tạo ra các biến trực giao và tối đa hóa biến đổi của các biến sau khi chiếu (hoặc một cách tương đương giảm thiểu sai số tái cấu trúc).

PCA được thể hiện sau đây, mặc dù cũng dựa trên sự phân tích thành các giá trị đặc trưng như KLT, sao cho ma trận hiệp biến được ước tính $\hat{R}_{xx}$ được tính toán từ N vector được quan sát $\mathbf{x}_i$, $i = 1 \dots N$ của chiều n :

$$\hat{R}_{xx} = \frac{1}{N-1} \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T$$

giả sử rằng các vector này được chỉnh tâm:

$$\mathbf{m}_x = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i = 0$$

Sự phân tích thành các giá trị đặc trưng của $\widehat{R}_{xx}$ ở dạng $\widehat{R}_{xx} = V\Lambda V^T$ cho phép tính toán các thành phần chính: $y_n = V^T x_n$.

PCA là phép biến đổi bởi ma trận V^T mà chiều dữ liệu trong cơ sở mới để tối đa hóa biến đổi của các biến sau khi chiếu.

Lưu ý rằng PCA cũng có thể thu được từ SVD của tín hiệu x_i được đặt ở dạng của ma trận X có kích thước $n \times N$. Trong trường hợp này, ta có thể viết:

$$X = UDV^T$$

Ta xác minh rằng $XX^T = UDD^T U^T$, tương ứng với sự chéo hóa của XX^T . Do đó các vectơ chiếu của PCA tương ứng với các vectơ cột là U và phép chiếu cho $U^T X = DV^T$ dưới dạng kết quả.

Ta cũng sẽ lưu ý rằng PCA được xem xét tổng quát là kỹ thuật giảm chiều, để "nén" tập dữ liệu có số chiều cao thành tập bao gồm một số thành phần chính. Theo sáng chế, PCA có lợi để khiến nó có thể giải tương quan tín hiệu đầu vào đa chiều, nhưng sự giảm các kênh (do đó làm giảm số lượng các kênh) được tránh để tránh đưa vào các thành phần lạ. Điều này buộc tốc độ bit mã hóa tối thiểu, để tránh "cắt ngắn" hình ảnh không gian, ngoại trừ các biến cụ thể trong đó các giá trị đặc trưng quá thấp để tỷ lệ không có thể được cho phép (ví dụ để mã hóa tốt hơn âm thanh ambisonic được tạo ra nhân tạo với một nguồn được tổng hợp theo không gian).

Giờ ta tham chiếu đến FIG.2 để mô tả các nguyên lý chung của các bước mà được thực thi theo phương pháp trong ý nghĩa của sáng chế, đối với khung hiện thời t. Bước S1 bao gồm bước thu được các tín hiệu tương ứng của các kênh ambisonic (ở đây có bốn kênh W, Y, Z, X theo ví dụ được mô tả, sử dụng quy ước thứ bậc kênh ACN (Ambisonics Channel Number - số lượng kênh ambisonic) cho mỗi khung t. Các tín hiệu này có thể được đặt ở dạng của ma trận $n \times L$ (đối với n kênh ambisonic (ở đây là 4) và L mẫu trên mỗi khung).

Ở bước tiếp theo S2, các tín hiệu của các kênh này có thể được xử lý trước một cách tùy chọn, ví dụ bởi bộ lọc thông cao như được mô tả sau đây với tham chiếu đến FIG.3.

Ở bước tiếp theo S3, phân tích thành phần chính PCA hoặc theo cách thức tương đương biến đổi Karhunen-Loève KLT được áp dụng cho các tín hiệu này, để thu được các giá trị đặc trưng và ma trận của các vector đặc trưng từ ma trận hiệp biến của n kênh. Theo các biến thể của sáng chế, SVD có thể được sử dụng.

Ở bước S4, ma trận này của các vector đặc trưng, thu được đối với khung hiện thời t , trải qua các hoán vị dấu để được căn chỉnh nhất có thể với ma trận có cùng bản chất với của khung trước $t-1$. Theo nguyên lý, ta đảm bảo rằng trục của các vector cột trong ma trận của các vector đặc trưng tương ứng nhiều nhất có thể với trục của các vector cột ở cùng một chỗ trong ma trận của khung trước, và nếu không, các vị trí của các vector đặc trưng của ma trận của khung hiện thời t không tương ứng được hoán vị. Sau đó, ta cũng đảm bảo rằng các hướng của các vector đặc trưng từ một ma trận đến ma trận khác cũng trùng nhau. Nói cách khác, ban đầu ta quan tâm đến các đường thẳng chứa các vector đặc trưng (chỉ là định hướng, không phải là hướng) và cho mỗi đường ta tìm kiếm đường gần nhất trong ma trận của khung trước $t-1$. Để thực hiện điều này, các vector được hoán vị trong ma trận của khung hiện thời. Sau đó, ở bước thứ hai, ta thử khớp hướng của các vector (có hướng). Để thực hiện điều này ta đảo dấu của các vector đặc trưng mà có thể không có hướng bên phải.

Phương án như vậy giúp nó có thể đảm bảo tính nhất quán tối đa giữa hai ma trận và do đó tránh được các tiếng kêu lách tách giữa hai khung trong khi phát lại âm thanh.

Ở bước S5, ta cũng đảm bảo rằng ma trận của các vector đặc trưng của khung hiện thời t , do đó được hiệu chỉnh bởi các hoán vị dấu, thực sự biểu diễn việc áp dụng phép quay (của góc cho $n=2$ kênh, có ba góc Euler, của trục và góc, hoặc của bộ bốn cho $n=3$ tương ứng với biểu diễn ambisonic phẳng bậc thứ nhất W, Y, Z, và hai bộ bốn cho $n=4$ trong kiểu biểu diễn ambisonic bậc thứ nhất W, Y, Z, X).

Để đảm bảo rằng nó thực sự là phép quay, định thức của ma trận của các vector đặc trưng của khung hiện thời t , được hiệu chỉnh bởi các hoán vị, phải dương và bằng (hoặc, thực tế, gần với) $+1$ ở bước S6. Nếu nó bằng (hoặc gần) -1 , sau đó ta cần:

- hoán vị lại hai vector đặc trưng (ví dụ được liên kết với các kênh năng lượng thấp, do đó không mang tính đại diện), hoặc

- tốt hơn là đảo dấu tất cả các phần tử của cột (ví dụ được liên kết với kênh năng lượng thấp) ở bước S6.

Sau đó ta thu được ma trận của các vector đặc trưng đối với khung hiện thời t hiệu quả tương ứng với phép quay ở bước S7.

Các tham số của ma trận này (ví dụ như giá trị góc, giá trị của trục và góc, hoặc của (các) bộ bốn của ma trận này) sau đó có thể được mã hóa trong số lượng các bit được cấp phát đối với mục đích này ở bước S8. Theo tùy chọn khác nhưng theo phương án có lợi, trong trường hợp trong đó chênh lệch tín hiệu được quan sát ở bước S9 (lớn hơn ngưỡng ví dụ) giữa ma trận quay được ước tính đối với khung hiện thời t và ma trận quay của khung trước $t-1$, biến số của các khung con nội suy có thể được xác định: nói cách khác số lượng các khung con này được cố định ở giá trị định trước. Bước S10 bao gồm:

- chia các khung hiện thời thành các khung con, và
- nội suy các ma trận cần được áp dụng cho các khung con liên tiếp từ ma trận của khung trước $t-1$ đến ma trận của khung hiện thời t , để làm phẳng sự chênh lệch giữa hai ma trận theo thời gian.

Ở bước S11, các ma trận quay được nội suy được áp dụng cho ma trận $n \times (L/K)$ biểu diễn mỗi trong số K khung con của các tín hiệu của các kênh ambisonic của bước S1 (hoặc một cách tùy chọn S2) để giải tương quan các tín hiệu này nhiều nhất có thể trước khi đa-đơn mã hóa của bước S14. Ta sẽ nhớ lại thực tế là ta muốn giải tương quan các tín hiệu này nhiều nhất có thể trước phép biến đổi đa-đơn này, theo cách thức thông thường. Cấp phát bit cho các kênh riêng biệt được hoàn thành ở bước S12 và được mã hóa ở bước S13.

Ở bước S14, trước khi tiến hành ghép kênh của bước S15 và do đó kết thúc phương pháp mã hóa nén, có thể quyết định về số lượng các bit được cấp phát trên mỗi kênh là hàm của sự biểu diễn của kênh này và của tốc độ bit khả dụng đối với mạng RES (FIG.7). Theo một phương án, năng lượng trong mỗi kênh được ước tính

đối với khung hiện thời và năng lượng này được nhân với số điểm định trước đối với kênh này và đối với tốc độ bit đã cho (điểm số này là ví dụ của điểm số MOS được giải thích sau đây với tham chiếu đến FIG.3). Số lượng các bit cần được cấp phát cho mỗi kênh do đó có trọng số. Phương án như vậy có lợi khi nó có thể là đối tượng bảo vệ riêng biệt trong ngữ cảnh ambisonic.

Thiết bị mã hóa DCOD và thiết bị giải mã DDEC được minh họa trên FIG.7 trong ý nghĩa của sáng chế, các thiết bị này là kép so với nhau (có nghĩa là "đảo ngược") và được nối với nhau bằng mạng truyền thông RES.

Thiết bị mã hóa DCOD bao gồm mạch xử lý điển hình bao gồm:

- bộ nhớ MEM1 để lưu trữ dữ liệu lệnh của chương trình máy tính trong ý nghĩa của sáng chế (các lệnh này có thể được phân bố giữa bộ mã hóa DCOD và bộ giải mã DDEC);
- giao diện INT1 để nhận các tín hiệu ambisonic được phân bố qua các kênh khác nhau (ví dụ bốn kênh bậc thứ nhất W, Y, Z, X) với việc xem xét tới việc mã hóa nén của chúng trong ý nghĩa của sáng chế;
- bộ xử lý PROC1 để nhận các tín hiệu này và xử lý chúng bằng cách thực thi các lệnh chương trình máy tính được lưu trữ trong bộ nhớ MEM1, với việc xem xét tới việc mã hóa của chúng; và
- giao diện truyền thông COM1 để truyền các tín hiệu được mã hóa qua mạng.

Thiết bị giải mã DDEC bao gồm chính mạch xử lý của nó, điển hình bao gồm:

- bộ nhớ MEM2 để lưu trữ dữ liệu lệnh của chương trình máy tính trong ý nghĩa của sáng chế (các lệnh này có thể được phân bố giữa bộ mã hóa DCOD và bộ giải mã DDEC như được chỉ báo ở trên);
- giao diện COM2 để nhận các tín hiệu được mã hóa từ mạng RES với việc xem xét tới việc giải mã của chúng từ việc nén trong ý nghĩa của sáng chế;

- bộ xử lý PROC2 để xử lý các tín hiệu này bằng cách thực thi các lệnh chương trình máy tính được lưu trữ trong bộ nhớ MEM2, với việc xem xét tới việc giải mã của chúng; và

- giao diện đầu ra INT2 để chuyển các tín hiệu giải mã ở dạng của các kênh ambisonic W' , Y' , Z' , X' , ví dụ với việc xem xét tới việc phát lại của chúng.

Tất nhiên, FIG.7 này minh họa một ví dụ của phương án cấu trúc của bộ mã hóa-giải mã (bộ mã hóa hoặc bộ giải mã) trong ý nghĩa của sáng chế. Các hình vẽ từ FIG.3 đến FIG.6, được thảo luận sau đây, các phương án chi tiết của các bộ mã hóa-giải mã này có nhiều chức năng hơn.

Giờ tham chiếu được tạo ra với FIG.3 để mô tả thiết bị mã hóa trong ý nghĩa của sáng chế.

Chiến lược của bộ mã hóa để giải tương quan các kênh của tín hiệu ambisonic nhiều nhất có thể và để mã hóa chúng với bộ mã hóa-giải mã lõi. Chiến lược này khiến nó có thể giới hạn các thành phần lạ trong tín hiệu ambisonic giải mã. Cụ thể hơn, ở đây ta thấy áp sự giải tương quan tối ưu của các kênh đầu vào trước khi đa-đơn mã hóa. Ngoài ra, phép nội suy là chi phí tính toán được giới hạn đối với bộ mã hóa và bộ giải mã vì nó được tiến hành trong miền cụ thể (góc trong 2D, bộ bốn trong 3D, cặp bộ bốn trong 4D) khiến nó có thể nội suy các ma trận hiệp biến được tính toán đối với việc phân tích PCA/KLT tốt hơn là lặp lại sự phân tích thành các giá trị đặc trưng và các vector đặc trưng, nhiều lần trên mỗi khung.

Tuy nhiên, trước khi thảo luận việc mã hóa lõi được thực hiện trong ý nghĩa của sáng chế, một số dấu hiệu của bộ mã hóa mà có lợi được thể hiện ở đây, cụ thể như việc tối ưu hóa của ngân sách bit được cấp phát để mã hóa là hàm của các tiêu chuẩn cảm giác, được thấy sau đây.

Theo phương án của bộ mã hóa được mô tả ở đây, trường hợp sau thông thường có thể là việc mở rộng của bộ mã hóa 3GPP EVS chuẩn hóa (đối với "các dịch vụ giọng nói tăng cường (Enhanced Voiced Services)"). Một cách thuận lợi, các tốc độ bit mã hóa EVS có thể được sử dụng sau đó không sửa đổi cấu trúc của dòng bit EVS. Do đó, các chức năng đa-đơn mã hóa (khối 340 của FIG.3 được mô tả sau đây) ở đây với sự cấp phát khả thi với mỗi kênh được biến đổi, được giới hạn

với các tốc độ bit sau đây để mã hóa trong dải audio siêu rộng: 9,6; 13,2; 16,4; 24,4; 32; 48; 64; 96 và 128 kbps.

Tất nhiên, có thể thêm các tốc độ bit bổ sung (để có độ chi tiết cụ thể hơn trong việc cấp phát) bằng cách sửa đổi bộ mã hóa-giải mã EVS. Cũng có thể sử dụng bộ mã hóa-giải mã ngoài EVS, ví dụ bộ mã hóa-giải mã OPUS®.

Nói chung, hãy nhớ rằng độ chi tiết mã hóa càng mịn, thì càng nhiều bit phải được dành để biểu diễn các kết hợp khả thi của các tốc độ bit. Sự thỏa thuận giữa độ mịn trong việc cấp phát và thông tin bổ sung mô tả cấp phát bit phải được tạo ra. Sự cấp phát này được tối ưu hóa ở đây bởi khối 320 của FIG.3, được mô tả sau đây. Đây là dấu hiệu có lợi trong và của chính nó và độc lập với sự phân tích thành các vectơ đặc trưng để thiết lập ma trận quay trong ý nghĩa của sáng chế. Như vậy, cấp phát bit được thực hiện bởi khối 320 có thể là đối tượng bảo vệ riêng biệt.

Tham chiếu đến FIG.3, khối 300 nhận tín hiệu đầu vào Y trong khung hiện thời của chỉ số t . Chỉ số này được thể hiện ở đây để không làm phức tạp các nhãn. Đây là ma trận có kích thước $n \times L$. Theo phương án tương thích đối với ngữ cảnh ambisonic bậc thứ nhất, ta có $n=4$ kênh W, Y, Z, X (do đó được xác định theo bậc ACN) mà có thể được chuẩn hóa theo quy ước SN3D. Theo một biến thể, bậc của các kênh nói cách khác có thể là ví dụ W, X, Y, Z (tuân theo quy ước FuMa) và sự chuẩn hóa có thể khác nhau (N3D hoặc FuMa). Do đó các kênh W, Y, Z, X tương ứng với các hàng liên tiếp: $y_{1,l}, y_{2,l}, y_{3,l}, y_{4,l}$ sẽ được biểu thị ở dạng của các tín hiệu một chiều $y_i(l), l = 1, \dots, L$. Do đó đây là sự liên tiếp của các mẫu từ 1 đến L chiếm dụng khung t .

Giả sử rằng tín hiệu (trong mỗi kênh) được lấy mẫu ở 48 kHz, không mất tính tổng quát. Độ dài khung được cố định ở 20 ms, tức là $L = 960$ mẫu liên tiếp, không mất tính tổng quát. Nói cách khác, ví dụ, có thể sử dụng độ dài khung là $L = 640$ mẫu để lấy mẫu ở 32 kHz.

Phân tích PCA/KLT và phép biến đổi PCA/KLT mà được mô tả sau đây được thực hiện trong miền thời gian. Do đó được hiểu là chúng giữ nguyên ở đây trong miền thời gian mà không cần thiết phải thực hiện việc biến đổi dải phụ hoặc tổng quát hơn biến đổi tần số.

Tại mỗi khung, khối 300 của bộ mã hóa áp dụng tiền xử lý (tùy chọn) để thu tín hiệu đầu vào được xử lý trước được biểu thị Y . Nó có thể là bộ lọc thông cao (với tần số ngưỡng điển hình ở 20Hz) của mỗi khung 20 ms mới của các kênh tín hiệu đầu vào. Phép toán này cho phép loại bỏ thành phần liên tục có thể làm sai lệch ước tính của ma trận hiệp biến để đầu ra tín hiệu từ khối 300 có thể được xem xét để có số trung bình không. Hàm truyền được biểu thị $H_{pre}(z)$, để ta có cho mỗi kênh: $X_i(z) = H_{pre}(z)Y_i(z)$. Nếu khối 300 không được áp dụng ta có $\mathbf{X} = \mathbf{Y}$. Bộ lọc thông thấp trong khối 340 cũng có thể được áp dụng để thực hiện đa-đơn mã hóa nhưng khi khối 300 được áp dụng, việc lọc thông cao trong quá trình tiền xử lý đơn mã hóa có thể được sử dụng trong khối 340 tốt hơn là vô hiệu, để tránh việc lặp lại cùng việc tiền xử lý và do đó làm giảm độ phức tạp tổng thể.

Hàm truyền được biểu thị $H_{pre}(z)$ ở trên có thể thuộc loại:

$$H_{pre}(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

bằng cách áp dụng việc lọc này cho mỗi trong số n kênh của tín hiệu đầu vào, trong đó các hệ số có thể được thể hiện trong bảng sau đây:

	8 kHz	16 kHz	32 kHz	48 kHz
b_0	0,9889542480671 40	0,994461788958.1 95	0,9972270499044 70	0,9981505111904 52
b_1	- 1,9779084961342 80	- 1,9889235779163 90	- 1,9944540998089 40	- 1,9963010223809 04
b_2	0,9889542480671 40	0,9944617889581 95	0,9972270499044 70	0,9981505111904 52
a_1	1,9777864837767 64	1,9888929058996 53	1,9944464105419 27	1,9962976017691 22
a_2	- 0,9780305084917 96	- 0,9889542499331 27	- 0,9944617890759 54	- 0,9963044429926 86

Nói cách khác, kiểu bộ lọc khác có thể được sử dụng, ví dụ bộ lọc Butterworth bậc sáu có tần số là 50 Hz.

Theo một số biến thể, việc tiền xử lý có thể bao gồm bước ma trận cố định mà có thể duy trì cùng số lượng các kênh hoặc làm giảm số lượng các kênh. Ví dụ về ma trận được áp dụng cho bốn kênh của tín hiệu ambisonic ở định dạng B được cho sau đây:

$$M_{B \rightarrow A} = \begin{bmatrix} 1/2 & \frac{1}{\sqrt{6}} & 0 & \frac{1}{\sqrt{12}} \\ 1/2 & \frac{-1}{\sqrt{6}} & 0 & \frac{1}{\sqrt{12}} \\ 1/2 & 0 & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{12}} \\ 1/2 & 0 & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{12}} \end{bmatrix}$$

Lưu ý rằng trong trường hợp này việc tiền xử lý này sẽ phải được đảo ngược khi giải mã bằng cách áp dụng ma trận của tín hiệu được giải mã qua $M_{A \rightarrow B} = M_{B \rightarrow A}^{-1}$, để tìm các kênh trong định dạng ban đầu.

Khối 310 tiếp theo ước tính, tại mỗi khung t , ma trận biến đổi thu được bằng cách xác định các vector đặc trưng bởi PCA/KLT và xác minh rằng ma trận biến đổi được tạo ra bởi các vector đặc trưng này thực sự đặc trưng cho phép quay. Các chi tiết của hoạt động của khối 310 được đưa ra thêm sau đây với tham chiếu đến FIG.4. Ma trận biến đổi này thực hiện ma trận của các kênh để giải tương quan chúng, khiến nó có thể áp dụng kiểu đa-đơn mã hóa độc lập bởi khối 340. Như được thể hiện chi tiết sau đây, khối 310 gửi các chỉ số lượng tử hóa của bộ ghép kênh biểu diễn ma trận biến đổi và, một cách tùy chọn, mã hóa thông tin số lượng các phép nội suy của ma trận biến đổi, trên mỗi khung con của khung hiện thời t , cũng như được thể hiện chi tiết sau đây.

Khối 320 xác định cấp phát tốc độ bit tối ưu cho mỗi kênh (sau phép biến đổi PCA/KLT) dựa trên ngân sách B bit đã cho. Khối này tìm kiếm sự phân bố của tốc độ bit giữa các kênh bằng cách tính toán điểm số cho mỗi kết hợp khả thi của các tốc độ bit; cấp phát tối ưu được tìm thấy bằng cách tìm kiếm kết hợp để tối đa hóa điểm số này.

Một vài tiêu chuẩn có thể được sử dụng để xác định điểm số cho mỗi kết hợp.

Ví dụ, số lượng các tốc độ bit khả thi đối với đơn mã hóa của kênh có thể được giới hạn với chín tốc độ bit riêng biệt của bộ mã hóa-giải mã EVS có dải audio siêu rộng: 9,6; 13,2; 16,4; 24,4; 32; 48; 64; 96 và 128 kbps. Tuy nhiên, nếu bộ mã hóa-giải mã theo sáng chế hoạt động ở tốc độ bit đã cho được liên kết với ngân sách của B bit trong khung hiện thời của chỉ số t, nói chung chỉ tập con có các tốc độ bit được liệt kê này có thể được sử dụng. Ví dụ, nếu bộ mã hóa-giải mã tốc độ bit được cố định ở $4 \times 13,2 = 52,8$ kbps để biểu diễn bốn kênh và nếu mỗi kênh nhận ngân sách tối thiểu là 9,6 kbps để đảm bảo dải siêu rộng cho mỗi trong số các kênh, các kết hợp khả thi của các tốc độ bit để mã hóa các kênh riêng biệt phải tương ứng với giới hạn mà tốc độ bit được sử dụng còn lại thấp hơn tốc độ bit khả dụng mà tương ứng với:

$$B_{multimono} = B - B_{overhead},$$

trong đó $B_{overhead}$ là ngân sách bit đối với thông tin bổ sung được mã hóa trên mỗi khung (cấp phát bit + dữ liệu quay) như được mô tả sau đây. Ví dụ, $B_{overhead}$ có thể theo bậc của $B_{overhead} = 55$ bit trên mỗi khung 20 ms (tức là 2,75 kbps) đối với trường hợp mã hóa ambisonic bốn kênh; nó bao gồm 51 bit để mã hóa ma trận quay và 4 bit (như được mô tả sau đây) để mã hóa cấp phát bit để mã hóa các kênh riêng biệt. Đối với tốc độ bit tổng thể là $4 \times 13,2 = 52,8$ kbps, do đó nó để lại ngân sách là $B_{multimono} = 50,05$ kbps.

Về các tốc độ bit trên mỗi kênh, điều này cho các hoán vị sau đây của các tốc độ bit trên mỗi kênh:

- Singleton (9,6, 9,6, 9,6, 9,6) - tổng = 38,4
- Các hoán vị của (13,2, 9,6, 9,6, 9,6) - tổng = 42 kbps
- Các hoán vị của (13,2, 13,2, 9,6, 9,6) - tổng = 45,6 kbps
- Các hoán vị của (13,2, 13,2, 13,2, 9,6) - tổng = 49,2 kbps
- Các hoán vị của (16,4, 9,6, 9,6, 9,6) - tổng = 45,2 kbps

- Các hoán vị của (16,4, 13,2, 9,6, 9,6) - tổng = 48,8 kbps

Ta có thể thấy rằng một số kết hợp chú ý đến giới hạn ngân sách tối đa có tốc độ bit thấp hơn nhiều hơn kết hợp khác, và cuối cùng chỉ hai kết hợp liên quan có thể được giữ lại:

- Các hoán vị của (13,2, 13,2, 13,2, 9,6) – 4 trường hợp và tốc độ bit chưa được sử dụng là $50,5 - 49,2 = 1,3$ kbps

- và các hoán vị của (16,4, 13,2, 9,6, 9,6) – 12 trường hợp và tốc độ bit chưa được sử dụng là $50,5 - 48,8 = 1,7$ kbps

Điều này có thể minh họa rằng mười sáu kết hợp được quan tâm đặc biệt và có thể được mã hóa trong 4 bit (16 giá trị). Ngoài ra, số lượng các bit nhất định vẫn có khả năng chưa được sử dụng phụ thuộc vào việc cấp phát được chọn.

Ta có thể thấy rằng việc mã hóa ma trận hóa tương thích dựa trên việc xử lý PCA/KLT và cho phép cấp phát bit linh hoạt có thể khiến các bit chưa được sử dụng và, đối với một số kênh, tốc độ bit thấp hơn (ví dụ 9,6 kbps) so với tốc độ bit được phân bố đồng đều trong số mỗi trong số các kênh (ví dụ 13,2 kbps trên mỗi kênh).

Để cải thiện tình huống này, sau đó khối 320 có thể đánh giá tất cả các kết hợp khả thi (liên quan) của các tốc độ bit đối với 4 kênh là kết quả của phép biến đổi PCA/KLT (xuất ra từ khối 310) và chỉ định điểm số cho chúng. Điểm số này được tính toán dựa trên:

- năng lượng của mỗi kênh, và

- điểm số trung bình có thể được lưu trữ từ trước và là kết quả của các kiểm tra chủ quan hoặc khách quan; điểm số này, được biểu thị MOS (đối với "Mean Opinion Score" (điểm số ý kiến trung bình), là điểm số trung bình đối với nhóm người thử nghiệm), được liên kết với tốc độ bit được cấp phát.

Điểm số này sau đó có thể được xác định bởi phương trình

$$S(b_{t,1}, \dots, b_{t,n}) = \sum_{i=1}^n Q(b_{t,i}) \cdot E_i$$

trong đó E_i là năng lượng trong khung hiện thời (của chỉ số t) của tín hiệu $s(l)$, $l = \dots, L - 1$ trên kênh i , với:

$$E_i = \sum_{l=0}^{L-1} s^2(l)$$

Cấp phát tối ưu có thể là sao cho:

$$b_{t,1}^{opt}, \dots, b_{t,n}^{opt} = \arg \max_{b_{t,1}, \dots, b_{t,n} | \sum_{i=1}^n b_{t,i} \leq B} S(b_{t,1}, \dots, b_{t,n})$$

Nói cách khác, hệ số E_i có thể được cố định ở giá trị được lấy bởi giá trị đặc trưng được liên kết với kênh i là kết quả từ sự phân tích thành các giá trị đặc trưng của tín hiệu là đầu vào đến khối 310 và sau hoán vị dấu khả thi.

Điểm số MOS $Q(b_i)$ tốt hơn là điểm số chất lượng chủ quan của bộ mã hóa-giải mã được sử dụng để đa-đơn mã hóa trong khối 340 đối với ngân sách b_i (trong số lượng các bit) trên mỗi khung 20 ms tương ứng với tốc độ bit $R_i = 50 b_i$ (trong bit/giây). Để bắt đầu với, ta có thể sử dụng các điểm số MOS chủ quan (trung bình) của bộ mã hóa chuẩn hóa EVS được cho bởi:

κ_i	0	1	2	3	4	5	6	7	8
b_i	192	264	328	488	640	960	1280	1920	2560
R_i	9600	13200	16400	24400	32000	48000	64000	96000	128000
$Q(b_i)$	3,62	3,79	4,25	4,60	4,53	4,82	4,83	4,85	4,87

Nói cách khác, các giá trị điểm số MOS khác đối với mỗi trong số các tốc độ bit được liệt kê có thể được khai triển từ các kiểm tra khác (chủ quan hoặc khách quan) dự báo chất lượng của bộ mã hóa-giải mã. Cũng có thể tương thích với các điểm số MOS được sử dụng trong khung hiện thời, theo sự phân loại của kiểu tín hiệu (ví dụ tín hiệu tiếng nói không có nhiều nền, hoặc tiếng nói với nhiều môi trường, hoặc âm nhạc hoặc nội dung được pha trộn), bằng cách sử dụng lại các phương pháp phân loại được thực thi bằng bộ mã hóa-giải mã EVS và bằng cách áp dụng chúng cho W kênh của tín hiệu đầu vào ambisonic trước khi thực hiện cấp phát bit. Điểm số MOS cũng có thể tương ứng với điểm số trung bình là kết quả của

các loại phương pháp và các tỷ lệ đánh giá khác nhau: MOS (tuyệt đối) từ 1 đến 5, DMOS (từ 1 đến 5), MUSHRA (từ 0 đến 100).

Theo một biến thể trong đó bộ mã hóa EVS được thay thế bởi bộ mã hóa-giải mã khác, danh sách của các tốc độ bit b_i và các số điểm $Q(b_i)$ có thể được thay thế trên cơ sở của bộ mã hóa-giải mã khác này. Cũng có thể thêm các tốc độ bit mã hóa bổ sung cho bộ mã hóa EVS và do đó bổ sung danh sách các tốc độ bit và các điểm số MOS, hoặc thậm chí để sửa đổi bộ mã hóa EVS và có khả năng là các điểm số MOS kết hợp.

Theo phương án thay thế khác, việc cấp phát giữa các kênh được làm mịn bằng cách lấy trọng số năng lượng với lũy thừa α trong đó α lấy giá trị giữa 0 và 1. Bằng cách thay đổi giá trị của α , do đó ta có thể điều khiển sự ảnh hưởng của năng lượng trong việc cấp phát: α càng gần với 1, năng lượng càng quan trọng hơn trong điểm số, và do đó cấp phát giữa kênh càng không đồng đều. Ngược lại, α càng gần đến 0, năng lượng càng ít quan trọng và sự cấp phát giữa kênh được phân bố đồng đều hơn. Do đó, điểm số này được thể hiện ở dạng:

$$S(b_{t,1}, \dots, b_{t,n}) = \sum_{i=1}^n Q(b_{t,i}) \cdot E_i^\alpha$$

Theo phương án thay thế khác, để khiến việc cấp phát bền hơn, trọng số thứ hai có thể được thêm vào hàm điểm số để phạt các thay đổi của tốc độ bit liên khung. Điểm phạt được thêm vào điểm số nếu kết hợp tốc độ bit không giống trong khung t cũng như trong khung $t - 1$. Sau đó điểm số được thể hiện ở dạng:

$$S(b_{t,1}, \dots, b_{t,n}) = \sum_{i=1}^n Q(b_{t,i}) \cdot E_i^\alpha \cdot (1 + \beta_i)$$

trong đó β_i có hằng số định trước là giá trị của nó (ví dụ 0,1) khi $b_{t,i} = b_{t-1,i}$, và $\beta_i = 0$ khi $b_{t,i} \neq b_{t-1,i}$.

Trọng số bổ sung này khiến nó có thể giới hạn các dao động quá thường xuyên trong tốc độ bit giữa kênh. Với trọng số này, chỉ có các thay đổi đáng kể trong năng lượng mang lại sự thay đổi trong tốc độ bit. Ngoài ra, giá trị của hằng số có thể được thay đổi để điều chỉnh độ ổn định của việc cấp phát.

Tham chiếu lại đến FIG.3, ngay khi tốc độ bit được tính toán cho mỗi khung, tốc độ bit này được mã hóa bởi khối 330, ví dụ hoàn toàn cho tất cả kết hợp tốc độ bit. Trong trường hợp có 9 tốc độ bit và 4 kênh, tốc độ bit được yêu cầu là $\lceil \log_2(9^4) \rceil = 13$ bit, trong đó $\lceil . \rceil$ tương ứng với làm tròn lên tới số nguyên tiếp theo. Kết hợp của 4 tốc độ bit có thể được mã hóa ở dạng của chỉ số: $\sum_{i=1}^n 9^i \kappa_i$. Tuy nhiên, người ta có thể muốn liệt kê (ban đầu, ngoại tuyến) các kết hợp khác nhau của các tốc độ bit liên quan đối với ngân sách bit đã cho và để sử dụng tốc độ bit tối thiểu để biểu diễn các kết hợp này. Chỉ số sau đó có thể được biểu diễn bởi loại "mã hoán vị" + "độ dịch kết hợp" của mã hóa; ví dụ, theo ví dụ trong đó ta sử dụng chỉ số 4 bit để mã hóa 16 kết hợp tốc độ bit bao gồm 4 hoán vị của (13,2, 13,2, 13,2, 9,6) và 12 hoán vị của (16,4, 13,2, 9,6, 9,6), ta có thể sử dụng các chỉ số 0-3 để mã hóa 4 hoán vị khả thi thứ nhất (với độ dịch tại 0 và mã nằm trong khoảng từ 0 đến 3) và các chỉ số 4-15 để mã hóa 12 hoán vị khả thi khác (với độ dịch tại 4 và mã từ 0 đến 11).

Tham chiếu lại đến FIG.3, khối ghép kênh 350 lấy đầu vào n kênh ma trận đến từ khối 310 và các tốc độ bit được cấp phát cho mỗi kênh đến từ khối 320 để sau đó mã hóa tách biệt các kênh khác nhau với bộ mã hóa-giải mã lỗi mà tương ứng với bộ mã hóa-giải mã EVS, chẳng hạn. Nếu bộ mã hóa-giải mã lỗi được sử dụng cho phép mã hóa âm thanh nổi hoặc đa kênh, cách thức đa-đơn có thể được thay thế bằng mã hóa đa âm thanh nổi hoặc đa kênh. Ngay khi các kênh được mã hóa, dòng bit kết hợp được gửi tới bộ ghép kênh (khối 350).

Trong các khung trong đó một phần của ngân sách tổng thể không được sử dụng hoàn toàn, bộ ghép kênh (khối 350) có thể áp dụng nhồi bit 0 để tiếp cận ngân sách bit được cấp phát cho khung hiện thời, tức là $B - \sum_{i=1}^n b_{t,i}^{opt}$ bit. Nói cách khác, ngân sách bit còn lại có thể được phân bổ lại để mã hóa các kênh được biến đổi để sử dụng toàn bộ ngân sách khả dụng và nếu đa-đơn mã hóa dựa trên công nghệ loại EVS, thì thuật toán mã hóa 3GPP EVS được chỉ định có thể được sửa đổi để đưa vào các tốc độ bit bổ sung. Trong trường hợp này, cũng có thể tích hợp các tốc độ bit bổ sung này trong bảng xác định sự tương ứng giữa b_i và $Q(b_i)$.

Bit cũng có thể được dành để có thể chuyển giữa hai chế độ mã hóa:

- mã hóa theo sáng chế với việc mã hóa ma trận quay, và

- mã hóa theo sáng chế với ma trận quay được giới hạn với ma trận đơn vị (do đó không được truyền) mà số lượng đó dùng cho đa-đơn mã hóa trực tiếp nếu ma trận quay của khung trước cũng là ma trận đơn vị (ví dụ khi tín hiệu ambisonic bao gồm các nguồn âm thanh rất khuếch tán hoặc nhiều nguồn phát tán theo không gian xung quanh các hướng ưu tiên nhất định, trong đó trường hợp mà các kênh ambisonic ít tương quan hơn đối với việc trộn âm thanh với nhiều nguồn điểm cô lập hơn).

Lựa chọn giữa hai chế độ này suy ra việc sử dụng bit trong dòng để chỉ báo xem khung hiện thời có sử dụng ma trận quay được giới hạn với ma trận đơn vị mà không có việc truyền của các tham số quay (bit = 0) hoặc nếu ma trận quay được mã hóa (bit = 1). Khi bit = 0, có thể theo một số biến thể để sử dụng việc cấp phát của các bit cố định cho các kênh riêng biệt và không truyền cấp phát bit.

Giờ tham chiếu được tạo ra với FIG.4 để mô tả chi tiết khối 310 mà áp dụng sự phân tích và biến đổi PCA/KLT. Trong khối này, bộ mã hóa tính toán ma trận hiệp biến từ các kênh ambisonic (được xử lý trước) trong khối 400:

$$C = \frac{1}{L-1} \mathbf{X}\mathbf{X}^T$$

Nói cách khác, ma trận này có thể được thay thế bằng ma trận tương quan, trong đó các kênh được chuẩn hóa trước bằng độ lệch chuẩn tương ứng của chúng, hoặc nói chung các trọng số phản ánh tầm quan trọng tương đối có thể được áp dụng cho mỗi trong số các kênh; hơn nữa, số hạng chuẩn hóa $1/(L-1)$ có thể được lược bỏ hoặc được thay thế bởi giá trị khác (ví dụ $1/L$). Các giá trị C_{ij} tương ứng với sự thay đổi giữa x_i và x_j .

Sau đó, bộ mã hóa thực hiện, trong khối 410, sự phân tích thành các giá trị đặc trưng (EVD đối với "Phân tích giá trị đặc trưng"), bằng cách tính toán các giá trị đặc trưng và các vector đặc trưng của ma trận C . Các vector đặc trưng được biểu thị V_t ở đây để chỉ báo chỉ số của khung t vì các vector đặc trưng V_{t-1} thu được trong khung trước của chỉ số $t-1$ tốt hơn là được lưu trữ và được sử dụng liên tiếp. Các giá trị đặc trưng được biểu thị $\lambda_1, \lambda_2, \dots, \lambda_n$.

Nói cách khác, sự phân tích giá trị suy biến (singular value decomposition, SVD) của các kênh X được xử lý trước có thể được sử dụng. Do đó, ta thu được các vector suy biến (U ở bên trái và V ở bên phải) và các giá trị suy biến σ_i . Trong trường hợp này, ta có thể xét đến các giá trị đặc trưng λ_i là $\lambda_i = \sigma_i^2$ và các vector đặc trưng V_t được cho bởi n vector suy biến (cột) ở bên trái U .

Sau đó, bộ mã hóa áp dụng, trong khối 420, hoán vị đầu thứ nhất của các cột của ma trận biến đổi đối với khung t (trong đó các cột là các vector đặc trưng) để tránh chênh lệch quá nhiều với ma trận biến đổi của khung trước $t-1$, sẽ phát sinh vấn đề với các lần nháp ở biên với khung trước.

Do đó, ngay khi bản nháp thô của ma trận biến đổi thu được đối với khung t , khối 430 lấy n vector đặc trưng được ước tính $V_t = v_{t,0}, \dots, v_{t,n}$ từ khung hiện thời của chỉ số t và n vector đặc trưng V_{t-1} được lưu trữ từ khung trước của chỉ số $t-1$, và áp dụng hoán vị đầu trên các vector V_t được ước tính để chúng gần nhất có thể với V_{t-1} . Do đó các vector đặc trưng của khung t được hoán vị để cơ sở kết hợp gần nhất có thể với cơ sở của khung $t-1$. Điều này có hiệu quả để cải thiện sự liên tiếp của các khung của các tín hiệu được biến đổi (sau khi ma trận biến đổi được áp dụng cho các kênh).

Giới hạn khác là ma trận biến đổi phải tương ứng với phép quay. Giới hạn này đảm bảo rằng bộ mã hóa có thể chuyển đổi ma trận biến đổi thành các góc Euler được tổng quát hóa (khối 430) để lượng tử hóa chúng (khối 440) với ngân sách bit định trước như được thấy ở trên. Với mục đích này, định thức của ma trận này phải dương (thường bằng +1).

Tốt hơn là, hoán vị đầu tối ưu thu được trong hai bước:

- Bước thứ nhất (S4 trên FIG.2 được thể hiện ở trên) khớp với các vector gần nhất giữa hai khung, chỉ chú ý đến trục và không phải hướng (định hướng) của trục. Vấn đề này có thể được đề ra dưới dạng vấn đề tổ hợp của việc phân phối tác vụ, trong đó mục tiêu là tìm kiếm cấu hình mà giảm thiểu chi phí. Chi phí này có thể được xác định ở đây là vết của giá trị tuyệt đối của tương quan liên kết giữa các ma trận vector đặc trưng của các khung t và $t-1$.

$$C_t = tr(abs(corr(V_t, V_{t-1})))$$

trong đó $tr(.)$ biểu thị vết của ma trận, các lượng $abs(.)$ để áp dụng phép toán giá trị tuyệt đối cho tất cả các hệ số của ma trận, và $corr(V1, V2)$ cho ma trận tương quan giữa các vector $V1$ và $V2$.

Theo một phương án, phương pháp “Hungarian” (hoặc “thuật toán Hungarian”) được sử dụng để xác định nhiệm vụ tối ưu mà cho hoán vị của các vector đặc trưng của khung t ;

- Bước thứ hai (S6 trên FIG.2) bao gồm xác định hướng/định hướng của mỗi vector đặc trưng được hoán vị. Khối 420 tính toán tương quan liên kết giữa các vector đặc trưng $\tilde{V}_t$ được hoán vị của khung t và vector đặc trưng của khung $t - 1$

$$\Gamma_t = corr(\tilde{V}_t, V_{t-1})$$

Nếu giá trị về sự chéo hóa của ma trận tương quan liên kết Γ_t là âm, điều này biểu thị sự thay đổi về dấu giữa các hướng của các vector đặc trưng. Sau đó, sự đảo dấu được thực hiện đối với vector đặc trưng tương ứng trong $\tilde{V}_t$.

Khi kết thúc hai bước, ma trận biến đổi tại khung t được chỉ định bởi V_t sao cho tại khung tiếp theo, ma trận được lưu trữ trở thành V_{t-1} .

Nói cách khác, việc tìm kiếm đối với hoán vị dấu tối ưu có thể được thực hiện bằng cách tính toán sự thay đổi của cơ sở ma trận $V_{t-1}^{-1}V_t$ hoặc $V_t V_{t-1}^{-1}$ được chuyển đổi thành 3D hoặc 4D và bằng cách chuyển đổi sự thay đổi này của ma trận cơ sở lần lượt thành một bộ bốn đơn vị hoặc hai bộ bốn đơn vị. Sau đó việc tìm kiếm này trở thành việc tìm kiếm lân cận gần nhất với từ điển biểu diễn tập hợp của các hoán vị dấu khả thi. Ví dụ, trong trường hợp 4D mười hai hoán vị chẵn khả thi (trong số tổng 24 hoán vị) của 4 giá trị được liên kết với các cặp bộ bốn đơn vị sau đây được viết dưới dạng các vector 4D:

- (1, 0, 0, 0) và (1, 0, 0, 0)
- (0, 0, 0, 1) và (0, 0, -1, 0)
- (0, 1, 0, 0) và (0, 0, 0, -1)
- (0, 0, 1, 0) và (0, -1, 0, 0)]

- $(0,5, -0,5, -0,5, -0,5)$ và $(0,5, 0,5, 0,5, 0,5)$
- $(0,5, 0,5, 0,5, 0,5)$ và $(0,5, -0,5, -0,5, -0,5)$
- $(0,5, -0,5, 0,5, -0,5)$ và $(0,5, -0,5, 0,5, 0,5)$
- $(0,5, -0,5, 0,5, 0,5)$ và $(0,5, -0,5, -0,5, 0,5)$
- $(0,5, 0,5, -0,5, 0,5)$ và $(0,5, 0,5, -0,5, -0,5)$
- $(0,5, -0,5, -0,5, 0,5)$ và $(0,5, 0,5, -0,5, 0,5)$
- $(0,5, 0,5, -0,5, -0,5)$ và $(0,5, 0,5, 0,5, -0,5)$
- $(0,5, 0,5, 0,5, -0,5)$ và $(0,5, -0,5, 0,5, -0,5)$

Việc tìm kiếm đối với hoán vị tối ưu (chẵn) có thể được thực hiện bằng cách sử dụng danh sách nêu trên dưới dạng từ điển của các cặp bộ bốn định trước và bằng cách thực hiện việc tìm kiếm lân cận gần nhất đối với cặp bộ bốn được liên kết với sự thay đổi của ma trận cơ sở. Ưu điểm của phương pháp này là việc sử dụng lại các tham số quay của bộ bốn và loại cặp bộ bốn.

Phép toán này được thực thi trong khối 460 tiếp theo giả sử rằng ma trận biến đổi sau khi hoán vị dấu thực sự là ma trận quay; ma trận biến đổi nhất thiết là ma trận unita, nhưng định thức của nó cũng phải bằng 1

$$\det(V_t) = 1$$

Tuy nhiên, ma trận biến đổi thu được từ các khối 410 và 420 (sau khi EVD và các hoán vị dấu) là ma trận trực giao (unita) có thể có định thức là -1 hoặc 1, có nghĩa là ma trận phản xạ hoặc ma trận quay.

Nếu ma trận biến đổi là ma trận phản xạ (nếu định thức của nó bằng -1), nó có thể được sửa đổi thành ma trận quay bằng cách đảo vector đặc trưng (ví dụ vector đặc trưng được liên kết với giá trị thấp nhất) hoặc bằng cách đảo hai cột (các vector đặc trưng).

Các phương pháp nhất định của việc phân tích vector đặc trưng (ví dụ bởi phép quay đã cho) hoặc của việc phân tích giá trị suy biến có thể dẫn đến các ma trận biến đổi về bản chất là các ma trận quay (với định thức là +1); trong trường hợp này, bước xác minh rằng định thức là +1 sẽ là tùy chọn.

Khối 430 chuyển đổi ma trận quay thành các tham số. Theo phương án ưu tiên, sự biểu diễn góc được sử dụng đối với sự lượng tử hóa (6 góc Euler suy rộng đối với trường hợp 4D, 3 góc Euler cho trường hợp 3D, và một góc trong 2D). Đối với trường hợp ambisonic (bốn kênh) ta thu được sáu góc Euler suy rộng theo phương pháp được mô tả trong bài báo "Generalization of Euler Angles to N-Dimensional Orthogonal Matrices" bởi David K. Hoffman, Richard C. Raffenetti, và Klaus Ruedenberg, được công bố trong Journal of Mathematical Physics 13, 528 (1972); đối với trường hợp của ambisonic phẳng (ba kênh) ta thu được ba góc Euler, và đối với trường hợp âm thanh nổi ta thu được góc quay theo các phương pháp đã biết trong tình trạng kỹ thuật. Các giá trị của các góc được lượng tử hóa trong khối 440 với ngân sách bit định trước. Theo phương án ưu tiên, sự lượng tử hóa vô hướng được sử dụng và kích thước bước lượng tử hóa là ví dụ giống nhau đối với mỗi góc. Ví dụ, trong trường hợp có 4 kênh ta mã hóa 6 góc Euler suy rộng với $3 \times (8+9) = 51$ bit (3 góc được xác định nằm trong khoảng $[-\pi/2, \pi/2]$ được mã hóa trong 8 bit có kích thước bước là $\pi/256$ và 3 góc khác được xác định trong khoảng $[-\pi, \pi]$ được mã hóa trong 9 bit có kích thước bước là $\pi/256$). Các chỉ số lượng tử hóa của ma trận biến đổi được gửi tới bộ ghép kênh (khối 350). Ngoài ra, khối 440 có thể chuyển đổi các tham số được lượng tử hóa thành ma trận quay lượng tử hóa $\hat{V}_t$, nếu các tham số được sử dụng để lượng tử hóa không khớp với các tham số được sử dụng cho phép nội suy.

Nói cách khác, các khối 430 và 440 có thể được thay thế như sau:

- Khối 430 có thể thực hiện chuyển đổi các ma trận quay thành cặp bộ bốn đơn vị (trường hợp có 4 kênh), thành bộ bốn đơn vị (trường hợp có 3 kênh), và thành góc (trường hợp có 2 kênh).

Sự chuyển đổi này thành cặp bộ bốn đối với trường hợp 4D có thể được tiến hành đối với ma trận quay mà có các hệ số được biểu thị $R[i,j]$, $i,j=0\dots3$, bằng mã giả sau đây:

Tính toán ma trận kết hợp $A[i,j]$ với:

$$A[0,0] = R[0,0] + R[1,1] + R[2,2] + R[3,3]$$

$$A[1,0] = R[1,0] - R[0,1] + R[3,2] - R[2,3]$$

$$A[2,0] = R[2,0] - R[3,1] - R[0,2] + R[1,3]$$

$$A[3,0] = R[3,0] + R[2,1] - R[1,2] - R[0,3]$$

$$A[0,1] = R[1,0] - R[0,1] - R[3,2] + R[2,3]$$

$$A[1,1] = -R[0,0] - R[1,1] + R[2,2] + R[3,3]$$

$$A[2,1] = -R[3,0] - R[2,1] - R[1,2] - R[0,3]$$

$$A[3,1] = R[2,0] - R[3,1] + R[0,2] - R[1,3]$$

$$A[0,2] = R[2,0] + R[3,1] - R[0,2] - R[1,3]$$

$$A[1,2] = R[3,0] - R[2,1] - R[1,2] + R[0,3]$$

$$A[2,2] = -R[0,0] + R[1,1] - R[2,2] + R[3,3]$$

$$A[3,2] = -R[1,0] - R[0,1] - R[3,2] - R[2,3]$$

$$A[0,3] = R[3,0] - R[2,1] + R[1,2] - R[0,3]$$

$$A[1,3] = -R[2,0] - R[3,1] - R[0,2] - R[1,3]$$

$$A[2,3] = R[1,0] + R[0,1] - R[3,2] - R[2,3]$$

$$A[3,3] = -R[0,0] + R[1,1] + R[2,2] - R[3,3]$$

$$A = A/4$$

Tính toán 2 bộ bốn từ ma trận kết hợp

$$A2 = \text{square}(A) \# \text{square of coefficients}$$

$$q1 = \text{sqrt}(A2.\text{sum}(\text{axis} = 1)) \# \text{sum the rows}$$

$$q2 = \text{sqrt}(A2.\text{sum}(\text{axis} = 0)) \# \text{sum the column}$$

Xác định các dấu

$$\text{Đối với } k=0..3 : \text{ nếu } \text{sign}(A[i,k]) < 0, \text{ thì } q2[k] = -q2[k]$$

$$\text{Đối với } k=0..3 : \text{ nếu } \text{sign}(A[k,j]) \neq \text{sign}(q1[k]*q2[j]), \text{ thì } q1[k] = -q1[k]$$

Sự chuyển đổi với bộ bốn cho trường hợp 3D có thể được tiến hành như sau
đối với ma trận $R[i,j]$ $i,j=0\dots 2$ có kích thước 3×3 :

Tính toán ma trận kết hợp được đơn giản hóa:

$$q[0] = (R[0,0] + R[1,1] + R[2,2] + 1)^2 + (R[2,1] - R[1,2])^2 + (R[0,2] - R[2,0])^2 + (R[1,0] - R[0,1])^2$$

$$q[1] = (R[2,1] - R[1,2])^2 + (R[0,0] - R[1,1] - R[2,2] + 1)^2 + (R[1,0] + R[0,1])^2 + (R[2,0] + R[0,2])^2$$

$$q[2] = (R[0,2] - R[2,0])^2 + (R[1,0] + R[0,1])^2 + (R[1,1] - R[0,0] - R[2,2] + 1)^2 + (R[2,1] + R[1,2])^2$$

$$q[3] = (R[1,0] - R[0,1])^2 + (R[2,0] + R[0,2])^2 + (R[2,1] + R[1,2])^2 + (R[2,2] - R[0,0] - R[1,1] + 1)^2$$

$$\text{Đối với } i=0..3 : q[i] = \sqrt{q[i]}/4$$

Tính toán bộ bốn q

$$\text{nếu } (R[2,1] - R[1,2]) < 0, q[1] = -q[1]$$

$$\text{nếu } (R[0,2] - R[2,0]) < 0, q[2] = -q[2]$$

$$\text{nếu } (R[1,0] - R[0,1]) < 0, q[3] = -q[3]$$

Đối với trường hợp của ma trận 2x2, góc đã được tính toán theo các phương pháp đã biết theo giải pháp kỹ thuật trước.

Theo một số biến thể, các bộ bốn đơn vị q_1 , q_2 (trường hợp 4D) và q (trường hợp 3D) có thể được chuyển đổi thành các phép biểu diễn trục-góc đã biết theo giải pháp kỹ thuật trước.

- Khối 440 có thể thực hiện lượng tử hóa trong miền được chỉ báo:

* Trường hợp có 4 kênh: cặp bộ bốn đơn vị q_1 và q_2 được lượng tử hóa bởi từ điển lượng tử hóa hình cầu theo chiều 4; theo quy ước, q_1 được lượng tử hóa với từ điển bán cầu (vì q_1 và $-q_1$ tương ứng với cùng một phép quay 3D) và q_2 được lượng tử hóa với từ điển hình cầu. Các ví dụ về các từ điển có thể được cho bởi các điểm định trước dựa trên các khối đa diện có 4 chiều; theo một số biến thể, có thể lượng tử hóa phép biểu diễn trục-góc kết hợp kép sẽ tương đương với cặp bộ bốn;

* Trường hợp có 3 kênh: bộ bốn đơn vị được lượng tử hóa bằng từ điển lượng tử hóa hình cầu trong 4 chiều - các ví dụ của các từ điển có thể được cho bởi các điểm định trước dựa trên các khối đa diện có 4 chiều;

* Trường hợp có 2 kênh: góc được lượng tử hóa bằng sự lượng tử hóa vô hướng đồng đều.

Giờ ta mô tả khối 460 đối với phép nội suy của các ma trận quay giữa hai khung liên tiếp. Nó làm mịn sự gián đoạn trong các kênh sau khi áp dụng các ma trận này. Thông thường, nếu hai tập hợp góc hoặc các bộ bốn quá khác với khung trước $t-1$ đến khung tiếp theo t , các tiếng kêu lách tách là mối quan tâm nếu sự chuyển đổi trơn tru không được áp dụng giữa hai khung này, trong các khung con giữa hai khung này. Sau đó, phép nội suy chuyển tiếp được tiến hành giữa ma trận quay được tính toán đối với khung $t-1$ và ma trận quay được tính toán cho khung t . Bộ mã hóa nội suy, trong khối 460, sự biểu diễn (được lượng tử hóa) của phép quay giữa khung hiện thời và khung trước để tránh các dao động quá nhanh của các kênh khác nhau sau phép biến đổi. Số lượng các phép nội suy có thể cố định (bằng giá trị định trước) hoặc tương thích. Sau đó, mỗi khung được chia thành các khung con là hàm của số lượng các phép nội suy được xác định trong khối 450. Do đó, nếu phép nội suy tương thích được sử dụng, khối 450 có thể mã hóa trong số lượng các bit được chọn, số lượng các phép nội suy cần được thực hiện, và do đó số lượng các khung con cần được cung cấp, trong trường hợp trong đó số lượng này được xác định tương thích; trong trường hợp của phép nội suy cố định, không có thông tin phải được mã hóa.

Tiếp theo, khối 460 chuyển đổi các ma trận quay thành miền xác định biểu diễn ma trận quay. Khung này được chia thành các khung con, và trong miền được chọn, phép nội suy được tiến hành cho mỗi khung con.

Đối với tín hiệu đầu vào ambisonic bậc thứ nhất (với 4 kênh W, X, Y, Z), trong khối 460, bộ mã hóa tái cấu trúc ma trận quay được lượng tử hóa 4D từ 6 góc Euler được lượng tử hóa và sau đó nó được chuyển đổi thành hai bộ bốn đơn vị cho các mục đích nội suy. Theo một biến thể trong đó đầu vào với bộ mã hóa là tín hiệu ambisonic phẳng (3 kênh W, X, Y), trong khối 460, bộ mã hóa tái cấu trúc ma trận

quay 3D được lượng tử hóa từ 3 góc Euler được lượng tử hóa và sau đó nó được chuyển đổi thành bộ bốn đơn vị cho các mục đích nội suy. Theo một biến thể trong đó đầu vào của bộ mã hóa là tín hiệu âm thanh nổi, bộ mã hóa sử dụng, trong khối 460, sự biểu diễn của phép quay 2D được lượng tử hóa với góc quay.

Theo phương án có 4 kênh, đối với phép nội suy của ma trận quay giữa khung t và khung $t - 1$, ma trận quay được tính toán đối với khung t được phân tích thành hai bộ bốn (cặp bộ bốn) bằng cách thừa số hóa của Cayley và ta sử dụng cặp bộ bốn được lưu trữ đối với khung trước $t-1$ và được biểu thị $(Q_{L,t-1}, Q_{R,t-1})$.

Cho mỗi khung con, các bộ bốn được nội suy theo từng cặp trong mỗi khung con.

Đối với bộ bốn bên trái $(Q_{L,t})$, khối xác định đường ngắn nhất giữa hai $(Q_{L,t}$ hoặc $-Q_{L,t}$) khả thi. Phụ thuộc vào trường hợp, dấu của bộ bốn của khung hiện thời được đảo. Sau đó, phép nội suy được tính toán đối với bộ bốn bên trái sử dụng phép nội suy tuyến tính hình cầu (spherical linear interpolation, SLERP):

$$Q_{L,interp}(\alpha) = Q_{L,t-1} \frac{\sin(1-\alpha)\Omega_L}{\sin \Omega_L} + Q_{L,t} \frac{\sin \alpha\Omega_L}{\sin \Omega_L}$$

trong đó α tương ứng với hệ số nội suy ($\alpha=1/K, 2/K, \dots 1$), và $\Omega_L = \arccos(Q_{L,t-1} \cdot Q_{L,t})$

Đối với bộ bốn bên phải $(Q_{R,t})$, nếu có sự đảo ngược đối bộ bốn bên trái thì ta phải duy trì tính chẵn lẻ và buộc dấu của bộ bốn bên phải. Sự ràng buộc về dấu này sau đây được gọi là "giới hạn đường ngắn nhất chung". Sau đó, phép nội suy được tính toán tương tự với bộ bốn bên trái:

$$Q_{R,interp}(\alpha) = Q_{R,t-1} \frac{\sin(1-\alpha)\Omega_R}{\sin \Omega_R} + Q_{R,t} \frac{\sin \alpha\Omega_R}{\sin \Omega_R}$$

trong đó α tương ứng với hệ số nội suy ($\alpha = 1/K, 2/K, \dots 1$) và $\Omega_R = \arccos(Q_{R,t-1} \cdot Q_{R,t})$

Ngay khi phép nội suy đã được tính toán đối với hai bộ bốn, ma trận quay có kích thước là 4x4 được tính toán (tương ứng với 3x3 đối với ambisonic phẳng hoặc 2x2 đối với trường hợp âm thanh nổi).

Sự chuyển đổi này thành ma trận quay có thể được tiến hành theo mã giả sau đây:

trường hợp 4D: đối với cặp bộ bốn

- Như được mô tả ở trên, các ma trận bộ bốn và phản bộ bốn được tính toán và tích ma trận được tính toán.

trường hợp 3D: đối với bộ bốn $q = (w, x, y, z)$ ta thu được ma trận $M[i,j]$, $i,j=0\dots2$, có kích thước 3x3

$$xy=2*x*y$$

$$xz=2*x*z$$

$$yz=2*y*z$$

$$wx=2*w*x$$

$$wy=2*w*y$$

$$wz=2*w*z$$

$$xx=2*x*x$$

$$yy=2*y*y$$

$$zz=2*z*z$$

$$M[0][0] = 1-(yy+zz)$$

$$M[0][1] = (xy-wz)$$

$$M[0][2] = (xz+wy)$$

$$M[1][0] = (xy+wz)$$

$$M[1][1] = 1-(xx+zz)$$

$$M[1][2] = (yz-wx)$$

$$M[2][0] = (xz-wy)$$

$$M[2][1] = (yz+wx)$$

$$M[2][2] = 1-(xx+yy);$$

Cuối cùng, các ma trận $V_t^{interp}(\alpha)$ (hoặc các chuyển vị của chúng) được tính toán trên mỗi khung con trong khối nội suy 460 sau đó được sử dụng trong khối biến đổi 470 tạo ra n kênh được biến đổi bằng cách áp dụng các ma trận quay do đó được thấy cho các kênh ambisonic được xử lý trước bởi khối 300.

Sau đây, ta trở lại số lượng K của các khung con cần được xác định trong khối 450 đối với trường hợp trong đó số lượng này là tương thích. Chênh lệch cuối cùng giữa khung hiện thời và khung trước được đo, hoặc được xác định trực tiếp từ chênh lệch góc của các tham số mô tả ma trận quay. Trong trường hợp sau, ta muốn đảm bảo rằng sự biến đổi góc giữa các khung con liên tiếp là không nhận biết được. Việc triển khai của số lượng các khung con tương thích đặc biệt có lợi để làm giảm độ phức tạp trung bình của bộ mã hóa-giải mã, nhưng nếu làm giảm độ phức tạp được chọn, có thể tốt hơn là sử dụng phép nội suy với số lượng các khung con cố định.

Chênh lệch cuối cùng giữa ma trận quay được hiệu chỉnh của khung t và ma trận quay của khung $t - 1$ cho phép đo của độ lớn của chênh lệch trong ma trận kênh giữa hai khung. Chênh lệch này càng lớn, số lượng các khung con đối với phép nội suy được hoàn thành trong khối 460 càng lớn. Để đo chênh lệch này, ta sử dụng tổng của giá trị tuyệt đối của ma trận tương quan liên kết giữa ma trận biến đổi của khung hiện thời và khung trước, như sau:

$$\delta_t = \|I_n - \text{corr}(V_t, V_{t-1})\|$$

trong đó I_n là ma trận đơn vị, V_t các vector đặc trưng của khung của chỉ số t , và $\|M\|$ là chuẩn của ma trận M mà ở đây tương ứng với tổng của các giá trị tuyệt đối của tất cả các hệ số. Các chuẩn ma trận khác có thể được sử dụng (ví dụ chuẩn Frobenius).

Nếu hai ma trận giống nhau thì hiệu số bằng 0. Các ma trận càng khác nhau, giá trị của chênh lệch δ_t càng lớn. Các ngưỡng định trước có thể được áp dụng cho

δ_t , mỗi ngưỡng được liên kết với số lượng các phép nội suy định trước, ví dụ theo logic quyết định sau đây:

Các ngưỡng: $\{4,0, 5,0, 6,0, 7,0\}$

Số lượng K của các khung con đối với phép nội suy: $\{10, 48, 96, 192\}$

Do đó, chỉ hai bit có thể đủ để mã hóa bốn giá trị khả thi đưa ra số lượng các phân vùng nhỏ (các khung con).

Số lượng K của các phép nội suy được xác định bởi khối 450 sau đó được gửi tới môđun nội suy 460, và trong trường hợp tương thích số lượng các khung con được mã hóa ở dạng chỉ số nhị phân mà được gửi tới bộ ghép kênh (khối 350).

Việc khai triển phép nội suy cho phép áp dụng cuối cùng sự tối ưu hóa của sự giải tương quan của các kênh đầu vào trước khi đa-đơn mã hóa. Thật sự, các ma trận quay được tính toán lần lượt cho khung trước $t-1$ và khung hiện thời t có thể rất khác nhau do sự tìm kiếm này để giải tương quan, nhưng thậm chí như vậy, phép nội suy có thể làm phẳng sự chênh lệch này.

Phép nội suy được sử dụng chỉ yêu cầu chi phí tính toán được giới hạn đối với bộ mã hóa và bộ giải mã vì nó được thực hiện trong miền xác định (góc trong 2D, bộ bốn trong 3D, cặp bộ bốn trong 4D). Cách thức này có lợi hơn so với nội suy các ma trận hiệp biến được tính toán để phân tích PCA/KLT và lặp lại loại EVD của việc phân tích giá trị đặc trưng nhiều lần trên mỗi khung.

Khối 470 sau đó thực hiện ma trận hóa các kênh ambisonic trên mỗi khung con, sử dụng các ma trận biến đổi được tính toán trong khối 460. Các lượng ma trận này để tính toán $V_t^{interp}(\alpha)^T X(\alpha)$ trên mỗi khung con, trong đó $X(\alpha)$ tương ứng với các khối con có kích thước $n \times (L/K)$ đối với $\alpha=1/K, 2/K, \dots 1$. Tín hiệu được chứa trong các kênh này sau đó được gửi tới khối 340 để đa-đơn mã hóa.

Giờ tham chiếu được tạo ra với FIG.5, để mô tả bộ giải mã theo phương án làm ví dụ của sáng chế.

Sau khi phân kênh dòng bit đối với khung hiện thời t bởi khối 500, thông tin cấp phát được giải mã (khối 510) để có thể phân kênh và giải mã (khối 520) (các dòng bit) nhận được cho mỗi trong số n kênh được biến đổi.

Khối 520 gọi nhiều trường hợp giải mã lỗi, được thực thi riêng biệt. Việc giải mã lỗi có thể là loại EVS, được sửa đổi tùy chọn để cải thiện hiệu năng của nó. Sử dụng cách thức đa-đơn, mỗi kênh được giải mã riêng biệt. Nếu việc mã hóa được sử dụng trước đó là mã hóa âm thanh nổi hoặc đa kênh, cách thức đa-đơn có thể được thay thế bằng đa âm thanh nổi hoặc đa kênh để giải mã. Các kênh do đó được giải mã được gửi tới khối 530 mà giải mã ma trận quay đối với khung hiện thời và tùy chọn số lượng K của các khung con cần được sử dụng đối với phép nội suy (nếu phép nội suy là tương thích). Đối với mỗi ma trận, khối nội suy 460 chia khung thành các khung con, trong đó số lượng K có thể được đọc trong dòng được mã hóa bởi khối 610 (FIG.6) và nội suy các ma trận quay, mục đích là tìm kiếm – khi không có lỗi truyền – cùng một ma trận như trong khối 460 của bộ mã hóa để có thể làm đảo ngược phép biến đổi được thực hiện trước đó trong khối 470.

Khối 530 thực hiện ma trận hóa để đảo ngược khối 470 để tái cấu trúc tín hiệu được giải mã, như được thể hiện chi tiết sau đây với tham chiếu đến FIG.6. Ma trận hóa này để tính $V_t^{interp}(\alpha)\hat{X}(\alpha)$ trên mỗi khung con, trong đó $\hat{X}(\alpha)$ tương ứng với các khối con liên tiếp có kích thước $n \times (L/K)$ đối với $\alpha=1/K, 2/K, \dots 1$.

Khối 530 nói chung thực hiện giải mã và tổng hợp PCA/KLT đảo ngược với những gì đã được thực hiện bởi khối 310 của FIG.3. Các chỉ số lượng tử hóa của các tham số lượng tử hóa quay trong khung hiện thời được giải mã trong khối 600. Số lượng tử hóa vô hướng có thể được sử dụng và kích thước bước lượng tử hóa là như nhau cho mỗi góc. Trong trường hợp thích hợp, số lượng các khung con nội suy được giải mã (khối 610) để tìm kiếm số lượng K của các khung con trong số tập hợp $\{10, 48, 96, 192\}$; theo một số biến thể trong đó độ dài khung L là khác, tập hợp của các giá trị này có thể thích hợp. Phép nội suy của bộ giải mã giống với phép nội suy được thực hiện trong bộ mã hóa (khối 460).

Khối 620 thực hiện ma trận hóa đảo ngược của các kênh ambisonic trên mỗi khung con, sử dụng các phép đảo ngược (thực tế là các chuyển vị) của các ma trận biến đổi được tính toán trong khối 460.

Do đó, sáng chế sử dụng toàn bộ phương thức khác so với bộ mã hóa-giải mã MPEG-H với phần xếp chồng-bổ sung dựa trên sự biểu diễn cụ thể của các ma trận

biến đổi được giới hạn với các ma trận quay từ một khung đến khung khác, trong miền thời gian, cụ thể cho phép phép nội suy của các ma trận biến đổi, với ánh xạ đảm bảo tính nhất quán theo hướng (bao gồm việc tính đến hướng do dấu).

Cách thức thông thường của sáng chế là mã hóa âm thanh ambisonic trong miền thời gian bởi PCA, cụ thể với các ma trận biến đổi PCA buộc phải là các ma trận quay và được nội suy bởi các khung con theo cách thức tối ưu hóa (cụ thể trong miền của các bộ bốn/cặp bộ bốn) để cải thiện chất lượng. Kích thước bước nội suy được cố định và tương thích phụ thuộc vào tiêu chuẩn khác biệt giữa ma trận tương quan liên kết và ma trận tham chiếu (đơn vị) hoặc giữa các ma trận cần được nội suy. Sự lượng tử hóa của các ma trận quay có thể được thực thi trong miền của các góc Euler suy rộng. Tuy nhiên, tốt hơn là có thể được chọn để định lượng các ma trận có các kích thước 3 và 4 trong miền của các bộ bốn và cặp bộ bốn (lần lượt), khiến nó có thể duy trì trong cùng một miền đối với sự lượng tử hóa và phép nội suy.

Ngoài ra, sự căn chỉnh của các vector đặc trưng được sử dụng để tránh các vấn đề về các lần nhấp và sự đảo kênh từ một khung đến khung khác.

Tất nhiên, sáng chế không bị giới hạn với các phương án được mô tả ở trên theo các ví dụ, và mở rộng đến các biến khác.

Do đó, phần mô tả ở trên được đề cập đến các trường hợp có bốn kênh.

Tuy nhiên, theo một số biến thể, cũng có thể mã hóa số lượng các kênh lớn hơn bốn. Việc khai triển vẫn giống (về mặt các khối chức năng) với trường hợp có $n = 4$, nhưng phép nội suy bởi cặp bộ bốn được thay thế bởi phương pháp thông thường sau đây.

Các ma trận biến đổi tại các khung $t - 1$ và t được biểu thị V_{t-1} và V_t . Phép nội suy có thể được thực hiện với hệ số α giữa V_{t-1} và V_t sao cho:

$$V_t^{interp}(\alpha) = V_{t-1}(V_{t-1}^T V_t)^\alpha$$

Số hạng $(V_{t-1}^T V_t)^\alpha$ có thể được tính toán trực tiếp bằng cách phân tích giá trị đặc trưng của $V_{t-1}^T V_t$. Thật sự, nếu $V_{t-1}^T V_t = QLQ^T$, ta có: $(V_{t-1}^T V_t)^\alpha = QL^\alpha Q^T$.

Cũng lưu ý rằng biến thể này cũng có thể thay thế phép nội suy bởi cặp bộ bốn đơn vị (trường hợp 4D), bộ bốn đơn vị (trường hợp 3D), hoặc góc, tuy nhiên điều này ít lợi hơn vì nó yêu cầu bước chéo hóa bổ sung và việc tính toán lũy thừa, trong khi phương án được mô tả ở trên hiệu quả hơn đối với các trường hợp này là 2, 3, hoặc 4 kênh.

YÊU CẦU BẢO HỘ

1. Phương pháp mã hóa đối với việc nén các tín hiệu audio tạo ra, theo thời gian, sự liên tiếp của các khung mẫu $(t-1, t)$, ở mỗi trong số N kênh trong biểu diễn ambisonic theo bậc lớn hơn 0, phương pháp này được thực thi bởi thiết bị mã hóa và bao gồm các bước:

- tạo ra, dựa trên các kênh và đối với khung hiện thời (t) , ma trận hiệp biến liên kênh, và tìm kiếm các vector đặc trưng của ma trận hiệp biến nêu trên với việc xem xét để thu được ma trận của các vector đặc trưng,

- kiểm tra ma trận của các vector đặc trưng để xác minh rằng nó biểu diễn phép quay trong không gian N chiều, và nếu không, thì hiệu chỉnh ma trận của các vector đặc trưng cho đến khi ma trận quay thu được, đối với khung hiện thời (t) , và

- áp dụng ma trận quay nêu trên cho các tín hiệu của N kênh trước khi mã hóa các tín hiệu nêu trên.

2. Phương pháp theo điểm 1, phương pháp này còn bao gồm bước lượng tử hóa ma trận quay (440), ma trận quay nêu trên được áp dụng cho các tín hiệu của N kênh trong biểu diễn được lượng tử hóa.

3. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, phương pháp này còn bao gồm bước lượng tử hóa ma trận quay (440) và bước nội suy của ma trận lượng tử hóa (460), ma trận quay nêu trên được áp dụng cho các tín hiệu của N kênh trong phép biểu diễn được lượng tử hóa và được nội suy.

4. Phương pháp theo điểm 3, trong đó phép nội suy được tiến hành trong miền của các bộ bốn.

5. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, phương pháp này còn bao gồm các bước:

- so sánh ma trận của các vector đặc trưng thu được đối với khung hiện thời (t) , với ma trận quay thu được đối với khung $(t-1)$ ở trước khung hiện thời (t) , và

- hoán vị các cột của ma trận của các vector đặc trưng của khung hiện thời (t) để đảm bảo tính thống nhất với ma trận quay của khung trước $(t-1)$.

6. Phương pháp theo điểm 5, trong đó sự hoán vị nêu trên của các cột khiến nó có thể đảm bảo tính nhất quán của các trục của các vector, và phương pháp này còn bao gồm các bước:

- xác minh, cho mỗi vector đặc trưng của khung hiện thời (t), tính nhất quán theo hướng với vector cột của vị trí tương ứng trong ma trận quay của khung trước ($t-1$), và

- trong trường hợp không có tính nhất quán, thì đảo dấu của các phần tử của vector đặc trưng này trong ma trận của các vector đặc trưng của khung hiện thời (t).

7. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, phương pháp này còn bao gồm các bước:

- ước tính chênh lệch giữa ma trận quay thu được đối với khung hiện thời (t) và ma trận quay thu được đối với khung ($t-1$) ở trước khung hiện thời,

- dựa trên chênh lệch được ước tính, xác định liệu ít nhất một phép nội suy có cần được thực hiện giữa ma trận quay của khung hiện thời (t) và ma trận quay của khung trước ($t-1$) hay không.

8. Phương pháp theo điểm 7, trong đó:

- dựa trên chênh lệch được ước tính, số lượng các phép nội suy cần được thực hiện giữa ma trận quay của khung hiện thời (t) và ma trận quay của khung trước ($t-1$) được xác định,

- khung hiện thời được chia thành số lượng các khung con tương ứng với số lượng các phép nội suy cần được thực hiện, và

- ít nhất số lượng các phép nội suy này được mã hóa với việc xem xét tới việc truyền qua mạng.

9. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó, với sự hoán vị giữa các cột của ma trận của các vector đặc trưng để đảo dấu của định thức của ma trận của các vector đặc trưng và định thức của ma trận quay bằng 1,

nếu định thức của ma trận của các vector đặc trưng bằng -1, các dấu của các phần tử của cột được chọn của ma trận của các vector đặc trưng được đảo để định thức bằng 1 và do đó ma trận quay được tạo ra.

10. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó biểu diễn ambisonic là bậc thứ nhất và số lượng N kênh là bốn, và trong đó ma trận quay của khung hiện thời được biểu diễn bằng hai bộ bốn.

11. Phương pháp theo điểm 10, kết hợp với điểm 9, trong đó mỗi phép nội suy cho khung con hiện thời là phép nội suy tuyến tính hình cầu (spherical linear interpolation, SLERP), được thực hiện là hàm của phép nội suy của khung con ở trước khung con hiện thời và dựa trên các bộ bốn của khung con trước.

12. Phương pháp theo điểm 11, trong đó phép nội suy tuyến tính hình cầu của khung con hiện thời được thực hiện để thu được các bộ bốn của khung con hiện thời, như sau:

$$Q_{L,interp}(\alpha) = Q_{L,t-1} \frac{\sin(1-\alpha)\Omega_L}{\sin \Omega_L} + Q_{L,t} \frac{\sin \alpha \Omega_L}{\sin \Omega_L}$$

$$Q_{R,interp}(\alpha) = Q_{R,t-1} \frac{\sin(1-\alpha)\Omega_R}{\sin \Omega_R} + Q_{R,t} \frac{\sin \alpha \Omega_R}{\sin \Omega_R}$$

trong đó:

$Q_{L,t-1}$ là một trong số các bộ bốn của khung con trước t-1,

$Q_{R,t-1}$ là bộ bốn khác của khung con trước t-1,

$Q_{L,t}$ là một trong số các bộ bốn của khung con hiện thời t,

$Q_{R,t}$ là bộ bốn khác của khung con hiện thời t,

$$\Omega_L = \text{Arccos} (Q_{L,t-1} \cdot Q_{L,t}); \Omega_R = \text{Arccos} (Q_{R,t-1} \cdot Q_{R,t})$$

và α tương ứng với hệ số nội suy.

13. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó bước tìm kiếm các vector đặc trưng được tiến hành bằng việc phân tích thành phần chính

(principal component analysis, PCA) hoặc bằng biến đổi Karhunen-Loève (Karhunen-Loève transform, KLT), trong miền thời gian.

14. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó bước dự báo trước ngân sách cấp phát bit cho mỗi kênh ambisonic được thực thi và bao gồm:

- cho mỗi kênh ambisonic, ước tính năng lượng âm thanh hiện tại trong kênh,
- chọn, trong bộ nhớ, điểm số chất lượng định trước (MOS), dựa trên kênh ambisonic này và đối với tốc độ bit hiện thời trong mạng,
- ước tính trọng số cần được áp dụng cho cấp phát bit cho kênh này, bằng cách nhân điểm số được chọn với năng lượng được ước tính.

15. Phương pháp giải mã các tín hiệu audio tạo ra, theo thời gian, sự liên tiếp của các khung mẫu $(t-1, t)$, ở mỗi trong số N kênh trong biểu diễn ambisonic của bậc cao hơn 0, phương pháp này được thực thi bởi thiết bị giải mã và bao gồm các bước:

- nhận, đối với khung hiện thời (t) , ngoài các tín hiệu của N kênh của khung hiện thời này, các tham số của ma trận quay,
- cấu trúc ma trận quay ngược từ các tham số nêu trên,
- áp dụng ma trận quay ngược nêu trên cho các tín hiệu từ N kênh nhận được, trước khi giải mã các tín hiệu nêu trên.

16. Thiết bị mã hóa bao gồm mạch xử lý để thực thi phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 14.

17. Thiết bị giải mã bao gồm mạch xử lý để thực thi phương pháp theo điểm 15.

18. Phương tiện đọc được bằng máy tính lưu trữ chương trình máy tính bao gồm các lệnh để thực thi phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 15, khi các lệnh nêu trên được thực thi bởi bộ xử lý của mạch xử lý.

1/6

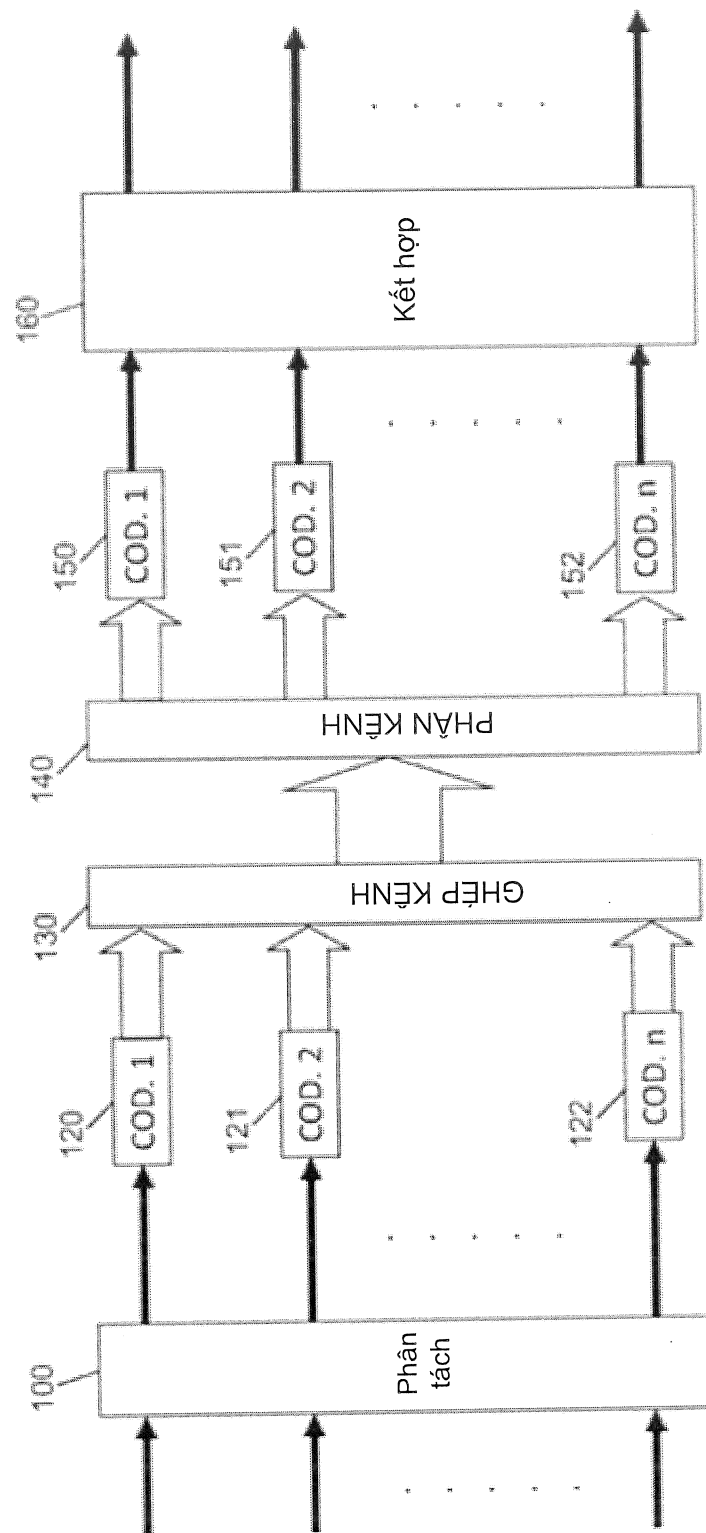
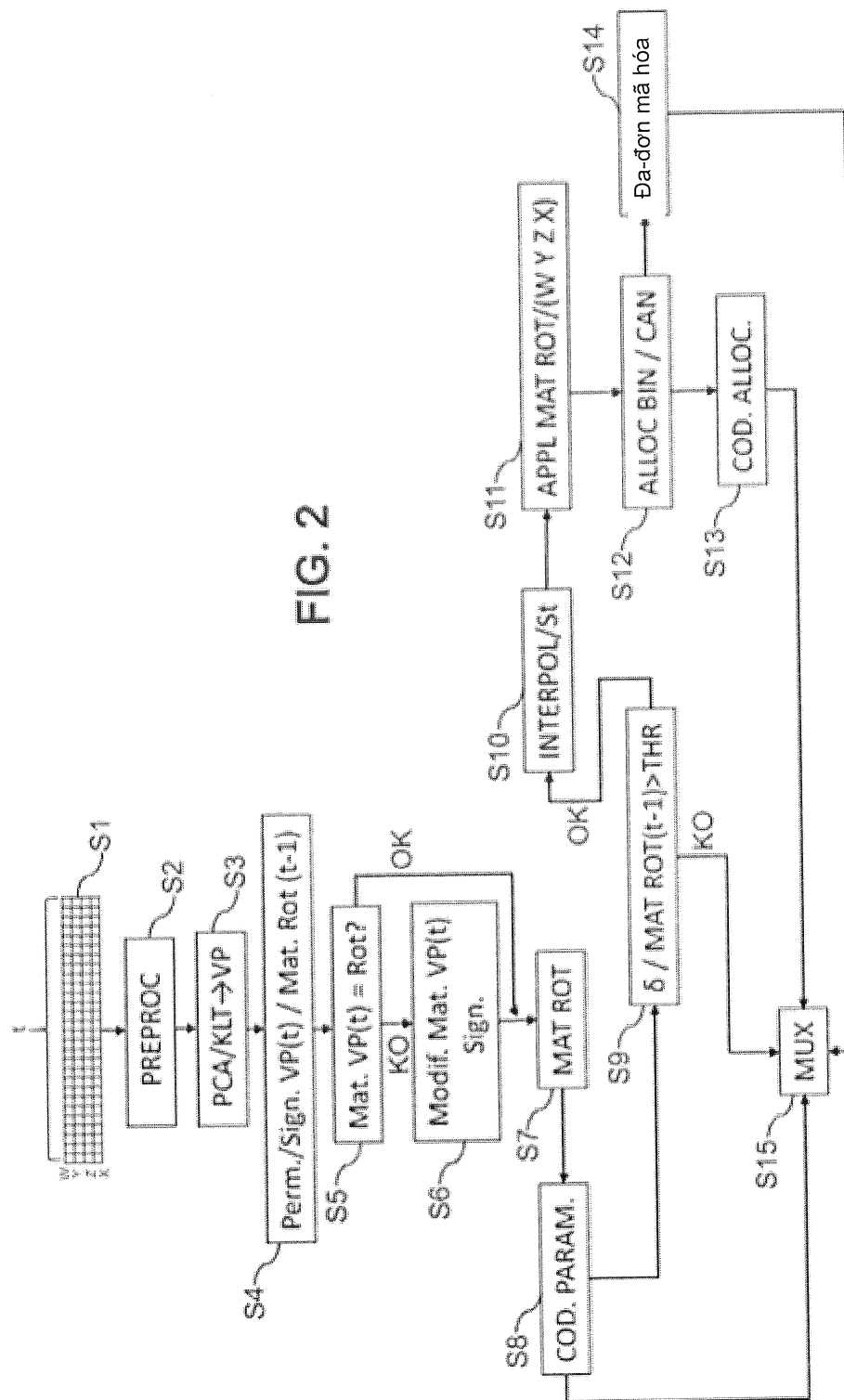


FIG. 1

FIG. 2



3/6

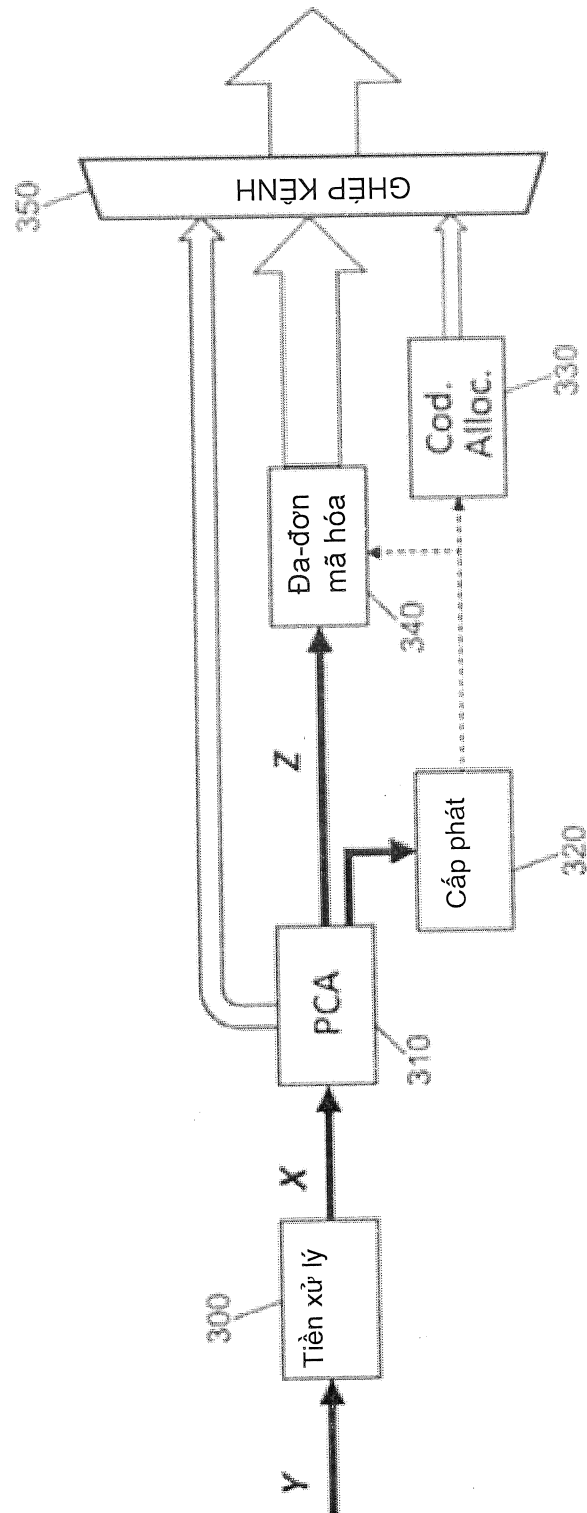


FIG. 3

4/6

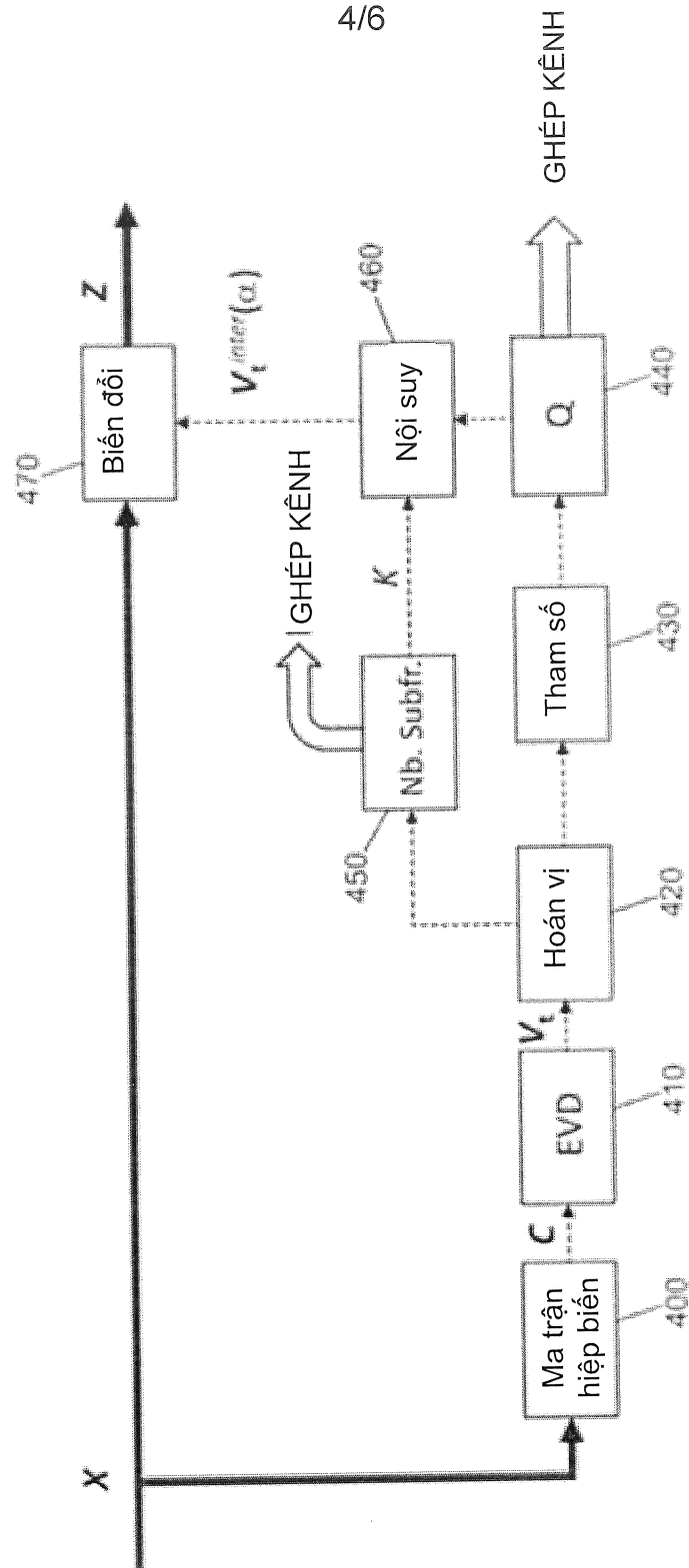
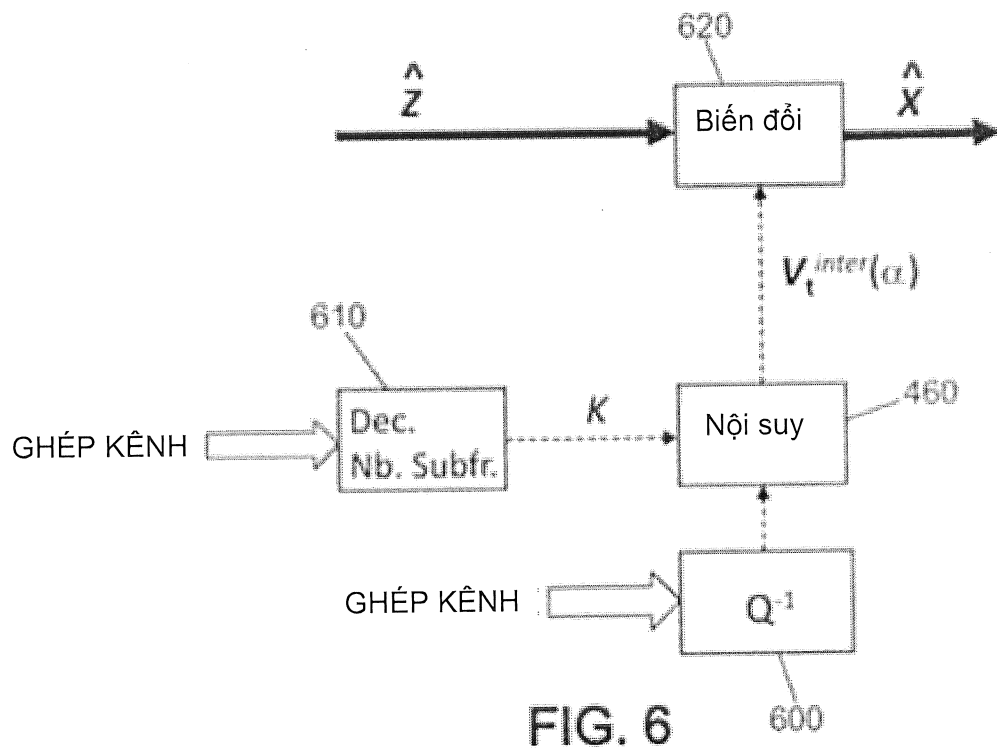
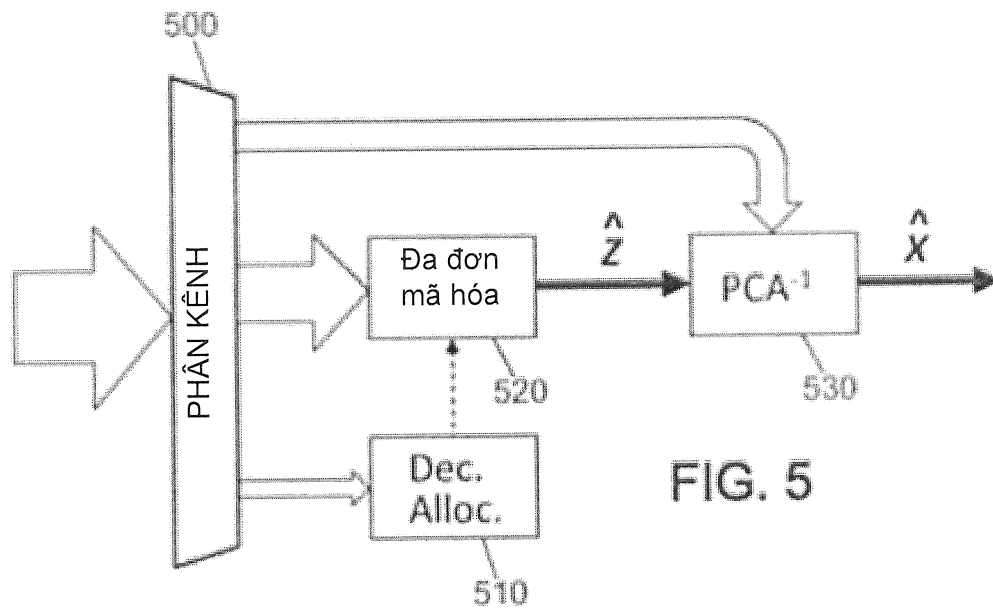


FIG. 4



6/6

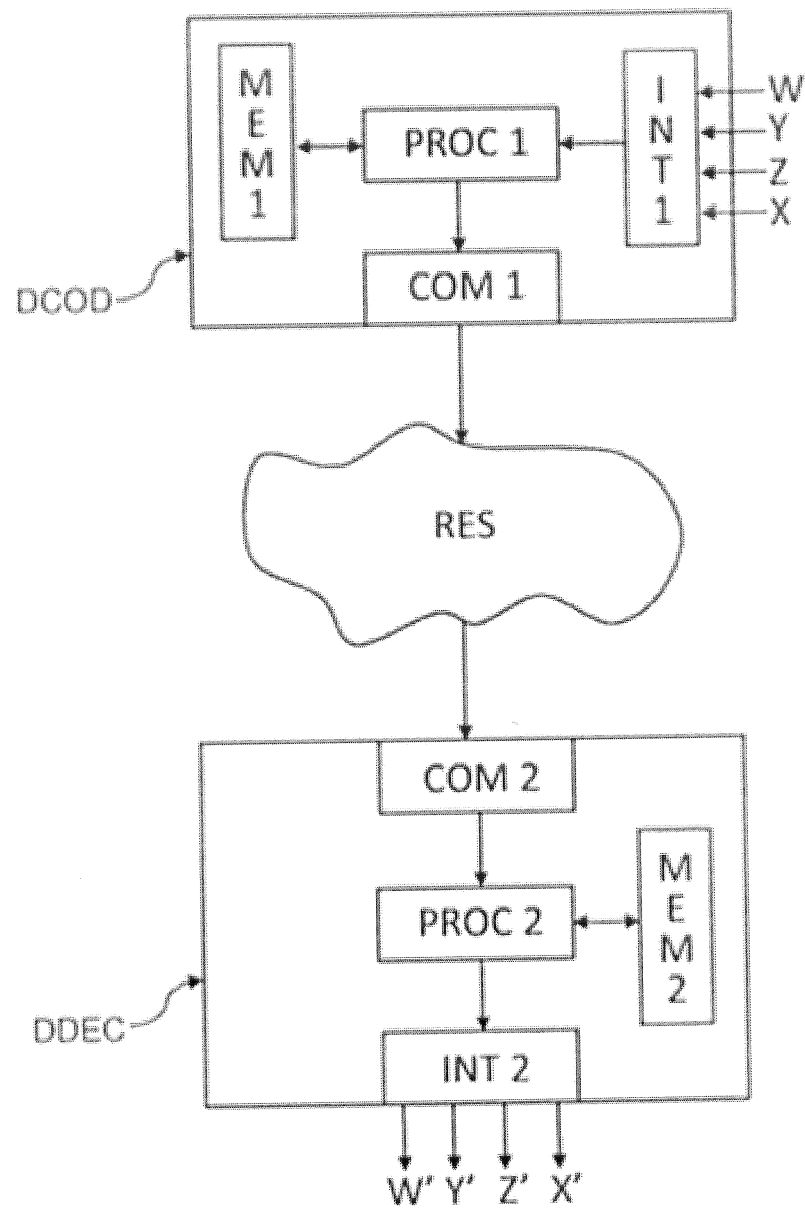


FIG. 7