



- (12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) 
CỤC SỞ HỮU TRÍ TUỆ 1-0045095
(51)^{2020.01} H04N 21/81; H04N 21/2343; H04N (13) B
21/6587; H04N 21/218; H04N 21/439
-

- (21) 1-2020-02532 (22) 11/10/2018
(86) PCT/EP2018/077770 11/10/2018 (87) WO 2019/072984 18/04/2019
(30) 17196259.0 12/10/2017 EP
(45) 25/04/2025 445 (43) 27/07/2020 388A
(71) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) MURTAZA, Adrian (RO); FUCHS, Harald (DE); CZELHAN, Bernd (DE);
PLOGSTIES, Jan (DE); AGNELLI, Matteo (IT); HOFMANN, Ingo (DE).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)
-

- (54) HỆ THỐNG VÀ PHƯƠNG PHÁP PHÂN PHỐI CÁC LUỒNG ÂM THANH VÀ
VIDEO ĐẾN MÁY KHÁCH CHO CÁC ỨNG DỤNG THỰC TẾ ẢO

(21) 1-2020-02532

(57) Sáng chế đề cập đến các kỹ thuật, hệ thống, phương pháp và các lệnh cho các môi trường thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc video 360 độ. Trong một ví dụ, hệ thống (102) bao gồm ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng. Hệ thống bao gồm ít nhất một bộ giải mã âm thanh (104) được tạo cấu hình để giải mã các tín hiệu âm thanh (108) từ ít nhất một luồng âm thanh (106). Hệ thống (102) được tạo cấu hình để yêu cầu (112) ít nhất một luồng âm thanh (106) và/hoặc một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh đến máy chủ (120) trên cơ sở ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng (110).

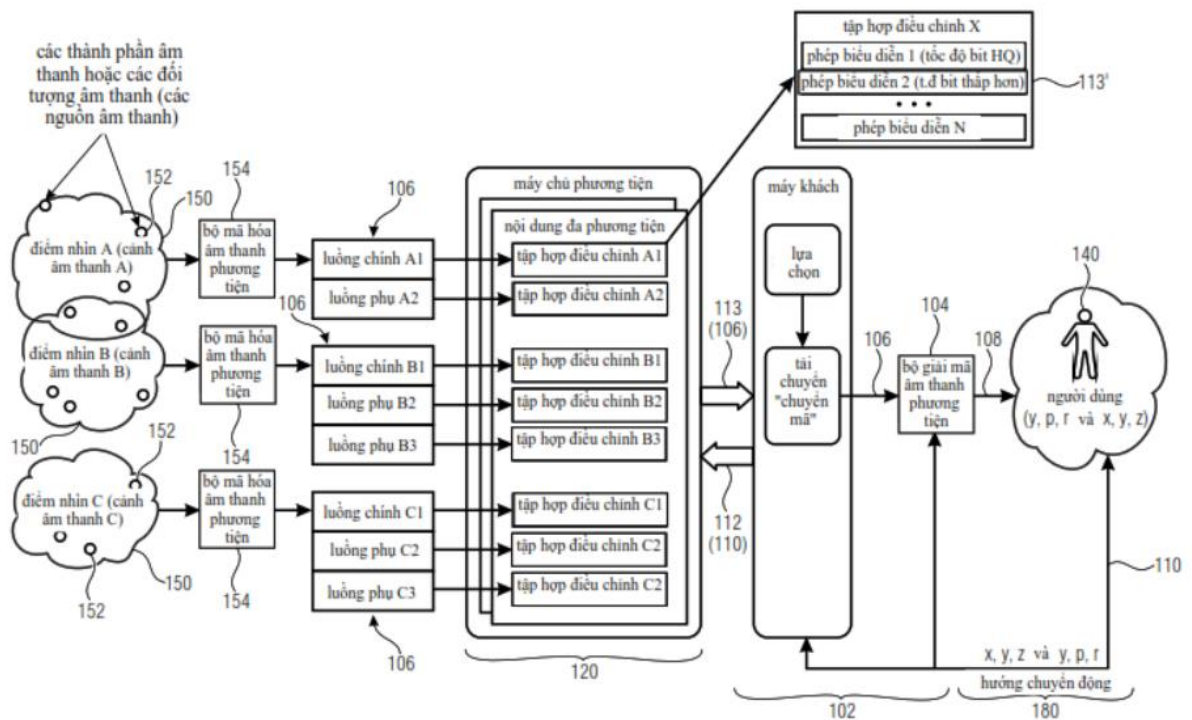


Fig. 1.1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực môi trường thực tế ảo (Virtual Reality - VR) hoặc tương tự trong môi trường thực tế tăng cường (Augmented Reality - AR) hoặc môi trường thực tế hỗn hợp (Mixed Reality - MR) hoặc môi trường video 360 độ, mà người dùng thường có thể hình dung toàn bộ nội dung 360 độ bằng cách sử dụng ví dụ như màn hình gắn trên đầu (Head Mounted Display - HMD) và nghe nó qua tai nghe (hoặc tương tự qua loa phóng thanh, chứa sự kết xuất chính xác tùy thuộc vào vị trí của nó).

Tình trạng kỹ thuật của sáng chế

Trong trường hợp sử dụng đơn giản, nội dung được tạo ra theo cách chỉ có một cảnh âm thanh/video (ví dụ, video 360 độ chẳng hạn) được đưa ra tại một thời điểm nhất định. Cảnh âm thanh/video có vị trí cố định (ví dụ hình cầu nếu người dùng đặt ở giữa), và người dùng không thể di chuyển trong cảnh, nhưng chỉ có thể xoay đầu theo nhiều hướng khác nhau (xoay, tiến-lùi, nghiêng). Trong trường hợp này, video và âm thanh khác nhau được phát lại (các công nhìn khác nhau được hiển thị) cho người dùng dựa trên hướng của đầu người đó.

Trong khi đối với video, nội dung video được phân phối cho toàn bộ cảnh 360 độ, cùng với siêu dữ liệu để mô tả quá trình kết xuất (ví dụ, ghép thông tin, ánh xạ chiếu, v.v.) và được chọn dựa trên chế độ xem của người dùng hiện tại, cho âm thanh nội dung là giống nhau cho toàn bộ cảnh. Dựa trên siêu dữ liệu, nội dung âm thanh được điều chỉnh phù hợp với chế độ xem của người dùng hiện tại (ví dụ, đối tượng âm thanh được kết xuất khác nhau dựa trên thông tin hướng của người dùng/công nhìn). Cần lưu ý rằng nội dung 360 độ đề cập đến bất kỳ loại nội dung nào có nhiều hơn một góc nhìn cùng một lúc, mà người dùng có thể chọn (ví dụ theo hướng đầu của người đó hoặc sử dụng thiết bị điều khiển từ xa).

Trong bối cảnh phức tạp hơn, khi người dùng có thể di chuyển trong cảnh VR hoặc "nhảy" từ cảnh này sang cảnh khác, nội dung âm thanh cũng có thể thay đổi (ví

dụ, các nguồn âm thanh không nghe thấy được trong một cảnh có thể nghe được ở cảnh tiếp theo - "cánh cửa được mở ra"). Với các hệ thống hiện có, các cảnh âm thanh hoàn chỉnh có thể được mã hóa thành một luồng và, nếu cần, thành các luồng bổ sung (phụ thuộc vào luồng chính). Các hệ thống như vậy được gọi là các hệ thống âm thanh thể hệ tiếp theo (ví dụ, âm thanh 3D MPEG-H). Các ví dụ về các trường hợp sử dụng như vậy có thể là:

- Ví dụ 1: Người dùng chọn để vào một phòng mới và toàn bộ cảnh âm thanh/video thay đổi
- Ví dụ 2: Người dùng di chuyển trong cảnh VR, mở cửa và đi qua, cần sự chuyển đổi âm thanh từ cảnh này sang cảnh tiếp theo

Với mục đích mô tả bối cảnh này, khái niệm điểm nhìn rời rạc trong không gian được đưa ra, dưới dạng vị trí riêng biệt trong không gian (hoặc trong môi trường VR), mà có sẵn nội dung âm thanh/video khác nhau.

Giải pháp "thăng về phía trước" là có bộ mã hóa thời gian thực thay đổi sự mã hóa (số thành phần âm thanh, thông tin không gian, v.v.) dựa trên phản hồi từ thiết bị phát lại về vị trí/hướng của người dùng. Giải pháp này có nghĩa là, ví dụ trong môi trường truyền phát, truyền thông rất phức tạp giữa máy khách và máy chủ:

- Máy khách (thường được cho là chỉ sử dụng logic đơn giản) sẽ yêu cầu các cơ chế nâng cao để truyền tải không chỉ các yêu cầu cho các luồng khác nhau, mà còn truyền tải thông tin phức tạp về các chi tiết mã hóa cho phép xử lý nội dung phù hợp dựa trên vị trí của người dùng.
- Máy chủ phương tiện thường có trước với các luồng khác nhau (được định dạng theo một cách cụ thể cho phép phân phối "theo phân đoạn") và chức năng chính của máy chủ là cung cấp thông tin về các luồng có sẵn và thực hiện phân phối khi được yêu cầu. Để kích hoạt các bối cảnh cho sự mã hóa dựa trên phản hồi từ thiết bị phát lại, máy chủ phương tiện sẽ yêu cầu các liên kết truyền thông nâng cao với nhiều bộ mã hóa phương tiện trực tiếp và khả năng tạo tất cả thông tin báo hiệu nhanh chóng (ví dụ, mô tả bản trình bày phương tiện) mà có thể thay đổi theo thời gian thực.

Mặc dù hệ thống như vậy có thể được tưởng tượng, nhưng các yêu cầu về tính toán và phức tạp của nó vượt ra ngoài chức năng và tính năng của thiết bị và hệ thống hiện có hoặc thậm chí sẽ được phát triển trong những thập kỷ tới.

Ngoài ra, nội dung biểu diễn môi trường VR hoàn chỉnh ("thế giới hoàn chỉnh") có thể được phân phối mọi lúc. Điều này sẽ giải quyết vấn đề, nhưng sẽ đòi hỏi một tốc độ bit khổng lồ vượt quá khả năng của các liên kết truyền thông có sẵn.

Điều này rất phức tạp đối với môi trường thời gian thực, và để cho phép các trường hợp sử dụng đó sử dụng các hệ thống có sẵn, các giải pháp thay thế được đề xuất cho phép chức năng này với độ phức tạp thấp.

Thuật ngữ và các định nghĩa

Thuật ngữ sau đây được sử dụng trong lĩnh vực kỹ thuật:

- Các thành phần âm thanh: các tín hiệu âm thanh có thể được biểu diễn ví dụ như các đối tượng âm thanh, kênh âm thanh, âm thanh dựa trên cảnh (Ambisonics bậc cao: Higher Order Ambisonics - HOA) hoặc bất kỳ sự kết hợp nào trong số chúng.

- Vùng quan tâm (Region-of-Interest - ROI): Một vùng của nội dung video (hoặc của môi trường được hiển thị hoặc mô phỏng) được người dùng quan tâm tại một thời điểm. Đây ví dụ có thể là vùng trên hình cầu, hoặc lựa chọn đa giác từ bản đồ 2D. ROI xác định một vùng cụ thể cho mục đích cụ thể, xác định các đường bao của đối tượng đang được xem xét.

- Thông tin vị trí người dùng: thông tin vị trí (ví dụ, các tọa độ x, y, z), thông tin hướng (xoay, tiến-lùi, nghiêng), hướng và tốc độ di chuyển, v.v..

- Cổng nhìn: Một phần của video hình cầu hiện đang được hiển thị và xem bởi người dùng.

- Điểm nhìn: điểm trung tâm của cổng nhìn.

- Video 360 độ (còn được gọi là video nhập vai hoặc video hình cầu): thể hiện trong bối cảnh của tài liệu này nội dung video chứa nhiều hơn một chế độ xem (tức là, cổng nhìn) theo một hướng cùng một lúc. Nội dung như vậy có thể được tạo ra, ví dụ, bằng cách sử dụng máy ảnh đa hướng hoặc nhiều máy ảnh. Trong khi phát lại, người xem có quyền kiểm soát hướng xem.

- Mô tả trình diễn phương tiện (Media Presentation Description - MPD) là cú pháp, ví dụ XML chứa thông tin về các phân đoạn của phương tiện truyền thông, các mối quan hệ và thông tin cần thiết của chúng để lựa chọn giữa chúng.

- Các tập hợp thích ứng chứa luồng phương tiện hoặc tập hợp các luồng phương tiện. Trong trường hợp đơn giản nhất, một tập hợp điều chỉnh chứa tất cả âm thanh và video cho nội dung, nhưng để giảm băng thông, mỗi luồng có thể được chia thành tập hợp điều chỉnh khác nhau. Trường hợp phổ biến là có một tập hợp điều chỉnh video, và nhiều tập hợp điều chỉnh âm thanh (một tập hợp cho mỗi ngôn ngữ được hỗ trợ). tập hợp điều chỉnh cũng có thể chứa phụ đề hoặc siêu dữ liệu tùy ý.

- Các sự biểu diễn cho phép tập hợp điều chỉnh chứa cùng một nội dung được mã hóa theo các cách khác nhau. Trong hầu hết các trường hợp, các sự biểu diễn sẽ được cung cấp ở nhiều tốc độ bit. Điều này cho phép khách hàng yêu cầu nội dung chất lượng cao nhất mà họ có thể phát mà không cần chờ vùng đệm. Các sự biểu diễn cũng có thể được mã hóa với các bộ mã hóa-giải mã khác nhau, cho phép hỗ trợ cho các máy khách với các bộ mã hóa-giải mã được hỗ trợ khác nhau.

Trong bối cảnh của ứng dụng này, các khái niệm về tập hợp điều chỉnh được sử dụng chung chung hơn, đôi khi thực sự đề cập đến các sự biểu diễn. Ngoài ra, các luồng phương tiện (các luồng âm thanh/video) thường được gói gọn trước tiên vào các phân đoạn của phương tiện là các tệp phương tiện thực tế do máy khách phát (ví dụ máy khách DASH). Các định dạng khác nhau có thể được sử dụng cho các phân đoạn của phương tiện, chẳng hạn như định dạng tệp phương tiện cơ sở ISO (ISO Base Media File Format - ISOBMFF) tương tự như định dạng bộ chứa MPEG-4 hoặc luồng truyền tải (Transport Stream - TS) MPEG-2. Việc gói gọn vào các phân đoạn của phương tiện và trong các sự biểu diễn/tập hợp điều chỉnh khác nhau độc lập với các phương pháp được mô tả ở đây, các phương pháp áp dụng cho tất cả các tùy chọn khác nhau.

Ngoài ra, mô tả về các phương pháp trong tài liệu này tập trung vào truyền thông chủ-khách DASH, nhưng các phương pháp này đủ chung để hoạt động với các môi trường phân phối khác, chẳng hạn như MMT, MPEG-2 TS, DASH-ROUTE, định dạng tệp cho tệp phát lại, v.v..

Nói chung, tập hợp điều chỉnh ở lớp cao hơn đối với luồng và có thể bao gồm siêu dữ liệu (ví dụ được liên kết với các vị trí). Luồng có thể bao gồm nhiều thành phần âm thanh. Cảnh âm thanh có thể liên quan đến nhiều luồng được phân phối như là một phần của nhiều tập hợp điều chỉnh.

Tài liệu tham khảo

[1] ISO/IEC 23008-3:2015, Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 3: 3D audio

[2] N16950, Study of ISO/IEC DIS 23000-20 Omnidirectional Media Format

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là khắc phục những hạn chế trên của các giải pháp hiện tại và tối ưu hóa việc phân phối âm thanh cho các ứng dụng thực tế ảo.

Theo phương án, hệ thống cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ có thể được tạo cấu hình để nhận các luồng video và âm thanh được sao chép trong thiết bị tiêu dùng phương tiện, trong đó hệ thống có thể bao gồm: ít nhất một bộ giải mã video đa phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video cho sự biểu diễn của các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng, và ít nhất một bộ giải mã âm thanh được tạo cấu hình để giải mã tín hiệu âm thanh từ ít nhất một luồng âm thanh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu ít nhất một luồng âm thanh và/hoặc một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh được thiết lập cho máy chủ trên cơ sở ít nhất là công nhìn hiện tại của người dùng và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để cung cấp cho máy chủ công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại để thu được ít nhất một luồng âm thanh và/hoặc một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh từ máy chủ.

Phương án có thể được tạo cấu hình sao cho ít nhất một cảnh được liên kết với ít nhất một thành phần âm thanh, mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan nơi thành phần âm thanh có thể nghe được, do đó các luồng âm thanh khác nhau được cung cấp cho các vị trí và/hoặc công nhìn khác nhau và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo của người dùng trong cảnh.

Theo một khía cạnh khác, hệ thống có thể được tạo cấu hình để quyết định xem có ít nhất một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh sẽ được đưa ra cho công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo của người dùng hiện tại trong cảnh hay không, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh ở vị trí ảo của người dùng hiện tại.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để quyết định theo cách dự đoán liệu ít nhất một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh sẽ trở nên phù hợp và/hoặc có thể nghe được dựa trên ít nhất là công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng hay không, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh và/hoặc luồng âm thanh và/hoặc tập hợp điều chỉnh ở vị trí ảo của người dùng cụ thể trước chuyển động và/hoặc tương tác của người dùng được dự đoán trong cảnh, trong đó hệ thống có thể được tạo cấu hình để tái tạo ít nhất trên thành phần âm thanh và/hoặc luồng âm thanh, khi nhận được, tại vị trí ảo của người dùng cụ thể sau chuyển động và/hoặc tương tác của người dùng trong cảnh.

Phương án của hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit và/hoặc mức chất lượng thấp hơn, tại vị trí ảo của người dùng trước sự di chuyển và/hoặc tương tác của người dùng trong cảnh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit và/hoặc mức chất lượng cao hơn, tại vị trí ảo của người dùng sau sự chuyển động và/hoặc tương tác của người dùng trong cảnh.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho ít nhất một thành phần âm thanh được liên kết với ít nhất một cảnh, mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan được liên kết với cảnh đó, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận luồng ở tốc độ bit và/hoặc chất lượng cao hơn cho các thành phần âm thanh gần với người dùng hơn so với các thành phần âm thanh ở xa người dùng hơn.

Theo một khía cạnh trong hệ thống, ít nhất một thành phần âm thanh có thể được liên kết với ít nhất một cảnh, thành phần âm thanh cuối cùng được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan liên quan đến cảnh đó, trong đó hệ thống có thể được tạo cấu hình để yêu cầu các luồng khác nhau ở các tốc độ bit và/hoặc mức chất lượng khác nhau cho các thành phần âm thanh dựa trên mức độ liên quan và/hoặc mức độ có thể kiểm tra của chúng tại mỗi vị trí ảo của người dùng trong cảnh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu luồng âm thanh ở tốc độ bit và/hoặc mức chất lượng cao hơn cho các thành phần âm thanh có liên quan hơn và/hoặc dễ nghe hơn ở vị trí ảo hiện tại của người dùng, và/hoặc luồng âm thanh ở tốc độ bit và/hoặc mức chất lượng thấp hơn cho các thành phần âm thanh ít liên quan hơn và/hoặc ít dễ nghe hơn ở vị trí ảo hiện tại của người dùng.

Theo phương án trong hệ thống, ít nhất một thành phần âm thanh có thể được liên kết với cảnh, mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan có liên quan tới cảnh đó, trong đó hệ thống có thể được tạo cấu hình để gửi định kỳ đến máy chủ cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng, để: đối với vị trí thứ nhất, luồng được cung cấp với tốc độ bit và/hoặc chất lượng cao hơn từ máy chủ, và đối với vị trí thứ hai, luồng được cung cấp ở tốc độ bit và/hoặc chất lượng thấp hơn từ máy chủ, trong đó vị trí thứ nhất gần với ít nhất một thành phần âm thanh hơn vị trí thứ hai.

Trong phương án của hệ thống, nhiều cảnh có thể được xác định cho nhiều môi trường trực quan như các môi trường liền kề và/hoặc lân cận, để các luồng thứ nhất được cung cấp liên quan đến cảnh thứ nhất, hiện tại, và trong trường hợp người dùng

chuyển sang cảnh thứ hai, cảnh xa hơn, để cung cấp cả các luồng liên quan đến cảnh thứ nhất và các luồng thứ hai liên quan đến cảnh thứ hai.

Trong phương án của hệ thống, nhiều cảnh có thể được xác định cho môi trường trực quan thứ nhất và thứ hai, môi trường thứ nhất và thứ hai là các môi trường liền kề và/hoặc lân cận, trong đó các luồng thứ nhất liên quan đến cảnh thứ nhất được cung cấp từ máy chủ cho việc tái tạo cảnh thứ nhất trong trường hợp vị trí của người dùng hoặc vị trí ảo ở môi trường thứ nhất liên quan đến cảnh thứ nhất, luồng thứ hai liên quan đến cảnh thứ hai được cung cấp từ máy chủ để tái tạo cảnh thứ hai trong trường hợp vị trí của người dùng hoặc vị trí ảo trong môi trường thứ hai liên quan đến cảnh thứ hai và cả luồng thứ nhất liên quan đến cảnh thứ nhất và luồng thứ hai liên quan đến cảnh thứ hai được cung cấp trong trường hợp vị trí hoặc vị trí ảo của người dùng ở vị trí chuyển tiếp giữa cảnh thứ nhất và cảnh thứ hai.

Trong phương án của hệ thống, nhiều cảnh có thể được xác định cho các môi trường trực quan thứ nhất và thứ hai là các môi trường liền kề và/hoặc lân cận, trong đó hệ thống được tạo cấu hình để yêu cầu và/hoặc nhận các luồng thứ nhất liên quan đến cảnh thứ nhất liên quan đến môi trường thứ nhất, để tái tạo cảnh thứ nhất trong trường hợp vị trí ảo của người dùng ở môi trường thứ nhất, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận luồng thứ hai liên quan đến cảnh thứ hai liên quan đến môi trường thứ hai, cho việc tái tạo cảnh thứ hai trong trường hợp vị trí ảo của người dùng ở môi trường thứ hai, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận cả luồng thứ nhất liên quan đến cảnh thứ nhất và luồng thứ hai liên quan đến cảnh thứ hai trong trường hợp vị trí ảo của người dùng ở vị trí chuyển tiếp giữa môi trường thứ nhất và môi trường thứ hai.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để các luồng thứ nhất liên quan đến cảnh thứ nhất thu được ở tốc độ bit và/hoặc chất lượng cao hơn khi người dùng ở môi trường thứ nhất liên quan đến cảnh thứ nhất, trong khi các luồng thứ hai liên quan đến cảnh thứ hai liên quan đến môi trường thứ hai thu được ở tốc độ bit và/hoặc chất lượng thấp hơn khi người dùng ở đầu vị trí chuyển tiếp từ cảnh thứ nhất sang cảnh thứ hai, và các luồng thứ nhất liên quan đến cảnh thứ nhất thu được tại tốc độ bit và/hoặc chất lượng thấp hơn và các luồng thứ hai liên quan đến cảnh thứ hai thu

được ở tốc độ bit và/hoặc chất lượng cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh thứ nhất sang cảnh thứ hai, trong đó tốc độ bit và/hoặc chất lượng thấp hơn là thấp hơn tốc độ bit và/hoặc chất lượng cao hơn.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho nhiều cảnh có thể được xác định cho nhiều môi trường chẳng hạn như các môi trường liền kề và/hoặc lân cận, để hệ thống có thể thu được các luồng liên quan đến cảnh hiện tại thứ nhất liên quan đến môi trường thứ nhất, hiện tại và, trong trường hợp khoảng cách của vị trí của người dùng hoặc vị trí ảo từ ranh giới của cảnh nằm dưới ngưỡng xác định trước, hệ thống có thể còn thu được các luồng âm thanh liên quan đến môi trường thứ hai, liền kề và/hoặc lân cận liên quan đến cảnh thứ hai.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho nhiều cảnh có thể được xác định cho nhiều môi trường trực quan, để hệ thống yêu cầu và/hoặc thu được các luồng liên quan đến cảnh hiện tại với tốc độ bit và/hoặc chất lượng cao hơn và các luồng liên quan đến cảnh thứ hai ở tốc độ bit và/hoặc chất lượng thấp hơn, trong đó tốc độ bit và/hoặc chất lượng thấp hơn là thấp hơn so với tốc độ bit và/hoặc chất lượng cao hơn.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để có thể xác định được N thành phần âm thanh, và trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của các thành phần âm thanh này lớn hơn ngưỡng định trước, N thành phần âm thanh được xử lý để thu được số lượng M của các thành phần âm thanh nhỏ hơn ($M < N$) liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của N thành phần âm thanh, để cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến N thành phần âm thanh, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của N thành phần âm thanh nhỏ hơn ngưỡng xác định trước, hoặc để cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến M thành phần âm thanh, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của N thành phần âm thanh lớn hơn ngưỡng xác định trước.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho ít nhất một cảnh môi trường trực quan liên quan đến ít nhất một số lượng lớn của N thành phần âm thanh ($N \geq 2$), mỗi thành phần âm thanh liên quan đến vị trí và/hoặc khu vực trong

môi trường trực quan, trong đó ít nhất một số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng cao, và trong đó ít nhất một số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng thấp, trong đó có ít nhất một sự biểu diễn thu được bằng cách xử lý N thành phần âm thanh để thu được số lượng nhỏ hơn của M thành phần âm thanh ($M < N$) liên quan đến vị trí hoặc khu vực gần vị trí hoặc khu vực của N thành phần âm thanh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc chất lượng cao hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh liên quan hơn và/hoặc dễ nghe hơn ở vị trí ảo của người dùng hiện tại trong cảnh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng thấp hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh ít liên quan hơn và/hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để trong trường hợp khoảng cách của người dùng và/hoặc mức độ liên quan và/hoặc mức độ dễ nghe và/hoặc hướng góc thấp hơn ngưỡng xác định trước, các luồng khác nhau sẽ thu được cho các thành phần âm thanh khác nhau.

Trong phương án, hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc thu được các luồng trên cơ sở hướng của người dùng và/hoặc hướng chuyển động của người dùng và/hoặc sự tương tác của người dùng trong cảnh.

Trong phương án của hệ thống, cổng nhìn có thể liên quan đến vị trí và/hoặc vị trí ảo và/hoặc dữ liệu chuyển động và/hoặc đầu.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho các thành phần âm thanh khác nhau được cung cấp ở các cổng nhìn khác nhau, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận, trong trường hợp một thành phần âm thanh thứ nhất nằm trong cổng nhìn, thành phần âm thanh thứ nhất tại tốc độ bit cao hơn so với thành phần âm thanh thứ hai không nằm trong cổng nhìn.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận luồng âm thanh thứ nhất và luồng âm thanh thứ hai, trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn và/hoặc dễ nghe hơn các

thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu và/hoặc nhận được ở tốc độ bit và/hoặc chất lượng cao hơn so với tốc độ bit và/hoặc chất lượng của các luồng âm thanh thứ hai.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình sao cho ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh thứ nhất liên quan đến môi trường trực quan thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh thứ hai liên quan đến môi trường trực quan thứ hai, trong đó hệ thống có thể được tạo cấu hình để thu được siêu dữ liệu mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với cảnh môi trường trực quan thứ hai, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất các thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở cảnh môi trường trực quan thứ nhất, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất các thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng nằm trong cảnh môi trường trực quan thứ hai, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất là các thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa cảnh môi trường trực quan thứ nhất và cảnh môi trường trực quan thứ hai.

Phương án của hệ thống có thể được tạo cấu hình để ít nhất một thành phần âm thanh thứ nhất được cung cấp trong ít nhất một luồng âm thanh và/hoặc tập hợp điều chỉnh, và ít nhất một thành phần âm thanh thứ hai được cung cấp trong ít nhất một luồng âm thanh thứ hai và/hoặc tập hợp điều chỉnh, và ít nhất một thành phần âm thanh thứ ba được cung cấp trong ít nhất một luồng âm thanh thứ ba và/hoặc tập hợp điều chỉnh, và trong đó ít nhất là cảnh môi trường trực quan thứ nhất được mô tả bằng siêu dữ liệu là cảnh hoàn chỉnh đòi hỏi ít nhất là các luồng âm thanh thứ nhất và thứ hai và/hoặc tập hợp điều chỉnh, và trong đó cảnh môi trường trực quan thứ hai được mô tả bởi siêu dữ liệu là cảnh không hoàn chỉnh mà yêu cầu ít nhất là luồng âm thanh thứ ba và/hoặc tập hợp điều chỉnh và ít nhất là luồng âm thanh thứ hai và/hoặc các tập hợp điều chỉnh liên quan đến ít nhất là cảnh môi trường trực quan thứ nhất, trong đó hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu, để cho phép gộp luồng âm thanh thứ hai thuộc về môi trường trực quan thứ nhất và luồng âm

thanh thứ ba liên quan đến môi trường trực quan thứ hai thành một luồng mới duy nhất, trong trường hợp vị trí ảo của người dùng nằm trong môi trường trực quan thứ hai.

Theo khía cạnh, hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên công nhìn hiện tại và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo của người dùng.

Theo khía cạnh, bộ xử lý siêu dữ liệu có thể được tạo cấu hình để kích hoạt và/hoặc vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên công nhìn và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng, trong đó bộ xử lý siêu dữ liệu có thể được tạo cấu hình để vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, trong trường hợp hệ thống quyết định rằng thành phần âm thanh đó sẽ không được tái tạo nữa do hệ quả của công nhìn hiện tại và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo, và trong đó bộ xử lý siêu dữ liệu có thể được tạo cấu hình để kích hoạt ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, trong trường hợp hệ thống quyết định phần tử âm thanh sẽ được tái tạo do hệ quả của công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để vô hiệu hóa việc giải mã các thành phần âm thanh được chọn trên cơ sở công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc vị trí ảo hiện tại của người dùng.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để gộp ít nhất một luồng âm thanh thứ nhất liên quan đến cảnh âm thanh hiện tại với ít nhất một luồng liên quan đến cảnh âm thanh lân cận, liền kề và/hoặc tương lai.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để thu và/hoặc thu thập dữ liệu thống kê hoặc tổng hợp trên công nhìn và/hoặc hướng đầu và/hoặc dữ liệu

chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại của người dùng, để truyền yêu cầu đến máy chủ liên quan đến dữ liệu thống kê hoặc tổng hợp.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để vô hiệu hóa việc giải mã và/hoặc tái tạo ít nhất một luồng trên cơ sở siêu dữ liệu liên quan đến ít nhất một luồng và trên cơ sở cổng nhìn và/hoặc hướng đầu và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại của người dùng.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để: thao tác siêu dữ liệu liên quan đến nhóm các luồng âm thanh được chọn, dựa trên ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại hoặc ước tính của người dùng, để: chọn và/hoặc kích hoạt và/hoặc vô hiệu hóa các thành phần âm thanh cấu thành cảnh âm thanh sẽ được tái tạo; và/hoặc cho phép gộp tất cả các luồng âm thanh đã chọn vào một luồng âm thanh duy nhất.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để kiểm soát yêu cầu của ít nhất một luồng đến máy chủ trên cơ sở khoảng cách của vị trí người dùng từ ranh giới của các môi trường lân cận và/hoặc liên kết liên quan đến các cảnh khác nhau hoặc các số liệu khác liên quan đến vị trí của người dùng trong môi trường hiện tại hoặc dự đoán về môi trường tương lai.

Theo một khía cạnh trong hệ thống, thông tin có thể được cung cấp từ hệ thống máy chủ, đối với từng thành phần âm thanh hoặc đối tượng âm thanh, trong đó thông tin bao gồm thông tin mô tả về các vị trí mà cảnh âm thanh hoặc các thành phần âm thanh đang hoạt động.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để lựa chọn giữa việc tái tạo một cảnh và biên soạn hoặc trộn hoặc ghép hoặc chồng hoặc kết hợp ít nhất hai cảnh trên cơ sở cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc vị trí ảo và/hoặc lựa chọn của người dùng hiện tại hoặc tương lai, hai cảnh liên quan đến các môi trường lân cận và/hoặc liên kết khác nhau.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để tạo hoặc sử dụng ít nhất là các tập hợp điều chỉnh sao cho: số tập hợp điều chỉnh liên quan đến một cảnh âm thanh; và/hoặc thông tin bổ sung được cung cấp mà từng tập hợp điều chỉnh có liên

quan đến một điểm nhìn, hoặc một cảnh âm thanh; và/hoặc thông tin bổ sung được cung cấp có thể bao gồm: thông tin về ranh giới của một cảnh âm thanh và/hoặc thông tin về mối quan hệ giữa một tập hợp điều chỉnh và một cảnh âm thanh (ví dụ, cảnh âm thanh được mã hóa thành ba luồng được gói gọn thành ba tập hợp điều chỉnh) và/hoặc thông tin về kết nối giữa ranh giới của cảnh âm thanh và nhiều tập hợp điều chỉnh.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để: nhận luồng cho cảnh liên quan đến môi trường lân cận hoặc liền kề; bắt đầu giải mã và/hoặc tái tạo luồng cho môi trường lân cận hoặc liền kề khi phát hiện sự chuyển đổi ranh giới giữa hai môi trường.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để hoạt động như máy khách và máy chủ được tạo cấu hình để phân phối các luồng video và/hoặc âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để: yêu cầu và/hoặc nhận ít nhất một tập hợp điều chỉnh thứ nhất bao gồm ít nhất một luồng âm thanh liên quan đến ít nhất một cảnh âm thanh thứ nhất; yêu cầu và/hoặc nhận ít nhất một tập hợp điều chỉnh thứ hai bao gồm ít nhất một luồng âm thanh thứ hai liên quan đến ít nhất hai cảnh âm thanh, gồm ít nhất một cảnh âm thanh thứ nhất; và cho phép gộp ít nhất một luồng âm thanh thứ nhất và ít nhất một luồng âm thanh thứ hai thành luồng âm thanh mới để được giải mã, dựa trên siêu dữ liệu có sẵn về cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại của người dùng và/hoặc thông tin mô tả mối quan hệ của ít nhất một tập hợp điều chỉnh thứ nhất với ít nhất một cảnh âm thanh thứ nhất và/hoặc mối quan hệ của ít nhất một tập hợp điều chỉnh thứ hai với ít nhất một cảnh âm thanh thứ nhất.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để nhận thông tin về cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại của người dùng và/hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng; và nhận thông tin về tính khả dụng của các tập hợp điều chỉnh và thông tin mô tả mối quan hệ của ít nhất một tập hợp điều chỉnh với ít nhất một cảnh và/hoặc điểm nhìn và/hoặc cổng nhìn và/hoặc vị trí và/hoặc vị trí ảo và/hoặc dữ liệu chuyển động và/hoặc hướng.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để quyết định xem có ít nhất một thành phần âm thanh từ ít nhất một cảnh âm thanh được lồng vào ít nhất một luồng và ít nhất một thành phần âm thanh bổ sung từ ít nhất một cảnh âm thanh bổ sung được lồng vào ít nhất một luồng bổ sung sẽ được tái tạo; và gây ra, trong trường hợp quyết định là có, thao tác gộp hoặc biên soạn hoặc trộn hoặc chồng hoặc kết hợp ít nhất một luồng bổ sung của cảnh âm thanh bổ sung vào ít nhất một luồng của ít nhất một cảnh âm thanh.

Theo một khía cạnh, hệ thống có thể được tạo cấu hình để thao tác siêu dữ liệu âm thanh liên quan đến các luồng âm thanh được chọn, dựa trên ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo hiện tại của người dùng, để: chọn và/hoặc kích hoạt và/hoặc vô hiệu hóa các thành phần âm thanh cấu thành cảnh âm thanh sẽ được tái tạo; và/hoặc cho phép gộp tất cả các luồng âm thanh đã chọn vào một luồng âm thanh duy nhất.

Theo một khía cạnh, máy chủ có thể được cung cấp để phân phối luồng âm thanh và video cho máy khách cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ, các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện, trong đó máy chủ có thể bao gồm bộ mã hóa để mã hóa và/hoặc lưu trữ để lưu trữ các luồng video để mô tả môi trường trực quan, môi trường trực quan được liên kết với cảnh âm thanh; trong đó máy chủ có thể còn bao gồm bộ mã hóa để mã hóa và/hoặc lưu trữ để lưu trữ nhiều luồng và/hoặc thành phần âm thanh và/hoặc tập hợp điều chỉnh được phân phối đến máy khách, các luồng và/hoặc các thành phần âm thanh và/hoặc các tập hợp điều chỉnh liên quan đến ít nhất một cảnh âm thanh, trong đó máy chủ được tạo cấu hình để: chọn và phân phối luồng video trên cơ sở yêu cầu từ máy khách, luồng video được liên kết với môi trường; chọn luồng âm thanh và/hoặc thành phần âm thanh và/hoặc tập hợp điều chỉnh trên cơ sở yêu cầu từ máy khách, yêu cầu liên quan đến ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng và với cảnh âm thanh liên quan đến môi trường; và phân phối luồng âm thanh cho máy khách.

Theo một khía cạnh, các luồng có thể được gói gọn thành các tập hợp điều chỉnh, mỗi tập hợp điều chỉnh bao gồm nhiều luồng liên quan đến các sự biểu diễn khác nhau, ở tốc độ bit và/hoặc chất lượng khác nhau, của cùng một nội dung âm thanh, trong đó tập hợp điều chỉnh đã chọn được chọn trên cơ sở của yêu cầu từ máy khách.

Theo một khía cạnh, hệ thống có thể hoạt động như máy khách và máy chủ.

Theo một khía cạnh, hệ thống có thể bao gồm một máy chủ.

Theo một khía cạnh, phương pháp có thể được đề xuất cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ được tạo cấu hình để nhận các luồng video và/hoặc âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện (ví dụ, thiết bị phát lại), bao gồm: giải mã các tín hiệu video từ các luồng video để biểu diễn các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng, và giải mã các tín hiệu âm thanh từ các luồng âm thanh, yêu cầu và/hoặc lấy từ máy chủ ít nhất một luồng âm thanh trên cơ sở cổng nhìn và/hoặc dữ liệu vị trí và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo và/hoặc siêu dữ liệu hiện tại của người dùng.

Theo một khía cạnh, chương trình máy tính có thể được đề xuất bao gồm các lệnh mà khi được chạy bởi bộ xử lý sẽ làm cho bộ xử lý thực hiện phương pháp trên.

Mô tả vắn tắt các hình vẽ

Các hình vẽ từ Fig.1.1 đến Fig.1.8 thể hiện các ví dụ của sáng chế.

Các hình vẽ từ Fig.2 đến Fig.6 thể hiện các bối cảnh của sáng chế.

Các hình vẽ từ Fig.7A đến Fig.8B thể hiện các phương pháp của sáng chế.

Mô tả chi tiết sáng chế

Sau đây (ví dụ, các hình vẽ từ Fig.1.1 đến Fig.1.8) có các ví dụ được bộc lộ của hệ thống theo các khía cạnh của sáng chế).

Các ví dụ về hệ thống theo sáng chế (có thể được thể hiện bằng các ví dụ khác nhau được bộc lộ dưới đây) được chỉ định chung bởi số tham chiếu 102. Ví dụ, hệ thống 102 có thể là hệ thống máy khách, vì nó có thể lấy từ hệ thống máy chủ (ví dụ

120) các luồng âm thanh và/hoặc video để biểu diễn các cảnh âm thanh và/hoặc các môi trường trực quan cho người dùng. Hệ thống máy khách 102 cũng có thể nhận siêu dữ liệu từ hệ thống máy chủ 120, ví dụ, thông tin phụ và/hoặc hỗ trợ liên quan đến các luồng âm thanh và/hoặc video.

Hệ thống 102 có thể được liên kết với (hoặc bao gồm trong một số ví dụ) thiết bị tiêu dùng phương tiện (MCD) tái tạo trên thực tế các tín hiệu âm thanh và/hoặc video cho người dùng. Trong một số ví dụ, người dùng có thể đeo MCD.

Hệ thống 102 có thể thực hiện các yêu cầu tới hệ thống máy chủ 120, các yêu cầu liên quan đến ít nhất một cổng nhìn và/hoặc hướng đầu (ví dụ, hướng góc) và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng 110 (một số số liệu có thể được cung cấp). Cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo 110 có thể được cung cấp trong phản hồi từ MCD tới hệ thống máy khách 102, mà lần lượt có thể cung cấp yêu cầu cho hệ thống máy chủ 120 trên cơ sở phản hồi này.

Trong một số trường hợp, yêu cầu (được biểu thị bởi số tham chiếu 112) có thể chứa cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo 110 (hoặc chỉ báo hoặc phiên bản được xử lý của nó). Trên cơ sở cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng 110, hệ thống máy chủ 120 sẽ cung cấp các luồng âm thanh và/hoặc video và/hoặc siêu dữ liệu cần thiết. Trong trường hợp này, hệ thống máy chủ 120 có thể có hiểu biết về vị trí của người dùng (ví dụ, trong môi trường ảo) và có thể liên kết các luồng chính xác với các vị trí của người dùng.

Trong các trường hợp khác, yêu cầu 112 từ hệ thống máy khách 102 có thể chứa các yêu cầu rõ ràng của các luồng âm thanh và/hoặc video cụ thể. Yêu cầu 112 có thể có trong trường hợp này dựa trên cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng 110. Hệ thống máy khách 102 có hiểu biết về tín hiệu âm thanh và video phải được kết xuất cho người dùng, ngay cả khi hệ thống máy khách 102 không có các

luồng cần thiết được lưu trữ trong đó. Ví dụ, hệ thống máy khách 102 có thể xử lý các luồng cụ thể trong hệ thống máy chủ 120.

Hệ thống máy khách 102 có thể là hệ thống cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống 102 bao gồm:

ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh 104 được tạo cấu hình để giải mã các tín hiệu âm thanh (108) từ ít nhất một luồng âm thanh 106,

trong đó hệ thống 102 được tạo cấu hình để yêu cầu 112 ít nhất một luồng âm thanh 106 và/hoặc một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh cho máy chủ 120 trên cơ sở ít nhất là công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng 110.

Cần lưu ý rằng trong các môi trường VR, AR, MR, người dùng 140 có thể có nghĩa là ở trong một môi trường cụ thể (ví dụ, một phòng cụ thể). Môi trường được mô tả với các tín hiệu video được mã hóa, ví dụ, ở phía máy chủ (phía hệ thống máy chủ 120 mà không nhất thiết phải bao gồm hệ thống máy chủ 120, nhưng có thể bao gồm bộ mã hóa khác đã được mã hóa trước đó các luồng video sau đó được lưu trữ trong bộ lưu trữ của máy chủ 120). Tại mỗi thời điểm, trong một số ví dụ, người dùng chỉ có thể thưởng thức một số tín hiệu video (ví dụ, công nhìn).

Nói chung, mỗi môi trường có thể được liên kết với cảnh âm thanh cụ thể. Cảnh âm thanh có thể được hiểu là tập hợp của tất cả các âm thanh sẽ được tái tạo cho người dùng trong môi trường cụ thể và trong một khoảng thời gian cụ thể.

Thông thường, các môi trường được hiểu là ở một số lượng riêng biệt. Theo đó, số lượng môi trường được hiểu là hữu hạn. Vì những lý do tương tự, số lượng cảnh âm

thanh được hiểu là hữu hạn. Do đó, trong kỹ thuật trước đây, các hệ thống VR, AR, MR đã được thiết kế sao cho:

- Người dùng dự định ở trong một môi trường duy nhất tại mỗi thời điểm; do đó, đối với từng môi trường:

- Hệ thống máy khách 102 yêu cầu hệ thống máy chủ 120 chỉ các luồng video được liên kết với môi trường duy nhất;
- Hệ thống máy khách 102 yêu cầu hệ thống máy chủ 120 chỉ các luồng âm thanh được liên kết với cảnh duy nhất.

Cách tiếp cận này đã dẫn đến các sự bất tiện.

Ví dụ, tất cả các luồng âm thanh sẽ được phân phối cùng nhau đến hệ thống máy khách 102 cho từng cảnh/môi trường, và các luồng âm thanh hoàn toàn mới cần được phân phối khi người dùng di chuyển đến môi trường khác (ví dụ, khi người dùng đi qua cửa, do đó bao hàm việc truyền môi trường/cảnh).

Hơn nữa, trải nghiệm không tự nhiên đã xảy ra trong một số trường hợp: ví dụ khi người dùng ở gần tường (ví dụ, tường ảo của phòng ảo), người đó nên trải nghiệm âm thanh phát ra từ phía bên kia của bức tường. Tuy nhiên, trải nghiệm này là không thể đối với các môi trường truyền thống: việc thu thập các luồng âm thanh liên quan đến cảnh hiện tại rõ ràng không chứa bất kỳ luồng nào liên quan đến các môi trường/cảnh liền kề.

Mặt khác, trải nghiệm của người dùng thường được cải thiện khi tốc độ bit của luồng âm thanh được tăng lên. Điều này có thể gây ra thêm các vấn đề: tốc độ bit càng cao, tải trọng mà hệ thống máy chủ cần phân phối cho hệ thống máy khách 102 càng cao. Ví dụ, khi cảnh âm thanh chứa nhiều nguồn âm thanh (được tải dưới dạng các thành phần âm thanh), một số trong số chúng nằm gần vị trí người dùng và các nguồn khác ở xa nó, các nguồn âm thanh ở xa sẽ khó nghe hơn. Do đó, việc phân phối tất cả các thành phần âm thanh ở cùng tốc độ bit hoặc mức chất lượng có thể dẫn đến tốc độ bit rất cao. Điều này chứng tỏ việc phân phối luồng âm thanh không hiệu quả. Nếu hệ thống máy chủ 120 phân phối các luồng âm thanh ở tốc độ bit cao nhất có thể, thì việc phân phối không hiệu quả sẽ xảy ra, vì các âm thanh có mức độ nghe được thấp hoặc mức độ liên quan thấp đến cảnh âm thanh tổng thể sẽ không đòi hỏi tốc độ bit cao,

tương tự như các âm thanh có liên quan được tạo ra gần hơn với người dùng. Do đó, nếu tất cả các luồng âm thanh của một cảnh được phân phối ở tốc độ bit cao nhất, truyền thông giữa hệ thống máy chủ 120 và hệ thống máy khách 102 sẽ tăng tải một cách không cần thiết. Nếu tất cả các luồng âm thanh của một cảnh được phân phối ở tốc độ bit thấp hơn, trải nghiệm của người dùng sẽ không thỏa đáng.

Các vấn đề về truyền thông làm trầm trọng thêm sự bất tiện đã thảo luận ở trên: khi người dùng đi qua cửa, người đó sẽ phải thay đổi ngay lập tức môi trường/cảnh, điều đó ngay lập tức sẽ yêu cầu hệ thống máy chủ 120 cung cấp tất cả các luồng cho hệ thống máy khách 102.

Do đó, thông thường, không thể giải quyết các vấn đề được thảo luận ở trên.

Tuy nhiên, với sáng chế, có thể giải quyết các vấn đề này: hệ thống máy khách 102 cung cấp yêu cầu cho hệ thống máy chủ 120 mà cũng có thể dựa trên công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng (và không chỉ dựa trên môi trường/cảnh). Theo đó, hệ thống máy chủ 120 có thể cung cấp, cho mỗi thời điểm, các luồng âm thanh sẽ được hiển thị ví dụ cho từng vị trí của người dùng.

Ví dụ, nếu người dùng không bao giờ đến gần tường, thì không cần hệ thống máy khách 102 để yêu cầu các luồng của môi trường lân cận (ví dụ, chúng có thể được yêu cầu bởi hệ thống máy khách 102 chỉ khi người dùng tiếp cận tường). Hơn nữa, các luồng phát ra từ bên ngoài bức tường có thể có tốc độ bit giảm, vì chúng có thể được nghe ở âm lượng thấp. Đáng chú ý, các luồng có liên quan hơn (ví dụ, các luồng đến từ các đối tượng âm thanh trong môi trường hiện tại) có thể được hệ thống máy chủ 120 phân phối đến hệ thống máy khách 102 ở tốc độ bit cao nhất và/hoặc mức chất lượng cao nhất (do hậu quả của việc thực tế là các luồng ít liên quan hơn ở tốc độ bit và/hoặc mức chất lượng thấp hơn, do đó để lại băng trống cho các luồng phù hợp hơn).

Ví dụ, mức chất lượng thấp hơn có thể đạt được bằng cách giảm tốc độ bit hoặc bằng cách xử lý các thành phần âm thanh theo cách giảm dữ liệu cần thiết để truyền đi, trong khi tốc độ bit được sử dụng trên mỗi tín hiệu âm thanh được giữ không đổi. Ví dụ, nếu số lượng 10 đối tượng âm thanh được đặt ở các vị trí khác nhau cách xa người

dùng, thì các đối tượng này có thể được trộn vào một số tín hiệu thấp hơn dựa trên vị trí của người dùng:

- Tại các vị trí rất xa vị trí người dùng (ví dụ cao hơn ngưỡng thứ nhất), các đối tượng được trộn thành 2 tín hiệu (có thể là các số lượng khác, dựa trên vị trí không gian và ngữ nghĩa của chúng) và được phân phối dưới dạng 2 "đối tượng ảo".
- Tại các vị trí gần vị trí người dùng hơn (ví dụ thấp hơn ngưỡng thứ nhất nhưng cao hơn ngưỡng thứ hai nhỏ hơn ngưỡng thứ nhất), các đối tượng được trộn thành 5 tín hiệu (dựa trên vị trí không gian và ngữ nghĩa của chúng) và được phân phối thành 5 (có thể là các số lượng khác) "đối tượng ảo".
- Tại các vị trí rất gần với vị trí của người dùng (thấp hơn ngưỡng thứ nhất và thứ hai), 10 đối tượng được phân phối dưới dạng 10 tín hiệu âm thanh được cung cấp chất lượng cao nhất.

Trong khi đối với chất lượng cao nhất, tất cả các tín hiệu âm thanh có thể được coi là rất quan trọng và có thể nghe được, người dùng có thể định vị riêng từng đối tượng. Đối với các mức chất lượng thấp hơn ở các vị trí ở xa, một số đối tượng âm thanh có thể trở nên ít liên quan hơn hoặc ít nghe thấy hơn, do đó, người dùng sẽ không thể định vị riêng các tín hiệu âm thanh trong không gian và do đó làm giảm mức chất lượng để phân phối các tín hiệu âm thanh này sẽ không dẫn đến giảm chất lượng trải nghiệm cho người dùng.

Một ví dụ khác là khi người dùng đi ra khỏi cửa: ở vị trí chuyển tiếp (ví dụ, tại ranh giới giữa hai môi trường/cảnh khác nhau), hệ thống máy chủ 120 sẽ cung cấp cả hai luồng của cả hai các cảnh/môi trường, nhưng ở các tốc độ bit thấp hơn. Điều này là do người dùng sẽ trải nghiệm các âm thanh đến từ hai môi trường khác nhau (các âm thanh có thể được gộp với nhau từ các luồng âm thanh khác nhau được liên kết với các cảnh/môi trường khác nhau) và không cần mức chất lượng cao nhất của từng nguồn âm thanh (hoặc thành phần âm thanh).

Theo quan điểm trên, sáng chế cho phép vượt xa cách tiếp cận truyền thống về số lượng môi trường trực quan và cảnh âm thanh riêng biệt, nhưng có thể cho phép biểu diễn dần dần các môi trường/cảnh khác nhau, mang lại trải nghiệm chân thực hơn cho người dùng.

Sau đây nó được coi là từng môi trường trực quan (ví dụ, môi trường ảo) được liên kết với cảnh âm thanh (các thuộc tính của các môi trường cũng có thể là các thuộc tính của cảnh). Mỗi môi trường/cảnh có thể được liên kết, ví dụ, với hệ tọa độ hình học (có thể là hệ tọa độ hình học ảo). Môi trường/cảnh có thể có các ranh giới, do đó, khi vị trí người dùng (ví dụ, vị trí ảo) vượt ra ngoài ranh giới, sẽ tới môi trường/cảnh khác. Các ranh giới có thể được dựa trên hệ tọa độ được sử dụng. Môi trường có thể bao gồm các đối tượng âm thanh (các thành phần âm thanh, nguồn âm thanh) có thể được định vị trong một số tọa độ cụ thể của môi trường/cảnh. Ví dụ, đối với vị trí và/hoặc hướng tương đối của người dùng đối với các đối tượng âm thanh (các thành phần âm thanh, nguồn âm thanh), hệ thống máy khách 102 có thể yêu cầu các luồng khác nhau và/hoặc hệ thống máy chủ 120 có thể cung cấp các luồng khác nhau (ví dụ, ở các tốc độ bit và/hoặc các mức chất lượng cao hơn/thấp hơn theo khoảng cách và/hoặc hướng).

Nói chung, hệ thống máy khách 102 có thể yêu cầu và/hoặc lấy từ hệ thống máy chủ 120 các luồng khác nhau (ví dụ, các sự biểu diễn khác nhau của cùng các âm thanh ở các tốc độ bit và/hoặc mức chất lượng khác nhau) trên cơ sở khả năng nghe thấy và/hoặc mức độ liên quan của chúng. Khả năng nghe thấy và/hoặc mức độ liên quan có thể được xác định, ví dụ, trên cơ sở ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng.

Trong một vài ví dụ, có thể có sự gộp các luồng khác nhau. Trong một số trường hợp, có thể có sự biên tập hoặc trộn hoặc ghép hoặc chồng hoặc kết hợp ít nhất hai cảnh. Ví dụ, có thể có việc sử dụng bộ trộn và/hoặc bộ kết xuất (mà ví dụ có thể được sử dụng hướng xuôi của nhiều bộ giải mã, mỗi bộ giải mã giải mã ít nhất một luồng âm thanh), hoặc thực hiện thao tác ghép luồng, ví dụ, hướng ngược giải mã các luồng. Trong các trường hợp khác, có thể có sự giải mã các luồng khác nhau và kết xuất chúng với các thiết lập loa phóng thanh khác nhau.

Cần lưu ý rằng sáng chế không nhất thiết phải từ chối khái niệm môi trường trực quan và cảnh âm thanh. Đặc biệt, với sáng chế, các luồng âm thanh và video liên quan đến một cảnh/môi trường cụ thể có thể được phân phối từ hệ thống máy chủ 120

đến hệ thống máy khách 102 khi người dùng vào môi trường/cảnh. Tuy vậy, trong cùng một môi trường/cảnh, các luồng âm thanh khác nhau và/hoặc các đối tượng âm thanh và/hoặc các tập hợp điều chỉnh có thể được yêu cầu, giải quyết và/hoặc phân phối. Cụ thể, có thể có khả năng là:

- ít nhất một số dữ liệu video liên quan đến môi trường trực quan được phân phối từ máy chủ 120 đến máy khách 102 tại lối vào của người dùng vào cảnh; và/hoặc

- ít nhất một số dữ liệu âm thanh (luồng, đối tượng, tập hợp điều chỉnh, ...) chỉ được phân phối đến hệ thống máy khách 102 trên cơ sở cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc vị trí ảo và/hoặc lựa chọn/tương tác của người dùng hiện tại (hoặc tương lai); và/hoặc

- (trong một số trường hợp): một số dữ liệu âm thanh được phân phối cho hệ thống máy khách 102 trên cơ sở cảnh hiện tại (không phân biệt vị trí hiện tại hoặc tương lai hoặc cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc vị trí ảo và/hoặc lựa chọn của người dùng), trong khi dữ liệu âm thanh còn lại được phân phối trên cơ sở hiện tại hoặc tương lai hoặc cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc vị trí ảo và/hoặc lựa chọn của người dùng.

Cần lưu ý rằng các thành phần khác nhau (hệ thống máy chủ, hệ thống máy khách, MCD, v.v.) có thể biểu diễn các thành phần trong các thiết bị phần cứng khác nhau hoặc thậm chí trong cùng một thiết bị (ví dụ, máy khách và MCD có thể được triển khai như một phần của cùng một điện thoại di động, hoặc tương tự, máy khách có thể ở trong PC được kết nối với màn hình thứ cấp bao gồm MCD).

Các ví dụ:

Một phương án của hệ thống 102 (máy khách) như được thể hiện trên Fig.1.1 được tạo cấu hình để nhận các luồng (âm thanh) 106 trên cơ sở vị trí xác định trong môi trường (ví dụ, môi trường ảo), có thể được hiểu là có liên quan đến cảnh video và âm thanh (sau đây đề cập là cảnh 150). Các vị trí khác nhau trong cùng một cảnh 150 nói chung bao hàm các luồng khác nhau 106 hoặc siêu dữ liệu khác nhau được liên kết với các luồng 106 được cung cấp cho bộ giải mã âm thanh 104 của hệ thống 102 (ví dụ, từ máy chủ phương tiện 120). Hệ thống 102 được kết nối với thiết bị tiêu dùng

phương tiện (MCD) mà từ đó nó nhận được phản hồi liên quan đến vị trí và/hoặc vị trí ảo của người dùng trong cùng một môi trường. Sau đây, vị trí của người dùng trong môi trường có thể liên quan đến cổng nhìn cụ thể mà người dùng mong muốn (cổng nhìn ví dụ được dự định là bề mặt giả sử như bề mặt hình chữ nhật được chiếu trên hình cầu, được biểu diễn cho người dùng).

Trong bối cảnh mẫu, khi người dùng di chuyển trong cảnh VR, AR và/hoặc MR 150, nội dung âm thanh có thể được tưởng tượng là gần như được tạo bởi một hoặc nhiều nguồn âm thanh 152 mà có thể thay đổi. Các nguồn âm thanh 152 có thể được hiểu là các nguồn âm thanh ảo, theo nghĩa là chúng có thể đề cập đến các vị trí trong môi trường ảo: việc kết xuất của từng nguồn âm thanh được điều chỉnh theo vị trí của người dùng (ví dụ, trong ví dụ minh họa được đơn giản hóa, mức độ của nguồn âm thanh cao hơn khi người dùng ở gần vị trí của nguồn âm thanh hơn, và thấp hơn khi người dùng ở xa nguồn âm thanh hơn). Mỗi thành phần âm thanh (nguồn âm thanh) tuy vậy không được mã hóa trong các luồng âm thanh được cung cấp cho bộ giải mã. Các luồng âm thanh có thể được liên kết với các vị trí và/hoặc các khu vực khác nhau trong cảnh. Ví dụ các nguồn âm thanh 152 không nghe được trong một cảnh có thể trở thành nghe được trong cảnh tiếp theo, ví dụ, khi cánh cửa trong cảnh VR, AR và/hoặc MR 150 được mở. Sau đó, người dùng có thể chọn để vào cảnh/môi trường mới 150 (ví dụ, phòng), và toàn bộ cảnh âm thanh thay đổi. Với mục đích mô tả bối cảnh này, khái niệm các điểm nhìn rời rạc trong không gian được sử dụng, dưới dạng vị trí riêng biệt trong không gian (hoặc trong môi trường VR), mà có sẵn nội dung âm thanh khác nhau.

Nói chung, máy chủ phương tiện 120 có thể cung cấp các luồng 106 liên quan đến cảnh 150 cụ thể trên cơ sở vị trí của người dùng trong cảnh 150. Các luồng 106 có thể được mã hóa bởi ít nhất một bộ mã hóa 154 và được cung cấp cho máy chủ phương tiện 120. Máy chủ phương tiện 120 có thể truyền các luồng 113 với các truyền thông 113 (ví dụ, thông qua mạng lưới truyền thông). Việc cung cấp các luồng 113 có thể dựa trên các yêu cầu 112 được đặt ra bởi hệ thống 102 trên cơ sở vị trí 110 của người dùng (ví dụ, trong môi trường ảo). Vị trí 110 của người dùng cũng có thể được hiểu là được liên kết với cổng nhìn mà người dùng mong muốn (đối với mỗi vị trí, có một hình chữ nhật duy nhất được biểu diễn) và với điểm nhìn (vì điểm nhìn là trung

tâm của cổng nhìn). Do đó, việc cung cấp cổng nhìn có thể, trong một số ví dụ, giống như việc cung cấp vị trí.

Hệ thống 102 như được thể hiện trên Fig.1.2 được tạo cấu hình để nhận các luồng (âm thanh) 113 trên cơ sở một cấu hình khác ở phía máy khách. Trong ví dụ về việc triển khai ở phía mã hóa này, phần lớn các bộ mã hóa phương tiện 154 được cung cấp có thể được sử dụng để tạo một hoặc nhiều luồng 106 cho mỗi cảnh 150 có sẵn liên quan đến một phần cảnh âm thanh của một điểm nhìn.

Máy chủ phương tiện 120 có thể lưu trữ nhiều tập hợp điều chỉnh âm thanh và video (không được thể hiện trên các hình vẽ) bao gồm các sự mã hóa khác nhau của cùng các luồng âm thanh và video ở các tốc độ bit khác nhau. Ngoài ra, máy chủ phương tiện có thể chứa thông tin mô tả của tất cả các tập hợp điều chỉnh, có thể bao gồm tính khả dụng của tất cả các tập hợp điều chỉnh đã tạo. Các tập hợp điều chỉnh cũng có thể chứa thông tin mô tả mối liên hệ của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn cụ thể. Theo cách này, mỗi tập hợp điều chỉnh có thể được liên kết với một trong số cảnh âm thanh có sẵn.

Các tập hợp điều chỉnh có thể chứa thêm thông tin mô tả ranh giới của từng cảnh âm thanh và/hoặc điểm nhìn mà có thể chứa ví dụ cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng âm thanh riêng lẻ. Ranh giới của một cảnh âm thanh có thể được xác định, ví dụ, dưới dạng các tọa độ hình học của một hình cầu (ví dụ, tâm và bán kính).

Hệ thống 102 ở phía máy khách có thể nhận thông tin về cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại của người dùng hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi các hành động của người dùng. Ngoài ra, hệ thống 102 cũng có thể nhận được thông tin về tính khả dụng của tất cả các tập hợp điều chỉnh và thông tin mô tả mối liên hệ của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn; và/hoặc thông tin mô tả "các ranh giới" của từng cảnh âm thanh và/hoặc điểm nhìn (mà có thể chứa ví dụ, cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng riêng lẻ). Ví dụ, thông tin như vậy có thể được cung cấp như một phần của cú pháp mô tả biểu diễn phương tiện (MPD) XML trong trường hợp môi trường phân phối DASH.

Hệ thống 102 có thể cung cấp tín hiệu âm thanh cho thiết bị tiêu dùng phương tiện (MCD) được sử dụng để tiêu dùng nội dung. Thiết bị tiêu dùng phương tiện cũng chịu trách nhiệm thu thập thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi hành động của người dùng) như là dữ liệu vị trí và chuyển tiếp 110.

Bộ xử lý công nhìn 1232 có thể được tạo cấu hình để nhận dữ liệu vị trí và chuyển tiếp 110 đã nêu từ phía thiết bị tiêu dùng phương tiện. Bộ xử lý công nhìn 1232 cũng có thể nhận thông tin về ROI được báo hiệu trong siêu dữ liệu và tất cả thông tin có sẵn ở đầu nhận (hệ thống 102). Bộ xử lý công nhìn 1232 sau đó có thể quyết định điểm nhìn âm thanh nào sẽ được tái tạo tại một thời điểm nhất định dựa trên tất cả thông tin nhận được và/hoặc được suy ra từ siêu dữ liệu đã nhận và/hoặc có sẵn. Ví dụ, bộ xử lý công nhìn 1232 có thể quyết định rằng một cảnh âm thanh hoàn chỉnh sẽ được tái tạo, một cảnh âm thanh mới 108 phải được tạo ra trong tất cả các cảnh âm thanh có sẵn, ví dụ, chỉ một số thành phần âm thanh từ nhiều cảnh âm thanh được tái tạo, trong khi các thành phần âm thanh còn lại khác của các cảnh âm thanh này sẽ không được tái tạo. Bộ xử lý công nhìn 1232 cũng có thể quyết định xem có sự chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh phải được tái tạo không.

Phần lựa chọn 1230 có thể được cung cấp để lựa chọn, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232, một hoặc nhiều tập hợp điều chỉnh ngoài các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được từ đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh cảnh âm thanh cần được tái tạo tại vị trí hiện tại của người dùng. Cảnh âm thanh này có thể là một cảnh âm thanh hoàn chỉnh như được xác định ở phía mã hóa hoặc cảnh âm thanh mới có thể phải được tạo ra trong tất cả các cảnh âm thanh có sẵn.

Ngoài ra, trong trường hợp khi sự chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh sắp xảy ra dựa trên chỉ báo của bộ xử lý công nhìn 1232, phần lựa chọn có thể được tạo cấu hình để chọn một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được bởi đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh cảnh âm thanh mà có thể được yêu cầu tái tạo trong tương lai gần (ví dụ, nếu người dùng đi theo hướng của cảnh âm thanh tiếp

theo với tốc độ nhất định, có thể dự đoán rằng cảnh âm thanh tiếp theo sẽ được yêu cầu và nó được chọn trước khi tái tạo).

Ngoài ra, một số tập hợp điều chỉnh tương ứng với các vị trí lân cận có thể được chọn trước tiên ở tốc độ bit thấp hơn và/hoặc mức chất lượng thấp hơn, ví dụ sự biểu diễn được mã hóa ở tốc độ bit thấp hơn được chọn ra các sự biểu diễn có sẵn trong một tập hợp điều chỉnh, và dựa trên các thay đổi vị trí, chất lượng được tăng lên bằng cách chọn tốc độ bit cao hơn cho các tập hợp điều chỉnh cụ thể đó, ví dụ sự biểu diễn được mã hóa ở tốc độ bit cao hơn được chọn ra trong số các sự biểu diễn có sẵn trong một tập hợp điều chỉnh.

Phân tải xuống và chuyển đổi 1234 có thể được cung cấp để yêu cầu, dựa trên chỉ báo nhận được từ phân lựa chọn, một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện, được tạo cấu hình để nhận một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện và trích xuất thông tin siêu dữ liệu từ tất cả các luồng âm thanh nhận được.

Bộ xử lý siêu dữ liệu 1236 có thể được cung cấp để nhận từ thông tin tải xuống và chuyển đổi về các luồng âm thanh nhận được, thông tin có thể chứa siêu dữ liệu âm thanh tương ứng với từng luồng âm thanh nhận được. Bộ xử lý siêu dữ liệu 1236 cũng có thể được tạo cấu hình để xử lý và thao tác siêu dữ liệu âm thanh được liên kết với từng luồng âm thanh 113, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232 có thể bao gồm thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng 110, để chọn/kích hoạt các thành phần âm thanh cần thiết 152 biên soạn cảnh âm thanh mới như được chỉ định bởi bộ xử lý công nhìn 1232, cho phép gộp tất cả các luồng âm thanh 113 vào một luồng âm thanh 106 duy nhất.

Bộ ghép/gộp luồng 1238 có thể được tạo cấu hình để gộp tất cả các luồng âm thanh đã chọn vào một luồng âm thanh 106 dựa trên thông tin nhận được từ bộ xử lý siêu dữ liệu 1236 mà có thể bao gồm siêu dữ liệu âm thanh được biến đổi và được xử lý tương ứng với tất cả các luồng âm thanh nhận được 113.

Bộ giải mã phương tiện 104 được tạo cấu hình để nhận và giải mã ít nhất một luồng âm thanh để tái tạo cảnh âm thanh mới như được chỉ định bởi bộ xử lý công

nhìn 1232 dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng.

Trong một phương án khác, hệ thống 102 như được thể hiện trên Fig.1.7 có thể được tạo cấu hình để nhận các luồng âm thanh 106 ở các tốc độ bit và/hoặc mức chất lượng âm thanh khác nhau. Cấu hình phần cứng của phương án này tương tự như trên Fig.1.2. Ít nhất một cảnh môi trường trực quan 152 có thể được liên kết với ít nhất một số lượng lớn của N thành phần âm thanh ($N \geq 2$), mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan. Ít nhất một số lượng lớn của N thành phần âm thanh 152 được cung cấp ít nhất một sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng cao, và trong đó ít nhất một số lượng lớn của N thành phần âm thanh 152 được cung cấp trong ít nhất một sự biểu diễn tại tốc độ bit và/hoặc mức chất lượng thấp, trong đó có ít nhất một sự biểu diễn thu được bằng cách xử lý N phần tử âm thanh 152 để thu được số lượng nhỏ hơn M của các thành phần âm thanh 152 ($M < N$) được liên kết với vị trí hoặc khu vực gần vị trí hoặc khu vực của N thành phần âm thanh 152.

Việc xử lý N thành phần âm thanh 152 có thể ví dụ là một sự bổ sung đơn giản các tín hiệu âm thanh hoặc có thể là một sự trộn giảm hoạt động dựa trên vị trí không gian 110 của chúng hoặc kết xuất các tín hiệu âm thanh bằng cách sử dụng vị trí không gian của chúng vào vị trí ảo mới nằm ở giữa các tín hiệu âm thanh. Hệ thống có thể được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng cao hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh liên quan hơn và/hoặc dễ nghe hơn ở vị trí ảo của người dùng hiện tại trong cảnh, trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng thấp hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh ít liên quan hơn và/hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh.

Fig.1.8 thể hiện ví dụ về hệ thống (có thể là hệ thống 102) thể hiện hệ thống 102 cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ được tạo cấu hình để nhận luồng video 1800 và luồng âm thanh 106 được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống 102 bao gồm:

ít nhất một bộ giải mã video phương tiện 1804 được tạo cấu hình để giải mã các tín hiệu video 1808 từ các luồng video 1800 để biểu diễn các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh 104 được tạo cấu hình để giải mã các tín hiệu âm thanh 108 từ ít nhất một luồng âm thanh 106.

Hệ thống 102 được tạo cấu hình để yêu cầu (112) ít nhất một luồng âm thanh 106 và/hoặc một thành phần âm thanh của luồng âm thanh và/hoặc một tập hợp điều chỉnh cho máy chủ (ví dụ 120) trên cơ sở ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng 110 (ví dụ, được cung cấp dưới dạng phản hồi từ thiết bị tiêu dùng phương tiện 180).

Hệ thống 102 có thể giống với hệ thống 102 theo các hình vẽ từ Fig.1.1 đến Fig.1.7 và/hoặc lấy các bối cảnh theo Fig.2a, Fig.2b.

Các ví dụ của sáng chế cũng đề cập đến phương pháp cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ được tạo cấu hình để nhận các luồng video và/hoặc âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện (ví dụ, thiết bị phát lại), phương pháp bao gồm các bước:

giải mã các tín hiệu video từ các luồng video để biểu diễn các cảnh VR, AR, MR hoặc môi trường video 360 độ cho người dùng, và

giải mã các tín hiệu âm thanh từ các luồng âm thanh,

yêu cầu và/hoặc lấy từ máy chủ, ít nhất một luồng âm thanh trên cơ sở cổng nhìn và/hoặc dữ liệu vị trí và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu và/hoặc dữ liệu vị trí ảo và/hoặc siêu dữ liệu hiện tại của người dùng.

Trường hợp 1

Các cảnh môi trường khác nhau 150 nói chung bao hàm việc nhận các luồng 106 khác nhau từ máy chủ 120. Tuy nhiên, các luồng 106 mà bộ giải mã âm thanh 104 nhận được cũng có thể được điều chỉnh theo vị trí của người dùng trong cùng cảnh 150.

Tại thời điểm thứ nhất (bắt đầu) ($t=t_1$) được thể hiện trên Fig.2a, người dùng được đặt ví dụ trong cảnh 150, có vị trí được xác định đầu tiên trong môi trường VR (hoặc môi trường AR, hoặc môi trường MR). Trong hệ tọa độ Đề-các XYZ (ví dụ, nằm ngang), cổng nhìn (vị trí) thứ nhất 110' của người dùng được liên kết với các tọa độ x'_u và y'_u (trục Z ở đây hướng ra bề mặt giấy). Trong cảnh thứ nhất 150 này, hai thành phần âm thanh 152-1 và 152-1 được đặt, có các tọa độ tương ứng x'_1 và y'_1 đối với thành phần âm thanh 1 (152-1) và x'_2 and y'_2 đối với thành phần âm thanh 2 (152-2). Khoảng cách d'_1 của người dùng đến thành phần âm thanh 1 (152-1) nhỏ hơn khoảng cách d'_2 (152-1) của người dùng đến thành phần âm thanh 2. Tất cả dữ liệu vị trí người dùng (cổng nhìn) được truyền từ MCD đến hệ thống 102.

Tại thời điểm mẫu thứ hai ($t=t_2$) được thể hiện trên Fig.2b, người dùng được đặt ví dụ trong cùng một cảnh 150, nhưng ở vị trí thứ hai hoặc vị trí khác. Trong hệ tọa độ Đề-các XY, cổng nhìn thứ hai (vị trí) 110" của người dùng được liên kết với các tọa độ mới x''_u và y''_u (trục Z ở đây hướng ra bề mặt giấy). Bây giờ khoảng cách d''_1 của người dùng từ thành phần âm thanh 1 (152-1) lớn hơn khoảng cách d''_2 của người dùng từ thành phần âm thanh 2 (152-2). Tất cả dữ liệu vị trí người dùng (cổng nhìn) được truyền một lần nữa từ MCD đến hệ thống 102.

Người dùng được trang bị MCD đã nêu để hiển thị cổng nhìn nhất định trong môi trường 360 độ có thể đang nghe, ví dụ, qua tai nghe. Người dùng có thể thưởng thức việc tái tạo các âm thanh khác nhau cho các vị trí khác nhau được mô tả trên Fig.2a và Fig.2b của cùng một cảnh 150.

Bất kỳ vị trí và/hoặc bất kỳ sự chuyển tiếp và/hoặc cổng nhìn và/hoặc vị trí ảo và/hoặc hướng đầu và/hoặc dữ liệu chuyển động trong cảnh ví dụ từ Fig.2a đến Fig.2b có thể được truyền định kỳ (ví dụ, trong phản hồi) từ MCD đến hệ thống 102 (máy khách) dưới dạng tín hiệu 110. Máy khách có thể truyền lại dữ liệu vị trí và chuyển tiếp 110' hoặc 110", (ví dụ, dữ liệu cổng nhìn) đến máy chủ 120. Máy khách 102 hoặc máy chủ 120 có thể quyết định dựa trên dữ liệu vị trí và chuyển tiếp 110' hoặc 110" (ví dụ, dữ liệu cổng nhìn) mà các luồng âm thanh 106 được yêu cầu để tái tạo cảnh âm thanh chính xác tại vị trí người dùng hiện tại. Máy khách có thể quyết định và truyền yêu cầu 112 cho luồng âm thanh 106 tương ứng trong khi máy chủ 120 có thể được tạo

cấu hình để phân phối (các) luồng tương ứng 106 tùy thuộc vào thông tin vị trí được cung cấp bởi máy khách (hệ thống 102). Ngoài ra, máy chủ 120 có thể quyết định và phân phối (các) luồng 106 phù hợp tùy thuộc vào thông tin vị trí được cung cấp bởi máy khách (hệ thống 102).

Máy khách (hệ thống 102) có thể yêu cầu truyền các luồng sẽ được giải mã để biểu diễn cảnh 150. Trong một số ví dụ, hệ thống 102 có thể truyền thông tin về mức chất lượng cao nhất được tái tạo trên MCD (trong các ví dụ khác, chính máy chủ 120 quyết định mức chất lượng sẽ được tái tạo trên MCD, trên cơ sở vị trí của người dùng trong cảnh). Để phản hồi lại, máy chủ 120 có thể chọn một trong vô số các sự biểu diễn được liên kết với cảnh âm thanh sẽ được biểu diễn, để phân phối ít nhất một luồng 106 theo vị trí 110' hoặc 110" của người dùng. Do đó, máy khách (hệ thống 102) có thể được tạo cấu hình để phân phối, ví dụ, thông qua bộ giải mã âm thanh 104, tín hiệu âm thanh 108 cho người dùng để tái tạo âm thanh liên quan đến vị trí thực tế (hiệu quả) 110' hoặc 110" của người đó. (các tập hợp điều chỉnh 113 có thể được sử dụng: các biến thể khác nhau của cùng các luồng, ví dụ, ở các tốc độ bit khác nhau, có thể được sử dụng cho các vị trí khác nhau của người dùng.)

Luồng 106 (có thể được xử lý trước hoặc được tạo khi đang di chuyển) có thể được truyền đến máy khách (hệ thống 102), và có thể được tạo cấu hình cho vô số điểm nhìn liên quan đến các cảnh âm thanh nhất định.

Cần lưu ý rằng các chất lượng khác nhau (ví dụ, các tốc độ bit khác nhau) có thể được cung cấp cho các luồng 106 khác nhau theo vị trí cụ thể (ví dụ, 110' hoặc 110") của người dùng trong môi trường (ví dụ, ảo). Ví dụ: Trong trường hợp có nhiều nguồn âm thanh 152-1 và 152-2, mỗi nguồn âm thanh 152-1 và 152-2 có thể được liên kết với vị trí cụ thể trong cảnh 150. Vị trí của người dùng 110' hoặc 110" càng gần với nguồn âm thanh thứ nhất 152-1, thì độ phân giải và/hoặc chất lượng cần thiết của luồng được liên kết với nguồn âm thanh thứ nhất 152-2 càng cao. Trường hợp mẫu này có thể áp dụng cho thành phần âm thanh 1 (152-1) trên Fig.2a cũng như thành phần âm thanh 2 (152-2) trên Fig.2b. Khoảng cách vị trí của người dùng 110 từ nguồn âm thanh thứ hai 152-2 càng lớn, thì độ phân giải cần thiết của luồng 106 được liên kết với nguồn âm thanh thứ hai 152-2 càng thấp. Trường hợp mẫu này có thể áp dụng cho

thành phần âm thanh 2 (152-2) trên Fig.2a cũng như thành phần âm thanh 1 (152-1) trên Fig.2b.

Trong thực tế, thứ nhất, nguồn âm thanh gần phải được nghe ở mức cao hơn (và do đó được cung cấp ở tốc độ bit cao hơn), trong khi thứ hai, nguồn âm thanh xa sẽ được nghe ở mức thấp hơn (có thể cho phép yêu cầu độ phân giải thấp hơn);

Do đó, trên cơ sở vị trí 110' hoặc 110" trong môi trường do khách hàng 102 cung cấp, máy chủ 120 có thể cung cấp các luồng 106 khác nhau ở các tốc độ bit khác nhau (hoặc chất lượng khác). Dựa trên thực tế là các thành phần âm thanh ở xa không yêu cầu các mức chất lượng cao, chất lượng trải nghiệm tổng thể của người dùng vẫn được duy trì ngay cả khi chúng được phân phối ở tốc độ bit hoặc mức chất lượng thấp hơn.

Do đó, các mức chất lượng khác nhau có thể được sử dụng cho một số thành phần âm thanh ở các vị trí khác nhau của người dùng, trong khi vẫn giữ được chất lượng trải nghiệm.

Nếu không có giải pháp này, tất cả các luồng 106 sẽ được máy chủ 120 cung cấp cho máy khách ở tốc độ bit cao nhất, điều này sẽ làm tăng tải trọng trong kênh truyền thông từ máy chủ 120 đến máy khách.

Trường hợp 2

Fig.3 (trường hợp 2) thể hiện phương án với một kịch bản mẫu khác (được biểu diễn trong mặt phẳng thẳng đứng XZ của không gian XYZ, trong đó trục Y được biểu diễn là hướng vào bề mặt giấy), trong đó người dùng di chuyển trong cảnh VR, AR, và/hoặc MR thứ nhất A (150A), mở cửa và đi qua (chuyển 150AB) qua, ngụ ý sự chuyển tiếp âm thanh từ cảnh thứ nhất 150A tại thời điểm t_1 qua vị trí tạm thời (150AB) tại thời điểm t_2 sang cảnh tiếp theo (thứ hai) (150B) tại thời điểm t_3 .

Tại thời điểm t_1 , người dùng có thể ở vị trí x_1 theo hướng x của cảnh VR, AR và/hoặc MR thứ nhất. Tại thời điểm t_3 , người dùng có thể ở cảnh VR, AR và/hoặc MR thứ hai B (150B) khác ở vị trí x_3 . Tại thời điểm tức thời t_2 , người dùng có thể ở vị trí chuyển tiếp 150AB, trong khi người đó đang mở cửa (ví dụ, cửa ảo) và đi qua. Do đó,

sự chuyển tiếp chỉ sự chuyển tiếp thông tin âm thanh từ cảnh thứ nhất 150A sang cảnh thứ hai 150B.

Trong bối cảnh này, người dùng đang thay đổi vị trí của mình 110 ví dụ từ môi trường VR thứ nhất (được đặc trưng bởi điểm nhìn thứ nhất (A) như được thể hiện trên Fig.1.1) đến môi trường VR thứ hai (được đặc trưng bởi điểm nhìn thứ hai (B) như được thể hiện trên Fig.1.1). Trong trường hợp cụ thể, ví dụ, trong quá trình chuyển tiếp qua cửa được đặt tại vị trí x_2 theo hướng x , một số thành phần âm thanh 152A và 152B có thể xuất hiện ở cả hai điểm nhìn (vị trí A và B).

Người dùng (được trang bị MCD) đang thay đổi vị trí 110 (x_1 - x_3) về phía cửa, điều này có thể ngụ ý rằng, ở vị trí chuyển tiếp x_2 , các thành phần âm thanh thuộc cả cảnh thứ nhất 150A và cảnh thứ hai 150B. MCD truyền dữ liệu vị trí mới và chuyển tiếp 110 đến máy khách, mà truyền lại nó đến máy chủ phương tiện 120. Người dùng có thể được kích hoạt để nghe các nguồn âm thanh phù hợp được xác định bởi vị trí trung gian x_2 giữa vị trí thứ nhất và thứ hai x_1 và x_3 .

Bất kỳ vị trí và bất kỳ sự chuyển tiếp nào từ vị trí thứ nhất (x_1) sang vị trí thứ hai (x_3) hiện được truyền định kỳ (ví dụ, liên tục) từ MCD đến máy khách. Máy khách 102 có thể truyền lại dữ liệu vị trí và chuyển tiếp 110 (x_1 - x_3) tới máy chủ phương tiện 120, được tạo cấu hình để phân phối một mục chuyên dụng tương ứng của ví dụ tập hợp các luồng mới được xử lý trước 106 dưới dạng tập hợp điều chỉnh được hiện thực hóa 113', tùy thuộc vào dữ liệu vị trí và chuyển tiếp 110 nhận được (x_1 - x_3).

Máy chủ phương tiện 120 có thể chọn một trong vô số các sự biểu diễn liên quan đến thông tin đã nói ở trên, không chỉ liên quan đến khả năng MCD hiển thị tốc độ bit cao nhất mà còn liên quan đến dữ liệu vị trí và chuyển tiếp 110 (x_1 - x_3) của người dùng trong khi người đó chuyển động từ vị trí này sang vị trí khác. (Trong bối cảnh này, có thể sử dụng các tập hợp điều chỉnh: máy chủ phương tiện 120 có thể quyết định tập hợp điều chỉnh 113' nào biểu diễn tối ưu sự chuyển tiếp ảo của người dùng, mà không can thiệp vào khả năng kết xuất của MCD.)

Do đó, máy chủ phương tiện 120 có thể phân phối luồng chuyên dụng 106 (ví dụ, dưới dạng tập hợp điều chỉnh 113' mới) theo sự chuyển tiếp các vị trí. Máy khách

102 có thể được tạo cấu hình để phân phối tương ứng tín hiệu âm thanh 108 cho người dùng 140, ví dụ, thông qua bộ giải mã âm thanh phương tiện 104.

Các luồng 106 (được tạo khi đang di chuyển và/hoặc được xử lý trước) có thể được truyền trong tập hợp điều chỉnh 113' được thực tế hóa định kỳ (ví dụ, liên tục) tới máy khách 102.

Khi người dùng đi qua cửa, máy chủ 120 có thể truyền cả các luồng 106 của cảnh thứ nhất 150A và các luồng 106 của cảnh thứ hai 150B. Điều này là để trộn hoặc ghép hoặc biên soạn hoặc tái tạo đồng thời các luồng 106 này, để tạo ấn tượng chân thực cho người dùng. Do đó, trên cơ sở vị trí 110 của người dùng (ví dụ, "vị trí tương ứng với cửa"), máy chủ 120 truyền các luồng 106 khác nhau đến máy khách.

Ngay cả trong trường hợp này, vì các luồng 106 khác nhau được nghe đồng thời, chúng có thể có các độ phân giải khác nhau và có thể được truyền từ máy chủ 120 đến máy khách ở các độ phân giải khác nhau. Khi người dùng đã hoàn thành sự chuyển tiếp và ở cảnh (vị trí) thứ hai 150A (và đã đóng cánh cửa phía sau người đó), sẽ có khả năng máy chủ 120 giảm hoặc kiểm chế truyền các luồng 106 của cảnh thứ nhất 150 (trong trường hợp máy chủ 120 đã cung cấp cho máy khách 102 các luồng, máy khách 102 có thể quyết định không sử dụng chúng).

Trường hợp 3

Fig.4 (trường hợp 3) thể hiện phương án với một kịch bản mẫu khác (được biểu diễn trong mặt phẳng thẳng đứng XZ của không gian XYZ, trong đó trục Y được biểu diễn là hướng vào bề mặt giấy), trong đó người dùng di chuyển trong cảnh VR, AR, và/hoặc MR 150A ngụ ý sự chuyển tiếp âm thanh từ một cảnh thứ nhất tại thời điểm t_1 tới vị trí thứ hai cũng trong cảnh thứ nhất 150A tại thời điểm t_2 . Người dùng ở vị trí thứ nhất có thể cách xa tường tại thời điểm t_1 ở khoảng cách d_1 so với tường; và có thể ở gần tường tại thời điểm t_2 , ở khoảng cách d_2 so với tường. Ở đây, $d_1 > d_2$. Trong khi ở khoảng cách d_1 , người dùng chỉ nghe thấy nguồn 152A của cảnh 150A, người đó cũng có thể nghe thấy nguồn 152B của cảnh 150B ở bên kia tường.

Khi người dùng ở vị trí thứ hai (d_2), máy khách 102 sẽ gửi đến máy chủ 120 dữ liệu liên quan đến vị trí của người dùng 110 (d_2) và nhận, từ máy chủ 120, không chỉ

các luồng âm thanh 106 của cảnh thứ nhất 150A, mà còn nhận các luồng âm thanh 106 của cảnh thứ hai 150B. Ví dụ, trên cơ sở siêu dữ liệu được cung cấp bởi máy chủ 120, máy khách 102 sẽ gây ra sự tái tạo, ví dụ, thông qua bộ giải mã 104, của các luồng 106 của cảnh thứ hai 150B (xuyên qua tường) ở mức âm lượng thấp.

Ngay cả trong trường hợp này, tốc độ bit (chất lượng) của các luồng 106 của cảnh thứ hai 150B có thể thấp, do đó yêu cầu giảm tải truyền từ máy chủ 120 đến máy khách. Đáng chú ý là vị trí 110 (d_1 , d_2) của máy khách (và/hoặc công nhin) xác định các luồng âm thanh 106 được cung cấp bởi máy chủ 120.

Ví dụ, hệ thống 102 có thể được tạo cấu hình để thu được các luồng liên quan đến cảnh hiện tại thứ nhất (150A) được liên kết với môi trường thứ nhất hoặc hiện tại, và trong trường hợp khoảng cách của vị trí người dùng hoặc vị trí ảo từ ranh giới (ví dụ, tương ứng với tường) của cảnh nằm dưới ngưỡng xác định trước (ví dụ, khi $d_2 < d_{\text{threshold}}$), hệ thống 102 tiếp tục thu được các luồng âm thanh liên quan đến môi trường thứ hai, liền kề và/hoặc lân cận liên quan đến cảnh thứ hai (150B).

Trường hợp 4

Fig.5a và Fig.5b thể hiện phương án với một bối cảnh mẫu khác (được biểu diễn trong mặt phẳng ngang XY của một không gian XYZ, trong đó trục Z được biểu diễn là hướng ra bề mặt giấy), trong đó người dùng được định vị trong một và cùng cảnh VR, AR và/hoặc MR 150 nhưng tại các thời điểm tức thời ở các khoảng cách khác nhau ví dụ tới hai thành phần âm thanh.

Tại thời điểm tức thời thứ nhất $t=t_1$ được thể hiện trên Fig.5a, người dùng được định vị ví dụ ở vị trí thứ nhất. Ở vị trí thứ nhất này, thành phần âm thanh thứ nhất 1 (152-1) và thành phần âm thanh thứ hai 2 (152-2) được đặt (ví dụ, ảo) ở khoảng cách d_1 và d_2 tương ứng từ người dùng được trang bị MCD. Cả hai khoảng cách d_1 và d_2 có thể lớn hơn trong trường hợp này so với khoảng cách ngưỡng $d_{\text{threshold}}$ xác định, và do đó hệ thống 102 được tạo cấu hình để nhóm cả hai thành phần âm thanh thành một nguồn ảo duy nhất 152-3. Vị trí và các thuộc tính (ví dụ, phạm vi không gian) của nguồn ảo duy nhất này có thể được tính toán dựa trên ví dụ các vị trí của hai nguồn ban đầu sao cho nó bắt chước tốt nhất có thể trường âm thanh ban đầu do hai nguồn tạo ra (ví dụ, hai nguồn điểm được đặt tốt nhất có thể được tái tạo ở giữa khoảng cách

giữa chúng như một nguồn duy nhất). Dữ liệu vị trí người dùng 110 (d_1, d_2) có thể được truyền từ MCD đến hệ thống 102 (máy khách) và sau đó đến máy chủ 120, mà có thể quyết định gửi luồng âm thanh thích hợp 106 sẽ được kết xuất bởi hệ thống máy chủ 120 (trong các phương án khác, đó là máy khách 102 quyết định các luồng nào được truyền từ máy chủ 120). Bằng cách nhóm cả hai thành phần âm thanh vào một nguồn ảo 152-3 duy nhất, máy chủ 120 có thể chọn một trong vô số các sự biểu diễn liên quan đến thông tin đã nói ở trên. (Ví dụ, có thể phân phối tương ứng luồng chuyên dụng 106 tập hợp điều chỉnh 113' được liên kết tương ứng với ví dụ một kênh duy nhất.). Do đó, người dùng có thể nhận được thông qua MCD tín hiệu âm thanh như được truyền từ thành phần âm thanh ảo duy nhất 152-3 được đặt giữa các thành phần âm thanh thực 1 (152-1) và 2 (152-2).

Tại thời điểm tức thời thứ hai $t=t_2$ được thể hiện trên Fig.5b, người dùng được đặt ví dụ trong cùng cảnh 150, có vị trí được xác định thứ hai trong cùng môi trường VR như trên Fig.5a. Ở vị trí thứ hai này, hai thành phần âm thanh 152-1 và 152-2 được đặt (ví dụ, ảo) ở các khoảng cách d_3 và d_4 tương ứng từ người dùng. Cả hai khoảng cách d_3 và d_4 có thể nhỏ hơn khoảng cách ngưỡng $d_{\text{threshold}}$, và do đó việc nhóm các thành phần âm thanh 152-1 và 152-2 thành một nguồn ảo 152-3 duy nhất không được sử dụng nữa. Dữ liệu vị trí người dùng được truyền từ MCD đến hệ thống 102 và sau đó đến máy chủ 120, mà có thể quyết định gửi một luồng âm thanh 106 thích hợp khác được kết xuất bởi máy chủ hệ thống 120 (trong các phương án khác, quyết định này được đưa ra bởi máy khách 102). Bằng cách tránh việc nhóm các thành phần âm thanh, máy chủ 120 có thể chọn sự biểu diễn khác liên quan đến thông tin đã nói ở trên để phân phối luồng 106 chuyên dụng phù hợp với tập hợp điều chỉnh 113' tương ứng liên quan đến các kênh khác nhau cho từng thành phần âm thanh. Do đó, người dùng có thể nhận được thông qua MCD tín hiệu âm thanh 108 như được truyền từ hai thành phần âm thanh 1 (152-1) và 2 (152-2). Do đó, vị trí của người dùng 110 càng gần các nguồn âm thanh 1 (152-1) và 2 (152-2), mức chất lượng cần thiết của luồng liên quan đến các nguồn âm thanh phải được chọn càng cao.

Trên thực tế, các nguồn âm thanh 1 (152-1) và 2 (152-2) được đặt tùy theo vị trí của người dùng càng gần, như được mô tả trên Fig.5b, mức độ phải điều chỉnh càng cao và do đó các tín hiệu âm thanh 108 có thể được kết xuất ở mức chất lượng càng

cao. Ngược lại, các nguồn âm thanh được định vị từ xa 1 và 2 được biểu diễn trên Fig.5b phải được nghe ở mức thấp hơn như được tái tạo bởi nguồn ảo duy nhất, do đó được kết xuất ví dụ ở mức chất lượng thấp hơn.

Trong cấu hình tương tự, vô số các thành phần âm thanh có thể được đặt ở phía trước người dùng, tất cả chúng đều được định vị ở các khoảng cách lớn hơn khoảng cách ngưỡng từ người dùng. Trong một phương án, hai nhóm của năm thành phần âm thanh, mỗi nhóm có thể được kết hợp trong hai nguồn ảo. Dữ liệu vị trí người dùng được truyền từ MCD đến hệ thống 102 và sau đó đến máy chủ 120, có thể quyết định gửi luồng âm thanh 106 phù hợp để được kết xuất bởi máy chủ hệ thống 120. Bằng cách nhóm tất cả 10 thành phần âm thanh vào chỉ hai nguồn ảo duy nhất, máy chủ 120 có thể chọn một trong vô số các sự biểu diễn liên quan đến thông tin đã nói ở trên để phân phối tương ứng luồng 106 chuyên dụng với tập hợp điều chỉnh 113' tương ứng được liên kết với ví dụ hai thành phần âm thanh duy nhất. Do đó, người dùng có thể nhận thông qua MCD tín hiệu âm thanh như được truyền từ hai thành phần âm thanh ảo riêng biệt được đặt trong cùng khu vực định vị với các thành phần âm thanh thực.

Tại thời điểm tức thời tiếp theo, người dùng đang tiếp cận nhiều (mười) thành phần âm thanh. Trong cảnh tiếp theo này, tất cả các thành phần âm thanh được đặt ở các khoảng cách nhỏ hơn khoảng cách ngưỡng $d_{\text{threshold}}$, và do đó, hệ thống 102 được tạo cấu hình để chấm dứt việc nhóm các thành phần âm thanh. Dữ liệu vị trí người dùng mới được truyền từ MCD đến hệ thống 102 và sau đó đến máy chủ 120, mà có thể quyết định gửi một luồng âm thanh thích hợp 106 khác sẽ được kết xuất bởi hệ thống máy chủ 120. Bằng cách không nhóm các thành phần âm thanh, máy chủ 120 có thể chọn sự biểu diễn khác liên quan đến thông tin đã nói ở trên để phân phối luồng 106 chuyên dụng tương ứng với tập hợp điều chỉnh 113' tương ứng liên quan đến các kênh khác nhau cho từng thành phần âm thanh. Do đó, người dùng có thể nhận thông qua MCD tín hiệu âm thanh như được truyền từ mười thành phần âm thanh khác nhau. Do đó, vị trí của người dùng càng gần các nguồn âm thanh thì độ phân giải cần thiết của luồng liên quan đến các nguồn âm thanh phải được chọn càng cao.

Trường hợp 5

Fig.6 (trường hợp 5) mô tả người dùng 140 được đặt ở một vị trí của một cảnh 150 duy nhất đeo thiết bị tiêu dùng truyền thông (MCD) mà có thể được hướng vào ba hướng khác nhau (mỗi hướng liên quan đến một cổng nhìn khác nhau 160-1, 160- 2, 160-3). Các hướng này như được thể hiện trên Fig.6 có thể có hướng (ví dụ, hướng góc) trong hệ tọa độ cực và/hoặc hệ Đề-các XY chỉ vào điểm nhìn thứ nhất 801 được đặt ví dụ ở 180° ở phần dưới cùng của Fig.6, vào điểm nhìn thứ hai 802 được đặt ví dụ 90° ở phía bên phải của Fig.6 và vào điểm nhìn thứ ba 803 được đặt ví dụ ở 0° ở phần trên của Fig.6. Mỗi điểm nhìn này liên quan đến hướng của người dùng 140 đeo thiết bị tiêu dùng phương tiện (MCD), người dùng được định vị ở trung tâm được cung cấp cổng nhìn cụ thể được hiển thị bởi MCD kết xuất tín hiệu âm thanh tương ứng 108 theo hướng của MCD.

Trong môi trường VR cụ thể này, thành phần âm thanh thứ nhất s1 (152) được đặt trong cổng nhìn thứ nhất 160-1, lân cận với điểm nhìn được đặt ví dụ tại 180° và thành phần âm thanh thứ hai s2 (152) được đặt trong cổng nhìn thứ ba 160-3 lân cận với điểm nhìn được đặt ví dụ ở 180° . Trước khi thay đổi hướng của mình, người dùng 140 trải nghiệm theo hướng thứ nhất về phía điểm nhìn 801 (cổng nhìn 160-1) âm thanh liên quan đến vị trí thực (có thật) của người đó to hơn từ thành phần âm thanh s1 so với thành phần âm thanh s2.

Bằng cách thay đổi hướng của mình, người dùng 140 có thể trải nghiệm theo hướng thứ hai về phía điểm nhìn 802, âm thanh liên quan đến vị trí thực 110 của người đó gần như có cùng âm lượng phát ra từ cả hai thành phần âm thanh s1 và s2.

Cuối cùng, bằng cách thay đổi hướng của mình, người dùng 140 có thể trải nghiệm theo hướng thứ ba về phía điểm nhìn 801 (cổng nhìn 160-3), âm thanh liên quan đến thành phần âm thanh 2 to hơn âm thanh liên quan đến thành phần âm thanh s1 (thực tế là âm thanh đến từ thành phần âm thanh 2 đến từ phía trước, trong khi âm thanh đến từ thành phần âm thanh 1 đến từ phía sau).

Do đó, các cổng nhìn và/hoặc các hướng và/hoặc dữ liệu vị trí ảo khác nhau có thể liên quan đến các tốc độ bit và/hoặc chất lượng khác nhau.

Các trường hợp và ví dụ khác

Fig.7 thể hiện phương án của phương pháp nhận các luồng âm thanh bởi hệ thống dưới dạng chuỗi các bước thao tác trong sơ đồ. Tại bất kỳ thời điểm nào người dùng của hệ thống 102 được liên kết với cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại của người đó. Tại thời điểm nhất định, hệ thống có thể xác định trong bước 701 của Fig.7A các thành phần âm thanh sẽ được tái tạo trên cơ sở cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại. Do đó, trong bước tiếp theo 703, mức độ liên quan và mức độ nghe thấy cho từng thành phần âm thanh có thể được xác định. Như được mô tả ở trên trong Fig.6, môi trường VR có thể có các thành phần âm thanh khác nhau nằm trong cảnh cụ thể 150 ở khu vực lân cận của người dùng hoặc xa hơn, nhưng cũng có hướng cụ thể trong 360 độ xung quanh. Tất cả các yếu tố này xác định mức độ liên quan và mức độ nghe thấy cho từng thành phần âm thanh nói trên.

Trong bước tiếp theo 705, hệ thống 102 có thể yêu cầu các luồng âm thanh theo mức độ liên quan và mức độ nghe thấy được xác định cho từng thành phần âm thanh từ máy chủ phương tiện 120.

Trong bước tiếp theo 707, hệ thống 102 có thể nhận các luồng âm thanh 113 được chuẩn bị phù hợp bởi máy chủ phương tiện 120, trong đó các luồng có các tốc độ bit khác nhau có thể phản ánh mức độ liên quan và mức độ nghe thấy được xác định trong các bước nêu trên.

Trong bước tiếp theo 709, hệ thống 102 (ví dụ bộ giải mã âm thanh) có thể giải mã các luồng âm thanh nhận được 113, sao cho ở bước 711, cảnh 150 cụ thể được tái tạo (ví dụ, bởi MCD), theo cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại.

Fig.7B mô tả sự tương tác giữa máy chủ phương tiện 120 và hệ thống 102 theo trình tự được mô tả ở trên của sơ đồ hoạt động. Tại thời điểm nhất định, máy chủ phương tiện có thể truyền luồng âm thanh 750 ở tốc độ bit thấp hơn, theo như đã xác định ở trên mức độ liên quan và mức độ nghe thấy của các thành phần âm thanh có liên quan của cảnh 150 đã nói ở trên. Hệ thống có thể xác định tại thời điểm tiếp theo 752 rằng xảy ra sự tương tác hoặc thay đổi dữ liệu vị trí. Sự tương tác như vậy có thể

dẫn đến hoặc ví dụ từ thay đổi trong dữ liệu vị trí trong cùng một cảnh 150 hoặc ví dụ kích hoạt tay nắm cửa trong khi người dùng cố gắng vào cảnh thứ hai tách biệt với cảnh thứ nhất bằng cửa được cung cấp bởi tay nắm cửa.

Việc thay đổi cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại có thể dẫn đến yêu cầu 754 được hệ thống 102 gửi đến máy chủ phương tiện 120. Yêu cầu này có thể phản ánh mức độ liên quan và mức độ nghe được cao hơn của các thành phần âm thanh có liên quan được xác định cho cảnh 150 tiếp theo. Để đáp ứng yêu cầu 754, máy chủ phương tiện có thể truyền luồng 756 ở tốc độ bit cao hơn, cho phép tái tạo cảnh 150 chân thực và hợp lý bởi hệ thống 102 tại bất kỳ vị trí ảo nào của người dùng hiện tại.

Fig.8 thể hiện một phương án khác của phương pháp nhận các luồng âm thanh bởi hệ thống cũng dưới dạng chuỗi các bước thao tác trong sơ đồ. Tại một thời điểm 801 nhất định, việc xác định cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại thứ nhất có thể được thực hiện. Bằng cách trừ đi trường hợp khẳng định, yêu cầu của các luồng liên quan đến vị trí thứ nhất được xác định bởi tốc độ bit thấp có thể được chuẩn bị và truyền bởi hệ thống 102 ở bước 803.

Bước xác định 805 có ba kết quả khác nhau có thể được thực hiện tại thời điểm tiếp theo. Một hoặc hai ngưỡng được xác định có thể có liên quan ở bước này để xác định ví dụ quyết định dự đoán liên quan đến cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo tiếp theo. Do đó, việc so sánh với ngưỡng thứ nhất và/hoặc ngưỡng thứ hai có thể được thực hiện, liên quan đến xác suất thay đổi thành vị trí thứ hai, dẫn đến ví dụ ba bước tiếp theo khác nhau sẽ được thực hiện.

Tại kết quả phản ánh, ví dụ, xác suất rất thấp (ví dụ, liên quan đến việc so sánh ở trên với ngưỡng được xác định trước thứ nhất), bước so sánh mới 801 sẽ được thực hiện.

Tại kết quả phản ánh xác suất thấp (ví dụ, cao hơn ngưỡng được xác định trước thứ nhất, nhưng ví dụ, thấp hơn ngưỡng được xác định trước thứ hai cao hơn ngưỡng

thứ nhất) có thể dẫn đến yêu cầu ở bước 809 cho các luồng âm thanh 113 ở tốc độ bit thấp.

Tại kết quả phản ánh xác suất cao (ví dụ, cao hơn ngưỡng được xác định trước thứ hai), ở bước 807, yêu cầu đối với các luồng âm thanh 113 ở tốc độ bit cao có thể được thực hiện. Do đó, bước tiếp theo sẽ được thực hiện sau khi chạy các bước 807 hoặc 809 có thể lại là bước xác định 801.

Fig.8B mô tả sự tương tác giữa máy chủ phương tiện 120 và hệ thống 102 chỉ theo một trong số trình tự được mô tả ở trên của sơ đồ hoạt động. Tại thời điểm nhất định, máy chủ phương tiện có thể truyền luồng âm thanh 850 ở tốc độ bit thấp, theo như đã xác định ở trên mức độ liên quan và mức độ nghe thấy thấp của các thành phần âm thanh của cảnh 150 đã nói ở trên. Hệ thống có thể xác định tại thời điểm tiếp theo 852 rằng sự tương tác dự báo sẽ xảy ra. Việc thay đổi dự báo của cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo hiện tại có thể dẫn đến yêu cầu 854 được hệ thống 102 gửi đến máy chủ phương tiện 120. Yêu cầu này có thể phản ánh một trong những trường hợp được mô tả ở trên liên quan đến xác suất cao đạt được vị trí thứ hai liên quan đến tốc độ bit cao theo mức độ nghe thấy của các thành phần âm thanh như được yêu cầu đối với cảnh 150 sau đó tương ứng. Để phản hồi lại, máy chủ phương tiện có thể truyền luồng 856 ở tốc độ bit cao hơn, cho phép tái tạo cảnh 150 chân thực và hợp lý bởi hệ thống 102 tại bất kỳ vị trí ảo nào của người dùng hiện tại.

Hệ thống 102 như được thể hiện trên Fig.1.3 được tạo cấu hình để nhận các luồng âm thanh 113 trên cơ sở một cấu hình khác ở phía máy khách, trong đó cấu trúc hệ thống có thể sử dụng các điểm nhìn riêng biệt dựa trên giải pháp sử dụng nhiều bộ giải mã âm thanh 1320, 1322. Ở phía máy khách, hệ thống 102 có thể bao gồm, ví dụ, các bộ phận của hệ thống được mô tả trên Fig.1.2 mà ngoài ra hoặc thay vào đó bao gồm nhiều bộ giải mã âm thanh 1320, 1322 có thể được tạo cấu hình để giải mã các luồng âm thanh riêng lẻ như được biểu thị bởi bộ xử lý siêu dữ liệu 1236 ví dụ với một số thành phần âm thanh bị vô hiệu hóa.

Bộ trộn/kết xuất 1238 có thể được đề xuất trong hệ thống 102 được tạo cấu hình để tái tạo cảnh âm thanh cuối cùng dựa trên thông tin về vị trí và/hoặc hướng và/hoặc

hướng chuyển động của người dùng, tức là, ví dụ, một số thành phần âm thanh không thể nghe thấy tại vị trí cụ thể đó nên bị vô hiệu hóa hoặc không được kết xuất.

Các phương án sau đây được thể hiện trên các hình vẽ Fig.1.4, Fig.1.5 và Fig.1.6 dựa trên các tập hợp điều chỉnh độc lập cho các điểm nhìn riêng biệt với các tập hợp điều chỉnh linh hoạt. Trong trường hợp khi người dùng di chuyển trong môi trường VR, cảnh âm thanh có thể thay đổi liên tục. Để đảm bảo trải nghiệm âm thanh tốt, tất cả các thành phần âm thanh soạn cảnh âm thanh tại thời điểm nhất định, có thể phải được cung cấp cho bộ giải mã phương tiện có thể sử dụng thông tin vị trí để tạo cảnh âm thanh cuối cùng.

Nếu nội dung được mã hóa trước, đối với một số vị trí được xác định trước, hệ thống có thể cung cấp phép tái tạo chính xác các cảnh âm thanh tại các vị trí cụ thể này với giả định rằng các cảnh âm thanh này không chồng lấp nhau và người dùng có thể "nhảy/chuyển" từ một vị trí đến một vị trí tiếp theo.

Nhưng trong các trường hợp khi người dùng "đi" từ một vị trí tới vị trí tiếp theo, các thành phần âm thanh từ hai (hoặc nhiều) cảnh âm thanh có thể được nghe thấy cùng một lúc. Giải pháp cho các trường hợp sử dụng này đã được đề xuất trong các ví dụ của các hệ thống trước đó, trong đó sự độc lập của các cơ chế được cung cấp để giải mã nhiều luồng âm thanh (sử dụng bộ ghép với một bộ giải mã phương tiện duy nhất hoặc nhiều bộ giải mã phương tiện với bộ trộn/bộ kết xuất bổ sung), các luồng âm thanh mô tả các cảnh âm thanh hoàn chỉnh phải được cung cấp cho máy khách.

Sự tối ưu hóa được cung cấp dưới đây bằng cách đưa ra khái niệm các thành phần âm thanh chung giữa nhiều luồng âm thanh.

Thảo luận về các khía cạnh và các ví dụ

Giải pháp 1: Các tập hợp điều chỉnh độc lập cho các vị trí riêng biệt (các điểm nhìn).

Một cách để giải quyết vấn đề được mô tả là sử dụng tập hợp điều chỉnh độc lập hoàn toàn cho từng vị trí. Để hiểu rõ hơn về giải pháp, Fig.1.1 được sử dụng làm bối cảnh ví dụ. Trong ví dụ này, ba điểm nhìn riêng biệt khác nhau (bao gồm ba cảnh âm

thanh khác nhau) được sử dụng để tạo môi trường VR hoàn chỉnh, mà người dùng sẽ có thể di chuyển vào. Do đó:

- Một vài cảnh âm thanh độc lập hoặc chồng chéo được mã hóa trong một số luồng âm thanh. Đối với mỗi cảnh âm thanh, một luồng chính có thể được sử dụng hoặc tùy thuộc vào trường hợp sử dụng, một luồng chính và các luồng phụ bổ sung (ví dụ, một số đối tượng âm thanh chứa các ngôn ngữ khác nhau có thể được mã hóa trong các luồng độc lập để phân phối hiệu quả). Trong ví dụ được cung cấp, cảnh âm thanh A được mã hóa thành hai luồng (A1 và A2), cảnh âm thanh B được mã hóa thành ba luồng (B1, B2 và B3) trong khi cảnh âm thanh C được mã hóa thành ba luồng (C1, C2 và C3). Cần lưu ý rằng cảnh âm thanh A và cảnh âm thanh B dùng chung một số thành phần chung (trong ví dụ này là hai đối tượng âm thanh). Vì mọi cảnh phải hoàn chỉnh và độc lập (để tái tạo độc lập, ví dụ như trên các thiết bị phát lại không VR), các thành phần chung phải được mã hóa hai lần cho mỗi cảnh.

- Tất cả các luồng âm thanh được mã hóa ở các tốc độ bit khác nhau (nghĩa là các sự biểu diễn khác nhau), cho phép điều chỉnh tốc độ bit hiệu quả tùy thuộc vào kết nối mạng (nghĩa là đối với những người dùng sử dụng kết nối tốc độ cao, phiên bản được mã hóa tốc độ bit cao được phân phối trong khi đối với những người dùng có kết nối mạng tốc độ thấp hơn, phiên bản tốc độ bit thấp hơn được phân phối).

- Các luồng âm thanh được lưu trữ trên máy chủ phương tiện, trong đó đối với mỗi luồng âm thanh, các sự mã hóa khác nhau ở các tốc độ bit khác nhau (nghĩa là, các sự biểu diễn khác nhau) được nhóm vào một tập hợp điều chỉnh với dữ liệu phù hợp báo hiệu sự sẵn có của tất cả các tập hợp điều chỉnh đã tạo.

- Ngoài ra, với các tập hợp điều chỉnh, máy chủ phương tiện nhận thông tin về vị trí "ranh giới" của từng cảnh âm thanh và mối quan hệ của nó với từng tập hợp điều chỉnh (có thể chứa ví dụ, cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng riêng lẻ). Theo cách này, mỗi tập hợp điều chỉnh có thể được liên kết với một trong số cảnh âm thanh có sẵn. Ranh giới của một cảnh âm thanh có thể được xác định, ví dụ, dưới dạng các tọa độ hình học của hình cầu (ví dụ, tâm và bán kính).

- Mỗi tập hợp điều chỉnh cũng chứa thông tin mô tả về các vị trí trong đó cảnh âm thanh hoặc các thành phần âm thanh đang hoạt động. Ví dụ, nếu một luồng phụ

chứa một hoặc một vài đối tượng, tập hợp điều chỉnh có thể chứa thông tin chẳng hạn như các vị trí mà các đối tượng có thể nghe được (ví dụ, các tọa độ của tâm của hình cầu và bán kính).

- Máy chủ phương tiện cung cấp thông tin về vị trí "ranh giới" được liên kết với từng tập hợp điều chỉnh cho máy khách, ví dụ máy khách DASH. Ví dụ, điều này có thể được lồng vào cú pháp mô tả biểu diễn phương tiện (MPD) XML trong trường hợp môi trường phân phối DASH.

- Máy khách nhận thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng).

- Máy khách nhận thông tin về từng tập hợp điều chỉnh, và dựa trên điều này và vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng, ví dụ bao gồm các tọa độ x, y, z và/hoặc các giá trị xoay, tiến-lùi, nghiêng), máy khách chọn một hoặc nhiều tập hợp điều chỉnh mô tả hoàn chỉnh cảnh âm thanh cần được tái tạo tại vị trí hiện tại của người dùng.

- Máy khách yêu cầu một hoặc nhiều tập hợp điều chỉnh

- Ngoài ra, máy khách có thể chọn nhiều tập hợp điều chỉnh hơn mô tả hoàn chỉnh nhiều hơn một cảnh âm thanh, và sử dụng các luồng âm thanh tương ứng với nhiều hơn một cảnh âm thanh để tạo cảnh âm thanh mới cần được tái tạo tại vị trí hiện tại của người dùng. Ví dụ, nếu người dùng đi bộ trong môi trường VR, và tại một thời điểm được đặt ở giữa (hoặc tại một vị trí nằm ở nơi có hai cảnh âm thanh có hiệu ứng có thể nghe được).

- Khi các luồng âm thanh khả dụng, nhiều bộ giải mã phương tiện có thể được sử dụng để giải mã các luồng âm thanh riêng biệt và bộ trộn/kết xuất 1238 bổ sung để tái tạo cảnh âm thanh cuối cùng dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng chuyển động của người dùng (tức là, ví dụ, một số thành phần âm thanh không thể nghe thấy tại vị trí cụ thể đó nên bị vô hiệu hóa hoặc không được kết xuất).

○ Ngoài ra, bộ xử lý siêu dữ liệu 1236 có thể được sử dụng để thao tác siêu dữ liệu âm thanh được liên kết với tất cả các luồng âm thanh, dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng, để:

- Chọn/kích hoạt các thành phần âm thanh 152 cần thiết để soạn cảnh âm thanh mới;

- Và để cho phép gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.

- Máy chủ phương tiện phân phối các tập hợp điều chỉnh cần thiết
- Ngoài ra, máy khách cung cấp thông tin về định vị của người dùng cho máy chủ phương tiện và máy chủ phương tiện cung cấp chỉ báo về tập hợp điều chỉnh cần thiết.

Fig.1.2 thể hiện một ví dụ khác về việc triển khai hệ thống như vậy bao gồm:

- ở phía mã hóa

○ nhiều bộ mã hóa phương tiện có thể được sử dụng để tạo một hoặc nhiều luồng âm thanh cho mỗi cảnh âm thanh có sẵn liên quan đến một phần cảnh âm thanh của một điểm nhìn

○ nhiều bộ mã hóa phương tiện có thể được sử dụng để tạo một hoặc nhiều luồng âm thanh cho mỗi cảnh video có sẵn liên quan đến một phần cảnh video của một điểm nhìn. Các bộ mã hóa video không được biểu diễn trong các hình vẽ để đơn giản.

○ Máy chủ phương tiện mà lưu trữ nhiều tập hợp điều chỉnh âm thanh và video bao gồm các sự mã hóa khác (tức là các sự biểu diễn khác nhau). Ngoài ra, máy chủ phương tiện chứa thông tin mô tả của tất cả các tập hợp điều chỉnh, có thể chứa

- tính khả dụng của tất cả các tập hợp điều chỉnh đã tạo;

- thông tin mô tả sự liên kết của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn; Theo cách này, mỗi tập hợp điều chỉnh có thể được liên kết với một trong các cảnh âm thanh có sẵn;

- thông tin mô tả "ranh giới" của từng cảnh âm thanh và/hoặc điểm nhìn (mà có thể chứa ví dụ cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng âm thanh riêng lẻ). Ranh giới của một cảnh âm thanh có thể được xác định, ví dụ, dưới dạng các tọa độ hình học của hình cầu (ví dụ, tâm và bán kính).

- ở phía máy khách hệ thống (hệ thống máy khách) có thể bao gồm bất kỳ trong số:

○ đầu nhận, mà có thể nhận:

- thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng)

- thông tin về tính khả dụng của tất cả các tập hợp điều chỉnh và thông tin mô tả mối liên hệ của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn; và/hoặc thông tin mô tả "ranh giới" của từng cảnh âm thanh và/hoặc điểm nhìn (mà có thể chứa ví dụ, cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng riêng lẻ). Ví dụ, thông tin như vậy có thể được cung cấp như một phần của cú pháp mô tả biểu diễn phương tiện (MPD) XML trong trường hợp môi trường phân phối DASH.

○ phía thiết bị tiêu dùng phương tiện được sử dụng để tiêu dùng nội dung (ví dụ dựa trên HMD). Thiết bị tiêu dùng phương tiện cũng chịu trách nhiệm thu thập thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi hành động của người dùng)

○ bộ xử lý công nhìn 1232 mà có thể được tạo cấu hình để

- nhận thông tin về công nhìn hiện tại mà có thể chứa vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi hành động của người dùng) từ phía thiết bị tiêu dùng phương tiện.

- nhận thông tin về ROI được báo hiệu trong siêu dữ liệu (các công nhìn video được báo hiệu như trong thông số kỹ thuật OMAF).

- nhận tất cả thông tin có sẵn ở đầu nhận;

- quyết định dựa trên tất cả thông tin nhận được và/hoặc được suy ra từ siêu dữ liệu đã nhận và/hoặc có sẵn, điểm nhìn âm thanh/video nào sẽ được tái tạo tại một thời điểm nhất định. Ví dụ, bộ xử lý công nhìn 1232 có thể quyết định rằng:

+ một cảnh âm thanh hoàn chỉnh sẽ được tái tạo

+ một cảnh âm thanh mới sẽ được tái tạo trong số tất cả các cảnh âm thanh có sẵn (ví dụ, chỉ một số thành phần âm thanh từ nhiều cảnh

âm thanh sẽ được tái tạo, trong khi các thành phần âm thanh còn lại khác của các cảnh âm thanh này sẽ không được tái tạo).

+ việc chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh phải được tái tạo

○ phần lựa chọn 1230 được tạo cấu hình để lựa chọn, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232, một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được từ đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh cảnh âm thanh cần được tái tạo tại vị trí hiện tại của người dùng. Cảnh âm thanh này có thể là một cảnh âm thanh hoàn chỉnh như được xác định ở phía mã hóa hoặc cảnh âm thanh mới phải được tạo ra trong tất cả các cảnh âm thanh có sẵn.

- Ngoài ra, trong trường hợp khi sự chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh sắp xảy ra dựa trên chỉ báo của bộ xử lý công nhìn 1232, phần lựa chọn có thể được tạo cấu hình để chọn một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được bởi đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh cảnh âm thanh mà có thể được yêu cầu tái tạo trong tương lai gần (ví dụ, nếu người dùng đi theo hướng của cảnh âm thanh tiếp theo với tốc độ nhất định, có thể dự đoán rằng cảnh âm thanh tiếp theo sẽ được yêu cầu và nó được chọn trước khi tái tạo).

- Ngoài ra, một số tập hợp điều chỉnh tương ứng với các vị trí lân cận có thể được chọn trước tiên ở tốc độ bit thấp hơn (tức là, sự biểu diễn được mã hóa ở tốc độ bit thấp hơn được chọn trong số các sự biểu diễn có sẵn trong một tập hợp điều chỉnh), và dựa trên các thay đổi vị trí, chất lượng được tăng lên bằng cách chọn tốc độ bit cao hơn cho các tập hợp điều chỉnh cụ thể đó (tức là, sự biểu diễn được mã hóa ở tốc độ bit cao hơn được chọn ra trong số các sự biểu diễn có sẵn trong một tập hợp điều chỉnh).

○ phần tải xuống và chuyển đổi có thể được tạo cấu hình thành:

- yêu cầu, dựa trên chỉ báo nhận được từ phần lựa chọn 1230, một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện 120;

- nhận, một hoặc nhiều tập hợp điều chỉnh (nghĩa là, một sự biểu diễn trong số tất cả các sự biểu diễn có sẵn trong mỗi tập hợp điều chỉnh) trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện 120;

- trích xuất thông tin siêu dữ liệu từ tất cả các luồng âm thanh nhận được

- bộ xử lý siêu dữ liệu 1236 mà có thể được tạo cấu hình để:

- nhận từ thông tin tải xuống và chuyển đổi về các luồng âm thanh nhận được, thông tin có thể chứa siêu dữ liệu âm thanh tương ứng với từng luồng âm thanh nhận được

- xử lý và thao tác siêu dữ liệu âm thanh liên quan đến từng luồng âm thanh, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232 có thể chứa thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng, để:

- chọn/kích hoạt các thành phần âm thanh 152 cần thiết soạn cảnh âm thanh mới như được chỉ định bởi bộ xử lý công nhìn 1232;

- cho phép gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.

- bộ ghép/gộp luồng 1238 có thể được tạo cấu hình để gộp tất cả các luồng âm thanh đã chọn vào một luồng âm thanh dựa trên thông tin nhận được từ bộ xử lý siêu dữ liệu 1236 mà có thể chứa siêu dữ liệu âm thanh được biến đổi và được xử lý tương ứng với tất cả các luồng âm thanh nhận được.

- bộ giải mã phương tiện được tạo cấu hình để nhận và giải mã ít nhất một luồng âm thanh để tái tạo cảnh âm thanh mới như được chỉ định bởi bộ xử lý công nhìn 1232 dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng.

Fig.1.3 thể hiện hệ thống bao gồm, ở phía máy khách, hệ thống (hệ thống máy khách) mà có thể bao gồm, ví dụ, các bộ phận của hệ thống được mô tả trên Fig.1.2 ngoài ra hoặc thay vào đó bao gồm:

- nhiều bộ giải mã phương tiện có thể được tạo cấu hình để giải mã các luồng âm thanh riêng lẻ như được chỉ định bởi bộ xử lý siêu dữ liệu 1236 (ví dụ với một số thành phần âm thanh bị vô hiệu hóa).

- bộ trộn/kết xuất 1238 có thể được tạo cấu hình để tái tạo cảnh âm thanh cuối cùng dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng chuyển động của người

dùng (tức là, ví dụ, một số thành phần âm thanh không thể nghe thấy tại vị trí cụ thể đó nên bị vô hiệu hóa hoặc không được kết xuất).

Giải pháp 2:

Các hình vẽ Fig.1.4, Fig.1.5 và Fig.1.6 đề cập đến các ví dụ theo giải pháp 2 của sáng chế (có thể là các phương án của các ví dụ theo Fig.1.1 và/hoặc Fig.1.2 và/hoặc Fig.1.3): Các tập hợp điều chỉnh độc lập cho các vị trí riêng biệt (điểm nhìn) với các tập hợp điều chỉnh linh hoạt.

Trong trường hợp khi người dùng di chuyển trong môi trường VR, cảnh âm thanh 150 có thể thay đổi liên tục. Để đảm bảo trải nghiệm âm thanh tốt, tất cả các thành phần âm thanh 152 soạn cảnh âm thanh 150 tại thời điểm nhất định, có thể phải có sẵn cho bộ giải mã phương tiện có thể sử dụng thông tin vị trí để tạo cảnh âm thanh cuối cùng.

Nếu nội dung được mã hóa trước, đối với một số vị trí được xác định trước, hệ thống có thể cung cấp phép tái tạo chính xác các cảnh âm thanh tại các vị trí cụ thể này với giả định rằng các cảnh âm thanh này không chồng lấp và người dùng có thể "nhảy/chuyển" từ một vị trí đến một vị trí tiếp theo.

Nhưng trong các trường hợp khi người dùng "đi" từ một vị trí tới vị trí tiếp theo, các thành phần âm thanh 152 từ hai (hoặc nhiều) cảnh âm thanh 150 có thể được nghe thấy cùng một lúc. Giải pháp cho các trường hợp sử dụng này đã được đề xuất trong các ví dụ của các hệ thống trước đó, trong đó sự độc lập của các cơ chế được cung cấp để giải mã nhiều luồng âm thanh (sử dụng bộ ghép với một bộ giải mã phương tiện duy nhất hoặc nhiều bộ giải mã phương tiện với bộ trộn/bộ kết xuất 1238 bổ sung), các luồng âm thanh mô tả các cảnh âm thanh hoàn chỉnh phải được cung cấp cho máy khách/hệ thống 102.

Sự tối ưu hóa được cung cấp dưới đây bằng cách đưa ra khái niệm các thành phần âm thanh 152 chung giữa nhiều luồng âm thanh.

Fig.1.4 thể hiện ví dụ trong đó các cảnh âm thanh khác nhau dùng chung ít nhất một thành phần âm thanh (đối tượng âm thanh, nguồn âm thanh, ...). Do đó, máy khách 102 có thể nhận, ví dụ, một luồng chính 106A chỉ được liên kết với một cảnh A

(ví dụ, được liên kết với môi trường mà người dùng hiện tại đang ở) và được liên kết với các đối tượng 152A, và một luồng phụ 106B được chia sẻ bởi cảnh B khác (ví dụ, luồng ở ranh giới giữa cảnh A mà người dùng hiện tại đang ở và luồng B lân cận hoặc liền kề có chung các đối tượng 152B) và được liên kết với các đối tượng 152B.

Hơn nữa, như được thể hiện trên Fig.1.4:

- Một vài cảnh âm thanh độc lập hoặc chồng chéo được mã hóa trong một số luồng âm thanh. Các luồng âm thanh 106 được tạo theo cách mà:

- đối với mỗi cảnh âm thanh 150, một luồng chính có thể được tạo bằng cách chỉ chứa các thành phần âm thanh 152 là một phần của cảnh âm thanh tương ứng, nhưng không phải là một phần của bất kỳ cảnh âm thanh nào khác; và/hoặc

- đối với tất cả các cảnh âm thanh 150 có chung thành phần âm thanh 152, thành phần âm thanh chung 152 chỉ có thể được mã hóa trong các luồng âm thanh phụ chỉ được liên kết với một trong các cảnh âm thanh và thông tin siêu dữ liệu phù hợp biểu thị mối liên hệ với các cảnh âm thanh khác được tạo. Hoặc được nêu khác nhau về siêu dữ liệu bổ sung biểu thị khả năng là một số luồng âm thanh có thể được sử dụng cùng với nhiều cảnh âm thanh; và/hoặc

- tùy thuộc vào trường hợp sử dụng, các luồng phụ bổ sung có thể được tạo (ví dụ, một số đối tượng âm thanh có chứa các ngôn ngữ khác nhau có thể được mã hóa trong các luồng độc lập để phân phối hiệu quả).

- Trong phương án được đề xuất:

- Cảnh âm thanh A được mã hóa thành:

- + luồng âm thanh chính (A1, 106A),

- + luồng âm thanh phụ (A2, 106B)

- + thông tin siêu dữ liệu có thể chỉ ra rằng một số thành phần âm thanh 152B từ cảnh âm thanh A không được mã hóa trong luồng âm thanh A này, mà được mã hóa trong luồng phụ A2 (106B) thuộc về cảnh âm thanh khác nhau (cảnh âm thanh B)

- Cảnh âm thanh B được mã hóa thành:

- + luồng âm thanh chính (B1, 106C),

- + luồng âm thanh phụ (B2),

- + luồng âm thanh phụ (B3),

+ thông tin siêu dữ liệu có thể chỉ ra rằng các thành phần âm thanh 152B từ luồng âm thanh B2 là thành phần âm thanh chung 152B cũng thuộc về cảnh âm thanh A.

- Cảnh âm thanh C được mã hóa thành ba luồng (C1, C2 và C3).

- Tất cả các luồng âm thanh 106 (106A, 106B, 106C...) có thể được mã hóa ở các tốc độ bit khác nhau (nghĩa là các sự biểu diễn khác nhau), cho phép điều chỉnh tốc độ bit hiệu quả, ví dụ, tùy thuộc vào kết nối mạng (nghĩa là đối với những người dùng sử dụng kết nối tốc độ cao, phiên bản được mã hóa tốc độ bit cao được phân phối trong khi đối với những người dùng có kết nối mạng tốc độ thấp hơn, phiên bản tốc độ bit thấp hơn được phân phối).
- Các luồng âm thanh 106 được lưu trữ trên máy chủ phương tiện 120, trong đó đối với mỗi luồng âm thanh, các sự mã hóa khác nhau ở các tốc độ bit khác nhau (nghĩa là, các sự biểu diễn khác nhau) được nhóm vào một tập hợp điều chỉnh với dữ liệu phù hợp báo hiệu sự sẵn có của tất cả các tập hợp điều chỉnh đã tạo. (Nhiều sự biểu diễn của các luồng được liên kết với cùng các tín hiệu âm thanh, nhưng ở các tốc độ bit và/hoặc chất lượng và/hoặc độ phân giải khác nhau có thể có trong cùng tập hợp điều chỉnh.)
- Ngoài ra, với các tập hợp điều chỉnh, máy chủ phương tiện 120 nhận thông tin về vị trí "ranh giới" của từng cảnh âm thanh và mối quan hệ của nó với từng tập hợp điều chỉnh (có thể chứa ví dụ, cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng riêng lẻ). Theo cách này, mỗi tập hợp điều chỉnh có thể được liên kết với một hoặc nhiều cảnh âm thanh 150 có sẵn. Ranh giới của một cảnh âm thanh có thể được xác định, ví dụ, dưới dạng các tọa độ hình học của hình cầu (ví dụ, tâm và bán kính).

○ Mỗi tập hợp điều chỉnh cũng chứa thông tin mô tả về các vị trí trong đó cảnh âm thanh hoặc các thành phần âm thanh 152 đang hoạt động. Ví dụ, nếu một luồng phụ (ví dụ, A2, 106B) chứa một hoặc một vài đối tượng, tập hợp điều chỉnh có thể chứa thông tin chẳng hạn như các vị trí mà các đối tượng có thể nghe được (ví dụ, các tọa độ của tâm của hình cầu và bán kính).

○ Thêm vào đó hoặc ngoài ra, mỗi tập hợp điều chỉnh (ví dụ, tập hợp điều chỉnh được liên kết với cảnh B) có thể chứa thông tin mô tả (ví dụ, siêu dữ liệu) có thể chỉ ra rằng các thành phần âm thanh (ví dụ, 152B) từ một cảnh âm thanh (ví dụ, B) (cũng

hoặc ngoài ra) được mã hóa trong các luồng âm thanh (ví dụ, 106B) thuộc về cảnh âm thanh khác (ví dụ, A).

- Máy chủ phương tiện 120 cung cấp thông tin về vị trí "ranh giới" được liên kết với từng tập hợp điều chỉnh cho hệ thống 102 (máy khách), ví dụ máy khách DASH. Ví dụ, điều này có thể được lồng vào cú pháp mô tả biểu diễn phương tiện (MPD) XML trong trường hợp môi trường phân phối DASH.

- Hệ thống 102 (máy khách) nhận thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng).

- Hệ thống 102 (máy khách) nhận thông tin về từng tập hợp điều chỉnh, và dựa trên điều này và vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng, ví dụ bao gồm các tọa độ x, y, z và/hoặc các giá trị xoay, tiến-lùi, nghiêng), hệ thống 102 (máy khách) chọn một hoặc nhiều tập hợp điều chỉnh mô tả hoàn chỉnh cảnh âm thanh 150 cần được tái tạo tại vị trí hiện tại của người dùng 140.

- Hệ thống 102 (máy khách) có thể yêu cầu một hoặc nhiều tập hợp điều chỉnh:

- Ngoài ra, hệ thống 102 (máy khách) có thể chọn một hoặc nhiều tập hợp điều chỉnh mô tả hoàn chỉnh hoặc một phần nhiều hơn một cảnh âm thanh 150, và sử dụng các luồng âm thanh 106 tương ứng với nhiều hơn một cảnh âm thanh 150 để tạo cảnh âm thanh 150 mới sẽ được tái tạo tại vị trí hiện tại của người dùng 140.

- Dựa trên siêu dữ liệu chỉ ra rằng thành phần âm thanh 152 là một phần của nhiều cảnh âm thanh 150, các thành phần âm thanh chung 152 chỉ có thể được yêu cầu một lần để tạo cảnh âm thanh mới, thay vì yêu cầu chúng hai lần, một lần cho mỗi cảnh âm thanh hoàn chỉnh.

- Khi các luồng âm thanh khả dụng cho hệ thống 102, ví dụ, một hoặc nhiều bộ giải mã phương tiện (104) có thể được sử dụng để giải mã các luồng âm thanh riêng biệt và/hoặc bộ trộn/kết xuất bổ sung để tái tạo cảnh âm thanh cuối cùng dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng chuyển động của người dùng (tức là, ví dụ, một số thành phần âm thanh không thể nghe thấy tại vị trí cụ thể đó nên bị vô hiệu hóa hoặc không được kết xuất).

○ Thay vào đó hoặc ngoài ra, bộ xử lý siêu dữ liệu có thể được sử dụng để thao tác siêu dữ liệu âm thanh được liên kết với tất cả các luồng âm thanh, dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng, để:

- Chọn/kích hoạt các thành phần âm thanh 152 (152A-152c) cần thiết để soạn cảnh âm thanh mới; và/hoặc
 - Và để cho phép gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.
- Máy chủ phương tiện 120 có thể phân phối các tập hợp điều chỉnh cần thiết
 - Ngoài ra, hệ thống 102 (máy khách) cung cấp thông tin về định vị của người dùng 140 cho máy chủ phương tiện 120 và máy chủ phương tiện cung cấp chỉ báo về các tập hợp điều chỉnh cần thiết.

Fig.1.5 thể hiện một triển khai ví dụ khác về hệ thống như vậy bao gồm:

- ở phía mã hóa
 - nhiều bộ mã hóa phương tiện 154 có thể được sử dụng để tạo một hoặc nhiều luồng âm thanh 106 lồng các thành phần âm thanh 152 từ một hoặc nhiều cảnh âm thanh 150 có sẵn liên quan đến một phần cảnh âm thanh của một điểm nhìn.
 - đối với mỗi cảnh âm thanh 150, một luồng chính có thể được tạo bằng cách chỉ chứa các thành phần âm thanh 152 là một phần của cảnh âm thanh 150 tương ứng, nhưng không phải là một phần của bất kỳ cảnh âm thanh nào khác
 - các luồng phụ bổ sung có thể được tạo cho cùng cảnh âm thanh (ví dụ, một số đối tượng âm thanh có chứa các ngôn ngữ khác nhau có thể được mã hóa trong các luồng độc lập để phân phối hiệu quả).
 - các luồng phụ bổ sung có thể được tạo chứa:
 - + các thành phần âm thanh 152 chung cho nhiều hơn một cảnh âm thanh 150
 - + thông tin siêu dữ liệu cho biết sự liên kết của luồng phụ này với tất cả các cảnh âm thanh 150 khác có chung các thành phần âm thanh 152 chung. Hoặc được nêu khác nhau về siêu dữ liệu biểu thị khả năng là một số luồng âm thanh có thể được sử dụng cùng với nhiều cảnh âm thanh.

○ nhiều bộ mã hóa phương tiện có thể được sử dụng để tạo một hoặc nhiều luồng âm thanh cho mỗi cảnh video có sẵn liên quan đến một phần cảnh video của một điểm nhìn. Các bộ mã hóa video không được biểu diễn trong các hình vẽ để đơn giản.

○ máy chủ phương tiện 120 mà lưu trữ nhiều tập hợp điều chỉnh âm thanh và video bao gồm các sự mã hóa khác nhau của cùng các luồng âm thanh và video tại các tốc độ bit khác nhau (tức là các sự biểu diễn khác nhau). Ngoài ra, máy chủ phương tiện 120 chứa thông tin mô tả của tất cả các tập hợp điều chỉnh, có thể chứa

- tính khả dụng của tất cả các tập hợp điều chỉnh đã tạo;

- thông tin mô tả sự liên kết của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn; Theo cách này, mỗi tập hợp điều chỉnh có thể được liên kết với một trong các cảnh âm thanh có sẵn;

- thông tin mô tả "ranh giới" của từng cảnh âm thanh và/hoặc điểm nhìn (mà có thể chứa ví dụ cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng âm thanh riêng lẻ). Ranh giới của một cảnh âm thanh có thể được xác định, ví dụ, dưới dạng các tọa độ hình học của hình cầu (ví dụ, tâm và bán kính).

- thông tin chỉ ra sự liên kết của một tập hợp điều chỉnh với nhiều hơn một cảnh âm thanh, mà có chung ít nhất một thành phần âm thanh chung.

- ở phía máy khách hệ thống (hệ thống máy khách) có thể bao gồm bất kỳ trong số:

- đầu nhận, mà có thể nhận:

- thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin nào đặc trưng cho các thay đổi được kích hoạt bởi hành động của người dùng)

- thông tin về tính khả dụng của tất cả các tập hợp điều chỉnh và thông tin mô tả mối liên hệ của một tập hợp điều chỉnh với một cảnh âm thanh và/hoặc điểm nhìn; và/hoặc thông tin mô tả "ranh giới" của từng cảnh âm thanh và/hoặc điểm nhìn (mà có thể chứa ví dụ, cảnh âm thanh hoàn chỉnh hoặc chỉ các đối tượng riêng lẻ). Ví dụ, thông tin như vậy có thể được cung cấp như một phần của cú pháp mô tả biểu diễn phương tiện (MPD) XML trong trường hợp môi trường phân phối DASH.

- thông tin chỉ ra sự liên kết của một tập hợp điều chỉnh với nhiều hơn một cảnh âm thanh, mà có chung ít nhất một thành phần âm thanh chung.

○ phía thiết bị tiêu dùng phương tiện được sử dụng để tiêu dùng nội dung (ví dụ dựa trên HMD). Thiết bị tiêu dùng phương tiện cũng chịu trách nhiệm thu thập thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi hành động của người dùng)

○ bộ xử lý công nhìn 1232 mà có thể được tạo cấu hình để

- nhận thông tin về công nhìn hiện tại mà có thể chứa vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng (hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi hành động của người dùng) từ phía thiết bị tiêu dùng phương tiện.

- nhận thông tin về ROI được báo hiệu trong siêu dữ liệu (các công nhìn video được báo hiệu như trong thông số kỹ thuật OMAF).

- nhận tất cả thông tin có sẵn ở đầu nhận;

- quyết định dựa trên tất cả thông tin nhận được và/hoặc được suy ra từ siêu dữ liệu đã nhận và/hoặc có sẵn, điểm nhìn âm thanh/video nào sẽ được tái tạo tại một thời điểm nhất định. Ví dụ, bộ xử lý công nhìn 1232 có thể quyết định rằng:

+ một cảnh âm thanh hoàn chỉnh sẽ được tái tạo

+ một cảnh âm thanh mới sẽ được tái tạo trong số tất cả các cảnh âm thanh có sẵn (ví dụ, chỉ một số thành phần âm thanh từ nhiều cảnh âm thanh sẽ được tái tạo, trong khi các thành phần âm thanh còn lại khác của các cảnh âm thanh này sẽ không được tái tạo)

+ việc chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh phải được tái tạo

○ phân lựa chọn 1230 được tạo cấu hình để lựa chọn, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232, một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được từ đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh hoặc một phần cảnh âm thanh cần được tái tạo tại vị trí hiện tại của người dùng. Cảnh âm thanh này có thể là một cảnh âm thanh hoàn chỉnh hoặc hoàn chỉnh một phần như được xác định ở phía mã hóa hoặc cảnh âm thanh mới phải được tạo ra trong tất cả các cảnh âm thanh có sẵn.

- Ngoài ra, trong trường hợp khi các thành phần âm thanh 152 thuộc nhiều hơn một cảnh âm thanh, ít nhất một tập hợp điều chỉnh được chọn dựa trên thông tin chỉ ra sự liên kết của ít nhất một tập hợp điều chỉnh với nhiều hơn một cảnh âm thanh, có chứa cùng các thành phần âm thanh 152.

- Ngoài ra, trong trường hợp khi sự chuyển tiếp giữa hai hoặc nhiều cảnh âm thanh sắp xảy ra dựa trên chỉ báo của bộ xử lý công nhìn 1232, phân lựa chọn có thể được tạo cấu hình để chọn một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn như được báo hiệu trong thông tin nhận được bởi đầu nhận; các tập hợp điều chỉnh được chọn mô tả hoàn chỉnh cảnh âm thanh mà có thể được yêu cầu tái tạo trong tương lai gần (ví dụ, nếu người dùng đi theo hướng của cảnh âm thanh tiếp theo với tốc độ nhất định, có thể dự đoán rằng cảnh âm thanh tiếp theo sẽ được yêu cầu và nó được chọn trước khi tái tạo).

- Ngoài ra, một số tập hợp điều chỉnh tương ứng với các vị trí lân cận có thể được chọn trước tiên ở tốc độ bit thấp hơn (tức là, sự biểu diễn được mã hóa ở tốc độ bit thấp hơn được chọn trong số các sự biểu diễn có sẵn trong một tập hợp điều chỉnh), và dựa trên các thay đổi vị trí, chất lượng được tăng lên bằng cách chọn tốc độ bit cao hơn cho các tập hợp điều chỉnh cụ thể đó (tức là, sự biểu diễn được mã hóa ở tốc độ bit cao hơn được chọn ra trong số các sự biểu diễn có sẵn trong một tập hợp điều chỉnh).

○ phân tải xuống và chuyển đổi có thể được tạo cấu hình thành:

- yêu cầu, dựa trên chỉ báo nhận được từ phân lựa chọn 1230, một hoặc nhiều tập hợp điều chỉnh trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện 120;

- nhận, một hoặc nhiều tập hợp điều chỉnh (nghĩa là, một sự biểu diễn trong số tất cả các sự biểu diễn có sẵn trong mỗi tập hợp điều chỉnh) trong số các tập hợp điều chỉnh có sẵn từ máy chủ phương tiện 120;

- trích xuất thông tin siêu dữ liệu từ tất cả các luồng âm thanh nhận được

○ bộ xử lý siêu dữ liệu 1236 mà có thể được tạo cấu hình để:

- nhận từ thông tin tải xuống và chuyển đổi về các luồng âm thanh nhận được, thông tin có thể chứa siêu dữ liệu âm thanh tương ứng với từng luồng âm thanh nhận được

- xử lý và thao tác siêu dữ liệu âm thanh liên quan đến từng luồng âm thanh, dựa trên thông tin nhận được từ bộ xử lý công nhìn 1232 có thể chứa thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng, để:

- chọn/kích hoạt các thành phần âm thanh 152 cần thiết soạn cảnh âm thanh mới như được chỉ định bởi bộ xử lý công nhìn 1232;

+ cho phép gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.

○ bộ ghép/gộp luồng 1238 có thể được tạo cấu hình để gộp tất cả các luồng âm thanh đã chọn vào một luồng âm thanh dựa trên thông tin nhận được từ bộ xử lý siêu dữ liệu 1236 mà có thể chứa siêu dữ liệu âm thanh được biến đổi và được xử lý tương ứng với tất cả các luồng âm thanh nhận được.

○ bộ giải mã phương tiện được tạo cấu hình để nhận và giải mã ít nhất một luồng âm thanh để tái tạo cảnh âm thanh mới như được chỉ định bởi bộ xử lý công nhìn 1232 dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng di chuyển của người dùng.

Fig.1.6 thể hiện hệ thống bao gồm, ở phía máy khách, hệ thống (hệ thống máy khách) mà có thể bao gồm, ví dụ, các bộ phận của hệ thống được mô tả trên Fig.1.5 ngoài ra hoặc thay vào đó bao gồm:

- nhiều bộ giải mã phương tiện có thể được tạo cấu hình để giải mã các luồng âm thanh riêng lẻ như được chỉ định bởi bộ xử lý siêu dữ liệu 1236 (ví dụ với một số thành phần âm thanh bị vô hiệu hóa).

- bộ trộn/kết xuất 1238 có thể được tạo cấu hình để tái tạo cảnh âm thanh cuối cùng dựa trên thông tin về vị trí và/hoặc hướng và/hoặc hướng chuyển động của người dùng (tức là, ví dụ, một số thành phần âm thanh không thể nghe thấy tại vị trí cụ thể đó nên bị vô hiệu hóa hoặc không được kết xuất).

Cập nhật định dạng tệp để phát lại tệp

Đối với trường hợp sử dụng định dạng tệp, nhiều luồng chính và luồng phụ có thể được gói gọn dưới dạng các rãnh ghi riêng biệt thành một tệp ISOBMFF duy nhất. Rãnh ghi đơn lẻ của tệp như vậy sẽ biểu diễn cho thành phần âm thanh đơn lẻ như đã đề cập trước đó. Vì không có MPD khả dụng mà có chứa thông tin cần thiết để phát đúng, nên thông tin cần được cung cấp ở cấp định dạng tệp, ví dụ bằng cách cung cấp/đưa ra hộp định dạng tệp cụ thể hoặc các hộp định dạng tệp cụ thể ở cấp rãnh ghi và phim. Tùy thuộc vào trường hợp sử dụng, có các thông tin khác nhau cần thiết để cho phép kết xuất chính xác các cảnh âm thanh được gói, tuy nhiên, tập hợp thông tin sau đây là cơ bản và do đó phải luôn có:

- Thông tin về các cảnh âm thanh được bao gồm, ví dụ "vị trí các ranh giới"
- Thông tin về tất cả các thành phần âm thanh khả dụng, đặc biệt là thành phần âm thanh nào được gói gọn trong rãnh ghi nào
 - Thông tin liên quan đến vị trí của các thành phần âm thanh được gói
 - Danh sách tất cả các thành phần âm thanh thuộc về một cảnh âm thanh, thành phần âm thanh có thể thuộc về nhiều cảnh âm thanh

Với thông tin này, tất cả các trường hợp sử dụng được đề cập, chứa trường hợp có bộ xử lý siêu dữ liệu bổ sung và mã hóa chung cũng sẽ hoạt động trong môi trường dựa trên tệp.

Xem xét thêm về các ví dụ trên

Trong các ví dụ (ví dụ, ít nhất một trong số các hình vẽ từ Fig.1.1 đến Fig.6), ít nhất một cảnh được liên kết với ít nhất một thành phần âm thanh (nguồn âm thanh 152), mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan nơi thành phần âm thanh có thể nghe được, sao cho các luồng âm thanh khác nhau được cung cấp từ hệ thống máy chủ 120 đến hệ thống máy khách 102 cho các vị trí và/hoặc công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo của người dùng khác nhau trong cảnh.

Trong các ví dụ, hệ thống máy khách 102 có thể được tạo cấu hình để quyết định xem có ít nhất một thành phần âm thanh 152 của luồng âm thanh (ví dụ, A1, A2) và/hoặc một tập hợp điều chỉnh sẽ được đưa ra cho công nhìn và/hoặc hướng đầu

và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc vị trí ảo của người dùng hiện tại trong cảnh hay không, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh ở vị trí ảo của người dùng hiện tại.

Trong các ví dụ, hệ thống máy khách (ví dụ 102) có thể được tạo cấu hình để quyết định một cách dự đoán liệu ít nhất một thành phần âm thanh (152) của luồng âm thanh và/hoặc một tập hợp điều chỉnh sẽ trở nên phù hợp và/hoặc có thể nghe được dựa trên ít nhất là công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo (110) hiện tại của người dùng hay không, và trong đó hệ thống được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh và/hoặc luồng âm thanh và/hoặc tập hợp điều chỉnh ở vị trí ảo của người dùng cụ thể trước chuyển động và/hoặc tương tác của người dùng được dự đoán trong cảnh, trong đó hệ thống có thể được tạo cấu hình để tái tạo ít nhất là trên thành phần âm thanh và/hoặc luồng âm thanh, khi nhận được, tại vị trí ảo của người dùng cụ thể sau chuyển động và/hoặc tương tác của người dùng trong cảnh. Xem ví dụ Fig.8A và Fig.8B ở trên. Trong một số ví dụ, ít nhất một trong các hoạt động của hệ thống 102 hoặc 120 có thể được thực hiện trên cơ sở dữ liệu dự đoán và/hoặc thống kê và/hoặc tổng hợp.

Trong các ví dụ, hệ thống máy khách (ví dụ 102) có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh (ví dụ 1152) ở tốc độ bit và/hoặc mức chất lượng thấp hơn, tại vị trí ảo của người dùng trước sự di chuyển và/hoặc tương tác của người dùng trong cảnh, trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit và/hoặc mức chất lượng cao hơn, tại vị trí ảo của người dùng sau sự chuyển động và/hoặc tương tác của người dùng trong cảnh. Xem, ví dụ, Fig.7B.

Trong các ví dụ, ít nhất một thành phần âm thanh có thể được liên kết với ít nhất một cảnh, thành phần âm thanh cuối cùng được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan liên quan đến cảnh đó, trong đó hệ thống được tạo cấu hình để yêu cầu các luồng khác nhau ở các tốc độ bit và/hoặc mức chất lượng khác nhau cho các thành phần âm thanh dựa trên mức độ liên quan và/hoặc mức độ có thể nghe được của chúng tại mỗi vị trí ảo của người dùng trong cảnh, trong đó hệ thống

được tạo cấu hình để yêu cầu luồng âm thanh ở tốc độ bit và/hoặc mức chất lượng cao hơn cho các thành phần âm thanh có liên quan hơn và/hoặc dễ nghe hơn ở vị trí ảo của người dùng hiện tại, và/hoặc luồng âm thanh ở tốc độ bit và/hoặc mức chất lượng thấp hơn cho các thành phần âm thanh ít liên quan hơn và/hoặc ít dễ nghe hơn ở vị trí ảo của người dùng hiện tại. Xem một cách tổng quát trên Fig.7A. Cũng xem Fig.2a và Fig.2b (trong đó các nguồn có liên quan và/hoặc dễ nghe hơn có thể gần với người dùng hơn), Fig.3 (trong đó nguồn có liên quan và/hoặc dễ nghe hơn là nguồn của cảnh 150a khi người dùng ở vị trí x_1 , và nguồn có liên quan và/hoặc dễ nghe hơn là nguồn của cảnh 150b khi người dùng ở vị trí x_3), Fig.4 (trong đó, tại thời điểm tức thời t_2 , các nguồn có liên quan và/hoặc dễ nghe hơn có thể là các nguồn của cảnh thứ nhất), Fig.6 (trong đó các nguồn dễ nghe hơn có thể là các nguồn được người dùng xem trước).

Trong các ví dụ, ít nhất một thành phần âm thanh (152) được liên kết với cảnh, mỗi thành phần âm thanh được liên kết với vị trí và/hoặc khu vực trong môi trường trực quan được liên kết với cảnh đó, trong đó hệ thống máy khách 102 được tạo cấu hình để gửi định kỳ đến hệ thống máy chủ 120 chế độ xem và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo (110) hiện tại của người dùng, sao cho: đối với vị trí gần hơn với ít nhất một thành phần âm thanh (152), luồng ở tốc độ bit và/hoặc chất lượng cao hơn được cung cấp từ máy chủ, và đối với vị trí cách xa ít nhất một thành phần âm thanh (152), luồng có tốc độ bit và/hoặc chất lượng thấp hơn được cung cấp từ máy chủ. Xem, ví dụ, Fig.2a và Fig.2b.

Trong phương án của hệ thống, nhiều cảnh (ví dụ 150A, 150B) có thể được xác định cho nhiều môi trường trực quan như các môi trường liền kề và/hoặc lân cận, sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh thứ nhất, hiện tại (ví dụ 150A), và trong trường hợp người dùng chuyển (ví dụ 150AB) sang cảnh thứ hai, cảnh xa hơn (ví dụ 150B), để cung cấp cả các luồng liên quan đến cảnh thứ nhất và các luồng thứ hai liên quan đến cảnh thứ hai. Xem, ví dụ, Fig.3.

Trong các ví dụ, nhiều cảnh có thể được xác định cho môi trường trực quan thứ nhất và thứ hai, môi trường thứ nhất và thứ hai là các môi trường liền kề và/hoặc lân cận, trong đó các luồng thứ nhất liên quan đến cảnh thứ nhất được cung cấp, từ máy chủ, cho việc tái tạo cảnh thứ nhất trong trường hợp vị trí ảo của người dùng ở môi

trường thứ nhất liên quan đến cảnh thứ nhất, các luồng thứ hai liên quan đến cảnh thứ hai được cung cấp, từ máy chủ, cho việc tái tạo cảnh thứ hai trong trường hợp vị trí ảo của người dùng trong môi trường thứ hai liên quan đến cảnh thứ hai, và cả luồng thứ nhất liên quan đến cảnh thứ nhất và luồng thứ hai liên quan đến cảnh thứ hai được cung cấp trong trường hợp vị trí ảo của người dùng ở vị trí chuyển tiếp giữa cảnh thứ nhất và cảnh thứ hai. Xem, ví dụ, Fig.3.

Trong các ví dụ, các luồng thứ nhất được liên kết với cảnh thứ nhất thu được ở tốc độ bit và/hoặc chất lượng cao hơn khi người dùng ở môi trường thứ nhất liên quan đến cảnh thứ nhất, trong khi các luồng thứ hai liên quan đến cảnh thứ hai liên quan đến môi trường thứ hai thu được ở tốc độ bit và/hoặc chất lượng thấp hơn khi người dùng ở đầu vị trí chuyển tiếp từ cảnh thứ nhất sang cảnh thứ hai, và các luồng thứ nhất liên quan đến cảnh thứ nhất thu được tại tốc độ bit và/hoặc chất lượng thấp hơn và các luồng thứ hai liên quan đến cảnh thứ hai thu được ở tốc độ bit và/hoặc chất lượng cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh thứ nhất sang cảnh thứ hai. Đây có thể là trường hợp, ví dụ, của Fig.3.

Trong các ví dụ, nhiều cảnh (ví dụ, 150A, 150B) được xác định cho nhiều môi trường trực quan (ví dụ, các môi trường liền kề), để hệ thống 102 có thể yêu cầu và/hoặc thu được các luồng liên quan đến cảnh hiện tại ở tốc độ bit và/hoặc chất lượng cao hơn và các luồng liên quan đến cảnh thứ hai ở tốc độ bit và/hoặc chất lượng thấp hơn. Xem, ví dụ, Fig.4.

Trong các ví dụ, số lượng lớn của N thành phần âm thanh được xác định, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của các thành phần âm thanh này lớn hơn ngưỡng định trước, N thành phần âm thanh được xử lý để thu được số lượng M nhỏ hơn của các thành phần âm thanh ($M < N$) liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của N thành phần âm thanh, để cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến N thành phần âm thanh, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của N thành phần âm thanh nhỏ hơn ngưỡng xác định trước, hoặc để cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến M thành phần âm thanh, trong trường hợp khoảng cách

của người dùng đến vị trí hoặc khu vực của N thành phần âm thanh lớn hơn ngưỡng xác định trước. Xem, ví dụ, Fig.1.7.

Trong các ví dụ, ít nhất một cảnh môi trường trực quan liên quan đến ít nhất một số lượng lớn của N thành phần âm thanh ($N \geq 2$), mỗi thành phần âm thanh liên quan đến vị trí và/hoặc khu vực trong môi trường trực quan, trong đó ít nhất một số lượng lớn của N thành phần âm thanh có thể được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng cao, và trong đó ít nhất một số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng thấp, trong đó có ít nhất một sự biểu diễn thu được bằng cách xử lý N thành phần âm thanh để thu được số lượng nhỏ hơn của M thành phần âm thanh ($M < N$) liên quan đến vị trí hoặc khu vực gần vị trí hoặc khu vực của N thành phần âm thanh, trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng cao hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh liên quan hơn và/hoặc dễ nghe hơn ở vị trí ảo của người dùng hiện tại trong cảnh, trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit và/hoặc mức chất lượng thấp hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh ít liên quan hơn và/hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh. Xem, ví dụ, Fig.1.7.

Trong các ví dụ, trong trường hợp khoảng cách của người dùng và/hoặc mức độ liên quan và/hoặc mức độ dễ nghe và/hoặc hướng góc thấp hơn ngưỡng xác định trước, các luồng khác nhau sẽ thu được cho các thành phần âm thanh khác nhau. Xem, ví dụ, Fig.1.7.

Trong các ví dụ, các thành phần âm thanh khác nhau được cung cấp ở các cổng nhìn khác nhau, sao cho, trong trường hợp một thành phần âm thanh thứ nhất nằm trong cổng nhìn hiện tại, thành phần âm thanh thứ nhất thu được ở tốc độ bit cao hơn thành phần âm thanh thứ hai không nằm trong cổng nhìn. Xem, ví dụ, Fig.6.

Trong các ví dụ, ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh thứ nhất liên quan đến môi trường trực quan thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh thứ hai liên quan đến môi trường trực quan thứ hai, trong đó hệ

thống 102 được tạo cấu hình để thu được siêu dữ liệu mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với cảnh môi trường trực quan thứ hai, và trong đó hệ thống được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất là các thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở môi trường trực quan thứ nhất, và trong đó hệ thống có thể được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất là các thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng nằm trong cảnh môi trường trực quan thứ hai, và trong đó hệ thống được tạo cấu hình để yêu cầu và/hoặc nhận ít nhất là các thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển tiếp giữa cảnh môi trường trực quan thứ nhất và cảnh môi trường trực quan thứ hai. Xem, ví dụ, Fig.1.4. Điều này có thể cũng áp dụng cho Fig.3.

Trong các ví dụ, ít nhất một thành phần âm thanh thứ nhất được cung cấp trong ít nhất một luồng âm thanh và/hoặc tập hợp điều chỉnh, và ít nhất một thành phần âm thanh thứ hai được cung cấp trong ít nhất một luồng âm thanh và/hoặc tập hợp điều chỉnh thứ hai, và ít nhất một thành phần âm thanh thứ ba được cung cấp trong ít nhất một luồng âm thanh và/hoặc tập hợp điều chỉnh thứ ba, và trong đó ít nhất là cảnh môi trường trực quan thứ nhất được mô tả bằng siêu dữ liệu là cảnh hoàn chỉnh yêu cầu ít nhất là các luồng âm thanh và/hoặc tập hợp điều chỉnh thứ nhất và thứ hai, và trong đó cảnh môi trường trực quan thứ hai được mô tả bởi siêu dữ liệu là cảnh không hoàn chỉnh mà yêu cầu ít nhất là luồng âm thanh và/hoặc tập hợp điều chỉnh thứ ba và ít nhất là luồng âm thanh và/hoặc các tập hợp điều chỉnh thứ hai liên quan đến ít nhất là cảnh môi trường trực quan thứ nhất, trong đó hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu, để cho phép gộp luồng âm thanh thứ hai thuộc về môi trường trực quan thứ nhất và luồng âm thanh thứ ba liên quan đến môi trường trực quan thứ hai thành một luồng mới duy nhất, trong trường hợp vị trí ảo của người dùng nằm trong môi trường trực quan thứ hai. Xem, ví dụ, các hình vẽ từ Fig.1,2 đến Fig.1,3, Fig.1,5, và Fig.1,6.

Trong các ví dụ, hệ thống 102 có thể bao gồm bộ xử lý siêu dữ liệu (ví dụ, 1236) được tạo cấu hình để thao tác siêu dữ liệu trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên công nhìn và/hoặc hướng đầu và/hoặc

dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng.

Trong các ví dụ, bộ xử lý siêu dữ liệu (ví dụ, 1236) có thể được tạo cấu hình để kích hoạt và/hoặc vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên công nhìn và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng, trong đó bộ xử lý siêu dữ liệu có thể được tạo cấu hình để vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, trong trường hợp hệ thống quyết định rằng thành phần âm thanh sẽ không được tái tạo nữa do hệ quả của công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại, và trong đó bộ xử lý siêu dữ liệu có thể được tạo cấu hình để kích hoạt ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, trong trường hợp hệ thống quyết định thành phần âm thanh sẽ được tái tạo do hệ quả của công nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại của người dùng.

Phía máy chủ

Ở đây cũng đề cập đến máy chủ (120) để phân phối các luồng âm thanh và video tới máy khách cho thực tế ảo VR, thực tế tăng cường AR, thực tế hỗn hợp MR, hoặc môi trường video 360 độ, các luồng video và âm thanh sẽ được tái tạo trong thiết bị tiêu dùng phương tiện, trong đó máy chủ (120) bao gồm và bộ mã hóa để mã hóa và/hoặc lưu trữ để lưu trữ các luồng video để mô tả môi trường trực quan, môi trường trực quan được liên kết với cảnh âm thanh; trong đó máy chủ còn bao gồm bộ mã hóa để mã hóa và/hoặc lưu trữ để lưu trữ nhiều luồng và/hoặc các thành phần âm thanh và/hoặc các tập hợp điều chỉnh sẽ được phân phối đến máy khách, các luồng và/hoặc các thành phần âm thanh và/hoặc các tập hợp điều chỉnh được liên kết với ít nhất một cảnh âm thanh, trong đó máy chủ được tạo cấu hình để:

chọn và phân phối luồng video trên cơ sở yêu cầu từ máy khách, luồng video liên quan đến môi trường;

chọn luồng âm thanh và/hoặc thành phần âm thanh và/hoặc tập hợp điều chỉnh trên cơ sở yêu cầu từ máy khách, yêu cầu liên quan đến ít nhất là cổng nhìn và/hoặc hướng đầu và/hoặc dữ liệu chuyển động và/hoặc siêu dữ liệu tương tác và/hoặc dữ liệu vị trí ảo hiện tại và liên quan đến cảnh âm thanh liên quan đến môi trường; và

phân phối luồng âm thanh cho máy khách.

Các phương án và biến thể khác nữa

Tùy thuộc vào các yêu cầu thực hiện nhất định, các ví dụ có thể được thực hiện trong phần cứng. Việc triển khai có thể được thực hiện bằng cách sử dụng phương tiện lưu trữ kỹ thuật số, ví dụ như đĩa mềm, đĩa đa năng kỹ thuật số (DVD), đĩa Blu-Ray, đĩa compac (CD), bộ nhớ chỉ đọc (ROM), bộ nhớ chỉ đọc có thể lập trình (PROM), bộ nhớ chỉ đọc có thể xóa và có thể lập trình (EPROM), bộ nhớ chỉ đọc có thể lập trình xóa được bằng điện (EEPROM) hoặc bộ nhớ flash, có các tín hiệu điều khiển có thể đọc được bằng điện tử được lưu trữ trên đó, mà kết hợp (hoặc có khả năng kết hợp) với hệ thống máy tính có thể lập trình sao cho phương pháp tương ứng được thực hiện. Do đó, vật ghi lưu trữ số có thể có khả năng đọc được bằng máy tính.

Nói chung, các ví dụ của sáng chế có thể được triển khai dưới dạng sản phẩm chương trình máy tính với các lệnh chương trình, các lệnh chương trình có tác dụng để thực hiện một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Các lệnh chương trình có thể ví dụ được lưu trữ trên vật mang có thể đọc được bằng máy.

Các ví dụ khác bao gồm chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây, được lưu trữ trên vật mang có thể đọc được bằng máy. Do đó, nói cách khác, ví dụ của phương pháp theo sáng chế là chương trình máy tính có các lệnh chương trình để thực hiện một trong số các phương pháp được mô tả ở đây, khi chương trình máy tính chạy trên máy tính.

Do đó, ví dụ khác của các phương pháp là phương tiện mang dữ liệu (hoặc phương tiện ghi lưu trữ số, hoặc vật ghi có thể đọc được bằng máy tính) bao gồm, chương trình máy tính đã được ghi lại trên đó để thực hiện một trong số các phương pháp được mô tả ở đây. Phương tiện mang dữ liệu, phương tiện lưu trữ kỹ thuật số

hoặc phương tiện được ghi lại là hữu hình và/hoặc không tạm thời, thay vì các tín hiệu vô hình và tạm thời.

Ví dụ khác bao gồm bộ xử lý, ví dụ máy tính, hoặc thiết bị logic lập trình được thực hiện một trong số các phương pháp được mô tả ở đây.

Ví dụ khác bao gồm máy tính có chương trình máy tính được cài đặt trên đó để thực hiện một trong số các phương pháp được mô tả ở đây.

Ví dụ khác bao gồm thiết bị hoặc hệ thống chuyển (ví dụ, bằng điện tử hoặc quang học) chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây đến bộ nhận. Bộ nhận có thể là, ví dụ, máy tính, thiết bị di động, thiết bị nhớ hoặc tương tự. Thiết bị hoặc hệ thống có thể, ví dụ, gồm có máy chủ tập tin để chuyển chương trình máy tính đến bộ nhận.

Trong một số ví dụ, thiết bị logic lập trình được (ví dụ, mảng cổng lập trình được dạng trường) có thể được sử dụng để thực hiện một số hoặc tất cả các chức năng của các phương pháp được mô tả ở đây. Trong một số ví dụ, mảng cổng lập trình được dạng trường có thể kết hợp với bộ vi xử lý để thực hiện một trong số các phương pháp được mô tả ở đây. Thông thường, các phương pháp có thể được thực hiện bởi thiết bị phần cứng thích hợp.

Các ví dụ được mô tả ở trên có thể minh họa cho các nguyên tắc được thảo luận ở trên. Cần hiểu rằng các biến đổi và biến thể của các phương án và các chi tiết được mô tả ở đây sẽ được bảo hộ. Do đó, mục đích của sáng chế chỉ bị giới hạn bởi phạm vi của các điểm yêu cầu bảo hộ sắp đưa ra dưới đây và không bị giới hạn bởi các chi tiết cụ thể được biểu diễn bằng cách mô tả và giải thích của các ví dụ ở đây.

YÊU CẦU BẢO HỘ

1. Hệ thống phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống bao gồm:

ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh được tạo cấu hình để giải mã các tín hiệu âm thanh từ các luồng âm thanh cho sự biểu diễn các cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu các luồng âm thanh thứ nhất và các luồng âm thanh thứ hai hoặc một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh cho máy chủ trên cơ sở ít nhất là công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng,

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai,

trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video bao gồm các môi trường video liên kề hoặc lân cận,

sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh âm thanh thứ nhất, và trong trường hợp người dùng chuyển sang cảnh âm thanh thứ hai, để cung cấp cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit cao hơn khi người dùng ở trong môi trường thứ nhất được liên kết với cảnh âm thanh thứ nhất,

trong khi các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai được thu ở tốc độ bit thấp hơn khi người dùng bắt đầu ở vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit thấp hơn và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được thu ở tốc độ bit cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn.

2. Hệ thống theo điểm 1, được tạo cấu hình để cung cấp cho máy chủ công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng để thu được ít nhất một luồng âm thanh hoặc một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh từ máy chủ.

3. Hệ thống theo điểm 1, trong đó ít nhất một cảnh âm thanh liên quan đến ít nhất một thành phần âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video nơi thành phần âm thanh có thể nghe được, sao cho các luồng âm thanh khác nhau được cung cấp cho các vị trí hoặc các công nhìn hoặc các hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo của người dùng khác nhau trong cảnh âm thanh.

4. Hệ thống theo điểm 1, được tạo cấu hình để quyết định xem ít nhất một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh sẽ được tái tạo cho công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc vị trí ảo của người dùng hiện tại trong cảnh âm thanh hay không, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở vị trí ảo của người dùng hiện tại.

5. Hệ thống theo điểm 1, được tạo cấu hình để quyết định một cách dự đoán xem ít nhất một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh sẽ trở nên có liên quan hoặc dễ nghe thấy dựa trên ít nhất là công nhìn hoặc hướng đầu

hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc vị trí ảo hiện tại của người dùng hay không,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh hoặc luồng âm thanh hoặc tập hợp điều chỉnh tại vị trí ảo của người dùng cụ thể trước sự chuyển động hoặc tương tác của người dùng được dự đoán trong cảnh âm thanh, và

trong đó hệ thống được tạo cấu hình để tái tạo ít nhất một thành phần âm thanh hoặc luồng âm thanh, khi nhận được, tại vị trí ảo cụ thể của người dùng sau sự chuyển động hoặc tương tác của người dùng trong cảnh âm thanh.

6. Hệ thống theo điểm 1, được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit thấp hơn, tại vị trí ảo của người dùng trước sự tương tác của người dùng, sự tương tác dẫn đến thay đổi dữ liệu vị trí trong cùng một cảnh âm thanh hoặc vào cảnh tiếp theo tách biệt với cảnh hiện tại,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit cao hơn, tại vị trí ảo của người dùng sau sự tương tác của người dùng trong cảnh âm thanh.

7. Hệ thống theo điểm 1, trong đó ít nhất một thành phần âm thanh liên quan đến ít nhất một cảnh âm thanh được liên kết với vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng ở tốc độ bit cao hơn đối với các thành phần âm thanh gần với người dùng hơn so với các thành phần âm thanh ở xa người dùng hơn.

8. Hệ thống theo điểm 1, trong đó ít nhất một thành phần âm thanh liên quan đến ít nhất một cảnh âm thanh, ít nhất một thành phần âm thanh được liên kết với vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu các luồng khác nhau ở các tốc độ bit khác nhau cho các thành phần âm thanh dựa trên mức độ liên quan hoặc mức độ dễ nghe thấy của chúng tại từng vị trí ảo của người dùng trong cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu luồng âm thanh ở tốc độ bit cao hơn cho các thành phần âm thanh liên quan hơn hoặc dễ nghe thấy hơn ở vị trí ảo của người dùng hiện tại, hoặc

luồng âm thanh ở tốc độ bit thấp hơn cho các thành phần âm thanh ít liên quan hơn hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại.

9. Hệ thống theo điểm 1, trong đó ít nhất một thành phần âm thanh liên quan đến cảnh âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để gửi định kỳ tới máy chủ cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng, sao cho:

đối với vị trí thứ nhất, luồng ở tốc độ bit cao hơn được cung cấp từ máy chủ, và

đối với vị trí thứ hai, luồng ở tốc độ bit thấp hơn được cung cấp từ máy chủ,

trong đó vị trí thứ nhất gần với ít nhất một thành phần âm thanh hơn so với vị trí thứ hai.

10. Hệ thống theo điểm 1, trong đó nhiều cảnh âm thanh được xác định cho các môi trường video thứ nhất và thứ hai, các môi trường video thứ nhất và thứ hai là các môi trường video liền kề hoặc lân cận,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được cung cấp từ máy chủ để tái tạo cảnh âm thanh thứ nhất trong trường hợp vị trí hoặc vị trí ảo của người dùng trong môi trường video thứ nhất được liên kết với cảnh âm thanh thứ nhất,

các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được cung cấp từ máy chủ để tái tạo cảnh âm thanh thứ hai trong trường hợp vị trí hoặc vị trí ảo của người dùng trong môi trường video thứ hai được liên kết với cảnh âm thanh thứ hai, và

cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được cung cấp trong trường hợp vị trí hoặc vị trí ảo của người dùng ở vị trí chuyển tiếp giữa cảnh âm thanh thứ nhất và cảnh âm thanh thứ hai.

11. Hệ thống theo điểm 1, trong đó nhiều cảnh âm thanh được xác định cho các môi trường video thứ nhất và thứ hai là các môi trường video liền kề hoặc lân cận,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường thứ nhất, để tái tạo cảnh âm thanh thứ nhất trong trường hợp vị trí ảo của người dùng ở môi trường thứ nhất,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai, để tái tạo cảnh âm thanh thứ hai trong trường hợp vị trí ảo của người dùng ở môi trường thứ hai, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai trong trường hợp vị trí ảo của người dùng ở vị trí chuyển tiếp giữa môi trường thứ nhất và môi trường thứ hai.

12. Hệ thống theo điểm 1, trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường gồm các môi trường liền kề hoặc lân cận,

sao cho hệ thống được tạo cấu hình để thu được các luồng âm thanh liên quan đến cảnh âm thanh hiện tại thứ nhất được liên kết với môi trường hiện tại, thứ nhất, và,

trong trường hợp khoảng cách của vị trí hoặc vị trí ảo của người dùng từ ranh giới của cảnh âm thanh nằm dưới ngưỡng xác định trước, hệ thống sẽ tiếp tục thu các luồng âm thanh liên quan đến môi trường video thứ hai, liền kề hoặc lân cận liên quan đến cảnh âm thanh thứ hai.

13. Hệ thống theo điểm 1, trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video,

sao cho hệ thống yêu cầu hoặc thu được các luồng âm thanh liên quan đến cảnh âm thanh hiện tại ở tốc độ bit cao hơn và các luồng âm thanh liên quan đến cảnh âm thanh thứ hai ở tốc độ bit thấp hơn,

trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn.

14. Hệ thống theo điểm 1, trong đó số lượng lớn của N thành phần âm thanh được xác định, và trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh lớn hơn ngưỡng xác định trước, số lượng lớn của N thành phần âm thanh được xử lý để thu được số lượng M nhỏ hơn các thành phần âm thanh liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của N thành phần âm thanh, để

cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến số lượng lớn của N thành phần âm thanh, trong trường hợp khoảng cách người dùng đến vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh nhỏ hơn ngưỡng xác định trước, hoặc

cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến M thành phần âm thanh, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh lớn hơn ngưỡng xác định trước.

15. Hệ thống theo điểm 1, trong đó ít nhất một môi trường âm thanh liên quan đến số lượng lớn của N thành phần âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video,

trong đó ít nhất một trong số số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit cao, và

trong đó ít nhất một trong số số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit thấp, trong đó ít nhất một sự biểu diễn thu được bằng cách xử lý N thành phần âm thanh để thu được số lượng M nhỏ hơn các thành phần âm thanh liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của N thành phần âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit cao hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh liên quan hơn hoặc dễ nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit thấp hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh ít liên

quan hơn hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh âm thanh.

16. Hệ thống theo điểm 9, trong đó, trong trường hợp khoảng cách của người dùng thấp hơn ngưỡng khoảng cách được xác định trước, hoặc mức độ liên quan thấp hơn ngưỡng liên quan được xác định trước, hoặc mức độ nghe thấy thấp hơn ngưỡng xác định trước, các luồng âm thanh khác nhau thu được cho các thành phần âm thanh khác nhau.

17. Hệ thống theo điểm 1, được tạo cấu hình để yêu cầu hoặc thu các luồng âm thanh trên cơ sở hướng của người dùng hoặc hướng chuyển động của người dùng hoặc sự tương tác của người dùng trong cảnh âm thanh.

18. Hệ thống theo điểm 1, trong đó cổng nhìn liên quan đến vị trí hoặc vị trí ảo hoặc dữ liệu chuyển động hoặc hướng đầu.

19. Hệ thống theo điểm 1, trong đó các thành phần âm thanh khác nhau được cung cấp ở các cổng nhìn khác nhau, trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận, trong trường hợp một thành phần âm thanh thứ nhất nằm trong cổng nhìn, thành phần âm thanh thứ nhất tại tốc độ bit cao hơn so với thành phần âm thanh thứ hai không nằm trong cổng nhìn.

20. Hệ thống theo điểm 1, được tạo cấu hình để vô hiệu hóa việc giải mã các thành phần âm thanh được chọn trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc vị trí ảo hiện tại của người dùng.

21. Hệ thống theo điểm 1, được tạo cấu hình để gộp ít nhất một luồng âm thanh thứ nhất liên quan đến cảnh âm thanh hiện tại với ít nhất một luồng liên quan đến cảnh âm thanh lân cận, liền kề hoặc tương lai.

22. Hệ thống theo điểm 1, được tạo cấu hình để lấy hoặc thu thập dữ liệu thống kê hoặc tổng hợp trên cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để truyền yêu cầu đến máy chủ liên quan đến dữ liệu thống kê hoặc tổng hợp.

23. Hệ thống theo điểm 1, được tạo cấu hình để vô hiệu hóa việc giải mã hoặc việc tái tạo ít nhất một luồng trên cơ sở siêu dữ liệu liên quan đến ít nhất một luồng và trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng.

24. Hệ thống theo điểm 1 còn được tạo cấu hình để:

thao tác siêu dữ liệu liên quan đến nhóm các luồng âm thanh đã chọn, dựa trên ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để:

chọn hoặc kích hoạt các thành phần âm thanh soạn cảnh âm thanh sẽ được tái tạo; hoặc

gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.

25. Hệ thống theo điểm 1, trong đó thông tin được cung cấp từ máy chủ, đối với từng thành phần âm thanh hoặc đối tượng âm thanh, trong đó thông tin chứa thông tin mô tả về các vị trí mà cảnh âm thanh hoặc các thành phần âm thanh đang hoạt động.

26. Hệ thống theo điểm 1, được tạo cấu hình để lựa chọn giữa việc tái tạo một cảnh âm thanh và biên soạn hoặc trộn hoặc ghép hoặc chồng hoặc kết hợp ít nhất hai cảnh âm thanh trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc vị trí ảo hoặc lựa chọn của người dùng hiện tại hoặc tương lai, hai cảnh âm thanh liên quan đến các môi trường video lân cận hoặc liền kề khác nhau.

27. Hệ thống theo điểm 1, được tạo cấu hình để tạo hoặc sử dụng ít nhất một tập hợp điều chỉnh sao cho:

một số tập hợp điều chỉnh được liên kết với một cảnh âm thanh; hoặc

thông tin bổ sung được cung cấp liên quan từng tập hợp điều chỉnh với một điểm nhìn hoặc một cảnh âm thanh; hoặc

thông tin bổ sung được cung cấp chứa

- thông tin về các ranh giới của một cảnh âm thanh, hoặc
- thông tin về mối quan hệ giữa một tập hợp điều chỉnh và một cảnh âm thanh, hoặc
- thông tin về kết nối giữa các ranh giới của cảnh âm thanh và nhiều tập hợp điều chỉnh.

28. Hệ thống theo điểm 1, được tạo cấu hình để:

nhận luồng cho cảnh âm thanh liên quan đến môi trường lân cận hoặc liền kề; bắt đầu giải mã hoặc tái tạo luồng âm thanh cho môi trường lân cận hoặc liền kề khi phát hiện sự chuyển tiếp ranh giới giữa hai môi trường.

29. Hệ thống theo điểm 1, trong đó hệ thống còn được tạo cấu hình để:

yêu cầu hoặc nhận ít nhất một tập hợp điều chỉnh thứ nhất bao gồm ít nhất một luồng âm thanh liên quan đến ít nhất một cảnh âm thanh thứ nhất;

yêu cầu hoặc nhận ít nhất một tập hợp điều chỉnh thứ hai bao gồm ít nhất một luồng âm thanh thứ hai liên quan đến ít nhất hai cảnh âm thanh, chứa ít nhất một cảnh âm thanh thứ nhất; và

gộp ít nhất một luồng âm thanh thứ nhất và ít nhất một luồng âm thanh thứ hai thành luồng âm thanh mới để được giải mã, dựa trên siêu dữ liệu có sẵn về cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc lựa chọn hiện tại của người dùng hoặc thông tin mô tả sự liên kết của ít nhất một tập hợp điều chỉnh thứ nhất với ít nhất một cảnh âm thanh thứ nhất hoặc sự liên kết của ít nhất một tập hợp điều chỉnh thứ hai với ít nhất một cảnh âm thanh thứ nhất.

30. Hệ thống theo điểm 1 được tạo cấu hình để:

nhận thông tin về cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi các hành động của người dùng; và

nhận thông tin về tính khả dụng của các tập hợp điều chỉnh và thông tin mô tả sự liên kết của ít nhất một tập hợp điều chỉnh với ít nhất một cảnh âm thanh hoặc điểm nhìn hoặc cổng nhìn hoặc vị trí hoặc vị trí ảo hoặc dữ liệu chuyển động hoặc hướng.

31. Hệ thống theo điểm 1 được tạo cấu hình để:

quyết định xem có ít nhất một thành phần âm thanh từ ít nhất một cảnh âm thanh được lồng vào ít nhất một luồng âm thanh và ít nhất một thành phần âm thanh bổ sung từ ít nhất một cảnh âm thanh bổ sung được lồng vào ít nhất một luồng âm thanh bổ sung sẽ được tái tạo; và

thực hiện, trong trường hợp quyết định là có, thao tác gộp hoặc biên soạn hoặc ghép hoặc chồng hoặc kết hợp ít nhất một luồng bổ sung của cảnh âm thanh bổ sung vào ít nhất một luồng của ít nhất một cảnh âm thanh.

32. Hệ thống theo điểm 1 được tạo cấu hình để:

thao tác siêu dữ liệu âm thanh liên quan đến các luồng âm thanh đã chọn, dựa trên ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để:

chọn hoặc cho phép hoặc kích hoạt các thành phần âm thanh soạn cảnh âm thanh được quyết định sẽ tái tạo; và

cho phép gộp tất cả các luồng âm thanh được chọn thành một luồng âm thanh duy nhất.

33. Hệ thống phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống bao gồm:

ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh được tạo cấu hình để giải mã các tín hiệu âm thanh từ ít nhất một luồng âm thanh cho sự biểu diễn cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu ít nhất một luồng âm thanh cho máy chủ trên cơ sở ít nhất là công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng, và

trong đó hệ thống được tạo cấu hình để kiểm soát yêu cầu của ít nhất một luồng âm thanh đến máy chủ trên cơ sở khoảng cách vị trí của người dùng từ các ranh giới của các môi trường video lân cận hoặc liên kề liên quan đến các cảnh âm thanh khác nhau,

trong đó ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường video thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường video thứ hai,

trong đó hệ thống được tạo cấu hình để thu được siêu dữ liệu tương tác mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với môi trường video thứ hai,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ nhất,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ hai, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa môi trường video thứ nhất và môi trường video thứ hai.

34. Hệ thống theo điểm 33, được tạo cấu hình để cung cấp cho máy chủ công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng để thu được ít nhất một luồng âm thanh hoặc một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh từ máy chủ.

35. Hệ thống theo điểm 33, trong đó ít nhất một cảnh âm thanh liên quan đến ít nhất một thành phần âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video nơi thành phần âm thanh có thể nghe được, sao cho các luồng âm thanh khác nhau được cung cấp cho các vị trí hoặc các cổng nhìn hoặc các hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo của người dùng khác nhau trong cảnh âm thanh.

36. Hệ thống theo điểm 33, được tạo cấu hình để quyết định xem ít nhất một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh sẽ được tái tạo cho cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc vị trí ảo của người dùng hiện tại hay không, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở vị trí ảo của người dùng hiện tại.

37. Hệ thống theo điểm 33, được tạo cấu hình để quyết định một cách dự đoán xem ít nhất một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh sẽ trở nên có liên quan hoặc dễ nghe thấy dựa trên ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc vị trí ảo hiện tại của người dùng hay không,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh hoặc luồng âm thanh hoặc tập hợp điều chỉnh tại vị trí ảo của người dùng cụ thể trước sự chuyển động hoặc tương tác của người dùng được dự đoán trong cảnh âm thanh, và

trong đó hệ thống được tạo cấu hình để tái tạo ít nhất một thành phần âm thanh hoặc luồng âm thanh, khi nhận được, tại vị trí ảo cụ thể của người dùng sau sự chuyển động hoặc tương tác của người dùng trong cảnh âm thanh.

38. Hệ thống theo điểm 33, được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit thấp hơn, tại vị trí ảo của người dùng trước sự tương tác của người dùng, sự tương tác dẫn đến thay đổi dữ liệu vị trí trong cùng một cảnh âm thanh hoặc vào cảnh tiếp theo tách biệt với cảnh hiện tại,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh ở tốc độ bit cao hơn, tại vị trí ảo của người dùng sau sự tương tác của người dùng trong cảnh âm thanh.

39. Hệ thống theo điểm 33, trong đó ít nhất một thành phần âm thanh liên quan đến ít nhất một cảnh âm thanh được liên kết với vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng ở tốc độ bit cao hơn đối với các thành phần âm thanh gần với người dùng hơn so với các thành phần âm thanh ở xa người dùng hơn.

40. Hệ thống theo điểm 33, trong đó ít nhất một thành phần âm thanh liên quan đến ít nhất một cảnh âm thanh, ít nhất một thành phần âm thanh được liên kết với vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu các luồng khác nhau ở các tốc độ bit khác nhau cho các thành phần âm thanh dựa trên mức độ liên quan hoặc mức độ dễ nghe thấy của chúng tại từng vị trí ảo của người dùng trong cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu luồng âm thanh ở tốc độ bit cao hơn cho các thành phần âm thanh liên quan hơn hoặc dễ nghe thấy hơn ở vị trí ảo của người dùng hiện tại, hoặc

luồng âm thanh ở tốc độ bit thấp hơn cho các thành phần âm thanh ít liên quan hơn hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại.

41. Hệ thống theo điểm 33, trong đó ít nhất một thành phần âm thanh liên quan đến cảnh âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video được liên kết với cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để gửi định kỳ tới máy chủ cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng, sao cho:

đối với vị trí thứ nhất, luồng ở tốc độ bit cao hơn được cung cấp từ máy chủ, và đối với vị trí thứ hai, luồng ở tốc độ bit thấp hơn được cung cấp từ máy chủ,

trong đó vị trí thứ nhất gần với ít nhất một thành phần âm thanh hơn so với vị trí thứ hai.

42. Hệ thống theo điểm 33, trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video gồm các môi trường video liên kề hoặc lân cận,

sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh âm thanh thứ nhất, hiện tại, và trong trường hợp người dùng chuyển sang cảnh âm thanh thứ hai, để cung cấp cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai.

43. Hệ thống theo điểm 33, trong đó nhiều cảnh âm thanh được xác định cho các môi trường video thứ nhất và thứ hai, các môi trường video thứ nhất và thứ hai là các môi trường video liên kề hoặc lân cận,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được cung cấp từ máy chủ để tái tạo cảnh âm thanh thứ nhất trong trường hợp vị trí hoặc vị trí ảo của người dùng trong môi trường video thứ nhất được liên kết với cảnh âm thanh thứ nhất,

các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được cung cấp từ máy chủ để tái tạo cảnh âm thanh thứ hai trong trường hợp vị trí hoặc vị trí ảo của người dùng trong môi trường video thứ hai được liên kết với cảnh âm thanh thứ hai, và

cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được cung cấp trong trường hợp vị trí hoặc vị trí ảo của người dùng ở vị trí chuyển tiếp giữa cảnh âm thanh thứ nhất và cảnh âm thanh thứ hai.

44. Hệ thống theo điểm 33, trong đó nhiều cảnh âm thanh được xác định cho các môi trường video thứ nhất và thứ hai là các môi trường video liên kề hoặc lân cận,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường thứ nhất, để tái tạo cảnh âm thanh thứ nhất trong trường hợp vị trí ảo của người dùng ở môi trường thứ nhất,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai, để tái tạo cảnh âm thanh thứ hai trong trường hợp vị trí ảo của người dùng ở môi trường thứ hai, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai trong trường hợp vị trí ảo của người dùng ở vị trí chuyển tiếp giữa môi trường thứ nhất và môi trường thứ hai.

45. Hệ thống theo điểm 33, trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường gồm các môi trường liền kề hoặc lân cận,

sao cho hệ thống được tạo cấu hình để thu được các luồng âm thanh liên quan đến cảnh âm thanh hiện tại thứ nhất được liên kết với môi trường hiện tại, thứ nhất, và,

trong trường hợp khoảng cách của vị trí hoặc vị trí ảo của người dùng từ ranh giới của cảnh âm thanh nằm dưới ngưỡng xác định trước, hệ thống sẽ tiếp tục thu các luồng âm thanh liên quan đến môi trường video thứ hai, liền kề hoặc lân cận liên quan đến cảnh âm thanh thứ hai.

46. Hệ thống theo điểm 33, trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video,

sao cho hệ thống yêu cầu hoặc thu được các luồng âm thanh liên quan đến cảnh âm thanh hiện tại ở tốc độ bit cao hơn và các luồng âm thanh liên quan đến cảnh âm thanh thứ hai ở tốc độ bit thấp hơn,

trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn.

47. Hệ thống theo điểm 33, trong đó số lượng lớn của N thành phần âm thanh được xác định, và trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của các thành phần âm thanh này lớn hơn ngưỡng xác định trước, số lượng lớn của N thành phần âm thanh được xử lý để thu được số lượng M nhỏ hơn các thành phần âm thanh liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh, để

cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến số lượng lớn của N thành phần âm thanh, trong trường hợp khoảng cách người dùng đến vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh nhỏ hơn ngưỡng xác định trước, hoặc

cung cấp cho hệ thống ít nhất một luồng âm thanh liên quan đến M thành phần âm thanh, trong trường hợp khoảng cách của người dùng đến vị trí hoặc khu vực của số lượng lớn của N thành phần âm thanh lớn hơn ngưỡng xác định trước.

48. Hệ thống theo điểm 33, trong đó ít nhất một môi trường âm thanh liên quan đến ít nhất một số lượng lớn của N thành phần âm thanh, mỗi thành phần âm thanh liên quan đến vị trí hoặc khu vực trong môi trường video,

trong đó ít nhất một trong số số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit cao, và

trong đó ít nhất một trong số số lượng lớn của N thành phần âm thanh được cung cấp trong ít nhất một sự biểu diễn ở tốc độ bit thấp, trong đó ít nhất một sự biểu diễn thu được bằng cách xử lý N thành phần âm thanh để thu được số lượng M nhỏ hơn các thành phần âm thanh liên quan đến vị trí hoặc khu vực gần với vị trí hoặc khu vực của N thành phần âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit cao hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh liên quan hơn hoặc dễ nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu sự biểu diễn ở tốc độ bit thấp hơn cho các thành phần âm thanh, trong trường hợp các thành phần âm thanh ít liên quan hơn hoặc ít nghe thấy hơn ở vị trí ảo của người dùng hiện tại trong cảnh âm thanh.

49. Hệ thống theo điểm 33, được tạo cấu hình để yêu cầu hoặc thu các luồng âm thanh trên cơ sở hướng của người dùng hoặc hướng chuyển động của người dùng hoặc sự tương tác của người dùng trong cảnh âm thanh.

50. Hệ thống theo điểm 33, trong đó công nhìn liên quan đến vị trí hoặc vị trí ảo hoặc dữ liệu chuyển động hoặc hướng đầu.

51. Hệ thống theo điểm 33, trong đó các thành phần âm thanh khác nhau được cung cấp ở các cổng nhìn khác nhau, trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận, trong trường hợp một thành phần âm thanh thứ nhất nằm trong cổng nhìn, thành phần âm thanh thứ nhất tại tốc độ bit cao hơn so với thành phần âm thanh thứ hai không nằm trong cổng nhìn.

52. Hệ thống theo điểm 33, được tạo cấu hình để yêu cầu hoặc nhận các luồng âm thanh thứ nhất và các luồng âm thanh thứ hai, trong đó các thành phần âm thanh thứ nhất trong luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe thấy hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai,

trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai.

53. Hệ thống theo điểm 33, trong đó hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng.

54. Hệ thống theo điểm 33, được tạo cấu hình để vô hiệu hóa việc giải mã các thành phần âm thanh được chọn trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc vị trí ảo hiện tại của người dùng.

55. Hệ thống theo điểm 33, được tạo cấu hình để gộp ít nhất một luồng âm thanh thứ nhất liên quan đến cảnh âm thanh hiện tại với ít nhất một luồng liên quan đến cảnh âm thanh lân cận, liền kề hoặc tương lai.

56. Hệ thống theo điểm 33, được tạo cấu hình để thu hoặc thu thập dữ liệu thống kê hoặc tổng hợp trên cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để truyền yêu cầu đến máy chủ liên quan đến dữ liệu thống kê hoặc tổng hợp.

57. Hệ thống theo điểm 33, được tạo cấu hình để vô hiệu hóa phép giải mã hoặc tái tạo ít nhất một luồng trên cơ sở siêu dữ liệu liên quan đến ít nhất một luồng và trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng.

58. Hệ thống theo điểm 33, còn được tạo cấu hình để:

thao tác siêu dữ liệu liên quan đến nhóm các luồng âm thanh đã chọn, dựa trên ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để:

chọn hoặc kích hoạt các thành phần âm thanh soạn cảnh âm thanh sẽ được tái tạo; hoặc

gộp tất cả các luồng âm thanh thành một luồng âm thanh duy nhất.

59. Hệ thống theo điểm 33, trong đó thông tin được cung cấp từ máy chủ, đối với từng thành phần âm thanh hoặc đối tượng âm thanh, trong đó thông tin chứa thông tin mô tả về các vị trí mà cảnh âm thanh hoặc các thành phần âm thanh đang hoạt động.

60. Hệ thống theo điểm 33, được tạo cấu hình để lựa chọn giữa việc tái tạo một cảnh âm thanh và biên soạn hoặc trộn hoặc ghép hoặc chồng hoặc kết hợp ít nhất hai cảnh âm thanh trên cơ sở cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc vị trí ảo hoặc lựa chọn của người dùng hiện tại hoặc tương lai, hai cảnh âm thanh liên quan đến các môi trường video lân cận hoặc liền kề khác nhau.

61. Hệ thống theo điểm 33, được tạo cấu hình để tạo hoặc sử dụng ít nhất một tập hợp điều chỉnh sao cho:

một số tập hợp điều chỉnh được liên kết với một cảnh âm thanh; hoặc

thông tin bổ sung được cung cấp liên quan từng tập hợp điều chỉnh với một điểm nhìn hoặc một cảnh âm thanh; hoặc

thông tin bổ sung được cung cấp chứa

- thông tin về các ranh giới của một cảnh âm thanh, hoặc

- thông tin về mối quan hệ giữa một tập hợp điều chỉnh và một cảnh âm thanh, trong đó cảnh âm thanh được mã hóa thành ba luồng được gói gọn thành ba tập hợp điều chỉnh, hoặc

- thông tin về kết nối giữa các ranh giới của cảnh âm thanh và nhiều tập hợp điều chỉnh.

62. Hệ thống theo điểm 33, được tạo cấu hình để:

nhận luồng cho cảnh âm thanh liên quan đến môi trường lân cận hoặc liền kề;

bắt đầu giải mã hoặc tái tạo luồng âm thanh cho môi trường lân cận hoặc liền kề khi phát hiện sự chuyển tiếp ranh giới giữa hai môi trường.

63. Hệ thống theo điểm 33, trong đó hệ thống còn được tạo cấu hình để:

yêu cầu hoặc nhận ít nhất một tập hợp điều chỉnh thứ nhất bao gồm ít nhất một luồng âm thanh liên quan đến ít nhất một cảnh âm thanh thứ nhất;

yêu cầu hoặc nhận ít nhất một tập hợp điều chỉnh thứ hai bao gồm ít nhất một luồng âm thanh thứ hai liên quan đến ít nhất hai cảnh âm thanh, chứa ít nhất một cảnh âm thanh thứ nhất; và

gộp ít nhất một luồng âm thanh thứ nhất và ít nhất một luồng âm thanh thứ hai thành luồng âm thanh mới để được giải mã, dựa trên siêu dữ liệu có sẵn về công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc lựa chọn hiện tại của người dùng hoặc thông tin mô tả sự liên kết của ít nhất một tập hợp điều chỉnh thứ nhất với ít nhất một cảnh âm thanh thứ nhất hoặc sự liên kết của ít nhất một tập hợp điều chỉnh thứ hai với ít nhất một lần cảnh âm thanh thứ nhất.

64. Hệ thống theo điểm 33, được tạo cấu hình để:

nhận thông tin về công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng hoặc bất kỳ thông tin đặc trưng cho các thay đổi nào được kích hoạt bởi các hành động của người dùng; và

nhận thông tin về tính khả dụng của các tập hợp điều chỉnh và thông tin mô tả sự liên kết của ít nhất một tập hợp điều chỉnh với ít nhất một cảnh âm thanh hoặc điểm nhìn hoặc công nhìn hoặc vị trí hoặc vị trí ảo hoặc dữ liệu chuyển động hoặc hướng.

65. Hệ thống theo điểm 33, được tạo cấu hình để:

quyết định xem có ít nhất một thành phần âm thanh từ ít nhất một cảnh âm thanh được lồng vào ít nhất một luồng âm thanh và ít nhất một thành phần âm thanh bổ sung từ ít nhất một cảnh âm thanh bổ sung được lồng vào ít nhất một luồng âm thanh bổ sung sẽ được tái tạo; và

thực hiện, trong trường hợp quyết định là có, thao tác gộp hoặc biên soạn hoặc ghép hoặc chồng hoặc kết hợp ít nhất một luồng bổ sung của cảnh âm thanh bổ sung vào ít nhất một luồng của ít nhất một cảnh âm thanh.

66. Hệ thống theo điểm 33 được tạo cấu hình để:

thao tác siêu dữ liệu âm thanh liên quan đến các luồng âm thanh đã chọn, dựa trên ít nhất là công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hiện tại của người dùng, để:

chọn hoặc cho phép hoặc kích hoạt các thành phần âm thanh soạn cảnh âm thanh được quyết định sẽ tái tạo; và

cho phép gộp tất cả các luồng âm thanh được chọn thành một luồng âm thanh duy nhất.

67. Hệ thống phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống bao gồm:

ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh được tạo cấu hình để giải mã các tín hiệu âm thanh từ các luồng âm thanh cho sự biểu diễn các cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu các luồng âm thanh thứ nhất và các luồng âm thanh thứ hai hoặc một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh cho máy chủ trên cơ sở ít nhất là công nhìn hoặc hướng đầu

hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng,

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai,

trong đó ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường video thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường video thứ hai,

trong đó hệ thống được tạo cấu hình để thu được siêu dữ liệu tương tác mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với môi trường video thứ hai,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ nhất,

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ hai, và

trong đó hệ thống được tạo cấu hình để yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa môi trường video thứ nhất và môi trường video thứ hai.

68. Hệ thống theo điểm 67, trong đó ít nhất một thành phần âm thanh thứ nhất được cung cấp trong ít nhất một luồng âm thanh hoặc tập hợp điều chỉnh, và ít nhất một thành phần âm thanh thứ hai được cung cấp trong ít nhất một luồng âm thanh thứ hai hoặc tập hợp điều chỉnh, và ít nhất một thành phần âm thanh thứ ba được cung cấp trong ít nhất một luồng âm thanh thứ ba hoặc tập hợp điều chỉnh, và trong đó ít nhất một môi trường video thứ nhất được mô tả bởi siêu dữ liệu tương tác dưới dạng cảnh âm thanh mà có thể sử dụng ít nhất một luồng âm thanh thứ nhất và thứ hai hoặc tập hợp điều chỉnh, và trong đó môi trường video thứ hai được mô tả bởi siêu dữ liệu

tương tác dưới dạng cảnh âm thanh mà có thể sử dụng ít nhất một luồng âm thanh thứ ba hoặc tập hợp điều chỉnh, và ít nhất một luồng âm thanh thứ hai hoặc tập hợp điều chỉnh được liên kết với ít nhất một môi trường video thứ nhất,

trong đó hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu tương tác, để gộp luồng âm thanh thứ hai thuộc môi trường video thứ nhất và luồng âm thanh thứ ba được liên kết với môi trường video thứ hai thành một luồng đơn mới, trong trường hợp vị trí ảo của người dùng là trong môi trường video thứ hai.

69. Hệ thống phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó hệ thống bao gồm:

ít nhất một bộ giải mã video phương tiện được tạo cấu hình để giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng, và

ít nhất một bộ giải mã âm thanh được tạo cấu hình để giải mã các tín hiệu âm thanh từ các luồng âm thanh cho sự biểu diễn các cảnh âm thanh,

trong đó hệ thống được tạo cấu hình để yêu cầu các luồng âm thanh thứ nhất và các luồng âm thanh thứ hai hoặc một thành phần âm thanh của luồng âm thanh hoặc một tập hợp điều chỉnh cho máy chủ trên cơ sở ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng,

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai,

trong đó hệ thống bao gồm bộ xử lý siêu dữ liệu được tạo cấu hình để thao tác siêu dữ liệu trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng,

trong đó bộ xử lý siêu dữ liệu được tạo cấu hình để kích hoạt hoặc vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh, dựa trên cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng, trong đó

trong đó bộ xử lý siêu dữ liệu được tạo cấu hình để vô hiệu hóa ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh trong trường hợp hệ thống quyết định rằng thành phần âm thanh đó không được tái tạo nữa do kết quả của cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng, và trong đó

trong đó bộ xử lý siêu dữ liệu được tạo cấu hình để kích hoạt ít nhất một thành phần âm thanh trong ít nhất một luồng âm thanh trước ít nhất một bộ giải mã âm thanh trong trường hợp hệ thống quyết định rằng thành phần âm thanh đó sẽ được tái tạo do kết quả của cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng.

70. Máy chủ để phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ, các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó máy chủ bao gồm bộ mã hóa để mã hóa hoặc bộ nhớ để lưu trữ các luồng video để mô tả môi trường video, môi trường video được liên kết với cảnh âm thanh;

trong đó máy chủ còn bao gồm bộ mã hóa để mã hóa hoặc bộ nhớ để lưu trữ nhiều luồng hoặc các thành phần âm thanh hoặc các tập hợp điều chỉnh sẽ được phân phối đến máy khách, các luồng âm thanh hoặc các thành phần âm thanh hoặc các tập hợp điều chỉnh được liên kết với ít nhất một cảnh âm thanh,

trong đó máy chủ được tạo cấu hình để:

chọn và phân phối luồng video trên cơ sở yêu cầu từ máy khách, luồng video liên quan đến môi trường;

chọn luồng âm thanh hoặc thành phần âm thanh hoặc tập hợp điều chỉnh trên cơ sở yêu cầu từ máy khách, yêu cầu liên quan đến ít nhất là cổng nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ

liệu vị trí ảo hiện tại của người dùng và liên quan đến cảnh âm thanh liên quan đến môi trường; và

phân phối luồng âm thanh đến máy khách,

trong đó yêu cầu dựa trên khoảng cách của vị trí người dùng từ các ranh giới của môi trường video lân cận hoặc liền kề liên quan đến các cảnh âm thanh khác nhau,

trong đó ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường video thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường video thứ hai,

trong đó máy chủ được tạo cấu hình để cung cấp siêu dữ liệu tương tác mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với môi trường video thứ hai,

trong đó máy chủ được tạo cấu hình để cung cấp ít nhất một thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ nhất,

trong đó máy chủ được tạo cấu hình để cung cấp ít nhất một thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ hai, và

trong đó máy chủ được tạo cấu hình để cung cấp ít nhất một thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa môi trường video thứ nhất và môi trường video thứ hai.

71. Máy chủ theo điểm 70, trong đó các luồng âm thanh được gói gọn thành các tập hợp điều chỉnh, mỗi tập hợp điều chỉnh bao gồm nhiều luồng liên quan đến các sự biểu diễn khác nhau, ở tốc độ bit khác nhau, của cùng một nội dung âm thanh,

trong đó tập hợp điều chỉnh đã chọn được chọn trên cơ sở yêu cầu từ máy khách.

72. Máy chủ để phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ, các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện,

trong đó máy chủ bao gồm bộ mã hóa để mã hóa hoặc lưu trữ các luồng video để mô tả môi trường video, môi trường video được liên kết với cảnh âm thanh;

trong đó máy chủ còn bao gồm bộ mã hóa để mã hóa hoặc lưu trữ nhiều luồng âm thanh hoặc các thành phần âm thanh hoặc các tập hợp điều chỉnh sẽ được phân phối đến máy khách, các luồng âm thanh hoặc các thành phần âm thanh hoặc các tập hợp điều chỉnh được liên kết với ít nhất một cảnh âm thanh,

trong đó máy chủ được tạo cấu hình để:

chọn và phân phối luồng video trên cơ sở yêu cầu từ máy khách, luồng video liên quan đến môi trường;

chọn luồng âm thanh hoặc thành phần âm thanh hoặc tập hợp điều chỉnh trên cơ sở yêu cầu từ máy khách, yêu cầu liên quan đến ít nhất là công nhìn hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu tương tác hoặc dữ liệu vị trí ảo hiện tại của người dùng và liên quan đến cảnh âm thanh được liên kết với môi trường; và

phân phối luồng âm thanh đến máy khách,

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai,

trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video bao gồm các môi trường video liên kế hoặc lân cận,

sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh âm thanh thứ nhất, và trong trường hợp người dùng chuyển sang cảnh âm thanh thứ hai, để cung cấp cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất ở tốc độ bit cao hơn khi người dùng ở trong môi trường thứ nhất được liên kết với cảnh âm thanh thứ nhất,

trong khi các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai ở tốc độ bit thấp hơn khi người dùng bắt đầu ở vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất ở tốc độ bit thấp hơn và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai ở tốc độ bit cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai,

trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn

73. Máy chủ theo điểm 72, trong đó các luồng âm thanh được gói gọn thành các tập hợp điều chỉnh, mỗi tập hợp điều chỉnh bao gồm nhiều luồng liên quan đến các sự biểu diễn khác nhau, ở tốc độ bit khác nhau, của cùng một nội dung âm thanh,

trong đó tập hợp điều chỉnh đã chọn được chọn trên cơ sở yêu cầu từ máy khách.

74. Phương pháp phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện, phương pháp bao gồm các bước:

giải mã các tín hiệu video từ các luồng video để biểu diễn các môi trường VR, AR, MR hoặc video 360 độ cho người dùng,

giải mã các tín hiệu âm thanh từ các luồng âm thanh để biểu diễn các cảnh âm thanh,

yêu cầu hoặc lấy từ máy chủ ít nhất một luồng âm thanh trên cơ sở công nhìn hoặc dữ liệu vị trí hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc siêu dữ liệu hiện tại của người dùng, và

kiểm soát yêu cầu của ít nhất một luồng âm thanh đến máy chủ trên cơ sở khoảng cách vị trí của người dùng từ các ranh giới của các môi trường video lân cận hoặc liên kết liên quan đến các cảnh âm thanh khác nhau

trong đó ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường video thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường video thứ hai,

trong đó phương pháp bao gồm bước thu siêu dữ liệu tương tác mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với môi trường video thứ hai,

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ nhất,

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ hai, và

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa môi trường video thứ nhất và môi trường video thứ hai.

75. Phương pháp phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện, phương pháp bao gồm các bước:

giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng,

giải mã các tín hiệu âm thanh từ các luồng âm thanh để biểu diễn các cảnh âm thanh,

yêu cầu hoặc lấy từ máy chủ ít nhất một luồng âm thanh trên cơ sở công nhìn hoặc dữ liệu vị trí hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc siêu dữ liệu hiện tại của người dùng, và

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai

trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video bao gồm các môi trường video liên kế hoặc lân cận,

sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh âm thanh thứ nhất, và trong trường hợp người dùng chuyển sang cảnh âm thanh thứ hai, để cung cấp cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit cao hơn khi người dùng ở trong môi trường thứ nhất được liên kết với cảnh âm thanh thứ nhất,

trong khi các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai được thu ở tốc độ bit thấp hơn khi người dùng bắt đầu ở vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit thấp hơn và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được thu ở tốc độ bit cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn.

76. Phương tiện lưu trữ số không tạm thời có chương trình máy tính được lưu trữ trên đó để thực hiện phương pháp phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện, phương pháp bao gồm các bước:

giải mã các tín hiệu video từ các luồng video để biểu diễn các môi trường VR, AR, MR hoặc video 360 độ cho người dùng,

giải mã các tín hiệu âm thanh từ các luồng âm thanh để biểu diễn các cảnh âm thanh,

yêu cầu hoặc lấy từ máy chủ ít nhất một luồng âm thanh trên cơ sở công nhìn hoặc dữ liệu vị trí hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc siêu dữ liệu hiện tại của người dùng, và

kiểm soát yêu cầu của ít nhất một luồng âm thanh đến máy chủ trên cơ sở khoảng cách vị trí của người dùng từ các ranh giới của các môi trường video lân cận hoặc liên kết liên quan đến các cảnh âm thanh khác nhau

trong đó ít nhất hai cảnh môi trường trực quan được xác định, trong đó ít nhất một thành phần âm thanh thứ nhất và thứ hai liên quan đến cảnh âm thanh thứ nhất được liên kết với môi trường video thứ nhất, và ít nhất một thành phần âm thanh thứ ba liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường video thứ hai,

trong đó phương pháp bao gồm bước thu siêu dữ liệu tương tác mô tả rằng ít nhất một thành phần âm thanh thứ hai được liên kết bổ sung với môi trường video thứ hai,

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ nhất,

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng ở trong môi trường video thứ hai, và

trong đó phương pháp bao gồm bước yêu cầu hoặc nhận ít nhất một thành phần âm thanh thứ nhất và thứ hai và thứ ba, trong trường hợp vị trí ảo của người dùng đang chuyển đổi giữa môi trường video thứ nhất và môi trường video thứ hai.

77. Phương tiện lưu trữ số không tạm thời có chương trình máy tính được lưu trữ trên đó để thực hiện phương pháp phân phối các luồng âm thanh và video đến máy khách cho các môi trường thực tế ảo (VR), thực tế tăng cường (AR), thực tế hỗn hợp (MR), hoặc video 360 độ được tạo cấu hình để nhận các luồng video và âm thanh được tái tạo trong thiết bị tiêu dùng phương tiện, phương pháp bao gồm các bước:

giải mã các tín hiệu video từ các luồng video để biểu diễn VR, AR, MR hoặc các môi trường video 360 độ cho người dùng,

giải mã các tín hiệu âm thanh từ các luồng âm thanh để biểu diễn các cảnh âm thanh,

yêu cầu hoặc lấy từ máy chủ ít nhất một luồng âm thanh trên cơ sở công nhìn hoặc dữ liệu vị trí hoặc hướng đầu hoặc dữ liệu chuyển động hoặc siêu dữ liệu hoặc dữ liệu vị trí ảo hoặc siêu dữ liệu hiện tại của người dùng, và

trong đó các thành phần âm thanh thứ nhất trong các luồng âm thanh thứ nhất có liên quan hơn hoặc dễ nghe hơn các thành phần âm thanh thứ hai trong các luồng

âm thanh thứ hai, trong đó các luồng âm thanh thứ nhất được yêu cầu hoặc nhận được ở tốc độ bit cao hơn tốc độ bit của các luồng âm thanh thứ hai

trong đó nhiều cảnh âm thanh được xác định cho nhiều môi trường video bao gồm các môi trường video liền kề hoặc lân cận,

sao cho các luồng thứ nhất được cung cấp liên quan đến cảnh âm thanh thứ nhất, và trong trường hợp người dùng chuyển sang cảnh âm thanh thứ hai, để cung cấp cả các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai,

trong đó các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit cao hơn khi người dùng ở trong môi trường thứ nhất được liên kết với cảnh âm thanh thứ nhất,

trong khi các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được liên kết với môi trường thứ hai được thu ở tốc độ bit thấp hơn khi người dùng bắt đầu ở vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, và

các luồng thứ nhất liên quan đến cảnh âm thanh thứ nhất được thu ở tốc độ bit thấp hơn và các luồng thứ hai liên quan đến cảnh âm thanh thứ hai được thu ở tốc độ bit cao hơn khi người dùng ở cuối vị trí chuyển tiếp từ cảnh âm thanh thứ nhất sang cảnh âm thanh thứ hai, trong đó tốc độ bit thấp hơn thấp hơn tốc độ bit cao hơn.

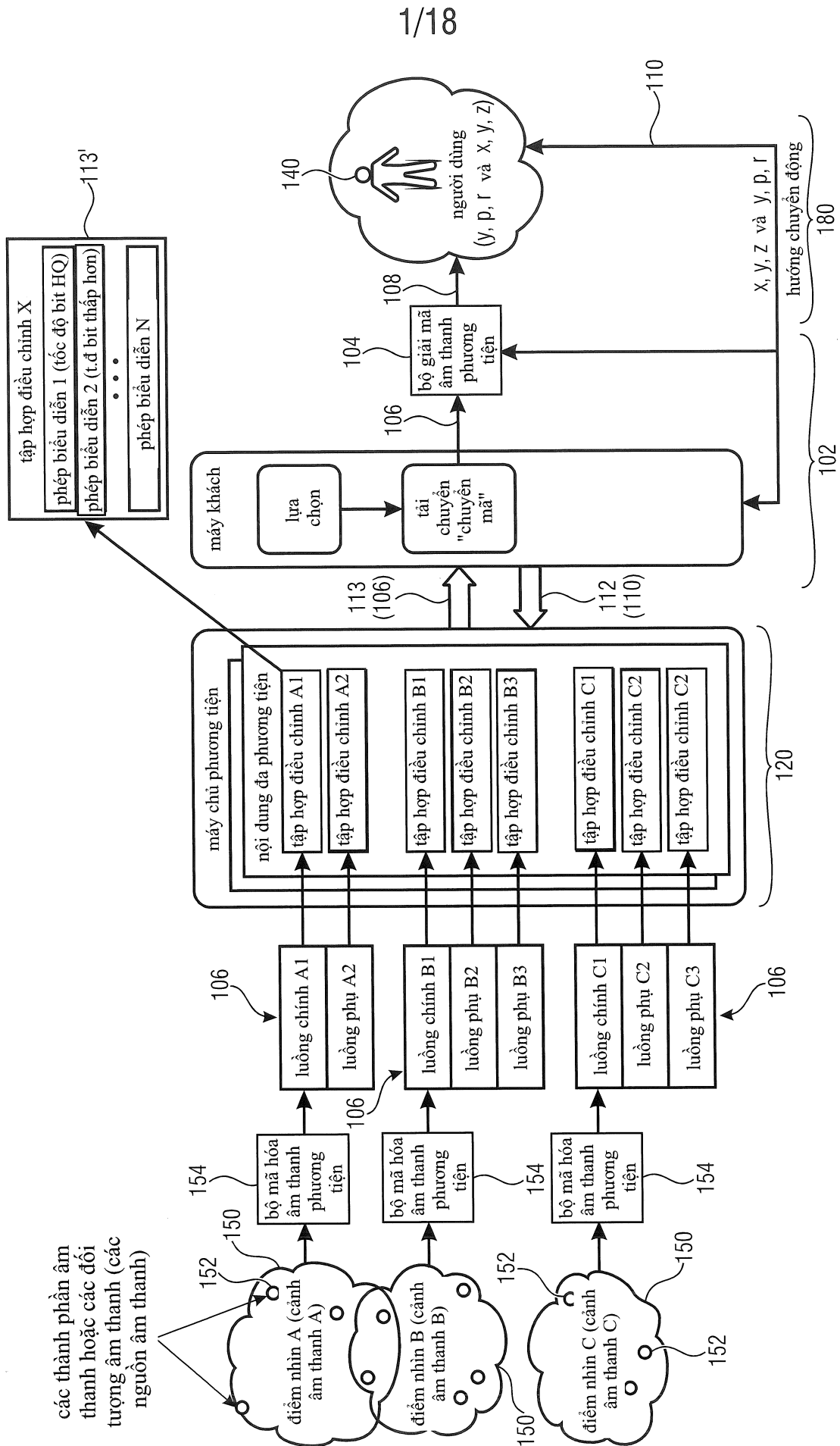


Fig. 1.1

2/18

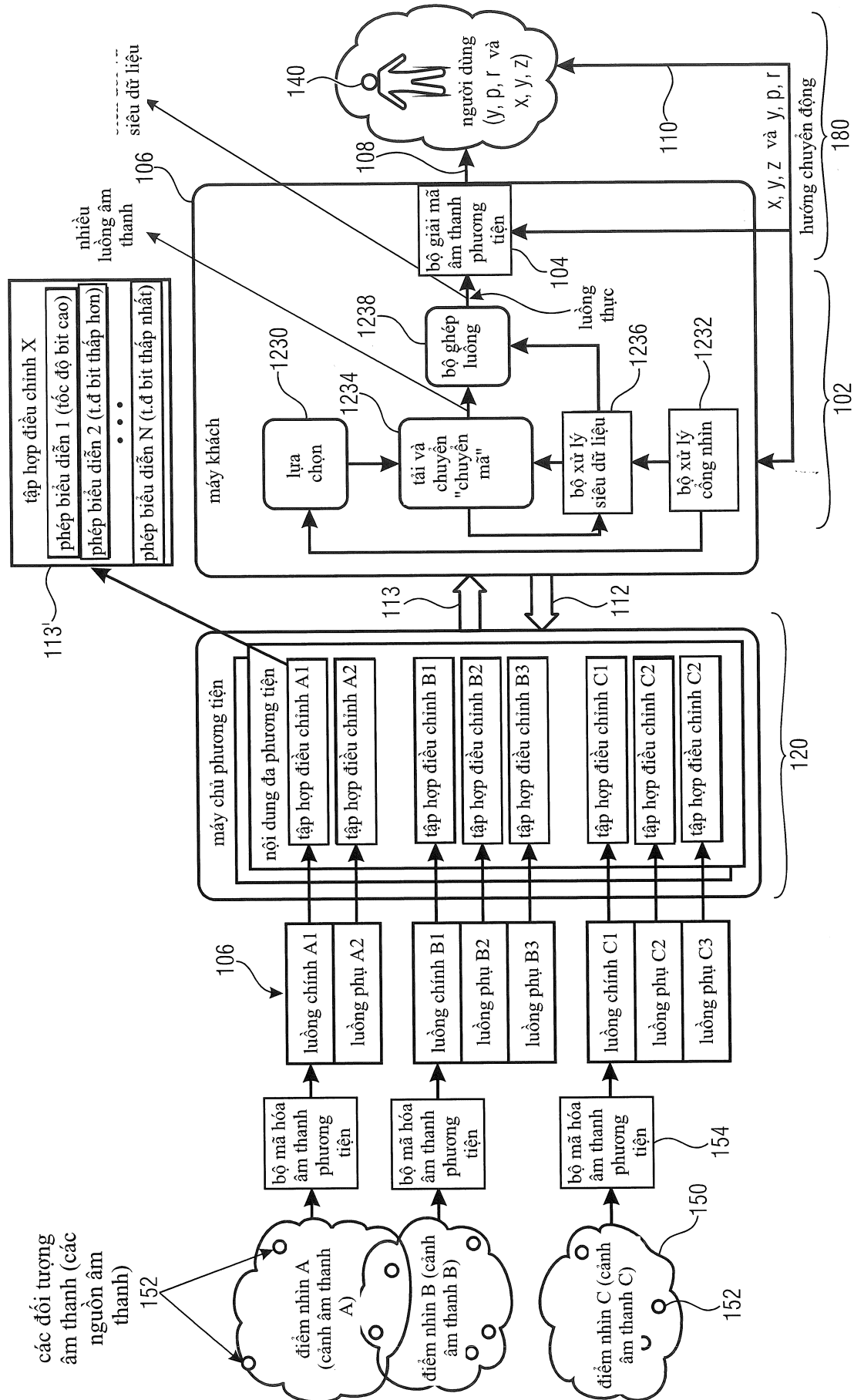


Fig. 1.2

3/18

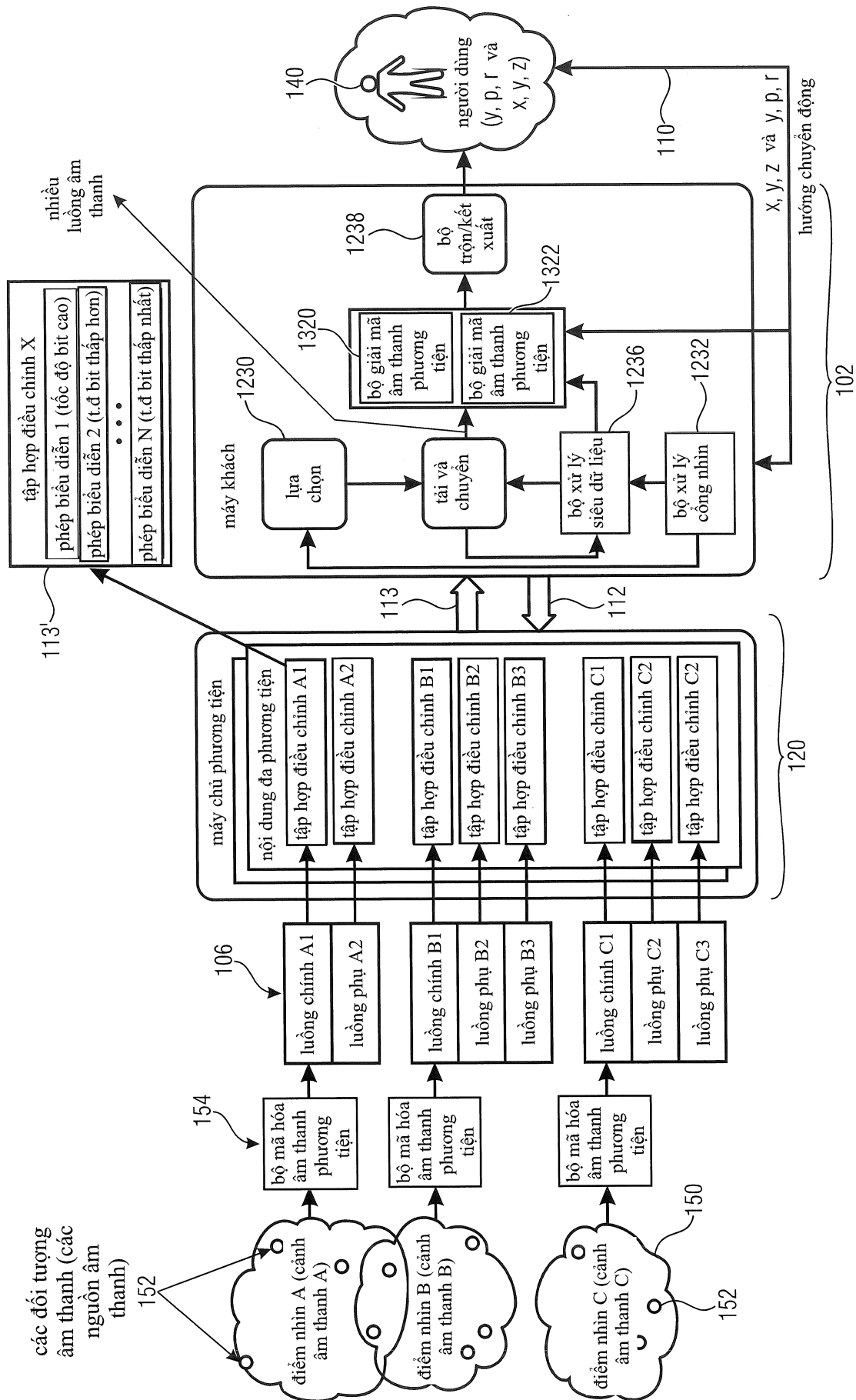


Fig. 1.3

4/18

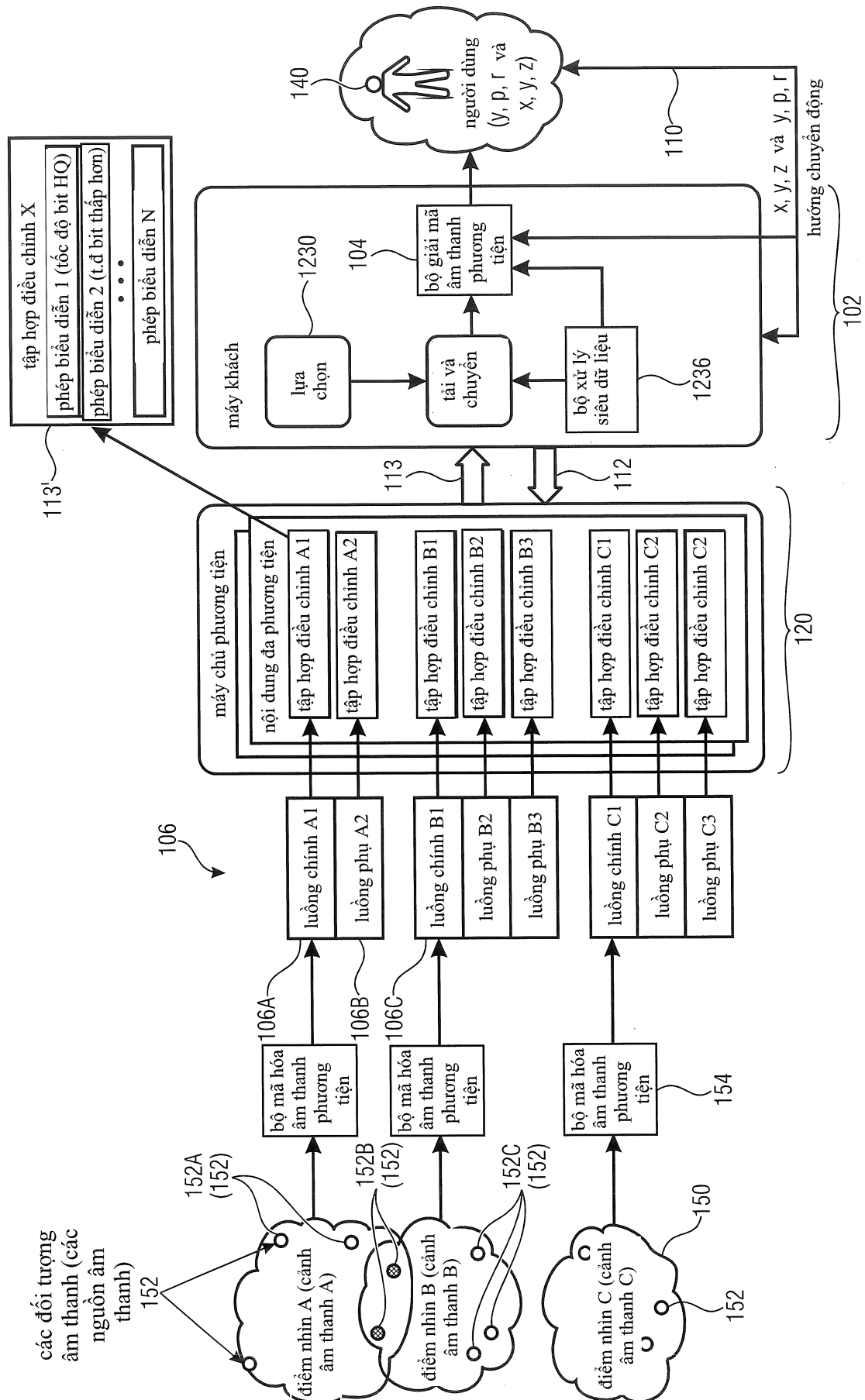


Fig. 1.4

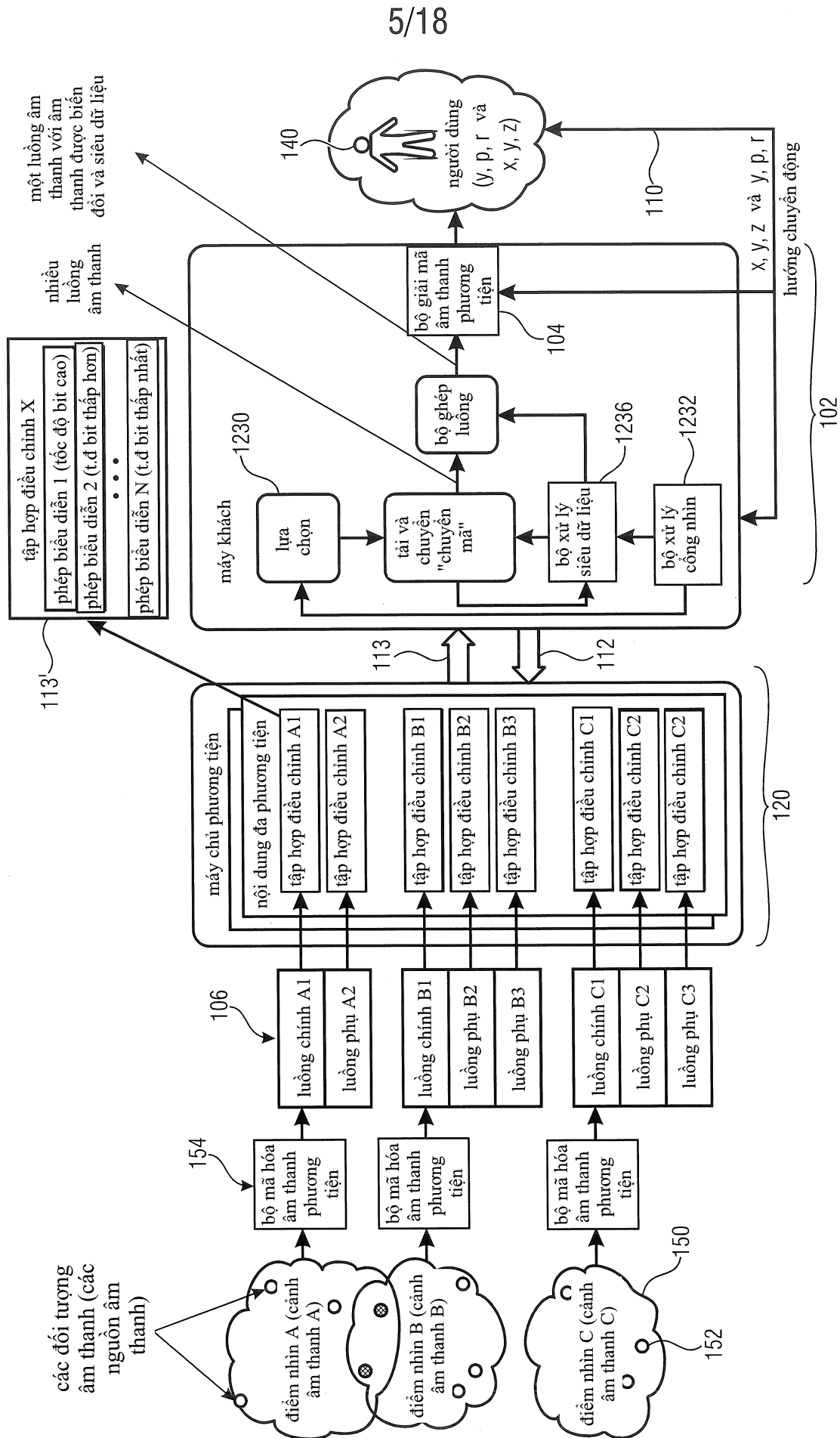


Fig. 1.5

6/18

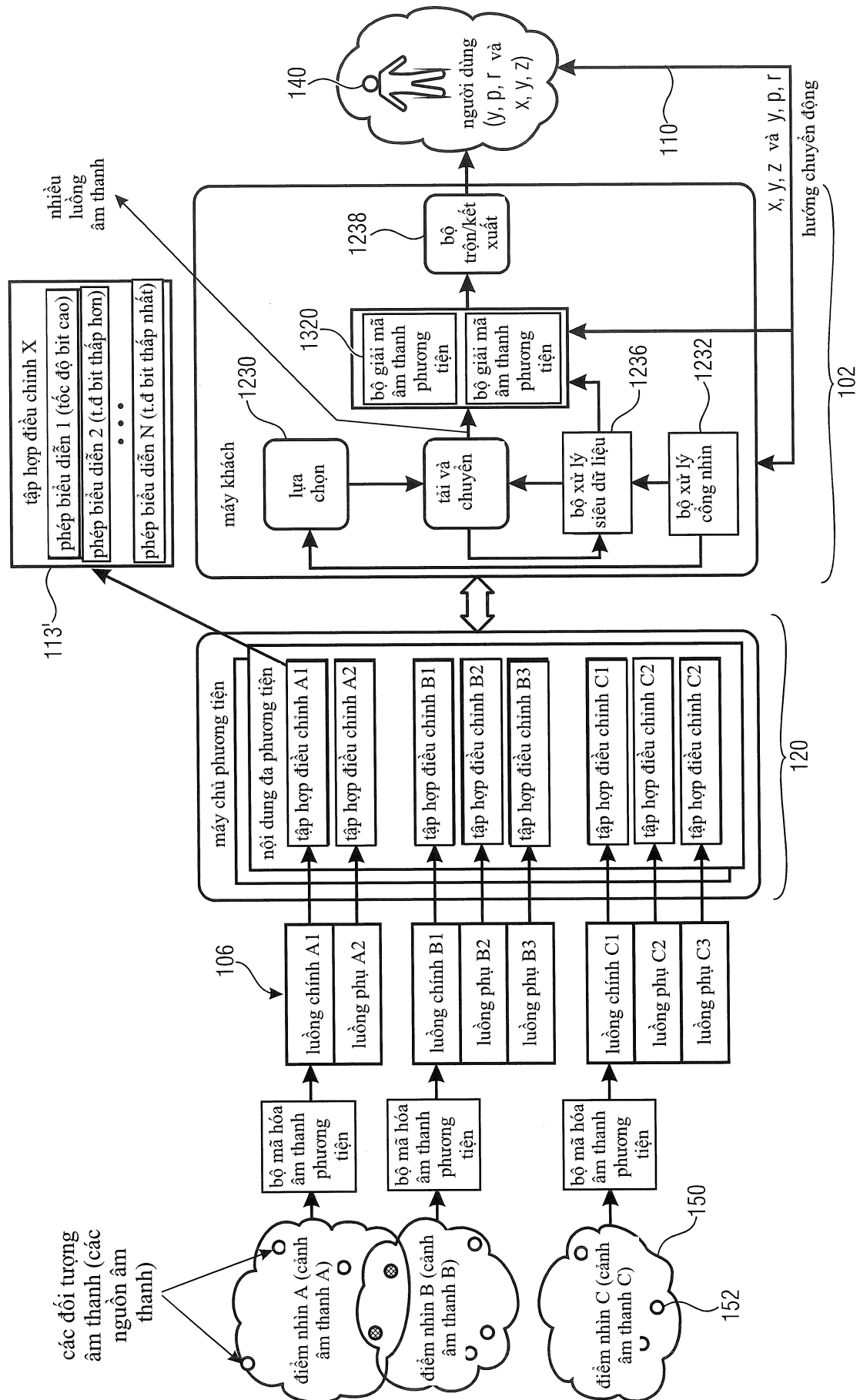


Fig. 1.6

7/18

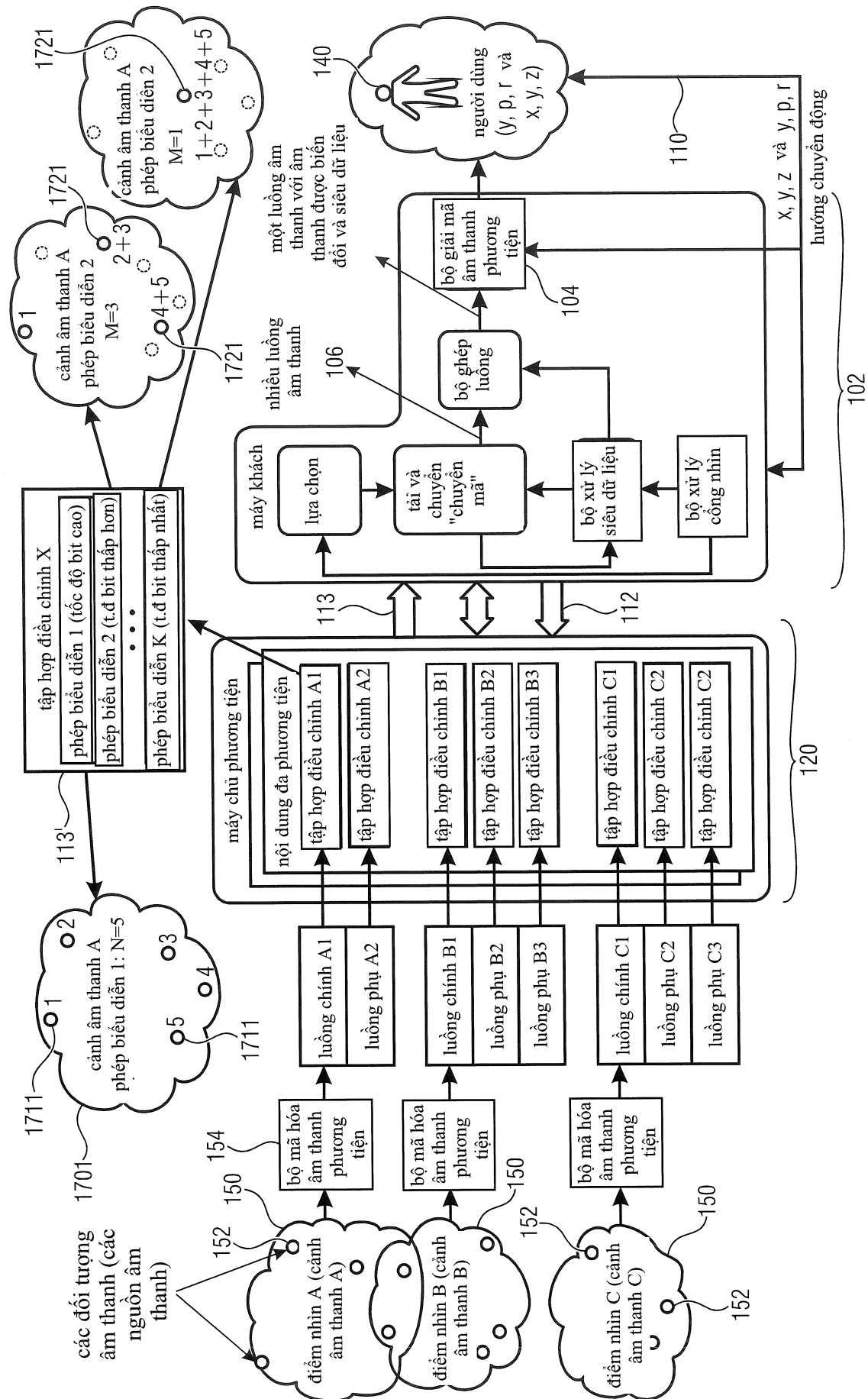


Fig. 1.7

8/18

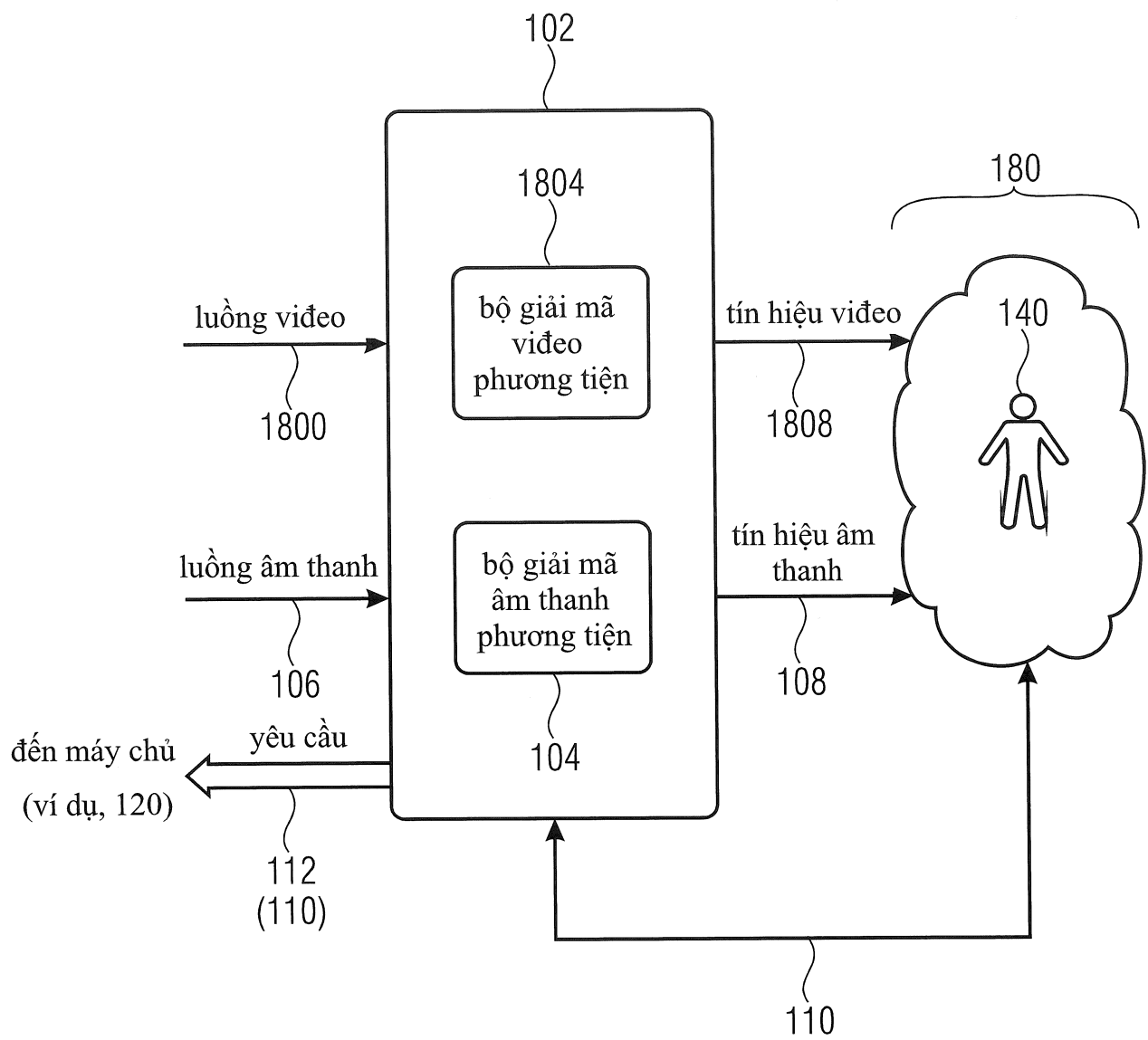


Fig. 1.8

9/18

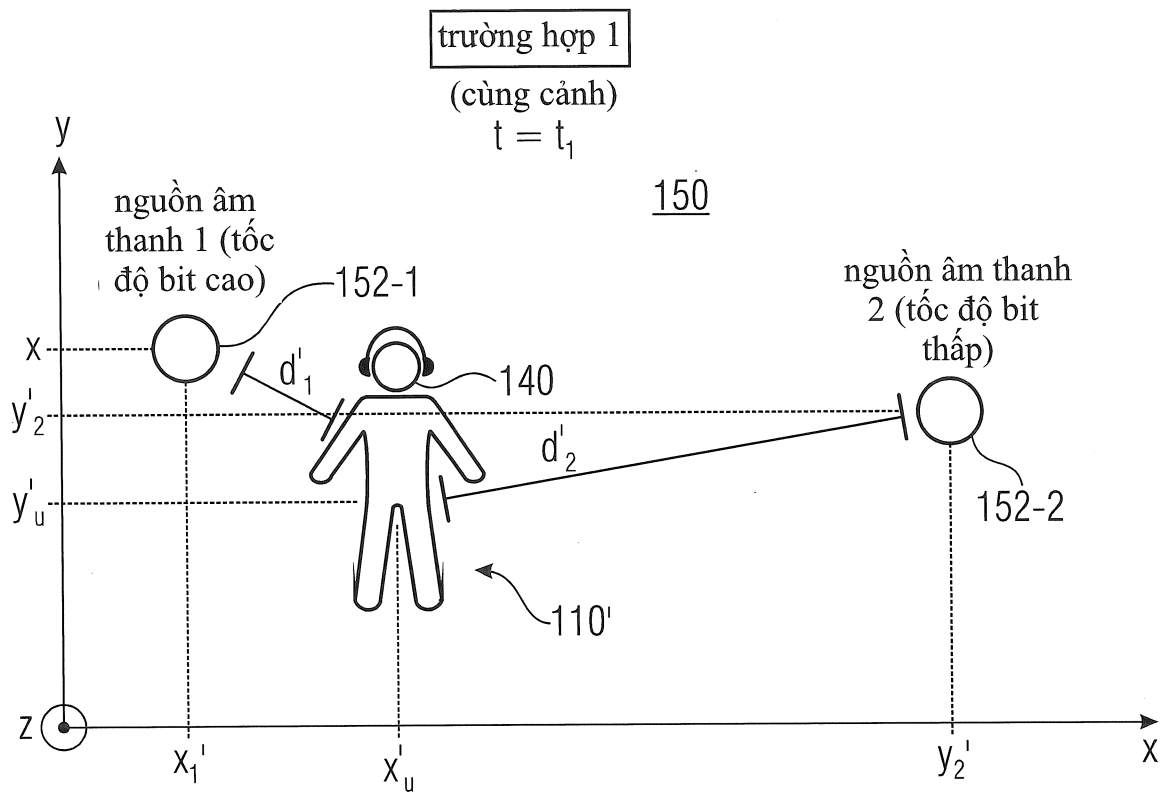


Fig. 2a

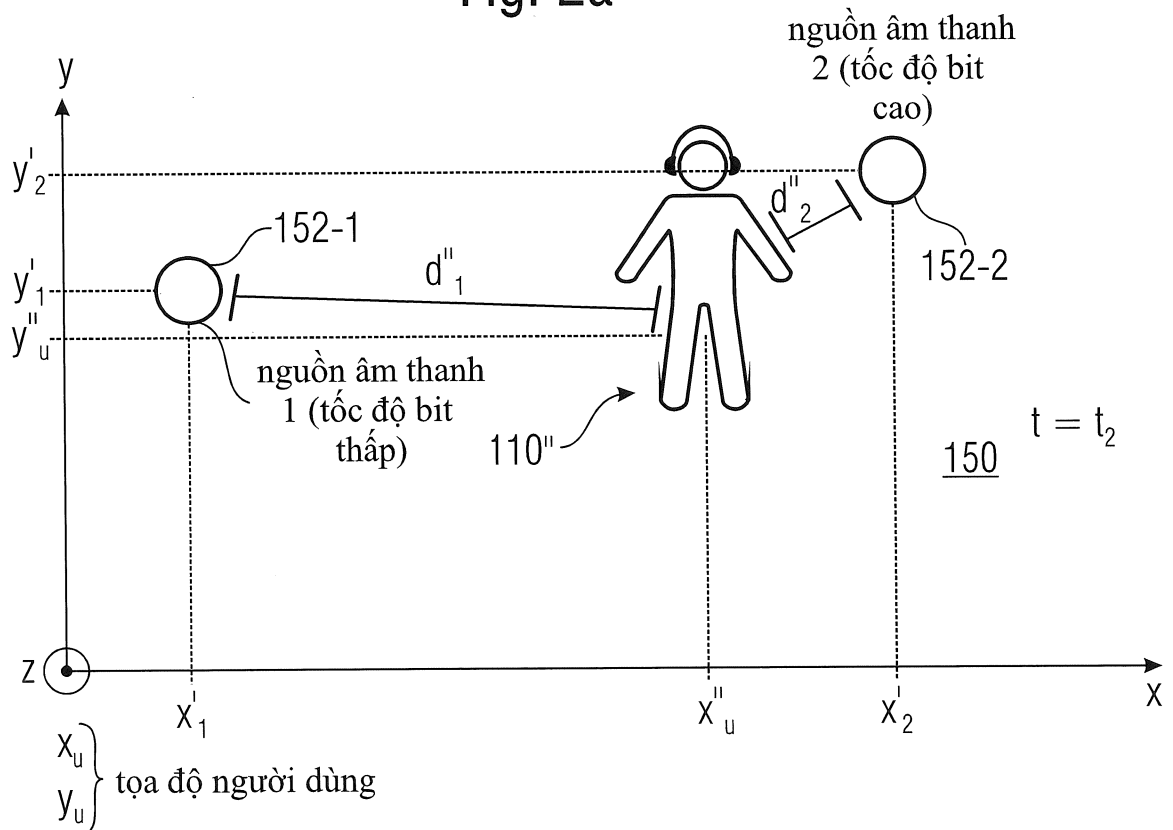


Fig. 2b

10/18

trường hợp 2

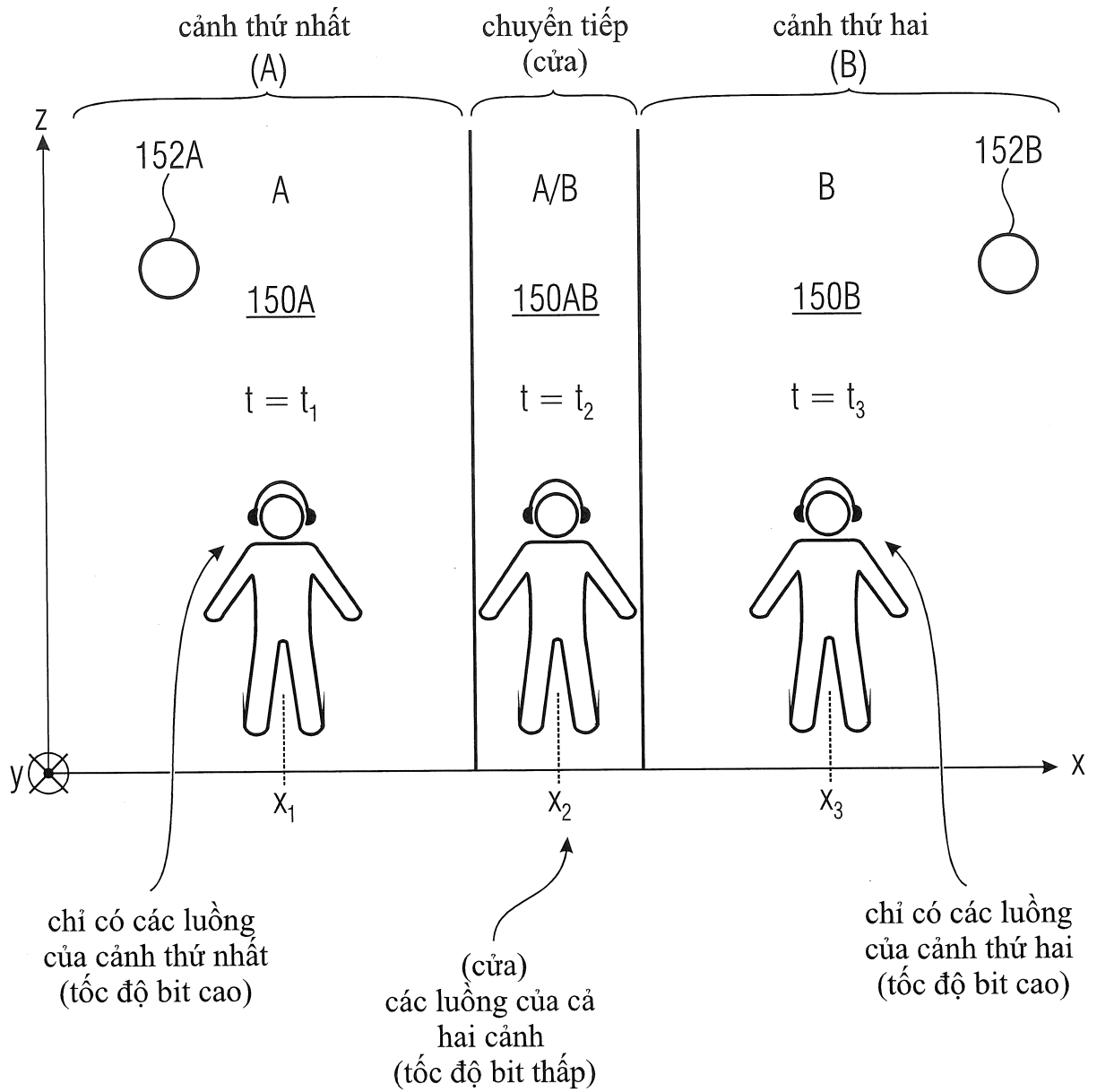


Fig. 3

11/18

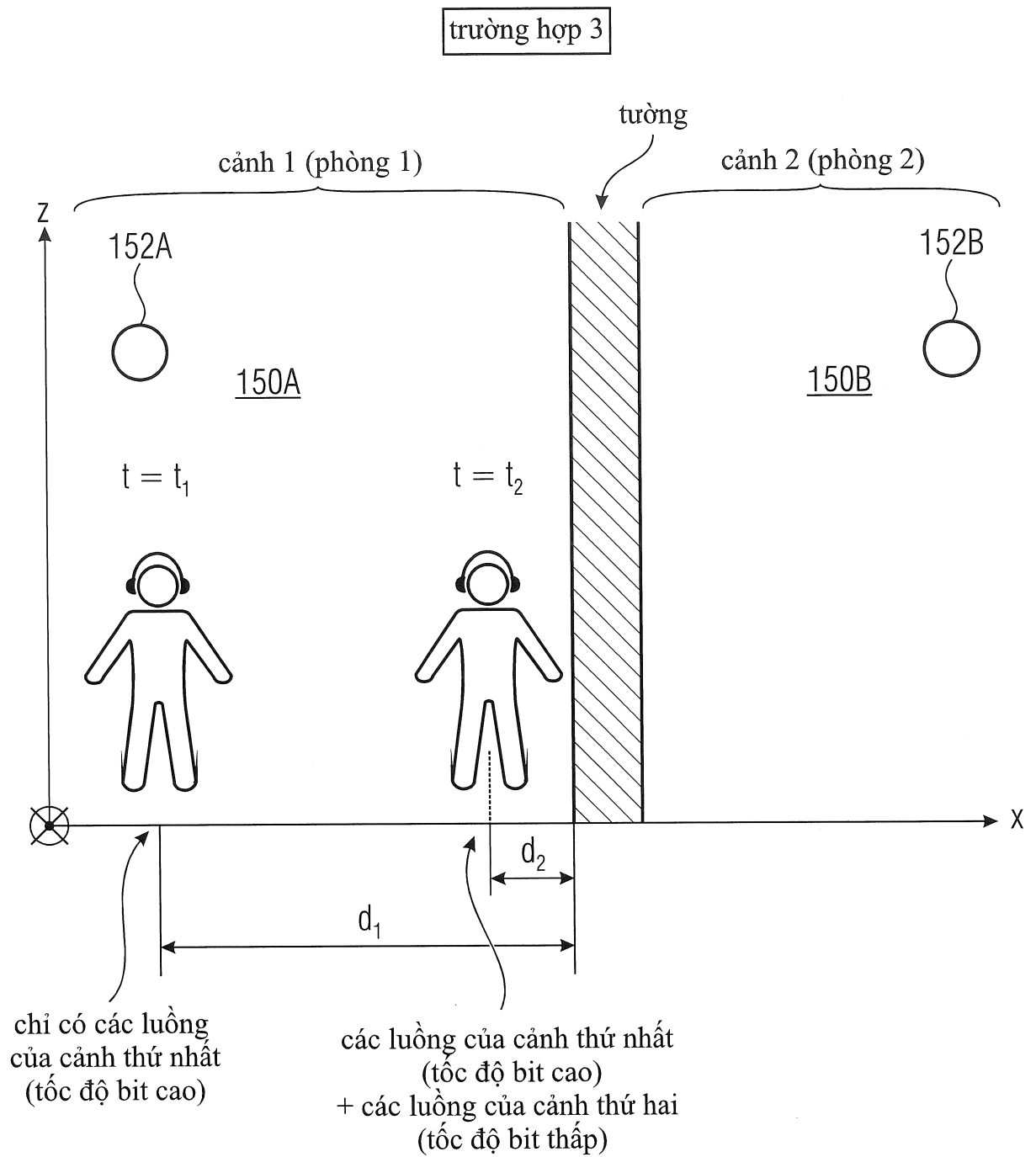


Fig. 4

12/18

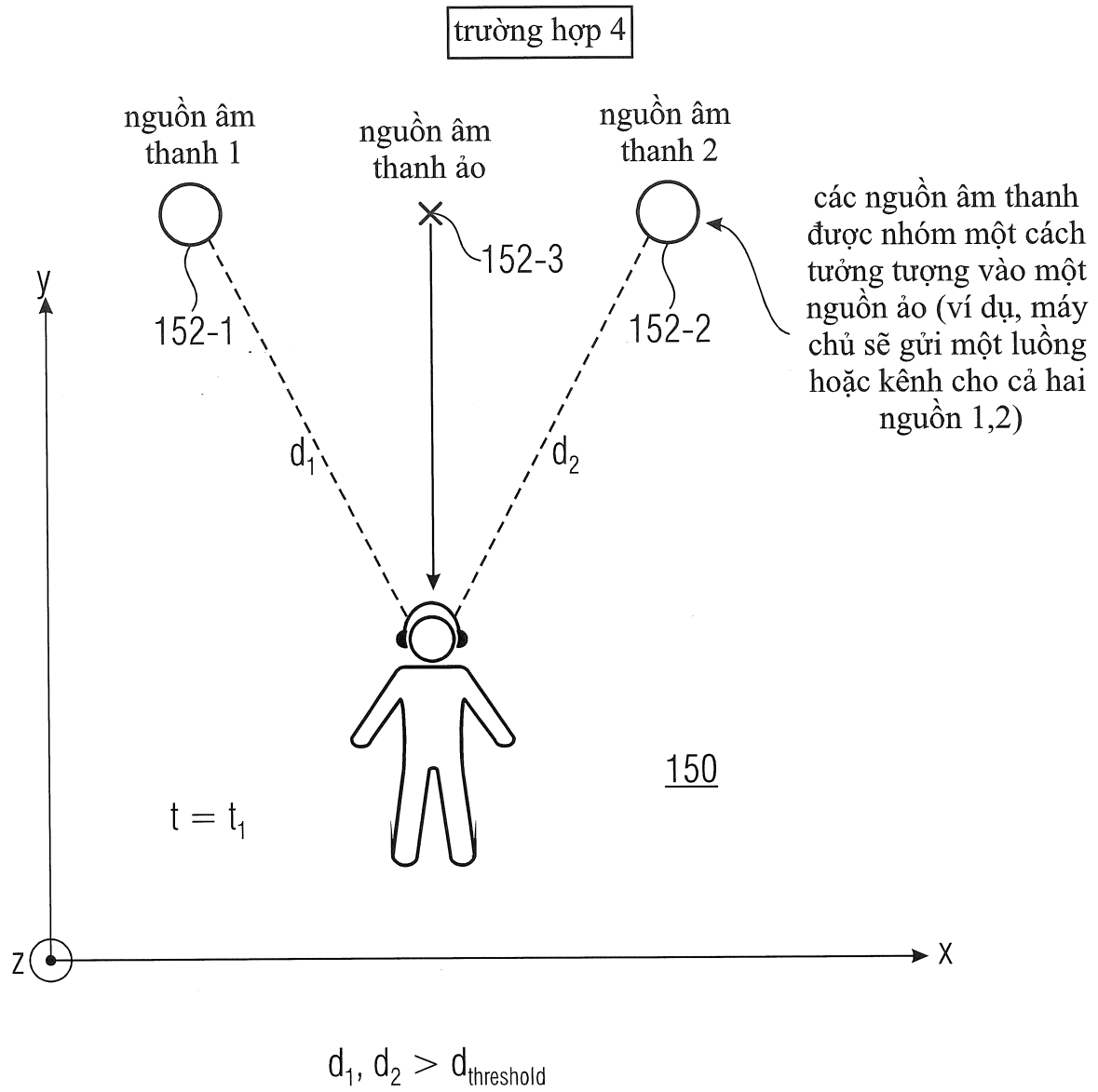


Fig. 5a

13/18

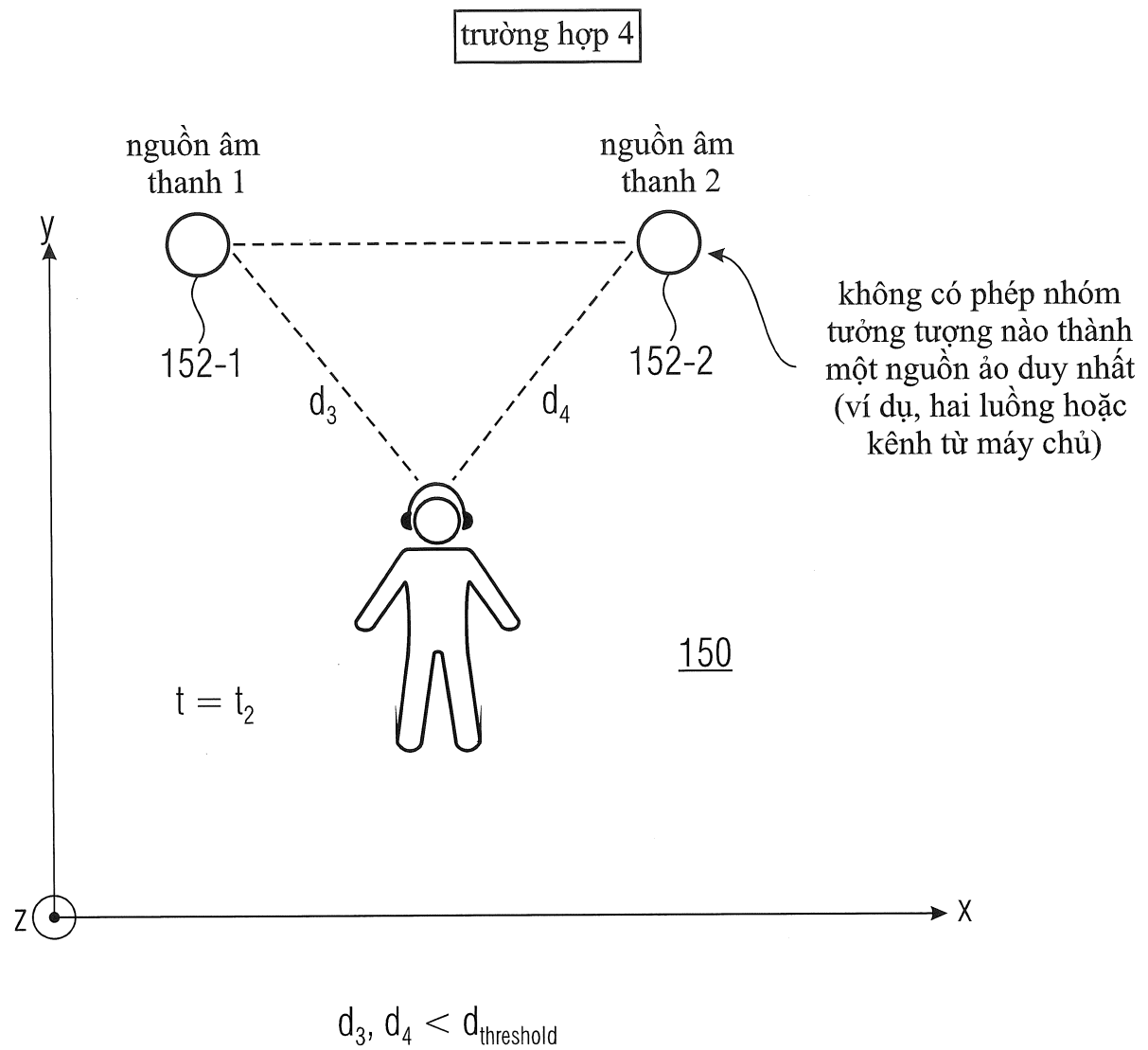


Fig. 5b

14/18

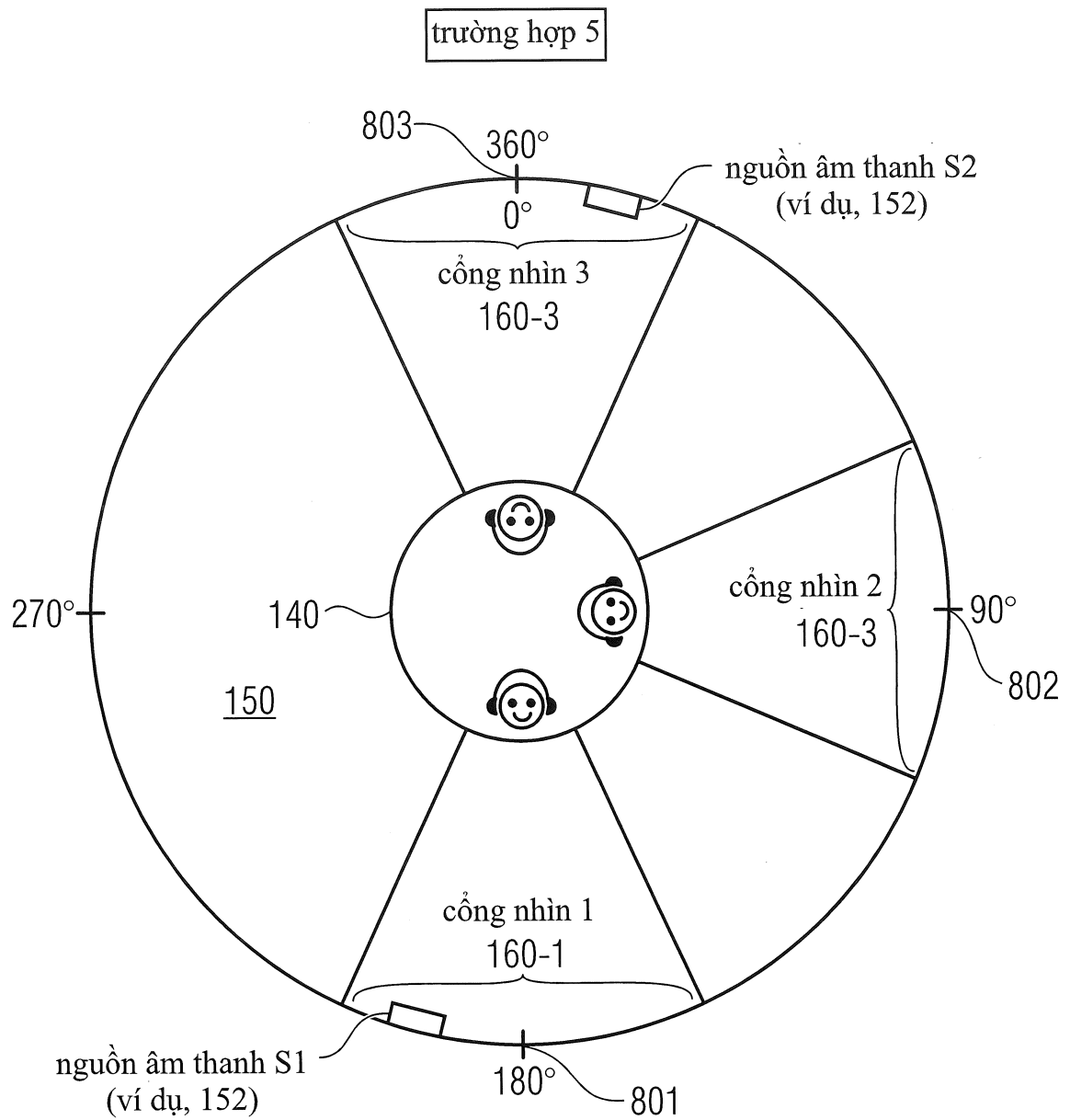


Fig. 6

15/18

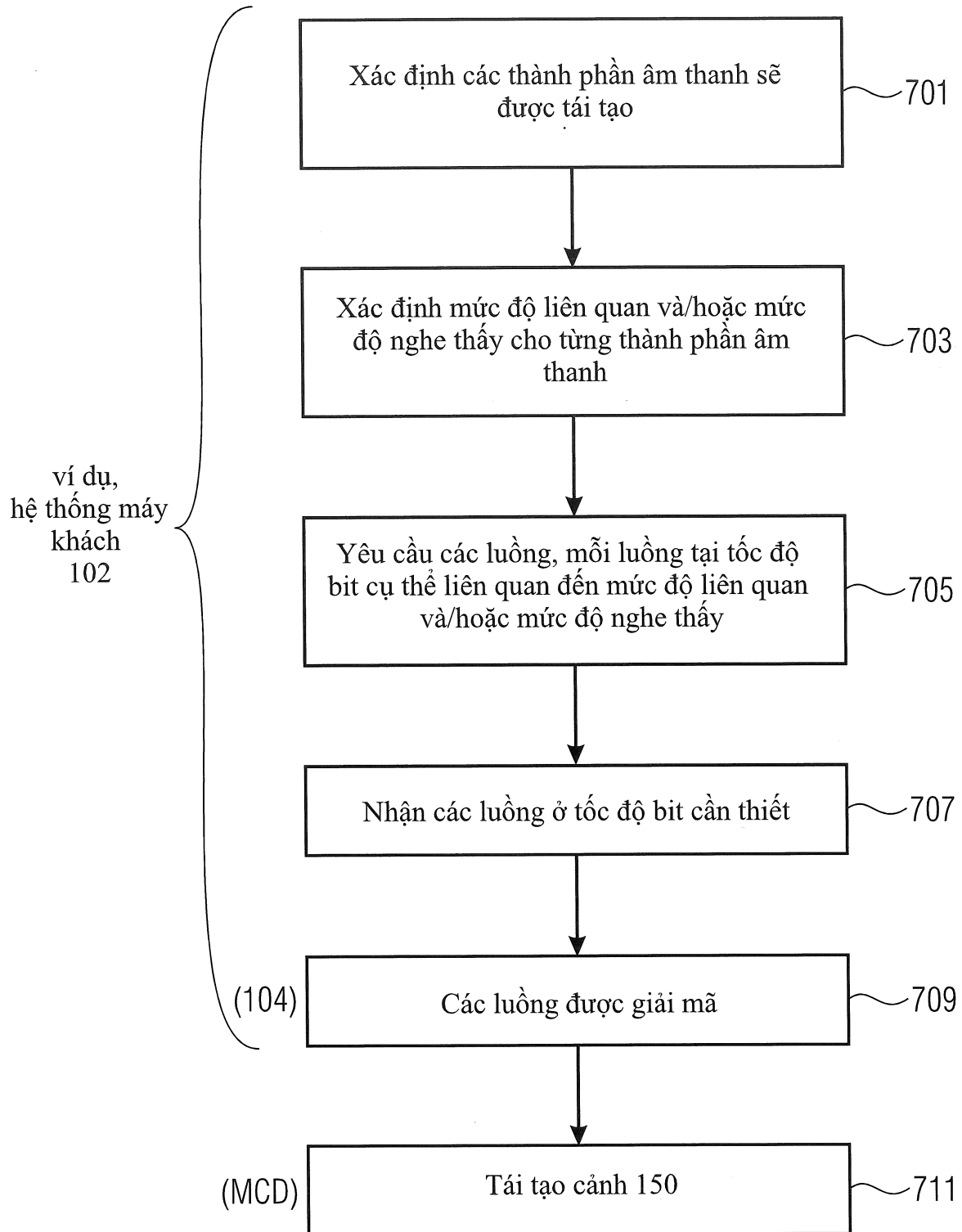


Fig. 7a

16/18

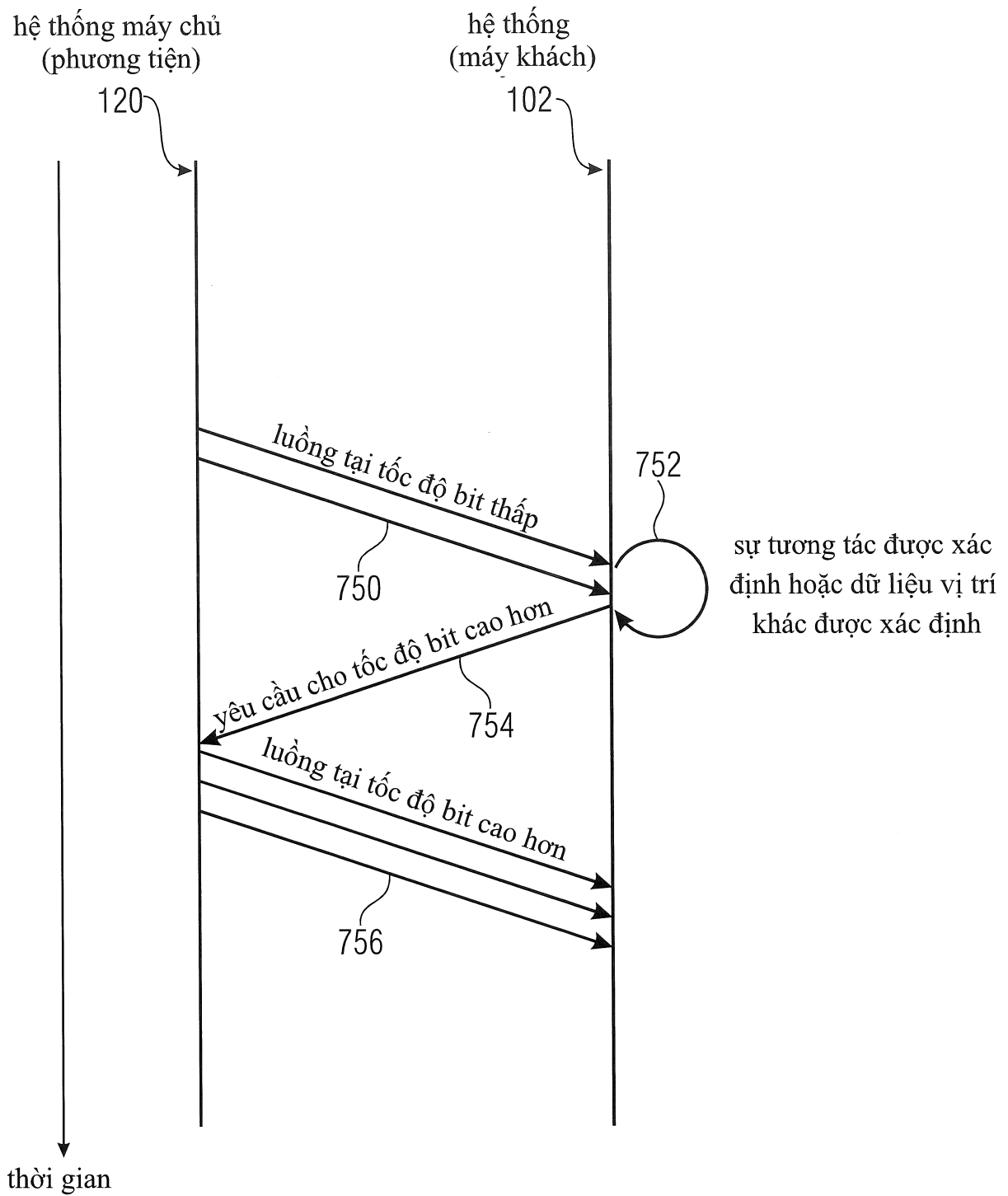


Fig. 7b

17/18

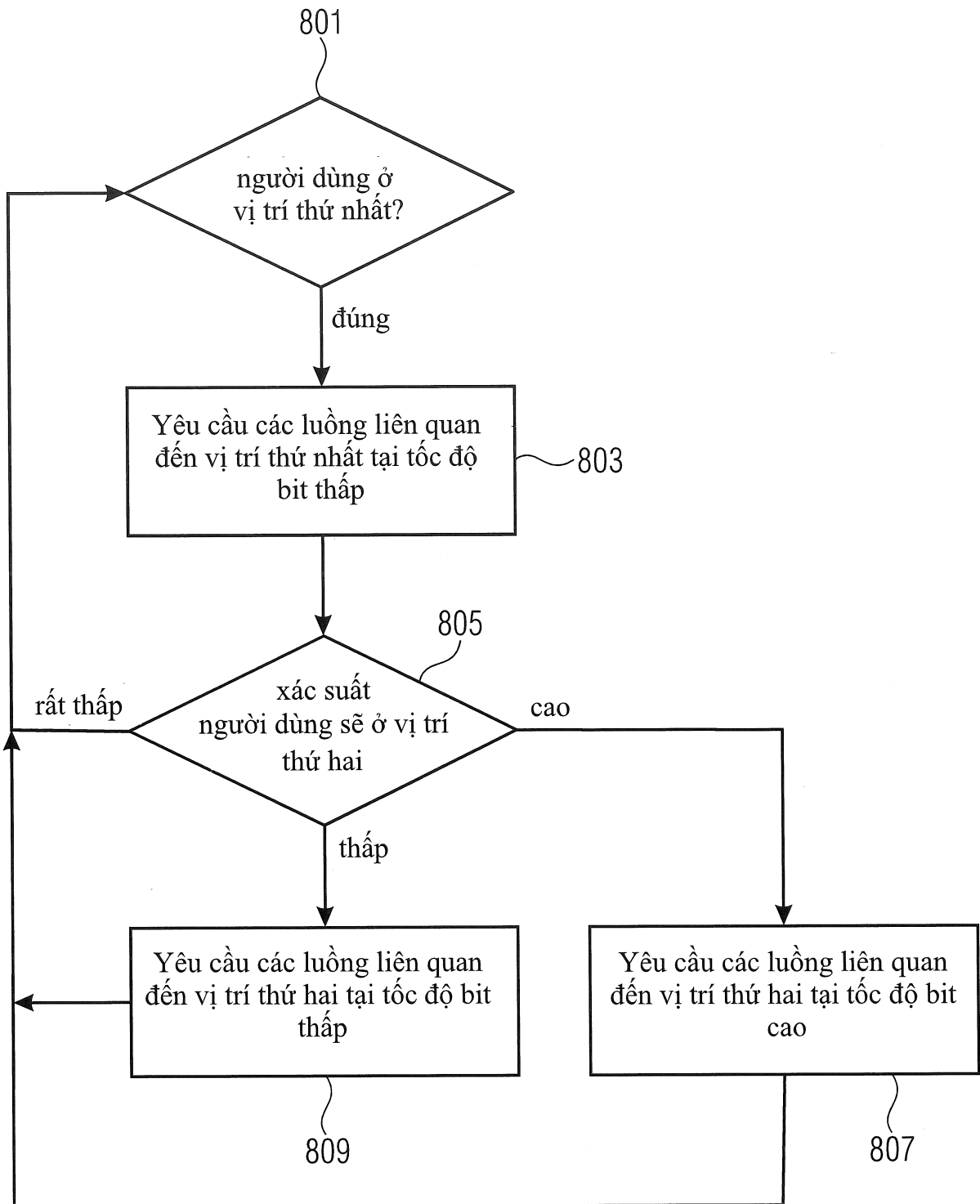


Fig. 8a

18/18

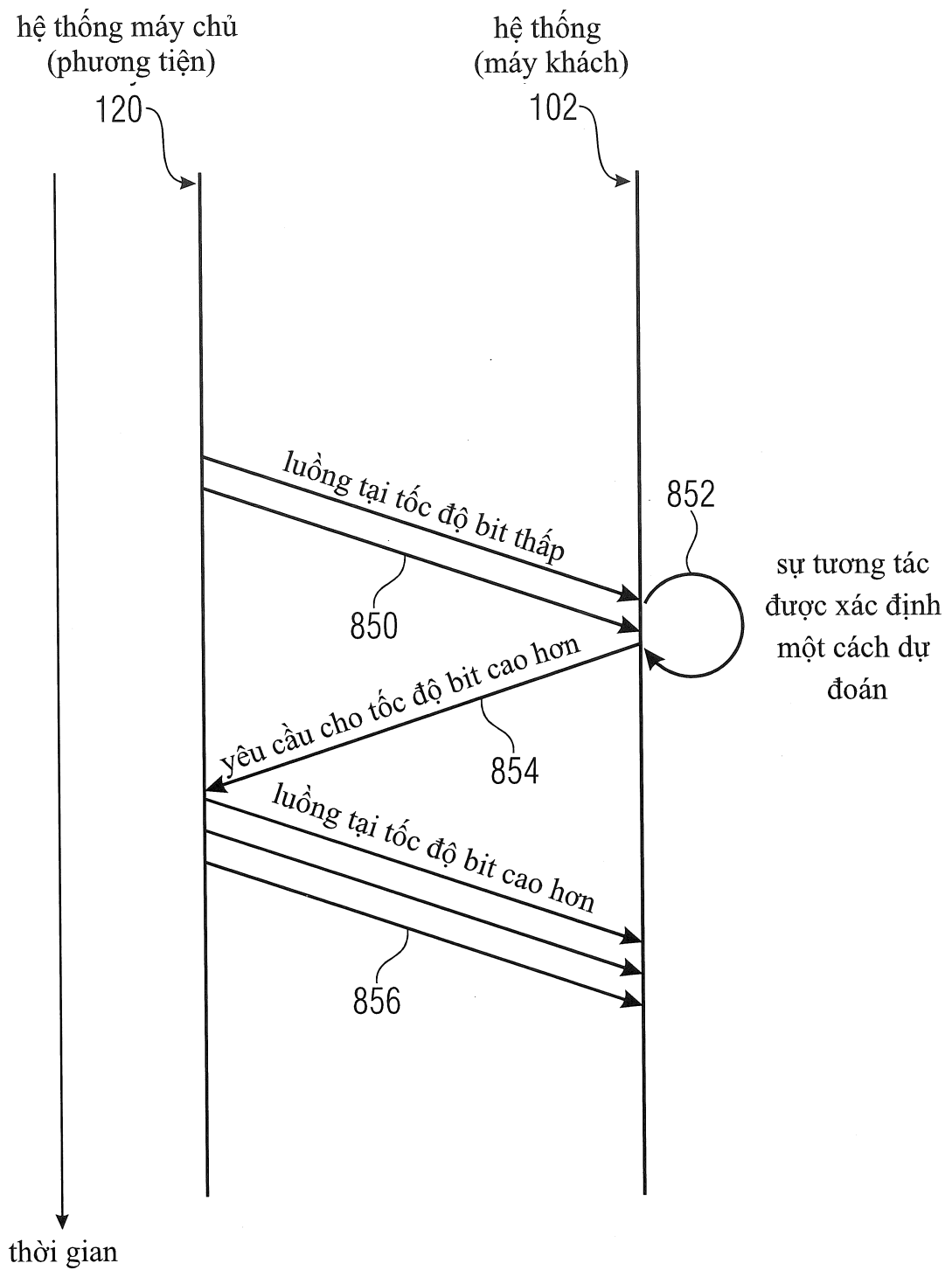


Fig. 8b