



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

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

(11)



1-0048844

(51)^{2020.01} H04S 7/00; G06F 3/01

(13) B

(21) 1-2020-05239

(22) 09/04/2019

(86) PCT/EP2019/058954 09/04/2019

(87) WO2019/197403 A1 17/10/2019

(30) 62/654,915 09/04/2018 US; 62/695,446 09/07/2018 US; 62/823,159 25/03/2019 US

(45) 25/07/2025 448

(43) 25/06/2021 399A

(73) DOLBY INTERNATIONAL AB (NL)

Apollo Building, 3E, Herikerbergweg 1-35, 1101 CN Amsterdam Zuidoost,
Netherlands

(72) FERSCH, Christof (DE); TERENTIV, Leon (DE); FISCHER, Daniel (DE).

(74) Công ty TNHH Quốc tế D & N (D&N INTERNATIONAL CO.,LTD.)

(54) PHƯƠNG PHÁP VÀ BỘ GIẢI MÃ ÂM THANH ĐỂ XỬ LÝ THÔNG TIN VỊ TRÍ
CHỈ BÁO VỊ TRÍ ĐỐI TƯỢNG CỦA ĐỐI TƯỢNG ÂM THANH VÀ PHƯƠNG
TIỆN ĐỌC ĐƯỢC BẰNG MÁY TÍNH

(21) 1-2020-05239

(57) Sáng chế đề cập đến phương pháp xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh, trong đó vị trí đối tượng có thể sử dụng để kết xuất đối tượng âm thanh, phương pháp này bao gồm các bước: thu nhận thông tin định hướng người nghe biểu thị sự định hướng của đầu người nghe; thu nhận thông tin dịch chuyển người nghe biểu thị sự dịch chuyển của đầu người nghe; xác định vị trí đối tượng từ thông tin vị trí; sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng; và sửa đổi thêm vị trí đối tượng được sửa đổi dựa vào thông tin định hướng người nghe. Sáng chế còn đề cập đến thiết bị tương ứng để xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh, trong đó vị trí đối tượng có thể sử dụng để kết xuất đối tượng âm thanh và hệ thống xử lý thông tin vị trí bao gồm các thiết bị này.

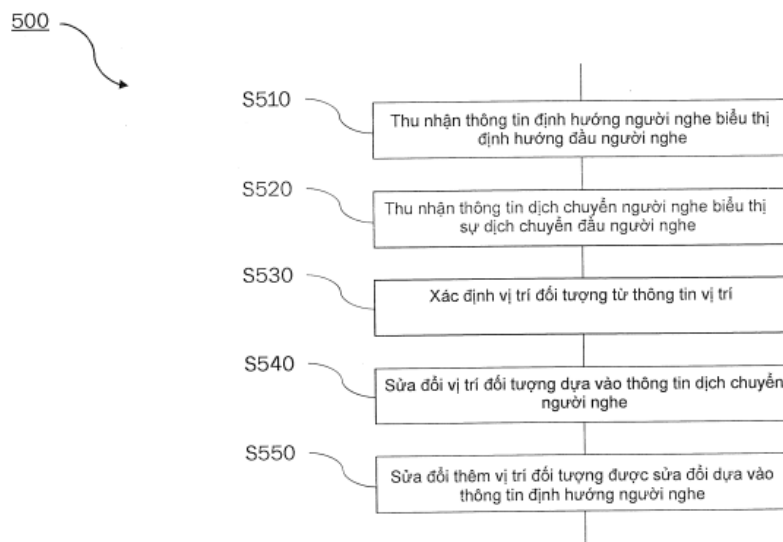


Fig.5

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp và thiết bị xử lý thông tin vị trí biểu thị vị trí đối tượng âm thanh, và thông tin biểu thị sự dịch chuyển vị trí của đầu người nghe.

Tình trạng kỹ thuật của sáng chế

Ấn bản đầu tiên (15 tháng 10 năm 2015) và các bản sửa đổi 1-4 của chuẩn âm 3D MPEG-H ISO/IEC 23008-3 không đưa ra để cho phép các chuyển động tịnh tiến nhỏ của đầu người dùng trong môi trường ba bậc tự do (Three Degrees of Freedom - 3DoF).

Bản chất kỹ thuật của sáng chế

Ấn bản đầu tiên (15 tháng 10 năm 2015) và các bản sửa đổi 1-4 của chuẩn âm 3D MPEG-H ISO/IEC 23008-3 cung cấp chức năng cho khả năng của môi trường 3DoF, trong đó người dùng (người nghe) thực hiện các hành động quay đầu. Tuy nhiên, chức năng này, ở mức tốt nhất chỉ hỗ trợ báo hiệu dịch chuyển cảnh quay và kết xuất tương ứng. Điều này có nghĩa là cảnh âm thanh có thể giữ cố định về mặt không gian dưới sự thay đổi theo định hướng đầu người nghe, điều này tương ứng với đặc tính 3DoF. Tuy nhiên, không có khả năng tính toán cho chuyển động tịnh tiến nhỏ của đầu người dùng trong hệ sinh thái âm thanh 3D MPEG-H hiện tại.

Do vậy, có nhu cầu về phương pháp và thiết bị xử lý thông tin vị trí của các đối tượng âm thanh mà có thể tính đến chuyển động tịnh tiến nhỏ của đầu người dùng, có khả năng kết hợp với chuyển động quay của đầu người dùng.

Sáng chế đề xuất thiết bị và hệ thống xử lý thông tin vị trí, có các dấu hiệu của các điểm yêu cầu bảo hộ độc lập và phụ thuộc tương ứng.

Theo một khía cạnh của sáng chế, sáng chế đề xuất phương pháp xử lý thông tin vị trí biểu thị vị trí đối tượng âm thanh, trong đó việc xử lý có thể phù hợp với chuẩn âm thanh 3D MPEG-H. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Đối tượng âm thanh có thể được chứa trong nội dung âm thanh dựa trên đối tượng, cùng với thông tin vị trí của nó. Thông tin vị trí có thể là (một phần của) siêu dữ liệu cho đối tượng âm thanh. Nội dung âm thanh (ví dụ, đối tượng âm thanh cùng với thông tin vị trí

của nó) có thể được truyền tải trong dòng bit âm thanh được mã hóa. Phương pháp có thể bao gồm bước nhận nội dung âm thanh (ví dụ, dòng bit âm thanh được mã hóa). Phương pháp có thể bao gồm bước thu được thông tin định hướng người nghe biểu thị sự định hướng của đầu người nghe. Người nghe có thể được gọi là người dùng, ví dụ, của bộ giải mã âm thanh thực hiện phương pháp. Sự định hướng của đầu người nghe (sự định hướng của người nghe) có thể là định hướng của đầu người nghe so với sự định hướng danh định. Phương pháp có thể còn bao gồm bước thu được thông tin dịch chuyển người nghe biểu thị sự dịch chuyển của đầu người nghe. Sự dịch chuyển của đầu người nghe có thể là sự dịch chuyển so với vị trí nghe danh định. Vị trí nghe danh định (hoặc vị trí người nghe danh định) có thể là vị trí mặc định (ví dụ, vị trí được xác định trước, vị trí dự tính cho đầu người nghe, hoặc điểm nghe hoàn hảo (sweet spot) theo sự bố trí loa). Thông tin định hướng người nghe và thông tin dịch chuyển người nghe có thể thu được thông qua giao diện đầu vào của bộ giải mã âm thanh 3D MPEG-H. Thông tin định hướng người nghe và thông tin dịch chuyển người nghe có thể được suy ra dựa vào thông tin cảm biến. Sự kết hợp của thông tin định hướng và thông tin vị trí có thể được gọi là thông tin sắp đặt. Phương pháp có thể còn bao gồm bước xác định vị trí đối tượng từ thông tin vị trí. Ví dụ, vị trí đối tượng có thể được trích xuất từ thông tin vị trí. Việc xác định (ví dụ, trích xuất) của vị trí đối tượng có thể còn được dựa vào thông tin về hình học của sự bố trí loa của một hoặc nhiều loa trong môi trường nghe. Vị trí đối tượng có thể còn được gọi là vị trí kênh của đối tượng âm thanh. Phương pháp có thể còn bao gồm bước sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng. Việc sửa đổi vị trí đối tượng có thể liên quan đến việc hiệu chỉnh vị trí đối tượng để dịch chuyển đầu người nghe từ vị trí nghe danh định. Nói cách khác, việc sửa đổi vị trí đối tượng có thể liên quan đến áp dụng phép bù dịch chuyển vị trí cho vị trí đối tượng. Phương pháp có thể còn bao gồm bước sửa đổi thêm vị trí đối tượng được sửa đổi dựa vào thông tin định hướng người nghe, ví dụ bằng cách áp dụng phép biến đổi quay cho vị trí đối tượng được sửa đổi (ví dụ, sự quay so với đầu người nghe hoặc vị trí nghe danh định). Việc sửa đổi thêm vị trí đối tượng được sửa đổi để kết xuất đối tượng âm thanh có thể liên quan đến sự dịch chuyển cảnh âm thanh quay.

Được tạo cấu hình như được mô tả trên đây, phương pháp đề xuất mang lại trải nghiệm nghe chân thực hơn đặc biệt là cho các đối tượng âm thanh được đặt gần đầu người

nghe. Ngoài ba bậc tự do (quay) thường mang lại cho người nghe trong môi trường 3DoF, phương pháp đề xuất có thể còn tính đến các chuyển động tịnh tiến của đầu người nghe. Điều này cho phép người nghe đến gần các đối tượng âm thanh gần từ các góc và thậm chí các phía khác nhau. Ví dụ, người nghe có thể nghe thấy đối tượng âm thanh “con muỗi” ở gần đầu người nghe từ các góc khác nhau bằng cách chuyển động một ít đầu của mình, có thể ngoài việc quay đầu của mình. Kết quả là, phương pháp đề xuất có thể cho phép trải nghiệm nghe được cải thiện, chân thực, đắm chìm hơn cho người nghe.

Theo một số phương án, việc sửa đổi vị trí đối tượng và việc sửa đổi thêm vị trí đối tượng đã được sửa đổi có thể được thực hiện sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm, được người nghe cảm nhận về mặt tâm thính học là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể sự dịch chuyển đầu người nghe từ vị trí nghe danh định và định hướng của đầu người nghe so với định hướng danh định. Theo đó, đối tượng âm thanh có thể được cảm nhận để chuyển động so với đầu người nghe khi đầu người nghe dịch chuyển từ vị trí nghe danh định. Tương tự, đối tượng âm thanh có thể được cảm nhận là quay so với đầu người nghe khi đầu người nghe thay đổi về sự định hướng từ định hướng danh định. Một hoặc nhiều loa có thể là một phần của tai nghe có micrô, ví dụ, hoặc có thể là một phần của sự bố trí loa (ví dụ, bố trí loa 2.1, 5.1, 7.1, v.v.).

Theo một số phương án, việc sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe có thể được thực hiện bằng cách tịnh tiến vị trí đối tượng theo vector tương quan dương với độ lớn và tương quan âm với hướng của vector dịch chuyển đầu người nghe từ vị trí nghe danh định.

Do đó, điều này đảm bảo rằng các đối tượng âm thanh gần được người nghe cảm nhận chuyển động phù hợp với sự chuyển động đầu của mình. Điều này góp phần vào trải nghiệm nghe chân thực hơn cho các đối tượng âm thanh này.

Theo một số phương án, thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển đầu người nghe từ vị trí nghe danh định một khoảng dịch chuyển vị trí nhỏ. Ví dụ, giá trị tuyệt đối của sự dịch chuyển có thể không lớn hơn 0,5 m. Sự dịch chuyển có thể được biểu diễn theo các tọa độ Đề Các (ví dụ, x, y, z) hoặc theo các tọa độ cầu (ví dụ, góc phương vị, góc ngẩng, bán kính).

Theo một số phương án, thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển của đầu người nghe từ vị trí nghe danh định mà có thể đạt được bằng cách người nghe chuyển động thân trên và/hoặc đầu của mình. Vì vậy, dịch chuyển có thể đạt được đối với người nghe mà không cần chuyển động thân dưới của mình. Ví dụ, sự dịch chuyển của đầu người nghe có thể đạt được khi người nghe đang ngồi trên ghế.

Theo một số phương án, thông tin vị trí có thể chứa thông tin chỉ báo về khoảng cách của đối tượng âm thanh từ vị trí nghe danh định. Khoảng cách (bán kính) có thể là nhỏ hơn 0,5 m. Ví dụ, khoảng cách có thể là nhỏ hơn 1 cm. Theo cách khác, khoảng cách của đối tượng âm thanh từ vị trí nghe danh định có thể được thiết lập là giá trị mặc định bởi bộ giải mã.

Theo một số phương án, thông tin định hướng người nghe có thể bao gồm thông tin về độ quay, độ nghiêng dọc, và độ nghiêng ngang của đầu người nghe. Độ quay, độ nghiêng dọc, và độ nghiêng ngang có thể được cho so với sự định hướng danh định (ví dụ, sự định hướng tham chiếu) của đầu người nghe.

Theo một số phương án, thông tin dịch chuyển người nghe có thể bao gồm thông tin về dịch chuyển đầu của người nghe từ vị trí nghe danh định được biểu thị trong các tọa độ Đề Các hoặc trong các tọa độ cầu. Vì vậy, sự dịch chuyển có thể được biểu thị theo các tọa độ x, y, z đối với các tọa độ Đề Các, và theo tọa độ góc phương vị, góc ngẩng, bán kính đối với tọa độ cầu.

Theo một số phương án, phương pháp có thể còn bao gồm bước việc phát hiện sự định hướng của đầu người nghe bởi thiết bị mang được và/hoặc cố định. Tương tự, phương pháp có thể còn bao gồm phát hiện sự dịch chuyển của đầu người nghe từ vị trí nghe danh định bởi thiết bị mang được và/hoặc cố định. Thiết bị mang được có thể là, tương ứng với, và/hoặc bao gồm, bộ nghe nói hoặc tai nghe có micrô thực tế tăng cường (augmented reality - AR)/ thực tế ảo (virtual reality - VR), chẳng hạn. Thiết bị cố định có thể là, tương ứng, và/hoặc bao gồm, các cảm biến camera, chẳng hạn. Điều này cho phép thu được thông tin chính xác về sự dịch chuyển và/hoặc định hướng của đầu người nghe, và do đó cho phép xử lý theo cách chân thực các đối tượng âm thanh ở gần phù hợp với sự định hướng và/hoặc sự dịch chuyển.

Theo một số phương án, phương pháp có thể còn bao gồm bước kết xuất đối tượng âm thanh cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm. Ví dụ, đối tượng âm thanh có thể được kết xuất cho loa trái và phải của tai nghe có micro.

Theo một số phương án, việc kết xuất có thể được thực hiện để tính đến sự bịt kín âm trong các khoảng cách nhỏ của đối tượng âm thanh từ đầu người nghe, dựa vào các hàm truyền liên quan đến đầu (head-related transfer function - HRTF) cho đầu người nghe. Do đó, việc kết xuất các đối tượng âm thanh ở gần sẽ được người nghe cảm nhận là thậm chí chân thực hơn.

Theo một số phương án, vị trí đối tượng được sửa đổi thêm có thể được điều chỉnh thành định dạng đầu vào được sử dụng bởi bộ kết xuất âm thanh 3D MPEG-H. Theo một số phương án, việc kết xuất có thể được thực hiện bằng cách sử dụng bộ kết xuất âm thanh 3D MPEG-H. Theo một số phương án, việc xử lý có thể được thực hiện bằng cách sử dụng bộ giải mã âm thanh 3D MPEG-H. Theo một số phương án, việc xử lý có thể được thực hiện bởi bộ dịch chuyển cảnh của bộ giải mã âm thanh 3D MPEG-H. Theo đó, phương pháp đề xuất cho phép thực hiện trải nghiệm sáu bậc tự do (Six Degrees of Freedom - 6DoF) giới hạn (tức là, 3DoF+) trong khuôn khổ của chuẩn âm thanh 3D MPEG-H.

Theo một khía cạnh khác, sáng chế đề xuất phương pháp khác để xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Phương pháp có thể bao gồm bước thu được thông tin dịch chuyển người nghe biểu thị sự dịch chuyển của đầu người nghe. Phương pháp này có thể còn bao gồm bước xác định vị trí đối tượng từ thông tin vị trí. Phương pháp này có thể còn bao gồm thêm bước sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng.

Được tạo cấu hình như được mô tả ở trên, phương pháp được đề xuất mang lại trải nghiệm nghe chân thực hơn đặc biệt là cho các đối tượng âm thanh được đặt gần đầu người nghe. Với việc có thể tính đến các chuyển động tịnh tiến nhỏ của đầu người nghe, phương pháp được đề xuất cho phép người nghe tiếp cận các đối tượng âm thanh ở gần từ các góc và thậm chí các phía khác nhau. Kết quả là, phương pháp đề xuất có thể cho phép trải nghiệm nghe được cải thiện, chân thực, đắm chìm hơn cho người nghe.

Theo một số phương án, việc sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe có thể được thực hiện sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng đã được sửa đổi, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể sự dịch chuyển của đầu người nghe từ vị trí nghe danh định.

Theo một số phương án, việc sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe có thể được thực hiện bằng cách tịnh tiến vị trí đối tượng theo vector tương quan dương với độ lớn và tương quan âm với hướng của vector dịch chuyển đầu người nghe từ vị trí nghe danh định.

Theo một khía cạnh khác, sáng chế đề xuất phương pháp khác để xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Phương pháp có thể bao gồm bước thu được thông tin định hướng người nghe biểu thị sự định hướng của đầu người nghe. Phương pháp có thể còn bao gồm bước xác định vị trí đối tượng từ thông tin vị trí. Phương pháp có thể vẫn bao gồm thêm bước sửa đổi vị trí đối tượng dựa vào thông tin định hướng người nghe, ví dụ bằng cách áp dụng phép biến đổi quay cho vị trí đối tượng (ví dụ, phép quay so với đầu người nghe hoặc vị trí nghe danh định).

Được tạo cấu hình như được mô tả ở trên, phương pháp đề xuất có thể tính đến sự định hướng của đầu người nghe để cung cấp trải nghiệm nghe chân thực hơn cho người nghe.

Theo một số phương án, việc sửa đổi vị trí đối tượng dựa vào thông tin định hướng người nghe có thể được thực hiện sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể hướng đầu người nghe so với sự định hướng danh định.

Theo một khía cạnh khác, sáng chế đề xuất thiết bị xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Thiết bị có thể bao gồm bộ xử lý và bộ nhớ được ghép nối với bộ xử lý. Bộ xử lý có thể được làm thích ứng để thu được thông tin định hướng người nghe biểu thị sự định hướng của đầu người nghe. Bộ xử lý có thể còn được làm thích ứng để thu được

thông tin dịch chuyển người nghe biểu thị sự dịch chuyển của đầu người nghe. Bộ xử lý có thể còn được làm thích ứng để xác định vị trí đối tượng từ thông tin vị trí. Bộ xử lý có thể còn được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng. Bộ xử lý có thể vẫn còn được làm thích ứng để sửa đổi thêm vị trí đối tượng được sửa đổi dựa vào thông tin định hướng người nghe, ví dụ bằng cách áp dụng phép biến đổi quay cho vị trí đối tượng được sửa đổi (ví dụ, phép quay so với đầu người nghe hoặc vị trí nghe danh định).

Theo một số phương án, bộ xử lý có thể được làm thích ứng để sửa đổi vị trí đối tượng và sửa đổi thêm vị trí đối tượng được sửa đổi sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể sự dịch chuyển của đầu người nghe từ vị trí nghe danh định và định hướng của đầu người nghe so với sự định hướng danh định.

Theo một số phương án, bộ xử lý có thể được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách tịnh tiến vị trí đối tượng theo vector tương quan dương với độ lớn và tương quan âm với hướng của vector dịch chuyển của đầu người nghe từ vị trí nghe danh định.

Theo một số phương án, thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển của đầu người nghe từ vị trí nghe danh định một khoảng dịch chuyển vị trí nhỏ.

Theo một số phương án, thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển của đầu người nghe từ vị trí nghe danh định mà có thể đạt được bằng việc người nghe chuyển động thân trên và/hoặc đầu của mình.

Theo một số phương án, thông tin vị trí có thể bao gồm chỉ báo về khoảng cách của đối tượng âm thanh từ vị trí nghe danh định.

Theo một số phương án, thông tin định hướng người nghe có thể bao gồm thông tin về độ quay, độ nghiêng dọc, và độ nghiêng ngang của đầu người nghe.

Theo một số phương án, thông tin dịch chuyển người nghe có thể bao gồm thông tin về dịch chuyển đầu của người nghe từ vị trí nghe danh định được biểu diễn theo các tọa độ Đề Các hoặc theo các tọa độ cầu.

Theo một số phương án, thiết bị có thể còn bao gồm thiết bị mang được và/hoặc cố định để phát hiện sự định hướng của đầu người nghe. Theo một số phương án, thiết bị có thể còn bao gồm thiết bị mang được và/hoặc cố định để phát hiện sự dịch chuyển của đầu người nghe từ vị trí nghe danh định.

Theo một số phương án, bộ xử lý có thể còn được làm thích ứng để kết xuất đối tượng âm thanh cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm.

Theo một số phương án, bộ xử lý có thể được làm thích ứng để thực hiện việc kết xuất có tính đến sự bịt kín âm trong các khoảng cách nhỏ của đối tượng âm thanh từ đầu người nghe, dựa vào các HRTF cho đầu người nghe.

Theo một số phương án, bộ xử lý có thể được làm thích ứng để điều chỉnh vị trí đối tượng được sửa đổi thêm thành định dạng đầu vào được sử dụng bởi bộ kết xuất âm thanh 3D MPEG-H. Theo một số phương án, việc kết xuất có thể được thực hiện bằng cách sử dụng bộ kết xuất âm thanh 3D MPEG-H. Tức là, bộ xử lý có thể cài đặt bộ kết xuất âm thanh 3D MPEG-H. Theo một số phương án, bộ xử lý có thể được làm thích ứng để cài đặt bộ giải mã âm thanh 3D MPEG-H. Theo một số phương án, bộ xử lý có thể được làm thích ứng để cài đặt bộ dịch chuyển cảnh của bộ giải mã âm thanh 3D MPEG-H.

Theo một khía cạnh khác, sáng chế đề xuất thiết bị khác xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Thiết bị có thể bao gồm bộ xử lý và bộ nhớ được ghép nối với bộ xử lý. Bộ xử lý có thể được làm thích ứng để thu được thông tin dịch chuyển người nghe biểu thị sự dịch chuyển của đầu người nghe. Bộ xử lý có thể còn được làm thích ứng để xác định vị trí đối tượng từ thông tin vị trí. Bộ xử lý có thể còn được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng.

Theo một số phương án, bộ xử lý có thể được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể sự dịch chuyển của đầu người nghe từ vị trí nghe danh định.

Theo một số phương án, bộ xử lý có thể được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách tịnh tiến vị trí đối tượng theo vector tương quan dương với độ lớn và tương quan âm với hướng của vector dịch chuyển của đầu người nghe từ vị trí nghe danh định.

Theo một khía cạnh khác, sáng chế đề xuất thiết bị khác xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh. Vị trí đối tượng có thể được sử dụng để kết xuất đối tượng âm thanh. Thiết bị có thể bao gồm bộ xử lý và bộ nhớ được ghép nối với bộ xử lý. Bộ xử lý có thể được làm thích ứng để thu được thông tin định hướng người nghe biểu thị sự định hướng của đầu người nghe. Bộ xử lý có thể còn được làm thích ứng để xác định vị trí đối tượng từ thông tin vị trí. Bộ xử lý có thể còn được làm thích ứng thêm để sửa đổi vị trí đối tượng dựa vào thông tin định hướng người nghe, ví dụ bằng cách áp dụng phép biến đổi quay cho vị trí đối tượng được sửa đổi (ví dụ, phép quay so với đầu người nghe hoặc vị trí nghe danh định).

Theo một số phương án, bộ xử lý có thể được làm thích ứng để sửa đổi vị trí đối tượng dựa vào thông tin định hướng người nghe sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể hướng đầu người nghe so với sự định hướng danh định.

Theo một khía cạnh khác nữa, sáng chế đề xuất hệ thống. Hệ thống này có thể bao gồm thiết bị theo khía cạnh bất kỳ trong số các khía cạnh trên đây và thiết bị mang được và/hoặc cố định có khả năng phát hiện sự định hướng của đầu người nghe và phát hiện sự dịch chuyển của đầu người nghe.

Cần hiểu rằng các bước của phương pháp và các dấu hiệu của thiết bị có thể được hoán đổi theo nhiều cách. Cụ thể, các chi tiết của phương pháp được bộc lộ có thể được thực hiện dưới dạng thiết bị được làm thích ứng để thực hiện một số hoặc tất cả các bước của phương pháp, và ngược lại, như người có hiểu biết trung bình hiểu rõ. Cụ thể, được hiểu rằng thiết bị theo sáng chế có thể liên quan đến thiết bị thực hiện hoặc thực thi phương pháp theo các phương án trên đây và biến thể của chúng, và rằng các nội dung trình bày tương ứng liên quan đến các phương pháp áp dụng tương tự cho thiết bị tương ứng. Tương tự, phải hiểu rằng phương pháp theo sáng chế có thể liên quan đến phương pháp vận hành

thiết bị theo các phương án trên đây và biến thể của chúng, và rằng các nội dung trình bày tương ứng liên quan đến thiết bị áp dụng tương tự cho các phương pháp tương ứng.

Mô tả vắn tắt các hình vẽ

Sáng chế được giải thích sau đây theo cách minh họa có dẫn chiếu đến các hình vẽ đính kèm, trong đó

Fig.1 minh họa dưới dạng giản đồ ví dụ về hệ thống âm thanh 3D MPEG-H;

Fig.2 minh họa dưới dạng giản đồ ví dụ về hệ thống âm thanh 3D MPEG-H theo sáng chế;

Fig.3 minh họa dưới dạng giản đồ ví dụ về hệ thống kết xuất âm thanh theo sáng chế;

Fig.4 minh họa dưới dạng giản đồ tập hợp ví dụ về trục tọa độ Đề Các và quan hệ của chúng với các tọa độ cầu; và

Fig.5 là lưu đồ minh họa dưới dạng giản đồ ví dụ về phương pháp xử lý thông tin vị trí cho đối tượng âm thanh theo sáng chế.

Mô tả chi tiết sáng chế

Như được sử dụng ở đây, 3DoF thường là hệ thống có thể xử lý chính xác chuyển động đầu của người dùng, cụ thể là quay đầu, được xác định bằng ba tham số (ví dụ, độ quay, độ nghiêng dọc, và độ nghiêng ngang). Hệ thống này thường có sẵn trong các hệ thống chơi trò chơi khác nhau, như hệ thống thực tế ảo (Virtual Reality-VR) / thực tế tăng cường (Augmented Reality - AR) / thực tế hỗn hợp (Mixed Reality - MR), hoặc trong các môi trường âm thanh khác của loại này.

Như được sử dụng ở đây, người dùng (ví dụ, bộ giải mã hoặc hệ thống tái tạo bao gồm bộ giải mã âm thanh) có thể còn được gọi là “người nghe”.

Như được sử dụng ở đây, 3DoF+ sẽ có nghĩa là, ngoài sự chuyển động đầu của người dùng, mà có thể được xử lý chính xác trong hệ thống 3DoF, các chuyển động tịnh tiến nhỏ có thể cũng được xử lý.

Như được sử dụng ở đây, “nhỏ” sẽ chỉ báo rằng các chuyển động được giới hạn ở dưới ngưỡng thường bằng 0,5 mét. Điều này có nghĩa là các chuyển động không lớn hơn

0,5 mét từ vị trí đầu ban đầu của người dùng. Ví dụ, các chuyển động của người dùng bị ràng buộc bởi chính người dùng ngồi trên ghế.

Như được sử dụng ở đây, “Âm thanh 3D MPEG-H” sẽ chỉ đặc điểm kỹ thuật như được chuẩn hóa trong ISO/IEC 23008-3 và/hoặc các bản sửa đổi, ấn bản bất kỳ trong tương lai hoặc các phiên bản khác của chúng trong chuẩn ISO/IEC 23008-3.

Trong ngữ cảnh của các chuẩn âm thanh do tổ chức MPEG cung cấp, sự khác biệt giữa 3DoF và 3DoF+ có thể được xác định như sau:

- 3DoF: cho phép người dùng trải nghiệm chuyển động quay, nghiêng dọc, và nghiêng ngang (ví dụ, của đầu người dùng);
- 3DoF+: cho phép người dùng trải nghiệm chuyển động quay, nghiêng dọc, và nghiêng ngang và chuyển động tịnh tiến giới hạn (ví dụ, của đầu người dùng), ví dụ trong khi ngồi trên ghế.

Các chuyển động đầu tịnh tiến (nhỏ) có giới hạn có thể là các chuyển động bị ràng buộc ở bán kính chuyển động nhất định. Ví dụ, các chuyển động có thể bị ràng buộc do người dùng đang ở vị trí ngồi, ví dụ, mà không sử dụng thân dưới. Chuyển động tịnh tiến của đầu nhỏ có thể liên quan hoặc tương ứng với dịch chuyển của đầu người dùng so với vị trí nghe danh định. Vị trí nghe danh định (hoặc vị trí người nghe danh định) có thể là vị trí mặc định (như, ví dụ, vị trí được xác định trước, vị trí dự tính đối với đầu người nghe, hoặc điểm nghe hoàn hảo theo sự bố trí loa).

Trải nghiệm 3DoF+ có thể sánh với trải nghiệm 6DoF bị hạn chế, trong đó các chuyển động tịnh tiến có thể được mô tả là các chuyển động giới hạn hoặc nhỏ của đầu. Theo một ví dụ, âm thanh còn được kết xuất dựa vào vị trí và định hướng đầu của người dùng, bao gồm sự bịt kín âm có thể xảy ra. Việc kết xuất có thể được thực hiện để tính đến sự bịt kín âm trong các khoảng cách nhỏ của đối tượng âm thanh từ đầu người nghe, ví dụ dựa vào các hàm truyền liên quan đến đầu (HRTF) cho đầu người nghe.

Về mặt phương pháp, hệ thống, máy và các thiết bị khác có thể thích hợp với chức năng được quy định trong chuẩn âm thanh 3D MPEG-H, điều đó có thể có nghĩa là 3DoF+ được cho phép cho (các) phiên bản trong tương lai bất kỳ của chuẩn MPEG, như các phiên bản tương lai của định dạng phương tiện truyền thông toàn hướng (ví dụ, như được chuẩn hóa trong các phiên bản tương lai của MPEG-I), và/hoặc trong các bản cập nhật bất kỳ cho

Âm thanh MPEG-H (ví dụ các bản sửa đổi hoặc các chuẩn mới hơn dựa vào chuẩn âm thanh 3D MPEG-H), hoặc các chuẩn liên quan hoặc hỗ trợ khác bất kỳ mà có thể yêu cầu cập nhật (ví dụ, các chuẩn chỉ rõ các loại siêu dữ liệu và bản tin SEI nhất định).

Ví dụ, bộ kết xuất âm thanh là tiêu chuẩn cho chuẩn âm thanh được quy định trong đặc điểm kỹ thuật âm thanh 3D MPEG-H, có thể được mở rộng để bao gồm việc kết xuất của cảnh âm thanh để tính đến tương tác của người dùng với cảnh âm thanh một cách chính xác, ví dụ, khi người dùng chuyển động đầu nhẹ qua một bên.

Sáng chế mang lại các ưu điểm kỹ thuật khác nhau, bao gồm ưu điểm cung cấp Âm thanh 3D MPEG-H có khả năng xử lý các trường hợp sử dụng 3DoF+. Sáng chế mở rộng chuẩn âm thanh 3D MPEG-H để hỗ trợ chức năng 3DoF+.

Để hỗ trợ chức năng 3DoF+, hệ thống kết xuất âm thanh cần tính đến các dịch chuyển vị trí nhỏ/ giới hạn của đầu người dùng/ người nghe. Các dịch chuyển vị trí cần được xác định dựa vào độ lệch tương đối từ vị trí ban đầu (tức là, vị trí mặc định/vị trí nghe danh định). Theo một ví dụ, độ lớn của độ lệch này (ví dụ, độ lệch theo bán kính có thể được xác định dựa vào $r_{\text{offset}} = \|P_0 - P_1\|$), trong đó P_0 là vị trí nghe danh định và P_1 là vị trí được dịch chuyển của đầu người nghe) là tối đa vào khoảng 0,5m. Theo một ví dụ khác, độ lớn của độ lệch bị giới hạn ở độ lệch mà chỉ có thể đạt được trong lúc người dùng ngồi trên ghế và không thực hiện chuyển động thân dưới (nhưng đầu của mình đang chuyển động so với thân của mình). Khoảng cách độ lệch (nhỏ) này dẫn đến chênh lệch mức độ (cảm thụ) và panning biên độ rất nhỏ cho các đối tượng âm thanh xa. Tuy nhiên, đối với các đối tượng gần, ngay cả khoảng cách độ lệch nhỏ này có thể trở nên liên quan về mặt cảm thụ. Thực vậy, việc chuyển động đầu người nghe có thể có ảnh hưởng cảm thụ đối với việc nhận biết vị trí định vị đối tượng âm thanh chính xác. Ảnh hưởng cảm thụ này có thể duy trì đáng kể (tức là, người dùng/người nghe có thể nhận thấy về mặt cảm thụ) miễn là tỷ lệ giữa (i) sự dịch chuyển đầu người dùng (ví dụ, $r_{\text{offset}} = \|P_0 - P_1\|$) và khoảng cách đến đối tượng âm thanh (ví dụ, r), về mặt lượng giác, dẫn đến các góc nằm trong phạm vi khả năng tâm thính của người dùng để phát hiện hướng âm thanh. Phạm vi như vậy có thể thay đổi cho các cài đặt bộ kết xuất âm thanh, vật liệu âm thanh và cấu hình phát lại khác nhau. Ví dụ, giả sử rằng phạm vi chính xác định vị là, ví dụ, $\pm 3^\circ$ với $\pm 0,25\text{m}$ tự do chuyển động bên này sang bên kia của đầu người nghe, điều này sẽ tương ứng $\sim 5\text{m}$ của khoảng cách đối tượng.

Đối với các đối tượng ở gần với người nghe, (ví dụ, các đối tượng ở khoảng cách $< 1\text{m}$ từ người dùng), việc xử lý thích hợp đối với sự dịch chuyển vị trí của đầu người nghe là quan trọng cho các kịch bản 3DoF+, vì có các ảnh hưởng cảm thụ đáng kể trong quá trình thay đổi mức độ và panning biên độ.

Một ví dụ về việc xử lý các đối tượng gần với người nghe là, ví dụ, khi đối tượng âm thanh (ví dụ, con muỗi) được đặt rất gần với mặt người nghe. Hệ thống âm thanh, như hệ thống âm thanh có khả năng VR/AR/MR, nên cho phép người dùng nhận biết đối tượng âm thanh này từ tất cả các phía và các góc thậm chí trong lúc người dùng đang có các chuyển động đầu tĩnh tiến nhỏ. Ví dụ, người dùng nên có khả năng nhận biết chính xác đối tượng (ví dụ, con muỗi) thậm chí trong lúc người dùng đang chuyển động đầu của mình mà không chuyển động thân dưới.

Tuy nhiên, hiện nay hệ thống tương thích với đặc điểm kỹ thuật âm thanh 3D MPEG-H hiện tại không có khả năng xử lý điều này một cách chính xác. Thay vào đó, việc sử dụng hệ thống tương thích với hệ thống âm thanh 3D MPEG-H dẫn đến “con muỗi” được nhận biết từ vị trí sai đối với người dùng. Trong các kịch bản mà liên quan đến hiệu suất 3DoF+, các chuyển động tĩnh tiến nhỏ sẽ dẫn đến các chênh lệch đáng kể trong việc nhận biết đối tượng âm thanh (ví dụ khi đầu người chuyển động sang bên trái, thì đối tượng âm thanh “con muỗi” sẽ được nhận biết từ phía bên phải so với đầu người dùng, v.v.).

Chuẩn âm thanh 3D MPEG-H bao gồm cú pháp dòng bit mà cho phép báo hiệu thông tin khoảng cách đối tượng thông qua cú pháp dòng bit, ví dụ, thông qua phần tử cú pháp *object_metadata()* (bắt đầu từ 0,5m).

Phần tử cú pháp *prodMetadataConfig()* có thể được đưa vào dòng bit được cung cấp bởi chuẩn âm thanh 3D MPEG-H, phần tử cú pháp này có thể được sử dụng để báo hiệu rằng khoảng cách đối tượng là rất gần với người nghe. Ví dụ, cú pháp *prodMetadataConfig()* có thể báo hiệu rằng khoảng cách giữa người dùng và đối tượng là nhỏ hơn khoảng cách ngưỡng nhất định (ví dụ, $< 1\text{cm}$).

Fig.1 và Fig.2 minh họa sáng chế dựa vào việc kết xuất qua tai nghe (tức là, trong đó các loa là chuyển động cùng với đầu người nghe).

Fig.1 thể hiện ví dụ về cách vận hành hệ thống 100 là phù hợp với hệ thống âm thanh 3D MPEG-H. Ví dụ này giả sử rằng đầu người nghe được đặt tại vị trí P_0 103 tại

thời điểm t_0 và di chuyển đến vị trí P_1 104 tại thời điểm $t_1 > t_0$. Các vòng tròn nét đứt xung quanh các vị trí P_0 và P_1 cho biết khu vực chuyển động 3DoF+ được phép (ví dụ, với bán kính 0,5 m). Vị trí A 101 cho biết vị trí đối tượng được báo hiệu (tại thời điểm t_0 và thời điểm t_1 , tức là, vị trí đối tượng được báo hiệu được giả định là không đổi theo thời gian). Vị trí A còn cho biết vị trí đối tượng được kết xuất bởi bộ kết xuất âm thanh 3D MPEG-H tại thời điểm t_0 . Vị trí B 102 cho biết vị trí đối tượng được kết xuất bởi âm thanh 3D MPEG-H tại thời điểm t_1 . Các đường thẳng mở rộng hướng lên trên từ các vị trí P_0 và P_1 cho biết các định hướng tương ứng (ví dụ, hướng quan sát) của đầu người nghe tại các thời điểm t_0 và t_1 . Sự dịch chuyển của đầu người dùng giữa vị trí P_0 và vị trí P_1 có thể được biểu diễn bởi $r_{\text{offset}} = \|P_0 - P_1\|$ 106. Với người nghe đang được đặt tại vị trí mặc định (vị trí nghe danh định) P_0 103 tại thời điểm t_0 , người nghe sẽ nhận biết đối tượng âm thanh (ví dụ, con muỗi) tại vị trí chính xác A 101. Nếu người dùng di chuyển đến vị trí P_1 104 tại thời điểm t_1 thì người dùng sẽ nhận biết đối tượng âm thanh ở vị trí B 102 nếu áp dụng quy trình xử lý âm thanh 3D MPEG-H như được chuẩn hóa hiện nay, việc này gây ra sai số được thể hiện δ_{AB} 105. Tức là, mặc dù đầu người nghe chuyển động, nhưng đối tượng âm thanh (ví dụ, con muỗi) sẽ vẫn được nhận biết như được đặt thẳng ở phía trước đầu người nghe (tức là, gần như di chuyển cùng với đầu người nghe). Đáng chú ý, sai số sinh ra δ_{AB} 105 xảy ra bất kể hướng của đầu người nghe.

Fig.2 thể hiện ví dụ về cách vận hành hệ thống so với hệ thống 200 của âm thanh MPEG-H 3D theo sáng chế. Trên Fig.2, đầu người nghe được đặt tại vị trí P_0 203 tại thời điểm t_0 và di chuyển đến vị trí P_1 204 tại thời điểm $t_1 > t_0$. Các vòng tròn nét đứt xung quanh các vị trí P_0 và P_1 một lần nữa cho biết khu vực chuyển động 3DoF+ được phép (ví dụ, với bán kính 0,5 m). Tại 201, được cho biết rằng vị trí A = B nghĩa là vị trí đối tượng được báo hiệu (tại thời điểm t_0 và thời điểm t_1 , tức là, vị trí đối tượng được báo hiệu được giả sử là không đổi theo thời gian). Vị trí A = B 201 còn cho biết vị trí của đối tượng được kết xuất bởi âm thanh 3D MPEG-H tại thời điểm t_0 và thời điểm t_1 . Các mũi tên thẳng đứng mở rộng hướng lên trên từ các vị trí P_0 203 và P_1 204 cho biết các định hướng tương ứng (ví dụ, hướng quan sát) của đầu người nghe tại các thời điểm t_0 và t_1 . Với việc người nghe đang được đặt tại vị trí ban đầu/mặc định (vị trí nghe danh định) P_0 203 tại thời điểm t_0 , người nghe sẽ nhận biết đối tượng âm thanh (ví dụ con muỗi) tại vị trí chính xác A 201. Nếu người dùng di chuyển đến vị trí P_1 203 tại thời điểm t_1 thì người dùng sẽ vẫn nhận

biết đối tượng âm thanh tại vị trí B 201 tương tự như (ví dụ, về cơ bản là bằng) với vị trí A 201 theo sáng chế. Vì vậy, sáng chế cho phép vị trí người dùng thay đổi theo thời gian (ví dụ, từ vị trí P₀ 203 đến vị trí P₁ 204) trong khi vẫn nhận biết âm thanh từ cùng vị trí (cố định về không gian) (ví dụ, vị trí A = B 201, v.v). Nói cách khác, đối tượng âm thanh (ví dụ, con muỗi) di chuyển tương đối với đầu người nghe, phù hợp với (ví dụ, tương quan âm với) chuyển động đầu người nghe. Điều này cho phép người dùng chuyển động xung quanh đối tượng âm thanh (ví dụ, con muỗi) và nhận biết đối tượng âm thanh từ các góc hoặc thậm chí các phía khác nhau. Sự dịch chuyển của đầu người dùng giữa vị trí P₀ và vị trí P₁ có thể được biểu diễn bởi $r_{\text{offset}} = ||P_0 - P_1||$ 206.

Fig.3 minh họa ví dụ của hệ thống kết xuất âm thanh 300 theo sáng chế. Hệ thống kết xuất âm thanh 300 có thể tương ứng hoặc bao gồm bộ giải mã, như bộ giải mã âm thanh 3D MPEG-H, chẳng hạn. Hệ thống kết xuất âm thanh 300 có thể bao gồm bộ dịch chuyển cảnh âm thanh 310 với giao diện xử lý dịch chuyển cảnh âm thanh tương ứng (ví dụ, giao diện cho dữ liệu dịch chuyển cảnh theo chuẩn âm thanh 3D MPEG-H). Bộ dịch chuyển cảnh âm thanh 310 có thể xuất ra các vị trí đối tượng 321 để kết xuất các đối tượng âm thanh tương ứng. Ví dụ, bộ dịch chuyển cảnh có thể xuất ra siêu dữ liệu vị trí đối tượng để kết xuất các đối tượng âm thanh tương ứng.

Hệ thống kết xuất âm thanh 300 có thể còn bao gồm bộ kết xuất đối tượng âm thanh 320. Ví dụ, bộ kết xuất có thể bao gồm phần cứng, phần mềm, và/hoặc chức năng xử lý một phần hoặc toàn bộ được thực hiện thông qua điện toán đám mây, bao gồm các dịch vụ khác nhau, như các nền tảng phát triển phần mềm, các máy chủ, bộ lưu trữ và phần mềm, trên Internet, thường được gọi là "đám mây", tương thích với đặc điểm kỹ thuật được quy định trong chuẩn âm thanh 3D MPEG-H. Bộ kết xuất đối tượng âm thanh 320 có thể kết xuất các đối tượng âm thanh vào một hoặc nhiều loa (thực hoặc ảo) theo các vị trí đối tượng tương ứng (các vị trí đối tượng này có thể là các vị trí đối tượng được sửa đổi hoặc được sửa đổi thêm được mô tả dưới đây). Bộ kết xuất đối tượng âm thanh 320 có thể kết xuất các đối tượng âm thanh vào các tai nghe và/hoặc loa phóng thanh. Tức là, bộ kết xuất đối tượng âm thanh 320 có thể tạo ra các dạng sóng đối tượng theo định dạng tái tạo cho trước. Kết quả là, bộ kết xuất đối tượng âm thanh 320 có thể sử dụng siêu dữ liệu đối tượng nén. Mỗi đối tượng có thể được kết xuất cho các kênh đầu ra nhất định theo vị trí đối tượng của nó (ví dụ, vị trí đối tượng được sửa đổi, hoặc vị trí đối tượng được sửa đổi thêm). Do

vậy, các vị trí đối tượng có thể còn được gọi là các vị trí kênh của các đối tượng âm thanh của chúng. Các vị trí đối tượng âm thanh 321 có thể được chứa trong siêu dữ liệu vị trí đối tượng hoặc siêu dữ liệu dịch chuyển cảnh xuất ra bởi bộ dịch chuyển cảnh 310.

Quy trình xử lý theo sáng chế có thể tuân theo với chuẩn âm thanh 3D MPEG-H. Như vậy, quy trình xử lý có thể được thực hiện bởi bộ giải mã âm thanh 3D MPEG-H, hoặc cụ thể hơn, bởi bộ dịch chuyển cảnh MPEG-H và/hoặc bộ kết xuất âm thanh 3D MPEG-H. Theo đó, hệ thống kết xuất âm thanh 300 trên Fig.3 có thể tương ứng với hoặc bao gồm bộ giải mã âm thanh 3D MPEG-H (tức là, bộ giải mã tuân theo đặc điểm kỹ thuật được quy định trong chuẩn âm thanh 3D MPEG-H). Theo một ví dụ, hệ thống kết xuất âm thanh 300 có thể là thiết bị bao gồm bộ xử lý và bộ nhớ được ghép nối với bộ xử lý, trong đó bộ xử lý được làm thích ứng để cài đặt bộ giải mã âm thanh 3D MPEG-H. Cụ thể, bộ xử lý có thể được làm thích ứng để cài đặt bộ dịch chuyển cảnh MPEG-H và/hoặc bộ kết xuất âm thanh 3D MPEG-H. Vì vậy, bộ xử lý có thể được làm thích ứng để thực hiện các bước xử lý được mô tả trong sáng chế (ví dụ, các bước S510 đến S560 của phương pháp 500 được mô tả dưới đây có tham chiếu đến Fig.5). Theo một ví dụ khác, việc xử lý hoặc hệ thống kết xuất âm thanh 300 có thể được thực hiện trong đám mây.

Hệ thống kết xuất âm thanh 300 có thể thu được (ví dụ, nhận) dữ liệu vị trí nghe 301. Hệ thống kết xuất âm thanh 300 có thể thu được dữ liệu vị trí nghe 301 thông qua giao diện đầu vào của bộ giải mã âm thanh 3D MPEG-H.

Dữ liệu vị trí nghe 301 có thể biểu thị định hướng và/hoặc vị trí (ví dụ, sự dịch chuyển) của đầu người nghe. Vì vậy, dữ liệu vị trí nghe 301 (có thể còn được gọi là thông tin sắp đặt) có thể bao gồm thông tin định hướng người nghe và/hoặc thông tin dịch chuyển người nghe.

Thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển của đầu người nghe (ví dụ, từ vị trí nghe danh định). Thông tin dịch chuyển người nghe có thể tương ứng hoặc bao gồm chỉ báo độ lớn của sự dịch chuyển của đầu người nghe từ vị trí nghe danh định, $r_{\text{offset}} = \|P_0 - P_1\|$ 206 như được minh họa trên Fig.2. Trong ngữ cảnh của sáng chế, thông tin dịch chuyển người nghe cho biết sự dịch chuyển vị trí *nhỏ* của đầu người nghe từ vị trí nghe danh định. Ví dụ, giá trị tuyệt đối của sự dịch chuyển có thể không lớn hơn 0,5 m. Thông thường, đây là sự dịch chuyển của đầu người nghe từ vị trí nghe danh định mà có

thể đạt được khi người nghe chuyển động thân trên và/hoặc đầu của mình. Tức là, sự dịch chuyển có thể đạt được đối với người nghe mà không cần chuyển động thân dưới của mình. Ví dụ, sự dịch chuyển của đầu người nghe có thể đạt được khi người nghe đang ngồi trên ghế, như nêu trên đây. Sự dịch chuyển có thể được biểu thị trong các hệ tọa độ khác nhau, như, ví dụ, trong các tọa độ Đề Các (ví dụ, theo x , y , z) hoặc trong các tọa độ cầu (ví dụ, theo góc phương vị, góc nâng, bán kính). Các hệ tọa độ khác để biểu thị sự dịch chuyển của đầu người nghe là cũng khả thi và nên được hiểu là được nằm trong sáng chế này.

Thông tin định hướng người nghe có thể biểu thị định hướng của đầu người nghe (ví dụ, định hướng của đầu người nghe so với sự định hướng danh định / sự định hướng tham chiếu của đầu người nghe). Ví dụ, thông tin định hướng người nghe có thể bao gồm thông tin về độ quay, độ nghiêng dọc, và độ nghiêng ngang của đầu người nghe. Ở đây, độ quay, độ nghiêng dọc, và độ nghiêng ngang có thể được đưa ra so với sự định hướng danh định.

Dữ liệu vị trí nghe 301 có thể được thu thập liên tục từ bộ thu mà có thể cung cấp thông tin về các chuyển động tịnh tiến của người dùng. Ví dụ, dữ liệu vị trí nghe 301 được sử dụng tại thời điểm nhất định có thể đã được thu thập gần đây từ bộ thu. Dữ liệu vị trí nghe có thể được suy ra/thu thập/tạo ra dựa vào thông tin cảm biến. Ví dụ, dữ liệu vị trí nghe 301 có thể được suy ra/thu thập/tạo ra bởi thiết bị mang được và/hoặc cố định có các bộ cảm biến thích hợp. Tức là, định hướng của đầu người nghe có thể được phát hiện bởi thiết bị mang được và/hoặc cố định. Tương tự, sự dịch chuyển của đầu người nghe (ví dụ, từ vị trí nghe danh định) có thể được phát hiện bởi thiết bị mang được và/hoặc cố định. Thiết bị mang được có thể là, tương ứng với, và/hoặc bao gồm, tai nghe có micrô (ví dụ, tai nghe có micrô AR/VR), chẳng hạn. Thiết bị cố định có thể là, tương ứng với, và/hoặc bao gồm, các cảm biến camera, chẳng hạn. Thiết bị cố định có thể được chứa trong máy thu hình hoặc đầu thu giải mã tín hiệu, chẳng hạn. Theo một số phương án, dữ liệu vị trí nghe 301 có thể được nhận từ bộ mã hóa âm thanh (ví dụ, bộ mã hóa tuân theo âm thanh 3D MPEG-H) có thể đã thu được (ví dụ, nhận) thông tin cảm biến.

Theo một ví dụ, thiết bị mang được và/hoặc cố định để phát hiện dữ liệu vị trí nghe 301 có thể được gọi là các thiết bị theo dõi hỗ trợ ước lượng/phát hiện vị trí của đầu và/hoặc ước lượng/phát hiện định hướng của đầu. Có các giải pháp khác nhau cho phép theo dõi các chuyển động của đầu người dùng một cách chính xác bằng cách sử dụng máy vi tính

hoặc các camera điện thoại thông minh (ví dụ, dựa vào việc nhận dạng và theo dõi khuôn mặt “FaceTrackNoIR”, “opentrack”). Ngoài ra, một số hệ thống thực tế ảo màn hình đeo trên đầu (Head-Mounted Display - HMD) (ví dụ, HTC VIVE, Oculus Rift) có công nghệ theo dõi đầu được tích hợp. Bất kỳ trong số các giải pháp này có thể được sử dụng trong ngữ cảnh của sáng chế.

Điều quan trọng cũng cần lưu ý là khoảng cách dịch chuyển đầu trong thế giới vật lý không nhất thiết tương ứng một - một với dịch chuyển được chỉ báo bởi dữ liệu vị trí nghe 301. Để đạt được hiệu quả siêu thực (ví dụ, hiệu ứng thị sai chuyển động của người dùng bị khuếch đại quá mức), các ứng dụng nhất định có thể sử dụng các cài đặt hiệu chuẩn cảm biến khác nhau hoặc chỉ ra các ánh xạ khác nhau giữa sự chuyển động trong các không gian thực và ảo. Do đó, người có hiểu biết trung bình có thể mong đợi rằng một chuyển động vật lý nhỏ dẫn đến sự dịch chuyển lớn hơn trong thực tế ảo trong một số trường hợp sử dụng. Trong trường hợp bất kỳ, có thể nói rằng các độ lớn của dịch chuyển trong thế giới vật lý và trong thực tế ảo (tức là, dịch chuyển được chỉ báo bởi dữ liệu vị trí nghe 301) là tương quan dương. Tương tự, các hướng dịch chuyển trong thế giới vật lý và trong thực tế ảo là tương quan dương.

Hệ thống kết xuất âm thanh 300 có thể còn nhận thông tin vị trí (đối tượng) (ví dụ, dữ liệu vị trí đối tượng) 302 và dữ liệu âm thanh 322. Dữ liệu âm thanh 322 có thể bao gồm một hoặc nhiều đối tượng âm thanh. Thông tin vị trí 302 có thể là một phần của siêu dữ liệu đối với dữ liệu âm thanh 322. Thông tin vị trí 302 có thể biểu thị các vị trí đối tượng tương ứng với một hoặc nhiều đối tượng âm thanh. Ví dụ, thông tin vị trí 302 có thể bao gồm chỉ báo về khoảng cách của các đối tượng âm thanh tương ứng so với vị trí nghe danh định của người dùng/người nghe. Khoảng cách (bán kính) có thể là nhỏ hơn 0,5 m. Ví dụ, khoảng cách có thể là nhỏ hơn 1 cm. Nếu thông tin vị trí 302 không bao gồm chỉ báo về khoảng cách của đối tượng âm thanh cho trước từ vị trí nghe danh định, thì hệ thống kết xuất âm thanh có thể cài đặt khoảng cách của đối tượng âm thanh này từ vị trí nghe danh định thành giá trị mặc định (ví dụ, 1 m). Thông tin vị trí 302 có thể còn bao gồm các chỉ báo về góc ngang và/hoặc góc phương vị của các đối tượng âm thanh tương ứng.

Mỗi vị trí đối tượng có thể dùng được để kết xuất đối tượng âm thanh tương ứng của nó. Theo đó, thông tin vị trí 302 và dữ liệu âm thanh 322 có thể được chứa trong, hoặc tạo thành, nội dung âm thanh dựa trên đối tượng. Nội dung âm thanh (ví dụ, các đối tượng

âm thanh / dữ liệu âm thanh 322 cùng với thông tin vị trí 302 của chúng) có thể được truyền tải trong dòng bit âm thanh được mã hóa. Ví dụ, nội dung âm thanh có thể ở trong định dạng của dòng bit được nhận từ cuộc truyền thông qua mạng. Trong trường hợp này, có thể nói rằng hệ thống kết xuất âm thanh nhận nội dung âm thanh (ví dụ, từ dòng bit âm thanh được mã hóa).

Theo một ví dụ của sáng chế, các tham số siêu dữ liệu có thể được sử dụng để sửa quy trình xử lý các trường hợp sử dụng với sự tăng cường tương thích ngược cho 3DoF và 3DoF+. Siêu dữ liệu có thể bao gồm thông tin dịch chuyển người nghe ngoài thông tin định hướng người nghe. Các tham số siêu dữ liệu này có thể được sử dụng bởi hệ thống được thể hiện trên các Fig.2 và Fig.3, cũng như các phương án khác bất kỳ của sáng chế.

Sự tăng cường tương thích ngược có thể cho phép sửa quy trình xử lý các trường hợp sử dụng (ví dụ, các phương án thực hiện sáng chế) dựa vào giao diện dịch chuyển cảnh âm thanh 3D MPEG-H tiêu chuẩn. Điều này có nghĩa bộ giải mã/bộ kết xuất âm thanh 3D MPEG-H kế thừa sẽ vẫn tạo ra đầu ra, ngay cả khi không chính xác. Tuy nhiên, bộ giải mã/bộ kết xuất âm thanh 3D MPEG-H được tăng cường theo sáng chế sẽ áp dụng chính xác dữ liệu mở rộng (ví dụ, siêu dữ liệu mở rộng) và quy trình xử lý và do đó có thể xử lý kịch bản các đối tượng được đặt gần với người nghe một cách chính xác.

Theo một ví dụ, sáng chế đề cập đến việc cung cấp dữ liệu cho các chuyển động tịnh tiến nhỏ của đầu người dùng ở các định dạng khác với định dạng được nêu dưới đây, và các công thức có thể được làm thích ứng sao cho phù hợp. Ví dụ, dữ liệu có thể được cung cấp ở định dạng như các tọa độ x, y, z (trong hệ tọa độ Đề Các) thay vì góc phương vị, góc ngẩng và bán kính (trong hệ tọa độ cầu). Ví dụ về các hệ tọa độ này so với nhau được thể hiện trên Fig.4.

Theo một ví dụ, sáng chế hướng đến việc cung cấp siêu dữ liệu (ví dụ, thông tin dịch chuyển người nghe được bao gồm trong dữ liệu vị trí nghe 301 được thể hiện trên Fig.3) để nhập chuyển động tịnh tiến của đầu người nghe. Siêu dữ liệu có thể được sử dụng, ví dụ, cho giao diện cho dữ liệu dịch chuyển cảnh. Siêu dữ liệu (ví dụ, thông tin dịch chuyển người nghe) có thể thu được bằng cách triển khai thiết bị theo dõi hỗ trợ theo dõi 3DoF+ hoặc 6DoF.

Theo một ví dụ, siêu dữ liệu (ví dụ, thông tin dịch chuyển người nghe, cụ thể là sự dịch chuyển của đầu người nghe, hoặc tương đương, sự dịch chuyển cảnh) có thể được biểu diễn bởi ba tham số sau đây **sd_azimuth**, **sd_elevation**, và **sd_radius**, liên quan đến góc phương vị, góc ngẩng và bán kính (các tọa độ cầu) của sự dịch chuyển của đầu người nghe (hoặc dịch chuyển cảnh).

Cú pháp cho các tham số này, được đưa ra bởi bảng sau.

Bảng 264b — Cú pháp của mpeg3daPositionalSceneDisplacementData()

Cú pháp	Số lượng bit	Gợi ý
mpeg3daPositionalScenedịch chuyểnData()		
{		
sd_azimuth;	8	Uimbsbf
sd_elevation;	6	Uimbsbf
sd_radius;	4	Uimbsbf
}		

sd_azimuth	Trường này xác định vị trí góc phương vị dịch chuyển cảnh. Trường này có thể nhận các giá trị từ -180 đến 180. $az_offset = (sd_azimuth - 128) \cdot 1,5$ $az_offset = \min(\max(az_offset, -180), 180)$
sd_elevation	Trường này xác định vị trí góc ngẩng dịch chuyển cảnh. Trường này có thể nhận các giá trị từ -90 đến 90. $el_offset = (sd_elevation - 32) \cdot 3,0$ $el_offset = \min(\max(el_offset, -90), 90)$
sd_radius	Trường này xác định bán kính dịch chuyển cảnh. Trường này có thể nhận các giá trị từ 0,015626 đến 0,25. $r_offset = (sd_radius + 1) / 16$

Theo một ví dụ khác, siêu dữ liệu (ví dụ, thông tin dịch chuyển người nghe) có thể được biểu diễn bởi ba tham số sau **sd_x**, **sd_y**, và **sd_z** trong các tọa độ Đề Các, các tham số này sẽ làm giảm việc xử lý dữ liệu từ tọa độ cầu thành các tọa độ Đề Các. Siêu dữ liệu có thể được đưa vào cú pháp sau:

Cú pháp	Số lượng bit	Gợi ý
mpeg3daPositionalScenedịch chuyểnDataTrans()		
{		
sd_x;	6	uimbsbf

<code>sd_y;</code>	6	<code>uimsbf</code>
<code>sd_z;</code>	6	<code>uimsbf</code>
<code>}</code>		

Như được mô tả ở trên, cú pháp ở trên hoặc các cú pháp tương đương của cú pháp có thể báo hiệu thông tin liên quan đến các chuyển động quay xung quanh các trục x , y , z .

Theo một ví dụ của sáng chế, việc xử lý các góc dịch chuyển cảnh cho các kênh và đối tượng có thể được cải thiện bằng cách mở rộng các phương trình tính đến việc thay đổi vị trí của đầu người dùng. Tức là, việc xử lý các vị trí đối tượng có thể tính đến (ví dụ, có thể dựa vào, ít nhất một phần) thông tin dịch chuyển người nghe.

Ví dụ về phương pháp 500 để xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh được minh họa trong lưu đồ trên Fig.5. Phương pháp này có thể được thực hiện bởi bộ giải mã, như bộ giải mã âm thanh 3D MPEG-H. Hệ thống kết xuất âm thanh 300 trên Fig.3 có thể là một ví dụ của bộ giải mã như vậy.

Bước thứ nhất (không được thể hiện trên Fig.5), nội dung âm thanh bao gồm đối tượng âm thanh và thông tin vị trí tương ứng được nhận, ví dụ từ dòng bit của âm thanh mã hóa. Sau đó, phương pháp này có thể còn bao gồm bước giải mã nội dung âm thanh mã hóa để thu được đối tượng âm thanh và thông tin vị trí.

Tại bước S510, thông tin định hướng người nghe được thu nhận (ví dụ, được nhận). Thông tin định hướng người nghe có thể biểu thị sự định hướng của đầu người nghe.

Tại bước S520, thông tin dịch chuyển người nghe được thu nhận (ví dụ, được nhận). Thông tin dịch chuyển người nghe có thể biểu thị sự dịch chuyển của đầu người nghe.

Tại bước S530, vị trí đối tượng được xác định từ thông tin vị trí. Ví dụ, vị trí đối tượng (ví dụ, theo góc phương vị, góc ngẩng, bán kính, hoặc x , y , z hoặc tương đương của chúng) có thể được trích xuất từ thông tin vị trí. Việc xác định vị trí đối tượng có thể còn được dựa, ít nhất một phần, vào thông tin về hình học bố trí loa của một hoặc nhiều loa (thực hoặc ảo) trong môi trường nghe. Nếu bán kính không được chứa trong thông tin vị trí đối với đối tượng âm thanh đó, thì bộ giải mã có thể cài đặt bán kính thành giá trị mặc định (ví dụ, 1 m). Theo một số phương án, giá trị mặc định có thể phụ thuộc vào hình học bố trí loa.

Lưu ý rằng, các bước S510, S520, và S520 có thể được thực hiện theo thứ tự bất kỳ.

Tại bước S540, vị trí đối tượng xác định tại bước S530 được sửa đổi dựa vào thông tin dịch chuyển người nghe. Điều này có thể được thực hiện bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng, theo thông tin dịch chuyển (ví dụ, theo sự dịch chuyển của đầu người nghe). Vì vậy, có thể cho rằng việc sửa đổi vị trí đối tượng là liên quan đến việc hiệu chỉnh vị trí đối tượng để dịch chuyển đầu người nghe (ví dụ, dịch chuyển từ vị trí nghe danh định). Cụ thể, việc sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe có thể được thực hiện bằng cách tịnh tiến vị trí đối tượng theo vector tương quan dương với độ lớn và tương quan âm với hướng của vector dịch chuyển đầu người nghe từ vị trí nghe danh định. Ví dụ về phép tịnh tiến này được minh họa dưới dạng biểu đồ trên Fig.2.

Tại bước S550, vị trí đối tượng sửa đổi thu được tại bước S540 được sửa đổi tiếp dựa vào thông tin định hướng người nghe. Ví dụ, điều này có thể được thực hiện bằng cách áp dụng phép biến đổi quay cho vị trí đối tượng được sửa đổi, theo thông tin định hướng người nghe. Chuyển động quay này có thể là quay so với đầu người nghe hoặc vị trí nghe danh định, chẳng hạn. Phép biến đổi quay có thể được thực hiện bởi thuật toán dịch chuyển cảnh.

Như được nêu trên đây, sự bù độ lệch người dùng (tức là, sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe) được xem xét khi áp dụng phép biến đổi quay. Ví dụ, việc áp dụng phép biến đổi quay có thể bao gồm:

- Tính ma trận biến đổi quay (dựa vào sự định hướng người dùng, ví dụ, thông tin định hướng người nghe),
- Chuyển đổi vị trí đối tượng từ tọa độ cầu sang tọa độ Đề Các,
- Áp dụng phép biến đổi quay cho các đối tượng âm thanh được bù độ lệch vị trí người dùng (tức là, đến vị trí đối tượng được sửa đổi), và
- Chuyển đổi vị trí đối tượng, sau khi biến đổi quay, ngược trở lại từ tọa độ Đề Các sang tọa độ cầu.

Bước S560 tiếp theo (không được thể hiện trên Fig.5), phương pháp 500 có thể bao gồm bước kết xuất đối tượng âm thanh vào một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm. Nhằm mục đích này, vị trí đối tượng được sửa đổi thêm có thể được điều chỉnh thành định dạng đầu vào được sử dụng bởi bộ kết xuất âm thanh 3D MPEG-H (ví dụ, bộ kết xuất đối tượng âm thanh 320 được mô tả trên đây). Một hoặc nhiều loa (thực hoặc ảo) nêu trên có thể là một phần của tai nghe có micro, ví dụ, hoặc có thể là một phần của bố trí loa (ví dụ, bố trí loa 2.1, bố trí loa 5.1, bố trí loa 7.1, v.v.). Theo một số phương án, đối tượng âm thanh có thể được kết xuất cho loa trái và loa phải của tai nghe có micro, chẳng hạn.

Mục đích của các bước S540 và S550 được mô tả ở trên là thực hiện sửa đổi vị trí đối tượng và sửa đổi thêm vị trí đối tượng đã được sửa đổi sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa (thực hoặc ảo) theo vị trí đối tượng được sửa đổi thêm, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định. Vị trí cố định này của đối tượng âm thanh sẽ được nhận biết về mặt tâm thính bất kể sự dịch chuyển của đầu người nghe từ vị trí nghe danh định và bất kể hướng đầu người nghe so với sự định hướng danh định. Nói cách khác, đối tượng âm thanh có thể được nhận biết là chuyển động (tịnh tiến) so với đầu người nghe khi đầu người nghe dịch chuyển từ vị trí nghe danh định. Tương tự, đối tượng âm thanh có thể được nhận biết là chuyển động (quay) so với đầu người nghe khi đầu người nghe thay đổi hướng từ sự định hướng danh định. Do đó, người nghe có thể nhận biết đối tượng âm thanh gần từ các góc và khoảng cách khác nhau, bằng cách chuyển động đầu của mình.

Việc sửa đổi vị trí đối tượng và việc sửa đổi thêm vị trí đối tượng đã sửa đổi lần lượt tại các bước S540 và S550, có thể được thực hiện trong ngữ cảnh dịch chuyển cảnh âm thanh (quay/tịnh tiến), ví dụ, bằng bộ dịch chuyển cảnh âm thanh 310 được mô tả ở trên.

Cần lưu ý là các bước nhất định có thể được bỏ qua, phụ thuộc vào trường hợp sử dụng cụ thể. Ví dụ, nếu dữ liệu vị trí nghe 301 chỉ bao gồm thông tin dịch chuyển người nghe (nhưng không bao gồm thông tin định hướng người nghe, hoặc chỉ có thông tin định hướng người nghe chỉ báo rằng không có sự lệch hướng của đầu người nghe từ sự định hướng danh định), thì bước S550 có thể được lược bỏ. Sau đó, việc kết xuất tại bước S560 sẽ được thực hiện theo vị trí đối tượng sửa đổi được xác định tại bước S540. Tương tự,

nếu dữ liệu vị trí nghe 301 chỉ bao gồm thông tin định hướng người nghe (nhưng không bao gồm thông tin dịch chuyển người nghe, hoặc chỉ có thông tin dịch chuyển người nghe chỉ báo rằng không có sự lệch vị trí của đầu người nghe từ vị trí nghe danh định), thì bước S540 có thể được lược bỏ. Sau đó, bước S550 liên quan đến việc sửa đổi vị trí đối tượng được xác định tại bước S530 dựa vào thông tin định hướng người nghe. Việc kết xuất tại bước S560 sẽ được thực hiện theo vị trí đối tượng được sửa đổi được xác định tại bước S550.

Nói chung, sáng chế đề xuất cập nhật vị trí của các vị trí đối tượng nhận được dưới dạng một phần của nội dung âm thanh dựa trên đối tượng (ví dụ, thông tin vị trí 302 cùng với dữ liệu âm thanh 322), dựa vào dữ liệu vị trí nghe 301 đối với người nghe.

Đầu tiên, vị trí đối tượng (hoặc vị trí kênh) $p = (az, el, r)$ được xác định. Điều này có thể được thực hiện trong ngữ cảnh của (ví dụ, như một phần của) bước 530 của phương pháp 500.

Đối với các tín hiệu dựa trên kênh, bán kính r có thể được xác định như sau:

— Nếu có loa phóng thanh dự định (của kênh của tín hiệu đầu vào dựa trên kênh) trong thiết lập loa phóng thanh tái tạo và khoảng cách của thiết lập tái tạo là được biết, thì bán kính r được đặt là khoảng cách loa phóng thanh (ví dụ, tính theo cm).

— Nếu không có loa phóng thanh dự định trong thiết lập loa phóng thanh tái tạo, nhưng khoảng cách của các loa phóng thanh tái tạo (ví dụ, từ vị trí nghe danh định) là được biết, thì bán kính r được đặt là khoảng cách loa phóng thanh tái tạo cực đại.

— Nếu không có loa phóng thanh dự định trong thiết lập loa phóng thanh tái tạo và không biết khoảng cách loa phóng thanh tái tạo, thì bán kính r được đặt là giá trị mặc định (ví dụ, 1023cm).

Đối với các tín hiệu dựa trên đối tượng, bán kính r được xác định như sau:

— Nếu biết khoảng cách đối tượng (ví dụ, từ các công cụ tạo ra và các định dạng tạo ra và được mang trong `prodMetadataConfig()`), thì bán kính r được đặt là khoảng cách đối tượng được biết đến (ví dụ, được báo hiệu bởi

goa_bsObjectDistance[] (theo cm) theo Bảng AMD5.7 của chuẩn âm thanh 3D MPEG-H).

Bảng AMD5.7 — Cú pháp của goa_Production_Metadata ()

Cú pháp	Số lượng bit	Gợi ý
goa_Production_Metadata() { /* CÂU HÌNH TẠO SIÊU DỮ LIỆU */ goa_hasObjectDistance; if (goa_hasObjectDistance) { for (o = 0; o < goa_numberOfOutputObjects; o++) { goa_bsObjectDistance[o] } } }	1 8	Bslbf Uimbsf

— Nếu biết khoảng cách đối tượng từ thông tin vị trí (ví dụ, từ siêu dữ liệu đối tượng và được mang trong object_metadata()), thì bán kính r được đặt là khoảng cách đối tượng được báo hiệu trong thông tin vị trí (ví dụ, bán kính[] (theo cm) được mang với siêu dữ liệu đối tượng). Bán kính r có thể được báo hiệu theo các phần: “Định tỷ lệ của siêu dữ liệu đối tượng” và “Giới hạn siêu dữ liệu đối tượng” được thể hiện dưới đây.

Định tỷ lệ của siêu dữ liệu đối tượng

Giống như một bước tùy chọn trong ngữ cảnh xác định vị trí đối tượng, vị trí đối tượng $p = (az, el, r)$ được xác định từ thông tin vị trí có thể được định tỷ lệ. Điều này có thể bao gồm việc áp dụng hệ số định tỷ lệ để đảo ngược việc định tỷ lệ mã hóa dữ liệu đầu vào cho mỗi thành phần. Điều này có thể được thực hiện cho mọi đối tượng. Việc định tỷ lệ thực tế vị trí đối tượng có thể được thực hiện theo mã giả dưới đây:

```
descscale_multidata()
{
  for (o = 0; o < num_objects; o++)
    azimuth[o] = azimuth[o] * 1.5;

  for (o = 0; o < num_objects; o++)
    elevation[o] = elevation[o] * 3.0;

  for (o = 0; o < num_objects; o++)
    radius[o] = pow(2.0, (radius[o] / 3.0)) / 2.0;
```

```

for (o = 0; o < num_objects; o++)
    gain[o] = pow(10.0, (gain[o] - 32.0) / 40.0);

if (uniform_spread == 1)
{
    for (o = 0; o < num_objects; o++)
        spread[o] = spread[o] * 1.5;
}
else
{
    for (o = 0; o < num_objects; o++)
        spread_width[o] = spread_width[o] * 1.5;

    for (o = 0; o < num_objects; o++)
        spread_height[o] = spread_height[o] * 3.0;

    for (o = 0; o < num_objects; o++)
        spread_depth[o] = (pow(2.0, (spread_depth[o] / 3.0)) / 2.0) - 0.5;
}
for (o = 0; o < num_objects; o++)
    dynamic_object_priority[o] = dynamic_object_priority[o];
}

```

Giới hạn siêu dữ liệu đối tượng

Là bước tùy chọn khác trong ngữ cảnh xác định vị trí đối tượng, vị trí đối tượng (có thể được định tỷ lệ) $p = (az, el, r)$ được xác định từ thông tin vị trí có thể được giới hạn. Điều này có thể bao gồm việc áp dụng giới hạn cho các giá trị được giải mã cho mỗi thành phần để giữ các giá trị trong khoảng hợp lệ. Điều này có thể được thực hiện cho mọi đối tượng. Giới hạn thực tế vị trí đối tượng có thể được thực hiện theo chức năng của mã giả dưới đây:

```

limit_range()
{
    minval = -180;
    maxval = 180;
    for (o = 0; o < num_objects; o++)
        azimuth[o] = MIN(MAX(azimuth[o], minval), maxval);

    minval = -90;
    maxval = 90;
    for (o = 0; o < num_objects; o++)
        elevation[o] = MIN(MAX(elevation[o], minval), maxval);
}

```

```

minval = 0,5;
maxval = 16;
for (o = 0; o < num_objects; o++)
    radius[o] = MIN(MAX(radius[o], minval), maxval);

minval = 0,004;
maxval = 5,957;
for (o = 0; o < num_objects; o++)
    gain[o] = MIN(MAX(gain[o], minval), maxval);

if (uniform_spread == 1)
{
    minval = 0;
    maxval = 180;
    for (o = 0; o < num_objects; o++)
        spread[o] = MIN(MAX(spread[o], minval), maxval);
}
else
{
    minval = 0;
    maxval = 180;
    for (o = 0; o < num_objects; o++)
        spread_width[o] = MIN(MAX(spread_width[o], minval), maxval);

    minval = 0;
    maxval = 90;
    for (o = 0; o < num_objects; o++)
        spread_height[o] = MIN(MAX(spread_height[o], minval), maxval);

    minval = 0;
    maxval = 15,5;
    for (o = 0; o < num_objects; o++)
        spread_depth[o] = MIN(MAX(spread_depth[o], minval), maxval);
}
minval = 0;
maxval = 7;
for (o = 0; o < num_objects; o++)
    dynamic_object_priority[o] = MIN(MAX(dynamic_object_priority[o],
minval),
    maxval);
}

```

Sau đó, vị trí đối tượng được xác định (và tùy chọn, được định tỷ lệ và/hoặc giới hạn) $p = (az, el, r)$ có thể được chuyển đổi sang hệ tọa độ được xác định trước, như ví dụ hệ tọa độ theo ‘quy ước chung’ trong đó góc phương vị 0° nằm ở tai phải (các giá trị dương theo hướng ngược chiều kim đồng hồ) và góc ngẩng 0° là đỉnh của đầu (các giá trị dương

theo hướng xuống). Vì vậy, vị trí đối tượng p có thể được chuyển đổi thành vị trí p' theo quy ước "chung". Điều này dẫn đến vị trí đối tượng p' với

$$p' = (az', el', r)$$

$$az' = az + 90^\circ$$

$$el' = 90^\circ - el$$

với bán kính r không thay đổi.

Đồng thời, sự dịch chuyển đầu người nghe được chỉ báo bởi thông tin dịch chuyển người nghe ($az_{offset}, el_{offset}, r_{offset}$) có thể được chuyển đổi sang hệ tọa độ được xác định trước. Bằng cách sử dụng 'quy ước chung' điều này có nghĩa là

$$az'_{offset} = az_{offset} + 90^\circ$$

$$el'_{offset} = 90^\circ - el_{offset}$$

với bán kính r_{offset} không thay đổi.

Đáng chú ý, sự chuyển đổi sang hệ tọa độ được xác định trước cho cả vị trí đối tượng và sự dịch chuyển đầu người nghe có thể được thực hiện trong ngữ cảnh của bước S530 hoặc bước S540.

Việc cập nhật vị trí thực tế có thể được thực hiện trong ngữ cảnh của (ví dụ, như một phần của) bước S540 trong phương pháp 500. Sự cập nhật vị trí có thể bao gồm các bước sau đây:

Tại bước thứ nhất, vị trí p hoặc, nếu việc chuyển sang hệ tọa độ được xác định trước đã được thực hiện, thì vị trí p' , được truyền đến hệ tọa độ Đề Các (x, y, z). Trong phần sau, không dự định giới hạn, quy trình sẽ được mô tả cho vị trí p' trong hệ tọa độ được xác định trước. Ngoài ra, không dự định giới hạn, định hướng/hướng sau đây của các trục tọa độ có thể được giả định: trục x chỉ bên phải (nhìn từ đầu người nghe khi ở định hướng danh định), trục y chỉ phía trước, và trục z chỉ thẳng lên trên. Đồng thời, sự dịch chuyển đầu người nghe được chỉ báo bởi thông tin dịch chuyển người nghe ($az'_{offset}, el'_{offset}, r_{offset}$) được chuyển đổi sang các tọa độ Đề Các.

Tại bước thứ hai, vị trí đối tượng trong các tọa độ Đề Các được dịch (tịnh tiến) theo sự dịch chuyển của đầu người nghe (dịch chuyển cảnh), theo cách được mô tả ở trên. Điều này có thể tiếp tục thông qua

$$\begin{aligned}
x &= r \cdot \sin(el') \cdot \cos(az') + r_{offset} \cdot \sin(el'_{offset}) \cdot \cos(az'_{offset}) \\
y &= r \cdot \sin(el') \cdot \sin(az') + r_{offset} \cdot \sin(el'_{offset}) \cdot \sin(az'_{offset}) \\
z &= r \cdot \cos(el') + r_{offset} \cdot \cos(el'_{offset})
\end{aligned}$$

Phép tịnh tiến trên đây là ví dụ của sự sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe trong bước S540 của phương pháp 500.

Vị trí đối tượng được dịch trong các tọa độ Đề Các được chuyển đổi sang các tọa độ cầu và có thể được gọi là p'' . Vị trí đối tượng được dịch có thể được biểu diễn, trong hệ tọa độ được xác định trước theo quy ước chung là $p'' = (az'', el'', r')$.

Khi có các dịch chuyển đầu của người nghe dẫn đến thay đổi tham số bán kính *nhỏ* (tức là $r' \approx r$), thì vị trí được sửa đổi p'' của đối tượng có thể được xác định lại là $p'' = (az'', el'', r)$.

Theo một ví dụ khác, khi có các dịch chuyển của đầu người nghe là *lớn* mà có thể làm thay đổi đáng kể tham số bán kính (tức là $r' \gg r$), thì vị trí được sửa đổi p'' của đối tượng có thể còn được xác định là $p'' = (az'', el'', r')$ thay vì $p'' = (az'', el'', r)$ với tham số bán kính được sửa đổi r' .

Giá trị tương ứng của tham số bán kính được sửa đổi r' có thể thu được từ khoảng cách dịch chuyển đầu của người nghe (tức là, $r_{offset} = \|P_0 - P_1\|$) và tham số bán kính ban đầu (tức là, $r = \|P_0 - A\|$), (xem ví dụ, Fig.1 và Fig.2). Ví dụ, tham số bán kính sửa đổi r' có thể được xác định dựa vào mối quan hệ lượng giác sau đây ^{$\frac{1}{2}$} :

$$r' = (r^2 + r_{offset}^2)^{1/2}$$

Ảnh xạ tham số bán kính sửa đổi r' này thành các hệ số khuếch đại đối tượng/kênh và ứng dụng của chúng để kết xuất âm thanh tiếp theo có thể cải thiện đáng kể các hiệu quả cảm thụ của việc thay đổi mức độ do các chuyển động người dùng. Việc cho phép sửa đổi tham số bán kính r' như vậy cho phép “điểm nghe hoàn hảo thích ứng”. Điều này sẽ có nghĩa là hệ thống kết xuất MPEG điều chỉnh động vị trí điểm nghe hoàn hảo theo vị trí hiện tại của người nghe. Nói chung, việc kết xuất đối tượng âm thanh theo vị trí đối tượng được sửa đổi (hoặc được sửa đổi thêm) có thể được dựa vào tham số bán kính sửa đổi r' . Cụ thể, các hệ số khuếch đại đối tượng/kênh để kết xuất đối tượng âm thanh có thể được dựa vào (ví dụ, được sửa đổi dựa vào) tham số bán kính sửa đổi r' .

Theo một ví dụ khác, trong suốt quá trình thiết lập tái tạo loa phóng thanh và kết xuất (ví dụ, tại bước S560 ở trên), việc dịch chuyển cảnh có thể bị vô hiệu hóa. Tuy nhiên, tùy chọn cho phép dịch chuyển cảnh có thể khả dụng. Điều này cho phép bộ kết xuất 3DoF+ tạo ra điểm nghe hoàn hảo có thể điều chỉnh động theo vị trí và định hướng hiện tại của người nghe.

Đáng chú ý, bước chuyển đổi vị trí đối tượng và sự dịch chuyển của đầu người nghe sang các tọa độ Đề Các là tùy chọn và sự tịnh tiến/dịch (sửa đổi) theo sự dịch chuyển của đầu người nghe (dịch chuyển cảnh) có thể được thực hiện trong hệ tọa độ thích hợp bất kỳ. Nói cách khác, việc lựa chọn các tọa độ Đề Các ở trên cần được hiểu là một ví dụ không làm giới hạn sáng chế.

Theo một số phương án, việc xử lý dịch chuyển cảnh (bao gồm sửa đổi vị trí đối tượng và/hoặc sửa đổi thêm vị trí đối tượng được sửa đổi) có thể được cho phép hoặc bị vô hiệu hóa bởi cờ (trường, phần tử, tập bit) trong dòng bit (ví dụ, phần tử **useTrackingMode**). Các mục nhỏ “17.3 Giao diện để thiết lập và kết xuất loa phóng thanh cục bộ” và “17.4 Giao diện cho các đáp ứng xung trong phòng âm lập thể (BRIRs)” trong ISO/IEC 23008-3 chứa các mô tả của phần tử **useTrackingMode** kích hoạt việc xử lý dịch chuyển cảnh. Trong ngữ cảnh của sáng chế, phần tử **useTrackingMode** sẽ xác định (mục nhỏ 17.3) xem việc xử lý các giá trị dịch chuyển cảnh được gửi thông qua các giao diện `mpegh3daSceneDisplacementData()` và `mpegh3daPositionalSceneDisplacementData()` sẽ xảy ra hoặc không. Ngoài ra hoặc theo cách khác (mục nhỏ 17.4) trường **useTrackingMode** sẽ xác định nếu thiết bị theo dõi được kết nối và việc kết xuất âm lập thể sẽ được xử lý trong chế độ theo dõi đầu đặc biệt, nghĩa là việc xử lý các giá trị dịch chuyển cảnh được gửi thông qua các giao diện `mpegh3daSceneDisplacementData()` và `mpegh3daPositionalSceneDisplacementData()` sẽ xảy ra.

Phương pháp và hệ thống mô tả ở đây có thể được cài đặt dưới dạng phần mềm, firmware và/hoặc phần cứng. Một số thành phần nhất định có thể, ví dụ, được cài đặt dưới dạng phần mềm chạy trên bộ xử lý hoặc bộ vi xử lý tín hiệu số. Các thành phần khác có thể, ví dụ, được cài đặt dưới dạng phần cứng và hoặc dạng mạch tích hợp chuyên dụng. Các tín hiệu có trong các phương pháp và hệ thống được mô tả có thể được lưu trữ trên phương tiện như bộ nhớ truy cập ngẫu nhiên hoặc phương tiện lưu trữ quang. Chúng có thể được truyền qua các loại mạng, như mạng vô tuyến, mạng vệ tinh, mạng không dây

hoặc mạng có dây, ví dụ mạng Internet. Thiết bị thông thường sử dụng các phương pháp và hệ thống mô tả trong sáng chế là thiết bị điện tử di động hoặc trang thiết bị người dùng khác được sử dụng để lưu trữ và/hoặc kết xuất tín hiệu âm thanh.

Mặc dù tài liệu này có tham chiếu đến MPEG và cụ thể là đến âm thanh 3D MPEG-H, nhưng sáng chế sẽ không được hiểu là bị giới hạn ở các chuẩn này. Thay vào đó, như sẽ được hiểu rõ bởi người có hiểu biết trung bình trong lĩnh vực này, sáng chế có thể cũng có ứng dụng có lợi ở các chuẩn mã hóa âm thanh khác.

Hơn nữa, mặc dù tài liệu này có tham chiếu thường xuyên đến sự dịch chuyển vị trí nhỏ của đầu người nghe (ví dụ, từ vị trí nghe danh định), nhưng sáng chế không bị giới hạn ở các dịch chuyển vị trí nhỏ và có thể, nói chung, được áp dụng vào sự dịch chuyển vị trí đầu người nghe tùy chọn.

Cần lưu ý rằng phần mô tả và hình vẽ chỉ minh họa các nguyên lý của phương pháp, hệ thống và thiết bị được đề xuất. Người có hiểu biết trong lĩnh vực này sẽ có thể thực hiện các bố trí khác nhau mà, mặc dù không được mô tả và thể hiện rõ ràng ở đây, thực thi các nguyên lý của sáng chế và được bao gồm trong bản chất và phạm vi của sáng chế. Hơn thế nữa, tất cả các ví dụ và phương án được nêu trong sáng chế được dự định chủ yếu chỉ nhằm mục đích giải thích để giúp người đọc hiểu các nguyên lý của phương pháp đề xuất. Hơn thế nữa, tất cả các nội dung trình bày ở đây bộc lộ các nguyên lý, khía cạnh và phương án của sáng chế, cũng như các ví dụ cụ thể của chúng, được dự tính bao gồm cả các dạng tương đương của chúng.

Ngoài ra các nội dung trên, các phương án thực hiện ví dụ khác nhau và các phương án ví dụ của sáng chế sẽ trở nên rõ ràng qua các phương án ví dụ được đánh số (enumerated example embodiment - EEE) được liệt kê dưới đây, mà không phải là các điểm yêu cầu bảo hộ.

EEE thứ nhất đề cập đến phương pháp giải mã dòng bit tín hiệu âm thanh mã hóa, phương pháp này bao gồm các bước: nhận, bởi thiết bị giải mã âm thanh 300, dòng bit tín hiệu âm thanh mã hóa (302, 322), trong đó dòng bit tín hiệu âm thanh mã hóa bao gồm dữ liệu âm thanh mã hóa (322) và siêu dữ liệu tương ứng với ít nhất một tín hiệu âm thanh đối tượng (302); giải mã, bởi thiết bị giải mã âm thanh (300), dòng bit tín hiệu âm thanh mã hóa (302, 322) để thu được dạng biểu diễn của nhiều nguồn âm; nhận, bởi thiết bị giải

mã âm thanh (300), dữ liệu vị trí nghe (301); tạo ra, bởi thiết bị giải mã âm thanh (300), dữ liệu vị trí đối tượng âm thanh (321), trong đó dữ liệu vị trí đối tượng âm thanh (321) mô tả nhiều nguồn âm thanh liên quan đến vị trí nghe dựa vào dữ liệu vị trí nghe (301).

EEE thứ hai đề cập đến phương pháp theo EEE thứ nhất, trong đó dữ liệu vị trí nghe (301) được dựa vào tập hợp thứ nhất gồm dữ liệu vị trí tịnh tiến thứ nhất và tập hợp thứ hai gồm dữ liệu vị trí và định hướng tịnh tiến thứ hai.

EEE thứ ba đề cập đến phương pháp theo EEE thứ hai, trong đó dữ liệu vị trí tịnh tiến thứ nhất hoặc dữ liệu vị trí tịnh tiến thứ hai được dựa vào ít nhất một trong số tập hợp tọa độ cầu hoặc tập hợp tọa độ Đề Các.

EEE thứ tư đề cập đến phương pháp theo EEE thứ nhất, trong đó dữ liệu vị trí nghe (301) thu được thông qua giao diện đầu vào của bộ giải mã âm thanh 3D MPEG-H.

EEE thứ năm đề cập đến phương pháp theo EEE thứ nhất, trong đó dòng bit tín hiệu âm thanh mã hóa bao gồm các phần tử cú pháp dòng bit âm thanh 3D MPEG-H, và trong đó các phần tử cú pháp dòng bit âm thanh 3D MPEG-H bao gồm dữ liệu âm thanh mã hóa (322) và siêu dữ liệu tương ứng với ít nhất một tín hiệu âm thanh đối tượng (302).

EEE thứ sáu đề cập đến phương pháp theo EEE thứ nhất, còn bao gồm bước kết xuất, bởi thiết bị giải mã âm thanh (300) vào nhiều loa phóng thanh nhiều nguồn âm, trong đó quy trình kết xuất tuân theo ít nhất chuẩn âm thanh 3D MPEG-H.

EEE thứ bảy đề cập đến phương pháp theo EEE thứ nhất, còn bao gồm bước chuyển đổi, bởi thiết bị giải mã âm thanh (300), dựa vào phép tịnh tiến của dữ liệu vị trí nghe (301), vị trí p tương ứng với ít nhất một tín hiệu âm thanh đối tượng (302) đến vị trí thứ hai p'' tương ứng với các vị trí đối tượng âm thanh (321).

EEE thứ tám đề cập đến phương pháp theo EEE thứ bảy, trong đó vị trí p' của các vị trí đối tượng âm thanh trong hệ tọa độ được xác định trước (ví dụ, theo quy ước chung) được xác định dựa vào:

$$p' = (az', el', r)$$

$$az' = az + 90^\circ$$

$$el' = 90^\circ - el$$

$$az'_{offset} = az_{offset} + 90^\circ$$

$$el'_{offset} = 90^\circ - el_{offset}$$

trong đó az tương ứng với tham số góc phương vị thứ nhất, el tương ứng với tham số góc ngẩng thứ nhất và r tương ứng với tham số bán kính thứ nhất, ở đây az' tương ứng với tham số góc phương vị thứ hai, el' tương ứng với tham số góc ngẩng thứ hai và r' tương ứng với tham số bán kính thứ hai, trong đó az_{offset} tương ứng với tham số góc phương vị thứ ba, el_{offset} tương ứng với tham số góc ngẩng thứ ba, và az'_{offset} trong đó tương ứng với tham số góc phương vị thứ tư, el'_{offset} tương ứng với tham số góc ngẩng thứ tư.

EEE thứ chín đề cập đến phương pháp theo EEE thứ tám, trong đó vị trí đối tượng âm thanh được dịch p'' (321) của vị trí đối tượng âm thanh (302) được xác định, trong các tọa độ Đề Các (x, y, z), dựa vào:

$$x = r \cdot \sin(el') \cdot \cos(az') + x_{offset}$$

$$y = r \cdot \sin(el') \cdot \sin(az') + y_{offset}$$

$$z = r \cdot \cos(el') + z_{offset}$$

trong đó vị trí trong hệ Đề Các (x, y, z) bao gồm các tham số x, y và z và trong đó x_{offset} liên quan đến tham số độ lệch trục x thứ nhất, y_{offset} liên quan đến tham số độ lệch trục y thứ nhất, và z_{offset} liên quan đến tham số độ lệch trục z thứ nhất.

EEE thứ mười đề cập đến phương pháp theo EEE thứ chín, trong đó trong các tham số x_{offset} , y_{offset} và z_{offset} được dựa vào

$$x_{offset} = r_{offset} \cdot \sin(el'_{offset}) \cdot \cos(az'_{offset})$$

$$y_{offset} = r_{offset} \cdot \sin(el'_{offset}) \cdot \sin(az'_{offset})$$

$$z_{offset} = r_{offset} \cdot \cos(el'_{offset})$$

EEE thứ mười một đề cập đến phương pháp theo EEE thứ bảy, trong đó tham số góc phương vị az_{offset} liên quan đến vị trí góc phương vị dịch chuyển cảnh và được dựa vào:

$$az_{offset} = (sd_azimuth - 128) \cdot 1,5$$

$$az_{offset} = \min(\max(az_{offset}, -180), 180)$$

trong đó $sd_azimuth$ là tham số siêu dữ liệu góc phương vị chỉ báo dịch chuyển cảnh góc phương vị 3DA MPEG, trong đó tham số góc ngẩng el_{offset} liên quan đến vị trí góc ngẩng dịch chuyển cảnh và được dựa vào:

$$el_{offset} = (sd_elevation - 32) \cdot 3$$

$$el_{offset} = \min(\max(el_{offset}, -90), 90)$$

trong đó $sd_elevation$ là tham số siêu dữ liệu góc ngẩng cho biết dịch chuyển cảnh góc ngẩng 3DA MPEG-H, trong đó tham số bán kính r_{offset} liên quan đến bán kính dịch chuyển cảnh và được dựa vào:

$$r_{offset} = (sd_radius + 1) / 16$$

trong đó sd_radius là tham số siêu dữ liệu bán kính cho biết dịch chuyển cảnh bán kính 3DA MPEG-H, và trong đó các tham số X và Y là các biến vô hướng.

EEE thứ mười hai đề cập đến phương pháp theo EEE thứ mười, trong đó tham số x_{offset} liên quan đến vị trí độ lệch dịch chuyển cảnh sd_x trong hướng theo trục x; tham số y_{offset} liên quan đến vị trí độ lệch dịch chuyển cảnh sd_y trong hướng theo trục y; và tham số z_{offset} liên quan đến vị trí độ lệch dịch chuyển cảnh sd_z trong hướng theo trục z.

EEE thứ mười ba đề cập đến phương pháp theo EEE thứ nhất, còn bao gồm việc nội suy, bởi thiết bị giải mã âm thanh, dữ liệu vị trí thứ nhất liên quan đến dữ liệu vị trí nghe (301) và tín hiệu âm thanh đối tượng (102) tại tốc độ cập nhật.

EEE thứ mười bốn đề cập đến phương pháp theo EEE thứ nhất, còn bao gồm bước xác định, bởi thiết bị giải mã âm thanh 300, mã hóa entropy hiệu quả của dữ liệu vị trí nghe (301).

EEE thứ mười lăm đề cập đến phương pháp theo EEE thứ nhất, trong đó dữ liệu vị trí liên quan đến vị trí nghe (301) được suy ra dựa vào thông tin cảm biến.

YÊU CẦU BẢO HỘ

1. Phương pháp (500) xử lý thông tin vị trí biểu thị vị trí đối tượng của đối tượng âm thanh, trong đó việc xử lý được thực hiện bằng cách sử dụng bộ giải mã âm thanh 3D MPEG-H, trong đó vị trí đối tượng là có thể sử dụng để kết xuất đối tượng âm thanh, phương pháp này bao gồm các bước:

thu nhận (S510) thông tin định hướng người nghe biểu thị sự định hướng đầu của người nghe;

thu nhận (S520) thông tin dịch chuyển người nghe biểu thị sự dịch chuyển đầu của người nghe so với vị trí nghe danh định, thông qua giao diện đầu vào của bộ giải mã âm thanh 3D MPEG-H;

xác định (S530) vị trí đối tượng từ thông tin vị trí;

sửa đổi (S540) vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng; và

sửa đổi thêm (S550) vị trí đối tượng được sửa đổi dựa vào thông tin định hướng người nghe, trong đó khi thông tin dịch chuyển người nghe biểu thị dịch chuyển đầu của người nghe từ vị trí nghe danh định một khoảng dịch chuyển vị trí nhỏ, khoảng dịch chuyển vị trí nhỏ này có giá trị tuyệt đối bằng 0,5 mét hoặc nhỏ hơn 0,5 mét, khoảng cách giữa vị trí đối tượng âm thanh được sửa đổi và vị trí nghe sau khi phần đầu của người nghe dịch chuyển được giữ bằng với khoảng cách gốc giữa vị trí đối tượng âm thanh và vị trí nghe danh định.

2. Phương pháp (500) theo điểm 1, trong đó:

bước sửa đổi (S540) vị trí đối tượng và sửa đổi thêm (S550) vị trí đối tượng được sửa đổi được thực hiện sao cho đối tượng âm thanh, sau khi được kết xuất cho một hoặc nhiều loa thực hoặc ảo theo vị trí đối tượng được sửa đổi thêm, được người nghe cảm nhận về mặt tâm thính là bắt nguồn từ vị trí cố định so với vị trí nghe danh định, bất kể sự dịch chuyển đầu của người nghe từ vị trí nghe danh định và định hướng của đầu người nghe so với sự định hướng danh định.

3. Phương pháp (500) theo điểm 1 hoặc 2, trong đó:

bước sửa đổi (S540) vị trí đối tượng dựa vào thông tin dịch chuyển người nghe được thực hiện bằng cách tịnh tiến vị trí đối tượng một khoảng bằng khoảng dịch chuyển đầu của người nghe từ vị trí nghe danh định, nhưng theo hướng ngược lại.

4. Phương pháp (500) theo điểm bất kỳ trong số các điểm từ 1 đến 3, trong đó:

thông tin dịch chuyển người nghe biểu thị sự dịch chuyển đầu của người nghe từ vị trí nghe danh định có thể đạt được bằng cách người nghe chuyển động thân trên và/hoặc đầu của mình.

5. Phương pháp (500) theo điểm bất kỳ trong số các điểm từ 1 đến 4, phương pháp này còn bao gồm bước:

phát hiện sự định hướng của đầu người nghe bởi thiết bị mang được và/hoặc cố định.

6. Phương pháp (500) theo điểm bất kỳ trong số các điểm từ 1 đến 5, phương pháp này còn bao gồm bước:

phát hiện sự dịch chuyển đầu của người nghe từ vị trí nghe danh định bởi thiết bị mang được và/hoặc cố định.

7. Phương pháp (500) theo điểm bất kỳ trong số các điểm từ 1 đến 6, trong đó khoảng cách giữa vị trí đối tượng âm thanh được sửa đổi và vị trí nghe sau khi dịch chuyển được ánh xạ thành các hệ số khuếch đại để sửa đổi mức âm thanh.

8. Bộ giải mã âm thanh 3D MPEG-H (300) xử lý thông tin vị trí biểu thị vị trí đối tượng (321) của đối tượng âm thanh, trong đó vị trí đối tượng là có thể sử dụng để kết xuất đối tượng âm thanh, bộ giải mã này bao gồm bộ xử lý và bộ nhớ được ghép nối với bộ xử lý, trong đó bộ xử lý được làm thích ứng để:

thu nhận thông tin định hướng người nghe biểu thị sự định hướng đầu của người nghe;

thu nhận thông tin dịch chuyển người nghe biểu thị sự dịch chuyển đầu của người nghe so với vị trí nghe danh định, thông qua giao diện đầu vào của bộ giải mã âm thanh 3D MPEG-H;

xác định vị trí đối tượng từ thông tin vị trí;

sửa đổi vị trí đối tượng dựa vào thông tin dịch chuyển người nghe bằng cách áp dụng phép tịnh tiến cho vị trí đối tượng; và

sửa đổi thêm vị trí đối tượng được sửa đổi dựa vào thông tin định hướng người nghe, trong đó

khi thông tin dịch chuyển người nghe biểu thị dịch chuyển đầu của người nghe từ vị trí nghe danh định một khoảng dịch chuyển vị trí nhỏ, khoảng dịch chuyển vị trí nhỏ có giá trị tuyệt đối bằng 0,5 mét hoặc nhỏ hơn 0,5 mét, bộ xử lý được tạo cấu hình để giữ khoảng cách giữa vị trí đối tượng âm thanh được sửa đổi và vị trí nghe sau khi phân đầu của người nghe dịch chuyển bằng với khoảng cách gốc giữa vị trí đối tượng âm thanh và vị trí nghe danh định.

9. Phương tiện đọc được bằng máy tính lưu trữ trên đó phần mềm máy tính bao gồm các lệnh, mà khi phần mềm được thực thi bởi bộ xử lý hoặc bộ vi xử lý tín hiệu số, khiến cho bộ xử lý hoặc bộ vi xử lý tín hiệu số thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 7.

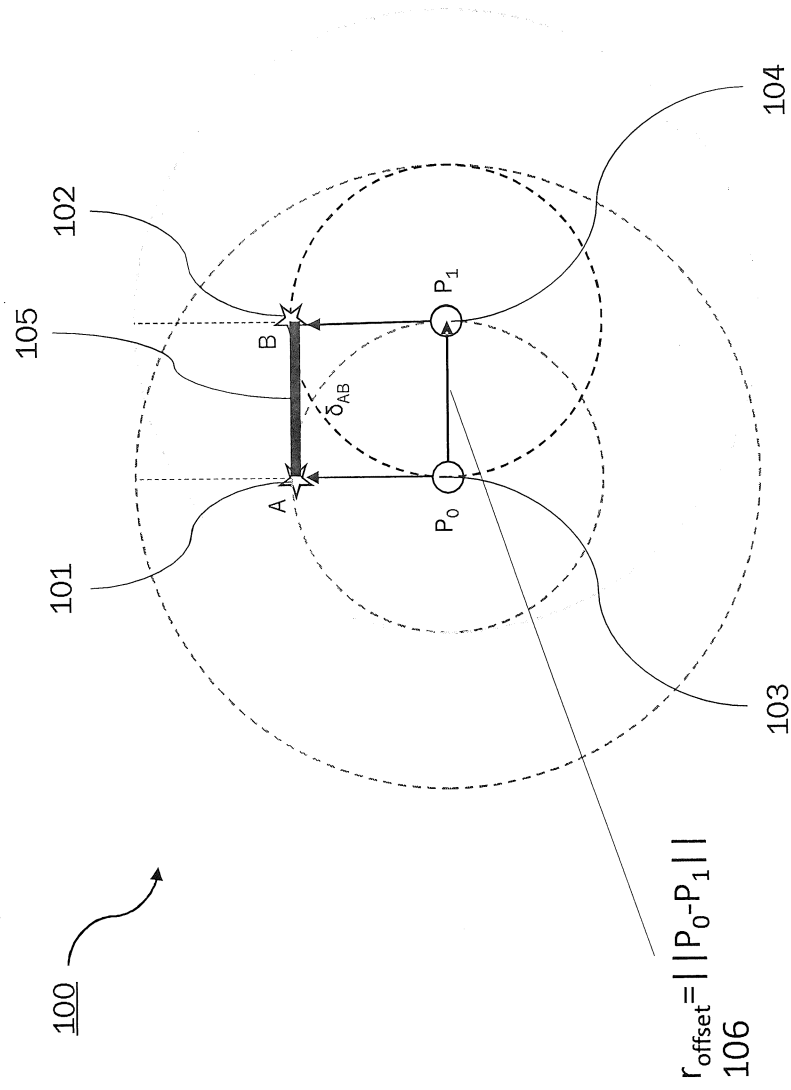


Fig.1

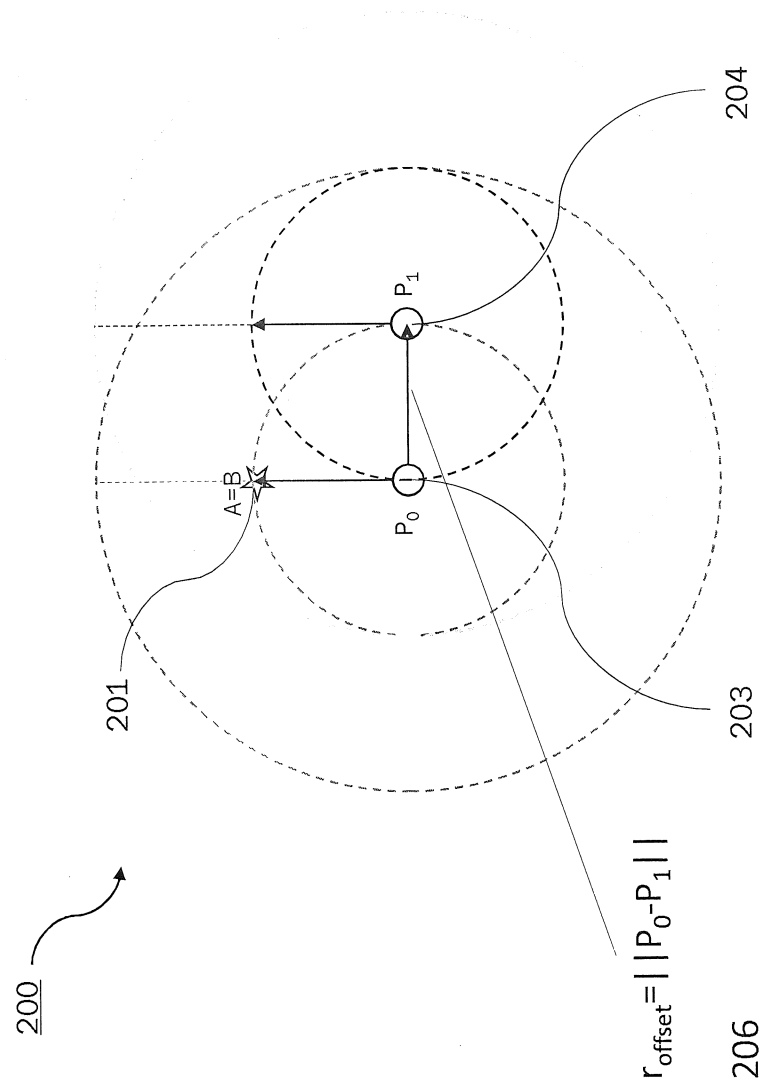


Fig.2

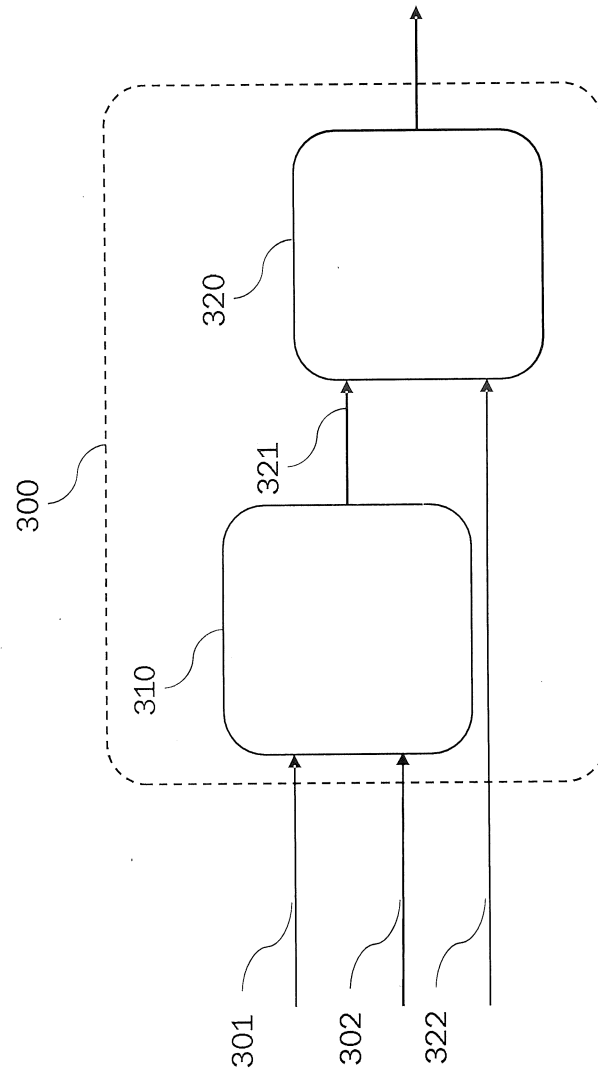


Fig.3

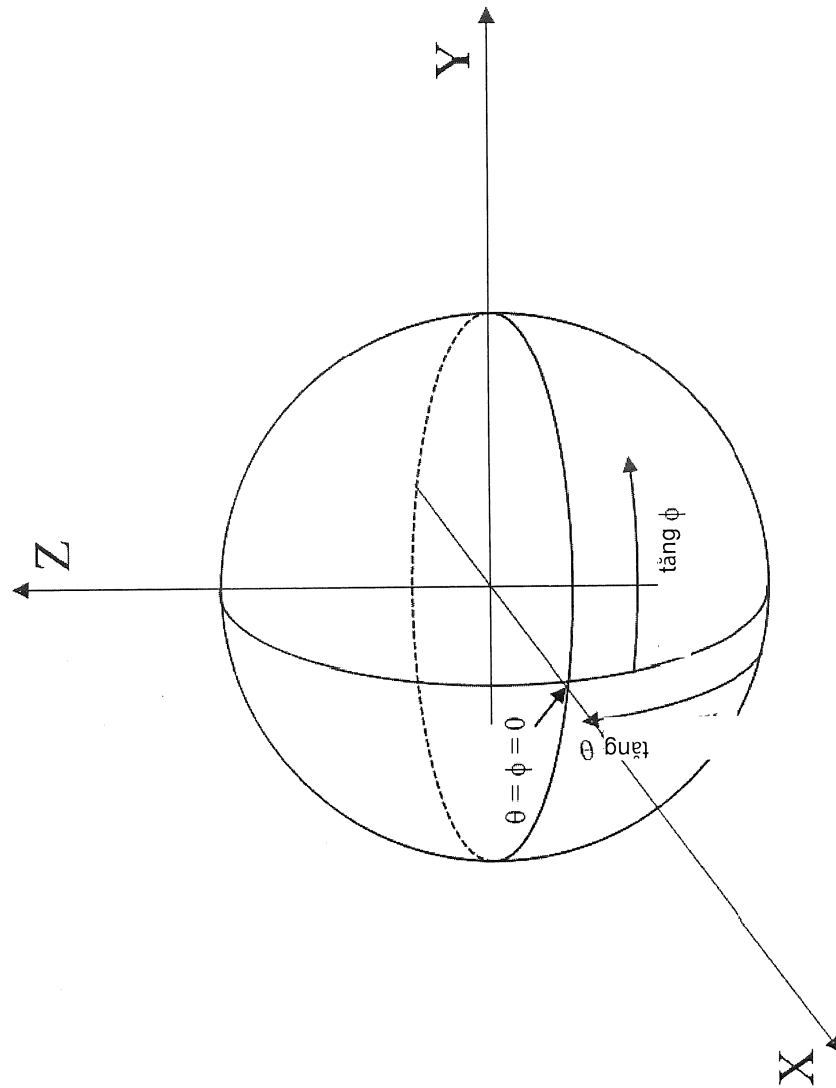


Fig.4

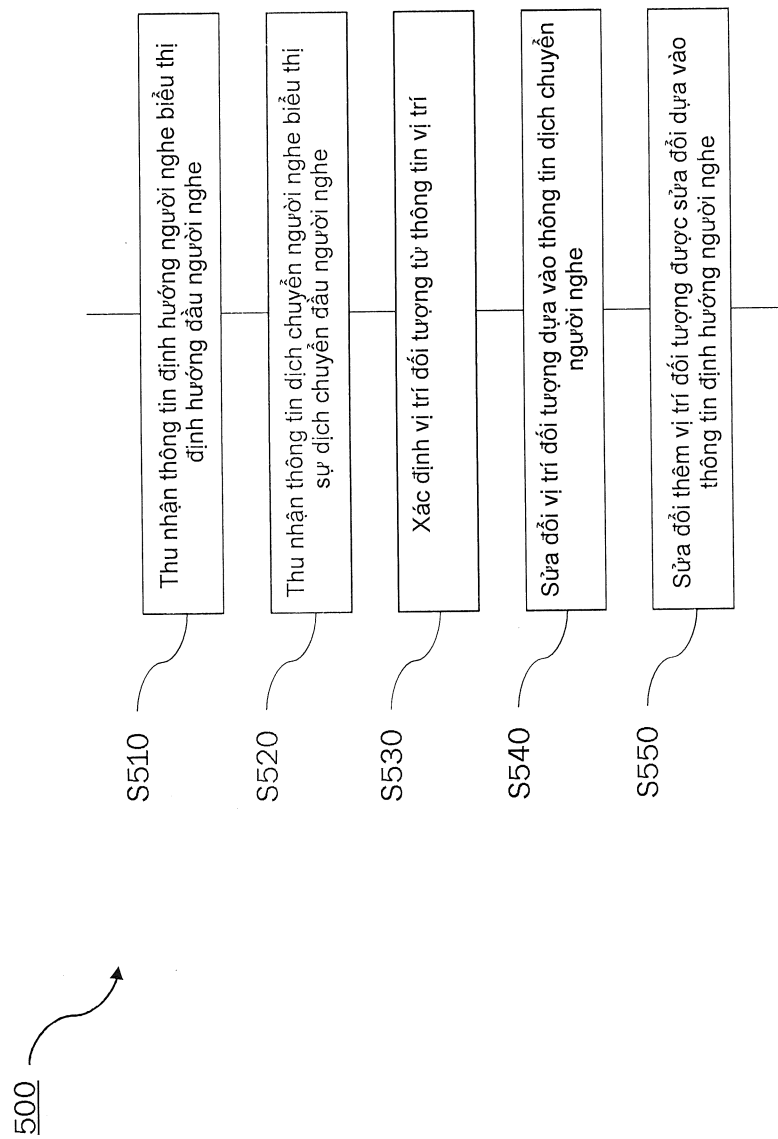


Fig.5