



(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)



1-0042489

(51)^{2019.01} G10L 19/008; H04S 7/00; G10L
21/0272; G06F 3/01

(13) B

(21) 1-2020-00767

(22) 13/07/2018

(86) PCT/EP2018/069147 13/07/2018

(87) WO 2019/012135 17/01/2019

(30) 17181481.7 14/07/2017 EP

(45) 27/01/2025 442

(43) 25/05/2020 386

(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany

(72) HABETS, Emanuel (NL); HERRE, Juergen (DE); PLINGE, Axel (DE);
THIERGART, Oliver (DE); KUECH, Fabian (DE).

(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP TẠO MÔ TẢ TRƯỜNG ÂM THANH TĂNG
CƯỜNG, THIẾT BỊ VÀ PHƯƠNG PHÁP TẠO MÔ TẢ TRƯỜNG ÂM THANH
BIẾN ĐỔI

(21) 1-2020-00767

(57) Sáng chế đề cập đến thiết bị và phương pháp tạo mô tả trường âm thanh tăng cường, thiết bị và phương pháp tạo mô tả trường âm thanh biến đổi. Thiết bị tạo mô tả trường âm thanh tăng cường bao gồm: bộ tạo trường âm thanh (100) để tạo ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một vị trí tham chiếu; và bộ tạo siêu dữ liệu (110) để tạo ra siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh, trong đó ít nhất một mô tả trường âm thanh và siêu dữ liệu cấu thành mô tả trường âm thanh tăng cường. Siêu dữ liệu có thể là bản đồ chiều sâu kết hợp thông tin khoảng cách với hướng trong toàn bộ băng hoặc băng con, tức là, ngăn thời gian tần số.

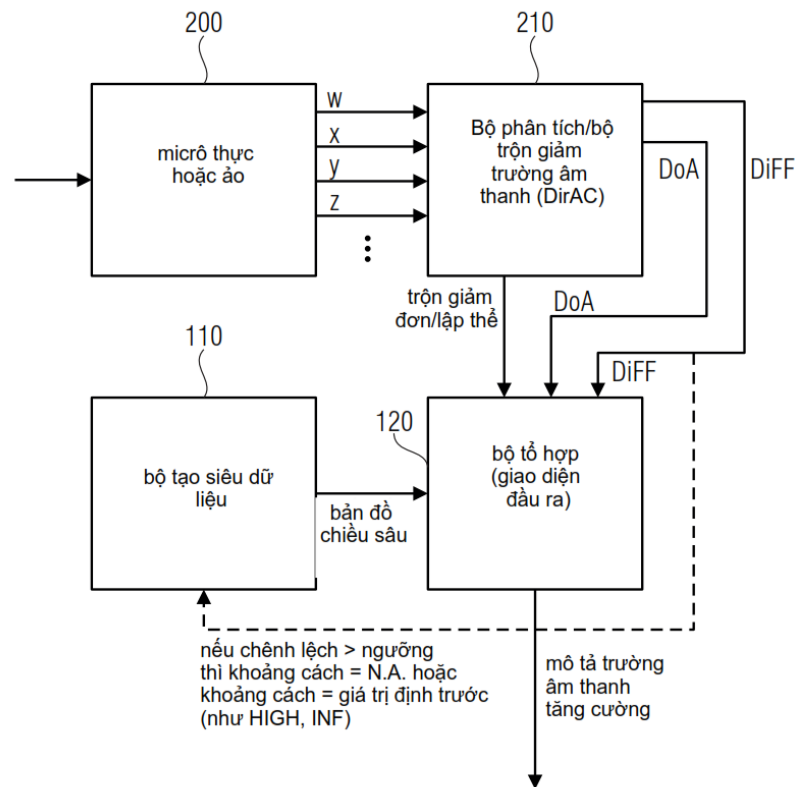


Fig. 2

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc xử lý âm thanh và, cụ thể, đề cập đến xử lý âm thanh liên quan đến trường âm thanh mà được xác định đối với vị trí tham chiếu như micro hoặc vị trí micro ảo.

Tình trạng kỹ thuật của sáng chế

Các tín hiệu Ambisonic (định dạng âm thanh vòm dạng hình cầu) bao gồm phép phân tích sóng hài hình cầu rút gọn của trường âm thanh. Ambisonic đến theo nhiều kiểu khác nhau. Có Ambisonic 'truyền thống' [31] mà ngày nay được biết đến là 'Ambisonic cấp một' (FOA - First-Order Ambisonic) và bao gồm bốn tín hiệu (nghĩa là, một tín hiệu đẳng hướng và lên đến ba tín hiệu định hướng hình số tám). Các biến thể Ambisonic gần đây được biết đến là 'Ambisonic cấp cao' (HOA - Higher-Order Ambisonic) và cung cấp độ phân giải không gian tăng cường và vùng điểm lý tưởng của người nghe lớn hơn với phí tổn của việc mang nhiều tín hiệu hơn. Nói chung, sự biểu diễn HOA cấp N được xác định đầy đủ bao gồm $(N+1)^2$ tín hiệu.

Liên quan đến ý tưởng Ambisonic, sự biểu diễn mã hóa âm thanh định hướng (DirAC - Directional Audio Coding) đã được nhận thức để biểu diễn cảnh âm thanh FOA hoặc HOA theo kiểu tham số gọn hơn. Cụ thể hơn, cảnh âm thanh không gian được biểu diễn bởi một (hoặc nhiều) kênh âm thanh được truyền mà biểu diễn sự trộn giảm của cảnh âm thanh và thông tin phụ liên quan của hướng và sự khuếch tán trong mỗi ngăn thời gian-tần số (TF - time-frequency). Nhiều thông tin hơn về DirAC có thể được tìm thấy trong [32, 33].

DirAC [32] có thể được sử dụng với các hệ thống micro khác nhau và với các cài đặt loa phóng thanh tùy chọn. Mục đích của hệ thống DirAC là tái tạo ấn tượng không gian của môi trường âm thanh hiện có chính xác nhất có thể bằng cách sử dụng hệ thống đa kênh/loa phóng thanh 3D. Trong môi trường được chọn, các phản hồi (các phản hồi âm thanh liên tục hoặc các phản hồi xung) được đo bằng micro đẳng hướng (W) và với tập hợp các micro cho phép đo hướng đến của âm thanh và sự khuếch tán của âm thanh. Phương pháp phổ biến là áp dụng ba micro hình số tám (X, Y, Z) được xếp hàng theo trục tọa độ Đề-các tương ứng [34]. Cách để thực hiện phương pháp này

là sử dụng micrô trường âm thanh, mang lại trực tiếp các phản hồi mong muốn. Các tín hiệu W, X, Y và Z cũng có thể được tính toán từ tập hợp các micrô đẳng hướng rời rạc.

Trong DirAC, tín hiệu âm thanh trước tiên được chia thành các kênh tần số theo tính chọn lọc tần số của hệ thống thính giác của người. Hướng và sự khuếch tán âm thanh được đo phụ thuộc vào thời gian tại mỗi kênh tần số. Trong quá trình truyền, một hoặc nhiều kênh âm thanh được gửi, cùng với dữ liệu về hướng và sự khuếch tán được phân tích. Trong quá trình tổng hợp, âm thanh mà được áp dụng cho các loa phóng thanh có thể là, ví dụ, kênh đẳng hướng W, hoặc âm thanh cho mỗi loa phóng thanh có thể được tính toán là tổng có trọng số của W, X, Y, và Z, mà tạo thành tín hiệu có các đặc trưng định hướng nhất định cho mỗi loa phóng thanh. Mỗi kênh âm thanh được chia thành các kênh tần số, mà sau đó được chia tùy ý cho các dòng khuếch tán và các dòng không khuếch tán tùy thuộc vào sự khuếch tán được phân tích. Dòng khuếch tán được tái tạo với kỹ thuật, mà tạo ra cảm quan khuếch tán của cảnh âm thanh, ví dụ, các kỹ thuật giải tương quan được sử dụng trong mã hóa tín hiệu điều khiển lập thể [35-37]. Âm thanh không khuếch tán được tái tạo với kỹ thuật nhằm tạo ra nguồn ảo giống như điểm theo dữ liệu hướng (ví dụ, VBAP [38]).

Ba kỹ thuật điều hướng trong 6DoF với mức độ tự do hạn chế được đề xuất trong [39]. Đưa ra tín hiệu Ambisonic đơn, tín hiệu Ambisonic đơn được tính toán bằng cách sử dụng: 1) mô phỏng phát lại HOA và chuyển động của người nghe trong dãy loa phóng thanh ảo, 2) tính toán và tịnh tiến dọc theo các sóng phẳng, và 3) mở rộng lại trường âm thanh xung quanh người nghe.

Hơn nữa, tham chiếu đến công nghệ DirAC như được mô tả, ví dụ, trong tài liệu công bố "mã hóa âm thanh định hướng - tái tạo dựa trên cảm quan của âm thanh không gian" (Directional Audio Coding – Perception-Based Reproduction of Spatial Sound), của tác giả V. Pulkki và cộng sự, hội thảo quốc tế về các nguyên tắc và ứng dụng của thính giác không gian, 11-13 tháng 11, 2009 tại Zao, Miyagi, Nhật Bản. Sự tham chiếu này mô tả việc mã hóa âm thanh định hướng làm ví dụ cho vị trí tham chiếu liên quan đến việc xử lý trường âm thanh đặc biệt là kỹ thuật được thúc đẩy theo cảm quan cho việc xử lý âm thanh không gian.

Sự tái tạo của cảnh âm thanh thường tập trung vào các cài đặt loa phóng thanh, vì đây là sự tái tạo điển hình riêng, ví dụ như phòng khách, và bối cảnh chuyên nghiệp

ví dụ như các rạp chiếu phim. Ở đây, mối quan hệ của cảnh với dạng hình tái tạo là tĩnh vì nó đi kèm hình ảnh hai chiều mà buộc người nghe phải nhìn theo hướng phía trước. Sau đó, mối quan hệ không gian của các đối tượng âm thanh và các đối tượng trực quan được định rõ và cố định tại thời điểm sản xuất.

Trong thực tế ảo (VR - virtual reality), sự đắm chìm đạt được rõ ràng bằng cách cho phép người dùng di chuyển tự do trong cảnh. Do đó, cần phải theo dõi chuyển động của người dùng và điều chỉnh sự tái tạo thị giác và thính giác tới vị trí của người dùng. Thông thường, người dùng đang đeo màn hình đeo trên đầu (HMD - head-mounted display) và tai nghe chụp đầu. Để có trải nghiệm đắm chìm với tai nghe chụp đầu, âm thanh phải được lập thể hóa. Sự lập thể hóa là sự mô phỏng về cách mà đầu, tai, và thân trên của con người thay đổi âm thanh của nguồn tùy thuộc vào hướng và khoảng cách của nó. Điều này đạt được bằng cách tích chập các tín hiệu với các hàm truyền liên quan đến đầu (HRTF - head-related transfer function) cho hướng tương đối của chúng [1, 2]. Sự lập thể hóa cũng làm cho âm thanh xuất hiện đến từ cảnh hơn là từ bên trong đầu [3]. Kịch bản phổ biến đã được giải quyết thành công là tái tạo video 360° [4, 5]. Ở đây, người dùng đang đeo HMD hoặc cầm máy tính bảng hoặc điện thoại trên tay. Bằng cách di chuyển đầu của họ hoặc thiết bị, người dùng có thể nhìn xung quanh theo hướng bất kỳ. Đây là kịch bản ba độ tự do (3DoF - three-degrees-of-freedom), vì người dùng có ba độ chuyển động (dọc, ngang, xoay). Một cách trực quan, điều này được nhận ra bằng cách chiếu video trên hình cầu xung quanh người dùng. Âm thanh thường được ghi lại bằng micrô không gian [6], ví dụ, Ambisonic cấp một (FOA), gắn với máy quay video. Trong miền Ambisonic, sự xoay đầu của người dùng được làm thích ứng theo cách đơn giản. Âm thanh sau đó được, ví dụ, kết xuất đến các loa phóng thanh được đặt xung quanh người dùng. Các tín hiệu loa phóng thanh ảo sau đó được lập thể hóa.

Các ứng dụng VR hiện đại cho phép sáu độ tự do (6DoF - six-degrees-of-freedom). Ngoài sự xoay đầu, người dùng có thể di chuyển xung quanh dẫn đến tịnh tiến vị trí của họ theo ba chiều theo không gian. Sự tái tạo 6DoF bị giới hạn bởi kích thước tổng thể của vùng đi bộ. Trong nhiều trường hợp, vùng này khá nhỏ, ví dụ, phòng khách thông thường. 6DoF thường gặp trong các trò chơi VR. Ở đây, toàn bộ cảnh được tổng hợp với hình ảnh do máy tính tạo ra (CGI - computer-generated imagery). Âm thanh thường được tạo bằng cách sử dụng việc kết xuất dựa trên đối

tượng trong đó mỗi đối tượng âm thanh được kết xuất với độ khuếch đại phụ thuộc vào khoảng cách và hướng tương đối từ người dùng dựa trên dữ liệu theo dõi. Hiện thực có thể được tăng cường bằng tiếng vang và nhiễu xạ nhân tạo [8, 9, 10].

Về nội dung được ghi lại, có một số thách thức riêng biệt để đảm bảo sự tái tạo 6DoF nghe nhìn. Ví dụ đã nêu về thao tác âm thanh không gian trong miền tịnh tiến không gian là các kỹ thuật 'phóng đại âm thanh' [11, 12]. Ở đây, vị trí người nghe hầu như được di chuyển vào cảnh thị giác được ghi lại, tương tự như phóng đại hình ảnh. Người dùng lựa chọn một hướng hoặc một phần hình ảnh và sau đó có thể nghe điều này từ điểm được tịnh tiến. Điều này đòi hỏi rằng tất cả các hướng đến (DoA - direction of arrival) đang thay đổi so với tái tạo ban đầu, không phóng đại.

Các phương pháp tái tạo 6DoF của nội dung được ghi lại đã sử dụng các vị trí ghi lại được phân bố theo không gian đã được đề xuất. Đối với video, các dây của các máy quay có thể được dùng để tạo kết xuất trường ánh sáng [13]. Đối với âm thanh, cài đặt tương tự dùng các dây micrô được phân bố hoặc các micrô Ambisonic. Đã được thể hiện rằng có thể tạo ra tín hiệu 'micrô ảo' được đặt tại vị trí tùy chọn từ các bản ghi như vậy [14].

Để nhận ra các biến đổi âm thanh không gian như vậy theo cách thuận tiện về mặt kỹ thuật, có thể dùng các kỹ thuật xử lý hoặc mã hóa âm thanh tham số (xem [15] để tham khảo tổng quát). Mã hóa âm thanh định hướng (DirAC - Directional audio coding) [16] là phương pháp phổ biến để biến đổi bản ghi thành sự biểu diễn mà bao gồm phổ âm thanh và thông tin phụ tham số về hướng âm thanh và sự khuếch tán. Nó được sử dụng cho phóng đại âm thanh [11] và các ứng dụng micrô ảo [14].

Phương pháp được đề xuất ở đây cho phép sự tái tạo 6DoF từ bản ghi của micrô FOA. Các bản ghi từ vị trí không gian đơn được sử dụng cho sự tái tạo 3DoF hoặc phóng đại âm thanh. Nhưng, theo sự hiểu biết của các tác giả sáng chế, cho đến nay, không có phương pháp nào tái tạo hoàn toàn 6DoF từ dữ liệu như vậy được đề xuất. Người ta nhận ra tái tạo 6DoF bằng cách tích hợp thông tin về khoảng cách của các nguồn âm thanh trong bản ghi. Thông tin khoảng cách này được hợp nhất vào sự biểu diễn tham số của DirAC, sao cho quan điểm thay đổi của người nghe được ánh xạ chính xác. Để đánh giá với thử nghiệm nghe, nhiều kích thích với tham chiếu ẩn và mô hình neo (MUSHRA) [17] được làm thích ứng cho VR. Bằng cách sử dụng CGI và âm thanh được tạo ra theo cách tổng hợp, có thể tạo ra sự tham chiếu dựa trên đối tượng

để so sánh. Việc ghi FOA ảo diễn ra ở vị trí được theo dõi của người dùng, thể hiện các tín hiệu được điều chỉnh 6DoF. Ngoài phương pháp được đề xuất, việc tái tạo mà không có thông tin khoảng cách và sự tịnh tiến được biểu diễn như các điều kiện trong thử nghiệm nghe

Không có biểu diễn trường âm thanh Ambisonic nào (có thể là Ambisonic FOA hoặc HOA thông thường hoặc là biểu diễn trường âm thanh tham số theo kiểu DirAC) cung cấp đủ thông tin để cho phép dịch chuyển tịnh tiến vị trí của người nghe khi nó cần cho các ứng dụng 6DoF vì không phải khoảng cách đối tượng cũng không phải các vị trí đối tượng tuyệt đối trong cảnh âm thanh được xác định trong các định dạng này. Cần lưu ý rằng sự dịch chuyển vị trí của người nghe có thể được tịnh tiến thành sự dịch chuyển tương đương của cảnh âm thanh theo hướng ngược lại.

Vấn đề diễn hình khi di chuyển trong 6DoF được minh họa trên Fig.1b. Giả sử rằng cảnh âm thanh được mô tả tại vị trí A bằng cách sử dụng Ambisonic. Trong trường hợp này, các âm thanh từ nguồn A và nguồn B đến từ cùng một hướng, tức là, chúng có cùng hướng đến (DOA - direction-of-arrival). Trong trường hợp một người di chuyển đến vị trí B, DOA của nguồn A và nguồn B là khác nhau. Sử dụng mô tả Ambisonic tiêu chuẩn của trường âm thanh, nghĩa là, không có thông tin bổ sung, không thể tính toán các tín hiệu Ambisonic tại vị trí B cho các tín hiệu Ambisonic tại vị trí A.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là một mặt đề xuất sự mô tả trường âm thanh tăng cường hoặc mặt khác tạo mô tả trường âm thanh biến đổi cho phép việc xử lý được cải thiện hoặc linh hoạt hoặc hiệu quả.

Mục đích của sáng chế được bởi các đối tượng trong các điểm yêu cầu bảo hộ độc lập.

Sáng chế dựa trên việc phát hiện rằng các mô tả trường âm thanh diễn hình mà liên quan đến vị trí tham chiếu cần thông tin bổ sung để các mô tả trường âm thanh này có thể được xử lý để mô tả trường âm thanh biến đổi mà không liên quan đến vị trí tham chiếu ban đầu nhưng liên quan đến vị trí tham chiếu khác có thể được tính toán. Để đạt được mục đích này, siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh được tạo và siêu dữ liệu cùng với sự mô tả trường âm thanh tương ứng với sự mô tả trường âm thanh tăng cường mà có thể, ví dụ, được truyền hoặc được lưu trữ.

Để tạo mô tả trường âm thanh biến đổi từ sự mô tả trường âm thanh và siêu dữ liệu và, cụ thể, siêu dữ liệu liên quan đến thông tin không gian của sự mô tả trường âm thanh, trường âm thanh biến đổi được tính toán bằng cách sử dụng thông tin không gian này, sự mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác. Do đó, sự mô tả trường âm thanh tăng cường bao gồm sự mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh này nằm dưới sự mô tả trường âm thanh được xử lý để thu được mô tả trường âm thanh biến đổi liên quan đến vị trí tham chiếu khác được định rõ bởi thông tin tịnh tiến bổ sung mà có thể, ví dụ, được cung cấp hoặc được sử dụng tại phía bộ giải mã.

Tuy nhiên, sáng chế không chỉ liên quan đến kịch bản bộ mã hóa/bộ giải mã, mà còn có thể được áp dụng trong ứng dụng mà tại đó cả hai việc tạo mô tả trường âm thanh tăng cường và tạo mô tả trường âm thanh biến đổi diễn ra một cách cơ bản trên một và cùng một vị trí. Sự mô tả trường âm thanh biến đổi có thể, ví dụ, là sự mô tả của chính trường âm thanh biến đổi hoặc thực tế trường âm thanh biến đổi trong các tín hiệu kênh, các tín hiệu lập thể hoặc, một lần nữa, trường âm thanh liên quan đến vị trí tham chiếu mà, tuy nhiên, liên quan đến vị trí tham chiếu mới hoặc khác hơn là vị trí tham chiếu ban đầu. Chẳng hạn, ứng dụng như vậy sẽ ở trong kịch bản thực tế ảo mà sự mô tả trường âm thanh cùng với siêu dữ liệu tồn tại và ở đó người nghe di chuyển ra khỏi vị trí tham chiếu mà trường âm thanh được đưa ra và di chuyển đến vị trí tham chiếu khác và tròn đó, sau đó, trường âm thanh cho người nghe di chuyển xung quanh trong khu vực ảo được tính toán tương ứng với trường âm thanh nhưng ngay lúc đó tại vị trí tham chiếu mà người dùng đã di chuyển tới.

Trong phương án ưu tiên, bộ tạo trường âm thanh tạo mô tả DirAC của trường âm thanh có một hoặc nhiều tín hiệu trộn giảm và dữ liệu hướng riêng và dữ liệu khuếch tán tùy chọn cho các ngăn thời gian-tần số khác nhau. Trong ngữ cảnh này, bộ tạo siêu dữ liệu được tạo cấu hình để tạo ra thông tin khoảng cách hoặc độ sâu riêng bổ sung cho các ngăn thời gian-tần số khác nhau như siêu dữ liệu. Cụ thể, và trong phương án ưu tiên, siêu dữ liệu liên quan đến thông tin không gian sẽ là bản đồ chiều sâu kết hợp khoảng cách nhất định với thông tin vị trí nhất định như thông tin hướng đến.

Trong các phương án, hướng đến được cho bởi chỉ góc nâng hoặc góc phương vị hoặc cả hai góc và bản đồ chiều sâu sau đó sẽ kết hợp, với từng thông tin vị trí hoặc thông tin hướng đến (thông tin DoA) thông tin khoảng cách nhất định như khoảng cách theo mét hoặc khoảng cách tương đối hoặc khoảng cách tuyệt đối hoặc tương đối được lượng tử hóa hoặc thông tin khoảng cách khác bất kỳ, mà từ đó, cuối cùng, khoảng cách đối với vị trí tham chiếu, mà trường âm thanh có liên quan đến có thể được suy ra.

Sau đó, các phương án triển khai ưu tiên khác được đưa ra.

Ambisonic trở thành một trong những định dạng được sử dụng phổ biến nhất cho âm thanh 3D trong bối cảnh của các ứng dụng thực tế ảo, tăng cường và hỗn hợp. Một loạt các công cụ tiếp nhận và sản xuất âm thanh được phát triển để tạo ra tín hiệu đầu ra ở định dạng Ambisonic. Để diễn tả nội dung được mã hóa Ambisonic trong các ứng dụng thực tế ảo (VR) tương tác, định dạng Ambisonic được chuyển đổi thành tín hiệu lập thể hoặc các kênh để tái tạo. Trong các ứng dụng được nêu ở trên, người nghe thường có thể thay đổi một cách tương hỗ sự định hướng của họ trong cảnh được diễn tả đến mức họ có thể xoay đầu trong cảnh âm thanh cho phép ba độ tự do (3DoF, nghĩa là, dọc, ngang, xoay) và vẫn trải nghiệm chất lượng âm thanh phù hợp. Điều này được thực hiện bằng cách xoay cảnh âm thanh trước khi kết xuất theo định hướng của đầu, mà có thể được thực hiện với độ phức tạp tính toán thấp và là lợi thế của sự biểu diễn Ambisonic. Tuy nhiên, trong các ứng dụng mới nổi, chẳng hạn như VR, mong muốn cho phép người dùng di chuyển tự do trong cảnh âm thanh hơn là chỉ thay đổi định hướng (được gọi là 'sáu độ tự do' hoặc 6DoF). Do đó, việc xử lý tín hiệu là cần thiết để thay đổi quan điểm về cảnh âm thanh (nghĩa là để di chuyển ảo trong cảnh âm thanh dọc theo các trục x, y, hoặc z). Tuy nhiên, nhược điểm chính của Ambisonic là định dạng mô tả trường âm thanh từ một quan điểm trong cảnh âm thanh. Cụ thể, nó không chứa thông tin về vị trí thực tế của các nguồn âm thanh trong cảnh âm thanh mà sẽ cho phép dịch chuyển cảnh âm thanh ('tịnh tiến') như được yêu cầu cho 6DoF. Mô tả sáng chế này cung cấp các mở rộng của Ambisonic để khắc phục vấn đề này và cũng tạo điều kiện cho sự tịnh tiến, và do đó cho phép 6DoF chính xác.

Những bản ghi Ambisonic cấp một (FOA) có thể được xử lý và được tái tạo qua các tai nghe chụp đầu. Chúng có thể được xoay để tính đến sự định hướng đầu của người nghe. Tuy nhiên, các hệ thống thực tế ảo (VR) cho phép người nghe di chuyển

trong sáu độ tự do (6DoF), nghĩa là, ba vòng xoay cộng với ba độ tịnh tiến tự do. Ở đây, các góc và khoảng cách rõ ràng của các nguồn âm thanh phụ thuộc vào vị trí của người nghe. Kỹ thuật để tạo điều kiện cho 6DoF được mô tả. Cụ thể, bản ghi FOA được mô tả sử dụng mô hình tham số, mà được biến đổi dựa trên vị trí của người nghe và thông tin về các khoảng cách tới các nguồn. Phương pháp được đánh giá thử nghiệm nghe, so sánh các kết xuất lập thể khác nhau của cảnh âm thanh tổng hợp trong đó người nghe có thể di chuyển tự do.

Trong các phương án ưu tiên tiếp theo, mô tả trường âm thanh tăng cường được đưa ra bởi giao diện đầu ra để tạo tín hiệu đầu ra để truyền hoặc lưu trữ, trong đó tín hiệu đầu ra bao gồm, đối với khung thời gian, một hoặc nhiều tín hiệu âm thanh được suy ra từ trường âm thanh và thông tin không gian cho khung thời gian. Đặc biệt, bộ tạo trường âm thanh trong các phương án tiếp theo thích ứng để suy ra dữ liệu hướng từ trường âm thanh, dữ liệu hướng liên quan đến hướng đến của âm thanh trong một khoảng thời gian hoặc ngăn tần số và bộ tạo siêu dữ liệu được tạo cấu hình để suy ra thông tin không gian như các mục dữ liệu kết hợp thông tin khoảng cách với dữ liệu hướng.

Đặc biệt, trong phương án như vậy, giao diện đầu ra được tạo cấu hình để tạo các tín hiệu đầu ra sao cho các mục dữ liệu đối với các khung thời gian được liên kết với các dữ liệu hướng cho các ngăn tần số khác nhau.

Trong phương án khác, bộ tạo trường âm thanh cũng được tạo cấu hình để tạo thông tin khuếch tán cho nhiều ngăn tần số của khung thời gian của trường âm thanh, trong đó bộ tạo siêu dữ liệu được tạo cấu hình để chỉ tạo thông tin khoảng cách cho ngăn tần số khác với giá trị được định trước, hoặc khác với vô hạn hoặc để tạo giá trị khoảng cách cho ngăn tần số, khi giá trị khuếch tán thấp hơn ngưỡng được định trước hoặc thích ứng. Do đó, đối với các ngăn thời gian/tần số mà có sự khuếch tán cao, mọi giá trị khoảng cách không được tạo hoặc giá trị khoảng cách được định trước được tạo mà được làm sáng tỏ bởi bộ giải mã theo cách nhất định. Do đó, chắc chắn rằng đối với các ngăn thời gian/tần số có sự khuếch tán cao, mọi kết xuất liên quan đến khoảng cách không được thực hiện, vì sự khuếch tán cao biểu thị rằng, đối với ngăn thời gian/tần số như vậy, âm thanh không đến từ nguồn được định vị trí nhất định mà đến từ hướng bất kỳ và, do đó, bất kể trường âm thanh được cảm nhận ở vị trí tham chiếu ban đầu hay vị trí tham chiếu khác hoặc mới.

Đối với bộ tính toán trường âm thanh, các phương án ưu tiên bao gồm giao diện tịnh tiến để cung cấp thông tin tịnh tiến hoặc thông tin xoay biểu thị sự xoay của người nghe dự định đến trường âm thanh biến đổi, bộ cung cấp siêu dữ liệu để cung cấp siêu dữ liệu cho bộ tính toán trường âm thanh và bộ cung cấp trường âm thanh để cung cấp mô tả trường âm thanh cho bộ tính toán trường âm thanh và, ngoài ra, giao diện đầu ra để đưa ra trường âm thanh biến đổi bao gồm mô tả trường âm thanh biến đổi và siêu dữ liệu biến đổi, siêu dữ liệu biến đổi được suy ra từ siêu dữ liệu bằng cách sử dụng thông tin tịnh tiến, hoặc giao diện đầu ra xuất ra nhiều kênh loa phóng thanh, mỗi kênh loa phóng thanh liên quan đến vị trí loa được định trước, hoặc giao diện đầu ra xuất ra sự biểu diễn lập thể của trường âm thanh biến đổi.

Trong một phương án, mô tả trường âm thanh bao gồm nhiều thành phần trường âm thanh. Nhiều thành phần trường âm thanh bao gồm thành phần đẳng hướng và ít nhất một thành phần định hướng. Ví dụ, mô tả trường âm thanh là mô tả trường âm thanh Ambisonic cấp một có thành phần đẳng hướng và ba thành phần định hướng X, Y, Z hoặc trường âm thanh như vậy là mô tả Ambisonic cấp cao bao gồm thành phần đẳng hướng, ba thành phần định hướng đối với các hướng X, Y, và Z và, ngoài ra, các thành phần định hướng tiếp theo mà có liên quan đến các hướng khác ngoài các hướng X, Y, Z.

Trong phương án, thiết bị bao gồm bộ phân tích để phân tích các thành phần trường âm thanh để suy ra, đối với các ngăn thời gian hoặc tần số khác nhau, thông tin hướng đến. Thiết bị ngoài ra có bộ biến đổi tịnh tiến để tính toán thông tin DoA biến đổi trên ngăn tần số hoặc thời gian bằng cách sử dụng thông tin DoA và siêu dữ liệu, ở đó siêu dữ liệu liên quan đến bản đồ chiều sâu kết hợp khoảng cách với thông tin DoA cho ngăn thời gian hoặc tần số.

Hơn nữa, bộ tính toán trường âm thanh có bộ bù khoảng cách để tính toán trường âm thanh biến đổi bằng cách sử dụng thông tin bù khoảng cách tùy thuộc từ khoảng cách được cung cấp từ bản đồ chiều sâu cho ngăn tần số hoặc thời gian, và từ khoảng cách mới kết hợp với ngăn thời gian hoặc tần số, khoảng cách mới có liên quan đến thông tin DoA biến đổi.

Trong phương án, bộ tính toán trường âm thanh tính toán vectơ thứ nhất chỉ từ vị trí tham chiếu tới nguồn âm thanh thu được bằng cách phân tích trường âm thanh. Hơn nữa, bộ tính toán trường âm thanh tính toán vectơ thứ hai chỉ từ vị trí tham chiếu khác

đến nguồn âm thanh và việc tính toán này được thực hiện bằng cách sử dụng vectơ thứ nhất và thông tin tịnh tiến, mà thông tin tịnh tiến định rõ vectơ tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác. Và, sau đó, khoảng cách từ vị trí tham chiếu khác tới nguồn âm thanh được tính toán bằng cách sử dụng vectơ thứ hai.

Hơn nữa, bộ tính toán trường âm thanh được tạo cấu hình để nhận, ngoài thông tin tịnh tiến, thông tin xoay biểu thị việc xoay của đầu người nghe theo một trong ba hướng xoay được đưa ra bởi dọc, ngang, xoay. Bộ tính toán trường âm thanh sau đó được tạo cấu hình để thực hiện biến đổi xoay để xoay dữ liệu hướng đến được đổi cho trường âm thanh bằng cách sử dụng thông tin xoay, trong đó dữ liệu hướng đến biến đổi được suy ra từ dữ liệu hướng đến thu được bởi sự phân tích âm thanh của mô tả trường âm thanh và thông tin tịnh tiến.

Trong phương án, bộ tính trường âm thanh được tạo cấu hình để xác định các tín hiệu nguồn từ mô tả trường âm thanh và các hướng của các tín hiệu nguồn liên quan đến vị trí tham chiếu bởi sự phân tích âm thanh.

Sau đó, các hướng mới của các nguồn âm thanh được tính toán mà liên quan đến vị trí tham chiếu khác và điều này được thực hiện bằng cách sử dụng siêu dữ liệu, và sau đó thông tin khoảng cách của các nguồn âm thanh liên quan đến vị trí tham chiếu khác được tính toán và, sau đó, trường âm thanh biến đổi được tổng hợp bằng cách sử dụng thông tin khoảng cách và các hướng mới của các nguồn âm thanh.

Trong phương án, phép tổng hợp trường âm thanh được thực hiện bằng cách quét các tín hiệu nguồn âm thanh sang hướng được cho bởi thông tin hướng mới liên quan đến cài đặt phát lại, và việc định tỷ lệ của các tín hiệu nguồn âm thanh được thực hiện bằng cách sử dụng thông tin khoảng cách trước khi thực hiện thao tác quét hoặc sau khi thực hiện thao tác quét.

Trong phương án khác, phần khuếch tán của tín hiệu nguồn âm thanh được bổ sung vào phần hướng của tín hiệu nguồn âm thanh, phần hướng được biến đổi bởi thông tin khoảng cách trước khi được bổ sung vào phần khuếch tán.

Cụ thể, tốt hơn là thực hiện tổng hợp nguồn âm thanh trong sự biểu diễn phổ ở đó thông tin hướng mới được tính toán cho mỗi ngăn tần số, ở đó thông tin khoảng cách được tính toán cho mỗi ngăn tần số, và ở đó phép tổng hợp hướng cho mỗi ngăn tần số sử dụng tín hiệu âm thanh cho ngăn tần số được thực hiện bằng cách sử dụng tín hiệu âm thanh cho ngăn tần số, độ khuếch đại quét cho ngăn tần số được suy ra từ

thông tin hướng mới và hệ số tỷ lệ cho ngăn tần số thu nhận từ thông tin khoảng cách cho ngăn tần số được thực hiện.

Hơn nữa, phép tổng hợp khuếch tán được thực hiện bằng cách sử dụng tín hiệu âm thanh khuếch tán được suy ra từ tín hiệu âm thanh từ ngăn tần số và bằng cách sử dụng tham số khuếch tán được suy ra bởi phép phân tích tín hiệu cho ngăn tần số và, sau đó, tín hiệu hướng và tín hiệu khuếch tán được kết hợp để thu được tín hiệu âm thanh được tổng hợp cho ngăn thời gian hoặc tần số, sau đó, phép chuyển đổi tần số-thời gian được thực hiện bằng cách sử dụng các tín hiệu âm thanh cho các ngăn thời gian/tần số để thu được tín hiệu âm thanh tổng hợp miền thời gian làm trường âm thanh biến đổi.

Do đó, nói chung, bộ tính trường âm thanh được tạo cấu hình để tổng hợp, đối với mỗi nguồn âm thanh, trường âm thanh liên quan đến vị trí tham chiếu khác bằng cách, ví dụ, xử lý, đối với mỗi nguồn, tín hiệu nguồn bằng cách sử dụng hướng mới cho tín hiệu nguồn để thu được mô tả trường âm thanh của tín hiệu nguồn liên quan đến vị trí tham chiếu khác/mới. Hơn nữa, tín hiệu nguồn được biến đổi trước khi xử lý tín hiệu nguồn hoặc sau khi xử lý tín hiệu nguồn bằng cách sử dụng thông tin hướng. Và, cuối cùng, các mô tả trường âm thanh đối với các nguồn được bổ sung cùng nhau để thu được trường âm thanh biến đổi liên quan đến vị trí tham chiếu khác.

Trong phương án nữa, bộ tính toán trường âm thanh thực hiện, ngoài phân tích DirAC hoặc phân tích nguồn âm thanh khác bất kỳ, thuật toán tách nguồn. Thuật toán tách nguồn thu được, lúc kết thúc, các tín hiệu nguồn âm thanh, ví dụ, trong miền thời gian hoặc trong miền tần số. Tín hiệu khuếch tán sau đó được tính toán bằng cách trừ các tín hiệu nguồn âm thanh khỏi trường âm thanh gốc sao cho trường âm thanh gốc được phân tích thành tín hiệu khuếch tán và một số tín hiệu nguồn âm thanh, trong đó từng tín hiệu nguồn âm thanh đã kết hợp với nó hướng nhất định.

Mô tả vắn tắt các hình vẽ

Các phương án được ưu tiên của sáng chế được thảo luận sau đây về các hình vẽ kèm theo, trong đó:

Fig.1a là phương án được ưu tiên của thiết bị tạo mô tả trường âm thanh tăng cường;

Fig.1b là sự minh họa giải thích vấn đề mẫu trong sáng chế;

Fig.2 là phương án triển khai ưu tiên của thiết bị tạo mô tả trường âm thanh tăng cường;

Fig.3a minh họa mô tả trường âm thanh tăng cường bao gồm dữ liệu âm thanh, và thông tin phụ cho dữ liệu âm thanh;

Fig.3b minh họa sự minh họa thêm về trường âm thanh tăng cường bao gồm dữ liệu âm thanh và siêu dữ liệu liên quan đến thông tin không gian như bản đồ chiều sâu;

Fig.3c minh họa định dạng khác để kết hợp siêu dữ liệu với mô tả trường âm thanh

Fig.3d minh họa định dạng nữa để kết hợp siêu dữ liệu với mô tả trường âm thanh

Fig.4a minh họa phương án triển khai của thiết bị tạo mô tả trường âm thanh biến đổi;

Fig.4b minh họa phương án triển khai nữa của thiết bị để tạo mô tả trường âm thanh biến đổi;

Fig.4c minh họa phương án nữa của thiết bị để tạo mô tả trường âm thanh biến đổi;

Fig.4d minh họa thiết bị làm mẫu để tạo mô tả trường âm thanh từ tín hiệu âm thanh như tín hiệu đơn và dữ liệu hướng đến;

Fig.5 minh họa phép tái tạo sáu DoF của âm thanh không gian;

Fig.6 minh họa phương án ưu tiên của phép tái tạo DoF thứ sáu bao gồm thiết bị tạo mô tả trường âm thanh tăng cường và thiết bị tạo mô tả trường âm thanh biến đổi;

Fig.7 minh họa sự biểu diễn cảnh thực tế ảo;

Fig.8 minh họa bảng MUSHRA trong thực tế ảo;

Fig.9 minh họa các đánh giá MUSHRA như các biểu đồ hộp;

Fig.10 minh họa các đường dẫn tín hiệu để hiển thị tham chiếu và DirAC;

Fig.11a minh họa phương án triển khai ưu tiên của thiết bị để tạo mô tả trường âm thanh biến đổi;

Fig.11b minh họa phương án triển khai ưu tiên để tính toán DoA mới và khoảng cách mới của nguồn âm thanh đối với vị trí tham chiếu mới/khác;

Fig.12 minh họa phương án triển khai ưu tiên nữa của thiết bị để tạo mô tả trường âm thanh biến đổi;

Fig.13a minh họa phương án triển khai phân tích DirAC của tình trạng kỹ thuật đã biết; và

Fig.13b minh họa phương án triển khai tổng hợp DirAC của tình trạng kỹ thuật đã biết.

Mô tả chi tiết sáng chế

Để kích hoạt các ứng dụng 6DoF cho các sự biểu diễn Ambisonic/DirAC được đề cập, cần phải mở rộng các sự biểu diễn này theo cách mà cung cấp thông tin còn thiếu để xử lý tịnh tiến. Cần lưu ý rằng sự mở rộng này có thể, ví dụ, 1) bổ sung khoảng cách hoặc các vị trí của các đối tượng vào sự biểu diễn cảnh hiện có, và/hoặc 2) bổ sung thông tin mà có thể tạo điều kiện thuận lợi cho quá trình tách các đối tượng riêng lẻ.

Ngoài ra, mục tiêu của các phương án là bảo toàn/sử dụng lại cấu trúc của các hệ thống Ambisonic (không theo tham số hoặc theo tham số) hiện có để cung cấp khả năng tương thích ngược với các sự biểu diễn/hệ thống này theo nghĩa

- các sự biểu diễn được mở rộng có thể được chuyển đổi thành các sự biểu diễn không mở rộng hiện có (ví dụ, để kết xuất), và
- cho phép sử dụng lại các triển khai phần mềm và phần cứng hiện có khi làm việc với sự biểu diễn được mở rộng.

Sau đây, một số phương pháp được mô tả, cụ thể là một phương pháp hạn chế (nhưng rất đơn giản) và ba định dạng Ambisonic mở rộng khác nhau để kích hoạt 6DoF.

Như được mô tả trong đoạn tình trạng kỹ thuật đã biết, DirAC truyền thống mang thông tin phụ tham số mà đặc trưng cho hướng và độ khuếch tán cho từng ngăn thời gian tần số (TF - Time Frequency). Sự mở rộng định dạng DirAC hiện có cung cấp thêm thông tin chiều sâu cho từng hoặc cho một số nhưng không phải tất cả các ngăn TF. Tương tự với thông tin hướng, sự liên quan của thông tin chiều sâu phụ thuộc vào độ khuếch tán thực tế. Độ khuếch tán cao có nghĩa là cả hướng và chiều sâu không liên quan (và có thể được bỏ qua trong thực tế đối với các giá trị độ khuếch tán rất cao).

Cần lưu ý rằng DirAC mở rộng chiều sâu không cung cấp giải pháp 6DoF đầy đủ vì có thể chỉ mang thông tin hướng và chiều sâu cho từng đối tượng trên mỗi ngăn TF.

Cần lưu ý rằng thông tin chiều sâu có thể được ước lượng từ các tín hiệu âm thanh hoặc từ các tín hiệu video (ví dụ, bản đồ chiều sâu thường được sử dụng trong ảnh/video lập thể (3D) hoặc kỹ thuật trường ánh sáng) hoặc có thể được bổ sung thủ công hoặc tự động một cách cụ thể khi trường âm thanh được tạo ra bởi phép tổng hợp âm thanh với các nguồn âm thanh được định vị trí.

Fig.1a minh họa thiết bị tạo mô tả trường âm thanh tăng cường bao gồm bộ tạo (mô tả) trường âm thanh 100 để tạo ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một vị trí tham chiếu. Hơn nữa, thiết bị bao gồm bộ tạo siêu dữ liệu 110 để tạo siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh. Siêu dữ liệu nhận, dưới dạng đầu vào, trường âm thanh hoặc ngoài ra hoặc bổ sung, thông tin riêng về các nguồn âm thanh.

Cả đầu ra của bộ tạo mô tả trường âm thanh 100 và bộ tạo siêu dữ liệu 110 cấu thành mô tả trường âm thanh tăng cường. Trong phương án, cả đầu ra bộ tạo mô tả trường âm thanh 100 và bộ tạo siêu dữ liệu 110 có thể được kết hợp trong bộ kết hợp 120 hoặc giao diện đầu ra 120 để thu mô tả trường âm thanh tăng cường mà bao gồm siêu dữ liệu không gian hoặc thông tin không gian của trường âm thanh như được tạo bởi bộ tạo siêu dữ liệu 110.

Fig.1b minh họa trường hợp mà được giải quyết bởi sáng chế. Ví dụ, vị trí A là ít nhất một vị trí tham chiếu và trường âm thanh được tạo bởi nguồn A và nguồn B và thực tế nhất định hoặc, ví dụ, micrô ảo được đặt tại vị trí A phát hiện âm thanh từ nguồn A và nguồn B. Âm thanh là sự chồng chất của âm thanh đến từ các nguồn âm thanh phát ra. Điều này biểu diễn mô tả trường âm thanh như được tạo bởi bộ tạo mô tả trường âm thanh.

Ngoài ra, bộ tạo siêu dữ liệu sẽ, bằng một số phương án triển khai nhất định có được thông tin không gian đối với nguồn A và thông tin không gian khác đối với nguồn B chẳng hạn như các khoảng cách của các nguồn này tới vị trí tham chiếu chẳng hạn như vị trí A.

Đương nhiên, vị trí tham chiếu có thể, thay vào đó, là vị trí B. Sau đó, micrô thực hoặc ảo sẽ được đặt ở vị trí B và mô tả trường âm thanh sẽ là trường âm thanh, ví dụ, được biểu diễn bằng các thành phần Ambisonic cấp một hoặc các thành phần Ambisonic cấp cao hoặc bất kỳ thành phần âm thanh nào khác có khả năng mô tả trường âm thanh đối với ít nhất một vị trí tham chiếu như vị trí B.

Sau đó, bộ tạo siêu dữ liệu có thể tạo ra, như thông tin về các nguồn âm thanh, khoảng cách của nguồn âm thanh A đến vị trí B hoặc khoảng cách của nguồn B đến vị trí B. Tất nhiên, thông tin thay thế về các nguồn âm thanh có thể là vị trí tuyệt đối hay tương đối so với vị trí tham chiếu. Vị trí tham chiếu có thể là nguồn gốc của hệ tọa độ chung hoặc có thể được đặt trong mối quan hệ xác định với nguồn gốc của hệ tọa độ chung.

Siêu dữ liệu khác có thể là vị trí tuyệt đối của một nguồn âm thanh và vị trí tương đối của nguồn âm thanh khác đối với nguồn âm thanh thứ nhất, v.v..

Fig.2 minh họa phương án triển khai ưu tiên của bộ tạo mô tả trường âm thanh. Bộ tạo mô tả trường âm thanh có thể, ví dụ, bao gồm micrô thực hoặc ảo 200 mà tạo ra, từ trường âm thanh đầu vào, các thành phần trường âm thanh như sự biểu diễn Ambisonic cấp một được minh họa trên Fig.2 như thành phần đẳng hướng w , và ba thành phần hướng x , y , z .

Dựa trên mô tả trường âm thanh này, bộ phân tích trường âm thanh 210 ngoài ra có thể bao gồm bộ trộn giảm, sẽ tạo mô tả trường âm thanh tham số bao gồm các tham số trộn giảm và bổ sung đơn hoặc lập thể như các tham số DoA hướng đến, ví dụ trên các khung thời gian hoặc các ngăn tần số, thông thường, các ngăn thời gian/tần số và, ngoài ra, thông tin khuếch tán cho số lượng các ngăn thời gian/tần số tương tự hoặc nhỏ hơn.

Hơn nữa, bộ tạo siêu dữ liệu 110 sẽ, ví dụ, được thực hiện như bộ tạo bản đồ chiều sâu mà tạo ra bản đồ chiều sâu mà kết hợp, với từng hướng đến hoặc thông tin DoA, khoảng cách nhất định theo các thuật ngữ tuyệt đối hoặc tương đối. Hơn nữa, bộ tạo siêu dữ liệu 110, trong phương án ưu tiên, cũng được điều khiển bởi tham số khuếch tán cho ngăn thời gian/tần số. Trong phương án triển khai này, bộ tạo siêu dữ liệu 110 sẽ được thực hiện để không tạo ra thông tin khoảng cách bất kỳ cho ngăn thời gian và/hoặc tần số mà có giá trị khuếch tán cao hơn ngưỡng được xác định trước hoặc ngưỡng thích ứng nhất định. Điều này là do thực tế là, khi ngăn thời gian hoặc tần số nhất định thể hiện độ khuếch tán cao thì có thể rút ra kết luận là trong ngăn thời gian hoặc tần số này, không tồn tại bất kỳ nguồn âm thanh được định vị trí nào, nhưng chỉ tồn tại âm thanh khuếch tán từ tất cả các hướng. Do đó, đối với ngăn thời gian hoặc tần số, bộ tạo siêu dữ liệu sẽ tạo ra, trong bản đồ chiều sâu, không phải giá trị nào như được biểu thị bởi "N.A" trên Fig.2, hoặc ngoài ra, bộ tạo siêu dữ liệu sẽ đưa vào giá trị

khoảng cách có giá trị được định trước như mã cho giá trị cao, giá trị vô cực hoặc giá trị khác mà sẽ được xác nhận bởi bộ giải mã như giá trị biểu thị nguồn âm thanh không được định vị trí cho ngắn thời gian/tần số.

Bản đồ chiều sâu và mô tả trường âm thanh được tạo ra bởi bộ phân tích trường âm thanh 210 tương ứng với sự biểu diễn trộn giảm đơn/lập thể cùng với các tham số không gian mà liên quan đến vị trí tham chiếu sau đó được tổ hợp trong bộ tổ hợp 120 để tạo mô tả trường âm thanh tăng cường.

Fig.3a minh họa ví dụ về dòng bit hoặc dòng dữ liệu chung bao gồm mô tả trường âm thanh tăng cường. Dòng dữ liệu sẽ bao gồm các khung thời gian i , $i+1$ và vvvv như được biểu thị bởi các số tham chiếu 302, 304, 306 và thông tin phụ liên quan cho khung thời gian tương ứng như được minh họa bởi các khối 310, 312. Trong phương án, thông tin phụ sẽ bao gồm thông tin hướng đến trên mỗi ngắn thời gian/tần số như được biểu thị ở 314, và, tùy chọn, giá trị khuếch tán trên mỗi ngắn thời gian/tần số như được minh họa bởi mục 316 và, tùy chọn, đối với khung i , bản đồ chiều sâu 320. Bản đồ chiều sâu cũng được minh họa trên Fig.3a được biểu thị bởi mục 322 và 324. Mục 322 minh họa thông tin DoA khác mở rộng, ví dụ, giữa 0° và 350° theo hướng góc phương vị với độ tăng ví dụ bằng 10° . Theo đó, mục 324 minh họa giá trị khoảng cách tương ứng kết hợp với giá trị DoA nhất định. Bản đồ chiều sâu này sẽ được tạo ra đối với từng độ tăng góc nâng sao cho, cuối cùng, bản đồ chiều sâu sẽ kết hợp, với từng tổ hợp của góc phương vị và góc nâng, tức là, với từng thông tin DoA, giá trị khoảng cách nhất định.

Đương nhiên, các cách hiệu quả hơn có khả năng khác để tạo và truyền bản đồ chiều sâu có thể được thực hiện trong đó, thông thường, đối với mỗi giá trị DoA xảy ra cho ngắn tần số trong khung thời gian nhất định mà có giá trị khuếch tán thấp hơn giá trị ngưỡng nhất định, khoảng cách sẽ có mặt.

Fig.3b minh họa phương án triển khai khác mà bộ phân tích trường âm thanh 210 trên Fig.2 không tạo ra thực tế sự trộn giảm, nhưng tạo ra sự biểu diễn đầy đủ theo định dạng B, định dạng A hoặc định dạng khác bất kỳ như sự biểu diễn cấp cao hơn cho khoảng thời gian nhất định. Sau đó, bản đồ chiều sâu 320a và bản đồ chiều sâu khác 320b sẽ được kết hợp với sự biểu diễn tín hiệu âm thanh biểu thị tại 326 trên Fig.3b. Khi, ví dụ, khoảng thời gian, mà định dạng B hoặc sự biểu diễn cấp cao hơn được kết hợp với một số khoảng thời gian riêng lẻ, và khi thay đổi vị trí của các nguồn

xảy ra trong các khoảng thời gian này, sau đó bản đồ chiều sâu 320a cho khoảng thời gian thứ nhất i và bản đồ chiều sâu khác cho khoảng thời gian $i + 1$ được biểu thị tại mục 320b sẽ được kết hợp với sự biểu diễn tín hiệu âm thanh 326. Như đã được nêu, sự biểu diễn tín hiệu âm thanh trên Fig.3b khác sự biểu diễn trên Fig.3a, vì Fig.3a chỉ có sự biểu diễn của trộn giảm hoặc thành phần w và, ngoài ra, sự biểu diễn trong các khung thời gian khác nhau, trong đó Fig.3b có sự biểu diễn tín hiệu âm thanh khác với sự biểu diễn thành phần hướng đầy đủ với ba hoặc nhiều hơn ba thành phần và không được tách thành các khung thời gian riêng rẽ nhưng được kết hợp với các bản đồ chiều sâu cho các lần tăng thời gian nhỏ hơn so với các khoảng thời gian, mà mục 326 được cho. Dù được minh họa trên Fig.3b như định dạng theo tuần tự, lưu ý là định dạng song song hoặc kết hợp giữa song song và theo dãy hoặc định dạng khác nhất định như định dạng vùng chứa MP4 có thể được sử dụng.

Fig.3c và Fig.3d minh họa các định dạng khác để kết hợp siêu dữ liệu với mô tả trường âm thanh dưới dạng định dạng B hoặc sự biểu diễn cấp cao hơn. Chỉ số i , $i+1$ viết tắt cho thời gian và chỉ số (1), (2), (Ni) viết tắt cho hướng.

Fig.4a minh họa phương án triển khai ưu tiên của thiết bị tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh. Đặc biệt, thiết bị bao gồm bộ tính toán trường âm thanh 420 mà tạo trường âm thanh biến đổi sử dụng siêu dữ liệu, mô tả trường âm thanh và thông tin tịnh tiến chỉ ra sự tịnh tiến từ vị trí tham chiếu tới vị trí tham chiếu khác.

Khi, ví dụ, trường âm thanh được cho đối với vị trí A trên Fig.1B, sau đó vị trí tham chiếu khác có thể là vị trí B và thông tin tịnh tiến sẽ, ví dụ, là vectơ biểu thị sự tịnh tiến của vị trí A sang vị trí B. Bộ tính toán trường âm thanh 420 sau đó sẽ tính toán trường âm thanh biến đổi như khi nó sẽ được cảm nhận bởi người nghe được định vị trí ở vị trí B, và đối với phép tính toán này, bộ tính toán trường âm thanh có sự mô tả trường âm thanh liên quan đến vị trí A và thông tin tịnh tiến và, ngoài ra, siêu dữ liệu liên quan đến vị trí không gian của nguồn A và nguồn B.

Trong phương án, bộ tính toán trường âm thanh 420 được kết nối với giao diện đầu vào 400 để nhận mô tả trường âm thanh tăng cường như, ví dụ, được thảo luận liên quan đến Fig.1a hoặc Fig.2 và giao diện đầu vào 400 sau đó phân tách mô tả trường âm thanh một mặt, nghĩa là, những gì đã được tạo ra bởi khối 100 trên Fig.1a

hoặc khối 210 trên Fig.2. Hơn nữa, giao diện đầu vào 400 phân tách siêu dữ liệu từ mô tả trường âm thanh tăng cường, nghĩa là, mục 310, 312 trên Fig.3a hoặc mục 320a, 320b trên Fig.3b.

Hơn nữa, giao diện tĩnh tiến 410 thu thông tin tĩnh tiến và/hoặc thông tin xoay bổ sung hoặc riêng biệt từ người nghe. Việc thực hiện của giao diện tĩnh tiến 410 có thể là bộ phận theo dõi đầu không chỉ theo dõi sự xoay của đầu trong môi trường thực tế ảo, mà còn là sự tĩnh tiến của đầu từ một vị trí, nghĩa là, vị trí A trên Fig.1b tới vị trí khác, nghĩa là, vị trí B trên Fig.1b.

Fig.4b minh họa phương án triển khai khác tương tự như trên Fig.1a, nhưng không liên quan đến kịch bản bộ mã hóa/bộ giải mã, nhưng liên quan đến kịch bản chung khi việc cung cấp siêu dữ liệu được chỉ định bởi bộ cung cấp siêu dữ liệu 402, việc cung cấp siêu dữ liệu được chỉ định bởi bộ cung cấp 404 được thực hiện mà không có giao diện đầu vào nhất định phân tách mô tả trường âm thanh được mã hóa hoặc tăng cường, nhưng tất cả đều được thực hiện, ví dụ, trong kịch bản thực tế hiện có, ví dụ, trong ứng dụng thực tế ảo. Tuy nhiên, sáng chế không bị giới hạn bởi các ứng dụng thực tế ảo, nhưng cũng có thể được thực hiện trong ứng dụng thực tế ảo khác bất kỳ, trong đó việc xử lý âm thanh không gian của các trường âm thanh mà liên quan đến vị trí tham chiếu là hữu ích để biến đổi trường âm thanh liên quan đến vị trí tham chiếu thứ nhất tới trường âm thanh khác liên quan đến vị trí tham chiếu thứ hai khác.

Bộ tính toán trường âm thanh 420 sau đó tạo mô tả trường âm thanh biến đổi hoặc, thay vào đó, tạo sự biểu diễn loa phóng thanh (ảo) hoặc tạo sự biểu diễn lập thể chẳng hạn như sự biểu diễn hai kênh để tái tạo tai nghe chụp đầu. Do đó, bộ tính toán trường âm thanh 420 có thể tạo, như trường âm thanh biến đổi, mô tả trường âm thanh biến đổi, về cơ bản giống như mô tả trường âm thanh gốc, nhưng bây giờ với vị trí tham chiếu mới. Trong phương án thay thế, sự biểu diễn loa phóng thanh thực hoặc ảo có thể được tạo cho cài đặt loa được phóng thanh xác định trước chẳng hạn như sơ đồ 5.1 hoặc cài đặt loa phóng thanh có nhiều loa phóng thanh hơn và, đặc biệt, có bố trí các loa phóng thanh ba chiều thay vì chỉ bố trí hai chiều, nghĩa là, sự bố trí loa phóng thanh có các loa phóng thanh được nâng lên đối với vị trí người dùng. Các ứng dụng khác mà đặc biệt hữu ích cho các ứng dụng thực tế ảo là các ứng dụng để tái tạo lập thể, nghĩa là, cho tai nghe chụp đầu mà có thể được áp dụng cho đầu của người dùng thực tế ảo.

Fig.4c minh họa phương án triển khai ưu tiên của sáng chế trong ngữ cảnh phân tích/tổng hợp DirAC. Để đạt được mục đích này, mô tả trường âm thanh thường được nhận bởi các micrô thực hoặc được tạo ra ban đầu bởi các micrô ảo hoặc được tổng hợp sớm hơn và ngay sau đó được sử dụng trong ứng dụng thực tế ảo hoặc trong ứng dụng khác bất kỳ được nhập vào bộ chuyển đổi thời gian thành tần số 421.

Sau đó, bộ phân tích DirAC 422 được tạo cấu hình để tạo ra, đối với từng ngăn thời gian/tần số, mục dữ liệu hướng đến và mục dữ liệu khuếch tán.

Sử dụng thông tin trường âm thanh không gian như được cho bởi bản đồ chiều sâu, ví dụ, khối 423 thực hiện biến đổi tịnh tiến và, tùy chọn, thông tin định tỷ lệ âm lượng, giá trị hướng đến mới được tính toán. Tốt hơn là, biến đổi xoay 424 cũng được thực hiện, tất nhiên, thông tin theo dõi liên quan đến thông tin tịnh tiến trên một mặt và thông tin xoay trên mặt khác được sử dụng trong các khối 423 đến 424 để tạo ra dữ liệu hướng đến như được nhập vào trong khối bộ tổng hợp DirAC 425. Ngoài ra, sau đó, thông tin định tỷ lệ phụ thuộc vào khoảng cách mới giữa nguồn âm thanh và vị trí tham chiếu mới được biểu thị bởi thông tin theo dõi cũng được tạo ra trong khối 423 và được sử dụng trong bộ tổng hợp DirAC 425 để cuối cùng thực hiện phép tổng hợp DirAC cho từng ngăn thời gian/tần số. Sau đó, trong khối 426, phép chuyển đổi thời gian/tần số được thực hiện, tốt hơn là đối với cài đặt loa phóng thanh ảo được xác định trước nhất định, sau đó, trong khối 427, việc kết xuất lập thể cho sự biểu diễn tai nghe chụp đầu lập thể được thực hiện.

Trong phương án nữa, bộ tổng hợp DirAC cung cấp trực tiếp các tín hiệu lập thể trong miền TF.

Phụ thuộc vào phương án triển khai của bộ phân tích DirAC, và, tất nhiên, phụ thuộc vào phương án triển khai của bộ tổng hợp DirAC 425, trường âm thanh ban đầu ở đầu vào của khối 421 hoặc ở đầu ra của khối 421 có thể được chuyển đến bộ tổng hợp DirAC 425 hoặc, ngoài ra, tín hiệu trộn giảm được tạo ra bởi bộ phân tích DirAC 422 được chuyển đến bộ tổng hợp DirAC.

Ví dụ, Fig.6 được mô tả sau đây minh họa trường hợp, trong đó bộ tổng hợp DirAC chỉ hoạt động trên thành phần trộn giảm như thành phần đẳng hướng hoặc thành phần áp lực, trong khi, trong phương án thay thế khác được minh họa ở Fig.13b, bộ tổng hợp DirAC hoạt động trên toàn bộ dữ liệu trường âm thanh, tức là toàn bộ sự

biểu diễn thành phần có, trong phương án này trên Fig.13b, mô tả trường với thành phần đẳng hướng w và ba thành phần hướng x, y, z .

Fig.4d minh họa phương án triển khai khác để thực hiện phép tổng hợp khác với bộ tổng hợp DirAC. Khi, ví dụ, bộ phân tích trường âm thanh tạo ra, đối với từng tín hiệu nguồn, tín hiệu đơn riêng S và hướng đến ban đầu và khi, phụ thuộc vào thông tin tĩnh tiến, hướng đến mới được tính toán, sau đó bộ tạo tín hiệu Ambisonic 430 của Fig.4d, ví dụ, sẽ được sử dụng để tạo mô tả trường âm thanh cho tín hiệu âm thanh, tức là, tín hiệu đơn S nhưng cho dữ liệu hướng đến (DoA) mới bao gồm góc ngang θ hoặc góc nâng θ và góc phương vị ϕ . Sau đó, thủ tục được thực hiện bởi bộ tính toán trường âm thanh 420 của Fig.4b sẽ tạo ra, ví dụ, sự biểu diễn trường âm thanh Ambisonic cấp một cho từng nguồn âm thanh với hướng đến mới và, sau đó, phép biến đổi trên mỗi nguồn âm thanh có thể được thực hiện sử dụng thừa số định tỷ lệ phụ thuộc vào khoảng cách của trường âm thanh tới vị trí tham chiếu mới và, sau đó, tất cả các trường âm thanh từ các nguồn riêng lẻ có thể được chồng lên nhau để cuối cùng thu được trường âm thanh biến đổi, một lần nữa, ví dụ, trong sự biểu diễn Ambisonic liên quan đến vị trí tham chiếu mới nhất định.

Khi giải thích rằng từng ngăn thời gian/tần số được xử lý bởi bộ phân tích DirAC 422 biểu diễn nguồn âm thanh nhất định (bảng thông được giới hạn), sau đó bộ tạo tín hiệu Ambisonic 430 có thể được sử dụng, thay vì bộ tổng hợp DirAC 425, để tạo, đối với từng ngăn thời gian/tần số, sự biểu diễn Ambisonic đầy đủ sử dụng tín hiệu trộn giảm hoặc tín hiệu áp lực hoặc thành phần đẳng hướng cho ngăn thời gian/tần số này như “tín hiệu đơn” của Fig.4d. Sau đó, phép chuyển đổi tần số-thời gian riêng lẻ trong bộ chuyển đổi tần số-thời gian 426 cho từng thành phần trong số các thành phần W, X, Y, Z sẽ dẫn đến mô tả trường âm thanh khác với những gì được minh họa trên Fig.4c.

Các phương án nữa được nêu sau đây. Mục đích là thu được tín hiệu lập thể ảo tại vị trí người nghe được cho ở vị trí ghi ban đầu và thông tin về các khoảng cách của các nguồn âm thanh từ vị trí ghi. Các nguồn vật lý được giả định để tách riêng được bởi góc của chúng về vị trí ghi.

Cảnh được ghi từ điểm nhìn (point of view - PoV) của micrô, mà vị trí được sử dụng như nguồn gốc của hệ tọa độ tham chiếu. Cảnh đã được tái tạo từ PoV của người

nghe, mà được theo dõi trong 6DoF, tham chiếu Fig.5. Nguồn âm thanh đơn lẻ được thể hiện ở đây để minh họa, mỗi liên hệ giữ cho từng ngăn thời gian-tần số.

Fig.5 minh họa phép tái tạo 6DoF của âm thanh không gian. Nguồn âm thanh được ghi lại bởi micrô với DoA $\mathbf{r}_r$ trong khoảng cách d_r liên quan đến vị trí và định hướng của các micrô (đường và hình cung màu đen). Nó phải được tái tạo tương ứng với người nghe di chuyển với DoA $\mathbf{r}_l$ và khoảng cách d_l (nét đứt). Điều này phải xem xét việc tịnh tiến $\mathbf{l}$ và xoay $\mathbf{o}$ của người nghe (chấm).

Nguồn âm thanh tại tọa độ $\mathbf{d}_r \in \mathbb{R}^3$ được ghi từ hướng đến (DoA) được biểu thị bằng vectơ đơn vị $\mathbf{r}_r = \mathbf{d}_r / \|\mathbf{d}_r\|$. DoA này có thể được ước lượng từ phân tích bản ghi. Nó đến từ khoảng cách $d_r = \|\mathbf{d}_r\|$. Giả định rằng thông tin này có thể được ước lượng tự động, ví dụ, sử dụng máy quay thời gian bay (time-of-flight), để thu được thông tin khoảng cách dưới dạng bản đồ chiều sâu $m(\mathbf{r})$ ánh xạ từng hướng $\mathbf{r}$ từ vị trí ghi đến khoảng cách của nguồn âm thanh gần nhất theo mét.

Người nghe được theo dõi trong 6DoF. Tại thời điểm được đưa ra, anh ta ở vị trí $\mathbf{l} \in \mathbb{R}^3$ so với micrô và có sự xoay $\mathbf{o} \in \mathbb{R}^3$ so với hệ thống tọa độ micrô. Vị trí ghi được chọn làm gốc của hệ tọa độ để đơn giản hóa ký hiệu.

Do đó âm thanh phải được tái tạo với khoảng cách d_l khác dẫn đến âm lượng thay đổi, và DoA $\mathbf{r}_l$ khác mà là kết quả của cả việc tịnh tiến và xoay sau đó.

Phương pháp để thu tín hiệu ảo từ quan điểm của người nghe bởi các phép biến đổi chuyên dụng dựa trên sự biểu diễn tham số, như được giải thích trong phần dưới đây, được đưa ra.

Phương pháp được đề xuất dựa trên phương pháp DirAC cơ bản để mã hóa âm thanh không gian tham số, xem [16]. Giả định rằng có một nguồn có hướng chi phối trên mỗi trường hợp thời gian – tần số của phổ được phân tích và chúng có thể được xử lý độc lập. Bản ghi được biến đổi thành sự biểu diễn thời gian – tần số bằng cách sử dụng biến đổi Fourier thời gian ngắn (STFT – short time Fourier transform). Chỉ số khung thời gian được ký hiệu là n và chỉ số tần số được ký hiệu là k . Bản ghi được biến đổi sau đó được phân tích, ước tính các hướng $\mathbf{r}_r(k, n)$ và độ khuếch tán $\psi(k, n)$ cho mỗi ngăn thời gian – tần số của phổ phức $P(k, n)$. Trong quá trình tổng hợp, tín hiệu được chia thành phần hướng và phần khuếch tán. Ở đây, tín hiệu loa phóng thanh được tính toán bằng cách quét phần hướng tùy thuộc vào các vị trí loa và bố

sung phần khuếch tán.

Phương pháp để biến đổi bản ghi tín hiệu FOA theo quan điểm của người nghe trong 6DoF có thể được chia thành năm bước, xem Fig.6.

Fig.6 minh họa phương pháp tái tạo 6DoF. Tín hiệu FOA được ghi trong định dạng B được xử lý bởi bộ mã hóa DirAC mà tính toán các giá trị hướng và độ khuếch tán cho mỗi ngăn thời gian – tần số của phổ phức. Véc tơ hướng sau đó được biến đổi bởi vị trí được theo dõi của người nghe và theo thông tin khoảng cách được cho trong bản đồ khoảng cách. Véc tơ hướng kết quả sau đó được xoay theo việc xoay đầu. Cuối cùng, các tín hiệu cho các kênh loa phóng thanh ảo 8+4 được tổng hợp trong bộ giải mã DirAC. Sau đó chúng được lập thể hóa.

Trong phương án, tín hiệu đầu ra được phân tích trong bộ mã hóa DirAC 422, thông tin khoảng cách được bổ sung từ bản đồ khoảng cách $m(r)$, sau đó việc tịnh tiến và xoay được theo dõi của người nghe được áp dụng trong các biến đổi mới 423 và 424. Bộ giải mã DirAC 425 tổng hợp các tín hiệu cho các loa phóng thanh ảo 8+4, lần lượt được lập thể hóa 427 cho việc phát lại tai nghe chụp đầu. Lưu ý rằng do việc xoay của cảnh âm thanh sau khi tịnh tiến là hoạt động độc lập, nó có thể được áp dụng thay thế trong bộ kết xuất lập thể. Tham số duy nhất được biến đổi cho 6DoF là véc tơ hướng. Theo định nghĩa mô hình, phần khuếch tán được giả định là đẳng hướng và đồng nhất và do đó được giữ không đổi.

Đầu vào của bộ mã hóa DirAC là tín hiệu âm thanh FOA trong sự biểu diễn định dạng B. Nó bao gồm bốn kênh, tức là, áp lực âm thanh đẳng hướng và ba gradient không gian cấp một, theo các giả định nhất định tỷ lệ thuận với vận tốc hạt. Tín hiệu này được mã hóa theo phương pháp tham số, xem [18]. Các tham số được suy ra từ áp lực âm thanh phức $P(k, n)$, mà là tín hiệu đẳng hướng được biến đổi và véc tơ vận tốc hạt phức $U(k, n) = [U_X(k, n), U_Y(k, n), U_Z(k, n)]^T$ tương ứng với các tín hiệu gradient được biến đổi.

Sự biểu diễn DirAC bao gồm tín hiệu $P(k, n)$, độ khuếch tán $\psi(k, n)$ và hướng $r(k, n)$ của sóng âm thanh tại mỗi ngăn thời gian – tần số. Để suy ra sau cùng, đầu tiên, véc tơ cường độ âm thanh hoạt động $I_a(k, n)$ được tính là phần thực (ký hiệu là $\text{Re}(\cdot)$) của sản phẩm của véc tơ áp lực với liên hợp phức (ký hiệu là $(\cdot)^*$) của véc tơ vận tốc [18]:

$$I_a(k, n) = \frac{1}{2} \text{Re}(P(k, n)U^*(k, n)). \quad (1)$$

Độ khuếch tán được ước lượng từ hệ số của biến thể của véc tơ này [18].

$$\psi(k, n) = \sqrt{1 - \frac{\|E\{I_a(k, n)\}\|}{E\{\|I_a(k, n)\|\}}} \quad (2)$$

trong đó E là toán tử mong muốn dọc theo các khung thời gian, được thực hiện như trung bình chuyển động.

Vì được dự định để thao tác âm thanh sử dụng bản đồ khoảng cách dựa trên hướng, biến thể của các phép ước lượng hướng có thể là thấp. Vì các khung thường ngắn nên không phải luôn luôn như vậy. Do đó, trong phương án tùy chọn, trung bình chuyển động được áp dụng để thu được ước lượng hướng nhẵn $\bar{I}_a(k, n)$. DoA của phần hướng của tín hiệu sau đó được tính toán như véc tơ độ dài đơn vị theo hướng đối diện:

$$\mathbf{r}_r(k, n) = - \frac{\bar{I}_a(k, n)}{\|\bar{I}_a(k, n)\|} \quad (3)$$

Khi hướng được mã hóa như véc tơ ba chiều của độ dài đơn vị cho từng ngăn thời gian-tần số, sẽ là đơn giản để tích hợp thông tin khoảng cách. Các véc tơ hướng được nhân với mục nhập bản đồ tương ứng của chúng sao cho độ dài véc tơ biểu diễn khoảng cách của nguồn âm thanh tương ứng $\mathbf{d}_r(k, n)$:

$$\begin{aligned} \mathbf{d}_r(k, n) &= \mathbf{r}_r(k, n)d_r(k, n) \\ &= \mathbf{r}_r(k, n)m(\mathbf{r}_r(k, n)) \end{aligned} \quad (4)$$

trong đó $\mathbf{d}_r(k, n)$ là véc tơ chỉ từ vị trí ghi của micro đến nguồn âm thanh hoạt động ở thời gian n và ngăn tần số k .

Vị trí người nghe được cho bởi hệ thống theo dõi cho khung xử lý hiện thời như $\mathbf{l}(n)$. Với sự biểu diễn véc tơ của các vị trí nguồn, có thể trừ véc tơ vị trí theo dõi $\mathbf{l}(n)$ để mang lại véc tơ hướng được tịnh tiến mới $\mathbf{d}_1(k, n)$ với độ dài $d_1(k, n) = \|\mathbf{d}_1(k, n)\|$, tham chiếu Fig.10. Các khoảng cách từ PoV của người nghe đến các nguồn âm thanh được suy ra, và các DoA được làm thích ứng trong một bước:

$$\mathbf{d}_1(k, n) = \mathbf{d}_r(k, n) - \mathbf{l}(n) \quad (5)$$

Khía cạnh quan trọng của phép tái tạo thực tế là làm suy giảm khoảng cách. Sự suy giảm được giả định là hàm của khoảng cách giữa nguồn âm thanh và người nghe

[19]. Độ dài của các vectơ hướng là để mã hóa sự suy giảm hoặc sự khuếch đại cho việc tái tạo. Khoảng cách đến vị trí ghi được mã hóa trong $\mathbf{d}_r(k, n)$ theo bản đồ khoảng cách, và khoảng cách sẽ được tái tạo được mã hóa trong $\mathbf{d}_l(k, n)$. Nếu chuẩn hóa các vectơ với độ dài đơn vị và sau đó nhân với tỉ số của khoảng cách cũ và mới, có thể thấy rằng độ dài yêu cầu được cho bằng cách chia $\mathbf{d}_l(k, n)$ cho độ dài vectơ ban đầu:

$$\mathbf{d}_v(k, n) = \frac{\mathbf{d}_l(k, n)}{\|\mathbf{d}_l(k, n)\|} \frac{\|\mathbf{d}_r(k, n)\|}{\|\mathbf{d}_r(k, n)\|} = \frac{\mathbf{d}_l(k, n)}{\|\mathbf{d}_r(k, n)\|} \quad (6)$$

Các thay đổi đối với định hướng của người nghe được áp dụng trong bước tiếp theo. Định hướng được đưa ra bởi việc theo dõi có thể được viết như vectơ gồm chiều dọc, ngang, và xoay $\mathbf{o}(n) = [o_x(n), o_z(n), o_y(n)]^T$ so với vị trí ghi làm gốc. Hướng nguồn được xoay theo định hướng người nghe, mà được thực hiện bằng cách sử dụng ma trận xoay 2D:

$$\mathbf{d}_p(k, n) = \mathbf{R}_Y(o_y(n))\mathbf{R}_Z(o_z(n))\mathbf{R}_X(o_x(n))\mathbf{d}_v(k, n) \quad (7)$$

DoA thu được cho người nghe sau đó được cho bởi vectơ được chuẩn hóa với độ dài đơn vị:

$$\mathbf{r}_p(k, n) = \frac{\mathbf{d}_p(k, n)}{\|\mathbf{d}_p(k, n)\|} \quad (8)$$

Vectơ hướng được biến đổi, độ khuếch tán, và phổ phức hợp được sử dụng cho các tín hiệu tổng hợp cho cài đặt loa phóng thanh ảo 8+4 được phân bố đồng đều. Tám loa ảo được định vị trí theo các bước góc phương vị 45° trên mặt phẳng người nghe (góc nâng 0°), và bốn loa ảo ở tạo hình chéo 90° trên góc nâng 45° . Phép tổng hợp được chia thành phần hướng và khuếch tán cho từng kênh loa phóng thanh $1 \leq i \leq I$, trong đó $I = 12$ là số lượng loa phóng thanh [16]:

$$Y_i(k, n) = Y_{i,S}(k, n) + Y_{i,D}(k, n) \quad (9)$$

Đối với phần hướng, việc quét biên độ mép mờ dần (edge fading amplitude panning - EFAP) được áp dụng để tái tạo âm thanh từ hướng phải cho hình học loa phóng thanh ảo [20]. Đã cho vectơ DoA $\mathbf{r}_p(k, n)$, điều này cung cấp độ khuếch đại quét $G_i(\mathbf{r})$ cho từng kênh loa phóng thanh ảo i . Độ khuếch đại phụ thuộc khoảng cách cho từng DoA được suy ra từ độ dài thu được của vectơ hướng, $\mathbf{d}_p(k, n)$. Phép tổng hợp hướng cho kênh i trở thành:

$$Y_{i,S}(k, n) = \sqrt{1 - \psi(k, n)}P(k, n) G_i(\mathbf{r}_p(k, n))(\|\mathbf{d}_p(k, n)\|)^{-\gamma} \quad (10)$$

trong đó số mũ γ là thừa số điều chỉnh mà thường được đặt tới khoảng 1 [19]. Lưu ý là với $\gamma = 0$ độ khuếch đại phụ thuộc khoảng cách được ngắt.

Áp lực $P(k, n)$ được sử dụng để tạo ra I các tín hiệu giải tương quan $\tilde{P}_i(k, n)$. Các tín hiệu được giải tương quan này được cộng với các kênh loa phóng thanh riêng lẻ như thành phần khuếch tán. Điều này cho phép phương pháp tiêu chuẩn [16]:

$$Y_{i,D}(k, n) = \sqrt{\psi(k, n)} \frac{1}{\sqrt{I}} \tilde{P}_i(k, n) \quad (11)$$

Phần khuếch tán và phần hướng của mỗi kênh được bổ sung cùng nhau, và các tín hiệu được biến đổi trở lại thành miền thời gian bởi STFT nghịch đảo. Các tín hiệu miền thời gian kênh này được tích hợp với các HRTF cho tai trái và tai phải tùy thuộc vào vị trí loa phóng thanh để tạo các tín hiệu được lập thể hóa.

Để đánh giá, cảnh đơn trong phòng khách ảo được tái tạo. Các điều kiện kết xuất khác nhau được sử dụng để tái tạo ba nguồn âm thanh hoạt động đồng thời. Kỹ thuật MUSHRA-VR mới đã được sử dụng để truy cập chất lượng với sự hỗ trợ của các đối tượng thử nghiệm.

Môi trường ảo trong thí nghiệm là phòng trong nhà với ba nguồn âm thanh ở các khoảng cách khác nhau từ vị trí ghi. Ở khoảng 50 cm có người nói, ở 1m là vô tuyến và ở 2m là cửa sổ mở, tham chiếu Fig.7. Fig.7 minh họa kịch bản VR, trong đó âm thanh đến từ người, vô tuyến và cửa sổ mở, từng nguồn được đánh dấu bởi vòng tròn. Người dùng có thể đi bộ trong vùng được đánh dấu bởi hình chữ nhật nét đứt xung quanh người về phía cửa sổ.

Việc kết xuất ảo được thực hiện bằng cách sử dụng phần tử đơn vị và HTC VIVE. Việc xử lý âm thanh được thực hiện với sự hỗ trợ của các plugin công nghệ phòng thu ảo (virtual studio technology - VST) và Max/MSP. Dữ liệu theo dõi và các điều kiện được thay đổi thông qua thông báo điều khiển âm thanh mở (open sound control - OSC). Vùng đi bộ bằng khoảng 2 x 2 m.

Trong khi có các tiêu chuẩn được thiết lập để đánh giá phép tái tạo âm thanh tĩnh, chúng thường không thể ứng dụng trực tiếp cho VR. Đặc biệt là đối với 6DoF, các phương pháp mới để đánh giá chất lượng âm thanh phải được phát triển như trải nghiệm được làm phức tạp hơn phép đánh giá chỉ âm thanh, và nội dung được biểu diễn phụ thuộc vào đường chuyển động đơn nhất của từng người nghe. Các phương pháp mới như tìm đường (wayfinding) trong VR [21] hoặc các phản hồi sinh lý tới các

trải nghiệm đắm chìm [22] được nghiên cứu tích cực, nhưng các phương pháp thử nghiệm tốt truyền thống có thể cũng được làm thích ứng với môi trường VR để hỗ trợ công việc phát triển được thực hiện ngày nay.

MUSHRA là phương pháp đánh giá chất lượng âm thanh được chấp nhận rộng rãi áp dụng cho khoảng rộng các trường hợp sử dụng từ đánh giá chất lượng tiếng nói đến cài đặt âm thanh không gian đa kênh [17]. Nó cho phép việc so sánh cạnh nhau của sự tham chiếu với nhiều phép kết xuất của cùng nội dung âm thanh và cung cấp tỷ lệ chất lượng tuyệt đối qua việc sử dụng của các mục thử nghiệm neo và tham chiếu ẩn. Trong thử nghiệm này, hệ phương pháp MUSHRA được chấp nhận vào thiết lập VR, và do đó một số độ lệch khỏi các triển khai khuyến cáo là cần thiết. Đặc biệt, phiên bản được thực hiện ở đây không cho phép lặp nội dung âm thanh, và mục neo là kết xuất 3DoF.

Các điều kiện khác nhau được gán ngẫu nhiên cho các điều kiện thử nghiệm trong từng lần chạy. Từng người tham gia được yêu cầu đánh giá chất lượng âm thanh của từng điều kiện và cho điểm theo thang từ 0 đến 100. Thực tế, họ biết rằng một trong số các điều kiện là giống với tham chiếu và do đó sẽ cho điểm bằng 100 điểm. Điều kiện “neo” tệ nhất là sẽ được cho điểm 20 (kém) hoặc thấp hơn; tất cả các điều kiện khác cần được cho điểm giữa thang đó.

Bảng MUSHRA trong VR được mô tả trên Fig.8. Con chạy trên Fig.8 để đánh giá bốn điều kiện có thể được tạo màu, dưới chúng là giá trị số và từng nút để chọn điều kiện. Bảng được thiết kế theo cách mà các đánh giá của thử nghiệm dưới các hệ thống có thể được thực hiện ở thời điểm bất kỳ trong khi có giao diện kín đáo trong môi trường ảo. Bằng cách nhấn nút trên bộ điều khiển giữ tay, giao diện bán trong suốt được tạo nấp ở mức của mắt trong trường nhìn (field of view - FoV) của người dùng, ở khoảng cách thích hợp cho việc nhìn tự nhiên. Con trỏ laser có mặt mà sao chép các trạng thái trên con chuột (không hoạt động, hoạt động, được nhấn, được làm nổi bật) cho các nút để hỗ trợ giao diện. Nhấn nút tương tự trên bộ điều khiển giữ tay loại bỏ bảng nhưng duy trì tất cả các đánh giá hiện thời và phát lại lựa chọn điều kiện. Tất cả các đánh giá được ghi trong thời gian thực vào tệp tin bao gồm ghi chú về sự ngẫu nhiên hóa của các điều kiện.

Tổng bốn điều kiện khác nhau được thực hiện cho thí nghiệm.

REF Kết xuất dựa trên đối tượng. Đây là điều kiện tham chiếu. Định dạng B

được tạo ra nhanh chóng cho vị trí hiện tại của người nghe và sau đó được kết xuất thông qua loa ảo.

C1 Tái tạo 3DoF. Vị trí người nghe được bỏ qua, tức là $l(n) = 0$, nhưng sự xoay đầu của người nghe $o(n)$ vẫn được áp dụng. Độ khuếch đại được đặt đến độ khuếch đại của các nguồn trong khoảng cách bằng 2 mét đến người nghe. Điều kiện này được sử dụng như neo.

C2 Phương pháp được đề xuất cho phép tái tạo 6DoF mà không có thông tin khoảng cách. Vị trí người nghe được sử dụng để thay đổi vectơ hướng. Tất cả các nguồn được định vị trí ở hình cầu phía ngoài diện tích đi bộ. Bán kính của hình cầu được cố định tới 2m, tức là, $m(r) = 2\forall$, và độ khuếch đại phụ thuộc khoảng cách được áp dụng ($\gamma = 1$).

C3 Phương pháp được đề xuất của phép tái tạo 6DoF với thông tin khoảng cách. Vị trí người nghe $l(n)$ được sử dụng để thay đổi vectơ hướng. Thông tin khoảng cách $m(r)$ được sử dụng để tính toán DoA đúng ở vị trí người nghe (5), và độ khuếch đại phụ thuộc khoảng cách (6) được áp dụng ($\gamma = 1$).

Đường liên hợp xử lý tín hiệu tương tự được sử dụng cho tất cả các điều kiện. Điều này được thực hiện để đảm bảo rằng việc so sánh được tập trung vào chỉ phép tái tạo không gian và kết quả không bị ảnh hưởng bởi màu sắc hoặc hiệu ứng khác. Đường liên hợp được thể hiện trên Fig.10. Fig.10 minh họa các đường tín hiệu cho việc kết xuất tham chiếu và DirAC. Trong trường hợp tham chiếu, dữ liệu theo dõi được sử dụng để thay đổi việc định vị trí và sự xoay của phép tổng hợp định dạng B dựa trên đối tượng (bên trái phía trên). Trong các điều kiện C1-C3 khác, dữ liệu theo dõi được áp dụng trong miền DirAC (bên phải).

Hai tín hiệu định dạng B được tính toán từ ba tín hiệu nguồn đơn. Tín hiệu trực tiếp (khô) được tính toán trực tuyến. Tín hiệu âm vang (ẩm) được tính toán trước ngoại tuyến. Chúng được cộng với nhau và được xử lý bởi DirAC mà kết xuất tới loa phóng thanh ảo, sau đó được nhị phân hóa. Chênh lệch nằm trong ứng dụng của dữ liệu theo dõi. Trong trường hợp tham chiếu, nó được áp dụng trước khi tổng hợp tín hiệu định dạng B, sao cho nó được ghi ảo ở vị trí người nghe. Trong các trường hợp khác, nó được áp dụng trong miền DirAC.

Việc kết xuất dựa trên đối tượng được sử dụng như kịch bản tham chiếu. Hầu như, người nghe được trang bị micrô định dạng B trên đầu của họ và thực hiện việc

ghi ở vị trí và sự xoay đầu của họ. Việc này được thực hiện đơn giản như sau: các đối tượng được đặt tương ứng với vị trí người nghe được theo dõi. Tín hiệu FOA được tạo ra từ từng nguồn với sự suy giảm khoảng cách. Tín hiệu định dạng B trực tiếp tổng hợp s_i cho tín hiệu nguồn $s_i(t)$ ở khoảng cách d_i , hướng với góc phương vị θ và góc nâng ψ là:

$$s_i(t) = 1/d_i \begin{pmatrix} 1 \\ \cos \theta \cos \phi \\ \sin \theta \cos \phi \\ \sin \phi \end{pmatrix} s_i(t - \frac{d_i}{c}) \quad (12)$$

trong đó c là tốc độ âm thanh theo m/giây. Sau đó, sự xoay được theo dõi được áp dụng trong miền FOA [7].

Âm vang nhân tạo được thêm vào tín hiệu nguồn theo cách bất biến theo thời gian để nâng cao tính chân thực của cảnh âm thanh trong nhà được kết xuất. Những phản xạ sớm từ ranh giới của căn phòng hình hộp giày được thêm vào với độ trễ, hướng và độ suy giảm chính xác. Âm vang muộn được tạo ra với mạng trễ phản hồi không gian (feedback delay network -FDN) mà phân phối đầu ra đa kênh cho cài đặt loa ảo [23]. Thời gian vang phụ thuộc tần số T_{60} là nằm giữa 90 và 150 mili giây với trung bình là 110 mili giây. Bộ lọc hiệu chỉnh âm với đặc tính thông thấp được áp dụng sau đó.

Tín hiệu vang sau đó được chuyển đổi từ cài đặt loa ảo 8+4 sang định dạng B bằng cách nhân từng tín hiệu loa ảo với mẫu định dạng B của DoA của chúng như trong (12). Tín hiệu định dạng B vang lại được thêm vào tín hiệu trực tiếp

Định dạng B tóm tắt được xử lý trong miền DirAC. Việc mã hóa được thực hiện bằng cách sử dụng giàn lọc bộ lọc gương cầu phương (quadrature mirror filter - QMF) với 128 băng, được chọn cho độ phân giải thời gian cao và tạo rãnh cửa theo thời gian thấp. Cả hướng và độ khuếch tán được ước tính với độ mịn trung bình di chuyển là 42 mili giây. Việc giải mã đang tạo ra tín hiệu loa ảo 8+4. Các tín hiệu 8+4 này sau đó được kết hợp với các HRTF để phát lại lập thể.

Tổng cộng 19 đối tượng đánh giá cảnh. Họ có độ tuổi từ 23 đến 41, ba trong số họ là nữ, tất cả đều báo cáo không bị khiếm thính. Hầu hết những người tham gia cần ít hơn mười phút để đánh giá. Các đối tượng mất nhiều thời gian hơn khi không quen với việc đánh giá âm thanh thực tế ảo, trong đó âm thanh và thị giác không phải luôn trùng khớp.

Fig.6 thể hiện điểm số kết quả là các ô hình hộp, tức là, các đánh giá MUSHRA (N = 19) dưới dạng các ô hình hộp. Đường chấm chấm biểu thị điểm trung bình, các hộp ở phần tư thứ nhất đến thứ ba, các râu ở mức $\pm 1,5$ độ trải giữa (inter-quartile range - IQR).

Có thể thấy rằng tất cả các đối tượng xác định chính xác tham chiếu là tốt nhất, mặc dù 4 trong số họ đánh giá nó dưới 100. Trong khi nghe có vẻ giống với vị trí ghi, nhưng các khác biệt với các điều kiện khác là rõ ràng đối với tất cả những người tham gia. Phép tái tạo 6DoF được đề xuất trong miền DirAC với thông tin khoảng cách (C3) có điểm tổng thể cao thứ hai. Phép tái tạo mà không có thông tin khoảng cách (C2) hoặc thậm chí không có theo dõi vị trí (C1) được hầu hết mọi người tham gia ghi điểm thấp hơn. Có thể thấy rằng những người tham gia đã không đồng ý về giá trị được gán cho điều kiện neo (C1). Trong khi 13 điểm dưới 30, sáu người còn lại không chắc chắn và chọn giá trị lên tới 70.

Hiệu quả chính đáng kể của điều kiện đã được tìm thấy ($p < 0,001$, $F = 43,75$) theo phân tích phương sai số đo lặp lại một chiều (ANOVA). Như phép phân tích hậu kiểm (post hoc analysis), nhiều so sánh của Tukey có trung bình 95% mức tin cậy theo gia đình được thực hiện. Tất cả các cặp điều kiện được tìm thấy khác nhau đáng kể, mạnh nhất là ($p < 0,001$), chỉ có C2-C3 không rõ ràng ($p < 0,04$).

Ngay cả khi các điều kiện được tìm thấy là khác nhau đáng kể, phương sai trong các phản hồi là tương đối lớn. Một lý do cho điều này có thể là các mức kinh nghiệm khác nhau của các đối tượng thử nghiệm với VR. Có thể nên làm thử nghiệm trước làm quen hoặc nhóm các đối tượng theo kinh nghiệm. Tuy nhiên, việc sử dụng một loạt người mới làm chuyên gia về các thử nghiệm VR và thử nghiệm nghe trong khi vẫn tạo ra hiệu ứng đáng kể cho thấy kết quả nắm giữ các yếu tố này.

Một số người tham gia gặp khó khăn khi có điều kiện 3DoF là neo. Điều này cũng có thể phản ánh sự thiếu kinh nghiệm trong âm thanh VR. Tuy nhiên, nó có thể đơn giản hóa thủ tục và giúp thống nhất để cung cấp neo bổ sung, không theo không gian, chẳng hạn như hỗn hợp đơn của các nguồn âm thanh.

Về phương pháp tái tạo được đề xuất, người ta thấy rằng nó cho phép tái tạo nội dung FOA, được ghi lại tại một điểm trong không gian, trong 6DoF. Trong khi hầu hết những người tham gia thử nghiệm đánh giá sự tham chiếu tín hiệu B-Format lý tưởng cao hơn, phương pháp được đề xuất đã đạt được điểm trung bình cao nhất để tái tạo

trong số các điều kiện khác. Phương pháp được đề xuất hoạt động ngay cả khi các nguồn âm thanh trong bản ghi được đặt ở các khoảng cách khác nhau từ micrô. Trong trường hợp đó, các khoảng cách phải được ghi lại dưới dạng siêu dữ liệu sẽ được tái. Kết quả cho thấy việc tái tạo khoảng cách giúp tăng cường tính chân thực của trải nghiệm. Hiệu ứng có thể mạnh hơn nếu khu vực đi bộ cho phép người dùng đi bộ xung quanh tất cả các nguồn âm thanh.

Phương pháp tái tạo âm thanh mới trong sáu độ tự do (6DoF) đã được đề xuất. Âm thanh được ghi dưới dạng Ambisonic (FOA) cấp một tại một vị trí và dữ liệu khoảng cách cho các nguồn âm thanh được lấy làm thông tin phụ. Sử dụng thông tin này, âm thanh được tái tạo đối với việc theo dõi trực tiếp của người nghe trong miền mã hóa âm thanh định hướng tham số (DirAC).

Thử nghiệm chủ quan cho thấy phương pháp được đề xuất được xếp hạng chặt chẽ với kết xuất dựa trên đối tượng. Điều này ngụ ý rằng phương pháp tái tạo được đề xuất có thể cung cấp thành công việc phát lại ảo vượt quá ba độ tự do khi thông tin xa được tính đến.

Fig.11a minh họa phương án triển khai nửa của bộ tính toán trường âm thanh 420.

Dựa trên mô tả trường âm thanh, hướng đến của băng toàn bộ hoặc hướng đến trên mỗi băng được xác định trong 1100. Các thông tin hướng đến biểu diễn dữ liệu hướng đến của trường âm thanh. Dựa trên dữ liệu hướng đến, thông tin tịnh tiến được thực hiện trong khối 1110. Để đạt được điều này, bản đồ chiều sâu 1120 được chứa như siêu dữ liệu cho mô tả trường âm thanh được sử dụng. Dựa trên bản đồ chiều sâu 1120, khối 1110 tạo ra dữ liệu hướng đến mới cho trường âm thanh mà, trong phương án triển khai này, chỉ phụ thuộc vào sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác. Để đạt được điều này, khối 1110 nhận thông tin tịnh tiến được tạo ra, ví dụ, bằng cách theo dõi trong ngữ cảnh thực hiện thực tế ảo.

Tốt hơn là hoặc theo cách khác, dữ liệu xoay cũng được sử dụng. Để đạt được điều này, khối 1130 thực hiện việc biến đổi xoay sử dụng thông tin xoay. Khi cả việc tịnh tiến và xoay được thực hiện, sau đó tốt hơn là thực hiện biến đổi xoay sau khi tính toán các DoA mới của trường âm thanh mà đã bao gồm thông tin từ phép tịnh tiến và bản đồ chiều sâu 1120.

Sau đó, trong khối 1140, mô tả trường âm thanh mới được tạo ra. Để đạt được điều này, mô tả trường âm thanh gốc có thể được sử dụng hoặc, theo cách khác, các nguồn tín hiệu mà đã được tách từ mô tả trường âm thanh bởi thuật toán tách nguồn có thể được sử dụng hoặc ứng dụng khác bất kỳ có thể được sử dụng. Về cơ bản, mô tả trường âm thanh mới có thể, ví dụ, mô tả trường âm thanh định hướng như thu được bởi bộ tạo Ambisonic 430 hoặc như được tạo ra bởi bộ tổng hợp DirAC 425 hoặc có thể là sự biểu diễn lập thể được tạo ra từ sự biểu diễn loa ảo trong việc kết xuất lập thể sau đó.

Tốt hơn là, như được minh họa trên Fig.11a, khoảng cách trên hướng đến cũng được sử dụng để tạo mô tả trường âm thanh mới để làm thích ứng âm lượng hoặc độ to của nguồn âm thanh nhất định với vị trí mới, tức là vị trí tham chiếu mới khác.

Dù Fig.11a minh họa trường hợp, trong đó phép biến đổi xoay được thực hiện sau thông tin tịnh tiến, lưu ý là thứ tự có thể là khác nhau. Cụ thể, phép biến đổi xoay có thể được áp dụng cho các DoA của trường âm thanh như được tạo ra bởi khối 1100 và, sau đó, phép biến đổi tịnh tiến bổ sung được áp dụng mà do sự tịnh tiến của đối tượng từ vị trí tham chiếu đến vị trí tham chiếu khác.

Tuy nhiên, lưu ý là các DoA của trường âm thanh phải được sử dụng để tìm thấy thông tin khoảng cách tương ứng từ bản đồ chiều sâu 1120 hơn là các DoA được xoay. Do đó, ngay sau khi các DoA của trường âm thanh đã được xác định bởi khối 1100, thông tin khoảng cách được yêu cầu bằng cách sử dụng bản đồ chiều sâu 1120 và thông tin khoảng cách này sau đó được sử dụng bằng cách tạo mô tả trường âm thanh mới trong khối 1140 để tính khoảng cách thay đổi và, do đó, âm lượng thay đổi của nguồn nhất định đối với vị trí tham chiếu nhất định. Về cơ bản, có thể nói rằng trong trường hợp khoảng cách trở nên lớn hơn thì tín hiệu nguồn âm thanh cụ thể bị suy giảm trong khi, khi khoảng cách trở nên ngắn hơn thì tín hiệu nguồn âm thanh được khuếch đại. Đương nhiên, sự suy giảm và khuếch đại của nguồn âm thanh nhất định phụ thuộc vào khoảng cách được tạo ra tỷ lệ với sự thay đổi khoảng cách, nhưng, trong các phương án khác, các phép toán ít phức tạp hơn có thể được áp dụng cho sự khuếch đại hoặc suy giảm tín hiệu nguồn âm thanh này với mức tăng khá thấp. Ngay cả việc thực hiện ít phức tạp như vậy cũng mang lại kết quả vượt trội so với trường hợp mà sự thay đổi khoảng cách bất kỳ được bỏ qua hoàn toàn.

Fig.11b minh họa phương án triển khai ưu tiên của bộ tính toán trường âm thanh 420. Trong khối 1102, việc tách nguồn và việc tính toán thông tin hướng đến hoặc hướng chung cho từng nguồn được thực hiện. Sau đó, trong khối 1104, vectơ hướng đến được nhân với vectơ thông tin khoảng cách, tức là, vectơ từ vị trí tham chiếu ban đầu đến nguồn âm thanh, tức là, ví dụ, vectơ từ mục 520 đến mục 510 của Fig.5. Sau đó, trong khối 1106, thông tin tịnh tiến, tức là, vectơ từ mục 520 đến mục 500 của Fig.5 được tính đến để tính toán vectơ hướng được tịnh tiến mới tức là vectơ từ vị trí người nghe 500 đến vị trí nguồn âm thanh 510. Sau đó, vectơ hướng đến mới với độ dài đúng được biểu thị bởi d_v được tính toán trong khối 1108. Vectơ này được định hướng theo cùng hướng như d_r , nhưng có độ dài khác, vì độ dài của vectơ này phản ánh thực tế là nguồn âm thanh 510 được ghi trong trường âm thanh gốc với âm lượng nhất định và, do đó, độ dài của d_v lớn hơn hoặc nhỏ hơn biểu thị sự thay đổi âm lượng. Điều này thu được bằng cách chia vectơ d_i cho khoảng cách ghi d_r , tức là, độ dài của vectơ d_r từ micrô 520 đến nguồn âm thanh 510.

Như trên Fig.5, khi khoảng cách được tái tạo lớn hơn khoảng cách được ghi, độ dài của d_v sẽ thấp hơn phần tử đơn vị. Điều này dẫn đến sự suy giảm nguồn âm thanh 510 cho phép tái tạo ở vị trí người nghe mới. Tuy nhiên, khi khoảng cách được tái tạo d_i nhỏ hơn khoảng cách được ghi, sau đó độ dài của d_v như được tính toán bởi khối 1108 sẽ lớn hơn 1 và thừa số định tỷ lệ tương ứng sẽ dẫn đến sự khuếch đại của nguồn âm thanh.

Fig.12 minh họa phương án triển khai ưu tiên nữa của bộ tính toán trường âm thanh.

Trong khối 1200, các nguồn riêng lẻ từ trường âm thanh được xác định, ví dụ, giống như trên mỗi băng hoặc băng đầy đủ. Khi việc xác định trên mỗi khung và băng được thực hiện thì việc này có thể được thực hiện bởi phép phân tích DirAC. Nếu việc xác định băng đầy đủ hoặc băng con được thực hiện thì điều này có thể được thực hiện bởi loại thuật toán tách nguồn băng đầy đủ hoặc băng con bất kỳ.

Trong khối 1210, sự tịnh tiến hoặc sự xoay của người nghe được xác định, ví dụ, bằng cách theo dõi đầu.

Trong khối 1220, khoảng cách cũ cho từng nguồn được xác định bằng cách sử dụng siêu dữ liệu và, ví dụ, bằng cách sử dụng bản đồ chiều sâu trong việc thực hiện phép phân tích DirAC. Do đó, từng băng được xem xét để là nguồn nhất định (miễn là

độ khuếch tán thấp hơn ngưỡng nhất định), và sau đó, khoảng cách nhất định cho từng ngắn thời gian/tần số có giá trị khuếch tán thấp được xác định.

Sau đó, trong khối 1230, khoảng cách mới trên mỗi nguồn thu được, ví dụ, bởi phép tính toán vectơ trên mỗi băng mà, ví dụ, được mô tả trong ngữ cảnh của Fig.11b.

Hơn nữa, như được minh họa trong khối 1240, hướng cũ cho mỗi nguồn được xác định, ví dụ, bởi phép tính DoA thu được trong phép phân tích DirAC hoặc bởi hướng đến hoặc phép phân tích thông tin hướng trong thuật toán tách nguồn, ví dụ.

Sau đó, trong khối 1250, hướng mới cho mỗi nguồn được xác định, ví dụ, bằng cách thực hiện phép tính vectơ cho mỗi băng hoặc băng đầy đủ.

Sau đó, trong khối 1260, trường âm thanh mới được tạo cho người nghe được tịnh tiến và xoay. Điều này có thể được thực hiện, ví dụ, bằng cách định tỷ lệ phần hướng cho mỗi kênh trong phép tổng hợp DirAC. Dựa trên phương án triển khai cụ thể, phép biến đổi khoảng cách có thể được thực hiện trong các khối 1270a, 1270b hoặc 1270c ngoài ra hoặc thay thế để thực hiện phép biến đổi khoảng cách trong khối 1260.

Ví dụ, khi được xác định rằng trường âm thanh chỉ có duy nhất một nguồn âm thanh, thì phép biến đổi khoảng cách có thể sẵn sàng được thực hiện trong khối 1270a.

Thay vào đó, khi các tín hiệu nguồn riêng lẻ được tính toán bởi khối 1200, thì phép biến đổi khoảng cách có thể được thực hiện cho các nguồn riêng lẻ trong khối 1270b, trước khi trường âm thanh mới thực sự được tạo trong khối 1260.

Ngoài ra, khi việc tạo trường âm thanh trong khối 1260, ví dụ, không kết xuất tín hiệu cài đặt loa phóng thanh hoặc tín hiệu lập thể, mà mô tả trường âm thanh khác, ví dụ, bằng cách sử dụng bộ mã hóa hoặc bộ tính toán Ambisonic 430, thì phép biến đổi khoảng cách cũng có thể được thực hiện sau khi tạo trong khối 1260, mà có nghĩa là trong khối 1270c. Tùy thuộc vào phương án triển khai, phép biến đổi khoảng cách cũng có thể được phân bổ cho một số bộ biến đổi để, đến cuối cùng, nguồn âm thanh nhất định có cường độ âm nhất định được định hướng bởi chênh lệch giữa khoảng cách ban đầu giữa nguồn âm thanh và vị trí tham chiếu và khoảng cách mới giữa nguồn âm thanh và vị trí tham chiếu khác.

Fig.13a minh họa bộ phân tích DirAC như được bộc lộ ban đầu, ví dụ, trong tài liệu tham khảo được trích dẫn trước đó "Mã hóa âm thanh có hướng" ("Directional Audio Coding") từ IWPASH năm 2009.

Bộ phân tích DirAC bao gồm bộ lọc băng tần 1310, bộ phân tích năng lượng 1320, bộ phân tích cường độ 1330, khối lấy trung bình theo thời gian 1340 và bộ tính toán độ khuếch tán 1350 và bộ tính toán hướng 1360.

Trong DirAC, cả phép phân tích và tổng hợp đều được thực hiện trong miền tần số. Có một số phương pháp để chia âm thanh thành các băng tần, trong từng thuộc tính riêng biệt. Các biến đổi tần số được sử dụng phổ biến nhất bao gồm biến đổi Fourier thời gian ngắn (STFT - short time Fourier transform), và giàn lọc gương cầu phương (QMF - Quadrature mirror filter). Ngoài ra, có toàn quyền tự do để thiết kế giàn lọc với các bộ lọc tùy ý được tối ưu hóa cho bất kỳ mục đích cụ thể nào. Mục đích của phân tích hướng là ước tính, ở mỗi băng tần, hướng đến của âm thanh, cùng với ước tính nếu âm thanh đến từ một hoặc nhiều hướng cùng một lúc. Về nguyên tắc, điều này có thể được thực hiện với một số kỹ thuật, tuy nhiên, phân tích năng lượng của trường âm thanh đã được tìm thấy là phù hợp, mà minh họa trên Fig.13a. Phân tích năng lượng có thể được thực hiện, khi tín hiệu áp lực và các tín hiệu vận tốc ở một, hai hoặc ba chiều được thu ở một vị trí duy nhất. Trong các tín hiệu định dạng B cấp một, tín hiệu đẳng hướng được gọi là tín hiệu W, mà được thu nhỏ bởi căn bậc hai của hai. Áp lực âm thanh có thể được ước tính là $P = \sqrt{2} * W$ được biểu diễn trong miền STFT.

Các kênh X, Y, Z có mẫu định hướng của lưỡng cực được định hướng dọc theo trục Đề-các, tạo thành vectơ $U = [X, Y, Z]$. Vectơ ước tính vectơ vận tốc trường âm thanh, và cũng được biểu diễn trong miền STFT. Năng lượng E của trường âm thanh được tính toán. Việc thu các tín hiệu định dạng B có thể thu được hoặc bằng cách định vị ngẫu nhiên các micrô có hướng, hoặc với tập hợp các micrô đẳng hướng có khoảng cách gần nhau. Trong một số ứng dụng, các tín hiệu micrô có thể được tạo thành trong miền tính toán, nghĩa là, được mô phỏng.

Hướng của âm thanh được xác định là hướng ngược lại của vectơ cường độ I. Hướng được biểu thị là góc phương vị và giá trị độ cao tương ứng trong siêu dữ liệu được truyền. Độ khuếch tán của trường âm thanh cũng được tính toán bằng cách sử dụng toán tử kỳ vọng của vectơ cường độ và năng lượng. Kết quả của phương trình này là một số có giá trị thực nằm giữa 0 và 1, đặc trưng nếu năng lượng âm thanh đến từ một hướng duy nhất (độ khuếch tán bằng 0), hoặc từ tất cả các hướng (độ khuếch tán bằng 1). Quy trình này phù hợp trong trường hợp có sẵn thông tin vận tốc 3 chiều (3D) đầy đủ hoặc ít hơn.

Fig.13b minh họa phép tổng hợp DirAC, một lần nữa có giàn của các bộ lọc băng 1370, khối micrô ảo 1400, khối bộ tổng hợp hướng/khuếch tán 1450, và cài đặt loa phóng thanh nhất định hoặc cài đặt loa phóng thanh dự định ảo 1460. Ngoài ra, bộ biến đổi độ khuếch đại khuếch tán 1380, khối bảng khuếch đại quét biên độ dựa trên véctor (VBAP - vector based amplitude panning) 1390, khối bù micrô 1420, khối lấy trung bình độ khuếch đại loa phóng thanh 1430 và bộ phân phối 1440 cho các kênh khác được sử dụng.

Trong phép tổng hợp DirAC này, với các loa phóng thanh, phiên bản chất lượng cao của tổng hợp DirAC được thể hiện trên Fig.13b nhận tất cả các tín hiệu định dạng B, trong đó tín hiệu micrô ảo được tính toán cho mỗi hướng loa phóng thanh của cài đặt loa phóng thanh 1460. Mẫu định hướng được sử dụng thường là lưỡng cực. Các tín hiệu micrô ảo sau đó được sửa đổi theo kiểu phi tuyến tính, tùy thuộc vào siêu dữ liệu. Phiên bản tốc độ bit thấp của DirAC không được thể hiện trên Fig.12b, tuy nhiên, trong tình huống này, chỉ một kênh của âm thanh được truyền như được minh họa trên Fig.6. Sự khác biệt trong bước xử lý là tất cả các tín hiệu micrô ảo sẽ được thay thế bằng một kênh âm thanh duy nhất được nhận. Các tín hiệu micrô ảo được chia thành hai dòng: dòng khuếch tán và dòng không khuếch tán, mà được xử lý riêng.

Âm thanh không khuếch tán được tái tạo dưới dạng các nguồn điểm bằng cách sử dụng sự quét biên độ dựa trên véctor (VBAP - vector base amplitude panning). Khi quét, tín hiệu âm thanh đơn được áp dụng cho tập con của các loa phóng thanh sau khi nhân với các hệ số khuếch đại dành riêng cho loa phóng thanh. Các hệ số khuếch đại được tính toán bằng cách sử dụng thông tin cài đặt loa phóng thanh, và hướng quét được chỉ định. Trong phiên bản tốc độ bit thấp, tín hiệu đầu vào được quét đơn giản đến các hướng được ám chỉ bởi siêu dữ liệu. Trong phiên bản chất lượng cao, mỗi tín hiệu micrô ảo được nhân với hệ số khuếch đại tương ứng, tạo ra hiệu ứng tương tự với việc quét, tuy nhiên nó ít bị ảnh hưởng bởi bất kỳ thành phần lạ phi tuyến nào.

Trong nhiều trường hợp, siêu dữ liệu định hướng phải tuân theo các thay đổi đột ngột theo thời gian. Để tránh các thành phần lạ, các hệ số khuếch đại cho các loa phóng thanh được tính toán với VBAP được làm nhẵn bằng cách tích hợp tạm thời với các hằng số thời gian phụ thuộc vào tần số tương đương với khoảng 50 chu kỳ ở mỗi băng tần. Điều này có hiệu quả loại bỏ các thành phần lạ, tuy nhiên, các thay đổi theo

hướng không được cảm nhận là chậm hơn so với không lấy trung bình trong hầu hết các trường hợp.

Mục đích của phép tổng hợp âm thanh khuếch tán là để tạo cảm nhận âm thanh bao quanh người nghe. Trong phiên bản tốc độ bit thấp, dòng khuếch tán được tái tạo bằng cách giải tương quan tín hiệu đầu ra và tái tạo nó từ mỗi loa. Trong phiên bản chất lượng cao, các tín hiệu micrô ảo của dòng khuếch tán đã không nhất quán ở một mức độ nào đó, và chúng cần được khử tương quan một cách vừa phải. Phương pháp này cung cấp chất lượng không gian tốt hơn cho âm vang xung quanh và âm thanh không gian xung quanh so với phiên bản tốc độ bit thấp.

Đối với phép tổng hợp DirAC với tai nghe chụp đầu, DirAC được tạo ra với lượng loa phóng thanh ảo nhất định xung quanh người nghe cho dòng không khuếch tán và lượng loa phóng thanh nhất định cho dòng khuếch tán. Các loa phóng thanh ảo được thực hiện dưới dạng tích chập của các tín hiệu đầu vào với các hàm truyền liên quan đến đầu (HRTFs - head-related transfer functions) đo được.

Mặc dù một số khía cạnh được mô tả trong ngữ cảnh của thiết bị, rõ ràng là các khía cạnh này cũng biểu diễn sự mô tả của các phương pháp tương ứng, trong đó khối hoặc thiết bị tương ứng với bước phương pháp hoặc đặc điểm của bước phương pháp. Tương tự, các khía cạnh được mô tả trong ngữ cảnh của bước phương pháp cũng biểu diễn sự mô tả của khối hoặc mục tương ứng hoặc đặc điểm của thiết bị tương ứng.

Mô tả trường âm thanh tăng cường theo sáng chế có thể được lưu trữ trong vật ghi lưu trữ số hoặc vật ghi lưu trữ không chuyển tiếp hoặc có thể được truyền trên vật ghi truyền chẳng hạn như vật ghi truyền không dây hoặc vật ghi truyền có dây chẳng hạn như Internet.

Tùy thuộc vào các yêu cầu thực hiện nhất định, các phương án theo sáng chế có thể được thực hiện trong phần cứng hoặc phần mềm. Phương án triển khai có thể được thực hiện bằng cách sử dụng vật ghi lưu trữ số, ví dụ như đĩa mềm, DVD, CD, ROM, PROM, EPROM, EEPROM, 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.

Một số phương án theo sáng chế bao gồm vật mang dữ liệu không chuyển tiếp có các tín hiệu điều khiển có thể đọc được bằng điện tử, có khả năng kết hợp với hệ

thống máy tính có thể lập trình, chẳng hạn như một trong số các phương pháp được mô tả ở đây được thực hiện.

Nói chung, các phương án theo sáng chế có thể được thực hiện dưới dạng sản phẩm chương trình máy tính với mã chương trình, mã chương trình có tác dụng thực hiện một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Mã chương trình, ví dụ, có thể được lưu trữ trên vật mang có thể đọc được bằng máy.

Các phương án 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.

Nói cách khác, phương án theo phương pháp sáng chế, do đó, là chương trình máy tính có mã 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, phương án nữa của các phương pháp sáng chế là vật mang dữ liệu (hoặc vật ghi lưu trữ số, hoặc vật ghi đọc được bằng máy tính) bao gồm, được ghi trong đó, chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây.

Do đó, phương án nữa của phương pháp sáng chế là dòng dữ liệu hoặc chuỗi tín hiệu biểu diễn chương trình máy tính để thực hiện một trong số các phương pháp được mô tả trong tài liệu này. Dòng dữ liệu hoặc chuỗi tín hiệu có thể ví dụ được tạo cấu hình để biến đổi thông qua kết nối truyền thông dữ liệu, ví dụ thông qua Internet.

Phương án nữa của sáng chế bao gồm phương tiện xử lý, ví dụ máy tính, hoặc thiết bị logic lập trình, được tạo cấu hình hoặc được điều chỉnh để thực hiện một trong số các phương pháp được mô tả trong tài liệu này.

Phương án nữa của sáng chế bao gồm máy tính được cài đặt trong đó chương trình máy tính để thực hiện một trong số các phương pháp được mô tả trong tài liệu này.

Trong một số phương án, 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 trong tài liệu này. Trong một số phương án, 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 trong tài liệu này. Nói chung, các phương pháp tốt hơn là được thực hiện bởi thiết bị phần cứng bất kỳ.

Các phương án được mô tả ở trên chỉ mang tính minh họa cho các nguyên tắc của sáng chế. Điều này được hiểu rằng các biến đổi và biến thể của các sắp xếp và các chi tiết được mô tả trong tài liệu này sẽ là rõ ràng đối với những người có hiểu biết trung bình trong lĩnh vực kỹ thuật tương ứng. Do đó, sáng chế chỉ bị giới hạn bởi phạm vi của các điểm yêu cầu bảo hộ mà không phải bởi các chi tiết cụ thể được mô tả và giải thích theo các phương án theo sáng chế trong tài liệu này.

Tài liệu tham khảo

- [1] Liitola, T., *Headphone sound externalization*, Ph.D. thesis, Helsinki University of Technology. Department of Electrical and Communications Engineering Laboratory of Acoustics and Audio Signal Processing., 2006.
- [2] Blauert, J., *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, The MIT Press, 1996, ISBN 0262024136.
- [3] Zhang, W., Samarasinghe, P. N., Chen, H., and Abhayapala, T. D., “Surround by Sound: A Review of Spatial Audio Recording and Reproduction,” *Applied Sciences*, 7(5), p. 532, 2017.
- [4] Bates, E. and Boland, F., “Spatial Music, Virtual Reality, and 360 Media,” in *Audio Eng. Soc. Int. Conf. on Audio for Virtual and Augmented Reality*, Los Angeles, CA, U.S.A., 2016.
- [5] Anderson, R., Gallup, D., Barron, J. T., Kontkanen, J., Snavely, N., Esteban, C. H., Agarwal, S., and Seitz, S. M., “Jump: Virtual Reality Video,” *ACM Transactions on Graphics*, 35(6), p. 198, 2016.
- [6] Merimaa, J., *Analysis, Synthesis, and Perception of Spatial Sound: Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*, Ph.D. thesis, Helsinki University of Technology, 2006.
- [7] Kronlachner, M. and Zotter, F., “Spatial Transformations for the Enhancement of Ambisonic Recordings,” in *2nd International Conference on Spatial Audio*, Erlangen, Germany, 2014.
- [8] Tsingos, N., Gallo, E., and Drettakis, G., “Perceptual Audio Rendering of Complex Virtual Environments,” *ACM Transactions on Graphics*, 23(3), pp. 249–258, 2004.
- [9] Taylor, M., Chandak, A., Mo, Q., Lauterbach, C., Schissler, C., and Manocha, D., “Guided multi-view ray tracing for fast auralization,” *IEEE Trans.*

Visualization & Comp. Graphics, 18, pp. 1797– 1810, 2012.

[10] Rungta, A., Schissler, C., Rewkowski, N., Mehra, R., and Manocha, D., “Diffraction Kernels for Interactive Sound Propagation in Dynamic Environments,” *IEEE Trans. Visualization & Comp. Graphics*, 24(4), pp. 1613–1622, 2018.

[11] Thiergart, O., Kowalczyk, K., and Habets, E. A. P., “An Acoustical Zoom based on Informed Spatial Filtering,” in *Int. Workshop on Acoustic Signal Enhancement*, pp. 109–113, 2014.

[12] Khaddour, H., Schimmel, J., and Rund, F., “A Novel Combined System of Direction Estimation and Sound Zooming of Multiple Speakers,” *Radioengineering*, 24(2), 2015.

[13] Ziegler, M., Keinert, J., Holzer, N., Wolf, T., Jaschke, T., op het Veld, R., Zakeri, F. S., and Foessel, S., “Immersive Virtual Reality for Live-Action Video using Camera Arrays,” in *IBC*, Amsterdam, Netherlands, 2017.

[14] Thiergart, O., Galdo, G. D., Taseska, M., and Habets, E. A. P., “Geometry-Based Spatial Sound Acquisition using Distributed Microphone Arrays,” *IEEE Trans. Audio, Speech, Language Process.*, 21(12), pp. 2583–2594, 2013.

[15] Kowalczyk, K., Thiergart, O., Taseska, M., Del Galdo, G., Pulkki, V., and Habets, E. A. P., “Parametric Spatial Sound Processing: A Flexible and Efficient Solution to Sound Scene Acquisition, Modification, and Reproduction,” *IEEE Signal Process. Mag.*, 32(2), pp. 31–42, 2015.

[16] Pulkki, V., “Spatial Sound Reproduction with Directional Audio Coding,” *J. Audio Eng. Soc.*, 55(6), pp. 503–516, 2007.

[17] International Telecommunication Union, “ITU-R BS.1534-3, Method for the subjective assessment of intermediate quality level of audio systems,” 2015.

[18] Thiergart, O., Del Galdo, G., Kuech, F., and Prus, M., “Three-Dimensional Sound Field Analysis with Directional Audio Coding Based on Signal Adaptive Parameter Estimators,” in *Audio Eng. Soc. Conv. Spatial Audio: Sense the Sound of Space*, 2010.

[19] Kuttruff, H., *Room Acoustics*, Taylor & Francis, 4 edition, 2000.

[20] Borß, C., “A polygon-based panning method for 3D loudspeaker setups,” in *Audio Eng. Soc. Conv.*, pp. 343–352, Los Angeles, CA, USA, 2014.

[21] Rummukainen, O., Schlecht, S., Plinge, A., and Habets, E. A. P.,

“Evaluating Binaural Reproduction Systems from Behavioral Patterns in a Virtual Reality – A Case Study with Impaired Binaural Cues and Tracking Latency,” in *Audio Eng. Soc. Conv. 143*, New York, NY, USA, 2017.

[22] Engelke, U., Darcy, D. P., Mulliken, G. H., Bosse, S., Martini, M. G., Arndt, S., Antons, J.-N., Chan, K. Y., Ramzan, N., and Brunnström, K., “Psychophysiology-Based QoE Assessment: A Survey,” *IEEE Selected Topics in Signal Processing*, 11(1), pp. 6–21, 2017.

[23] Schlecht, S. J. and Habets, E. A. P., “Sign-Agnostic Matrix Design for Spatial Artificial Reverberation with Feedback Delay Networks,” in *Proc. Audio Eng. Soc. Conf.*, pp. 1–10– accepted, Tokyo, Japan, 2018

[31] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Acoust. Soc. Am.*, vol. 21, no. 1, pp. 2-10, 1973.

[32] V. Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing," in *Proc. of the 28th AES International Conference*, 2006.

[33] —, "Spatial sound reproduction with directional audio coding," *Journal Audio Eng. Soc.*, vol. 55, no. 6, pp. 503-516, Jun. 2007.

[34] C. G. and G. M., "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," U.S. Patent 4 042 779, 1977.

[35] C. Faller and F. Baumgarte, "Binaural cue coding - part ii: Schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, Nov. 2003.

[36] C. Faller, "Parametric multichannel audio coding: Synthesis of coherence cues," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 1, Jan. 2006.

[37] H. P. J. E. E. Schuijers, J. Breebaart, "Low complexity parametric stereo coding," in *Proc. of the 116th AES Convention*, Berlin, Germany, 2004.

[38] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Acoust. Soc. Am.*, vol. 45, no. 6, pp. 456-466, Jun. 1997.

[39] J. G. Tylka and E. Y. Choueiri, "Comparison of techniques for binaural navigation of higher- order Ambisonic sound fields," in *Proc. of the AES International Conference on Audio for Virtual and Augmented Reality*, New York, Sep. 2016.

YÊU CẦU BẢO HỘ

1. Thiết bị tạo mô tả trường âm thanh tăng cường, thiết bị bao gồm:

bộ tạo trường âm thanh để tạo ra ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một vị trí tham chiếu; và

bộ tạo siêu dữ liệu để tạo ra siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh,

trong đó ít nhất một mô tả trường âm thanh và siêu dữ liệu cấu thành mô tả trường âm thanh tăng cường.

trong đó bộ tạo trường âm thanh được tạo cấu hình để tạo mô tả mã hóa âm thanh định hướng (DirAC - Directional Audio Coding) của trường âm thanh, mô tả DirAC bao gồm một hoặc nhiều tín hiệu trộn giảm và dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau, và

trong đó bộ tạo siêu dữ liệu được tạo cấu hình để tạo ra thông tin chiều sâu hoặc vị trí riêng bổ sung cho các ngăn thời gian-tần số khác nhau như siêu dữ liệu.

2. Thiết bị theo điểm 1,

trong đó bộ tạo trường âm thanh được tạo cấu hình để tạo mô tả DirAC của trường âm thanh sao cho mô tả DirAC khuếch tán thêm dữ liệu cho các ngăn thời gian-tần số khác nhau.

3. Thiết bị theo điểm 1,

trong đó bộ tạo trường âm thanh được tạo cấu hình để ước lượng thông tin chiều sâu từ các tín hiệu âm thanh được sử dụng bởi bộ tạo trường âm thanh hoặc từ các tín hiệu video kết hợp với các tín hiệu âm thanh hoặc từ bản đồ chiều sâu được sử dụng trong ảnh/video lập thể (ba chiều) hoặc kỹ thuật trường ánh sáng hoặc từ thông tin hình học của cảnh đồ họa máy tính.

4. Thiết bị theo điểm 1,

trong đó bộ tạo siêu dữ liệu được tạo cấu hình để tạo ra, như thông tin chiều sâu hoặc vị trí riêng bổ sung, bản đồ chiều sâu bao gồm, đối với dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau, thông tin khoảng cách tương ứng cho các ngăn thời gian-tần số khác nhau.

5. Thiết bị theo điểm 1, trong đó thiết bị còn bao gồm giao diện đầu ra để tạo ra tín hiệu đầu ra để truyền hoặc lưu trữ, tín hiệu đầu ra bao gồm, đối với khung thời gian,

một hoặc nhiều tín hiệu âm thanh được suy ra từ trường âm thanh và thông tin chiều sâu hoặc vị trí riêng bổ sung cho các ngăn thời gian-tần số khác nhau.

6. Thiết bị theo điểm 1,

trong đó bộ tạo trường âm thanh được tạo cấu hình để suy ra dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau từ trường âm thanh, dữ liệu hướng cho các ngăn thời gian-tần số khác nhau liên quan đến hướng đến của âm thanh đối với ngăn thời gian-tần số của các ngăn thời gian-tần số khác nhau, và trong đó bộ tạo siêu dữ liệu được tạo cấu hình để suy ra thông tin chiều sâu hoặc vị trí riêng bổ sung cho các ngăn thời gian-tần số khác nhau như các mục dữ liệu kết hợp thông tin khoảng cách với dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau.

7. Thiết bị theo điểm 6,

trong đó bộ tạo trường âm thanh được tạo cấu hình để suy ra dữ liệu hướng riêng cho các ngăn tần số khác nhau trên mỗi khung thời gian của mô tả trường âm thanh,

trong đó bộ tạo siêu dữ liệu được tạo cấu hình để suy ra các mục dữ liệu kết hợp thông tin khoảng cách với dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau cho khung thời gian, và

trong đó giao diện đầu ra được tạo cấu hình để tạo ra tín hiệu đầu ra bao gồm mô tả trường âm thanh tăng cường sao cho các mục dữ liệu cho khung thời gian được nối với dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau.

8. Thiết bị theo một trong số các điểm nêu trên,

trong đó thông tin chiều sâu hoặc vị trí riêng bổ sung cho các ngăn thời gian-tần số khác nhau là bản đồ chiều sâu gồm các mục dữ liệu hướng đến như dữ liệu riêng, và các khoảng cách kết hợp, sao cho từng mục dữ liệu hướng đến trong số các mục dữ liệu hướng đến có khoảng cách kết hợp.

9. Thiết bị theo điểm 1,

trong đó bộ tạo trường âm thanh được tạo cấu hình để tạo ra thông tin khuếch tán cho các ngăn tần số của khung thời gian của trường âm thanh, và

trong đó bộ tạo siêu dữ liệu được tạo cấu hình để chỉ tạo ra thông tin khoảng cách cho ngăn tần số, khi giá trị khuếch tán cho ngăn tần số thấp hơn ngưỡng khuếch tán, hoặc trong đó bộ tạo siêu dữ liệu được tạo cấu hình để chỉ tạo ra siêu dữ liệu

khoảng cách khác giá trị được định trước, khi giá trị khuếch tán cho ngăn tần số thấp hơn giá trị khuếch tán ngưỡng.

10. Thiết bị tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, thiết bị bao gồm:

bộ tính toán trường âm thanh để tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó thông tin bao gồm thông tin chiều sâu, và trong đó mô tả trường âm thanh bao gồm mô tả DirAC có dữ liệu hướng cho các ngăn thời gian-tần số khác nhau,

trong đó bộ tính toán trường âm thanh được tạo cấu hình

để tính toán dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau sử dụng dữ liệu hướng, thông tin chiều sâu và thông tin tịnh tiến, và

để kết xuất mô tả DirAC sử dụng dữ liệu hướng biến đổi cho mô tả âm thanh bao gồm các kênh âm thanh để truyền hoặc lưu trữ mô tả DirAC sử dụng dữ liệu hướng biến đổi cho các ngăn tần số khác nhau thay vì dữ liệu hướng cho các ngăn thời gian-tần số khác nhau như được chứa trong mô tả DirAC.

11. Thiết bị theo điểm 10,

trong đó bộ tính toán trường âm thanh được tạo cấu hình để truyền hoặc lưu trữ mô tả DirAc sử dụng thêm dữ liệu khuếch tán như được chứa bởi mô tả DirAC.

12. Thiết bị theo điểm 10, trong đó bộ tính toán trường âm thanh được tạo cấu hình để xác định, cho ngăn thời gian-tần số, để duy trì dữ liệu hướng hoặc để tính toán dữ liệu hướng biến đổi dựa trên dữ liệu khuếch tán cho ngăn thời gian tần số, trong đó dữ liệu hướng biến đổi chỉ được tính toán cho dữ liệu khuếch tán biểu thị độ khuếch tán thấp hơn mức khuếch tán được định trước hoặc mức khuếch tán thích ứng.

13. Thiết bị theo một điểm 10, trong đó thiết bị còn bao gồm:

giao diện tịnh tiến để cung cấp thông tin tịnh tiến hoặc thông tin xoay biểu thị sự xoay của người nghe dự định cho mô tả trường âm thanh biến đổi;

bộ cấp siêu dữ liệu để cấp siêu dữ liệu cho bộ tính toán trường âm thanh;

bộ cấp trường âm thanh để cấp mô tả trường âm thanh cho bộ tính toán trường âm thanh; và

giao diện đầu ra để xuất ra mô tả trường âm thanh biến đổi và siêu dữ liệu biến đổi, siêu dữ liệu biến đổi được suy ra từ siêu dữ liệu sử dụng thông tin tịnh tiến, hoặc để xuất ra các kênh loa phóng thanh, từng kênh loa phóng thanh liên quan đến vị trí loa phóng thanh được định trước, hoặc để xuất ra sự biểu diễn lập thể của mô tả trường âm thanh biến đổi.

14. Thiết bị theo điểm 10, trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để tính toán vectơ thứ nhất chỉ từ vị trí tham chiếu đến nguồn âm thanh thu được bởi sự phân tích của trường âm thanh;

để tính toán vectơ thứ hai chỉ từ vị trí tham chiếu khác đến nguồn âm thanh sử dụng vectơ thứ nhất và thông tin tịnh tiến, thông tin tịnh tiến xác định vectơ tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác ; và

để tính toán giá trị biến đổi khoảng cách sử dụng vị trí tham chiếu khác, vị trí của nguồn âm thanh, và vectơ thứ hai, hoặc sử dụng khoảng cách từ vị trí tham chiếu khác đến vị trí của nguồn âm thanh và vectơ thứ hai.

15. Thiết bị theo điểm 10,

trong đó dữ liệu hướng cho các ngăn thời gian-tần số khác nhau bao gồm dữ liệu hướng đến, và trong đó dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau bao gồm dữ liệu hướng đến biến đổi,

trong đó bộ tính toán trường âm thanh được tạo cấu hình để nhận, ngoài thông tin tịnh tiến, thông tin xoay, và

trong đó bộ tính toán trường âm thanh được tạo cấu hình để thực hiện sự biến đổi xoay để xoay dữ liệu hướng đến biến đổi cho mô tả trường âm thanh sử dụng thông tin xoay, trong đó dữ liệu hướng đến biến đổi được suy ra từ dữ liệu hướng đến thu được bởi sự phân tích trường âm thanh của mô tả trường âm thanh và sử dụng thông tin tịnh tiến.

16. Thiết bị theo điểm 10, trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để xác định các nguồn từ mô tả trường âm thanh và các hướng cho các nguồn bởi việc phân tích trường âm thanh;

để xác định, đối với từng nguồn, khoảng cách của nguồn từ vị trí tham chiếu sử dụng siêu dữ liệu;

để xác định hướng mới của nguồn liên quan đến vị trí tham chiếu khác sử dụng hướng cho nguồn và thông tin tịnh tiến;

để xác định thông tin khoảng cách mới cho nguồn liên quan đến vị trí tham chiếu khác; và

để tạo trường âm thanh biến đổi sử dụng hướng mới của nguồn, thông tin khoảng cách mới, và mô tả trường âm thanh hoặc các tín hiệu nguồn tương ứng với các nguồn được suy ra từ mô tả trường âm thanh.

17. Thiết bị tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, thiết bị bao gồm:

bộ tính toán trường âm thanh để tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó mô tả trường âm thanh bao gồm các thành phần trường âm thanh, các thành phần trường âm thanh bao gồm thành phần đẳng hướng và ít nhất một thành phần định hướng,

trong đó bộ tính toán trường âm thanh bao gồm:

bộ phân tích trường âm thanh để phân tích các thành phần trường âm thanh để suy ra, đối với các ngăn tần số khác nhau, thông tin hướng đến;

bộ biến đổi tịnh tiến để tính toán thông tin hướng đến biến đổi trên mỗi ngăn tần số sử dụng thông tin hướng đến và siêu dữ liệu, siêu dữ liệu bao gồm bản đồ chiều sâu kết hợp thông tin khoảng cách với thông tin hướng đến cho ngăn tần số; và

bộ bù khoảng cách để tính toán mô tả trường âm thanh biến đổi sử dụng thông tin bù khoảng cách phụ thuộc vào thông tin khoảng cách được cung cấp bởi bản đồ chiều sâu cho ngăn tần số, và phụ thuộc vào khoảng cách mới kết hợp với ngăn tần số có liên quan đến thông tin hướng đến biến đổi.

18. Thiết bị tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, thiết bị bao gồm:

bộ tính toán trường âm thanh để tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để tính toán vector thứ nhất chỉ từ vị trí tham chiếu đến vị trí của nguồn âm thanh thu được bởi sự phân tích mô tả trường âm thanh;

để tính toán vector thứ hai chỉ từ vị trí tham chiếu khác đến vị trí của nguồn âm thanh bằng cách sử dụng vector thứ nhất và thông tin tịnh tiến, thông tin tịnh tiến xác định vector tịnh tiến từ vị trí tham chiếu sang vị trí tham chiếu khác; và

để tính toán giá trị biến đổi khoảng cách bằng cách sử dụng vị trí tham chiếu khác, vị trí của nguồn âm thanh, và vector thứ hai, hoặc sử dụng khoảng cách từ vị trí tham chiếu khác đến vị trí của nguồn âm thanh và vector thứ hai,

trong đó vector thứ nhất được tính toán bằng cách nhân vector đơn vị hướng đến với khoảng cách được chứa bởi siêu dữ liệu, hoặc

trong đó vector thứ hai được tính toán bằng cách trừ vector tịnh tiến khỏi vector thứ nhất, hoặc

trong đó giá trị biến đổi khoảng cách được tính toán bằng cách chia vector thứ hai cho mức chuẩn của vector thứ nhất.

19. Thiết bị tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, thiết bị bao gồm:

bộ tính toán trường âm thanh để tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó thông tin không gian bao gồm thông tin chiều sâu;

trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để xác định các tín hiệu nguồn từ mô tả trường âm thanh và các hướng của các tín hiệu nguồn liên quan đến vị trí tham chiếu bởi sự phân tích âm thanh;

để tính toán các hướng mới của các tín hiệu nguồn liên quan đến vị trí tham chiếu khác sử dụng thông tin chiều sâu;

để tính toán thông tin khoảng cách cho các nguồn âm thanh kết hợp với các tín hiệu nguồn liên quan đến vị trí tham chiếu khác; và

để thu được mô tả trường âm thanh biến đổi được tổng hợp sử dụng thông tin khoảng cách, các tín hiệu nguồn và các hướng mới.

20. Thiết bị theo điểm 19, trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để thu được mô tả trường âm thanh biến đổi được tổng hợp bằng cách quét tín hiệu nguồn trong số các tín hiệu nguồn với hướng được cho bởi hướng mới liên quan đến cài đặt phát lại, và

bằng cách định tỷ lệ các tín hiệu nguồn trong số các tín hiệu nguồn sử dụng thông tin khoảng cách trước khi thực hiện quét hoặc sau khi thực hiện quét.

21. Thiết bị theo điểm 19,

trong đó bộ tính toán trường âm thanh được tạo cấu hình để cộng tín hiệu khuếch tán với phần hướng của tín hiệu nguồn trong số các tín hiệu nguồn, phần hướng được biến đổi bởi thông tin khoảng cách trước khi được cộng vào tín hiệu khuếch tán.

22. Thiết bị theo điểm 19,

trong đó thông tin không gian bao gồm thông tin chiều sâu cho các ngăn thời gian-tần số khác nhau, các ngăn thời gian-tần số khác nhau bao gồm các ngăn tần số khác nhau của khung thời gian,

trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để thực hiện, trong việc phân tích âm thanh, chuyển đổi thời gian-tần số của mô tả trường âm thanh và để tính toán các mục dữ liệu hướng đến cho các ngăn thời gian-tần số khác nhau của khung thời gian, các mục dữ liệu hướng đến là mô tả DirAC của mô tả trường âm thanh;

để tính toán hướng mới cho từng ngăn tần số trong số các ngăn tần số khác nhau của khung thời gian sử dụng thông tin chiều sâu cho ngăn tần số tương ứng,

để tính toán thông tin khoảng cách cho từng ngăn tần số trong số các ngăn tần số khác nhau của khung thời gian, và

để thực hiện tổng hợp hướng cho từng ngăn tần số trong số các ngăn tần số khác nhau của khung thời gian sử dụng tín hiệu âm thanh cho ngăn tần số tương ứng, độ khuếch đại quét cho ngăn tần số tương ứng được suy ra từ hướng mới cho ngăn tần số tương ứng, và vectơ định tỷ lệ cho ngăn tần số tương ứng được suy ra từ thông tin khoảng cách cho ngăn tần số tương ứng.

23. Thiết bị theo điểm 22, trong đó bộ tính toán trường âm thanh được tạo cấu hình:

để thực hiện sự tổng hợp khuếch tán sử dụng tín hiệu âm thanh khuếch tán được suy ra từ tín hiệu âm thanh cho ngăn tần số tương ứng và sử dụng tham số khuếch tán được suy ra bởi sự phân tích âm thanh cho ngăn tần số tương ứng và để tổ hợp tín hiệu

âm thanh trực tiếp và tín hiệu âm thanh khuếch tán để thu được tín hiệu âm thanh được tổng hợp cho ngăn tần số tương ứng; và

để thực hiện chuyển đổi tần số-thời gian bằng cách sử dụng các tín hiệu âm thanh được tổng hợp cho các ngăn tần số cho khung thời gian để thu được tín hiệu âm thanh được tổng hợp miền thời gian như mô tả trường âm thanh biến đổi.

24. Thiết bị theo điểm 19, trong đó bộ tính toán trường âm thanh được tạo cấu hình để tổng hợp, đối với từng nguồn âm thanh, trường âm thanh liên quan đến vị trí tham chiếu khác, trong đó sự tổng hợp bao gồm:

đối với từng nguồn, xử lý tín hiệu nguồn sử dụng hướng mới cho tín hiệu nguồn để thu được mô tả trường âm thanh của tín hiệu nguồn liên quan đến vị trí tham chiếu khác;

biến đổi tín hiệu nguồn trước khi xử lý tín hiệu nguồn hoặc biến đổi mô tả trường âm thanh sử dụng thông tin hướng, và

bổ sung sự mô tả trường âm thanh cho các nguồn để thu được trường âm thanh biến đổi liên quan đến vị trí tham chiếu khác.

25. Thiết bị theo điểm 19,

trong đó sự phân tích âm thanh được tạo cấu hình để thu được tín hiệu khuếch tán ngoài tín hiệu nguồn, và

trong đó bộ tính toán trường âm thanh được tạo cấu hình để bổ sung tín hiệu khuếch tán vào phần hướng được tính toán bằng cách sử dụng hướng mới và thông tin khoảng cách.

26. Thiết bị theo điểm 25,

trong đó sự phân tích âm thanh được tạo cấu hình để xác định các tín hiệu nguồn bởi thuật toán tách nguồn và để trừ ít nhất một số tín hiệu nguồn khỏi mô tả trường âm thanh để thu được tín hiệu khuếch tán.

27. Phương pháp tạo mô tả trường âm thanh tăng cường, phương pháp bao gồm các bước:

tạo ra ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một vị trí tham chiếu; và

tạo ra siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh,

trong đó ít nhất một mô tả trường âm thanh và siêu dữ liệu cấu thành mô tả trường âm thanh tăng cường,

trong đó bước tạo ra ít nhất một trường âm thanh bao gồm tạo ra mô tả DirAC của trường âm thanh, mô tả DirAC có một hoặc nhiều tín hiệu trộn giảm và dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau, và trong đó bước tạo ra siêu dữ liệu bao gồm tạo ra thông tin chiều sâu riêng bổ sung cho các ngăn thời gian-tần số khác nhau như siêu dữ liệu.

28. Phương pháp tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, phương pháp bao gồm:

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó thông tin không gian bao gồm thông tin chiều sâu, và trong đó mô tả trường âm thanh bao gồm mô tả DirAC có dữ liệu hướng cho các ngăn thời gian-tần số khác nhau; và

trong đó bước tính toán bao gồm:

tính toán dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau sử dụng dữ liệu hướng, thông tin chiều sâu và thông tin tịnh tiến, và

kết xuất mô tả DirAC sử dụng dữ liệu hướng biến đổi thành mô tả âm thanh bao gồm các kênh âm thanh hoặc truyền hoặc lưu trữ mô tả DirAC sử dụng dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau thay vì dữ liệu hướng cho các ngăn thời gian tần số khác nhau như được chứa bởi mô tả DirAC.

29. Vật ghi lưu trữ số không chuyển tiếp có chương trình máy tính được lưu trên đó để thực hiện, khi chương trình máy tính đã nêu được chạy bởi máy tính, phương pháp tạo mô tả trường âm thanh tăng cường, phương pháp bao gồm bước:

tạo ra ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một vị trí tham chiếu; và

tạo ra siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh,

trong đó ít nhất một mô tả trường âm thanh và siêu dữ liệu cấu thành mô tả trường âm thanh tăng cường,

trong đó bước tạo ra ít nhất một trường âm thanh bao gồm tạo ra mô tả DirAC của trường âm thanh, mô tả DirAC có một hoặc nhiều tín hiệu trộn giảm và dữ liệu hướng riêng cho các ngăn thời gian-tần số khác nhau, và trong đó bước tạo ra siêu dữ

liệu bao gồm tạo ra thông tin chiều sâu riêng bổ sung cho các ngăn thời gian-tần số khác nhau như siêu dữ liệu.

30. Vật ghi lưu trữ số không chuyển tiếp có chương trình máy tính được lưu trên đó để thực hiện, khi chương trình máy tính đã nêu được chạy bởi máy tính, phương pháp tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, phương pháp bao gồm bước:

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó thông tin không gian bao gồm thông tin chiều sâu, và trong đó mô tả trường âm thanh bao gồm mô tả DirAC có dữ liệu hướng cho các ngăn thời gian-tần số khác nhau; và

trong đó bước tính toán bao gồm:

tính toán dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau sử dụng dữ liệu hướng, thông tin chiều sâu và thông tin tịnh tiến, và

kết xuất mô tả DirAC sử dụng dữ liệu hướng biến đổi thành mô tả âm thanh bao gồm các kênh âm thanh hoặc truyền hoặc lưu trữ mô tả DirAC sử dụng dữ liệu hướng biến đổi cho các ngăn thời gian-tần số khác nhau thay vì dữ liệu hướng cho các ngăn thời gian tần số khác nhau như được chứa bởi mô tả DirAC.

31. Phương pháp tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, phương pháp bao gồm các bước:

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó mô tả trường âm thanh bao gồm các thành phần trường âm thanh, các thành phần trường âm thanh bao gồm thành phần đẳng hướng và ít nhất một thành phần định hướng,

trong đó bước tính toán bao gồm:

phân tích các thành phần trường âm thanh để suy ra, đối với các ngăn tần số khác nhau, thông tin hướng đến;

tính toán thông tin hướng đến biến đổi trên mỗi ngăn tần số sử dụng thông tin hướng đến và siêu dữ liệu, siêu dữ liệu bao gồm bản đồ chiều sâu kết hợp thông tin khoảng cách với thông tin hướng đến cho ngăn tần số; và

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin bù khoảng cách phụ thuộc vào thông tin khoảng cách được cung cấp bởi bản đồ chiều sâu cho ngăn tần số và phụ thuộc vào khoảng cách mới kết hợp với ngăn tần số có liên quan đến thông tin hướng đến biến đổi.

32. Phương pháp tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, phương pháp bao gồm các bước:

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó bước tính toán bao gồm:

tính toán vector thứ nhất chỉ từ vị trí tham chiếu đến vị trí của nguồn âm thanh thu được bởi sự phân tích mô tả trường âm thanh;

tính toán vector thứ hai chỉ từ vị trí tham chiếu khác đến vị trí của nguồn âm thanh bằng cách sử dụng vector thứ nhất và thông tin tịnh tiến, thông tin tịnh tiến xác định vector tịnh tiến từ vị trí tham chiếu sang vị trí tham chiếu khác; và

tính toán giá trị biến đổi khoảng cách bằng cách sử dụng vị trí tham chiếu khác, vị trí của nguồn âm thanh, và vector thứ hai, hoặc sử dụng khoảng cách từ vị trí tham chiếu khác đến vị trí của nguồn âm thanh và vector thứ hai,

trong đó vector thứ nhất được tính toán bằng cách nhân vector đơn vị hướng đến với khoảng cách được chứa trong siêu dữ liệu, hoặc

trong đó vector thứ hai được tính toán bằng cách trừ vector tịnh tiến khỏi vector thứ nhất, hoặc

trong đó giá trị biến đổi khoảng cách được tính toán bằng cách chia vector thứ hai cho mức chuẩn của vector thứ nhất.

33. Phương pháp tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, phương pháp bao gồm bước:

tính toán mô tả trường âm thanh biến đổi sử dụng thông tin không gian, mô tả trường âm thanh và thông tin tịnh tiến biểu thị sự tịnh tiến từ vị trí tham chiếu đến vị trí tham chiếu khác,

trong đó bước tính toán bao gồm:

xác định các tín hiệu nguồn từ mô tả trường âm thanh và các hướng của các tín hiệu nguồn liên quan đến vị trí tham chiếu bởi sự phân tích âm thanh;

tính toán các hướng mới của tín hiệu nguồn liên quan đến vị trí tham chiếu khác sử dụng thông tin chiều sâu;

tính toán thông tin khoảng cách cho các nguồn âm thanh kết hợp với các tín hiệu nguồn liên quan đến vị trí tham chiếu khác; và

thu được mô tả trường âm thanh biến đổi được tổng hợp sử dụng thông tin khoảng cách, các tín hiệu nguồn và các hướng mới.

34. Vật ghi lưu trữ số không chuyển tiếp có chương trình máy tính được lưu trên đó để thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 31 đến 33 khi chương trình máy tính đã nêu được chạy bởi máy tính.

1/17

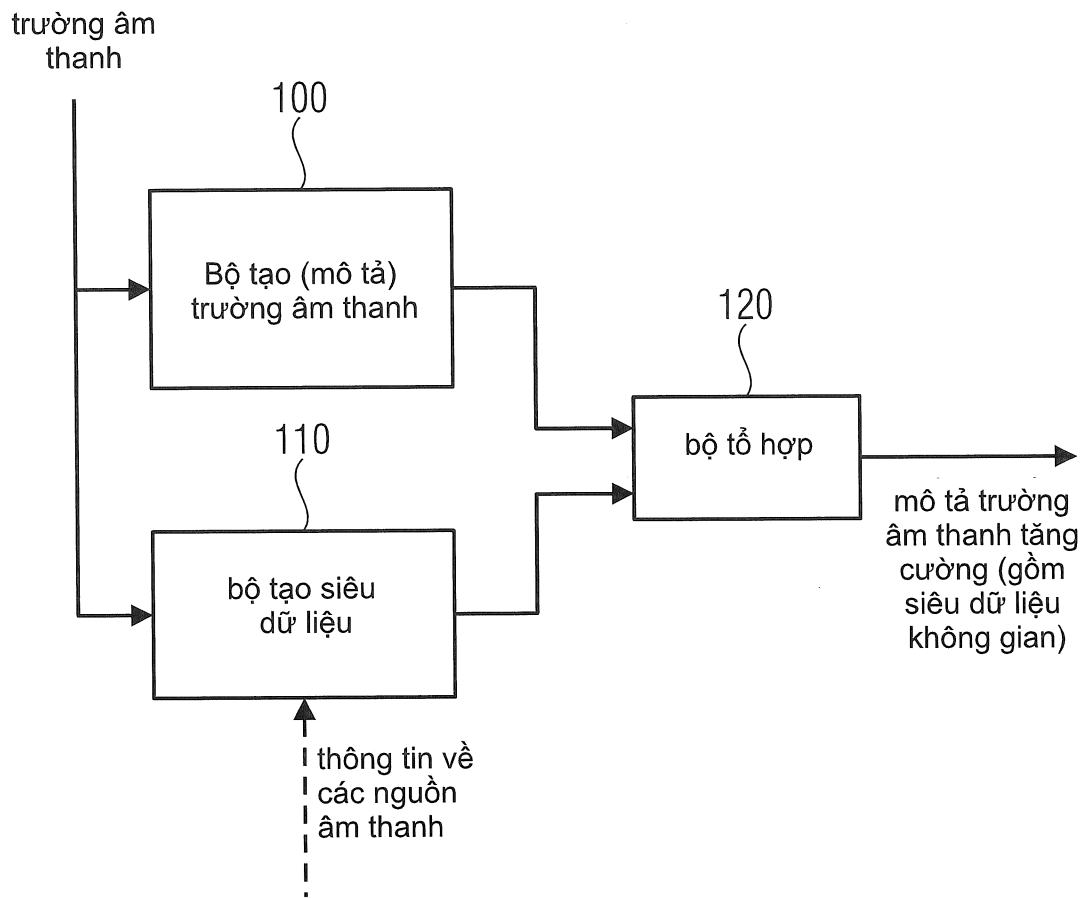


Fig. 1a

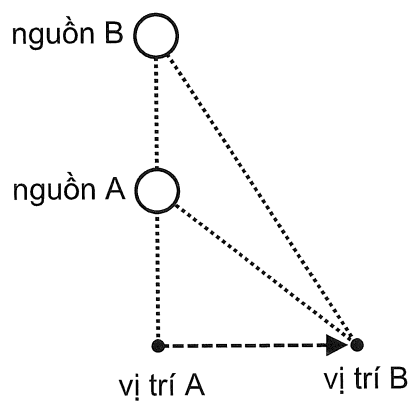


Fig. 1b

2/17

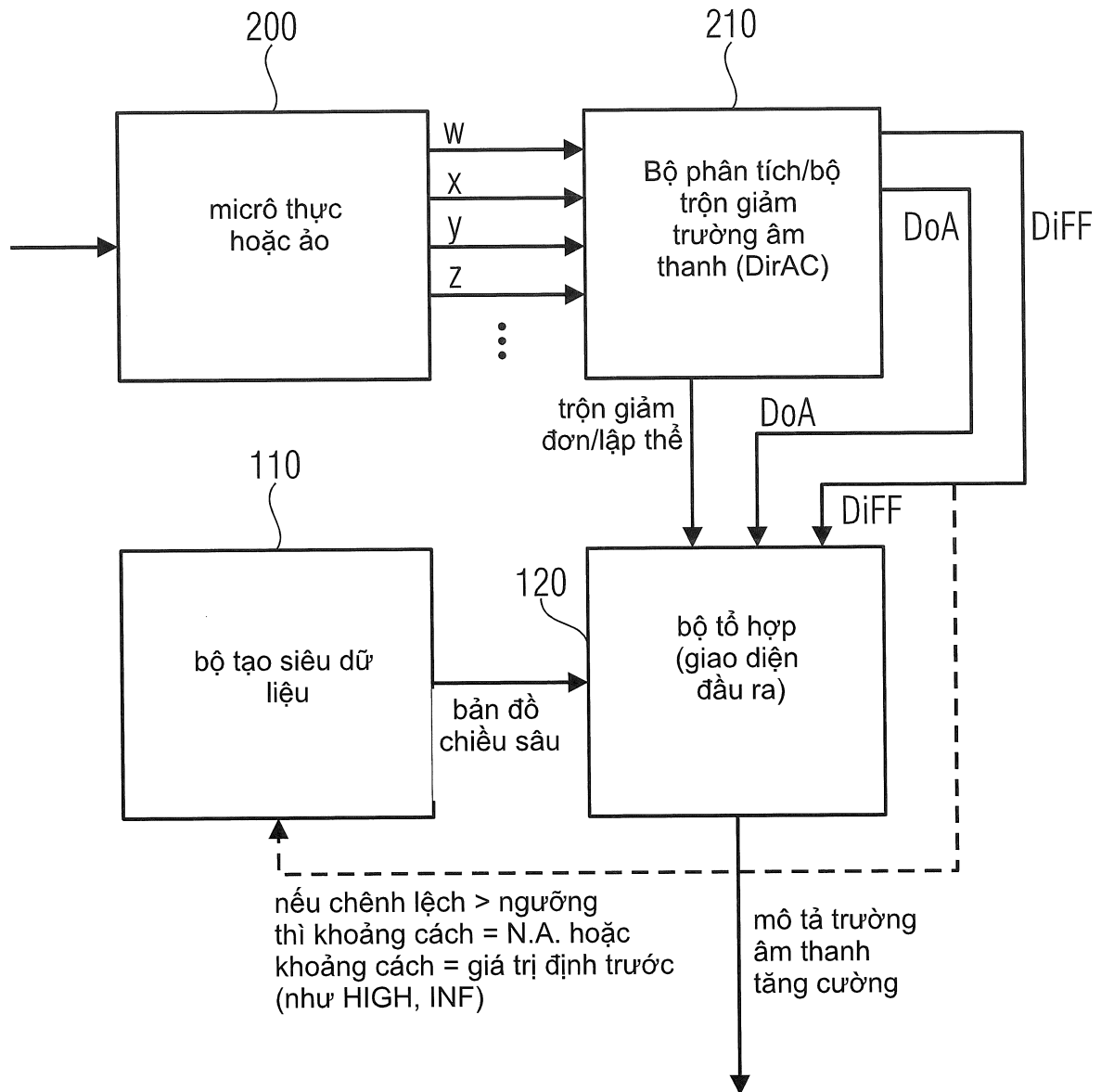


Fig. 2

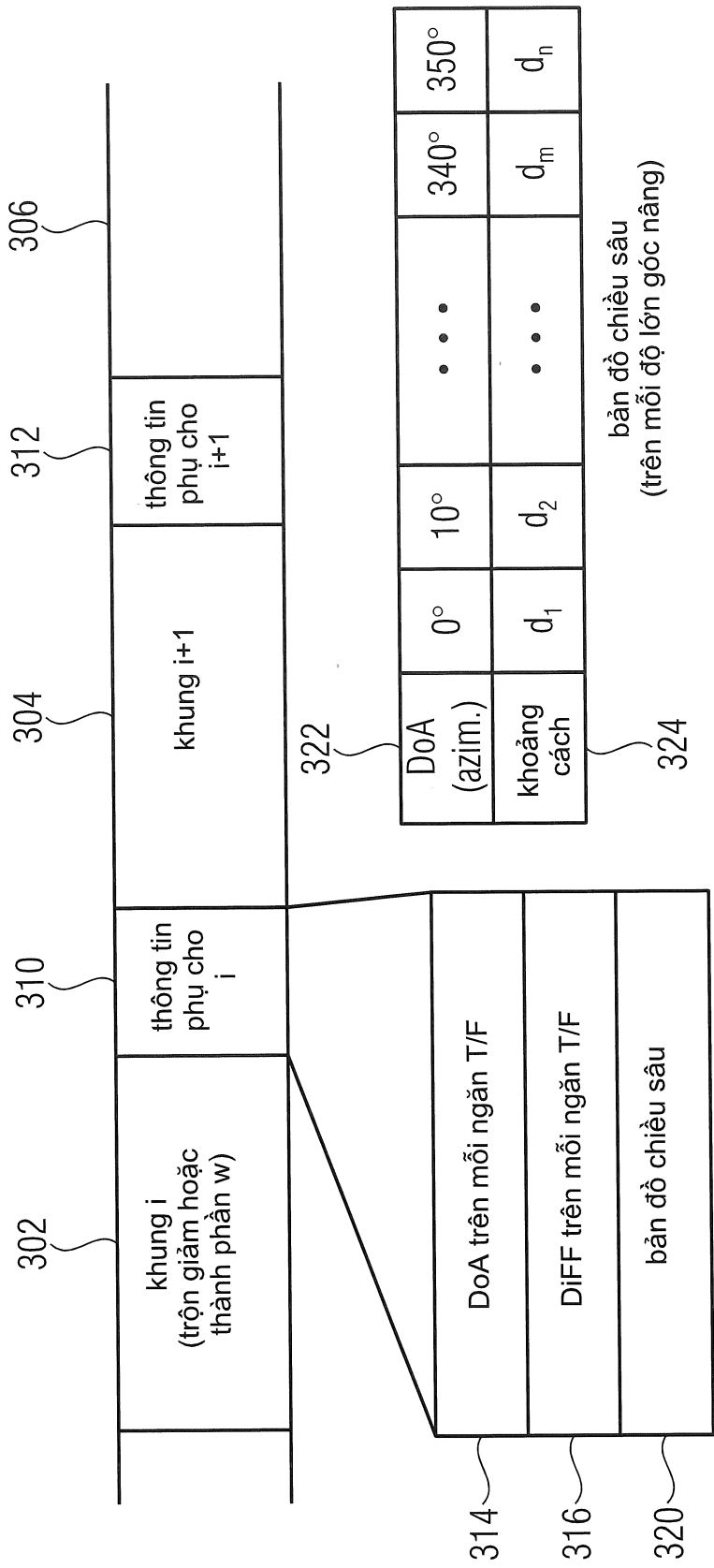


Fig. 3a

4/17

	326	320a	320b
	định dạng B hoặc sự biểu diễn cấp cao hơn cho các khoảng thời gian ví dụ i và $i+1$	bản đồ chiều sâu trên mỗi khoảng thời gian i	bản đồ chiều sâu cho khoảng thời gian $i+1$

Fig. 3b

	định dạng B hoặc sự biểu diễn cấp cao hơn cho khoảng thời gian i			định dạng B hoặc sự biểu diễn cấp cao hơn cho khoảng thời gian $i+1$
	bản đồ chiều sâu $i^{(1)}$	bản đồ chiều sâu $i^{(2)}$	...	bản đồ chiều sâu $i^{(Ni)}$

Fig. 3c

	định dạng B hoặc sự biểu diễn cấp cao hơn cho khoảng thời gian i	định dạng B hoặc sự biểu diễn cấp cao hơn cho khoảng thời gian $i+1$	
	bản đồ chiều sâu cho khoảng thời gian i	bản đồ chiều sâu cho khoảng thời gian $i+1$	

Fig. 3d

5/17

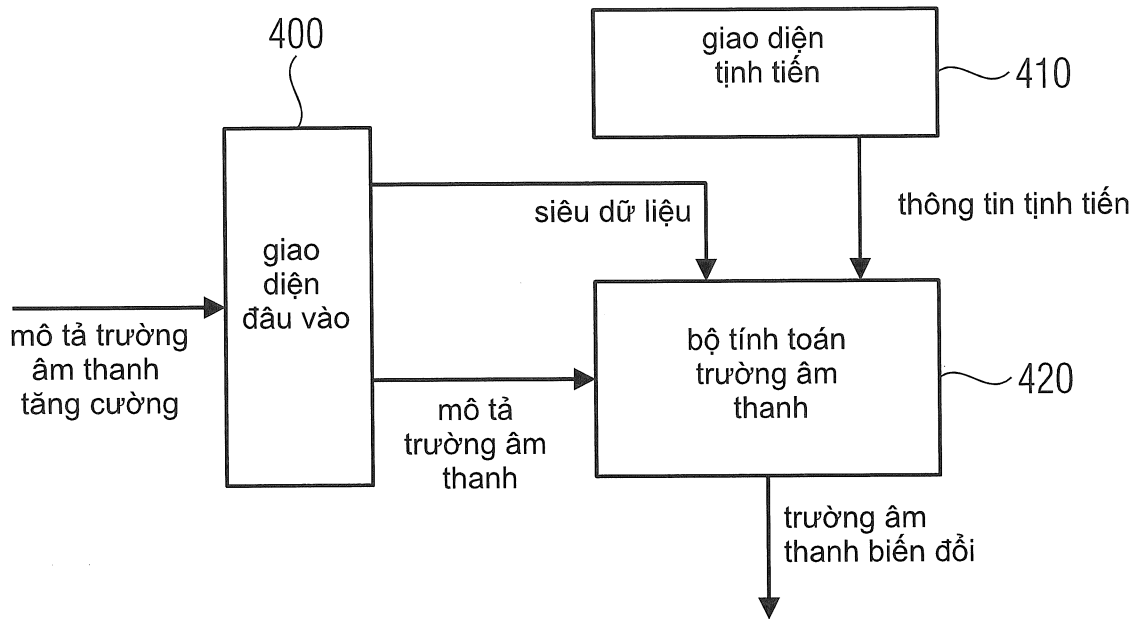


Fig. 4a

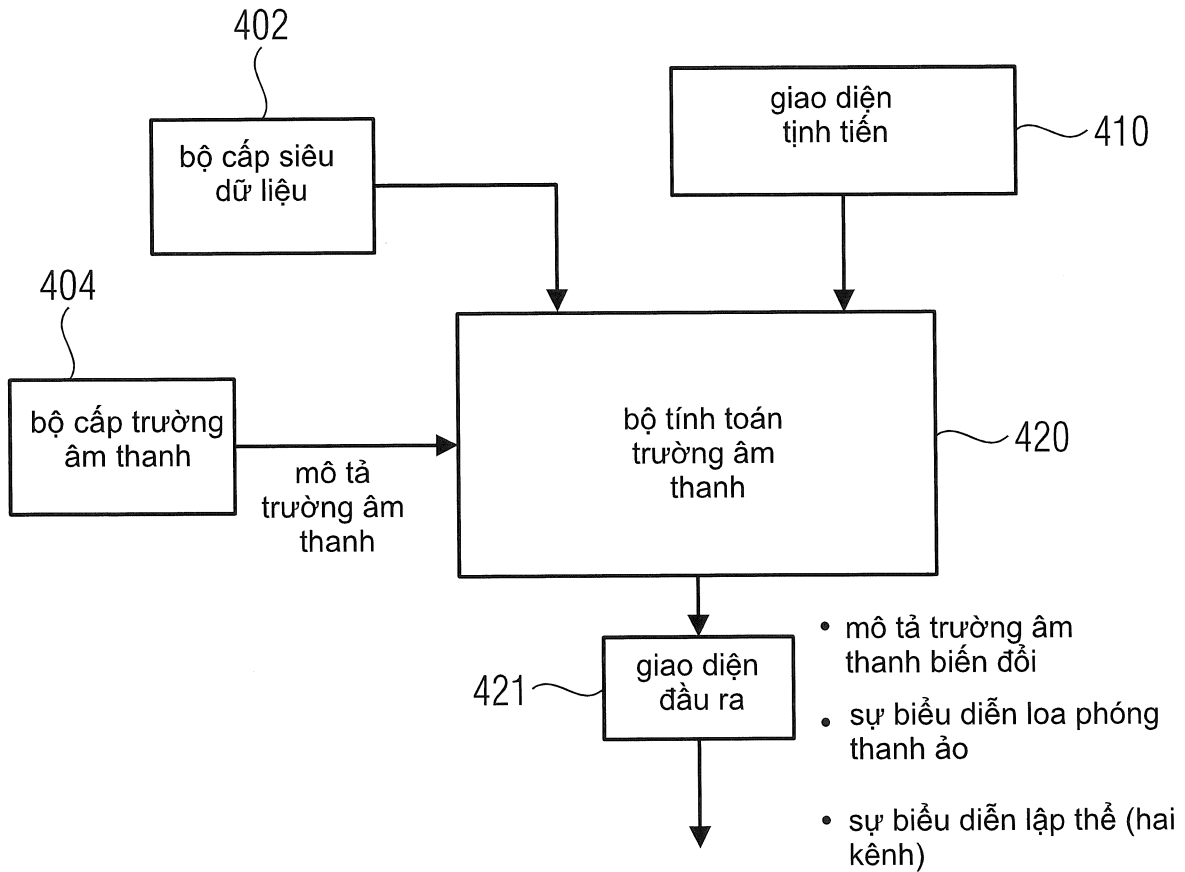


Fig. 4b

6/17

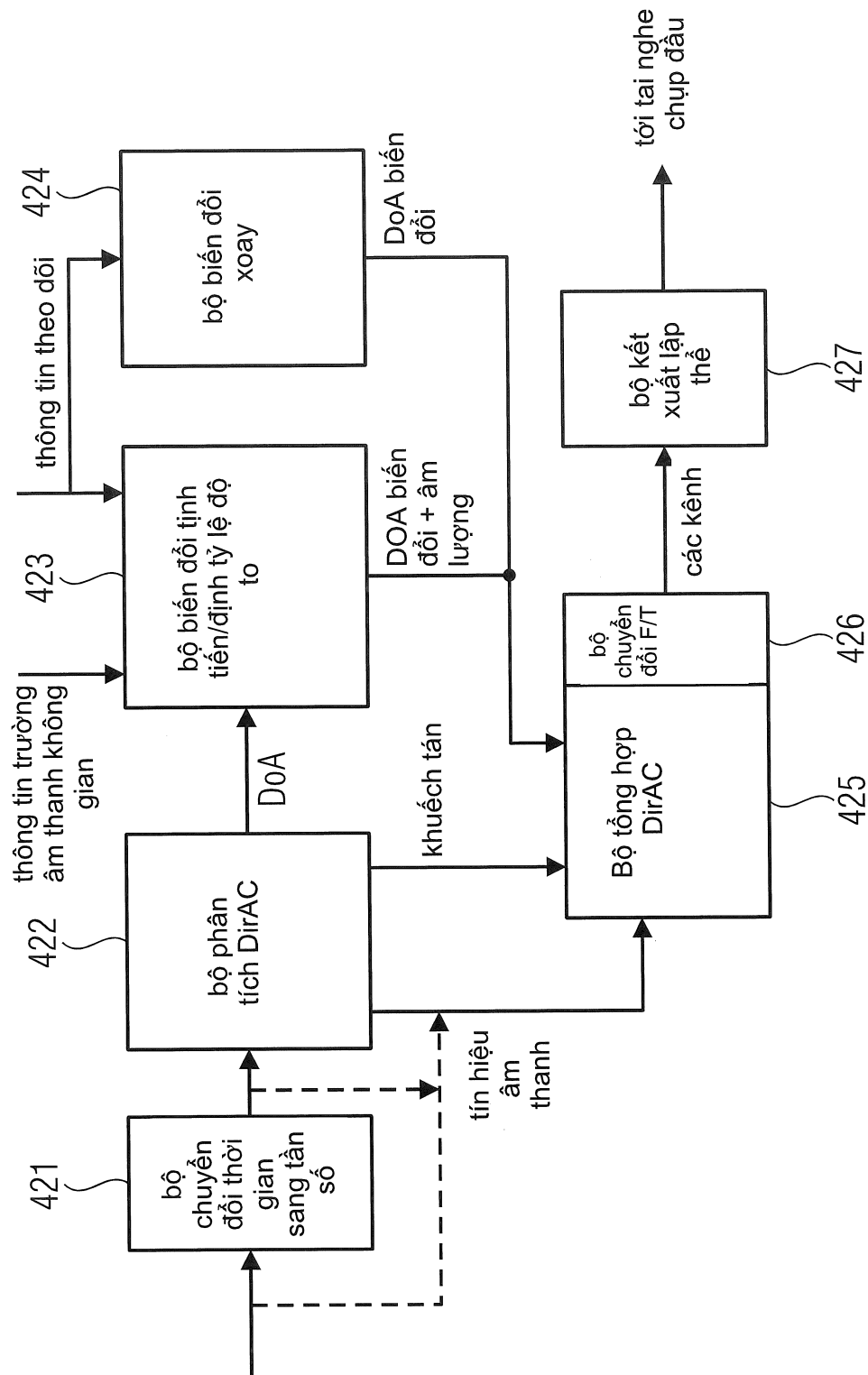
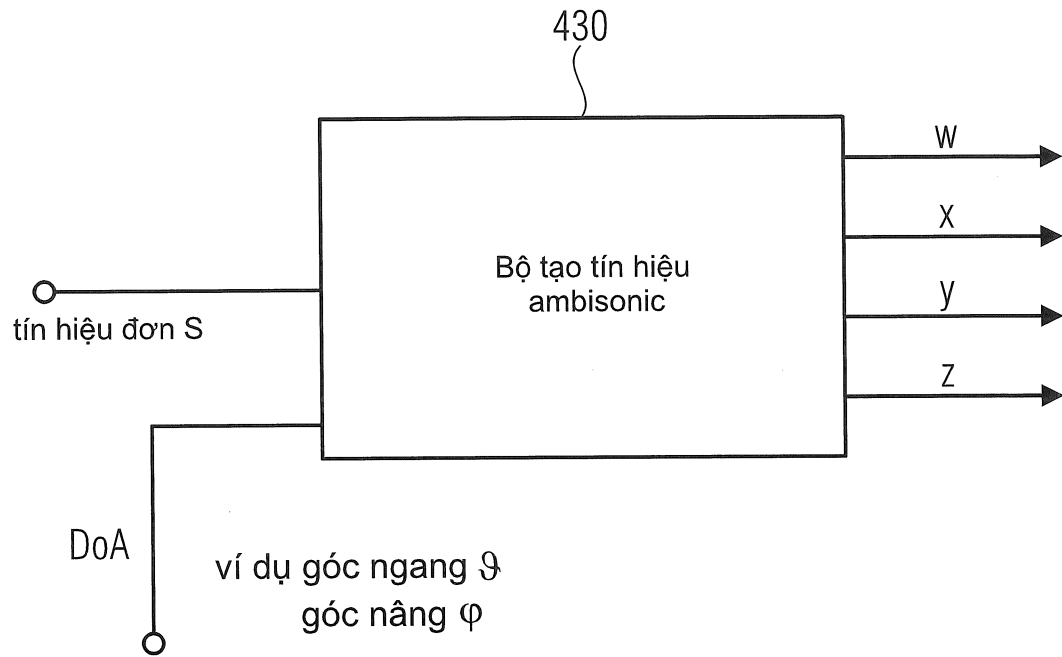


Fig. 4c

7/17



thành phần đẳng hướng $W = S \cdot \sqrt{\frac{1}{2}}$

$$\left\{ \begin{array}{l} X = S \cdot \cos \vartheta \cdot \cos \varphi \\ Y = S \cdot \sin \vartheta \cdot \cos \varphi \\ Z = S \cdot \sin \varphi \end{array} \right.$$

các thành phần hướng

Fig. 4d

8/17

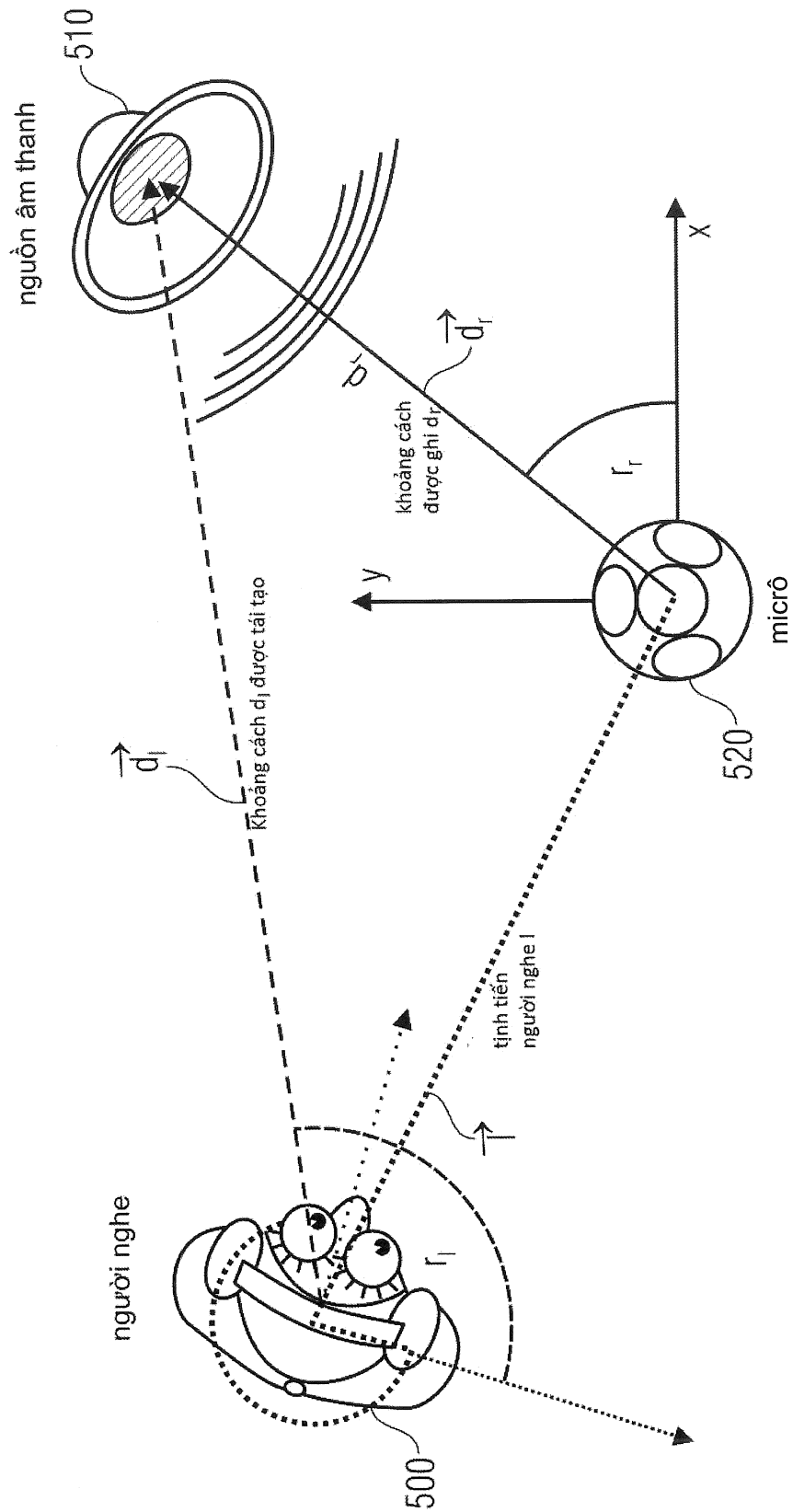


Fig. 5

9/17

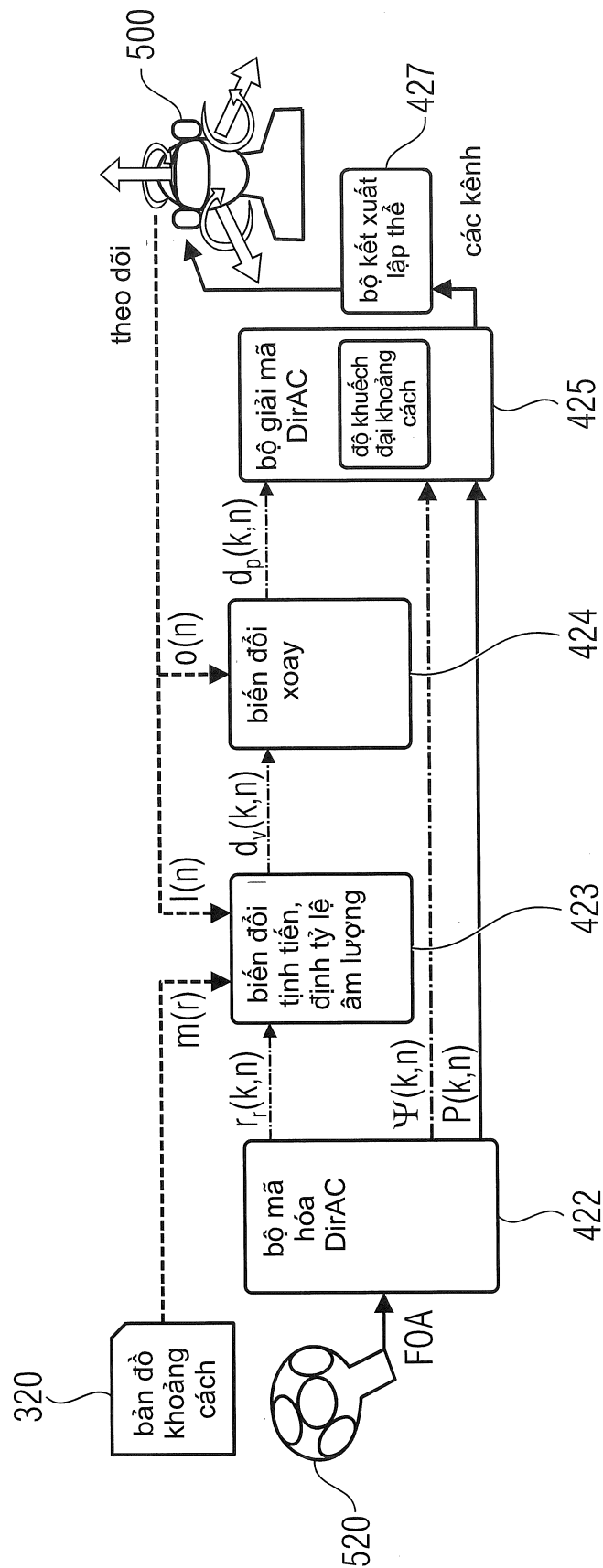


Fig. 6

10/17

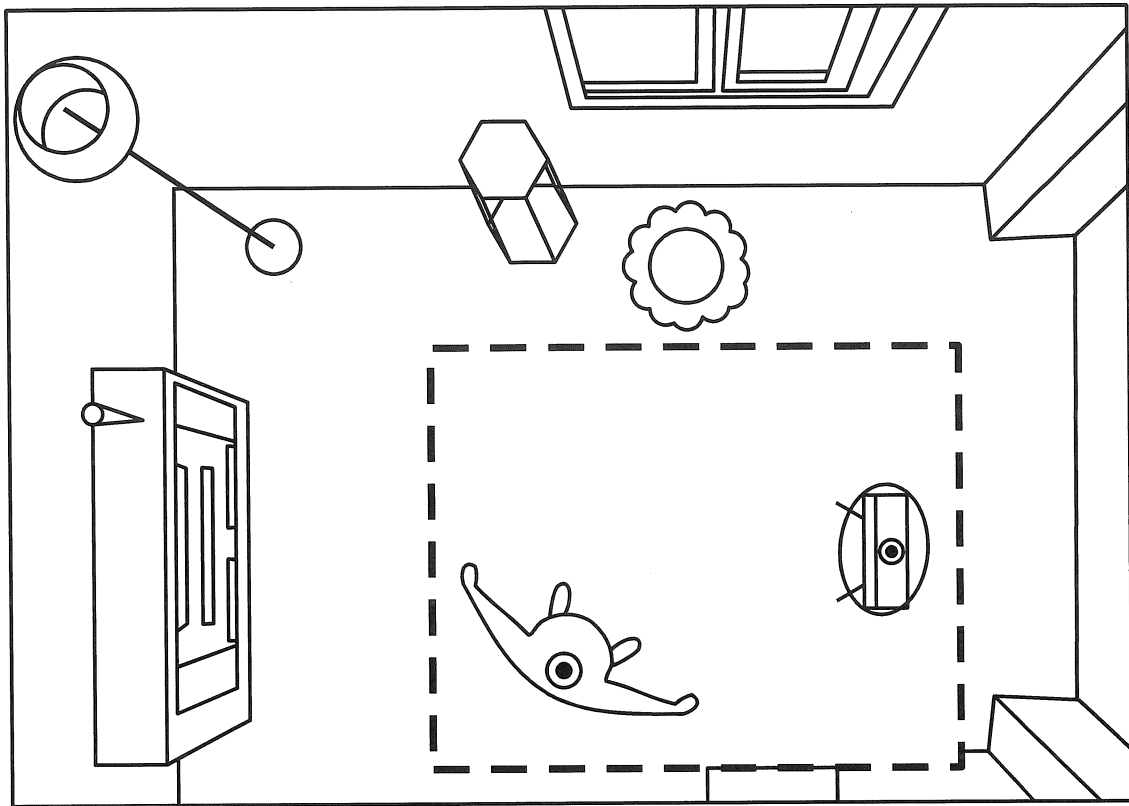


Fig. 7

MUSHRA-VR		BẢNG CHỌN				
Tái tạo không gian						
Nội dung ambisonic cấp một được tái tạo sử dụng các phương pháp kết xuất khác nhau	tuyệt vời					
	tốt					
	hợp lý					
	kém					
	tệ					
Đi bộ xung quanh và đánh giá sự tái tạo						
Mục hiện tại:	1	100	100	100	100	
Điều kiện được chọn:	REF					
		REF	A	B	C	D

Fig. 8

11/17

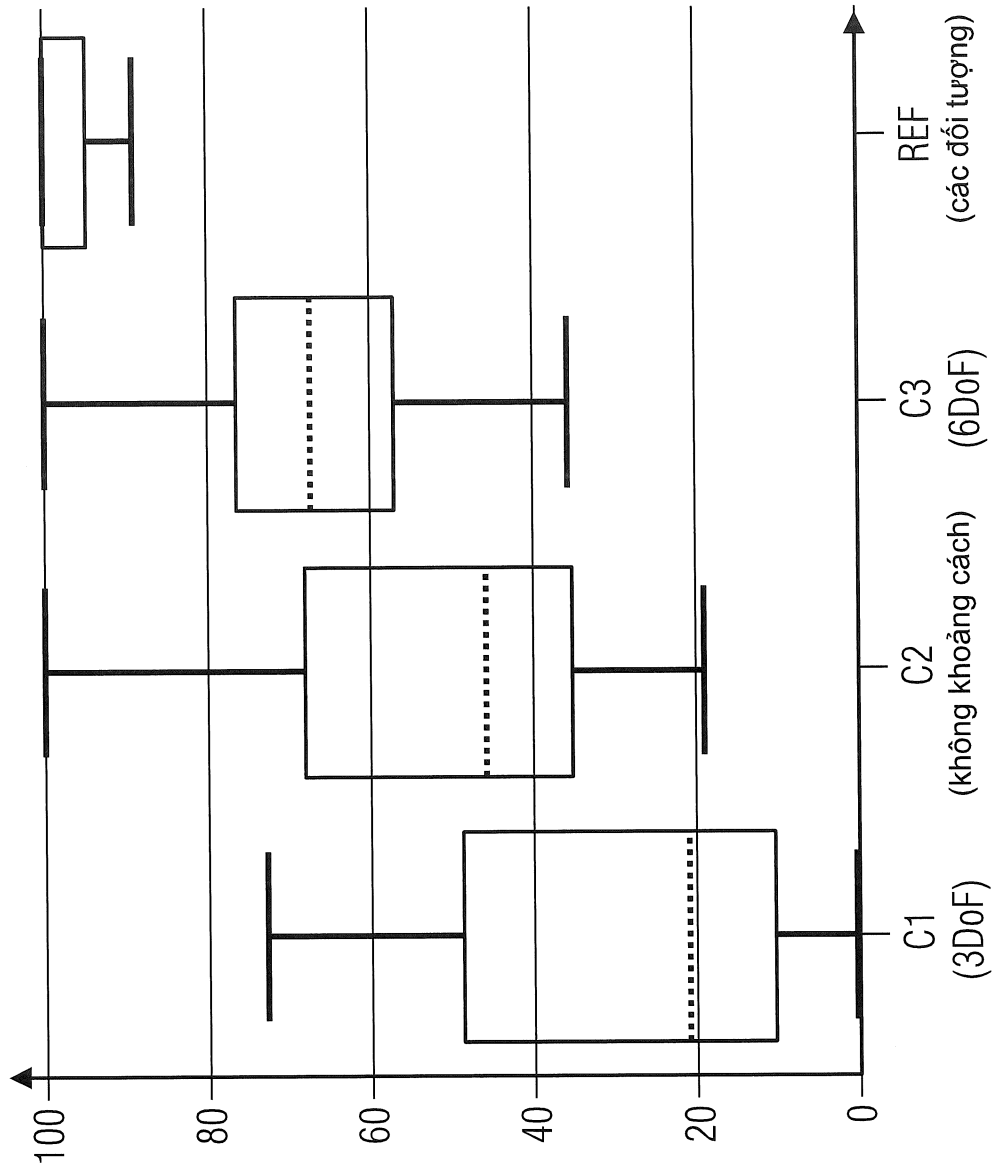


Fig. 9

12/17

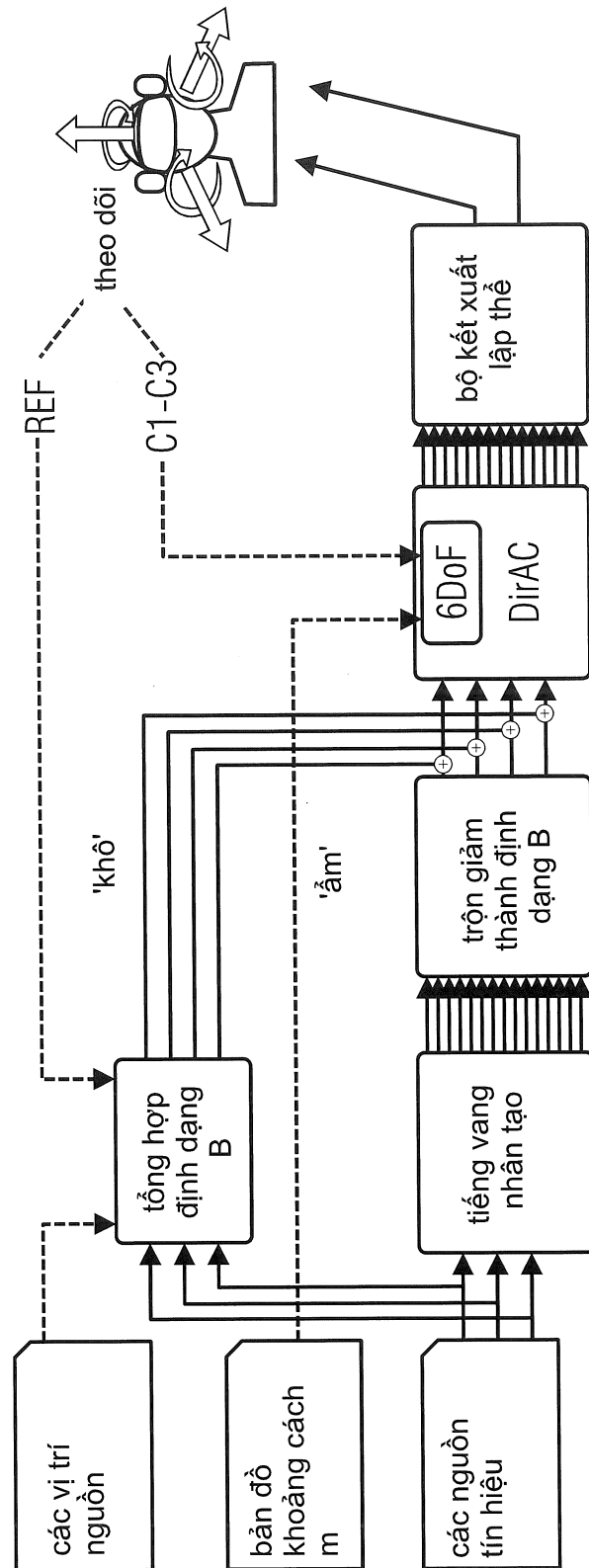


Fig. 10

13/17

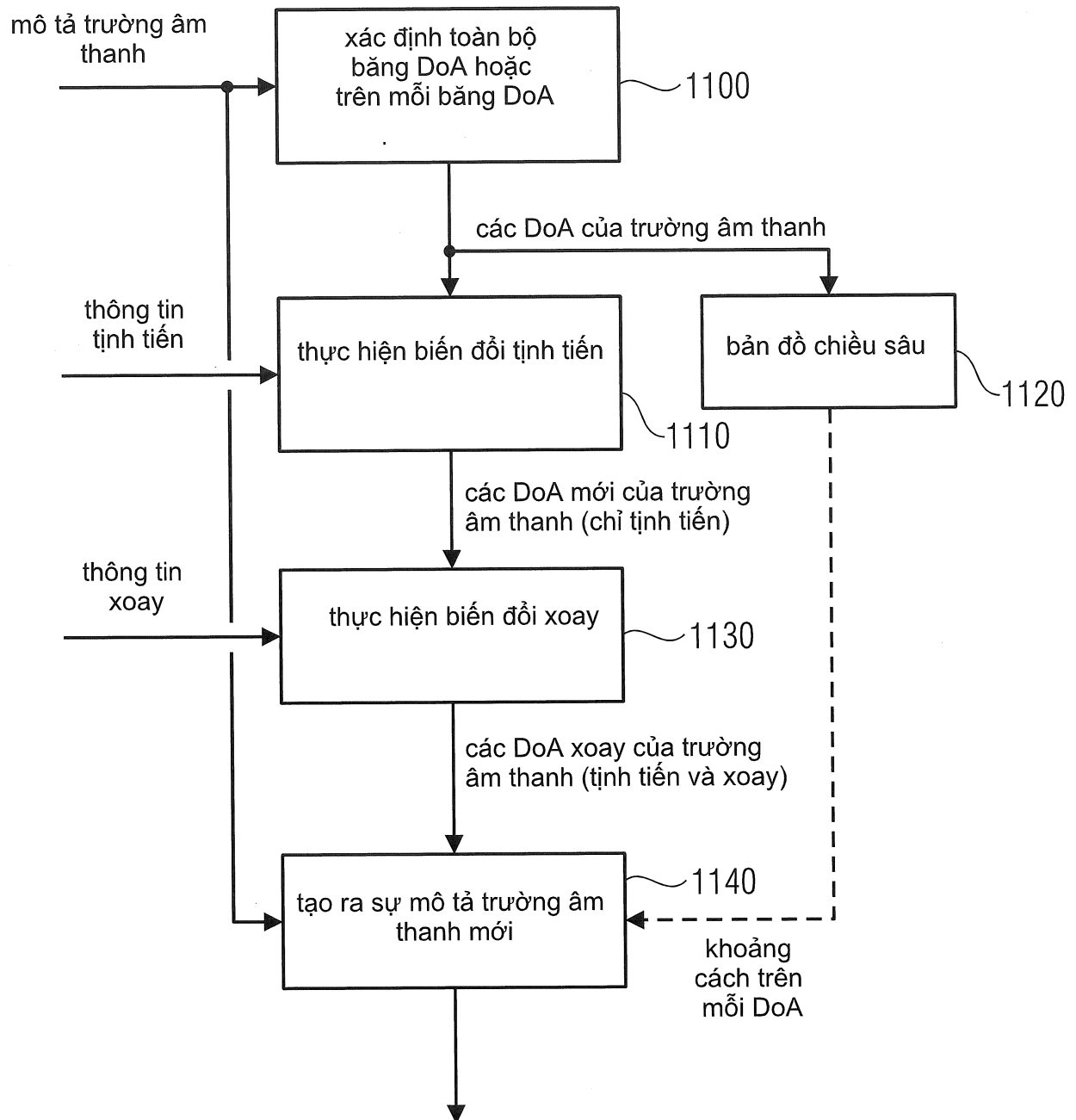


Fig. 11a

14/17

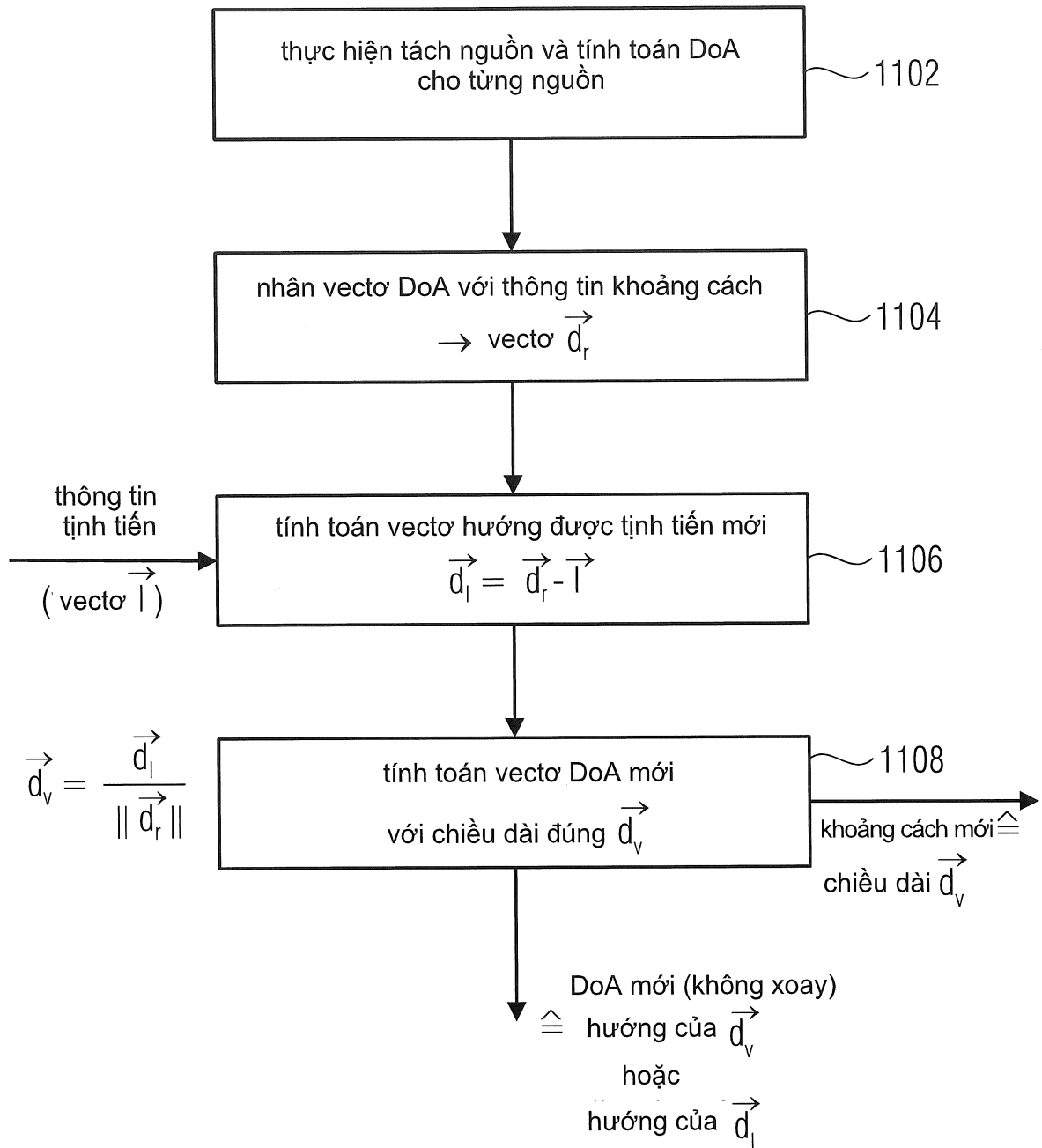


Fig. 11b

15/17

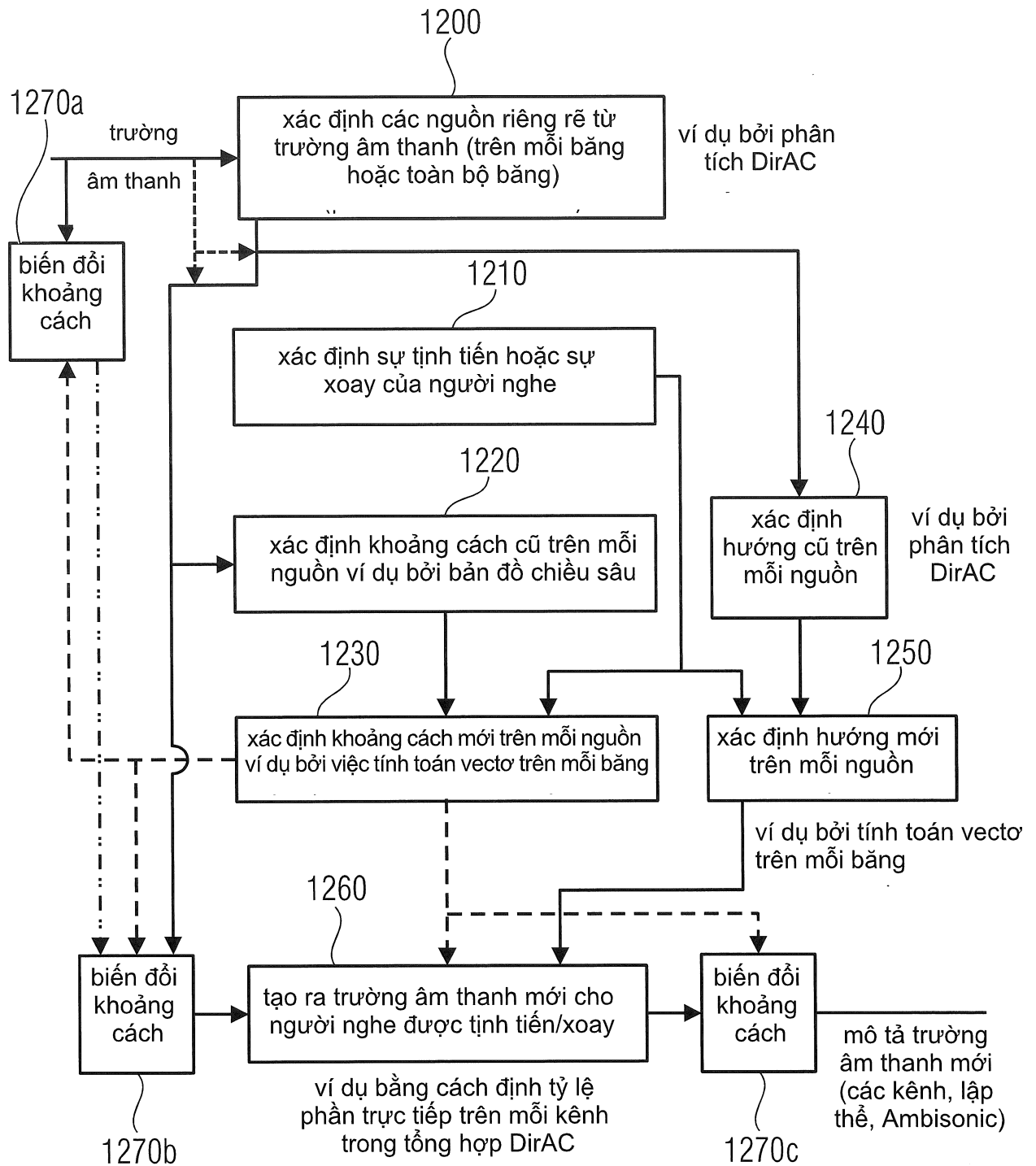
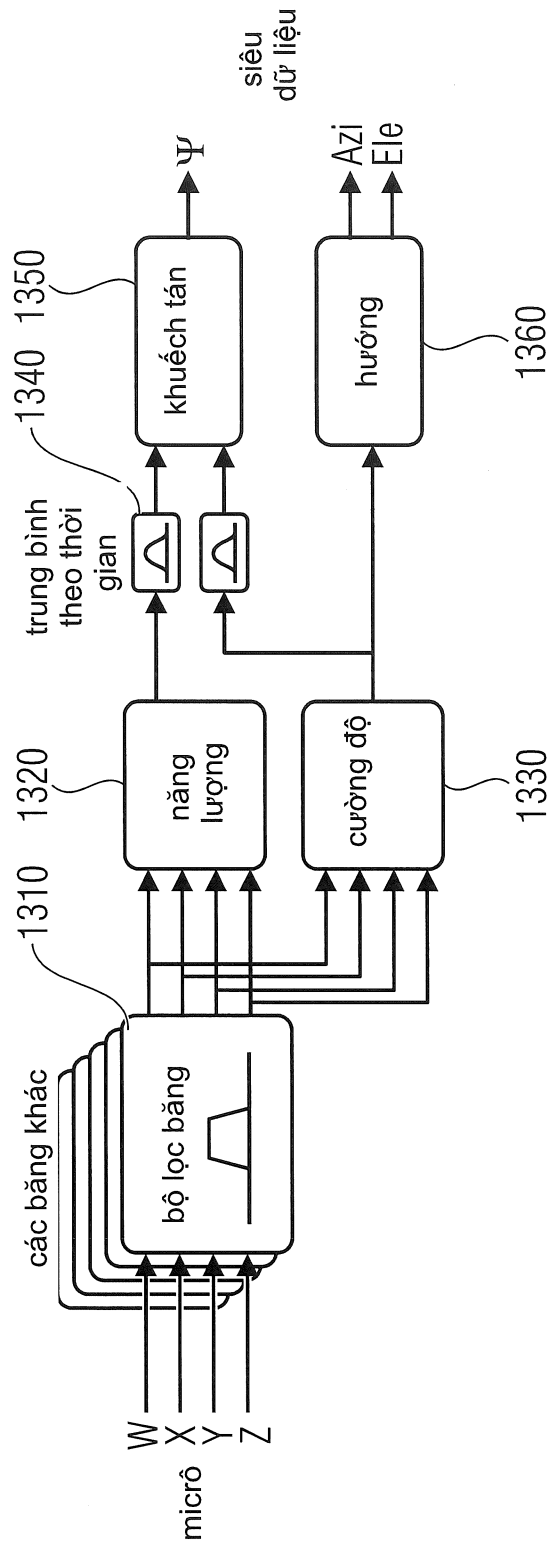


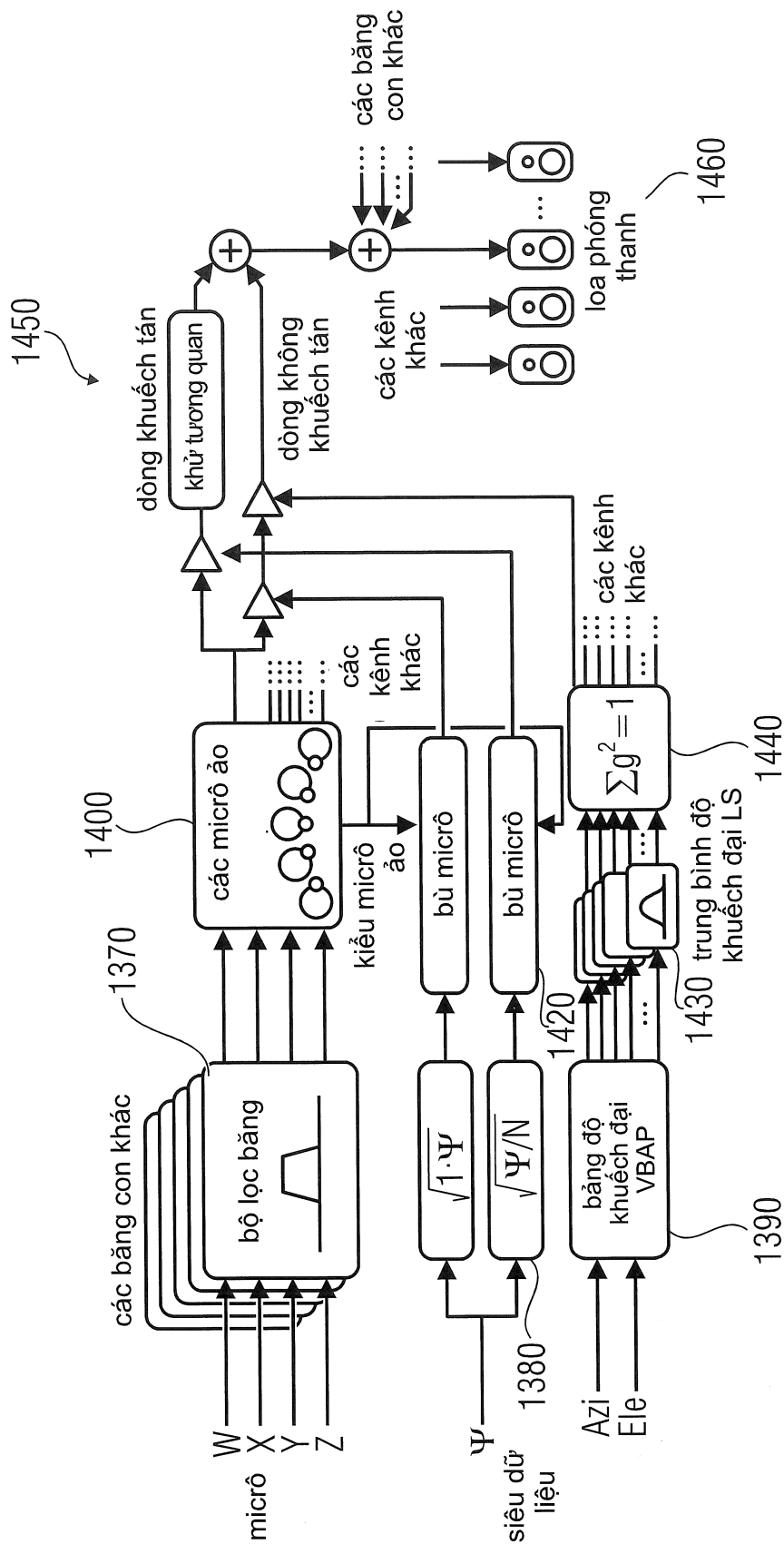
Fig. 12



Phân tích DirAC

Fig. 13a

17/17



Tổng hợp DirAC

Fig. 13b