



(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)



1-0043610

(51)^{2020.01} G10L 19/008; G10L 19/16

(13) B

(21) 1-2020-02203

(22) 01/10/2018

(86) PCT/EP2018/076641 01/10/2018

(87) WO 2019/068638 11/04/2019

(30) 17194816.9 04/10/2017 EP

(45) 25/02/2025 443

(43) 25/09/2020 390A

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

(72) FUCHS, Guillaume (FR); HERRE, Juergen (DE); KUECH, Fabian (DE); DOEHLA, Stefan (DE); MULTRUS, Markus (DE); THIERGART, Oliver (DE); WUEBBOLT, Oliver (DE); GHIDO, Florin (RO); BAYER, Stefan (AT); JAEGER, Wolfgang (DE).

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

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP TẠO MÔ TẢ CẢNH ÂM THANH KẾT HỢP

(21) 1-2020-02203

(57) Sáng chế đề cập đến thiết bị và phương pháp tạo mô tả cảnh âm thanh kết hợp. Thiết bị tạo mô tả cảnh âm thanh kết hợp bao gồm: giao diện đầu vào (100) để nhận mô tả thứ nhất của cảnh thứ nhất ở định dạng thứ nhất và mô tả thứ hai của cảnh thứ hai ở định dạng thứ hai, trong đó định dạng thứ hai khác với định dạng thứ nhất; bộ chuyển đổi định dạng (120) để chuyển đổi mô tả thứ nhất thành định dạng chung và để chuyển đổi mô tả thứ hai thành định dạng chung, khi định dạng thứ hai khác với định dạng chung; và bộ kết hợp định dạng (140) để kết hợp mô tả thứ nhất ở định dạng chung và mô tả thứ hai ở định dạng chung để thu được cảnh âm thanh kết hợp.

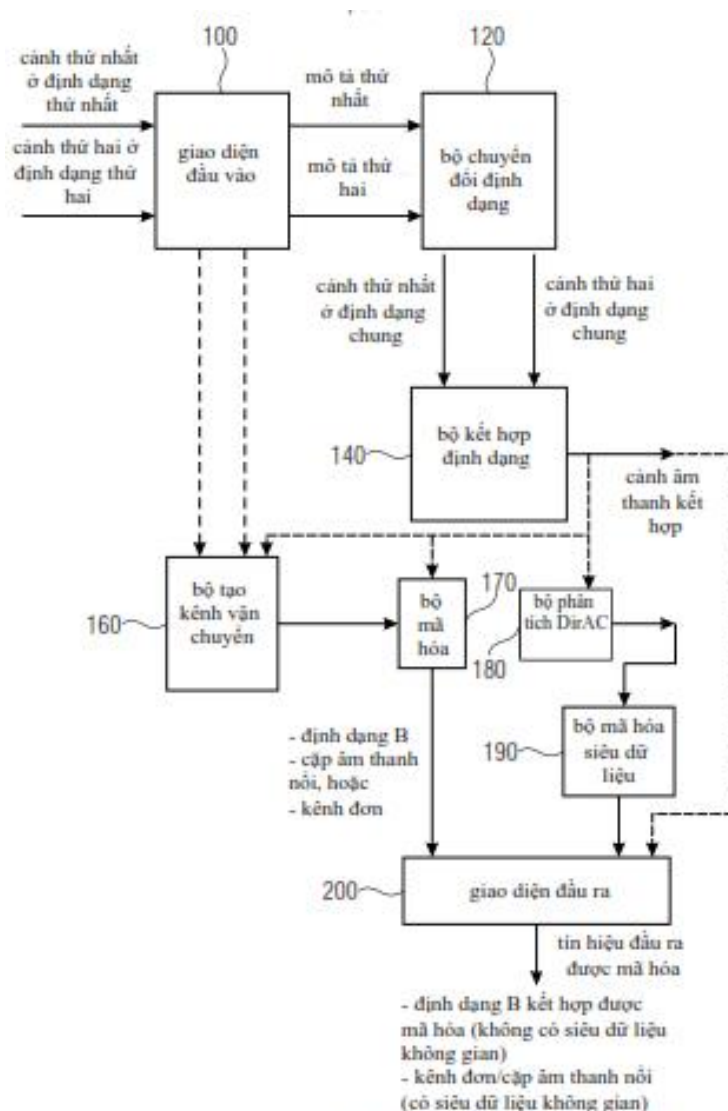


Fig. 1a

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc xử lý tín hiệu âm thanh và cụ thể là đề cập đến việc xử lý tín hiệu âm thanh của các mô tả âm thanh của các cảnh âm thanh.

Tình trạng kỹ thuật của sáng chế

Việc truyền cảnh âm thanh theo ba chiều đòi hỏi phải xử lý nhiều kênh thường truyền một lượng lớn dữ liệu để truyền. Hơn nữa, âm thanh 3D có thể được biểu diễn theo nhiều cách khác nhau: âm thanh dựa trên kênh truyền thống trong đó mỗi kênh truyền được liên kết với vị trí loa phóng thanh; âm thanh được truyền qua các đối tượng âm thanh, mà có thể được định vị theo ba chiều độc lập với các vị trí của loa phóng thanh; và dựa trên cảnh (hoặc Ambisonics), trong đó cảnh âm thanh được biểu diễn bởi tập hợp các tín hiệu hệ số là các trọng số tuyến tính của các hàm cơ sở trực giao không gian, ví dụ, các sóng hài hình cầu. Trái ngược với sự biểu diễn dựa trên kênh, sự biểu diễn dựa trên cảnh độc lập với thiết lập loa phóng thanh cụ thể, và có thể được tái tạo trên bất kỳ thiết lập loa phóng thanh nào với chi phí của quá trình kết xuất bổ sung tại bộ giải mã.

Đối với mỗi định dạng này, các sơ đồ mã hóa chuyên dụng đã được phát triển để lưu trữ hoặc truyền hiệu quả ở các tốc độ bit thấp các tín hiệu âm thanh. Ví dụ, âm thanh vòm MPEG (Moving Picture Experts Group: nhóm các chuyên gia hình ảnh động) là sơ đồ mã hóa theo tham số cho âm thanh vòm dựa trên kênh, trong khi mã hóa đối tượng âm thanh không gian MPEG (Spatial Audio Object Coding - SAOC) là phương pháp mã hóa theo tham số dành riêng cho âm thanh dựa trên đối tượng. Kỹ thuật mã hóa theo tham số cho bậc Ambisonics cao cũng được đề xuất trong giai đoạn MPEG-H tiêu chuẩn 2 gần đây.

Trong bối cảnh này, nơi cả ba sự biểu diễn của cảnh âm thanh, âm thanh dựa trên kênh, dựa trên đối tượng và dựa trên cảnh, được sử dụng và cần được hỗ trợ, cần phải thiết kế sơ đồ phổ quát cho phép mã hóa theo tham số hiệu quả cả ba sự biểu diễn âm thanh 3D. Ngoài ra, cần phải có khả năng mã hóa, truyền và tái tạo các cảnh âm thanh phức tạp gồm hỗn hợp các sự biểu diễn âm thanh khác nhau.

Kỹ thuật mã hóa âm thanh định hướng (Directional Audio Coding - DirAC) [1] là cách tiếp cận hiệu quả để phân tích và tái tạo âm thanh không gian. DirAC sử dụng sự biểu diễn thúc đẩy theo cảm giác của trường âm thanh dựa trên hướng đến (direction of arrival - DOA) và độ khuếch tán đo được trên mỗi dải tần. Nó được xây dựng dựa trên giả định rằng tại một thời điểm tức thời và tại một dải tần hạn, độ phân giải không gian của hệ thống thính giác bị giới hạn để giải mã một tín hiệu cho hướng và một tín hiệu khác cho sự kết hợp giữa các âm thanh. Sau đó, âm thanh không gian được biểu diễn trong miền tần số bằng cách làm nhiễu xuyên kênh hai luồng: luồng khuếch tán vô hướng và luồng không khuếch tán có hướng.

DirAC ban đầu được dự định cho âm thanh định dạng B được ghi nhưng cũng có thể dùng làm định dạng chung để trộn các định dạng âm thanh khác nhau. DirAC đã được mở rộng để xử lý định dạng âm thanh vòm thông thường 5.1 trong tài liệu [3]. Nó cũng được đề xuất để trộn sắp nhiều luồng DirAC trong tài liệu [4]. Ngoài ra, DirAC đã được mở rộng còn để hỗ trợ các đầu vào micrô ngoài định dạng B [6].

Tuy nhiên, khái niệm phổ dụng bị thiếu để biến DirAC thành sự biểu diễn phổ dụng của các cảnh âm thanh trong 3D mà cũng có thể hỗ trợ khái niệm các đối tượng âm thanh.

Một vài cân nhắc trước đó đã được thực hiện để xử lý các đối tượng âm thanh trong DirAC. DirAC được sử dụng trong tài liệu [5] dưới dạng đầu trước âm thanh cho bộ giải mã âm thanh không gian SAOC, như phép phân tách nguồn khó thấy để trích xuất một số người nói từ hỗn hợp các nguồn. Tuy nhiên, người ta không hình dung sử dụng chính DirAC làm sơ đồ mã hóa âm thanh không gian và xử lý trực tiếp các đối tượng âm thanh cùng với siêu dữ liệu của chúng và có khả năng kết hợp chúng với nhau và với các sự biểu diễn âm thanh khác.

Tài liệu tham khảo

[1] V. Pulkki, M-V Laitinen, J Vilkkamo, J Ahonen, T Lokki và T Pihlajamäki, “Directional audio coding - perception-based reproduction of spatial sound”, Hội thảo quốc tế về các nguyên tắc và ứng dụng về nghe không gian, tháng 11 năm 2009, Zao; Miyagi, Nhật Bản.

[2] Ville Pulkki. "Virtual source positioning using vector base amplitude panning", J. Audio Eng. Soc., 45(6):456-466, tháng 6 năm 1997.

[3] M. V. Laitinen và V. Pulkki, "Converting 5.1 audio recordings to B-format for directional audio coding reproduction", Hội thảo quốc tế về âm thanh, tiếng nói và xử lý tín hiệu IEEE 2011 (ICASSP), Prague, 2011, pp. 61-64.

[4] G. Del Galdo, F. Kuech, M. Kallinger và R. Schultz-Amling, "Efficient merging of multiple audio streams for spatial sound reproduction in Directional Audio Coding," Hội thảo quốc tế về âm thanh, tiếng nói và xử lý tín hiệu IEEE 2009, Taipei, 2009, pp. 265-268.

[5] Juergen Herre, Cornelia Falch, Dirk Mahne, Giovanni Del Galdo, Markus Kallinger, và Oliver Thiergart, "Interactive Teleconferencing Combining Spatial Audio Object Coding and DirAC Technology", J. Audio Eng. Soc., Vol. 59, No. 12, tháng 12 năm 2011.

[6] R. Schultz-Amling, F. Kuech, M. Kallinger, G. Del Galdo, J. Ahonen, V. Pulkki, "Planar Microphone Array Processing for the Analysis and Reproduction of Spatial Audio using Directional Audio Coding," Hội nghị kỹ thuật âm thanh 124, Amsterdam, Hà Lan, 2008.

[7] Daniel P. Jarrett and Oliver Thiergart and Emanuel A. P. Habets và Patrick A. Naylor, "Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain", Hội nghị kỹ sư điện và điện tử ở Israel IEEE lần thứ 27 (IEEEI), 2012.

[8] Bằng sáng chế Mỹ số 9,015,051.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương thức cải tiến về việc xử lý và tiến hành các cảnh âm thanh và các mô tả cảnh âm thanh.

Mục đích này đạt được bởi các đối tượng trong các điểm yêu cầu bảo hộ độc lập.

Các phương án của sáng chế liên quan đến sơ đồ mã hóa theo tham số phổ dụng cho cảnh âm thanh 3D được xây dựng xung quanh mô hình mã hóa âm thanh định

hướng (DirAC), kỹ thuật được thúc đẩy theo cảm giác để xử lý âm thanh không gian. DirAC ban đầu được thiết kế để phân tích bản ghi định dạng B của cảnh âm thanh. Sáng chế nhằm mục đích mở rộng khả năng xử lý hiệu quả mọi định dạng âm thanh không gian chẳng hạn như âm thanh dựa trên kênh, Ambisonics, các đối tượng âm thanh, hoặc hỗn hợp của chúng

Phép tái tạo DirAC có thể dễ dàng được tạo ra cho bố trí loa phóng thanh và tai nghe tùy ý. Sáng chế cũng mở rộng khả năng này để xuất ra thêm Ambisonics, các đối tượng âm thanh hoặc hỗn hợp định dạng. Quan trọng hơn, sáng chế cho phép người dùng có thể điều khiển các đối tượng âm thanh và để đạt được, ví dụ, tăng cường hội thoại ở đầu bộ giải mã.

Bối cảnh: Tổng quan hệ thống của bộ giải mã âm thanh không gian DirAC

Sau đây, tổng quan về hệ thống mã hóa âm thanh không gian mới dựa trên DirAC được thiết kế cho dịch vụ âm thanh và tiếng nói chìm (Immersive Voice and Audio Services - IVAS) sẽ được trình bày. Mục tiêu của một hệ thống như vậy là có thể xử lý các định dạng âm thanh không gian khác nhau biểu diễn cảnh âm thanh và mã hóa chúng ở các tốc độ bit thấp và tái tạo cảnh âm thanh gốc một cách trung thực nhất có thể sau khi truyền.

Hệ thống có thể chấp nhận làm các sự biểu diễn khác nhau đầu vào cho các cảnh âm thanh. Cảnh âm thanh đầu vào có thể được ghi lại bằng các tín hiệu đa kênh nhằm được tái tạo ở các vị trí loa phóng thanh khác nhau, các đối tượng thính giác cùng với siêu dữ liệu mô tả các vị trí của các đối tượng theo thời gian, hoặc định dạng Ambisonics bậc một hoặc bậc cao hơn biểu diễn trường âm thanh tại vị trí người nghe hoặc vị trí tham chiếu.

Tốt hơn là hệ thống này dựa trên dịch vụ thoại cải tiến 3GPP (Enhanced Voice Services - EVS) vì giải pháp này dự kiến sẽ hoạt động với độ trễ thấp để kích hoạt các dịch vụ đàm thoại trên các mạng di động.

Fig.9 là phía mã hóa của mã hóa âm thanh không gian dựa trên DirAC hỗ trợ các định dạng âm thanh khác nhau. Như được thể hiện trên Fig.9, bộ mã hóa (bộ mã hóa IVAS) có khả năng hỗ trợ các định dạng âm thanh khác nhau được trình bày cho

hệ thống một cách riêng biệt hoặc cùng một lúc. Các tín hiệu âm thanh có thể là âm thanh trong tự nhiên, được thu nhận bởi micro, hoặc tín hiệu điện trong tự nhiên, được cho là được truyền đến loa phóng thanh. Các định dạng âm thanh được hỗ trợ có thể là tín hiệu đa kênh, các thành phần Ambisonics bậc một và bậc cao hơn, và các đối tượng âm thanh. Cảnh âm thanh phức tạp cũng có thể được mô tả bằng cách kết hợp các định dạng đầu vào khác nhau. Tất cả các định dạng âm thanh sau đó được truyền đến phép phân tích DirAC 180 trích xuất sự biểu diễn theo tham số của cảnh âm thanh hoàn chỉnh. Hướng đến và độ khuếch tán được đo trên mỗi đơn vị thời gian-tần số tạo thành các tham số. Phép phân tích DirAC theo sau là bộ mã hóa siêu dữ liệu không gian 190 định lượng và mã hóa các tham số DirAC để thu được sự biểu diễn theo tham số tốc độ bit thấp.

Cùng với các tham số, tín hiệu trộn giảm có nguồn gốc 160 từ các nguồn hoặc tín hiệu đầu vào âm thanh khác nhau được mã hóa để truyền bởi bộ mã hóa lõi âm thanh thông thường 170. Trong trường hợp này, một bộ mã hóa âm thanh dựa trên EVS được chấp nhận để mã hóa tín hiệu trộn giảm. Tín hiệu trộn giảm bao gồm các kênh khác nhau, được gọi là các kênh vận chuyển: tín hiệu có thể là, ví dụ, bốn tín hiệu hệ số tạo thành tín hiệu định dạng B, cặp âm thanh nổi hoặc trộn giảm đơn âm phụ thuộc vào tốc độ bit được đặt ra. Các tham số không gian được mã hóa và dòng bit âm thanh được mã hóa được ghép kênh trước khi được truyền qua kênh truyền thông.

Fig.10 là bộ giải mã của mã hóa âm thanh không gian dựa trên DirAC cung cấp các định dạng âm thanh khác nhau. Trong bộ giải mã như được thể hiện trên Fig.10, các kênh vận chuyển được giải mã bởi bộ giải mã lõi 1020, trong khi siêu dữ liệu DirAC được giải mã trước tiên 1060 trước khi được chuyển với các kênh vận chuyển được giải mã tới phép tổng hợp DirAC 220, 240. Ở giai đoạn này (1040), các lựa chọn khác nhau có thể được xem xét. Có thể yêu cầu phát cảnh âm thanh trực tiếp trên bất kỳ cấu hình loa phóng thanh hoặc tai nghe nào như thường có trong hệ thống DirAC thông thường (MC trên Fig.10). Ngoài ra, cũng có thể yêu cầu hiển thị cảnh ở định dạng Ambisonics cho các thao tác khác, chẳng hạn như xoay, dời lại hoặc chuyển động của cảnh (FOA/HOA trên Fig.10). Cuối cùng, bộ giải mã có thể phân phối các đối tượng riêng lẻ như chúng đã có mặt ở phía bộ mã hóa (các đối tượng trên Fig.10).

Các đối tượng âm thanh cũng có thể được phục hồi nhưng điều thú vị hơn đối với người nghe là điều chỉnh bản phối được kết xuất bằng thao tác tương tác của các đối tượng. Các thao tác đối tượng điển hình là điều chỉnh mức độ, sự cân bằng hoặc vị trí không gian của đối tượng. Việc tăng cường đối thoại dựa trên đối tượng trở nên, ví dụ, có khả năng được đưa ra bởi tính năng tương tác này. Cuối cùng, có thể xuất các định dạng ban đầu như chúng đã có mặt ở đầu vào bộ mã hóa. Trong trường hợp này, nó có thể là sự pha trộn của các kênh và đối tượng âm thanh hoặc Ambisonics và các đối tượng. Để đạt được việc truyền riêng các thành phần đa kênh và Ambisonics, một vài trường hợp của hệ thống được mô tả có thể được sử dụng.

Sáng chế này có lợi ở chỗ, đặc biệt là theo khía cạnh thứ nhất, khung được thiết lập để kết hợp các mô tả cảnh khác nhau thành cảnh âm thanh kết hợp theo định dạng chung, cho phép kết hợp các mô tả cảnh âm thanh khác nhau.

Ví dụ, định dạng phổ biến này có thể là định dạng B hoặc có thể là định dạng biểu diễn tín hiệu áp suất/vận tốc, hoặc tốt hơn là, cũng có thể là định dạng biểu diễn tham số DirAC.

Định dạng này là định dạng nhỏ gọn, ngoài ra, cho phép một lượng đáng kể tương tác của người dùng một mặt và mặt khác là hữu ích đối với tốc độ bit cần thiết để biểu diễn tín hiệu âm thanh.

Theo khía cạnh khác nữa của sáng chế, việc tổng hợp nhiều cảnh âm thanh có thể được thực hiện một cách thuận lợi bằng cách kết hợp hai hoặc nhiều mô tả DirAC khác nhau. Cả hai mô tả DirAC khác nhau này đều có thể được xử lý bằng cách kết hợp các cảnh trong miền tham số hoặc, bằng cách khác, kết xuất riêng từng cảnh âm thanh và sau đó kết hợp các cảnh âm thanh đã được kết xuất từ các mô tả DirAC riêng lẻ trong miền phổ hoặc, thay vào đó, đã có trong miền thời gian.

Quy trình này cho phép xử lý rất hiệu quả và chất lượng cao các cảnh âm thanh khác nhau được kết hợp thành sự biểu diễn cảnh duy nhất, và đặc biệt là tín hiệu âm thanh miền thời gian duy nhất.

Khía cạnh khác của sáng chế có lợi ở chỗ dữ liệu âm thanh đặc biệt hữu ích được chuyển đổi để chuyển đổi siêu dữ liệu đối tượng thành siêu dữ liệu DirAC được

lấy từ nơi bộ chuyển đổi dữ liệu âm thanh này có thể được sử dụng trong khung của khía cạnh thứ nhất, thứ hai hoặc thứ ba hoặc cũng có thể được áp dụng độc lập với nhau. Bộ chuyển đổi dữ liệu âm thanh cho phép chuyển đổi hiệu quả dữ liệu đối tượng âm thanh, ví dụ, tín hiệu dạng sóng cho đối tượng âm thanh, và dữ liệu vị trí tương ứng, thông thường, liên quan đến thời gian để biểu diễn quỹ đạo nhất định của đối tượng âm thanh trong quá trình tái tạo thiết lập thành mô tả cảnh âm thanh rất hữu ích và nhỏ gọn, và đặc biệt là định dạng mô tả cảnh âm thanh DirAC. Trong khi mô tả đối tượng âm thanh diễn hình với tín hiệu dạng sóng đối tượng âm thanh và siêu dữ liệu vị trí đối tượng âm thanh có liên quan đến thiết lập tái tạo cụ thể hoặc nói chung, có liên quan đến hệ tọa độ tái tạo nhất định, mô tả DirAC đặc biệt hữu ích ở chỗ nó có liên quan đến vị trí của người nghe hoặc micrô và hoàn toàn không có bất kỳ giới hạn nào đối với thiết lập loa phóng thanh hoặc thiết lập tái tạo.

Do đó, mô tả DirAC được tạo từ các tín hiệu siêu dữ liệu của đối tượng âm thanh cũng cho phép kết hợp các đối tượng âm thanh rất hữu ích và nhỏ gọn và chất lượng cao khác với các kỹ thuật kết hợp đối tượng âm thanh khác chẳng hạn như mã hóa đối tượng âm thanh không gian hoặc quét biên độ của các đối tượng trong thiết lập tái tạo.

Bộ mã hóa cảnh âm thanh theo khía cạnh khác nữa của sáng chế đặc biệt hữu ích trong việc cung cấp sự biểu diễn kết hợp của cảnh âm thanh có siêu dữ liệu DirAC và, ngoài ra, cung cấp đối tượng âm thanh có siêu dữ liệu đối tượng âm thanh.

Đặc biệt, trường hợp này đặc biệt hữu ích và có lợi cho tính tương tác cao để tạo ra mô tả siêu dữ liệu kết hợp có siêu dữ liệu DirAC trên một mặt và song song với đó là siêu dữ liệu đối tượng trên mặt khác. Do đó, ở khía cạnh này, siêu dữ liệu đối tượng không được kết hợp với siêu dữ liệu DirAC, mà được chuyển đổi thành siêu dữ liệu giống như DirAC sao cho siêu dữ liệu đối tượng bao gồm theo hướng hoặc, ngoài ra, khoảng cách và/hoặc độ khuếch tán của đối tượng riêng lẻ cùng với tín hiệu đối tượng. Do đó, tín hiệu đối tượng được chuyển đổi thành sự biểu diễn giống DirAC để có thể xử lý linh hoạt sự biểu diễn DirAC cho cảnh âm thanh thứ nhất và đối tượng bổ sung trong cảnh âm thanh thứ nhất này. Do đó, ví dụ, các đối tượng cụ thể có thể được

xử lý rất chọn lọc do thực tế là kênh vận chuyển tương ứng của chúng ở một mặt và các tham số kiểu DirAC ở mặt khác vẫn có sẵn.

Theo khía cạnh khác nữa của sáng chế, thiết bị hoặc phương pháp thực hiện phép tổng hợp dữ liệu âm thanh là đặc biệt hữu ích ở chỗ bộ thao tác được cung cấp để thao tác mô tả DirAC của một hoặc nhiều đối tượng âm thanh, mô tả DirAC về tín hiệu đa kênh hoặc mô tả DirAC về các tín hiệu Ambisonics bậc nhất hoặc các tín hiệu Ambisonics bậc cao hơn. Và, mô tả DirAC đã thao tác sau đó được tổng hợp bằng bộ tổng hợp DirAC.

Khía cạnh này có ưu điểm đặc biệt là mọi thao tác cụ thể đối với bất kỳ tín hiệu âm thanh nào đều được thực hiện rất hữu ích và hiệu quả trong miền DirAC, tức là bằng cách thao tác kênh vận chuyển của mô tả DirAC hoặc bằng cách thao tác thay thế dữ liệu tham số của mô tả DirAC. Việc sửa đổi này thực sự hiệu quả hơn và thực tế hơn để thực hiện trong miền DirAC so với thao tác trong các miền khác. Đặc biệt, các hoạt động gán trọng số phụ thuộc vào vị trí như các hoạt động thao tác được ưu tiên có thể được thực hiện cụ thể trong miền DirAC. Do đó, trong phương án cụ thể, việc chuyển đổi sự biểu diễn tín hiệu tương ứng trong miền DirAC và sau đó, thực hiện thao tác trong miền DirAC là trường hợp ứng dụng đặc biệt hữu ích để xử lý và thao tác cảnh âm thanh hiện đại.

Mô tả vắn tắt các hình vẽ

Các phương án được ưu tiên sau đó sẽ được thảo luận liên quan đến các hình vẽ đi kèm của chúng, trong đó:

Fig.1a là sơ đồ khối của cách triển khai được ưu tiên của thiết bị hoặc phương pháp tạo mô tả về cảnh âm thanh kết hợp theo khía cạnh thứ nhất của sáng chế;

Fig.1b là cách triển khai tạo cảnh âm thanh kết hợp, trong đó định dạng chung là sự biểu diễn áp suất/vận tốc;

Fig.1c là cách triển khai được ưu tiên của phép tạo cảnh âm thanh kết hợp, trong đó các tham số DirAC và mô tả DirAC là định dạng chung;

Fig.1d là cách triển khai được ưu tiên của bộ kết hợp trên Fig.1c minh họa hai phương án khác nhau để thực hiện bộ kết hợp các tham số DirAC của các cảnh âm thanh hoặc các mô tả cảnh âm thanh khác nhau;

Fig.1e là cách triển khai được ưu tiên của phép tạo cảnh âm thanh kết hợp trong đó định dạng chung là định dạng B làm ví dụ cho sự biểu diễn Ambisonics;

Fig.1f là hình minh họa đối tượng âm thanh/bộ chuyển đổi DirAC hữu ích trong bối cảnh, ví dụ, của Fig.1c hoặc Fig.1d hoặc hữu ích trong bối cảnh của khía cạnh thứ ba liên quan đến bộ chuyển đổi siêu dữ liệu;

Fig.1g là hình minh họa mẫu về tín hiệu đa kênh 5.1 thành mô tả DirAC;

Fig.1h là hình minh họa thêm cho việc chuyển đổi định dạng đa kênh thành định dạng DirAC trong bối cảnh của phía bộ mã hóa và phía bộ giải mã;

Fig.2a minh họa phương án của thiết bị hoặc phương pháp tổng hợp nhiều cảnh âm thanh theo khía cạnh thứ hai của sáng chế;

Fig.2b minh họa cách triển khai được ưu tiên của bộ tổng hợp DirAC của Fig.2a;

Fig.2c minh họa cách triển khai thêm của bộ tổng hợp DirAC với sự kết hợp các tín hiệu được kết xuất;

Fig.2d minh họa cách triển khai bộ thao tác chọn lọc được kết nối hoặc trước bộ kết hợp cảnh 221 của Fig.2b hoặc trước bộ kết hợp 225 của Fig.2c;

Fig.3a là cách triển khai được ưu tiên của thiết bị hoặc phương pháp thực hiện và chuyển đổi dữ liệu âm thanh theo khía cạnh thứ ba của sáng chế;

Fig.3b là cách triển khai được ưu tiên của bộ chuyển đổi siêu dữ liệu cũng được minh họa trên Fig.1f;

Fig.3c là lưu đồ thực hiện việc triển khai thêm của việc chuyển đổi dữ liệu âm thanh thông qua miền áp suất/vận tốc;

Fig.3d minh họa lưu đồ để thực hiện việc kết hợp trong miền DirAC;

Fig.3e minh họa cách triển khai được ưu tiên để kết hợp các mô tả DirAC khác nhau, ví dụ như được minh họa trên Fig.1d liên quan đến khía cạnh thứ nhất của sáng chế;

Fig.3f minh họa việc chuyển đổi dữ liệu vị trí đối tượng thành sự biểu diễn theo tham số DirAC;

Fig.4a minh họa cách triển khai được ưu tiên của bộ mã hóa cảnh âm thanh theo khía cạnh thứ tư của sáng chế để tạo mô tả siêu dữ liệu kết hợp bao gồm siêu dữ liệu DirAC và siêu dữ liệu đối tượng;

Fig.4b minh họa phương án được ưu tiên liên quan đến khía cạnh thứ tư của sáng chế;

Fig.5a là cách triển khai được ưu tiên của thiết bị thực hiện phép tổng hợp dữ liệu âm thanh hoặc phương pháp tương ứng theo khía cạnh thứ năm của sáng chế;

Fig.5b minh họa cách triển khai được ưu tiên của bộ tổng hợp DirAC của Fig.5a;

Fig.5c minh họa sự thay thế khác cho quy trình của bộ thao tác của Fig.5a;

Fig.5d minh họa quy trình khác để thực hiện bộ thao tác theo Fig.5a;

Fig.6 minh họa bộ chuyển đổi tín hiệu âm thanh để tạo, từ tín hiệu đơn âm và thông tin hướng đến, tức là, từ mô tả DirAC mẫu, độ khuếch tán ví dụ được đặt thành không, sự biểu diễn định dạng B bao gồm thành phần đa hướng và các thành phần có hướng theo các hướng X, Y và Z;

Fig.7a minh họa cách triển khai phép phân tích DirAC của tín hiệu micro định dạng B;

Fig.7b minh họa cách triển khai phép tổng hợp DirAC theo quy trình đã biết;

Fig.8 minh họa lưu đồ để minh họa các phương án tiếp theo, đặc biệt là phương án theo Fig.1a;

Fig.9 là phía bộ mã hóa của phép mã hóa âm thanh không gian dựa trên DirAC hỗ trợ các định dạng âm thanh khác nhau.

Fig.10 là bộ giải mã của phép mã hóa âm thanh không gian dựa trên DirAC cung cấp các định dạng âm thanh khác nhau.

Fig.11 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau theo định dạng B kết hợp;

Fig.12 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp trong các miền áp suất/vận tốc;

Fig.13 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau trong miền DirAC với khả năng thao tác đối tượng ở phía bộ giải mã;

Fig.14 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau ở phía bộ giải mã thông qua bộ kết hợp siêu dữ liệu DirAC;

Fig.15 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau ở phía bộ giải mã trong phép tổng hợp DirAC; và

Các hình vẽ từ Fig.16a đến Fig.16f minh họa một vài sự biểu diễn của các định dạng âm thanh hữu ích trong bối cảnh của các khía cạnh từ thứ nhất đến thứ năm của sáng chế.

Mô tả chi tiết sáng chế

Fig.1a minh họa phương án được ưu tiên của thiết bị tạo mô tả về cảnh âm thanh kết hợp. Thiết bị bao gồm giao diện đầu vào 100 để nhận mô tả thứ nhất về cảnh thứ nhất ở định dạng thứ nhất và mô tả thứ hai về cảnh thứ hai ở định dạng thứ hai, trong đó định dạng thứ hai khác với định dạng thứ nhất. Định dạng có thể là bất kỳ định dạng cảnh âm thanh nào, chẳng hạn như bất kỳ định dạng hoặc mô tả cảnh nào được minh họa trên các hình vẽ từ Fig.16a đến Fig.16f.

Ví dụ, Fig.16a minh họa mô tả đối tượng chứa thường là tín hiệu dạng sóng của đối tượng 1 (được mã hóa) chẳng hạn như kênh đơn và siêu dữ liệu tương ứng liên quan đến vị trí của đối tượng 1, trong đó thông tin này thường được cung cấp cho mỗi khung thời gian hoặc nhóm các khung thời gian, và tín hiệu dạng sóng của đối tượng 1

được mã hóa. Các sự biểu diễn tương ứng cho đối tượng thứ hai hoặc xa hơn có thể có như được minh họa trên Fig.16a.

Một cách khác có thể là mô tả đối tượng chứa trộn giảm đối tượng là tín hiệu đơn âm, tín hiệu âm thanh nổi có hai kênh hoặc tín hiệu có ba kênh trở lên và siêu dữ liệu đối tượng có liên quan chẳng hạn như năng lượng đối tượng, thông tin tương quan trên mỗi ngăn thời gian/tần số và một cách tùy chọn, các vị trí đối tượng. Tuy nhiên, các vị trí đối tượng cũng có thể được cung cấp ở phía bộ giải mã dưới dạng thông tin kết xuất điển hình và do đó, có thể được biến đổi bởi người dùng. Định dạng trên Fig.16b ví dụ có thể được triển khai dưới dạng định dạng SAOC (spatial audio object coding: mã hóa đối tượng âm thanh không gian) đã biết.

Một mô tả khác về cảnh được minh họa trên Fig.16c là mô tả đa kênh có sự biểu diễn được mã hóa hoặc không được mã hóa của kênh thứ nhất, kênh thứ hai, kênh thứ ba, kênh thứ tư, hoặc kênh thứ năm, trong đó kênh thứ nhất có thể là kênh bên trái L, kênh thứ hai có thể là kênh bên phải R, kênh thứ ba có thể là kênh trung tâm C, kênh thứ tư có thể là kênh vòm bên trái LS và kênh thứ năm có thể là kênh vòm bên phải RS. Đương nhiên, tín hiệu đa kênh có thể có số lượng kênh nhỏ hơn hoặc lớn hơn, chẳng hạn như chỉ có hai kênh cho âm thanh nổi hoặc sáu kênh cho định dạng 5.1 hoặc tám kênh cho định dạng 7.1, v.v.

Cách biểu diễn hiệu quả hơn của tín hiệu đa kênh được minh họa trên Fig.16d, trong đó trộn giảm kênh chẳng hạn như trộn giảm đơn âm, hoặc trộn giảm âm thanh nổi hoặc trộn giảm với nhiều hơn hai kênh có liên quan tới thông tin phụ theo tham số như siêu dữ liệu kênh thông thường cho mỗi ngăn thời gian và/hoặc tần số. Sự biểu diễn theo tham số như vậy ví dụ có thể được thực hiện theo tiêu chuẩn vòm MPEG.

Một sự biểu diễn khác của cảnh âm thanh ví dụ có thể là định dạng B chứa tín hiệu đa hướng W, và các thành phần có hướng X, Y, Z như được thể hiện trên Fig.16e. Đây sẽ là một tín hiệu bậc một hay tín hiệu FoA. Tín hiệu Ambisonics bậc cao hơn, tức là tín hiệu HoA có thể có các thành phần bổ sung như được biết đến trong kỹ thuật trước đó.

Trái ngược với sự biểu diễn theo Fig.16c và Fig.16d, sự biểu diễn theo Fig.16e là sự biểu diễn không phụ thuộc vào thiết lập loa phóng thanh nhất định, mà phụ thuộc

vào các mô tả trường âm thanh như được trải nghiệm ở vị trí (micro hoặc người nghe) nhất định.

Một mô tả trường âm thanh khác như vậy là định dạng DirAC ví dụ như được minh họa trên Fig.16f. Định dạng DirAC thường bao gồm tín hiệu trộn giảm DirAC là tín hiệu đơn âm hoặc âm thanh nổi hoặc bất kỳ tín hiệu trộn giảm hoặc tín hiệu vận chuyển nào và thông tin phụ theo tham số tương ứng. Thông tin phụ theo tham số này, ví dụ, là thông tin hướng đến trên mỗi ngăn thời gian/tần số và một cách tùy ý, thông tin độ khuếch tán trên mỗi ngăn thời gian/tần số.

Đầu vào vào giao diện đầu vào 100 của Fig.1a có thể ví dụ là bất kỳ một trong các định dạng được minh họa liên quan đến các hình vẽ từ Fig.16a đến Fig.16f. Giao diện đầu vào 100 chuyển tiếp các mô tả định dạng tương ứng sang bộ chuyển đổi định dạng 120. Bộ chuyển đổi định dạng 120 được tạo cấu hình để chuyển đổi mô tả thứ nhất thành định dạng chung và để chuyển đổi mô tả thứ hai thành định dạng chung giống như vậy, khi định dạng thứ hai khác với định dạng chung. Tuy nhiên, khi định dạng thứ hai đã ở định dạng chung, thì bộ chuyển đổi định dạng chỉ chuyển đổi mô tả thứ nhất thành định dạng chung, vì mô tả thứ nhất có định dạng khác với định dạng chung.

Do đó, ở đầu ra của bộ chuyển đổi định dạng hoặc, nói chung, ở đầu vào của bộ kết hợp định dạng, tồn tại sự biểu diễn cảnh thứ nhất ở định dạng chung và sự biểu diễn cảnh thứ hai trong cùng định dạng chung. Do thực tế là cả hai mô tả hiện có trong một và cùng một định dạng chung, bộ kết hợp định dạng giờ đây có thể kết hợp mô tả thứ nhất và mô tả thứ hai để thu được cảnh âm thanh kết hợp.

Theo phương án được minh họa trên Fig.1e, bộ chuyển đổi định dạng 120 được tạo cấu hình để chuyển đổi mô tả thứ nhất thành tín hiệu định dạng B thứ nhất, ví dụ như được minh họa ở 127 trên Fig.1e và để tính toán sự biểu diễn định dạng B cho mô tả thứ hai như được minh họa trên Fig.1e ở 128.

Sau đó, bộ kết hợp định dạng 140 được triển khai như một bộ cộng tín hiệu thành phần được minh họa ở 146a cho bộ cộng thành phần W, ở 146b cho bộ cộng thành phần X, được minh họa ở 146c cho bộ cộng thành phần Y và được minh họa ở 146d cho bộ cộng thành phần Z.

Do đó, trong phương án của Fig.1e, cảnh âm thanh kết hợp có thể là sự biểu diễn định dạng B và các tín hiệu định dạng B sau đó có thể hoạt động như các kênh vận chuyển và sau đó có thể được mã hóa thông qua bộ mã hóa kênh vận chuyển 170 của Fig.1a. Do đó, cảnh âm thanh kết hợp liên quan đến tín hiệu định dạng B có thể được nhập trực tiếp vào bộ mã hóa 170 của Fig.1a để tạo tín hiệu định dạng B được mã hóa mà sau đó có thể được xuất ra qua giao diện đầu ra 200. Trong trường hợp này, không cần bất kỳ siêu dữ liệu không gian nào, nhưng, với giá của sự biểu diễn được mã hóa của bốn tín hiệu âm thanh, tức là, thành phần đa hướng W và các thành phần có hướng X, Y, Z.

Thay vào đó, định dạng chung là định dạng áp suất/vận tốc như được minh họa trên Fig.1b. Để đạt được mục đích này, bộ chuyển đổi định dạng 120 bao gồm bộ phân tích thời gian/tần số 121 cho cảnh âm thanh thứ nhất và bộ phân tích thời gian/tần số 122 cho cảnh âm thanh thứ hai hoặc, nói chung, cảnh âm thanh có số lượng N, trong đó N là số nguyên.

Sau đó, với mỗi sự biểu diễn phổ như vậy được tạo bởi các bộ chuyển đổi phổ 121, 122, áp suất và vận tốc được tính như được minh họa ở 123 và 124, và sau đó, bộ kết hợp định dạng được tạo cấu hình để tính toán tín hiệu áp suất tổng trên một mặt bằng cách tính tổng các tín hiệu áp suất tương ứng được tạo bởi các khối 123, 124. Và ngoài ra, tín hiệu vận tốc riêng cũng được tính theo từng khối 123, 124 và các tín hiệu vận tốc có thể được cộng vào với nhau để thu được tín hiệu áp suất/vận tốc kết hợp.

Tùy thuộc vào cách triển khai, các quy trình trong các khối 142, 143 không nhất thiết phải được thực hiện. Thay vào đó, tín hiệu áp suất được kết hợp hoặc "được cộng" và tín hiệu vận tốc được kết hợp hoặc "được cộng" có thể được mã hóa theo cách tương tự như được minh họa trên Fig.1e của tín hiệu định dạng B và sự biểu diễn áp suất/vận tốc này có thể được mã hóa một lần nữa thông qua bộ mã hóa 170 của Fig.1a và sau đó có thể được truyền đến bộ giải mã mà không có bất kỳ thông tin phụ nào liên quan đến các tham số không gian, vì sự biểu diễn áp suất/vận tốc kết hợp đã bao gồm thông tin không gian cần thiết để thu được trường âm thanh chất lượng cao được kết xuất cuối cùng ở phía bộ giải mã.

Tuy nhiên, trong phương án, nên thực hiện phép phân tích DirAC theo sự biểu diễn áp suất/vận tốc được tạo bởi khối 141. Để đạt được mục đích này, vector cường độ 142 được tính toán và trong khối 143, các tham số DirAC từ vector cường độ được tính toán và sau đó, các tham số DirAC kết hợp được lấy làm sự biểu diễn theo tham số của cảnh âm thanh kết hợp. Để đạt được mục đích này, bộ phân tích DirAC 180 theo Fig.1a được triển khai để thực hiện chức năng của khối 142 và 143 của Fig.1b. Và tốt hơn là, dữ liệu DirAC còn phải trải qua thao tác mã hóa siêu dữ liệu trong bộ mã hóa siêu dữ liệu 190. Bộ mã hóa siêu dữ liệu 190 thường bao gồm bộ lượng tử hóa và bộ mã hóa entropy để làm giảm tốc độ bit cần thiết cho việc truyền các tham số DirAC.

Cùng với các tham số DirAC được mã hóa, kênh vận chuyển được mã hóa cũng được truyền đi. Kênh vận chuyển được mã hóa được tạo bởi bộ tạo kênh vận chuyển 160 của Fig.1a ví dụ có thể được triển khai như được minh họa trên Fig.1b bởi bộ tạo trộn giảm thứ nhất 161 để tạo ra trộn giảm từ cảnh âm thanh thứ nhất và bộ tạo trộn giảm thứ N 162 để tạo trộn giảm từ cảnh âm thanh thứ N.

Sau đó, các kênh trộn giảm được kết hợp trong bộ kết hợp 163 thông thường bằng cách bổ sung đơn giản và tín hiệu trộn giảm được kết hợp sau đó là kênh vận chuyển được mã hóa bởi bộ mã hóa 170 của Fig.1a. Ví dụ, trộn giảm được kết hợp có thể là cặp âm thanh nổi, tức là, kênh thứ nhất và kênh thứ hai của sự biểu diễn âm thanh nổi hoặc có thể là kênh đơn, tức là tín hiệu kênh đơn.

Theo phương án tiếp theo được minh họa trên Fig.1c, phép chuyển đổi định dạng trong bộ chuyển đổi định dạng 120 được thực hiện để chuyển đổi trực tiếp từng định dạng âm thanh đầu vào thành định dạng DirAC là định dạng phổ biến. Để đạt được mục đích này, bộ chuyển đổi định dạng 120 một lần nữa tạo phép chuyển đổi thời gian-tần số hoặc phép phân tích thời gian/tần số trong các khối tương ứng 121 cho cảnh thứ nhất và khối 122 cho cảnh thứ hai hoặc cảnh xa hơn. Sau đó, các tham số DirAC được suy ra từ các sự biểu diễn phổ của các cảnh âm thanh tương ứng được minh họa ở 125 và 126. Kết quả của quy trình trong các khối 125 và 126 là các tham số DirAC chứa thông tin năng lượng trên mỗi ô thời gian/tần số, thông tin hướng đến e_{DOA} trên mỗi ô thời gian/tần số và thông tin độ khuếch tán ψ cho mỗi ô thời gian/tần số. Sau đó, bộ kết hợp định dạng 140 được tạo cấu hình để thực hiện kết hợp trực tiếp

trong miền tham số DirAC để tạo các tham số DirAC kết hợp ψ cho độ khuếch tán và e_{DOA} cho hướng đến. Đặc biệt, bộ kết hợp 144 cần thông tin năng lượng E_1 và E_N nhưng không phải là một phần của sự biểu diễn theo tham số kết hợp cuối cùng được tạo bởi bộ kết hợp định dạng 140.

Do đó, so sánh Fig.1c với Fig.1e cho thấy rằng, khi bộ kết hợp định dạng 140 đã thực hiện kết hợp trong miền tham số DirAC, bộ phân tích DirAC 180 là không cần thiết và không được thực hiện. Thay vào đó, đầu ra của bộ kết hợp định dạng 140 là đầu ra của khối 144 trên Fig.1c được chuyển trực tiếp đến bộ mã hóa siêu dữ liệu 190 của Fig.1a và từ đó vào giao diện đầu ra 200 sao cho siêu dữ liệu không gian được mã hóa và đặc biệt là các tham số DirAC kết hợp được mã hóa có trong đầu ra tín hiệu đầu ra được mã hóa bởi giao diện đầu ra 200.

Hơn nữa, bộ tạo kênh vận chuyển 160 của Fig.1a có thể nhận, từ giao diện đầu vào 100, sự biểu diễn tín hiệu dạng sóng cho cảnh thứ nhất và sự biểu diễn tín hiệu dạng sóng cho cảnh thứ hai. Các sự biểu diễn này được đưa vào các khối tạo trộn giảm 161, 162 và các kết quả được cộng vào khối 163 để thu được trộn giảm kết hợp như được minh họa đối với Fig.1b.

Fig.1d minh họa sự biểu diễn tương tự với Fig.1c. Tuy nhiên, trên Fig.1d, dạng sóng của đối tượng âm thanh được đưa vào bộ chuyển đổi sự biểu diễn thời gian/tần số 121 cho đối tượng âm thanh 1 và 122 cho đối tượng âm thanh N. Ngoài ra, siêu dữ liệu được đưa vào, cùng với sự biểu diễn phổ vào bộ tính tham số DirAC 125, 126 như được minh họa trên Fig.1c.

Tuy nhiên, Fig.1d cung cấp sự biểu diễn chi tiết hơn về cách thức triển khai hoạt động của bộ kết hợp 144. Trong phương án thứ nhất, bộ kết hợp thực hiện phép cộng có trọng số năng lượng của độ khuếch tán riêng cho từng đối tượng hoặc cảnh riêng lẻ, và phép tính có trọng số năng lượng tương ứng của DoA kết hợp cho mỗi ô thời gian/tần số được thực hiện như được minh họa trong phương trình thấp hơn của phương án 1.

Tuy nhiên, các triển khai khác cũng có thể được thực hiện. Đặc biệt, một phép tính rất hiệu quả khác là đặt độ khuếch tán thành không cho siêu dữ liệu DirAC kết hợp, và chọn hướng đến được tính từ đối tượng âm thanh nhất định có năng lượng cao

nhất trong ô thời gian/tần số làm hướng đến cho mỗi ô thời gian/tần số. Tốt hơn là, quy trình theo Fig.1d thích hợp hơn khi đầu vào vào giao diện đầu vào là các đối tượng âm thanh riêng lẻ tương ứng biểu thị dạng sóng hoặc tín hiệu đơn cho từng đối tượng và siêu dữ liệu tương ứng chẳng hạn như thông tin vị trí được minh họa liên quan đến Fig.16a hoặc Fig.16b.

Tuy nhiên, trong phương án của Fig.1c, cảnh âm thanh có thể là sự biểu diễn khác bất kỳ được minh họa trên Fig.16c, 16d, 16e hoặc 16f. Sau đó, có thể có siêu dữ liệu hoặc không, tức là siêu dữ liệu trên Fig.1c là tùy ý. Tuy nhiên, sau đó, độ khuếch tán hữu ích điển hình được tính cho mô tả cảnh nhất định, chẳng hạn như mô tả cảnh Ambisonics trên Fig.16e và sau đó, phương án thứ nhất về cách kết hợp các tham số được ưu tiên so với phương án thứ hai của Fig.1d. Do đó, theo sáng chế, bộ chuyển đổi định dạng 120 được tạo cấu hình để chuyển đổi định dạng Ambisonics bậc cao hoặc định dạng Ambisonics bậc nhất thành định dạng B, trong đó định dạng Ambisonics bậc cao được cắt bớt trước khi chuyển đổi sang định dạng B.

Trong phương án khác, bộ chuyển đổi định dạng được tạo cấu hình để chiếu đối tượng hoặc kênh trên sóng hài hình cầu ở vị trí tham chiếu để thu được các tín hiệu được chiếu, và trong đó bộ kết hợp định dạng được tạo cấu hình để kết hợp các tín hiệu chiếu để thu được các hệ số định dạng B, trong đó đối tượng hoặc kênh được đặt trong không gian tại vị trí xác định và có khoảng cách riêng tùy chọn từ vị trí tham chiếu. Quy trình này đặc biệt hoạt động tốt cho việc chuyển đổi các tín hiệu đối tượng hoặc các tín hiệu đa kênh thành các tín hiệu Ambisonics bậc nhất hoặc bậc cao.

Ngoài ra, bộ chuyển đổi định dạng 120 được tạo cấu hình để thực hiện phân tích DirAC bao gồm phân tích thời gian-tần số của các thành phần định dạng B và xác định các vectơ áp suất và vận tốc và trong đó bộ kết hợp định dạng sau đó được tạo cấu hình để kết hợp các vectơ áp suất/vận tốc khác nhau và trong đó bộ kết hợp định dạng còn bao gồm bộ phân tích DirAC 180 để lấy siêu dữ liệu DirAC từ dữ liệu áp suất/vận tốc kết hợp.

Trong phương án khác, bộ chuyển đổi định dạng được tạo cấu hình để trích xuất các tham số DirAC trực tiếp từ siêu dữ liệu đối tượng của định dạng đối tượng âm thanh thành định dạng thứ nhất hoặc thứ hai, trong đó vectơ áp suất cho sự biểu diễn

DirAC là tín hiệu dạng sóng đối tượng và hướng được suy ra từ vị trí đối tượng trong không gian hoặc độ khuếch tán được đưa ra trực tiếp trong siêu dữ liệu đối tượng hoặc được đặt thành giá trị mặc định, chẳng hạn như giá trị không.

Trong phương án khác, bộ chuyển đổi định dạng được tạo cấu hình để chuyển đổi các tham số DirAC được suy ra từ định dạng dữ liệu đối tượng thành dữ liệu áp suất/vận tốc và bộ kết hợp định dạng được tạo cấu hình để kết hợp dữ liệu áp suất/vận tốc với dữ liệu áp suất/vận tốc lấy từ mô tả khác nhau của một hoặc nhiều đối tượng âm thanh khác nhau.

Tuy nhiên, trong triển khai được ưu tiên được minh họa liên quan đến Fig.1c và Fig.1d, bộ kết hợp định dạng được tạo cấu hình để kết hợp trực tiếp các tham số DirAC được suy ra bởi bộ chuyển đổi định dạng 120 sao cho cảnh âm thanh kết hợp được tạo bởi khối 140 của Fig.1a đã là kết quả cuối cùng và bộ phân tích DirAC 180 được minh họa trên Fig.1a là không cần thiết, vì đầu ra dữ liệu của bộ kết hợp định dạng 140 đã có trong định dạng DirAC.

Trong triển khai tiếp theo, bộ chuyển đổi định dạng 120 đã bao gồm bộ phân tích DirAC cho Ambisonics bậc nhất hoặc định dạng đầu vào Ambisonics bậc cao hoặc định dạng tín hiệu đa kênh. Hơn nữa, bộ chuyển đổi định dạng bao gồm bộ chuyển đổi siêu dữ liệu để chuyển đổi siêu dữ liệu đối tượng thành siêu dữ liệu DirAC, và bộ chuyển đổi siêu dữ liệu đó ví dụ được minh họa trên Fig.1f ở 150 mà một lần nữa hoạt động trên phép phân tích thời gian/tần số trong khối 121 và tính toán năng lượng trên mỗi dải trên mỗi khung thời gian được minh họa ở 147, hướng đến được minh họa ở khối 148 của Fig.1f và độ khuếch tán được minh họa ở khối 149 của Fig.1f. Và, siêu dữ liệu được kết hợp bởi bộ kết hợp 144 để kết hợp các luồng siêu dữ liệu DirAC riêng lẻ, tốt hơn là bằng phép cộng có trọng số như được minh họa mẫu bằng một trong hai phương án của phương án theo Fig.1d.

Các tín hiệu kênh đa kênh có thể được chuyển đổi trực tiếp sang định dạng B. Định dạng B thu được sau đó có thể được xử lý bằng DirAC thông thường. Fig.1g minh họa phép chuyển đổi 127 sang định dạng B và phép xử lý DirAC 180 tiếp theo.

Tài liệu tham khảo [3] phác thảo các cách để thực hiện chuyển đổi từ tín hiệu đa kênh sang định dạng B. Về nguyên tắc, việc chuyển đổi tín hiệu âm thanh đa kênh

thành định dạng B rất đơn giản: loa phóng thanh ảo được xác định là ở các vị trí khác nhau của sơ đồ bố trí loa phóng thanh. Ví dụ đối với sơ đồ bố trí 5.0, loa phóng thanh được đặt trên mặt phẳng ngang ở các góc phương vị $\pm 30^\circ$ và $\pm 110^\circ$. Sau đó, micro định dạng B ảo được xác định là nằm ở giữa các loa phóng thanh, và phép ghi ảo được thực hiện. Do đó, kênh W được tạo bằng cách tính tổng tất cả các kênh loa phóng thanh của tệp âm thanh 5.0. Sau đó, có thể tóm tắt quá trình nhận W và các hệ số định dạng B khác như sau:

$$W = \sum_{i=1}^k \sqrt{\frac{1}{2}} w_i s_i$$

$$X = \sum_{i=1}^k w_i s_i (\cos(\theta_i) \cos(\varphi_i))$$

$$Y = \sum_{i=1}^k w_i s_i (\sin(\theta_i) \cos(\varphi_i))$$

$$Z = \sum_{i=1}^k w_i s_i (\sin(\varphi_i))$$

trong đó s_i là các tín hiệu đa kênh nằm trong không gian tại các vị trí loa phóng thanh được xác định bởi góc phương vị θ_i và góc nâng φ_i của mỗi loa phóng thanh và w_i là hàm trọng số của khoảng cách. Nếu khoảng cách không có sẵn hoặc đơn giản là bị bỏ qua, thì $w_i = 1$. Mặc dù vậy, kỹ thuật đơn giản này bị hạn chế vì đây là quá trình không thể đảo ngược. Ngoài ra, do loa phóng thanh thường được phân phối không đồng đều, nên cũng có sự sai lệch trong ước tính được thực hiện bởi phép phân tích DirAC tiếp theo theo hướng có mật độ loa phóng thanh cao nhất. Ví dụ, trong sơ đồ bố trí 5.1, sẽ thiên về phía trước vì có nhiều loa phóng thanh ở phía trước hơn ở phía sau.

Để giải quyết vấn đề này, kỹ thuật tiếp theo đã được đề xuất trong tài liệu tham khảo [3] để xử lý tín hiệu đa kênh 5.1 với DirAC. Sơ đồ mã hóa cuối cùng sau đó nhìn sẽ như được minh họa trên Fig.1h thể hiện bộ chuyển đổi định dạng B 127, bộ phân tích DirAC 180 như được mô tả chung về phần tử 180 trên Fig.1 và các phần tử khác 190, 1000, 160, 170, 1020, và/hoặc 220, 240.

Trong phương án khác, giao diện đầu ra 200 được tạo cấu hình để thêm mô tả đối tượng riêng biệt cho đối tượng âm thanh vào định dạng kết hợp, trong đó mô tả đối tượng bao gồm ít nhất một trong số hướng, khoảng cách, độ khuếch tán hoặc bất kỳ thuộc tính đối tượng nào khác, trong đó đối tượng này có hướng duy nhất trong tất cả các dải tần và hoặc là tĩnh hoặc di chuyển chậm hơn ngưỡng vận tốc.

Đặc tính này còn được xây dựng chi tiết hơn đối với khía cạnh thứ tư của sáng chế được thảo luận liên quan đến Fig.4a và Fig.4b.

Phương án mã hóa thứ nhất: Kết hợp và xử lý các sự biểu diễn âm thanh khác nhau thông qua định dạng B hoặc sự biểu diễn tương đương.

Việc thực hiện thứ nhất của bộ mã hóa được hình dung là có thể đạt được bằng cách chuyển đổi tất cả định dạng đầu vào thành định dạng B kết hợp như được mô tả trên Fig.11.

Fig.11: Tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau theo định dạng B kết hợp

Do DirAC ban đầu được thiết kế để phân tích tín hiệu định dạng B, nên hệ thống chuyển đổi các định dạng âm thanh khác nhau thành tín hiệu định dạng B kết hợp. Các định dạng trước tiên được chuyển đổi riêng lẻ 120 thành tín hiệu định dạng B trước khi được kết hợp với nhau bằng cách tính tổng các thành phần định dạng B W, X, Y, Z của chúng. Các thành phần Ambisonics bậc nhất (First Order Ambisonics - FOA) có thể được chuẩn hóa và sắp xếp lại thành định dạng B. Giả sử FOA ở định dạng ACN/N3D, bốn tín hiệu của đầu vào định dạng B thu được bằng:

$$\begin{cases} W = Y_0^0 \\ X = \sqrt{\frac{2}{3}} Y_1^1 \\ Y = \sqrt{\frac{2}{3}} Y_1^{-1} \\ Z = \sqrt{\frac{2}{3}} Y_1^0 \end{cases}$$

Trong đó Y_m^l biểu thị thành phần Ambisonics của thứ tự l và chỉ số m , $-l \leq m \leq +l$. Do các thành phần FOA có đầy đủ trong định dạng Ambisonics bậc

cao hơn, định dạng HOA chỉ cần được cắt bớt trước khi được chuyển đổi thành định dạng B.

Vì các đối tượng và kênh đã xác định vị trí trong không gian, nên có thể chiếu từng đối tượng và kênh riêng lẻ trên Harmonics hình cầu (spherical Harmonics - SH) tại vị trí trung tâm chẳng hạn như vị trí ghi hoặc tham chiếu. Tổng các phép chiếu cho phép kết hợp các đối tượng khác nhau và nhiều kênh ở định dạng B đơn và sau đó có thể được xử lý bằng phép phân tích DirAC. Các hệ số định dạng B (W, X, Y, Z) sau đó được đưa ra bởi:

$$W = \sum_{i=1}^k \sqrt{\frac{1}{2}} w_i s_i$$

$$X = \sum_{i=1}^k w_i s_i (\cos(\theta_i) \cos(\varphi_i))$$

$$Y = \sum_{i=1}^k w_i s_i (\sin(\theta_i) \cos(\varphi_i))$$

$$Z = \sum_{i=1}^k w_i s_i (\sin(\varphi_i))$$

trong đó s_i là các tín hiệu độc lập nằm trong không gian tại các vị trí được xác định bởi góc phương vị θ_i và góc nâng φ_i , và w_i là hàm trọng số của khoảng cách. Nếu khoảng cách không có sẵn hoặc đơn giản là bị bỏ qua, thì $w_i = 1$. Ví dụ, các tín hiệu độc lập có thể tương ứng với các đối tượng âm thanh được đặt tại vị trí đã cho hoặc tín hiệu được liên kết với kênh loa phóng thanh ở vị trí xác định.

Trong các ứng dụng yêu cầu sự biểu diễn Ambisonics ở các bậc cao hơn bậc nhất, việc tạo các hệ số Ambisonics được trình bày ở trên cho bậc nhất được mở rộng bằng cách xem xét thêm các thành phần bậc cao hơn.

Bộ tạo kênh vận chuyển 160 có thể trực tiếp nhận tín hiệu đa kênh, các tín hiệu dạng sóng đối tượng, và các thành phần Ambisonics bậc cao hơn. Bộ tạo kênh vận chuyển sẽ giảm số lượng kênh đầu vào để truyền bằng cách trộn giảm chúng. Các kênh có thể được trộn với nhau như trong âm thanh vòm MPEG trong trộn giảm đơn âm hoặc âm thanh nổi, trong khi các tín hiệu dạng sóng đối tượng có thể được tổng hợp

theo cách thụ động thành trộn giảm đơn âm. Ngoài ra, từ Ambisonics bậc cao hơn, có thể trích xuất sự biểu diễn bậc thấp hơn hoặc tạo bằng cách tạo chùm trộn giảm âm thanh nổi hoặc bất kỳ phần nào khác của không gian. Nếu các trộn giảm thu được từ định dạng đầu vào khác nhau tương thích với nhau, chúng có thể được kết hợp với nhau bằng thao tác cộng đơn giản.

Ngoài ra, bộ tạo kênh vận chuyển 160 có thể nhận định dạng B kết hợp giống như định dạng được chuyển đến phép phân tích DirAC. Trong trường hợp này, tập hợp con các thành phần hoặc kết quả của phép tạo chùm (hoặc phép xử lý khác) tạo thành các kênh vận chuyển sẽ được mã hóa và truyền đến bộ giải mã. Trong hệ thống được đề xuất, cần có phép mã hóa âm thanh thông thường mà có thể dựa trên, nhưng không bị giới hạn, bộ mã hóa-giải mã 3GPP EVS tiêu chuẩn. 3GPP EVS là lựa chọn bộ mã hóa-giải mã được ưu tiên vì khả năng mã hóa tín hiệu tiếng nói hoặc âm nhạc ở tốc độ bit thấp với chất lượng cao trong khi cần độ trễ tương đối thấp cho phép truyền thông theo thời gian thực.

Ở tốc độ bit rất thấp, số lượng kênh cần truyền phải được giới hạn ở một và do đó chỉ tín hiệu micro đa hướng W của định dạng B được truyền đi. Nếu tốc độ bit cho phép, số lượng kênh vận chuyển có thể tăng lên bằng cách chọn tập hợp con các thành phần định dạng B. Ngoài ra, các tín hiệu định dạng B có thể được kết hợp thành bộ tạo chùm 160 được định hướng theo các phân vùng cụ thể của không gian. Ví dụ như hai đường caddioit có thể được thiết kế để chỉ ở hai hướng ngược nhau, ví dụ ở bên trái và bên phải của cảnh không gian:

$$\begin{cases} L = \sqrt{2}W + Y \\ R = \sqrt{2}W - Y \end{cases}$$

Hai kênh âm thanh nổi L và R này sau đó có thể được mã hóa một cách hiệu quả 170 bằng phép mã hóa âm thanh nổi chung. Hai tín hiệu sau đó sẽ được khai thác hợp lý bởi phép tổng hợp DirAC ở phía bộ giải mã để kết xuất cảnh âm thanh. Các phép tạo chùm khác có thể được hình dung, ví dụ, micro caddioit ảo có thể được chỉ về phía hướng bất kỳ của góc phương vị θ và góc nâng φ đã cho:

$$C = \sqrt{2}W + \cos(\theta) \cos(\varphi) X + \sin(\theta) \cos(\varphi) Y + \sin(\varphi) Z$$

Các cách khác để tạo các kênh truyền có thể được hình dung là mang nhiều thông tin không gian hơn kênh truyền đơn âm.

Thay vào đó, 4 hệ số của định dạng B có thể được truyền trực tiếp. Trong trường hợp đó, siêu dữ liệu DirAC có thể được trích xuất trực tiếp ở phía bộ giải mã mà không cần truyền thêm thông tin cho siêu dữ liệu không gian.

Fig.12 thể hiện phương pháp thay thế khác để kết hợp các định dạng đầu vào khác nhau. Fig.12 còn là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp trong miền áp suất/vận tốc;

Cả thành phần tín hiệu đa kênh và Ambisonics đều được đưa vào phép phân tích DirAC 123, 124. Đối với mỗi định dạng đầu vào, phép phân tích DirAC được thực hiện gồm phép phân tích thời gian-tần số của các thành phần định dạng B $w^i(n), x^i(n), y^i(n), z^i(n)$ và phép xác định các vector áp suất và vận tốc:

$$P^i(n, k) = W^i(k, n)$$

$$U^i(n, k) = X^i(k, n)e_x + Y^i(k, n)e_y + Z^i(k, n)e_z$$

trong đó i là chỉ số của đầu vào và, k và n là các chỉ số thời gian và tần số của ô thời gian-tần số, và e_x, e_y, e_z biểu diễn các vector đơn vị Đề-các.

$P(n, k)$ và $U(n, k)$ là cần thiết để tính toán các tham số DirAC, cụ thể là DOA và độ khuếch tán. Bộ kết hợp siêu dữ liệu DirAC có thể khai thác rằng N nguồn cùng phát dẫn đến sự kết hợp tuyến tính giữa các áp suất và vận tốc phần tử sẽ được đo khi chúng được phát một mình. Các đại lượng kết hợp sau đó được suy ra bởi:

$$P(n, k) = \sum_{i=1}^N P^i(n, k)$$

$$U(n, k) = \sum_{i=1}^N U^i(n, k)$$

Các tham số DirAC kết hợp được tính 143 thông qua phép tính của vector cường độ kết hợp:

$$I(k, n) = \frac{1}{2} \Re\{P(k, n) \cdot \overline{U(k, n)}\},$$

trong đó ψ biểu thị liên hợp phức. Độ khuếch tán của trường âm thanh kết hợp được cho bởi:

$$\psi(k, n) = 1 - \frac{\|E\{I(k, n)\}\|}{cE\{E(k, n)\}}$$

trong đó $E\{.\}$ biểu thị toán tử trung bình theo thời gian, c tốc độ âm thanh và $E(k, n)$ năng lượng trường âm thanh được cho bởi:

$$E(k, n) = \frac{\rho_0}{4} \|U(k, n)\|^2 + \frac{1}{\rho_0 c^2} |P(k, n)|^2$$

Hướng đến (DOA) được biểu thị bằng phương tiện của vector đơn vị $e_{DOA}(k, n)$, được xác định là

$$e_{DOA}(k, n) = -\frac{I(k, n)}{\|I(k, n)\|}$$

Nếu đối tượng âm thanh là đầu vào, các tham số DirAC có thể được trích xuất trực tiếp từ siêu dữ liệu đối tượng trong khi vector áp suất $P^i(k, n)$ là tín hiệu cốt yếu (dạng sóng) của đối tượng. Chính xác hơn, hướng được lấy thẳng từ vị trí đối tượng trong không gian, trong khi độ khuếch tán được cho trực tiếp trong siêu dữ liệu đối tượng hoặc - nếu không có sẵn - có thể được đặt mặc định thành không. Từ các tham số DirAC, các vector áp suất và vận tốc được cho trực tiếp bởi:

$$\hat{P}^i(k, n) = \sqrt{1 - \psi^i(k, n)} P^i(k, n)$$

$$\hat{U}^i(k, n) = -\frac{1}{\rho_0 c} \hat{P}^i(k, n) \cdot e_{DOA}^i(k, n)$$

Sự kết hợp của các đối tượng hoặc sự kết hợp của một đối tượng với các định dạng đầu vào khác nhau sau đó thu được bằng cách tính tổng các vector áp suất và vận tốc như đã giải thích ở trên.

Tóm lại, sự kết hợp của các đóng góp đầu vào khác nhau (Ambisonics, kênh, đối tượng) được thực hiện trong miền áp suất/vận tốc và sau đó kết quả được chuyển đổi thành các tham số DirAC hướng/độ khuếch tán. Việc hoạt động ở miền áp suất/vận tốc theo lý thuyết là tương đương với hoạt động ở định dạng B. Lợi ích chính của phương án này so với phương án trước đó là khả năng tối ưu hóa phép phân tích

DirAC theo từng định dạng đầu vào như được đề xuất trong tài liệu [3] cho định dạng vòm 5.1.

Hạn chế chính của sự kết hợp trong định dạng B kết hợp hoặc miền áp suất/vận tốc như vậy là phép chuyển đổi xảy ra ở đầu trước của chuỗi xử lý là đường thất nghẽn đối với toàn bộ hệ thống mã hóa. Thật vậy, việc chuyển đổi các sự biểu diễn âm thanh từ Ambisonics bậc cao, các đối tượng hoặc các kênh thành tín hiệu định dạng B (bậc nhất) đã làm mất rất nhiều độ phân giải không gian mà không thể phục hồi sau đó.

Phương án mã hóa thứ 2: kết hợp và xử lý trong miền DirAC

Để tránh các giới hạn của việc chuyển đổi tất cả các định dạng đầu vào thành tín hiệu định dạng B kết hợp, phương án này đề xuất lấy các tham số DirAC trực tiếp từ định dạng ban đầu và sau đó kết hợp chúng trong miền tham số DirAC. Tổng quan chung về hệ thống như vậy được đưa ra trên Fig.13. Fig.13 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau trong miền DirAC với khả năng thao tác đối tượng ở phía bộ giải mã.

Sau đây, cũng có thể coi các kênh riêng lẻ của tín hiệu đa kênh là đầu vào đối tượng âm thanh cho hệ thống mã hóa. Siêu dữ liệu đối tượng sau đó tính theo thời gian và biểu thị vị trí loa phóng thanh và khoảng cách liên quan đến vị trí người nghe.

Mục tiêu của giải pháp thay thế này là để tránh sự kết hợp có hệ thống của các định dạng đầu vào khác nhau thành định dạng B kết hợp hoặc sự biểu diễn tương đương. Mục đích là để tính toán các tham số DirAC trước khi kết hợp chúng. Phương pháp sau đó tránh bất kỳ sai lệch nào trong việc ước tính hướng và độ khuếch tán do sự kết hợp. Hơn nữa, nó có thể khai thác tối ưu các đặc tính của từng sự biểu diễn âm thanh trong quá trình phân tích DirAC hoặc trong khi xác định các tham số DirAC.

Sự kết hợp của siêu dữ liệu DirAC xảy ra sau khi xác định 125, 126, 126a cho từng định dạng đầu vào các tham số DirAC, độ khuếch tán, hướng cũng như áp suất trong các kênh vận chuyển được truyền. Phép phân tích DirAC có thể ước tính các tham số từ định dạng B trung gian, thu được bằng cách chuyển đổi định dạng đầu vào như đã giải thích trước đó. Ngoài ra, các tham số DirAC có thể được ước tính một cách thuận lợi mà không cần thông qua định dạng B mà trực tiếp từ định dạng đầu vào,

điều này có thể cải thiện hơn nữa độ chính xác của việc ước tính. Ví dụ, trong tài liệu [7] đề xuất việc ước tính độ khuếch tán trực tiếp từ Ambisonics bậc cao hơn. Trong trường hợp các đối tượng âm thanh, bộ chuyển đổi siêu dữ liệu đơn giản 150 trên Fig.15 có thể trích xuất từ hướng siêu dữ liệu đối tượng và độ khuếch tán cho từng đối tượng.

Sự kết hợp 144 của một số luồng siêu dữ liệu Dirac thành một luồng siêu dữ liệu DirAC kết hợp duy nhất có thể đạt được như đề xuất trong tài liệu [4]. Đối với một số nội dung, tốt hơn nhiều là ước tính trực tiếp các tham số DirAC từ định dạng ban đầu thay vì chuyển đổi nó sang định dạng B kết hợp trước khi thực hiện phân tích DirAC. Thật vậy, các tham số, hướng và độ khuếch tán, có thể bị sai lệch khi chuyển sang định dạng B [3] hoặc khi kết hợp các nguồn khác nhau.

Một cách khác đơn giản hơn có thể lấy trung bình các tham số của các nguồn khác nhau bằng cách gán trọng số chúng theo năng lượng của chúng:

$$\psi(k, n) = \frac{1}{\sum_{i=1}^N E^i(k, n)} \sum_{i=1}^N E^i(k, n) \psi^i(k, n)$$

$$e_{DOA}(k, n) = \frac{1}{\sum_{i=1}^N (1 - \psi^i(k, n)) E^i(k, n)} \sum_{i=1}^N (1 - \psi^i(k, n)) E^i(k, n) e_{DOA}^i(k, n)$$

Đối với mỗi đối tượng, có khả năng vẫn gửi hướng riêng và khoảng cách tùy chọn, độ khuếch tán hoặc bất kỳ thuộc tính đối tượng nào khác có liên quan như một phần của dòng bit được truyền từ bộ mã hóa đến bộ giải mã (xem ví dụ Fig.4a, Fig.4b). Thông tin phụ này sẽ làm phong phú thêm siêu dữ liệu DirAC kết hợp và sẽ cho phép bộ giải mã khôi phục và hoặc thao tác đối tượng riêng lẻ. Do đối tượng có một hướng duy nhất trong tất cả các dải tần và có thể được coi là tĩnh hoặc di chuyển chậm, thông tin bổ sung cần được cập nhật ít thường xuyên hơn các tham số DirAC khác và sẽ chỉ có tốc độ bit bổ sung rất thấp.

Ở phía bộ giải mã, phép lọc theo hướng có thể được thực hiện như được trình bày trong tài liệu [5] để thao tác với các đối tượng. Phép lọc theo hướng dựa trên kỹ thuật làm suy giảm phổ thời gian ngắn. Nó được thực hiện trong miền phổ bằng hàm khuếch đại không pha, phụ thuộc vào hướng của các đối tượng. Hướng có thể được

chứa trong dòng bit nếu hướng của các đối tượng được truyền dưới dạng thông tin phụ. Mặt khác, hướng cũng có thể được đưa ra tương tác bởi người dùng.

Phương án thứ 3: kết hợp ở phía bộ giải mã

Ngoài ra, việc kết hợp có thể được thực hiện ở phía bộ giải mã. Fig.14 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau ở phía bộ giải mã thông qua bộ kết hợp siêu dữ liệu DirAC. Trên Fig.14, sơ đồ mã hóa dựa trên DirAC hoạt động ở tốc độ bit cao hơn trước đó nhưng cho phép truyền siêu dữ liệu DirAC riêng lẻ. Các luồng siêu dữ liệu DirAC khác nhau được kết hợp 144 như ví dụ được đề xuất trong tài liệu [4] trong bộ giải mã trước khi tổng hợp DirAC 220, 240. Bộ kết hợp siêu dữ liệu DirAC 144 cũng có thể thu được vị trí của đối tượng riêng lẻ để thao tác tiếp theo của đối tượng trong phép phân tích DirAC.

Fig.15 là tổng quan hệ thống của bộ mã hóa/bộ giải mã dựa trên DirAC kết hợp các định dạng đầu vào khác nhau ở phía bộ giải mã trong phép tổng hợp DirAC. Nếu tốc độ bit cho phép, hệ thống có thể được nâng cao hơn nữa như được đề xuất trên Fig.15 bằng cách gửi cho từng thành phần đầu vào (FOA/HOA, MC, đối tượng) tín hiệu trộn giảm của chính nó cùng với siêu dữ liệu DirAC được liên kết của nó. Tuy nhiên, các luồng DirAC khác nhau có chung phép tổng hợp DirAC 220, 240 tại bộ giải mã để giảm độ phức tạp.

Fig.2a minh họa khái niệm để thực hiện phép tổng hợp nhiều cảnh âm thanh theo khía cạnh thứ hai, khác nữa của sáng chế. Thiết bị được minh họa trên Fig.2a bao gồm giao diện đầu vào 100 để nhận mô tả DirAC thứ nhất của cảnh thứ nhất và để nhận mô tả DirAC thứ hai của cảnh thứ hai và một hoặc nhiều kênh vận chuyển.

Ngoài ra, bộ tổng hợp DirAC 220 được đề xuất để tổng hợp nhiều cảnh âm thanh trong miền phổ để thu được tín hiệu âm thanh miền phổ biểu diễn nhiều cảnh âm thanh. Ngoài ra, bộ chuyển đổi phổ-thời gian 214 được đề xuất để chuyển đổi tín hiệu âm thanh miền phổ thành miền thời gian để xuất ra tín hiệu âm thanh miền thời gian mà ví dụ có thể được phát ra bởi loa. Trong trường hợp này, bộ tổng hợp DirAC được tạo cấu hình để thực hiện kết xuất tín hiệu đầu ra loa phóng thanh. Ngoài ra, tín hiệu âm thanh có thể là tín hiệu âm thanh nổi có thể được phát ra tai nghe. Một lần nữa,

thay vào đó, đầu ra tín hiệu âm thanh của bộ chuyển đổi phổ-thời gian 214 có thể là mô tả trường âm thanh định dạng B. Tất cả các tín hiệu này, nghĩa là các tín hiệu loa phóng thanh cho nhiều hơn hai kênh, các tín hiệu tai nghe hoặc các mô tả trường âm thanh là tín hiệu miền thời gian để xử lý thêm như xuất ra bởi loa hoặc tai nghe hoặc để truyền hoặc lưu trữ trong trường hợp các mô tả trường âm thanh chẳng hạn như các tín hiệu Ambisonics bậc nhất hoặc các tín hiệu Ambisonics bậc cao hơn.

Hơn nữa, thiết bị theo Fig.2a còn bao gồm giao diện người dùng 260 để điều khiển bộ tổng hợp DirAC 220 trong miền phổ. Ngoài ra, một hoặc nhiều kênh vận chuyển có thể được cung cấp cho giao diện đầu vào 100 được sử dụng cùng với các mô tả DirAC thứ nhất và thứ hai, trong trường hợp này, các mô tả theo tham số cung cấp, cho mỗi ô thời gian/tần số, thông tin hướng đến và, một cách tùy ý, còn có thông tin độ khuếch tán.

Thông thường, hai mô tả DirAC khác nhau được nhập vào giao diện 100 trên Fig.2a mô tả hai cảnh âm thanh khác nhau. Trong trường hợp này, bộ tổng hợp DirAC 220 được tạo cấu hình để thực hiện việc kết hợp các cảnh âm thanh này. Một phương án khác của phép kết hợp được minh họa trên Fig.2b. Ở đây, bộ kết hợp cảnh 221 được tạo cấu hình để kết hợp hai mô tả DirAC trong miền tham số, tức là, các tham số được kết hợp để thu được các tham số hướng đến (DoA) kết hợp và các tham số độ khuếch tán một cách tùy chọn ở đầu ra của khối 221. Dữ liệu này sau đó được đưa vào bộ kết xuất DirAC 222 mà nhận, ngoài ra, một hoặc nhiều kênh vận chuyển để chuyển sang các kênh để thu được tín hiệu âm thanh miền phổ 222. Sự kết hợp của dữ liệu tham số DirAC tốt hơn là được thực hiện như được minh họa trên Fig.1d và, như được mô tả liên quan đến hình vẽ này và, đặc biệt, đối với phương án thứ nhất.

Nếu ít nhất một trong hai mô tả được đưa vào bộ kết hợp cảnh 221 bao gồm các giá trị độ khuếch tán bằng 0 hoặc không có giá trị độ khuếch tán nào, thì, ngoài ra, phương án thứ hai có thể được áp dụng cũng như đã thảo luận trong ngữ cảnh của Fig.1d.

Một phương án khác được minh họa trên Fig.2c. Trong quy trình này, các mô tả DirAC riêng lẻ được kết xuất bằng phương tiện của bộ kết xuất DirAC thứ nhất 223 cho mô tả thứ nhất và bộ kết xuất DirAC thứ hai cho mô tả thứ hai và ở đầu ra của các

khối 223 và 224, có tín hiệu âm thanh miền phổ thứ nhất và thứ hai, và các tín hiệu âm thanh miền phổ thứ nhất và thứ hai này được kết hợp trong bộ kết hợp 225 để thu được, ở đầu ra của bộ kết hợp 225, tín hiệu kết hợp miền phổ.

Ví dụ, bộ kết xuất DirAC thứ nhất 223 và bộ kết xuất DirAC thứ hai được tạo cấu hình để tạo tín hiệu âm thanh nổi có kênh trái L và kênh phải R. Sau đó, bộ kết hợp 225 được tạo cấu hình để kết hợp kênh trái từ khối 223 và kênh phải từ khối 224 để thu được kênh trái kết hợp. Ngoài ra, kênh phải từ khối 223 được thêm vào với kênh phải từ khối 224, và kết quả là kênh bên phải được kết hợp ở đầu ra của khối 225.

Đối với các kênh riêng lẻ của tín hiệu đa kênh, quy trình tương tự được thực hiện, tức là, các kênh riêng lẻ được thêm vào một cách riêng lẻ, do đó, luôn luôn có cùng một kênh từ bộ kết xuất DirAC 223 được thêm vào cùng một kênh tương ứng của bộ kết xuất DirAC khác, v.v. Quy trình tương tự cũng được thực hiện đối với, ví dụ, tín hiệu Ambisonics định dạng B hoặc bậc cao hơn. Ví dụ, khi bộ kết xuất DirAC thứ nhất 223 xuất ra các tín hiệu W, X, Y, Z, và bộ kết xuất DirAC thứ hai 224 xuất ra định dạng tương tự, thì bộ kết hợp kết hợp hai tín hiệu đa hướng để thu được tín hiệu đa hướng kết hợp W, và quy trình tương tự cũng được thực hiện cho các thành phần tương ứng để cuối cùng thu được thành phần kết hợp X, Y và Z.

Hơn nữa, như đã được phác thảo liên quan đến Fig.2a, giao diện đầu vào được tạo cấu hình để nhận siêu dữ liệu đối tượng âm thanh bổ sung cho đối tượng âm thanh. Đối tượng âm thanh này đã có thể đã có trong mô tả DirAC thứ nhất hoặc thứ hai hoặc tách biệt với mô tả DirAC thứ nhất và thứ hai. Trong trường hợp này, bộ tổng hợp DirAC 220 được tạo cấu hình để xử lý có chọn lọc siêu dữ liệu đối tượng âm thanh bổ sung hoặc dữ liệu đối tượng liên quan đến siêu dữ liệu đối tượng âm thanh bổ sung này, ví dụ, thực hiện lọc theo hướng dựa trên siêu dữ liệu đối tượng âm thanh bổ sung hoặc dựa trên thông tin hướng người dùng đã cho thu được từ giao diện người dùng 260. Thay vào đó hoặc ngoài ra, và như được minh họa trên Fig.2d, bộ tổng hợp DirAC 220 được cấu hình để thực hiện, trong miền phổ, chức năng khuếch đại pha không, chức năng khuếch đại pha không phụ thuộc vào hướng của đối tượng âm thanh, trong đó hướng có trong dòng bit nếu các hướng của các đối tượng được truyền dưới dạng thông tin bên, hoặc trong đó hướng nhận được từ giao diện người dùng 260. Đầu

vào siêu dữ liệu đối tượng âm thanh bổ sung vào giao diện 100 dưới dạng tính năng tùy chọn trên Fig.2a phản ánh khả năng vẫn gửi, cho mỗi đối tượng riêng lẻ theo hướng riêng và khoảng cách tùy ý, độ khuếch tán và bất kỳ thuộc tính đối tượng liên quan nào khác như một phần của dòng bit được truyền từ bộ mã hóa đến bộ giải mã. Do đó, siêu dữ liệu đối tượng âm thanh bổ sung có thể liên quan đến đối tượng đã có trong mô tả DirAC thứ nhất hoặc trong mô tả DirAC thứ hai hoặc là một đối tượng bổ sung không có trong mô tả DirAC thứ nhất và trong mô tả DirAC thứ hai.

Tuy nhiên, nên ưu tiên có siêu dữ liệu đối tượng âm thanh bổ sung theo kiểu DirAC, nghĩa là thông tin hướng đến và, một cách tùy ý, thông tin khuếch tán mặc dù các đối tượng âm thanh điển hình có độ khuếch tán bằng không, nghĩa là tập trung vào vị trí thực tế của chúng làm cho hướng đến tập trung và cụ thể là không đổi trên tất cả các dải tần và có nghĩa là, liên quan đến tốc độ khung, hoặc tĩnh hoặc di chuyển chậm. Do đó, do đối tượng có một hướng duy nhất trong tất cả các dải tần và có thể được coi là tĩnh hoặc di chuyển chậm, thông tin bổ sung cần được cập nhật ít thường xuyên hơn các tham số DirAC khác và do đó sẽ chỉ có tốc độ bit bổ sung rất thấp. Ví dụ, trong khi các mô tả DirAC thứ nhất và thứ hai có dữ liệu DoA và dữ liệu độ khuếch tán cho từng dải phổ và cho từng khung, siêu dữ liệu đối tượng âm thanh bổ sung chỉ cần một dữ liệu DoA duy nhất cho tất cả các dải tần và dữ liệu này chỉ cho mọi khung thứ hai hoặc, tốt hơn là mọi khung thứ ba, thứ tư, thứ năm hoặc thậm chí là thứ mười trong phương án được ưu tiên.

Hơn nữa, liên quan đến phép lọc theo hướng được thực hiện trong bộ tổng hợp DirAC 220 thường có trong bộ giải mã ở phía bộ giải mã của hệ thống mã hóa/giải mã, bộ tổng hợp DirAC có thể, trong phương án của Fig.2b, thực hiện phép lọc theo hướng trong miền tham số trước khi kết hợp cảnh hoặc một lần nữa thực hiện phép lọc định hướng sau khi kết hợp cảnh. Tuy nhiên, trong trường hợp này, phép lọc theo hướng được áp dụng cho cảnh kết hợp thay vì các mô tả riêng lẻ.

Hơn nữa, trong trường hợp đối tượng âm thanh không có trong mô tả thứ nhất hoặc mô tả thứ hai, nhưng có trong siêu dữ liệu đối tượng âm thanh của chính nó, việc lọc theo hướng như được minh họa bởi bộ thao tác có chọn lọc chỉ có thể được áp dụng một cách chọn lọc đối tượng âm thanh bổ sung, mà siêu dữ liệu đối tượng âm

thanh bổ sung tồn tại mà không ảnh hưởng đến mô tả DirAC thứ nhất hoặc thứ hai hoặc mô tả DirAC kết hợp. Đối với chính đối tượng âm thanh, tồn tại kênh vận chuyển riêng biệt biểu diễn tín hiệu dạng sóng đối tượng hoặc tín hiệu dạng sóng đối tượng có trong kênh vận chuyển được trộn giảm.

Ví dụ, việc thao tác có chọn lọc như được minh họa, ví dụ như trên Fig.2b, có thể tiến hành theo cách mà hướng đến nhất định được cho theo hướng của đối tượng âm thanh được đưa ra trên Fig.2d có trong dòng bit làm thông tin phụ hoặc nhận được từ giao diện người dùng. Sau đó, dựa trên thông tin hướng người dùng đã cho hoặc thông tin điều khiển, ví dụ phác thảo rằng, từ hướng nhất định, dữ liệu âm thanh sẽ được tăng cường hoặc bị suy giảm. Do đó, đối tượng (siêu dữ liệu) cho đối tượng đang xem xét được khuếch đại hoặc bị suy giảm.

Trong trường hợp dữ liệu dạng sóng thực tế là dữ liệu đối tượng được đưa vào bộ thao tác có chọn lọc 226 từ bên trái trên Fig.2d, dữ liệu âm thanh sẽ thực sự bị suy giảm hoặc được tăng cường tùy thuộc vào thông tin điều khiển. Tuy nhiên, trong trường hợp dữ liệu đối tượng có thông tin năng lượng tiếp theo, ngoài hướng đến và độ khuếch tán một cách tùy ý hoặc khoảng cách, thì thông tin năng lượng cho đối tượng sẽ bị giảm trong trường hợp suy giảm cần thiết cho đối tượng hoặc thông tin năng lượng sẽ được tăng lên trong trường hợp khuếch đại dữ liệu đối tượng cần thiết.

Do đó, phép lọc có hướng dựa trên kỹ thuật suy giảm phổ thời gian ngắn, và nó được thực hiện ở miền phổ bằng hàm khuếch đại pha không phụ thuộc vào hướng của các đối tượng. Hướng có thể có trong dòng bit nếu hướng của các đối tượng được truyền dưới dạng thông tin phụ. Mặt khác, hướng cũng có thể được đưa ra tương tác bởi người dùng. Đương nhiên, quy trình tương tự không thể chỉ được áp dụng cho từng đối tượng riêng lẻ được cho và phản ánh bởi siêu dữ liệu đối tượng âm thanh bổ sung thường được cung cấp bởi dữ liệu DoA cho tất cả các dải tần số và dữ liệu DoA với tỷ lệ cập nhật thấp đối với tốc độ khung và cũng được cung cấp bởi thông tin năng lượng cho đối tượng, mà bộ lọc theo hướng cũng có thể được áp dụng cho mô tả DirAC thứ nhất độc lập với mô tả DirAC thứ hai hoặc ngược lại hoặc cũng có thể được áp dụng cho mô tả DirAC kết hợp nếu có thể.

Hơn nữa, cần lưu ý rằng tính năng liên quan đến dữ liệu đối tượng âm thanh bổ sung cũng có thể được áp dụng trong khía cạnh thứ nhất của sáng chế được minh họa liên quan đến các hình vẽ từ Fig.1a đến Fig.1f. Sau đó, giao diện đầu vào 100 theo Fig.1a cũng nhận thêm dữ liệu đối tượng âm thanh như đã thảo luận liên quan đến Fig.2a, và bộ kết hợp định dạng có thể được triển khai như bộ tổng hợp DirAC trong miền phổ 220 được điều khiển bởi giao diện người dùng 260.

Hơn nữa, khía cạnh thứ hai của sáng chế như được minh họa trên Fig.2 khác với khía cạnh thứ nhất ở chỗ giao diện đầu vào nhận được hai mô tả DirAC, tức là các mô tả về trường âm thanh có cùng định dạng và do đó, đối với khía cạnh thứ hai, bộ chuyển đổi định dạng 120 của khía cạnh thứ nhất không nhất thiết phải có.

Mặt khác, khi đầu vào vào bộ kết hợp định dạng 140 theo Fig.1a gồm hai mô tả DirAC, thì bộ kết hợp định dạng 140 có thể được triển khai như đã thảo luận về khía cạnh thứ hai được minh họa trên Fig.2a, hoặc, thay vào đó, các thiết bị 220, 240 theo Fig.2a có thể được triển khai như đã thảo luận liên quan đến bộ kết hợp định dạng 140 theo Fig.1a về khía cạnh thứ nhất.

Fig.3a minh họa bộ chuyển đổi dữ liệu âm thanh bao gồm giao diện đầu vào 100 để nhận mô tả đối tượng của đối tượng âm thanh có siêu dữ liệu đối tượng âm thanh. Hơn nữa, sau giao diện đầu vào 100 là bộ chuyển đổi siêu dữ liệu 150 cũng tương ứng với các bộ chuyển đổi siêu dữ liệu 125, 126 được thảo luận liên quan đến khía cạnh thứ nhất của sáng chế để chuyển đổi siêu dữ liệu đối tượng âm thanh thành siêu dữ liệu DirAC. Đầu ra của bộ chuyển đổi âm thanh theo Fig.3a được cấu thành bởi giao diện đầu ra 300 để truyền hoặc lưu trữ siêu dữ liệu DirAC. Ngoài ra, giao diện đầu vào 100 có thể nhận thêm tín hiệu dạng sóng như được minh họa bằng đầu vào mũi tên thứ hai vào giao diện 100. Hơn nữa, giao diện đầu ra 300 có thể được triển khai để đưa ra, điển hình là sự biểu diễn được mã hóa của tín hiệu dạng sóng vào đầu ra tín hiệu đầu ra bởi khối 300. Nếu bộ chuyển đổi dữ liệu âm thanh được tạo cấu hình để chỉ chuyển đổi một mô tả đối tượng duy nhất bao gồm siêu dữ liệu, thì giao diện đầu ra 300 cũng cung cấp mô tả DirAC của đối tượng âm thanh duy nhất này cùng với tín hiệu dạng sóng được mã hóa thông thường như kênh vận chuyển DirAC.

Đặc biệt, siêu dữ liệu đối tượng âm thanh có vị trí đối tượng, và siêu dữ liệu DirAC có hướng đến đối với vị trí tham chiếu xuất phát từ vị trí đối tượng. Cụ thể, bộ chuyển đổi siêu dữ liệu 150, 125, 126 được tạo cấu hình để chuyển đổi các tham số DirAC được suy ra từ định dạng dữ liệu đối tượng thành dữ liệu áp suất/vận tốc, và bộ chuyển đổi siêu dữ liệu được tạo cấu hình để áp dụng phép phân tích DirAC cho dữ liệu áp suất/vận tốc này, ví dụ như được minh họa bằng lưu đồ của Fig.3c gồm khối 302, 304, 306. Với mục đích này, đầu ra các tham số DirAC theo khối 306 có chất lượng tốt hơn các tham số DirAC được suy ra từ siêu dữ liệu đối tượng thu được bởi khối 302, tức là, là các tham số DirAC nâng cao. Fig.3b minh họa việc chuyển đổi vị trí cho một đối tượng thành hướng đến đối với vị trí tham chiếu cho đối tượng cụ thể.

Fig.3f minh họa sơ đồ nguyên lý để giải thích chức năng của bộ chuyển đổi siêu dữ liệu 150. Bộ chuyển đổi siêu dữ liệu 150 nhận vị trí của đối tượng được biểu thị bởi vector P trong hệ tọa độ. Hơn nữa, vị trí tham chiếu mà siêu dữ liệu DirAC có liên quan được đưa ra bởi vector R trong cùng hệ tọa độ. Do đó, hướng đến của vector DoA kéo dài từ đầu vector R đến đầu vector B . Do đó, vector DoA thực tế thu được bằng cách trừ vector vị trí tham chiếu R khỏi vector vị trí đối tượng P .

Để có thông tin DoA được chuẩn hóa được chỉ định bởi vector DoA, chênh lệch vector được chia cho độ lớn hay chiều dài của vector DoA. Hơn nữa, và nếu điều này là cần thiết và có chủ đích, độ dài của vector DoA cũng có thể được đưa vào siêu dữ liệu được tạo bởi bộ chuyển đổi siêu dữ liệu 150 để ngoài ra, khoảng cách của đối tượng từ điểm tham chiếu cũng có trong siêu dữ liệu sao cho thao tác chọn lọc của đối tượng này cũng có thể được thực hiện dựa trên khoảng cách của đối tượng từ vị trí tham chiếu. Đặc biệt, khối hướng trích xuất 148 của Fig.1f cũng có thể hoạt động như được thảo luận liên quan đến Fig.3f, mặc dù các phương án khác để tính toán thông tin DoA và, một cách tùy chọn, thông tin khoảng cách cũng có thể được áp dụng. Hơn nữa, như đã thảo luận đối với Fig.3a, các khối 125 và 126 được minh họa trên Fig.1c hoặc Fig.1d có thể hoạt động theo cách tương tự như được thảo luận đối với Fig.3f.

Hơn nữa, thiết bị theo Fig.3a có thể được tạo cấu hình để nhận được nhiều mô tả đối tượng âm thanh, và bộ chuyển đổi siêu dữ liệu được tạo cấu hình để chuyển đổi trực tiếp từng mô tả siêu dữ liệu thành mô tả DirAC và sau đó, bộ chuyển đổi siêu dữ

liệu được tạo cấu hình để kết hợp các mô tả siêu dữ liệu DirAC riêng lẻ để thu được mô tả DirAC kết hợp như siêu dữ liệu DirAC được minh họa trên Fig.3a. Trong một phương án, việc kết hợp được thực hiện bằng cách tính hệ số trọng số 320 cho hướng đến thứ nhất sử dụng năng lượng thứ nhất và bằng cách tính hệ số trọng số 322 cho hướng đến thứ hai sử dụng năng lượng thứ hai, trong đó hướng đến được xử lý bởi các khối 320, 332 liên quan đến cùng một ngăn thời gian/tần số. Sau đó, trong khối 324, phép cộng có trọng số được thực hiện như được thảo luận liên quan đến mục 144 trên Fig.1d. Do đó, quy trình được minh họa trên Fig.3a thể hiện phương án của phương án thay thế thứ nhất theo Fig.1d..

Tuy nhiên, đối với phương án thay thế thứ hai, quy trình sẽ là tất cả độ khuếch tán được đặt thành không hoặc thành một giá trị nhỏ và, đối với ngăn thời gian/tần số, tất cả các giá trị hướng đến khác nhau được đưa ra cho ngăn thời gian/tần số này là được xem xét và giá trị hướng đến lớn nhất được chọn là giá trị hướng đến kết hợp cho ngăn thời gian/tần số này. Trong các phương án khác, người ta cũng có thể chọn giá trị thứ hai đến lớn nhất với điều kiện là thông tin năng lượng cho hai giá trị hướng đến này không quá khác nhau. Giá trị hướng đến được chọn có năng lượng là năng lượng lớn nhất trong số các năng lượng từ sự đóng góp khác nhau cho ngăn tần số thời gian này hoặc năng lượng cao thứ hai hoặc thứ ba.

Do đó, khía cạnh thứ ba như được mô tả liên quan đến các hình vẽ từ Fig.3a đến Fig.3f khác với khía cạnh thứ nhất ở chỗ khía cạnh thứ ba cũng hữu ích cho việc chuyển đổi một mô tả đối tượng duy nhất thành siêu dữ liệu DirAC. Ngoài ra, giao diện đầu vào 100 có thể nhận được một số mô tả đối tượng có cùng định dạng đối tượng/siêu dữ liệu. Do đó, bất kỳ bộ chuyển đổi định dạng nào như được thảo luận liên quan đến khía cạnh thứ nhất trên Fig.1a là không bắt buộc. Do đó, phương án theo Fig.3a có thể hữu ích trong bối cảnh nhận hai mô tả đối tượng khác nhau bằng cách sử dụng các tín hiệu dạng sóng đối tượng khác nhau và siêu dữ liệu đối tượng khác nhau làm mô tả cảnh thứ nhất và mô tả thứ hai làm đầu vào cho bộ kết hợp định dạng 140, và đầu ra của bộ chuyển đổi siêu dữ liệu 150, 125, 126 hoặc 148 có thể là sự biểu diễn DirAC với siêu dữ liệu DirAC và do đó, bộ phân tích DirAC 180 của Fig.1 cũng không bắt buộc. Tuy nhiên, các thành phần khác liên quan đến bộ tạo kênh vận chuyển 160 tương ứng với bộ trộn giảm 163 của Fig.3a có thể được sử dụng trong bối cảnh

của khía cạnh thứ ba cũng như bộ mã hóa kênh vận chuyển 170, bộ mã hóa siêu dữ liệu 190 và, trong bối cảnh này, giao diện đầu ra 300 của Fig.3a tương ứng với giao diện đầu ra 200 của Fig.1a. Do đó, tất cả các mô tả tương ứng được đưa ra liên quan đến khía cạnh thứ nhất cũng áp dụng cho khía cạnh thứ ba.

Fig.4a, Fig.4b minh họa khía cạnh thứ tư của sáng chế trong bối cảnh thiết bị thực hiện tổng hợp dữ liệu âm thanh. Đặc biệt, thiết bị có giao diện đầu vào 100 để nhận mô tả DirAC về cảnh âm thanh có siêu dữ liệu DirAC và ngoài ra để nhận tín hiệu đối tượng có siêu dữ liệu đối tượng. Bộ mã hóa cảnh âm thanh này được minh họa trên Fig.4b còn bao gồm bộ tạo siêu dữ liệu 400 để tạo mô tả siêu dữ liệu kết hợp bao gồm siêu dữ liệu DirAC trên một mặt và mặt khác là siêu dữ liệu đối tượng. Siêu dữ liệu DirAC bao gồm hướng đến cho các ô thời gian/tần số riêng lẻ và siêu dữ liệu đối tượng bao gồm hướng hoặc thêm vào đó là khoảng cách hoặc độ khuếch tán của đối tượng riêng lẻ.

Đặc biệt, giao diện đầu vào 100 được tạo cấu hình để nhận, ngoài ra, tín hiệu vận chuyển được liên kết với mô tả DirAC của cảnh âm thanh như được minh họa trên Fig.4b, và giao diện đầu vào còn được tạo cấu hình để nhận tín hiệu dạng sóng đối tượng được liên kết với tín hiệu đối tượng. Do đó, bộ mã hóa cảnh bao gồm thêm bộ mã hóa tín hiệu vận chuyển để mã hóa tín hiệu vận chuyển và tín hiệu dạng sóng đối tượng, và bộ mã hóa vận chuyển 170 có thể tương ứng với bộ mã hóa 170 theo Fig.1a.

Đặc biệt, bộ tạo siêu dữ liệu 140 tạo siêu dữ liệu kết hợp có thể được tạo cấu hình như đã thảo luận liên quan đến khía cạnh thứ nhất, khía cạnh thứ hai hoặc khía cạnh thứ ba. Và trong phương án được ưu tiên, bộ tạo siêu dữ liệu 400 được tạo cấu hình để tạo, cho siêu dữ liệu đối tượng, hướng băng rộng duy nhất mỗi lần, tức là cho một khung thời gian nhất định, và bộ tạo siêu dữ liệu được tạo cấu hình để làm mới hướng băng rộng duy nhất mỗi lần ít thường xuyên hơn siêu dữ liệu DirAC.

Quy trình được thảo luận liên quan đến Fig.4b cho phép có siêu dữ liệu kết hợp mà có siêu dữ liệu cho mô tả DirAC đầy đủ và ngoài ra, còn có siêu dữ liệu cho đối tượng âm thanh bổ sung, nhưng ở định dạng DirAC sao cho việc kết xuất DirAC rất hữu ích có thể được thực hiện, đồng thời, thực hiện việc lọc theo hướng có chọn lọc hoặc biến đổi như đã thảo luận về khía cạnh thứ hai.

Do đó, khía cạnh thứ tư của sáng chế và đặc biệt, bộ tạo siêu dữ liệu 400 biểu diễn bộ chuyển đổi định dạng cụ thể trong đó định dạng phổ biến là định dạng DirAC, và đầu vào là mô tả DirAC cho cảnh thứ nhất trong định dạng thứ nhất được thảo luận liên quan đến Fig.1a và cảnh thứ hai là tín hiệu đơn hoặc kết hợp chẳng hạn như tín hiệu đối tượng SAOC. Do đó, đầu ra của bộ chuyển đổi định dạng 120 biểu diễn đầu ra của bộ tạo siêu dữ liệu 400, nhưng ngược lại với sự kết hợp cụ thể thực tế của siêu dữ liệu bởi một trong hai phương án, ví dụ, như đã thảo luận đối với Fig.1d, siêu dữ liệu đối tượng có trong tín hiệu đầu ra, tức là "siêu dữ liệu kết hợp" tách biệt với siêu dữ liệu của mô tả DirAC để cho phép biến đổi có chọn lọc cho dữ liệu đối tượng.

Do đó, "hướng/khoảng cách/độ khuếch tán" được biểu thị ở mục 2 ở phía bên phải của Fig.4a tương ứng với đầu vào siêu dữ liệu đối tượng âm thanh bổ sung vào giao diện đầu vào 100 của Fig.2a, nhưng trong phương án của Fig.4a, chỉ cho một mô tả DirAC duy nhất. Do đó, theo một nghĩa nào đó, người ta có thể nói rằng Fig.2a biểu diễn việc triển khai phía bộ giải mã của bộ mã hóa được minh họa trên Fig.4a, Fig.4b với điều kiện là phía bộ giải mã của thiết bị theo Fig.2a chỉ nhận được một mô tả DirAC duy nhất và siêu dữ liệu đối tượng được tạo bởi bộ tạo siêu dữ liệu 400 trong cùng một dòng bit với "siêu dữ liệu đối tượng âm thanh bổ sung".

Do đó, phép biến đổi hoàn toàn khác nhau của dữ liệu đối tượng bổ sung có thể được thực hiện khi tín hiệu vận chuyển được mã hóa có sự biểu diễn riêng của tín hiệu dạng sóng đối tượng tách biệt với dòng vận chuyển DirAC. Và, tuy nhiên, bộ mã hóa vận chuyển 170 trộn giảm cả hai dữ liệu, tức là kênh vận chuyển cho mô tả DirAC và tín hiệu dạng sóng từ đối tượng, khi đó sự phân tách sẽ kém hoàn hảo hơn, nhưng bằng cách thêm thông tin năng lượng của đối tượng, thậm chí còn tách biệt với kênh trộn giảm kết hợp và phép biến đổi có chọn lọc của đối tượng liên quan đến mô tả DirAC.

Các hình vẽ từ Fig.5a đến Fig.5d thể hiện thêm khía cạnh thứ năm của sáng chế trong bối cảnh thiết bị thực hiện tổng hợp dữ liệu âm thanh. Để đạt được mục đích này, giao diện đầu vào 100 được cung cấp để nhận mô tả DirAC của một hoặc nhiều đối tượng âm thanh và/hoặc mô tả DirAC của tín hiệu đa kênh và/hoặc mô tả DirAC của tín hiệu Ambisonics bậc nhất và/hoặc tín hiệu Ambisonics bậc cao hơn, trong đó mô tả DirAC bao gồm thông tin vị trí của một hoặc nhiều đối tượng hoặc thông tin phụ

cho các tín hiệu Ambisonics bậc nhất hoặc các tín hiệu Ambisonics bậc cao hoặc thông tin vị trí cho tín hiệu đa kênh dưới dạng thông tin phụ hoặc từ giao diện người dùng.

Cụ thể, bộ thao tác 500 được tạo cấu hình để thao tác mô tả DirAC của một hoặc nhiều đối tượng âm thanh, mô tả DirAC của tín hiệu đa kênh, mô tả DirAC của các tín hiệu Ambisonics bậc nhất hoặc mô tả DirAC của các tín hiệu Ambisonics bậc cao để thu được mô tả DirAC được thao tác. Để tổng hợp mô tả DirAC được thao tác này, bộ tổng hợp DirAC 220, 240 được tạo cấu hình để tổng hợp mô tả DirAC được thao tác này để thu được dữ liệu âm thanh tổng hợp.

Trong phương án được ưu tiên, bộ tổng hợp DirAC 220, 240 bao gồm bộ kết xuất DirAC 222 như được minh họa trên Fig.5b và bộ chuyển đổi phổ-thời gian 240 được kết nối sau đó phát ra tín hiệu miền thời gian được thao tác. Đặc biệt, bộ thao tác 500 được tạo cấu hình để thực hiện thao tác gán trọng số phụ thuộc vào vị trí trước khi kết xuất DirAC.

Đặc biệt, khi bộ tổng hợp DirAC được tạo cấu hình để xuất ra nhiều đối tượng của các tín hiệu Ambisonics bậc nhất hoặc tín hiệu Ambisonics bậc cao hoặc tín hiệu đa kênh, bộ tổng hợp DirAC được tạo cấu hình để sử dụng bộ chuyển đổi phổ-thời gian riêng cho từng đối tượng hoặc từng thành phần của tín hiệu Ambisonics bậc nhất hoặc bậc cao hoặc cho từng kênh của tín hiệu đa kênh như được minh họa trên Fig.5d tại các khối 506, 508. Như được phác thảo trong khối 510, đầu ra của các chuyển đổi riêng biệt tương ứng được thêm vào với nhau miễn là tất cả các tín hiệu ở định dạng chung, tức là, ở định dạng tương thích.

Do đó, trong trường hợp giao diện đầu vào 100 của Fig.5a, nhận được nhiều hơn một, tức là hai hoặc ba sự biểu diễn, mỗi sự biểu diễn có thể được thao tác riêng như được minh họa trong khối 502 trong miền tham số như đã thảo luận về Fig.2b hoặc Fig.2c, và sau đó, phép tổng hợp có thể được thực hiện như được nêu trong khối 504 cho mỗi mô tả được thao tác, và phép tổng hợp sau đó có thể được thêm vào miền thời gian như đã thảo luận về khối 510 trên Fig.5d. Ngoài ra, kết quả của các quy trình tổng hợp DirAC riêng lẻ trong miền phổ có thể đã được thêm vào trong miền phổ và sau đó phép chuyển đổi miền thời gian đơn lẻ cũng có thể được sử dụng. Đặc biệt, bộ

thao tác 500 có thể được triển khai như bộ thao tác được thảo luận liên quan đến Fig.2d hoặc được thảo luận liên quan đến bất kỳ khía cạnh nào khác trước đó.

Do đó, khía cạnh thứ năm của sáng chế đề xuất tính năng quan trọng đối với thực tế, khi các mô tả DirAC riêng lẻ của các tín hiệu âm thanh rất khác nhau được đưa vào, và khi thao tác nhất định của các mô tả riêng lẻ được thực hiện như đã thảo luận về khối 500 của Fig.5a, trong đó đầu vào của bộ thao tác 500 có thể là mô tả DirAC ở bất kỳ định dạng nào, chỉ bao gồm một định dạng duy nhất, trong khi khía cạnh thứ hai tập trung vào việc tiếp nhận ít nhất hai mô tả DirAC khác nhau hoặc ví dụ về khía cạnh thứ tư có liên quan đến việc tiếp nhận mô tả DirAC một mặt và mặt khác mô tả tín hiệu đối tượng.

Sau đây tham chiếu đến Fig.6. Fig.6 minh họa một triển khai khác để thực hiện phép tổng hợp khác với tổng hợp DirAC. Ví dụ, khi bộ phân tích trường âm thanh tạo ra, đối với mỗi tín hiệu nguồn, tín hiệu đơn riêng biệt S và hướng đến ban đầu và khi, tùy thuộc vào thông tin dịch, hướng đến mới được tính, thì bộ tạo tín hiệu Ambisonics 430 theo Fig.6 ví dụ sẽ được sử dụng để tạo mô tả trường âm thanh cho tín hiệu nguồn 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 một góc phương vị ϕ . Sau đó, quy trình được thực hiện bởi bộ tính trường âm thanh 420 theo Fig.6 sẽ được tạo ra, ví dụ, sự biểu diễn trường âm thanh Ambisonics bậc nhất cho mỗi nguồn âm thanh với hướng đến mới và sau đó, phép biến đổi thêm cho mỗi âm thanh nguồn có thể được thực hiện bằng cách sử dụng hệ số tỷ lệ tùy thuộc vào khoảng cách của trường âm thanh đến vị trí tham chiếu mới và sau đó, tất cả các trường âm thanh từ các nguồn riêng lẻ có thể chồng lên nhau để cuối cùng thu được trường âm thanh được biến đổi, một lần nữa, trong ví dụ sự biểu diễn Ambisonics liên quan đến một vị trí tham chiếu mới nhất định.

Khi giải thích rằng mỗi ngăn thời gian/tần số được xử lý bởi bộ phân tích DirAC 422 biểu diễn nguồn âm thanh (giới hạn băng thông) nhất định, thì bộ tạo tín hiệu Ambisonics 430 có thể được sử dụng, thay vì bộ tổng hợp DirAC 425, để tạo, cho mỗi ngăn thời gian/tần số, sự biểu diễn Ambisonics đầy đủ sử dụng tín hiệu trộn giảm hoặc tín hiệu áp suất hoặc thành phần đa hướng cho ngăn thời gian/tần số này như là "tín hiệu đơn S" theo Fig.6. Sau đó, phép chuyển đổi tần số-thời gian riêng trong bộ

chuyển đổi tần số-thời gian 426 cho từng thành phần W, X, Y, Z sau đó sẽ dẫn đến mô tả trường âm thanh khác với mô tả được minh họa trên Fig.6.

Sau đây sẽ giải thích thêm về phép phân tích DirAC và phép tổng hợp DirAC được đưa ra như đã biết trong kỹ thuật. Fig.7a minh họa bộ phân tích DirAC như đã bộc lộ ban đầu, ví dụ trong tài liệu tham khảo “Directional Audio Coding” từ IWPASH năm 2009. Bộ phân tích DirAC bao gồm giàn lọc băng 1310, bộ phân tích năng lượng 1320, bộ phân tích cường độ 1330, khối trung bình theo thời gian 1340 và bộ tính độ khuếch tán 1350 và bộ tính định hướng 1360. Trong DirAC, cả phép phân tích và phép tổng hợp đều được thực hiện trong miền tần số. Có một số phương pháp để phân chia âm thanh thành các dải 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 (short time Fourier transform - STFT), và giàn lọc gương cầu phương (Quadrature mirror filter bank - QMF). Ngoài ra, hoàn toàn có thể tự do thiết kế giàn lọc với các bộ lọc tùy ý được tối ưu hóa cho bất kỳ mục đích cụ thể nào. Mục tiêu của phép phân tích theo hướng là ước tính ở mỗi dải 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, phép phân tích năng lượng của trường âm thanh được thấy là phù hợp được minh họa trên Fig.7a. Phép phân tích năng lượng có thể được thực hiện, khi tín hiệu áp suất và các tín hiệu vận tốc ở một, hai hoặc ba chiều được thu từ một vị trí. Trong các tín hiệu định dạng B bậc nhất, tín hiệu đa hướng được gọi là tín hiệu W, được thu nhỏ bởi căn bậc hai của hai. Áp suất âm thanh có thể được ước tính là $S = \sqrt{2} * W$, được thể hiện trong miền STFT.

Các kênh X-, Y- và Z có mô hình theo hướng của lưỡng cực có hướng dọc theo trục Đề-các, tạo với nhau thành vector $U = [X, Y, Z]$. Vector ước tính vector vận tốc trường âm thanh, và cũng được biểu thị 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 bằng cách định vị ngẫu nhiên các micrô theo hướng, hoặc với tập hợp có khoảng cách gần nhau của các micrô đa hướng. Trong một số ứng dụng, các tín hiệu micrô có thể được tạo thành trong miền tính toán, tức là được mô phỏng. Hướng của âm thanh được xác định là hướng ngược lại của vector cường độ I. Hướng được biểu thị là các giá trị góc phương vị và góc nâng 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 vector cường độ và năng lượng. Kết quả của phương trình này là số có giá trị thực giữa không và một, đặc trưng nếu năng lượng âm thanh đến từ một hướng duy nhất (độ khuếch tán là không), hoặc từ tất cả các hướng (độ khuếch tán là một). Quy trình này phù hợp trong trường hợp có sẵn thông tin vận tốc 3D đầy đủ hoặc có kích thước nhỏ hơn.

Fig.7b minh họa phép tổng hợp DirAC, một lần nữa có giàn lọc băng 1370, khối micrô ảo 1400, khối tổng hợp có hướng/khuếch tán 1450, và thiết lập loa phóng thanh nhất định hoặc thiết lập loa phóng thanh dự định ảo 1460. Ngoài ra, bộ biến đổi khuếch tán-khuếch đại 1380, khối băng khuếch đại quét biên độ dựa trên vector (vector based amplitude panning - VBAP) 1390, khối bù micrô 1420, khối trung bình khuếch đại loa phóng thanh 1430 và bộ phân phối 1440 cho các kênh khác được sử dụng. Trong quá trình tổng hợp DirAC với loa phóng thanh này, phiên bản tổng hợp DirAC chất lượng cao được thể hiện trên Fig.7b nhận tất cả tín hiệu định dạng B, trong đó tín hiệu micrô ảo được tính cho từng hướng loa phóng thanh của thiết lập loa phóng thanh 1460. Kiểu hướng được sử dụng thường là lưỡng cực. Các tín hiệu micrô ảo sau đó được biến đổi theo kiểu phi tuyến, tùy thuộc vào siêu dữ liệu. Phiên bản tốc độ bit thấp của DirAC không được thể hiện trên Fig.7b, tuy nhiên, trong tình huống này, chỉ có một kênh âm thanh được truyền đi như được minh họa trên Fig.6. Sự khác biệt trong quá trình xử lý là tất cả các tín hiệu micrô ảo sẽ được thay thế bằng một kênh âm thanh duy nhất nhận được. Các tín hiệu micro ảo được chia thành hai luồng: luồng khuếch tán và luồng không khuếch tán, được xử lý riêng.

Âm thanh không khuếch tán được tái tạo dưới dạng các nguồn điểm bằng cách sử dụng quét biên độ dựa trên vector (VBAP). Trong phép quét, tín hiệu âm thanh đơn âm được áp dụng cho tập hợp con 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 bằng cách sử dụng thông tin của thiết lập loa phóng thanh và hướng quét được chỉ định. Trong phiên bản tốc độ bit thấp, tín hiệu đầu vào chỉ được quét đơn giản tới các hướng được bao hàm 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 lậ phi tuyến tính nào.

Trong nhiều trường hợp, siêu dữ liệu định hướng có thể thay đổi đột ngột theo thời gian. Để tránh các thành phần lạ, các hệ số khuếch đại của loa phóng thanh được tính toán bằng VBAP được làm mịn bằng cách lấy tích phân theo thời gian với các hằng số thời gian phụ thuộc 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, những thay đổi theo hướng không được coi là chậm hơn so với không tính trung bình trong hầu hết các trường hợp. Mục đích của việc tổng hợp âm thanh khuếch tán là tạo ra nhận thức về âm thanh bao quanh người nghe. Trong phiên bản tốc độ bit thấp, luồng khuếch tán được tái tạo bằng cách giải tương quan tín hiệu đầu vào và tái tạo nó từ mọi loa phóng thanh. Trong phiên bản chất lượng cao, các tín hiệu micro ảo của luồng khuếch tán đã không nhất quán ở một mức độ nào đó, và chúng chỉ cần được giải tương quan một cách nhẹ nhàng. Cách tiếp cận này cung cấp chất lượng không gian tốt hơn cho âm vang vòm và âm thanh xung quanh so với phiên bản tốc độ bit thấp. Để tổng hợp DirAC với tai nghe, DirAC được tạo ra với một lượng loa phóng thanh ảo nhất định xung quanh người nghe cho luồng không khuếch tán và một số lượng loa phóng thanh nhất định cho luồng khuếch tán. Các loa phóng thanh ảo được triển khai dưới dạng tích chập các tín hiệu đầu vào với các hàm truyền liên quan đến đầu (head-related transfer function - HRTF) đo được.

Sau đây, mỗi quan hệ chung khác nữa liên quan đến các khía cạnh khác nhau và đặc biệt, liên quan đến việc triển khai thêm khía cạnh thứ nhất như được thảo luận liên quan đến Fig.1a được đưa ra. Nói chung, sáng chế đề cập đến sự kết hợp của các cảnh khác nhau ở các định dạng khác nhau bằng cách sử dụng định dạng chung, định dạng chung ví dụ có thể là miền định dạng B, miền áp suất/vận tốc hoặc miền siêu dữ liệu như được thảo luận, ví dụ, trong các mục 120, 140 của Fig.1a.

Khi sự kết hợp không được thực hiện trực tiếp theo định dạng DirAC chung, thì phép phân tích DirAC 802 được thực hiện theo một phương án trước khi truyền trong bộ mã hóa như đã thảo luận trước đó đối với mục 180 của Fig.1a.

Sau đó, sau khi phân tích DirAC, kết quả được mã hóa như đã thảo luận trước đó đối với bộ mã hóa 170 và bộ mã hóa siêu dữ liệu 190 và kết quả được mã hóa được truyền thông qua tín hiệu đầu ra được mã hóa được tạo bởi giao diện đầu ra 200. Tuy

nhiên, theo phương án khác, kết quả có thể được hiển thị trực tiếp bằng thiết bị theo Fig.1a khi đầu ra của khối 160 của Fig.1a và đầu ra của khối 180 của Fig.1a được chuyển tiếp đến bộ kết xuất DirAC. Do đó, thiết bị theo Fig.1a sẽ không phải là thiết bị mã hóa cụ thể mà sẽ là bộ phân tích và bộ kết xuất tương ứng.

Phương án khác được minh họa ở nhánh bên phải của Fig.8, trong đó việc truyền từ bộ mã hóa đến bộ giải mã được thực hiện và, như được minh họa trong khối 804, phép phân tích DirAC và phép tổng hợp DirAC được thực hiện sau khi truyền, tức là ở phía bộ giải mã. Quy trình này sẽ là trường hợp khi phương án theo Fig.1a được sử dụng, tức là tín hiệu đầu ra được mã hóa là tín hiệu định dạng B không có siêu dữ liệu không gian. Sau khối 808, kết quả có thể được kết xuất để phát lại hoặc, thay vào đó, kết quả thậm chí có thể được mã hóa và truyền lại. Do đó, rõ ràng là các quy trình theo sáng chế như được xác định và mô tả liên quan đến các khía cạnh khác nhau rất linh hoạt và có thể được điều chỉnh rất tốt cho các trường hợp sử dụng cụ thể.

Khía cạnh thứ nhất của sáng chế: Mã hóa/kết xuất âm thanh không gian dựa trên DirAC phổ dụng

Bộ mã hóa âm thanh không gian dựa trên Dirac có thể mã hóa các tín hiệu đa kênh, các định dạng Ambisonics và các đối tượng âm thanh một cách riêng biệt hoặc đồng thời.

Lợi ích và lợi thế so với kỹ thuật trước đó

- Sơ đồ mã hóa âm thanh không gian dựa trên DirAC phổ dụng cho hầu hết các định dạng đầu vào âm thanh nhập vai phù hợp
- Việc kết xuất âm thanh phổ dụng của các định dạng đầu vào khác nhau trên các định dạng đầu ra khác nhau

Khía cạnh thứ hai của sáng chế: Kết hợp hai hoặc nhiều mô tả DirAC trên bộ giải mã

Khía cạnh thứ hai của sáng chế liên quan đến sự kết hợp và kết xuất hai hoặc nhiều hơn hai mô tả DirAC trong miền phổ.

Lợi ích và lợi thế so với kỹ thuật trước đó

- Kết hợp luồng DirAC hiệu quả và chính xác

- Cho phép sử dụng DirAC biểu diễn phổ dụng cho bất kỳ cảnh nào và kết hợp hiệu quả các luồng khác nhau trong miền tham số hoặc miền phổ
- Việc thao tác cảnh hiệu quả và trực quan cho từng cảnh DirAC riêng lẻ hoặc cảnh kết hợp trong miền phổ và sau đó chuyển đổi thành miền thời gian của cảnh kết hợp được thao tác.

Khía cạnh thứ ba của sáng chế: Chuyển đổi các đối tượng âm thanh thành miền DirAC

Khía cạnh thứ ba của sáng chế liên quan đến việc chuyển đổi siêu dữ liệu đối tượng và các tín hiệu dạng sóng đối tượng tùy chọn trực tiếp vào miền DirAC và trong phương án kết hợp một số đối tượng thành sự biểu diễn đối tượng.

Lợi ích và lợi thế so với kỹ thuật trước đó

- Ước tính siêu dữ liệu DirAC hiệu quả và chính xác bằng bộ chuyển mã siêu dữ liệu đơn giản của siêu dữ liệu đối tượng âm thanh
- Cho phép DirAC mã hóa các cảnh âm thanh phức tạp liên quan đến một hoặc nhiều đối tượng âm thanh
- Phương pháp hiệu quả để mã hóa các đối tượng âm thanh thông qua DirAC trong một sự biểu diễn theo tham số duy nhất của cảnh âm thanh hoàn chỉnh.

Khía cạnh thứ tư của sáng chế: Kết hợp siêu dữ liệu đối tượng và siêu dữ liệu DirAC thông thường

Khía cạnh thứ ba của sáng chế đề cập đến việc sửa đổi siêu dữ liệu DirAC với các hướng và, một cách ối ưu, khoảng cách hoặc độ khuếch tán của các đối tượng riêng lẻ tạo thành cảnh âm thanh kết hợp được biểu thị bằng các tham số DirAC. Thông tin bổ sung này dễ dàng được mã hóa, vì nó chủ yếu bao gồm một hướng bằng rộng duy nhất cho mỗi đơn vị thời gian và có thể được làm mới ít thường xuyên hơn các tham số DirAC khác vì các đối tượng có thể được coi là tĩnh hoặc di chuyển với tốc độ chậm.

Lợi ích và lợi thế so với kỹ thuật trước đó

- Cho phép DirAC mã hóa cảnh âm thanh phức tạp liên quan đến một hoặc nhiều đối tượng âm thanh

- Ước tính siêu dữ liệu DirAC hiệu quả và chính xác bằng bộ chuyển mã siêu dữ liệu đơn giản của siêu dữ liệu đối tượng âm thanh.
- Phương pháp hiệu quả hơn để mã hóa các đối tượng âm thanh thông qua DirAC bằng cách kết hợp hiệu quả siêu dữ liệu của chúng trong miền DirAC
- Phương pháp hiệu quả để mã hóa các đối tượng âm thanh và thông qua DirAC bằng cách kết hợp hiệu quả các sự biểu diễn âm thanh của chúng trong một sự biểu diễn theo tham số duy nhất của cảnh âm thanh.

Khía cạnh thứ năm của sáng chế: Thao tác các cảnh MC đối tượng và FOA/HOA C trong tổng hợp DirAC

Khía cạnh thứ tư liên quan đến phía bộ giải mã và khai thác các vị trí đã biết của các đối tượng âm thanh. Các vị trí có thể được cung cấp bởi người dùng thông qua giao diện tương tác và cũng có thể được bao gồm dưới dạng thông tin phụ bổ sung bên trong dòng bit.

Mục đích là để có thể thao tác cảnh âm thanh đầu ra bao gồm một số đối tượng bằng cách thay đổi riêng lẻ các thuộc tính của các đối tượng, chẳng hạn như mức độ, cân bằng và/hoặc vị trí không gian. Nó cũng có thể được hình dung để lọc hoàn toàn đối tượng hoặc khôi phục các đối tượng riêng lẻ từ luồng kết hợp.

Việc thao tác của cảnh âm thanh đầu ra có thể đạt được bằng cách xử lý chung các tham số không gian của siêu dữ liệu DirAC, siêu dữ liệu đối tượng, đầu vào người dùng tương tác nếu có và các tín hiệu âm thanh được mang trong các kênh vận chuyển.

Lợi ích và lợi thế so với kỹ thuật trước đó

- Cho phép DirAC xuất ra các đối tượng âm thanh ở phía bộ giải mã như được trình bày ở đầu vào của bộ mã hóa.
- Cho phép tái tạo DirAC để thao tác đối tượng âm thanh riêng lẻ bằng cách áp dụng khuếch đại, xoay, ...
- Khả năng yêu cầu nỗ lực tính toán bổ sung tối thiểu vì nó chỉ yêu cầu thao tác gán trọng số phụ thuộc vào vị trí trước giàn lọc kết xuất và tổng hợp ở cuối quá trình tổng hợp DirAC (các đầu ra đối tượng bổ sung sẽ chỉ yêu cầu một giàn lọc tổng hợp bổ sung cho mỗi đầu ra đối tượng).

Sáng chế đề xuất các phương án thay thế khác trong các phương án tiếp theo, và đặc biệt liên quan đến khía cạnh thứ nhất và cũng liên quan đến các khía cạnh khác. Các phương án thay thế này như sau:

Thứ nhất, kết hợp các định dạng khác nhau trong miền định dạng B và thực hiện phân tích DirAC trong bộ mã hóa hoặc truyền các kênh kết hợp đến bộ giải mã và thực hiện phân tích và tổng hợp DirAC ở đó.

Thứ hai, kết hợp các định dạng khác nhau trong miền áp suất/vận tốc và thực hiện phân tích DirAC trong bộ mã hóa. Ngoài ra, dữ liệu áp suất/vận tốc được truyền đến bộ giải mã và phép phân tích DirAC được thực hiện trong bộ giải mã và phép tổng hợp cũng được thực hiện trong bộ giải mã.

Thứ ba, kết hợp các định dạng khác nhau trong miền siêu dữ liệu và truyền một luồng DirAC duy nhất hoặc truyền một số luồng DirAC đến bộ giải mã trước khi kết hợp chúng và thực hiện kết hợp trong bộ giải mã.

Hơn nữa, các phương án hoặc các khía cạnh của sáng chế có liên quan đến các khía cạnh sau:

Thứ nhất, kết hợp các định dạng âm thanh khác nhau theo ba phương án thay thế trên.

Thứ hai, việc tiếp nhận, kết hợp và kết xuất hai mô tả DirAC đã có cùng định dạng được thực hiện.

Thứ ba, đối tượng cụ thể đối với bộ chuyển đổi DirAC với "phép chuyển đổi trực tiếp" dữ liệu đối tượng thành dữ liệu DirAC được thực hiện.

Thứ tư, siêu dữ liệu đối tượng ngoài siêu dữ liệu DirAC thông thường và kết hợp cả hai siêu dữ liệu; cả hai dữ liệu đều tồn tại song song với nhau trong dòng bit, nhưng các đối tượng âm thanh cũng được mô tả theo kiểu siêu dữ liệu DirAC.

Thứ năm, các đối tượng và luồng DirAC được truyền riêng tới bộ giải mã và các đối tượng được thao tác có chọn lọc trong bộ giải mã trước khi chuyển đổi các tín hiệu âm thanh đầu ra (loa phóng thanh) sang miền thời gian.

Ở đây cần đề cập rằng tất cả các phương án thay thế hoặc các khía cạnh như đã thảo luận trước đây và tất cả các khía cạnh như được xác định bởi các yêu cầu bảo hộ trong phần yêu cầu bảo hộ sau đây đều có thể được sử dụng riêng lẻ, tức là không có bất kỳ phương án thay thế hoặc đối tượng nào khác ngoài phương án thay thế, đối tượng hoặc yêu cầu bảo hộ độc lập sắp đưa ra dưới đây. Tuy nhiên, trong các phương án khác, hai hoặc nhiều hơn hai phương án thay thế hoặc các khía cạnh hoặc các yêu cầu bảo hộ độc lập có thể được kết hợp với nhau và, trong các phương án khác, tất cả các khía cạnh hoặc phương án thay thế và tất cả các yêu cầu bảo hộ độc lập có thể được kết hợp với nhau.

Tín hiệu âm thanh được mã hóa theo sáng chế có thể được lưu trữ trên phương tiện lưu trữ kỹ thuật số hoặc phương tiện lưu trữ không tạm thời hoặc có thể được truyền trên phương tiện truyền dẫn chẳng hạn như phương tiện truyền dẫn không dây hoặc phương tiện truyền dẫn có dây như Internet.

Mặc dù một vài khía cạnh đã được mô tả trong ngữ cảnh của thiết bị, rõ ràng rằng các khía cạnh này cũng thể hiện mô tả của phương pháp tương ứng, trong đó khối hoặc thiết bị tương ứng với bước xử lý của phương pháp hoặc đặc điểm của bước xử lý của phương pháp. Tương tự, các khía cạnh được mô tả trong ngữ cảnh của bước xử lý của phương pháp cũng biểu diễn sự mô tả của khối hoặc mục hoặc dấu hiệu tương ứng của thiết bị tương ứng.

Phụ thuộc vào các yêu cầu thực thi nhất định, các phương án của sáng chế có thể được thực hiện trong phần cứng hoặc trong phần mềm. Phương án có thể được thực hiện sử dụng vật lưu trữ số, ví dụ, đĩa mềm, DVD, CD, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ FLASH, có các tín hiệu điều khiển có thể đọc được bằng điện được lưu trữ trên đó, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính khả trình sao cho phương pháp tương ứng được thực hiện.

Một số phương án theo sáng chế bao gồm vật mang dữ liệu có các tín hiệu điều khiển có thể đọc được bằng điện tử, mà có khả năng kết hợp với hệ thống máy tính có thể lập trình, sao cho một trong số các phương pháp được mô tả ở đây được thực hiện.

Nói chung, các phương án của sáng chế có thể được thực hiện như sản phẩm chương trình máy tính với mã chương trình, mã chương trình có tác dụng để thực hiện

một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Mã chương trình có thể, ví dụ, 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 hoặc vật ghi lưu trữ không tạm thời.

Do đó, nói cách khác, phương án của phương pháp theo sáng chế 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 khác của các phương pháp theo sáng chế là vật mang dữ liệu (hoặc vật ghi lưu trữ số, hoặc vật ghi có thể đọc được bằng máy tính) gồm chương trình máy tính đã được ghi lại trên đó để thực hiện một trong số các phương pháp được mô tả ở đây.

Do đó, phương án nữa của các phương pháp theo 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 đã được mô tả ở đây. Ví dụ, có thể tạo cấu hình dòng dữ liệu hoặc chuỗi tín hiệu để được truyền thông qua sự kết nối truyền thông dữ liệu, ví dụ thông qua Internet.

Phương án khác gồm có phương tiện xử lý, ví dụ máy tính, hoặc thiết bị logic lập trình được, được tạo cấu hình hoặc được làm thích ứng để thực hiện một trong số các phương pháp được mô tả ở đây.

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

Theo một số phương án, thiết bị logic lập trình được (ví dụ, mảng công lập trình được dạng trường) có thể được sử dụng để thực hiện một số hoặc tất cả các chức năng của các phương pháp được mô tả ở đây. Trong một số phương án, mảng công lập trình được dạng trường có thể kết hợp với bộ vi xử lý để thực hiện một trong số các phương pháp được mô tả ở đây. Thông thường, các phương pháp 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 lý của sáng chế. Cần hiểu rằng các biến thể và biến đổi của các phương án và các chi tiết

được mô tả ở đây sẽ là rõ ràng đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật tương ứng. Do đó, mục đích của sáng chế chỉ bị giới hạn bởi phạm vi của các điểm yêu cầu bảo hộ sắp đưa ra dưới đây và không bị giới hạn bởi các chi tiết cụ thể được biểu diễn bằng cách mô tả và giải thích của các phương án ở đây.

YÊU CẦU BẢO HỘ

1. Thiết bị tạo mô tả cảnh âm thanh kết hợp, bao gồm:

giao diện đầu vào (100) để nhận mô tả thứ nhất về cảnh thứ nhất ở định dạng thứ nhất và mô tả thứ hai về cảnh thứ hai ở định dạng thứ hai, trong đó định dạng thứ hai khác với định dạng thứ nhất;

bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi mô tả thứ nhất thành định dạng chung và để chuyển đổi mô tả thứ hai thành định dạng chung, khi định dạng thứ hai khác với định dạng chung; và

bộ kết hợp định dạng (140) để kết hợp mô tả thứ nhất ở định dạng chung và mô tả thứ hai ở định dạng chung để thu được mô tả của cảnh âm thanh kết hợp.

2. Thiết bị theo điểm 1,

trong đó định dạng thứ nhất được chọn từ nhóm các định dạng bao gồm định dạng Ambisonics bậc nhất, định dạng Ambisonics bậc cao, định dạng DirAC, định dạng đối tượng âm thanh, và định dạng đa kênh, và

trong đó định dạng thứ hai được chọn từ nhóm các định dạng bao gồm định dạng Ambisonics bậc nhất, định dạng Ambisonics bậc cao, định dạng chung, định dạng DirAC, định dạng đối tượng âm thanh và định dạng đa kênh.

3. Thiết bị theo điểm 1 hoặc điểm 2,

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi mô tả thứ nhất thành sự biểu diễn tín hiệu định dạng B thứ nhất và để chuyển đổi mô tả thứ hai thành sự biểu diễn tín hiệu định dạng B thứ hai, và

trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp sự biểu diễn tín hiệu định dạng B thứ nhất và sự biểu diễn tín hiệu định dạng B thứ hai bằng cách kết hợp riêng lẻ các thành phần riêng lẻ của sự biểu diễn tín hiệu định dạng B thứ nhất và sự biểu diễn tín hiệu định dạng B thứ hai.

4. Thiết bị theo một điểm bất kỳ trong số các điểm trên,

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi mô tả thứ nhất thành sự biểu diễn tín hiệu áp suất/vận tốc thứ nhất và để chuyển đổi mô tả thứ hai thành sự biểu diễn tín hiệu áp suất/vận tốc thứ hai, và

trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp sự biểu diễn tín hiệu áp suất/vận tốc thứ nhất và sự biểu diễn tín hiệu áp suất/vận tốc thứ hai bằng cách kết hợp riêng lẻ các thành phần riêng lẻ của các sự biểu diễn tín hiệu áp suất/vận tốc để thu được sự biểu diễn tín hiệu áp suất/vận tốc kết hợp.

5. Thiết bị theo một điểm bất kỳ trong số các điểm trên,

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi mô tả thứ nhất thành sự biểu diễn tham số DirAC thứ nhất và để chuyển đổi mô tả thứ hai thành sự biểu diễn tham số DirAC thứ hai, khi mô tả thứ hai khác với sự biểu diễn tham số DirAC, và

trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp sự biểu diễn tham số DirAC thứ nhất và sự biểu diễn tham số DirAC thứ hai bằng cách kết hợp riêng lẻ các thành phần riêng lẻ của các biểu diễn tham số DirAC thứ nhất và sự biểu diễn tham số DirAC thứ hai để thu được sự biểu diễn tham số DirAC kết hợp cho cảnh âm thanh kết hợp.

6. Thiết bị theo điểm 5,

trong đó bộ kết hợp định dạng (140) được tạo cấu hình để tạo các giá trị hướng đến cho các ô thời gian-tần số hoặc các giá trị hướng đến và các giá trị độ khuếch tán cho các ô thời gian-tần số biểu thị cảnh âm thanh kết hợp.

7. Thiết bị theo một điểm bất kỳ trong số các điểm trên,

còn bao gồm bộ phân tích DirAC (180) để phân tích cảnh âm thanh kết hợp để suy ra các tham số DirAC cho cảnh âm thanh kết hợp,

trong đó các tham số DirAC bao gồm các giá trị hướng đến cho các ô thời gian-tần số hoặc các giá trị hướng đến và các giá trị độ khuếch tán cho các ô thời gian-tần số biểu thị cảnh âm thanh kết hợp.

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

còn bao gồm bộ tạo kênh vận chuyển (160) để tạo tín hiệu kênh vận chuyển từ cảnh âm thanh kết hợp hoặc từ cảnh thứ nhất và cảnh thứ hai, và

bộ mã hóa kênh vận chuyển (170) để mã hóa lỗi tín hiệu kênh vận chuyển, hoặc trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để tạo tín hiệu âm thanh nổi từ cảnh thứ nhất hoặc cảnh thứ hai ở định dạng Ambisonics bậc nhất hoặc định dạng Ambisonics bậc cao hơn bằng cách sử dụng bộ tạo chùm tương ứng hướng đến vị trí bên trái hoặc vị trí bên phải, hoặc

trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để tạo tín hiệu âm thanh nổi từ cảnh thứ nhất hoặc cảnh thứ hai ở dạng sự biểu diễn đa kênh bằng cách trộn giảm ba hoặc nhiều hơn ba kênh của sự biểu diễn đa kênh, hoặc

trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để tạo tín hiệu âm thanh nổi từ cảnh thứ nhất hoặc cảnh thứ hai ở dạng sự biểu diễn đối tượng âm thanh bằng cách quét từng đối tượng sử dụng vị trí của đối tượng hoặc bằng cách trộn giảm các đối tượng vào trộn giảm âm thanh nổi sử dụng thông tin biểu thị đối tượng nào nằm trong kênh âm thanh nổi nào, hoặc

trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để chỉ thêm kênh bên trái của tín hiệu âm thanh nổi vào kênh vận chuyển trộn giảm bên trái và chỉ thêm kênh bên phải của tín hiệu âm thanh để thu được kênh vận chuyển bên phải, hoặc

trong đó định dạng chung là định dạng B, và trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để xử lý sự biểu diễn định dạng B kết hợp để suy ra tín hiệu kênh vận chuyển, trong đó quá trình xử lý bao gồm việc thực hiện thao tác tạo chùm hoặc trích xuất tập hợp con các thành phần của tín hiệu định dạng B chẳng hạn như thành phần đa hướng như kênh vận chuyển đơn, hoặc

trong đó quá trình xử lý bao gồm việc tạo chùm bằng cách sử dụng tín hiệu đa hướng và thành phần Y với các dấu hiệu ngược lại của định dạng B để tính toán các kênh trái và phải, hoặc

trong đó quá trình xử lý bao gồm thao tác tạo chùm bằng cách sử dụng các thành phần của định dạng B và góc phương vị đã cho và góc nâng đã cho, hoặc

trong đó bộ tạo kênh vận chuyển (160) được tạo cấu hình để cung cấp các tín hiệu định dạng B của cảnh âm thanh kết hợp tới bộ mã hóa kênh vận chuyển, trong đó bất kỳ siêu dữ liệu không gian nào không có trong đầu ra cảnh âm thanh kết hợp bởi bộ kết hợp định dạng (140).

9. Thiết bị theo một điểm bất kỳ trong số các điểm trên, thiết bị còn bao gồm:

bộ mã hóa siêu dữ liệu (190)

để mã hóa siêu dữ liệu DirAC được mô tả trong cảnh âm thanh kết hợp để thu được siêu dữ liệu DirAC được mã hóa, hoặc

để mã hóa siêu dữ liệu DirAC được suy ra từ cảnh thứ nhất để thu được siêu dữ liệu DirAC được mã hóa thứ nhất và để mã hóa siêu dữ liệu DirAC được suy ra từ cảnh thứ hai để thu được siêu dữ liệu DirAC được mã hóa thứ hai.

10. Thiết bị theo một điểm bất kỳ trong số các điểm trên, thiết bị còn bao gồm:

giao diện đầu ra (200) để tạo tín hiệu đầu ra được mã hóa biểu diễn cảnh âm thanh kết hợp, tín hiệu đầu ra bao gồm siêu dữ liệu DirAC được mã hóa và một hoặc nhiều kênh vận chuyển được mã hóa.

11. Thiết bị theo một điểm bất kỳ trong số các điểm trên,

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi định dạng Ambisonics bậc cao hoặc định dạng Ambisonics bậc nhất thành định dạng B, trong đó định dạng Ambisonics bậc cao được cắt bớt trước khi chuyển đổi sang định dạng B, hoặc

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chiếu đổi tượng hoặc kênh trên sóng hài hình cầu ở vị trí tham chiếu để thu được các tín hiệu được chiếu, và trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp các tín hiệu được chiếu để thu được các hệ số định dạng B, trong đó đổi tượng hoặc kênh được đặt trong không gian tại vị trí xác định và có khoảng cách riêng tùy chọn từ vị trí tham chiếu, hoặc

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để thực hiện phân tích DirAC bao gồm phân tích thời gian-tần số của các thành phần định dạng B và xác định các vector áp suất và vận tốc, và trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp các vector áp suất/vận tốc khác nhau và trong đó bộ kết hợp định dạng (140) còn bao gồm bộ phân tích DirAC để lấy siêu dữ liệu DirAC từ dữ liệu áp suất/vận tốc kết hợp, hoặc

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để trích xuất các tham số DirAC từ siêu dữ liệu đối tượng của định dạng đối tượng âm thanh thành định dạng thứ nhất hoặc thứ hai, trong đó vector áp suất là tín hiệu dạng sóng đối tượng và hướng được suy ra từ vị trí đối tượng trong không gian hoặc độ khuếch tán được đưa ra trực tiếp trong siêu dữ liệu đối tượng hoặc được đặt thành giá trị mặc định chẳng hạn như giá trị 0, hoặc

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để chuyển đổi các tham số DirAC được suy ra từ định dạng dữ liệu đối tượng thành dữ liệu áp suất/vận tốc và bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp dữ liệu áp suất/vận tốc với dữ liệu áp suất/vận tốc lấy từ mô tả khác nhau của một hoặc nhiều đối tượng âm thanh khác nhau, hoặc

trong đó bộ chuyển đổi định dạng (120) được tạo cấu hình để suy ra trực tiếp các tham số DirAC, và trong đó bộ kết hợp định dạng (140) được tạo cấu hình để kết hợp các tham số DirAC để thu được cảnh âm thanh kết hợp.

12. Thiết bị theo một điểm trong số các điểm trên, trong đó bộ chuyển đổi định dạng (120) bao gồm:

bộ phân tích DirAC (180) cho định dạng đầu vào Ambisonics bậc nhất hoặc định dạng đầu vào Ambisonics bậc cao hoặc định dạng tín hiệu đa kênh;

bộ chuyển đổi siêu dữ liệu (150, 125, 126, 148) để chuyển đổi siêu dữ liệu đối tượng thành siêu dữ liệu DirAC hoặc để chuyển đổi tín hiệu đa kênh có vị trí bất biến theo thời gian thành siêu dữ liệu DirAC; và

bộ kết hợp siêu dữ liệu (144) để kết hợp các luồng siêu dữ liệu DirAC riêng lẻ hoặc kết hợp siêu dữ liệu hướng đến từ một số luồng bằng phép cộng có trọng số,

trọng số của phép cộng có trọng số được thực hiện theo các năng lượng của các năng lượng tín hiệu áp suất liên quan hoặc, để kết hợp siêu dữ liệu khuếch tán từ một số luồng bằng phép cộng có trọng số, trọng số của phép cộng có trọng số được thực hiện theo các năng lượng của các năng lượng tín hiệu áp suất liên quan, hoặc

trong đó bộ kết hợp siêu dữ liệu (144) được tạo cấu hình để tính toán, cho ngăn thời gian/tần số của mô tả thứ nhất của cảnh thứ nhất, giá trị năng lượng, và giá trị hướng đến, và để tính toán cho ngăn thời gian/tần số của mô tả thứ hai của cảnh thứ hai, giá trị năng lượng và giá trị hướng đến, và trong đó bộ kết hợp định dạng (140) được tạo cấu hình để nhân năng lượng thứ nhất với giá trị hướng đến thứ nhất và để cộng kết quả nhân của giá trị năng lượng thứ hai và giá trị hướng đến thứ hai để thu được giá trị hướng đến kết hợp hoặc, thay vào đó, để chọn giá trị hướng đến trong số giá trị hướng đến thứ nhất và giá trị hướng đến thứ hai có liên quan tới năng lượng cao hơn làm giá trị hướng đến kết hợp.

13. Thiết bị theo một điểm bất kỳ trong số các điểm trên,

còn bao gồm giao diện đầu ra (200, 300) để thêm vào định dạng kết hợp, mô tả đối tượng riêng biệt cho đối tượng âm thanh, mô tả đối tượng bao gồm ít nhất một thuộc tính trong số hướng, khoảng cách, độ khuếch tán hoặc bất kỳ thuộc tính đối tượng nào khác, trong đó đối tượng này có hướng duy nhất trong tất cả các dải tần và hoặc là tĩnh hoặc di chuyển chậm hơn ngưỡng vận tốc.

14. Phương pháp tạo mô tả cảnh âm thanh kết hợp, phương pháp bao gồm các bước:

nhận mô tả thứ nhất của cảnh thứ nhất ở định dạng thứ nhất và nhận mô tả thứ hai của cảnh thứ hai ở định dạng thứ hai, trong đó định dạng thứ hai khác với định dạng thứ nhất;

chuyển đổi mô tả thứ nhất thành định dạng chung và chuyển đổi mô tả thứ hai thành định dạng chung, khi định dạng thứ hai khác với định dạng chung; và

kết hợp mô tả thứ nhất ở định dạng chung và mô tả thứ hai ở định dạng chung để thu được mô tả cảnh âm thanh kết hợp.

15. Vật ghi đọc được bằng máy tính chứa chương trình máy tính được tạo cấu hình để thực hiện phương pháp theo điểm 14 khi chạy trên máy tính hoặc bộ xử lý.

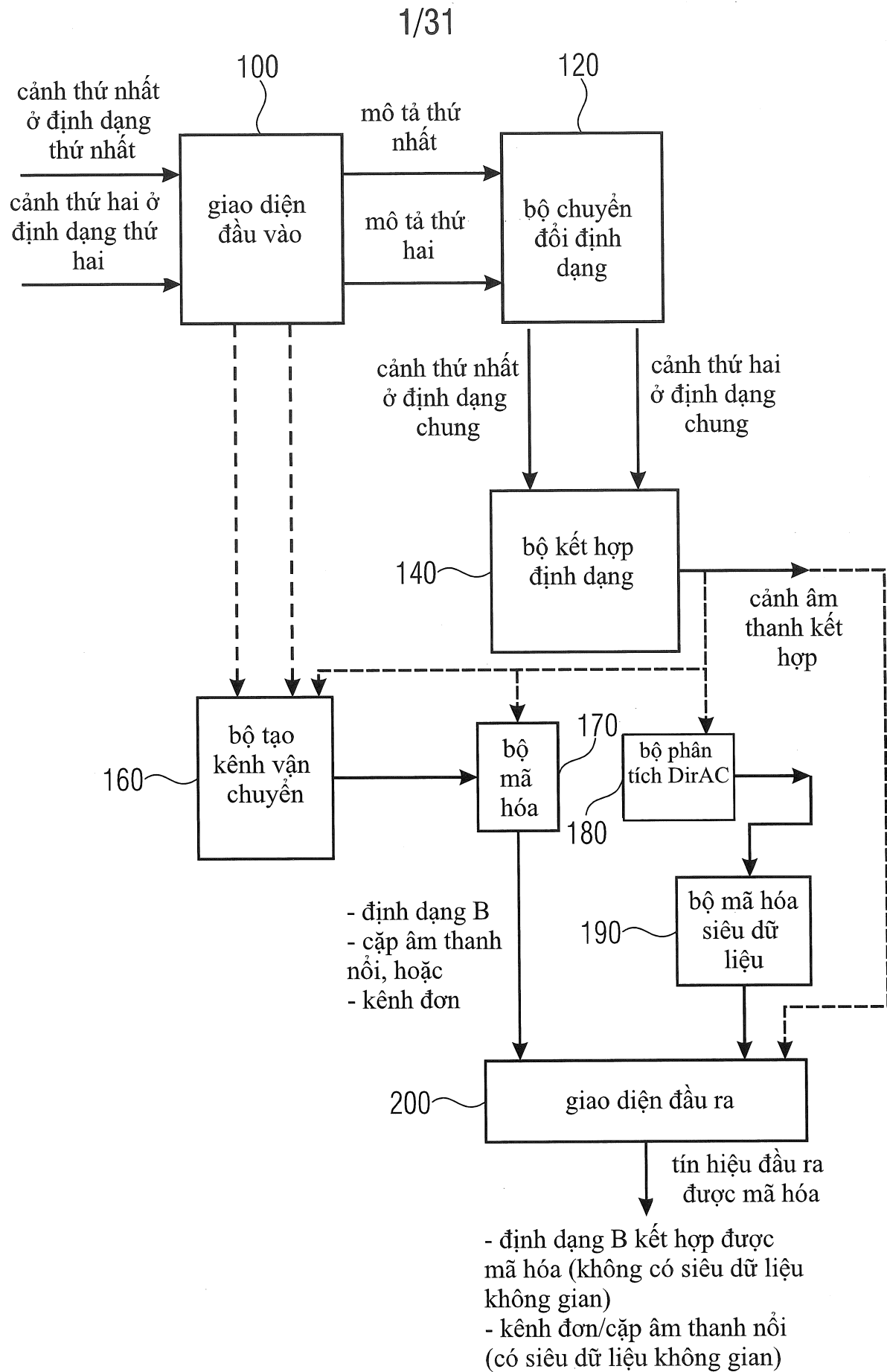
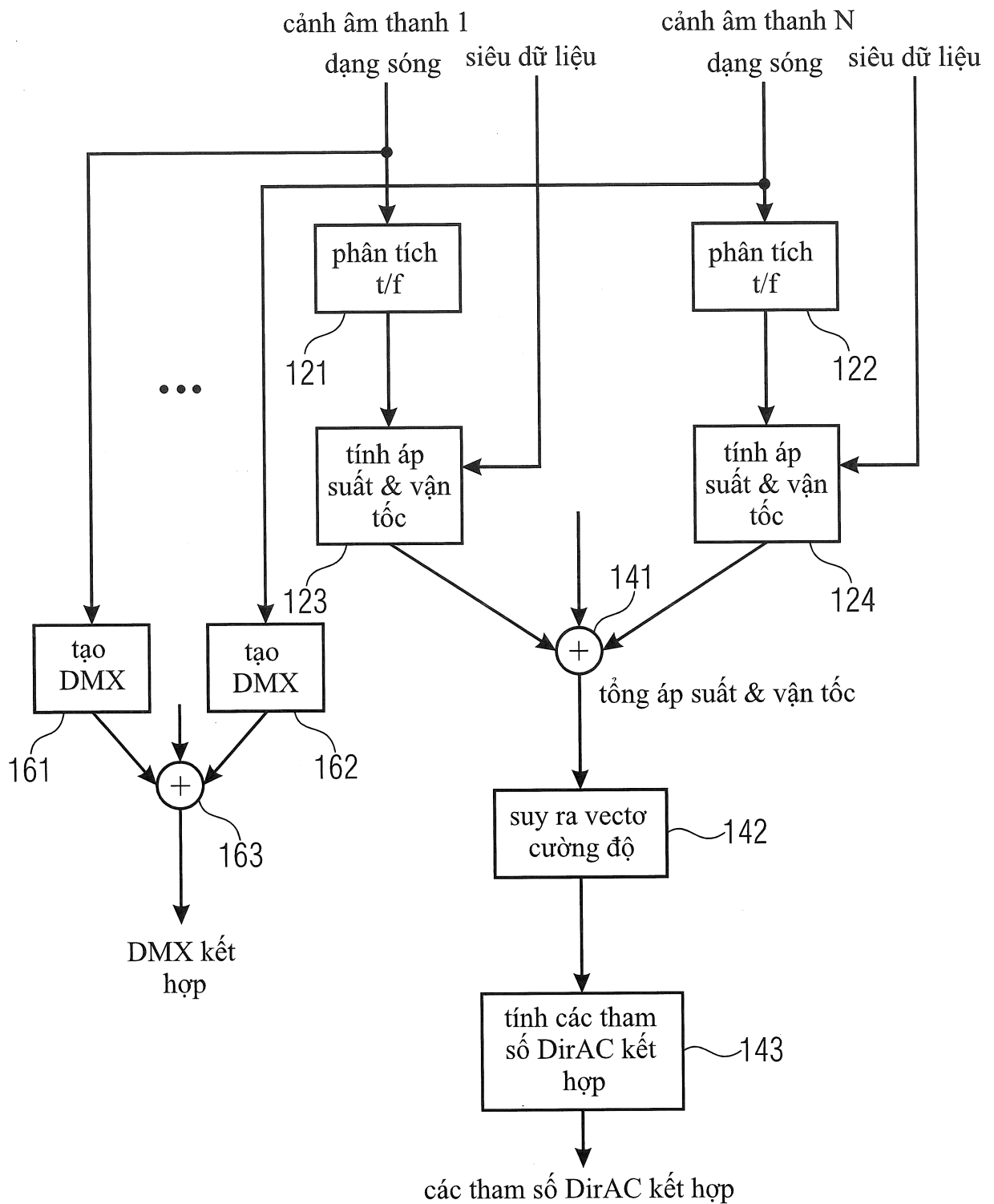


Fig. 1a

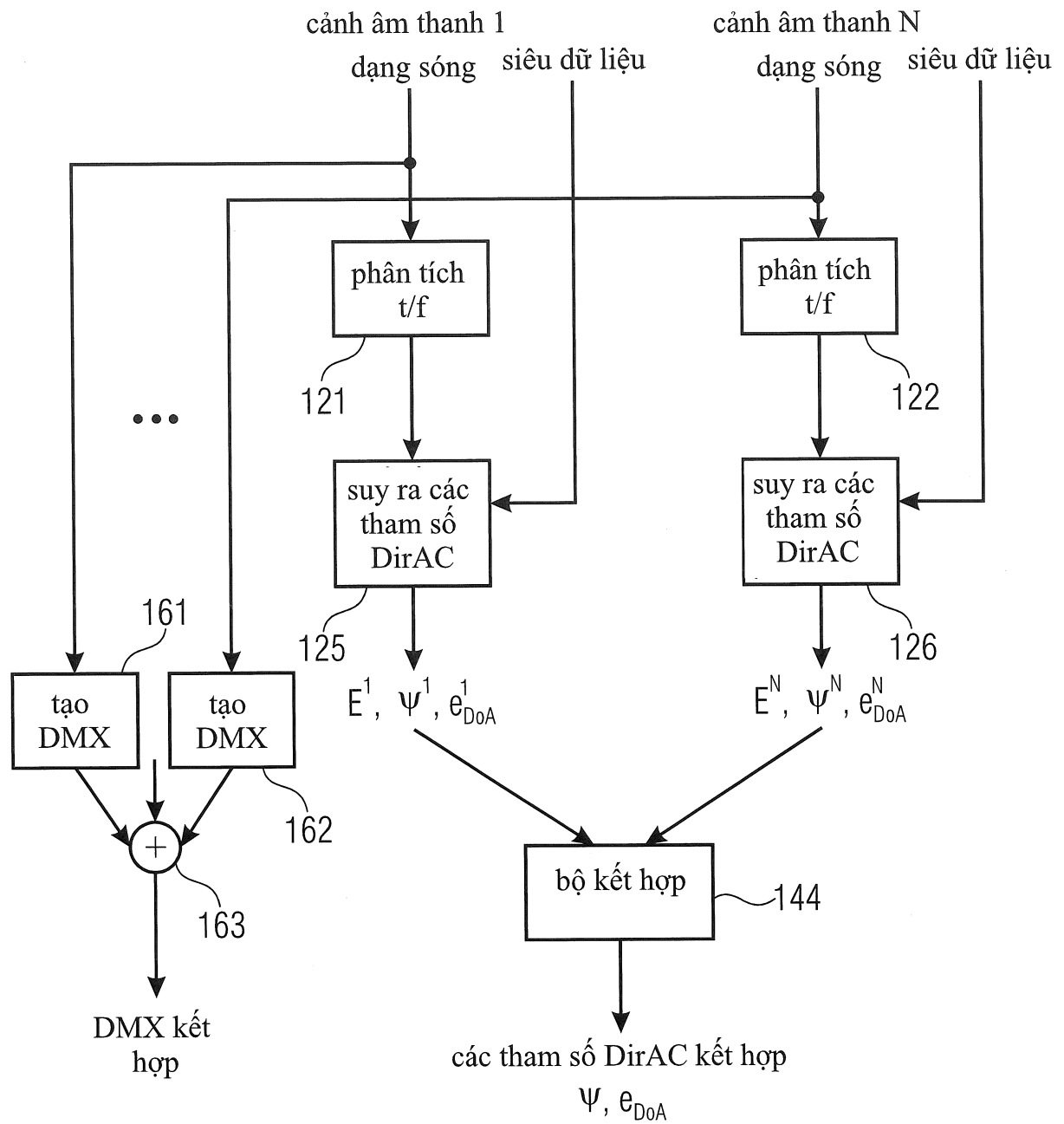
2/31



- sơ đồ khối về việc kết hợp các cảnh âm thanh khác nhau ở các định dạng khác nhau trong miền áp suất/vận tốc

Fig. 1b

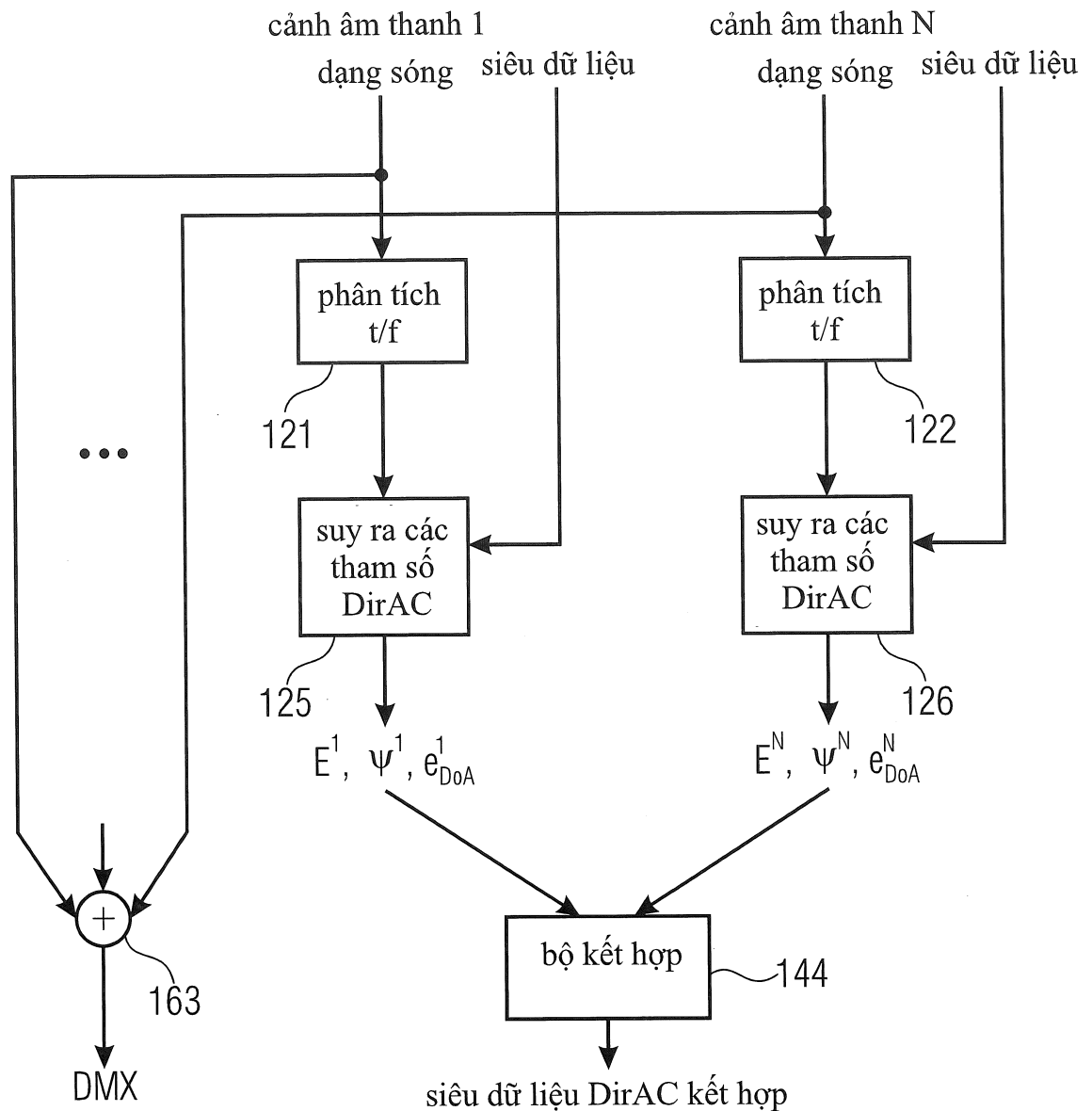
3/31



- sơ đồ khối về việc kết hợp các cảnh âm thanh khác nhau ở các định dạng khác nhau trong miền tham số DirAC

Fig. 1c

4/31



bộ kết hợp có thể là:

$$\text{alt. ①} \left\{ \begin{array}{l} \psi = \frac{1}{\sum E^i} \sum_{i=1}^N E^i \psi^i \\ e_{DoA} = \frac{1}{\sum (1 - \psi^i E^i)} \sum_{i=1}^N (1 - \psi^i E^i) E^i e_{DoA}^i \end{array} \right.$$

$$\text{alt. ②} \left\{ \begin{array}{l} \psi = 0 \quad \text{vì các đối tượng âm thanh thường không có sự khuếch tán} \\ e_{DoA} = e_{DoA}^{\text{argmax}(E^i)} \end{array} \right.$$

Fig. 1d

5/31

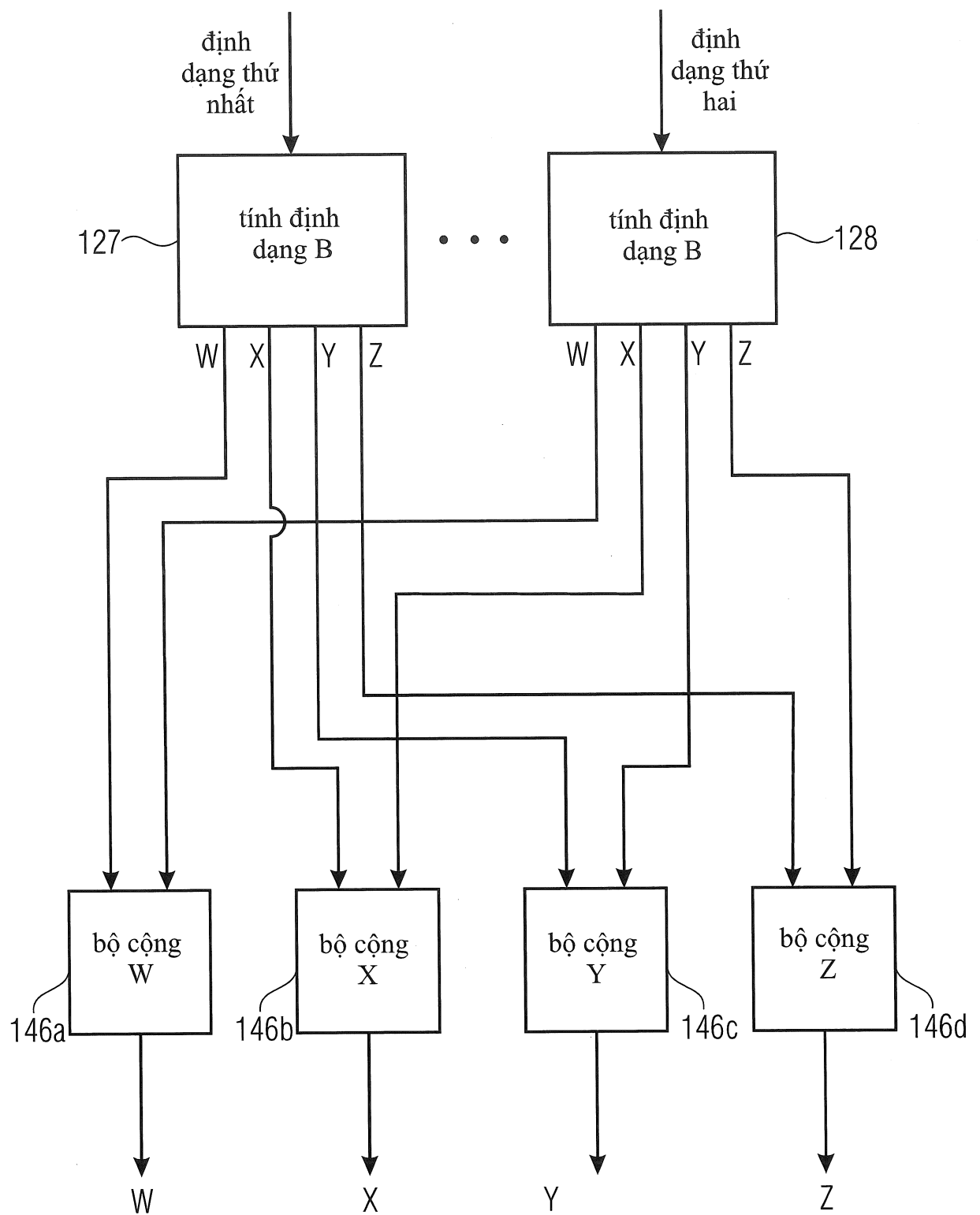
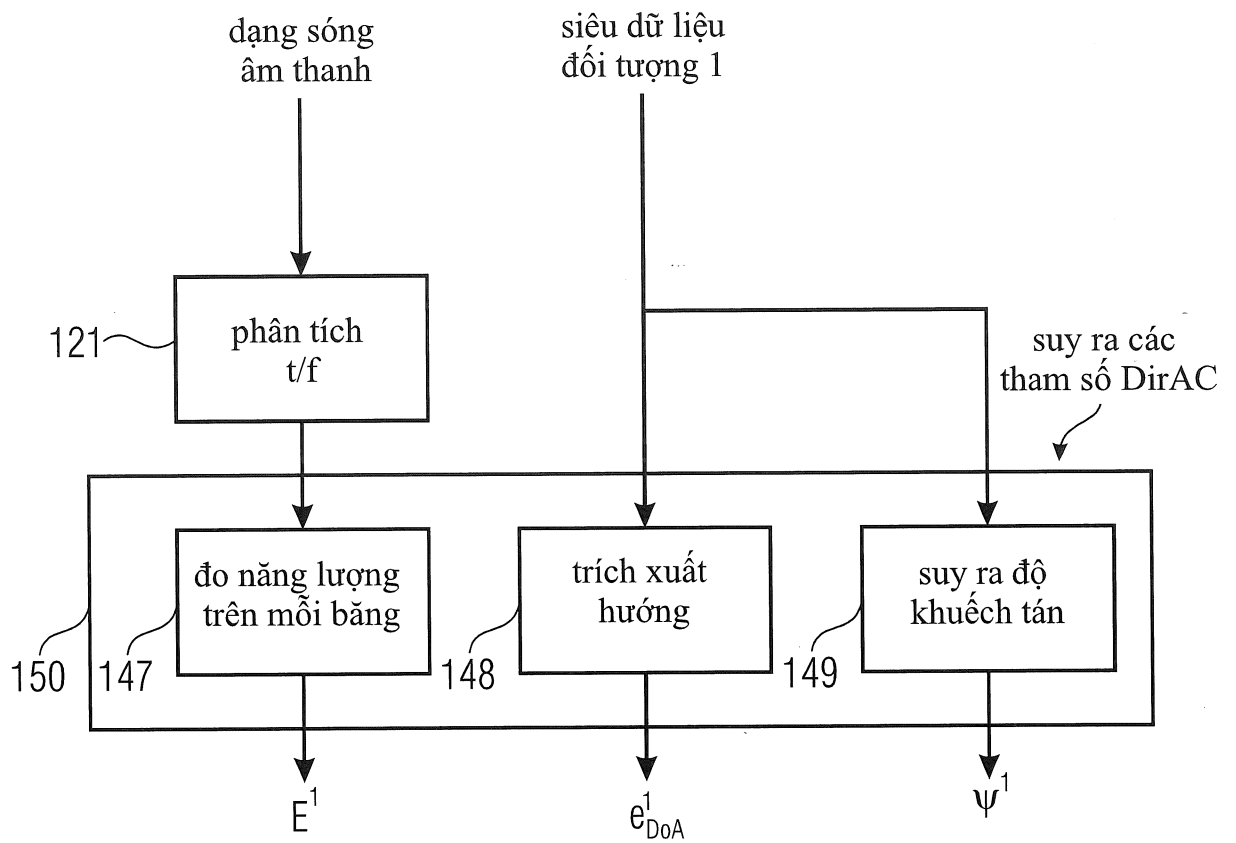


Fig. 1e

6/31



- sơ đồ khối về việc suy ra các tham số DirAC từ đối tượng âm thanh

Fig. 1f

7/31

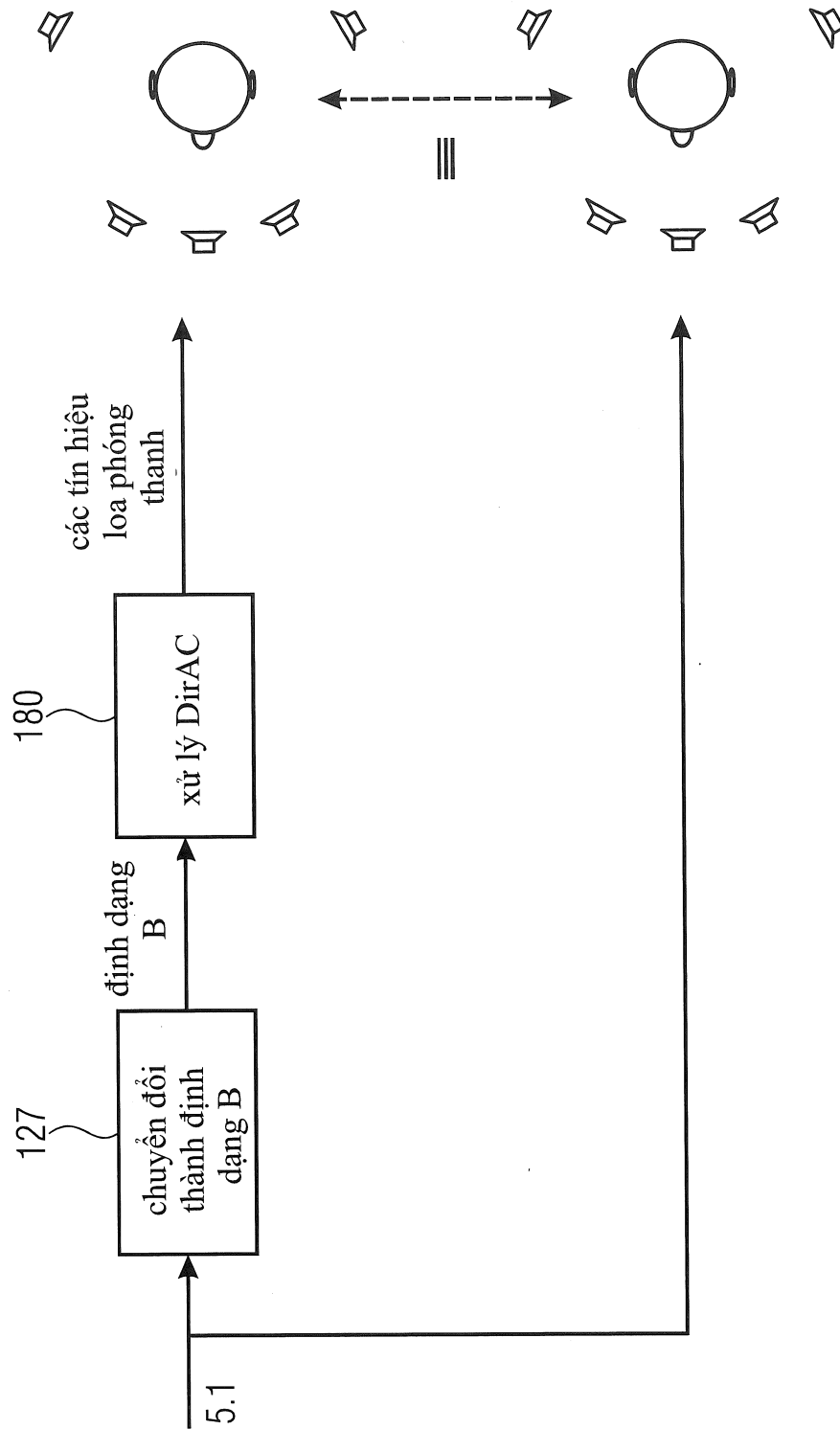


Fig. 1g

8/31

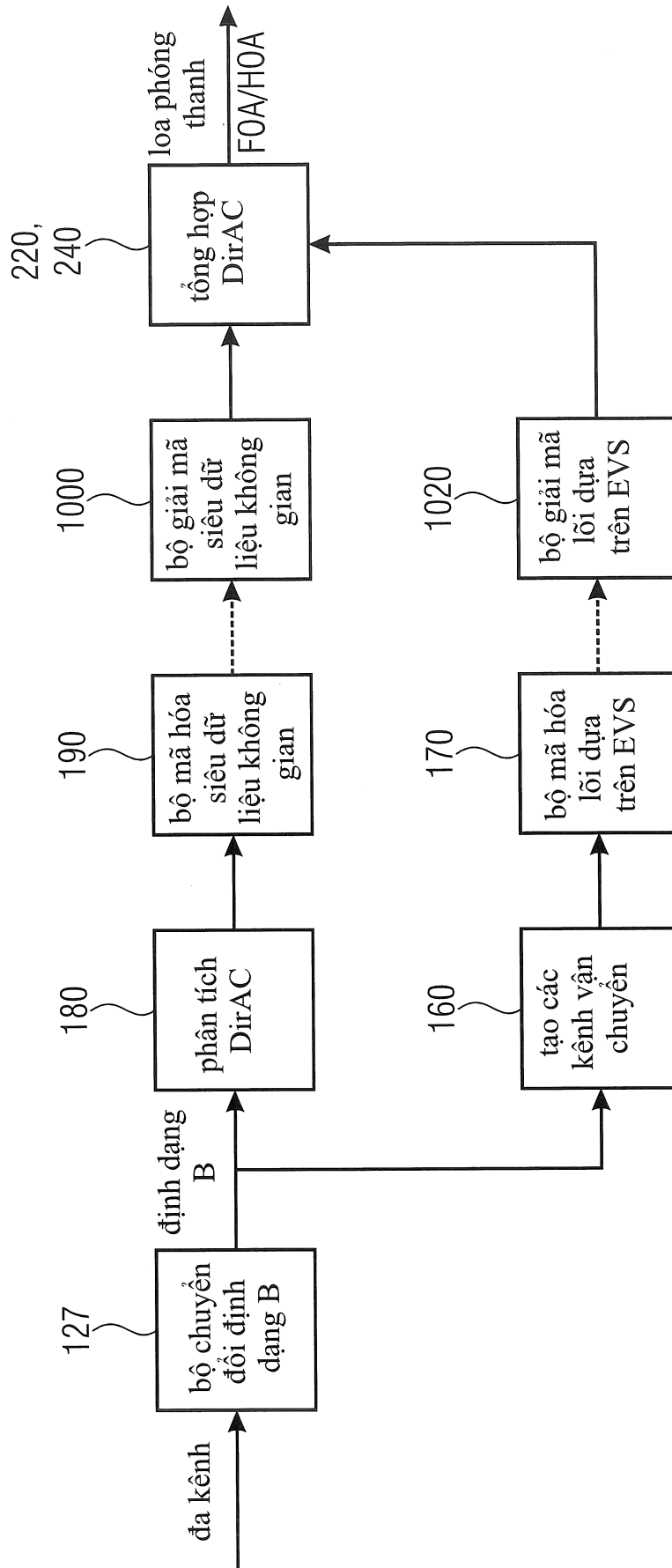
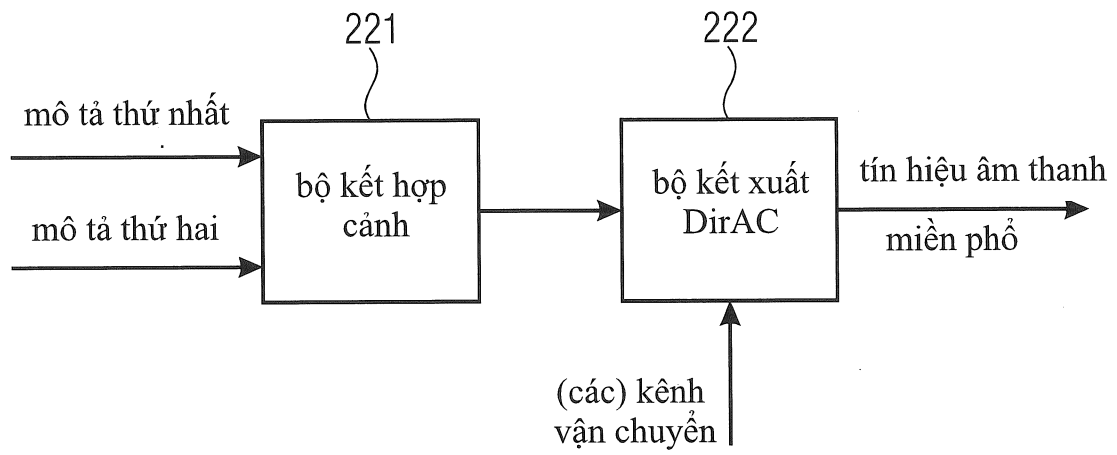


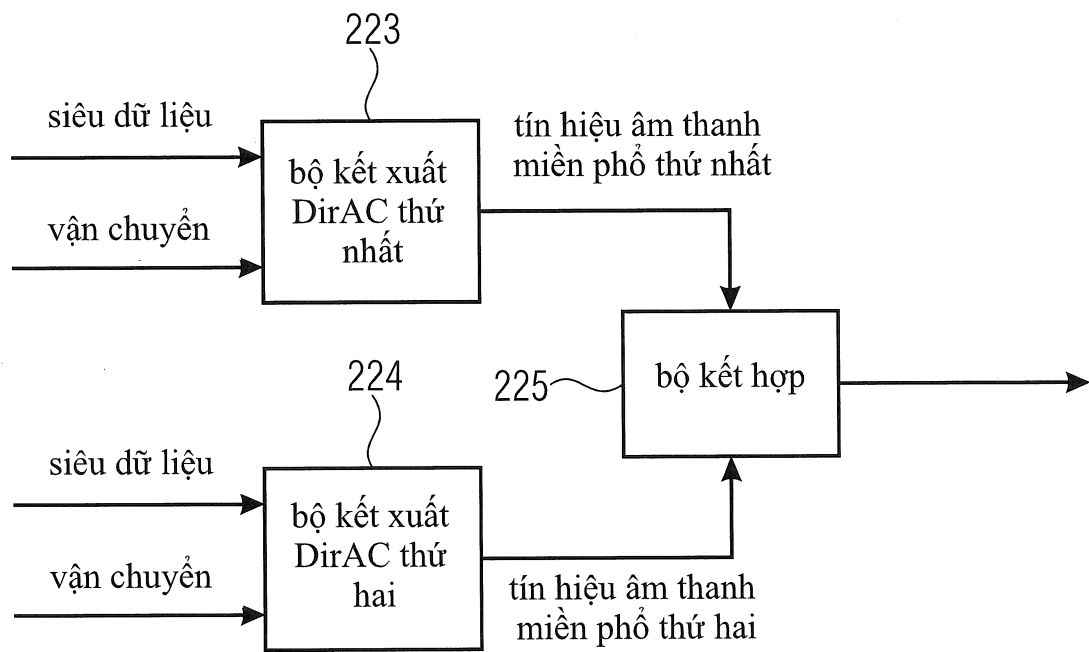
Fig. 1h

10/31



bộ tổng hợp DirAC

Fig. 2b



bộ tổng hợp DirAC

Fig. 2c

11/31

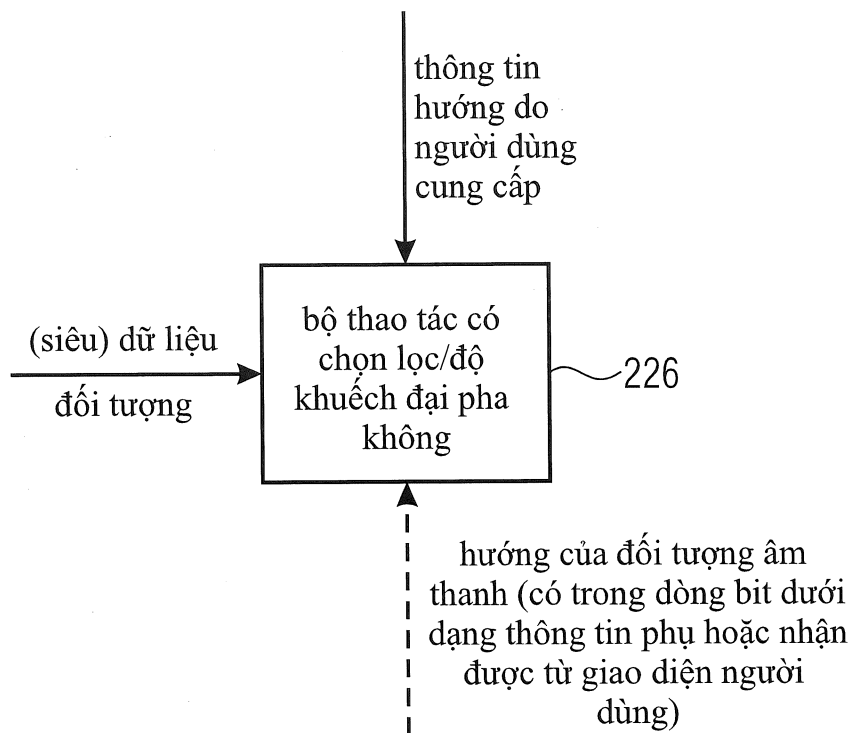


Fig. 2d

12/31

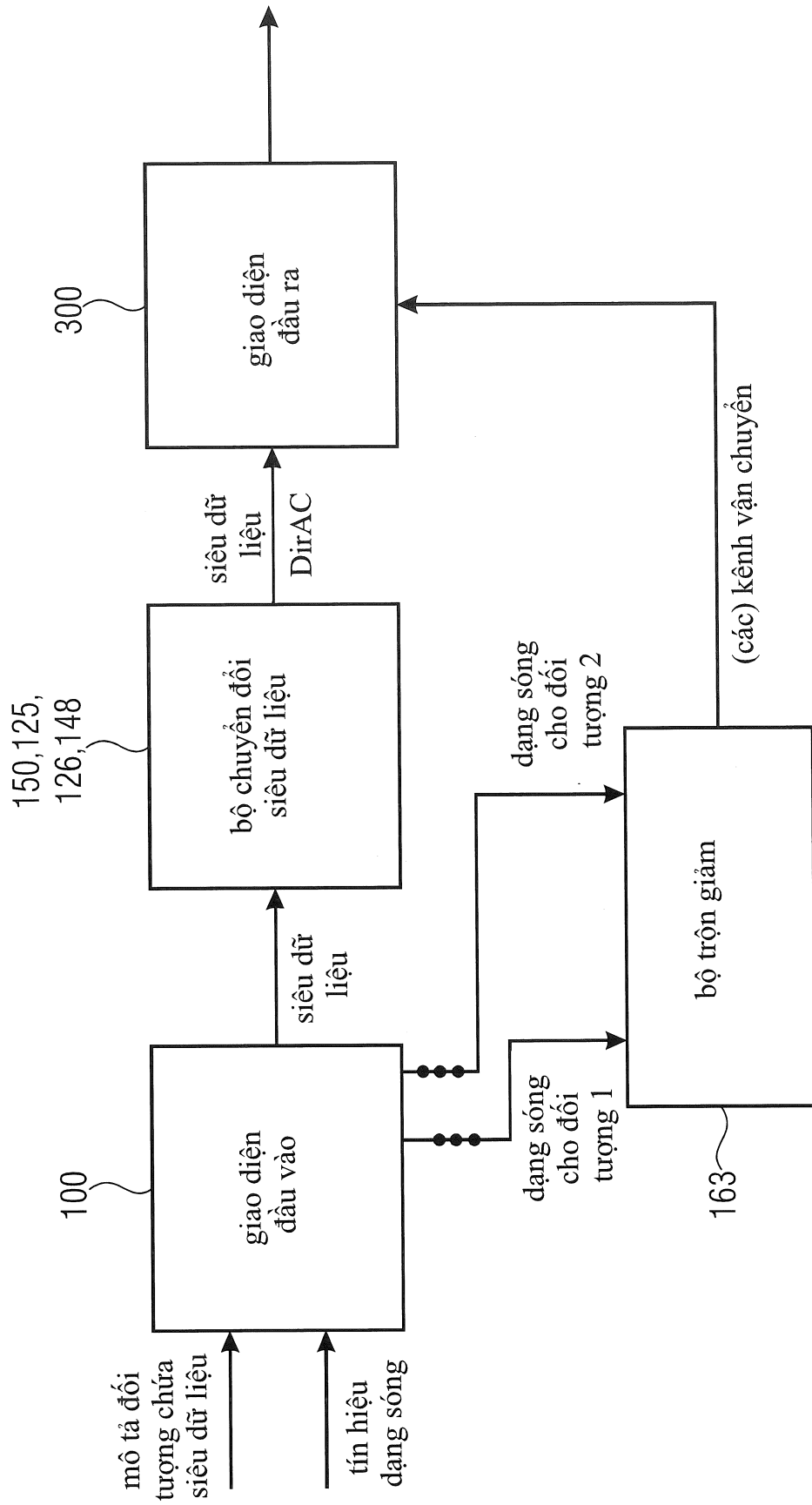


Fig. 3a

13/31

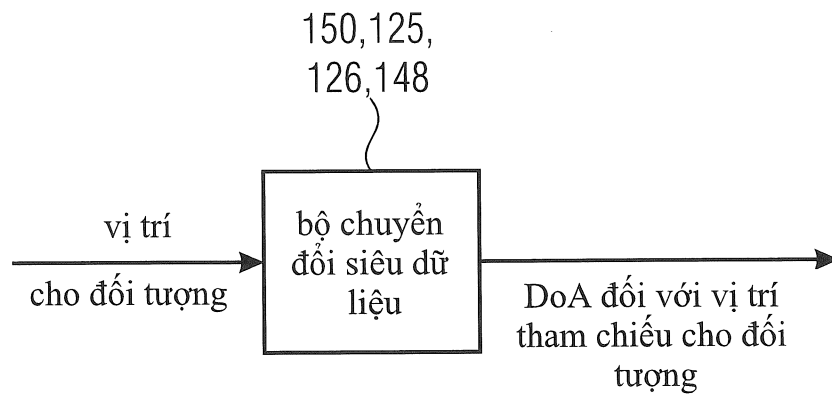


Fig. 3b

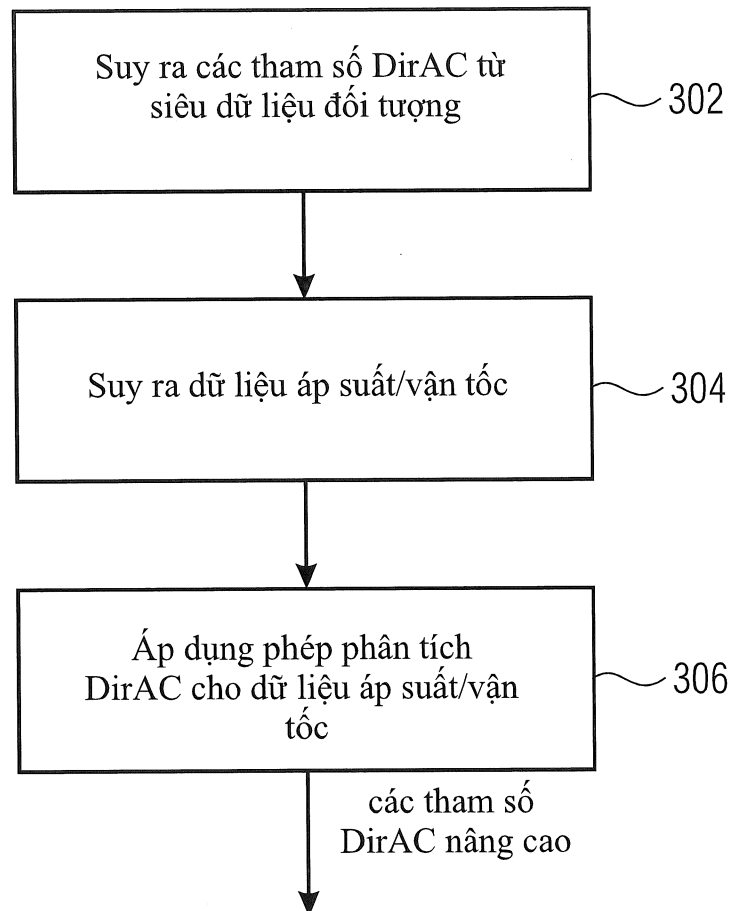


Fig. 3c

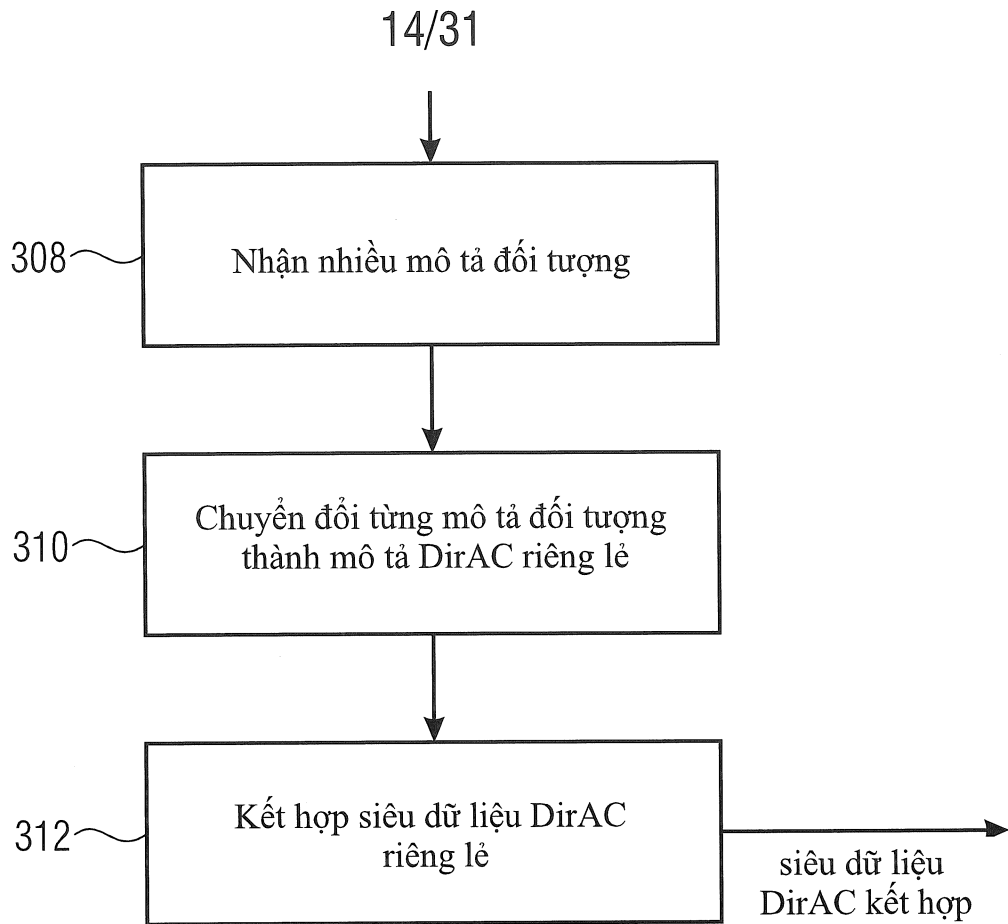


Fig. 3d

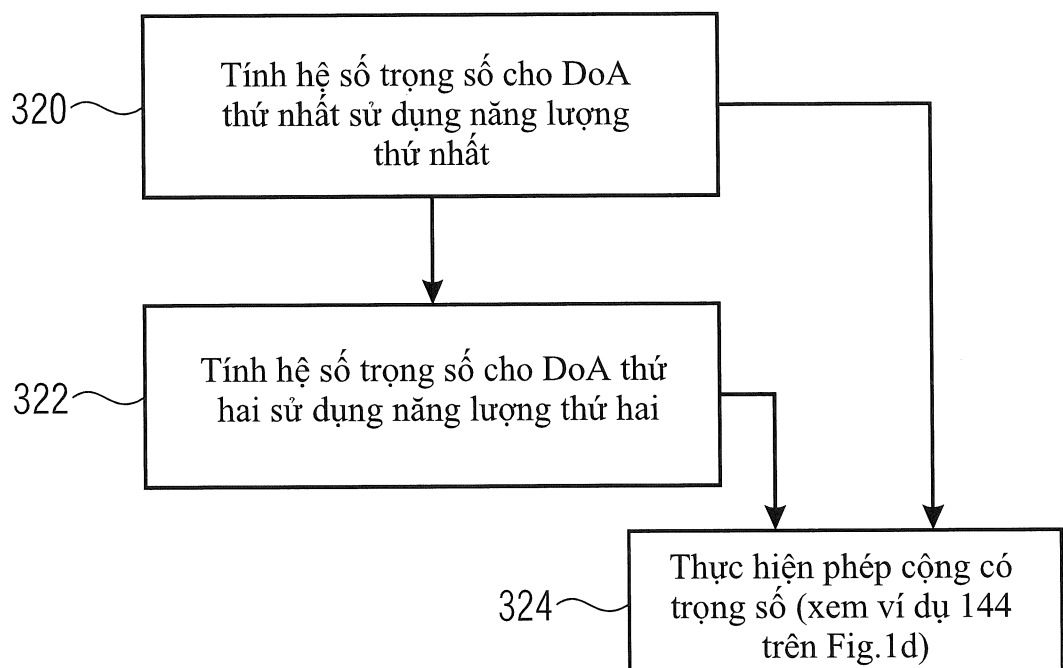


Fig. 3e

15/31

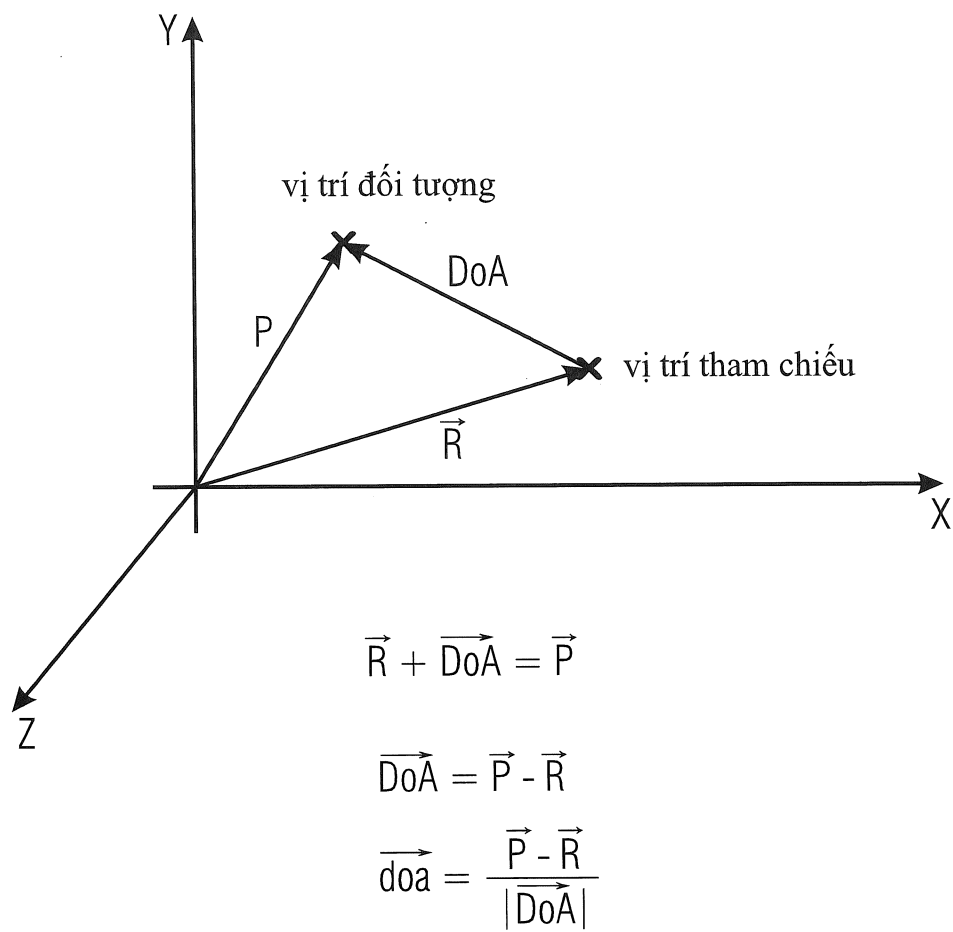


Fig. 3f

16/31

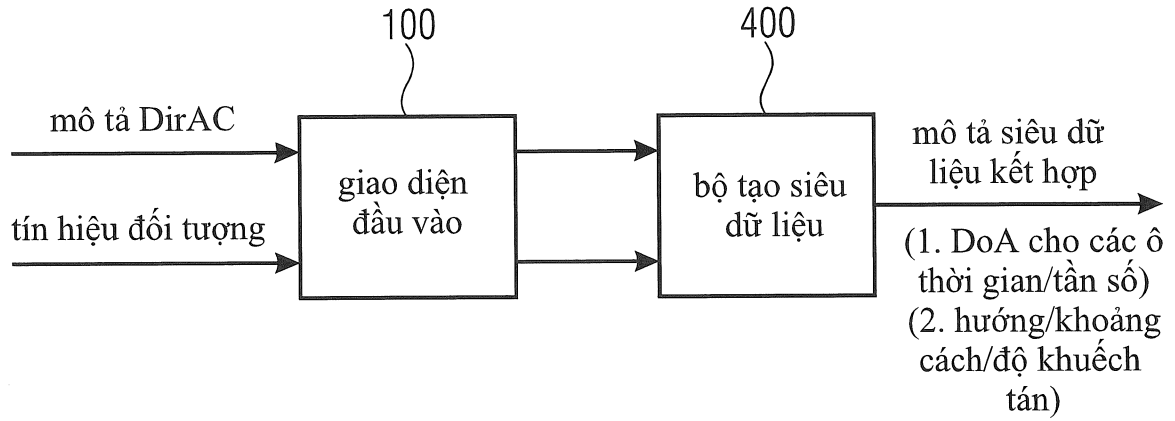


Fig. 4a

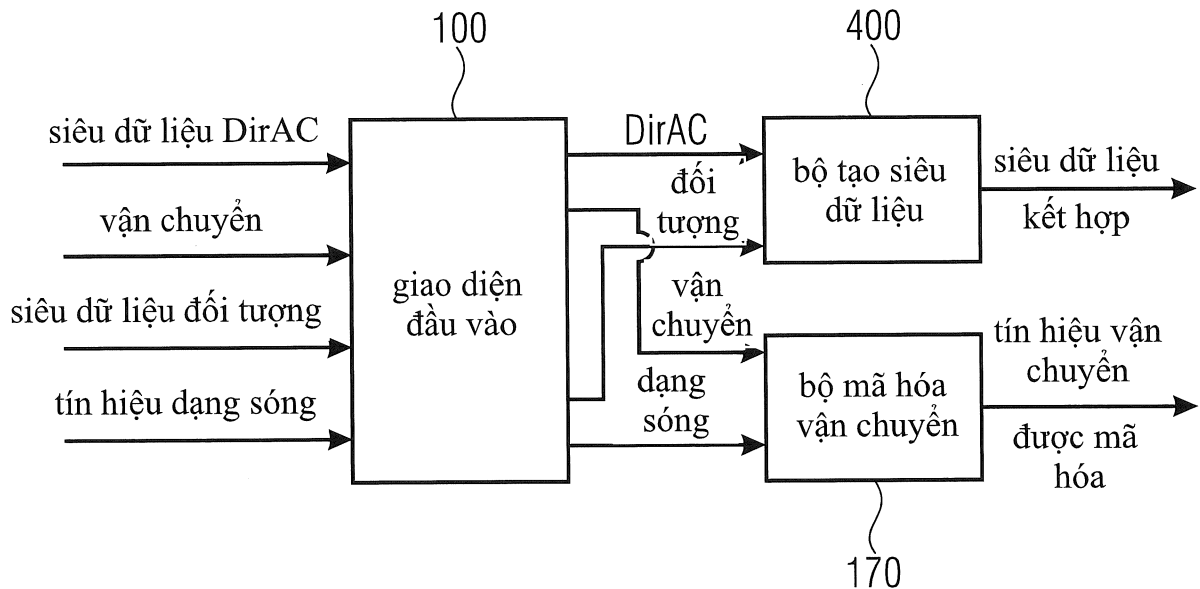


Fig. 4b

17/31

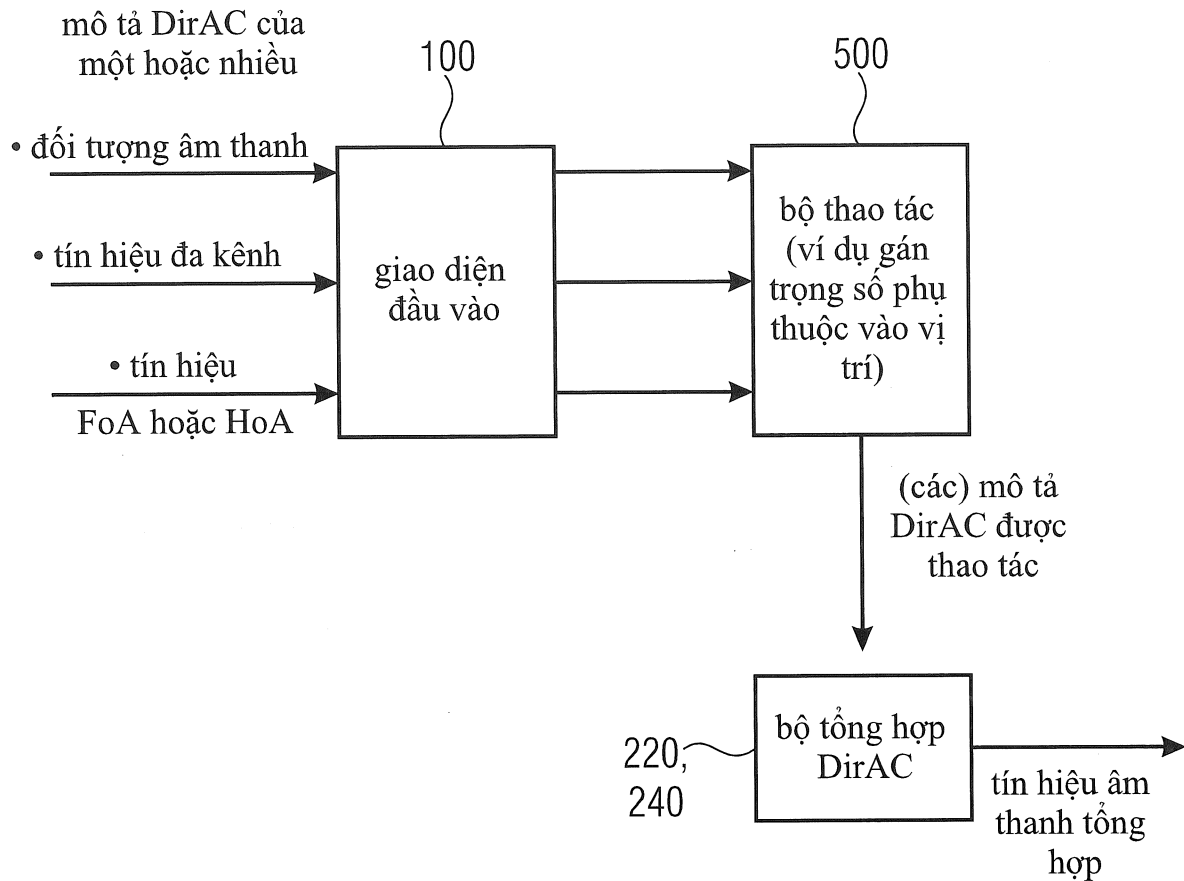


Fig. 5a

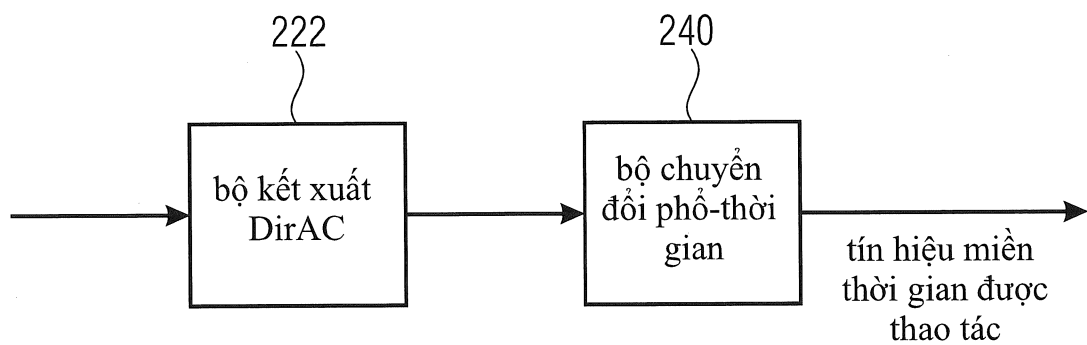


Fig. 5b

18/31

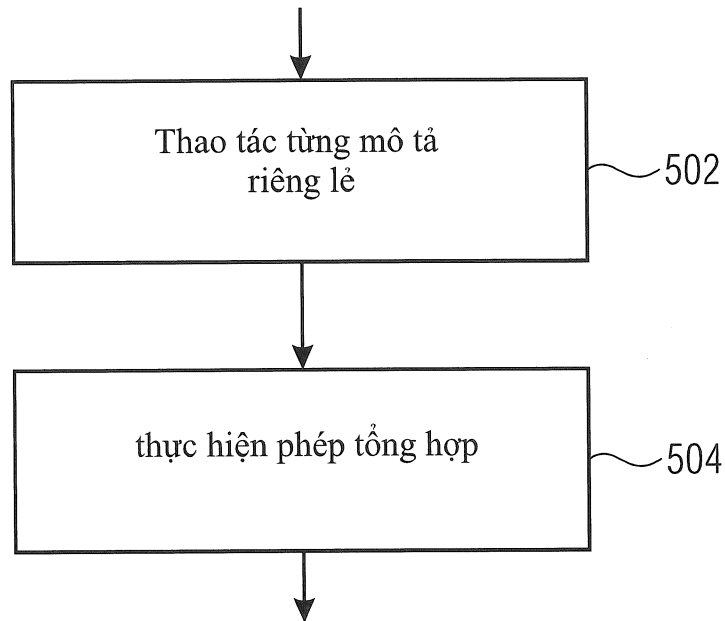


Fig. 5c

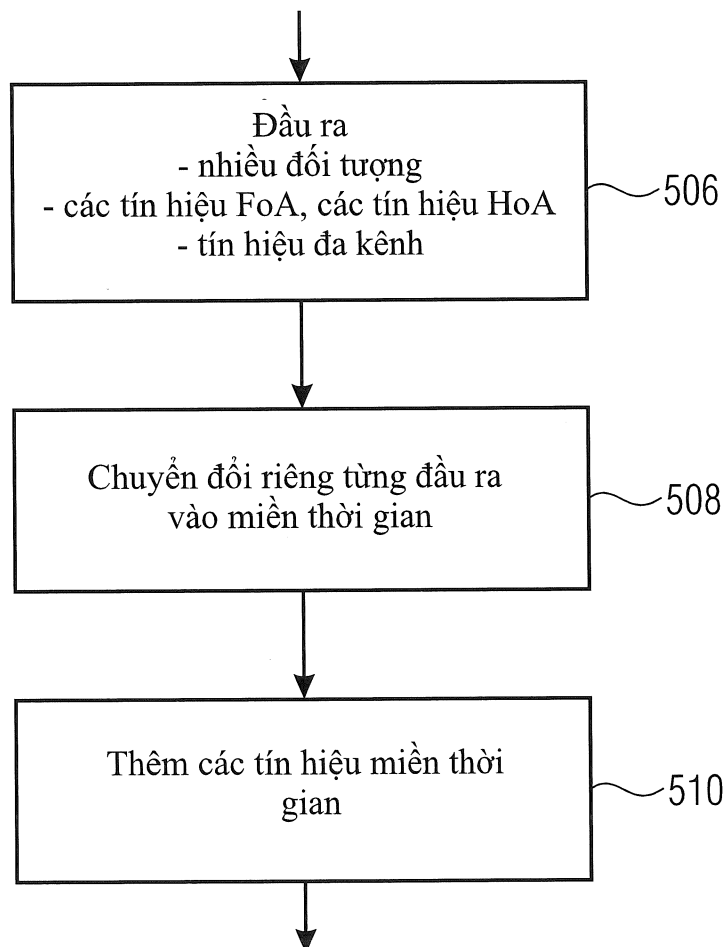
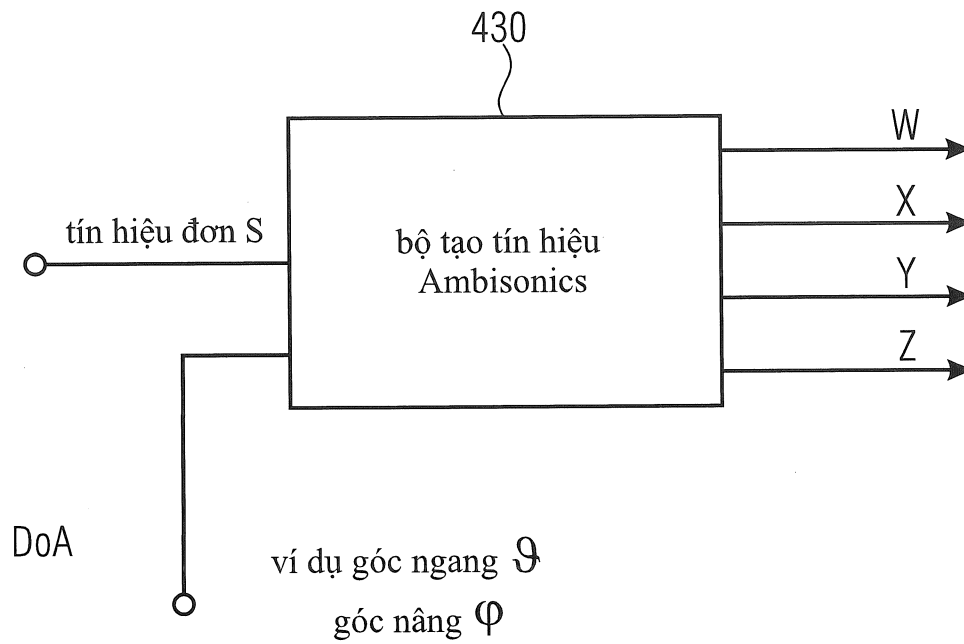


Fig. 5d

19/31



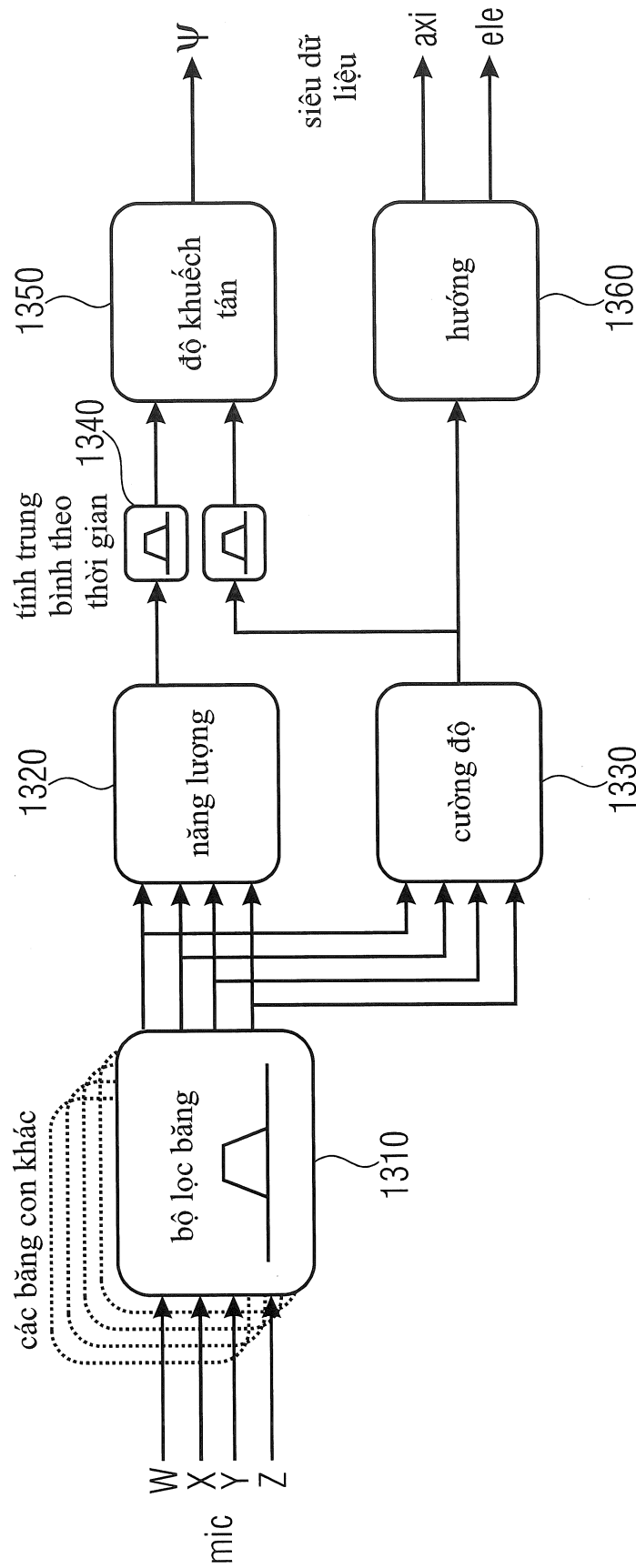
thành phần đa hướng $W = S \cdot \frac{1}{\sqrt{2}}$

các thành phần có hướng

$$\begin{cases} X = S \cdot \cos\vartheta \cos\varphi \\ Y = S \cdot \sin\vartheta \cos\varphi \\ Z = S \cdot \sin\varphi \end{cases}$$

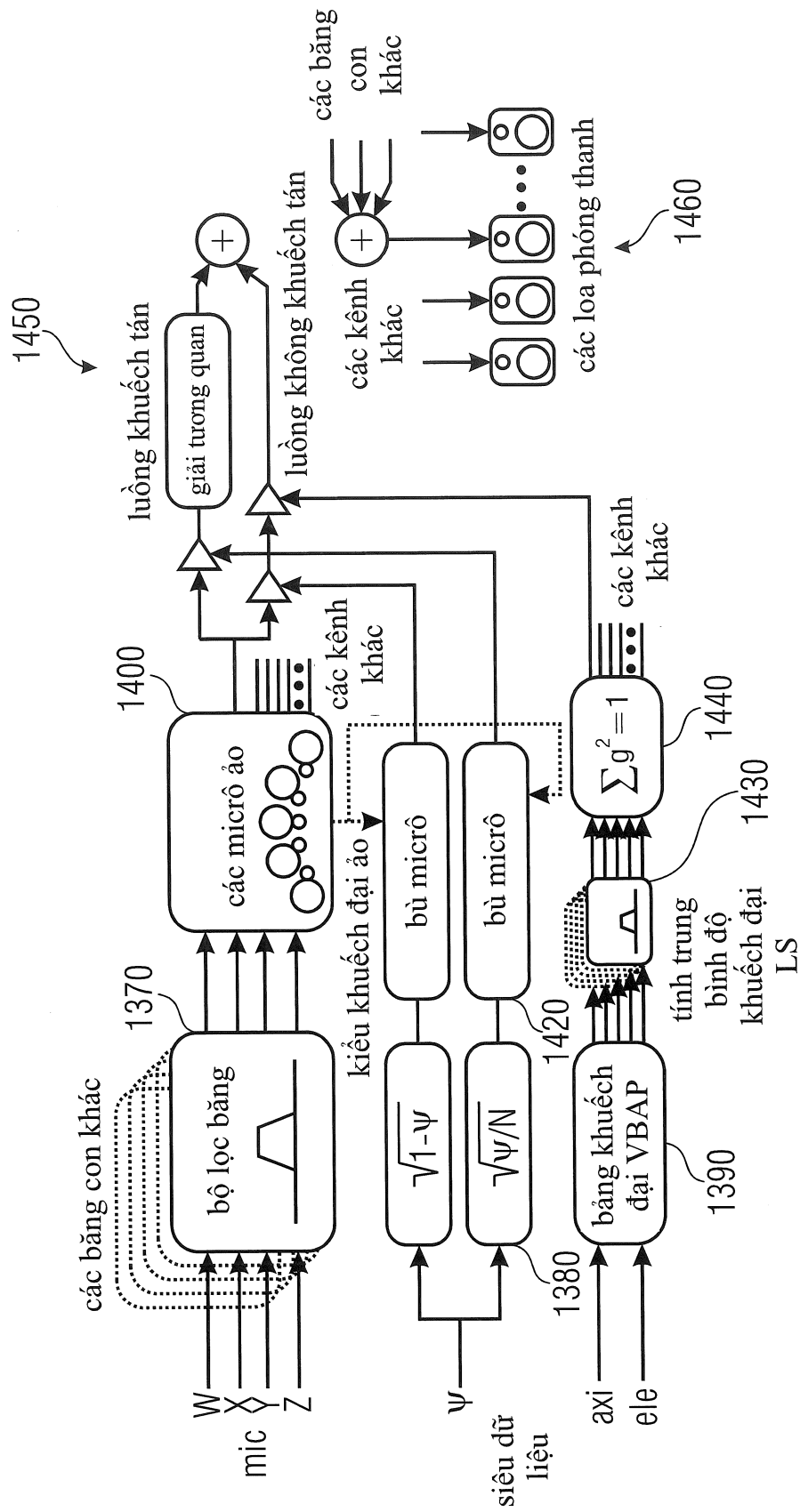
Fig. 6

20/31



bộ phân tích DirAC

Fig. 7a
(kỹ thuật trước đó)



bộ tổng hợp DirAC

Fig. 7b
(kỹ thuật trước đó)

22/31

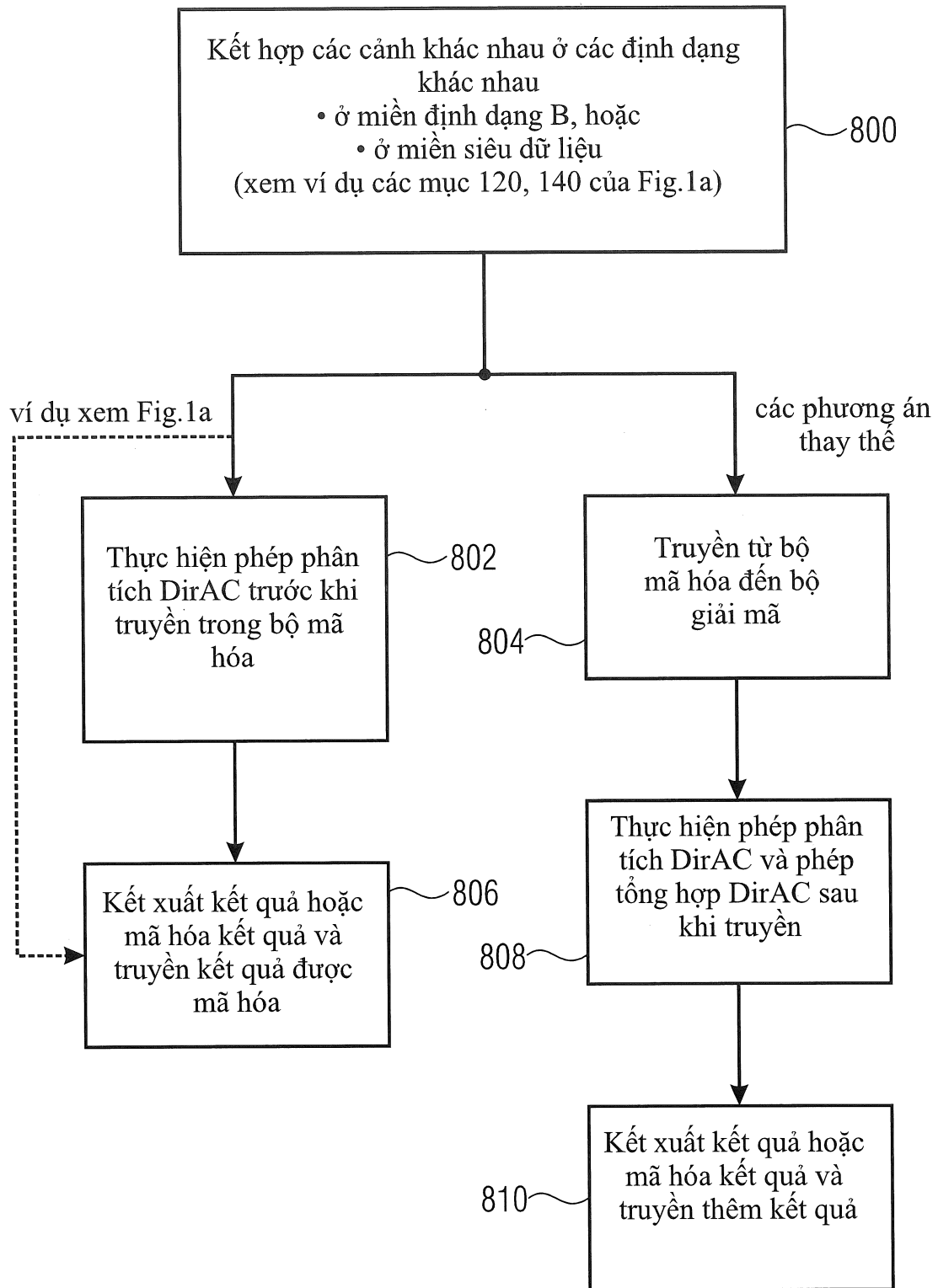


Fig. 8

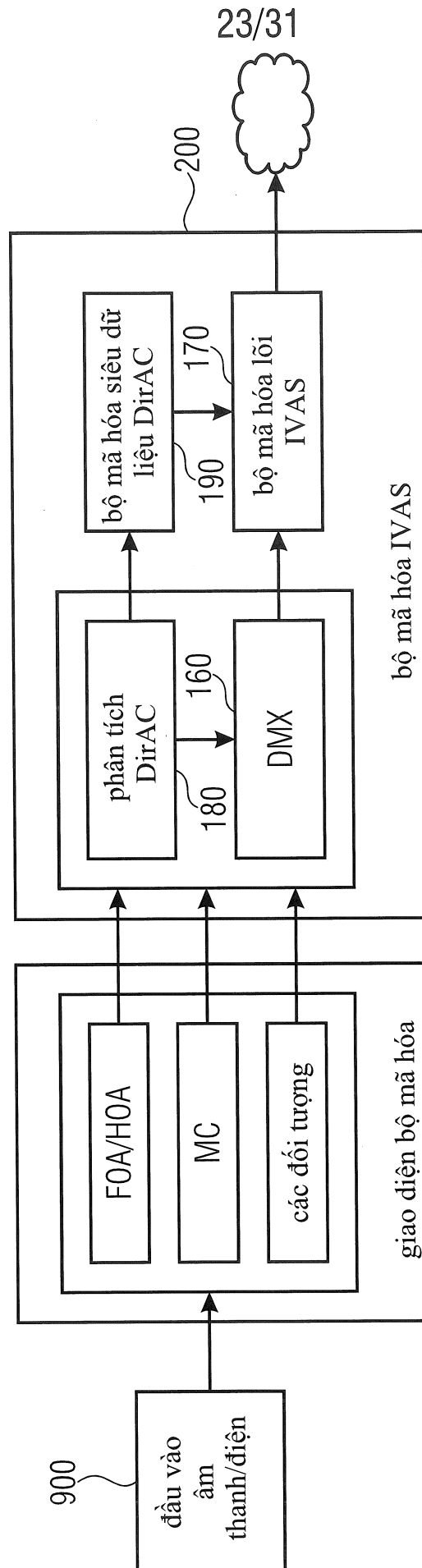


Fig. 9

24/31

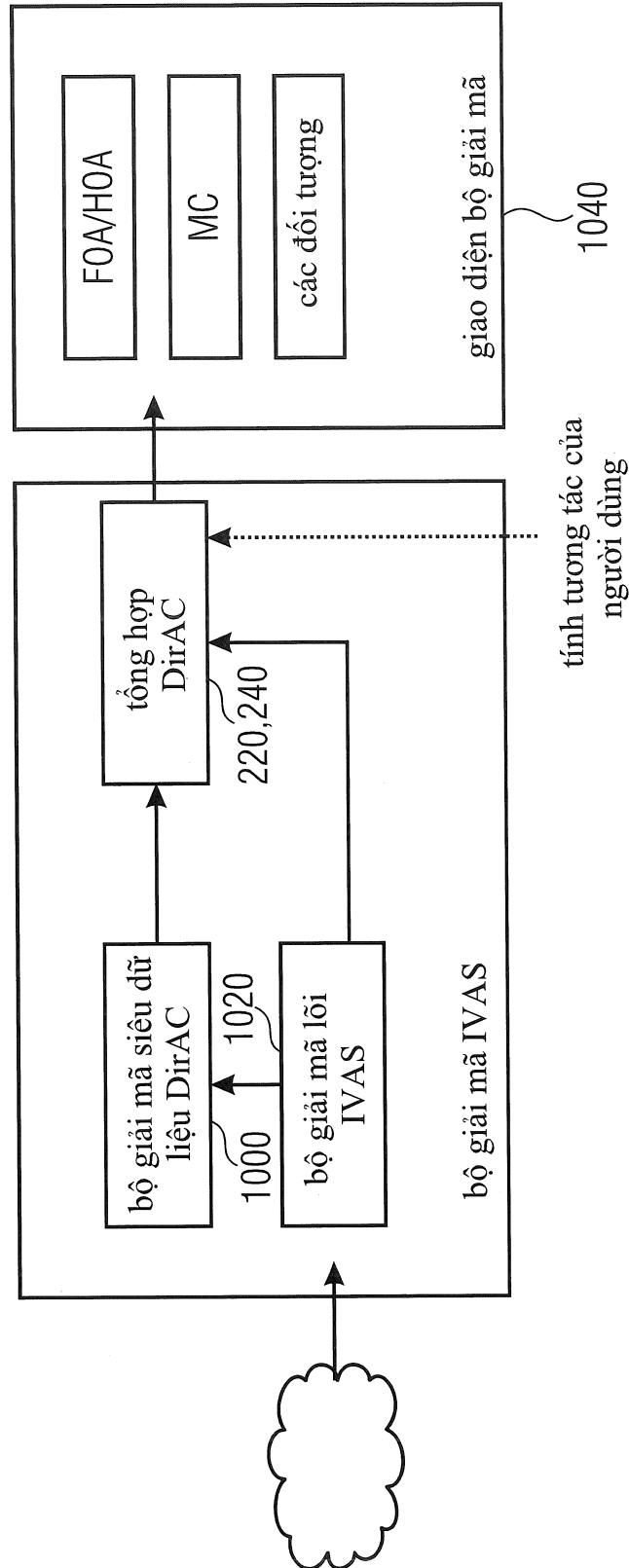


Fig. 10

25/31

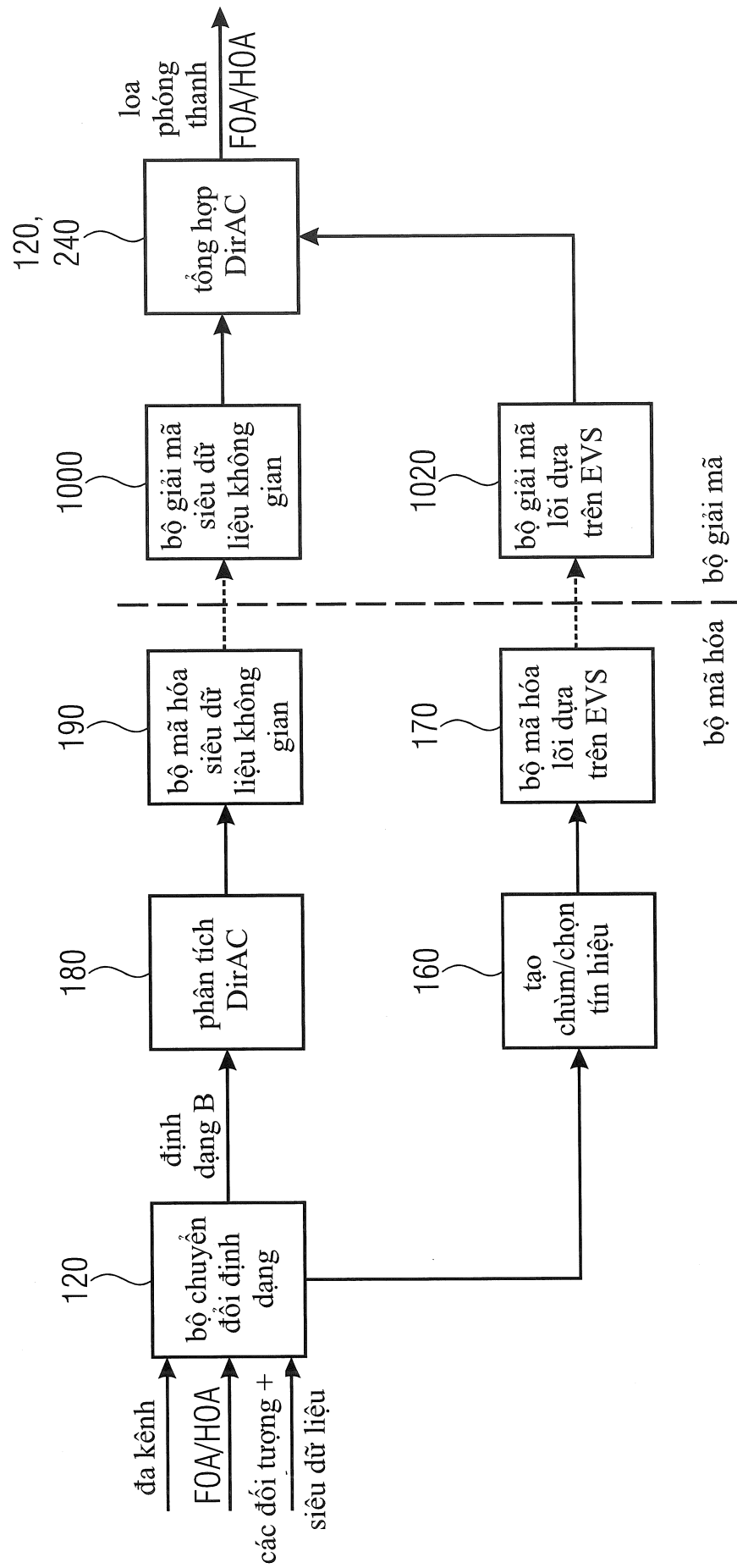


Fig. 11

27/31

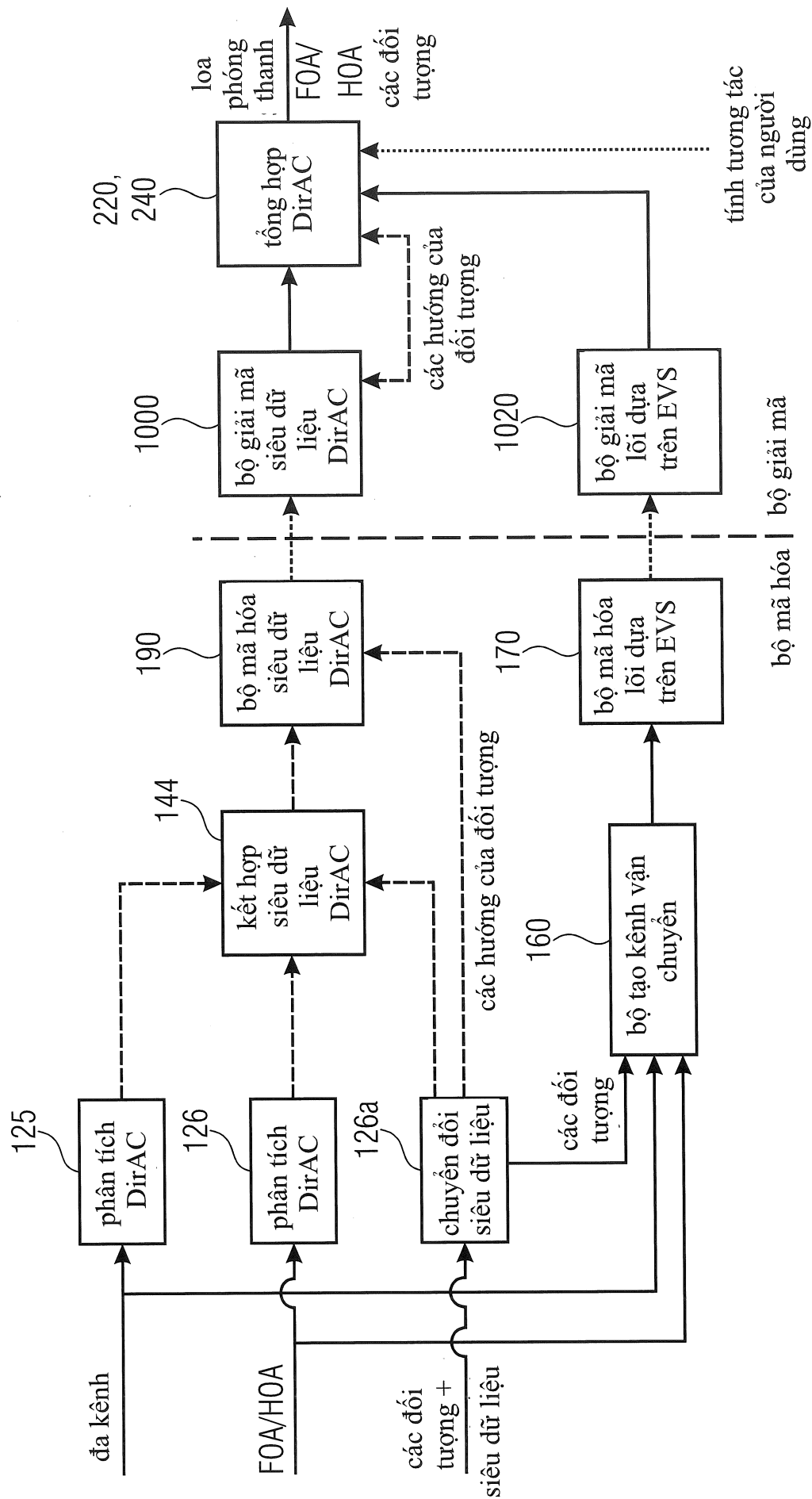


Fig. 13

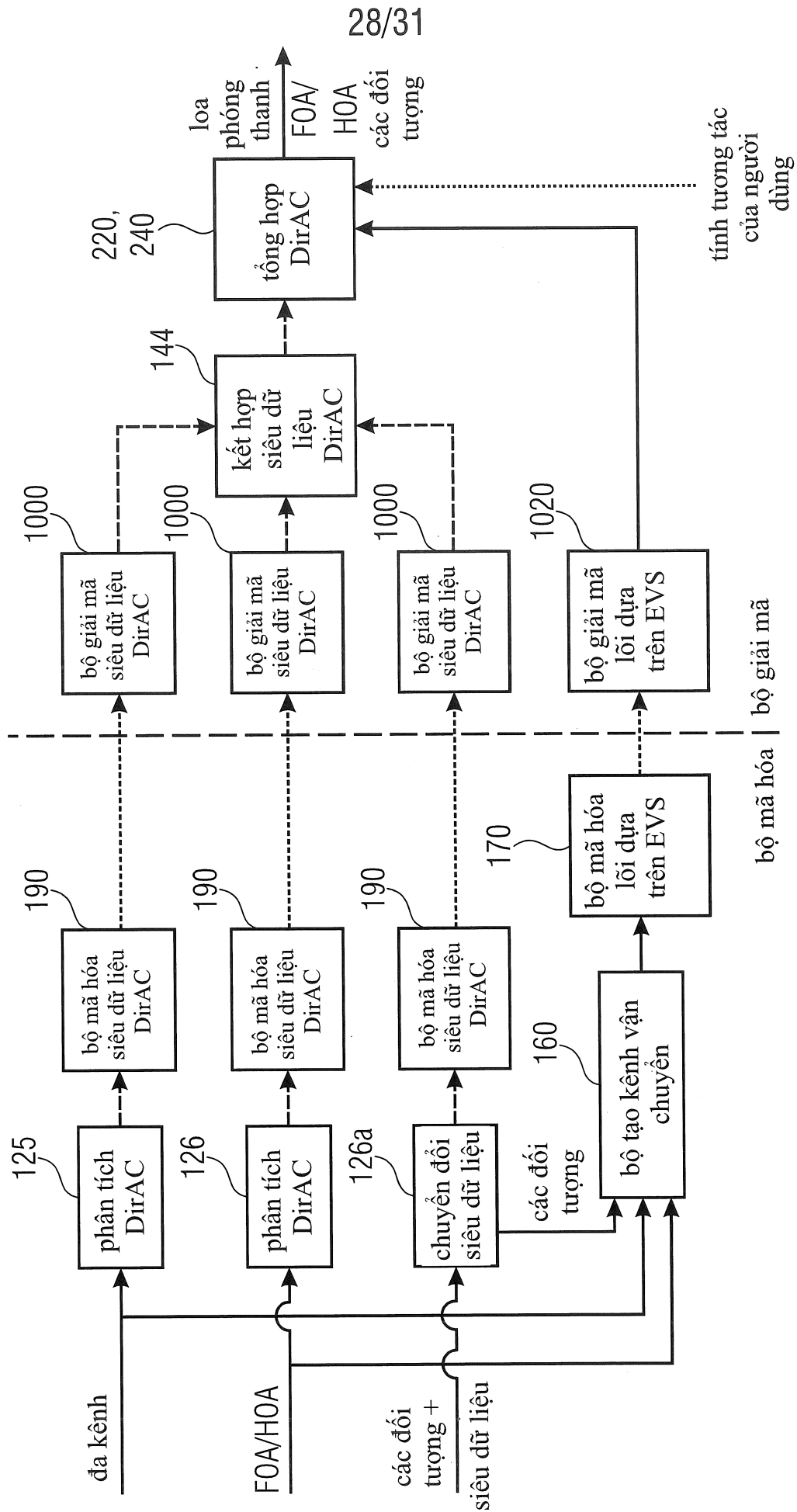


Fig. 14

29/31

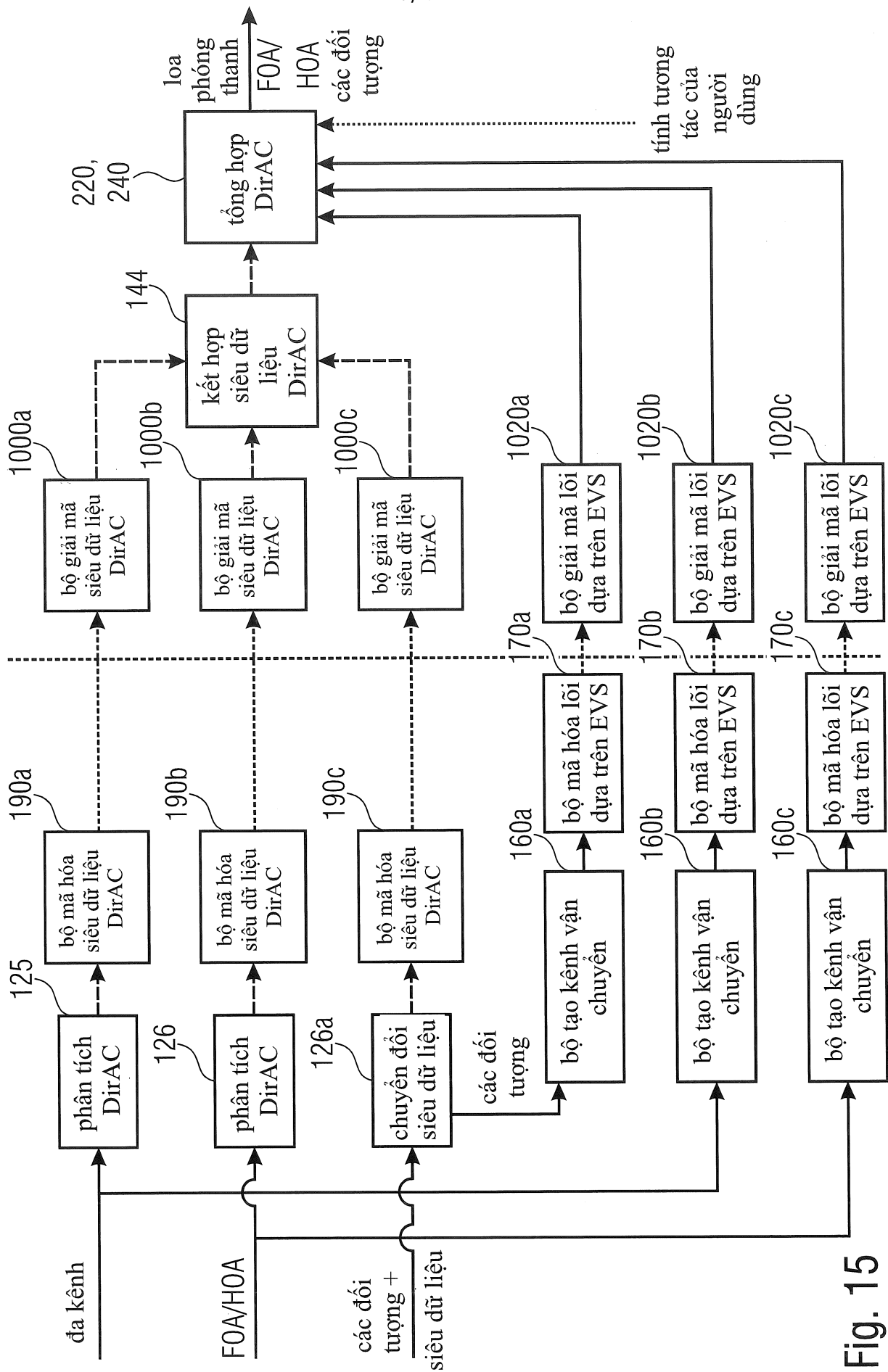


Fig. 15

	tín hiệu dạng sóng của đối tượng được mã hóa 1 (kênh đơn)	vị trí của đối tượng 1 trên mỗi khung thời gian	tín hiệu dạng sóng của đối tượng được mã hóa 1	vị trí của đối tượng 2	• • •
--	---	---	---	---------------------------	-------

Fig. 16a

	trộn giảm đối tượng (đơn âm/âm thanh nổi/...)	siêu dữ liệu đối tượng (ví dụ năng lượng đối tượng, sự tương quan trên mỗi ngăn thời gian/tần số	các vị trí của đối tượng (tùy chọn)	có thể được cho/biến đổi bởi người dùng ví dụ SAOC	
--	--	---	--	--	--

Fig. 16b

	kênh thứ nhất ví dụ L	kênh thứ hai ví dụ R	kênh thứ ba ví dụ C	kênh thứ tư ví dụ LS	kênh thứ năm ví dụ RS	ĐA KÊNH
--	--------------------------	-------------------------	------------------------	-------------------------	--------------------------	---------

Fig. 16c

	trộn giảm kênh (đơn âm/âm thanh nổi/...)	thông tin phụ theo tham số dưới dạng siêu dữ liệu kênh cho ngăn thời gian/tần số	
			ví dụ ÂM THANH VÒM MPEG

Fig. 16d

	W	X	Y	Z	các thành phần cao hơn tùy chọn	F0A H0A	31/31
--	---	---	---	---	---------------------------------	------------	-------

Fig. 16e

	trộn giảm DirAC (đơn âm hoặc âm thanh nổi ...)	thông tin phụ theo tham số (hướng đến, (tùy chọn) độ khuếch tán trên mỗi ngăn thời gian/tần số)	
			DirAC

Fig. 16f