



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) 
CỤC SỞ HỮU TRÍ TUỆ
1-0043547
(51)^{2019.01} **G10L 19/008**; H04S 7/00; G10L (13) **B**
21/0272; G06F 3/01

(21) 1-2020-00765 (22) 13/07/2018
(86) PCT/EP2018/069140 13/07/2018 (87) WO 2019/012131 17/01/2019
(30) 17181488.2 14/07/2017 EP
(45) 25/02/2025 443 (43) 25/05/2020 386A
(73) FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (DE)
Hansastraße 27c, 80686 Muenchen, DE
(72) HERRE, Juergen (DE); HABETS, Emanuel (NL).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP TẠO MÔ TẢ TRƯỜNG ÂM THANH NÂNG CAO

(21) 1-2020-00765

(57) Sáng chế đề cập đến thiết bị và phương pháp tạo mô tả trường âm thanh nâng cao, và vật ghi lưu trữ số không tạm thời có chương trình máy tính được lưu trữ trên đó để thực hiện phương pháp tạo mô tả trường âm thanh nâng cao. Thiết bị tạo mô tả trường âm thanh nâng cao, thiết bị bao gồm: bộ tạo trường âm thanh (100, 250, 260) để tạo ít nhất một mô tả trường âm thanh biểu thị trường âm thanh đối với ít nhất một địa điểm tham chiếu; và bộ tạo siêu dữ liệu (110) để tạo siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh, trong đó ít nhất một mô tả trường âm thanh và siêu dữ liệu tạo thành mô tả trường âm thanh nâng cao. Ít nhất một mô tả trường âm thanh bao gồm trong trường hợp nhất định mô tả trường âm thanh thứ nhất liên quan đến ít nhất một địa điểm tham chiếu và mô tả trường âm thanh thứ hai liên quan đến địa điểm tham chiếu tiếp theo khác với địa điểm tham chiếu, và siêu dữ liệu liên quan đến thông tin không gian biểu thị địa điểm tham chiếu và địa điểm tham chiếu tiếp theo hoặc một trong hai địa điểm tham chiếu và vectơ kéo dài giữa cả hai địa điểm tham chiếu.

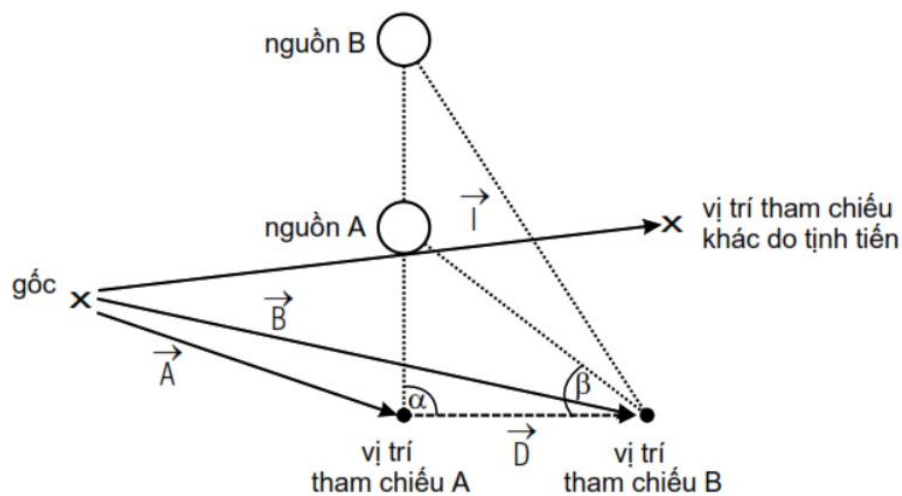


Fig. 4c

Lĩnh vực kỹ thuật được đề cập

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

Tình trạng kỹ thuật của sáng chế

Các tín hiệu ambisonic (âm thanh bao quanh hoặc âm thanh lan tỏa) bao gồm phép phân tích sóng hài hình cầu bị cắt vát của trường âm thanh. Ambisonic đến với nhiều vị khác nhau. Có ambisonic 'truyền thống' [31] mà ngày nay được biết đến là 'ambisonic cấp một' (FOA - First-Order Ambisonic) và bao gồm bốn tín hiệu (nghĩa là, một tín hiệu đẳng hướng và lên đến ba tín hiệu định hướng hình số tám). Nhiều biến thể ambisonic hơn gần đây được biết đến là 'ambisonic cấp cao' (HOA - Higher-Order Ambisonic) và cung cấp độ phân giải không gian nâng cao và vùng điểm lý tưởng người nghe lớn hơn với tùy chỉnh mang nhiều tín hiệu hơn. Nói chung, phép biểu diễn HOA cấp N được xác định đầy đủ bao gồm $(N+1)^2$ tín hiệu.

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

DirAC [32] có thể được sử dụng với các hệ thống micro khác nhau và với các thiết lập loa phát thanh tùy ý. Mục đích của hệ thống DirAC là tái tạo ấn tượng không gian của môi trường âm thanh hiện có chính xác nhất có thể sử dụng hệ thống đa kênh/loa phát thanh 3D. Trong môi trường được chọn, các phản hồi (âm thanh liên tục hoặc các phản hồi xung) được đo bằng micro đẳng hướng (W) và với bộ micro cho phép ước số hướng đến của âm thanh và sự khuếch tán của âm thanh. Phương pháp

phổ biến là áp dụng ba micrô hình số tám (X, Y, Z) được xếp hàng tương ứng theo trục tọa độ Đề-các [34]. Cách để thực hiện phương pháp này là sử dụng micrô trường âm thanh, trực tiếp mang lại các phản hồi mong muốn. Các tín hiệu W, X, Y và Z cũng có thể được tính toán từ bộ các micrô đẳng hướng rời rạc.

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

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

Hơn nữa, tham chiếu đến công nghệ DirAC như được mô tả, ví dụ, trong công bố "Mã hóa âm thanh định hướng - Tái tạo dựa trên cảm quan của âm thanh không gian" (Directional Audio Coding – Perception-Based Reproduction of Spatial Sound), V. Pulkki và cộng sự, Hội thảo Quốc tế về các nguyên tắc và ứng dụng của thính giác không gian, 11-13 tháng 11, 2009, Zao, Miyagi, Nhật Bản. Sự tham chiếu này mô tả việc mã hóa âm thanh định hướng làm ví dụ cho địa điểm tham chiếu liên quan đến việc xử lý trường âm thanh đặc biệt là kỹ thuật được thúc đẩy một cách cảm quan cho việc xử lý âm thanh không gian. Nó có các ứng dụng trong việc thu, mã hóa, và tổng

hợp lại âm thanh không gian, trong hội nghị viễn truyền, trong bộ lọc định hướng, và trong các môi trường thính giác ảo.

Sự tái tạo của cảnh âm thanh thường tập trung vào các thiết lập loa phát thanh, vì đây là sự tái tạo điển hình riêng ví dụ như phòng khách, và bối cảnh chuyên nghiệp ví dụ như các rạp chiếu phim. Ở đây, mối quan hệ của cảnh với dạng hình tái tạo là tĩnh vì nó đi kèm cùng với hình ảnh hai chiều mà buộc người nghe phải nhìn theo hướng phía trước. Sau đó, mối quan hệ không gian của các đối tượng âm thanh và hình ảnh được định rõ và cố định tại thời điểm sản xuất.

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

Các ứng dụng VR hiện đại cho phép sáu độ tự do (six-degrees-of-freedom - 6DoF). Ngoài sự xoay đầu, người dùng có thể di chuyển xung quanh dẫn đến tịnh tiến vị trí của anh/cô ấy theo ba chiều không gian. Sự tái tạo 6DoF bị giới hạn bởi kích thước tổng thể của vùng đi bộ. Trong nhiều trường hợp, vùng này khá nhỏ, ví dụ, phòng khách thông thường. 6DoF thường gặp trong các trò chơi VR. Ở đây, toàn bộ cảnh được tổng hợp với hình ảnh do máy tính tạo ra (computer-generated imagery - CGI). Âm thanh thường được tạo bằng cách sử dụng việc kết xuất dựa trên đối tượng trong đó mỗi đối tượng âm thanh được kết xuất với độ khuếch đại phụ thuộc vào khoảng cách và hướng tương đối từ người dùng dựa trên dữ liệu theo dõi. Hiện thực có thể được nâng cao bằng tiếng vang và nhiễu xạ nhân tạo [8, 9, 10].

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

Các phương pháp để tái tạo 6DoF của nội dung được ghi lại đã sử dụng các vị trí ghi lại được phân bố theo không gian được đề xuất. Đối với video, các dãy của các máy quay phim có thể được dùng để tạo kết xuất trường ánh sáng [13]. Đối với âm thanh, thiết lập tương tự dùng các dãy micro được phân bố hoặc các micro ambisonic. Nó đã được chứng minh rằng có thể tạo ra tín hiệu của 'micro ảo' được đặt tại vị trí tùy ý từ các bản ghi như vậy [14].

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

Phương pháp được đề xuất trong sáng chế này cho phép sự tái tạo 6DoF từ bản ghi của micrô FOA. Các bản ghi từ vị trí không gian đơn được sử dụng cho sự tái tạo 3DoF hoặc phóng đại âm thanh. Nhưng, theo sự hiểu biết của các tác giả sáng chế, cho đến nay, không có phương pháp nào tái tạo hoàn toàn 6DoF từ dữ liệu như vậy được đề xuất. Người ta nhận ra tái tạo 6DoF bằng cách tích hợp thông tin về khoảng cách của các nguồn âm thanh trong bản ghi. Thông tin khoảng cách này được hợp nhất vào phép biểu diễn tham số của DirAC, sao cho phối cảnh thay đổi của người nghe được ánh xạ chính xác.

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

Vấn đề điển hình khi di chuyển trong 6DoF được minh họa trong Fig.1b. Chúng ta hãy giả sử rằng cảnh âm thanh được mô tả tại Vị trí A sử dụng ambisonic. Trong trường hợp này, các âm thanh từ Nguồn A và Nguồn B đến từ cùng một hướng, tức là, chúng có cùng hướng đến (direction-of-arrival - DoA). Trong trường hợp một người di chuyển đến Vị trí B, DoA của Nguồn A và Nguồn B là khác nhau. Sử dụng mô tả ambisonic tiêu chuẩn của trường âm thanh, nghĩa là, không có thông tin bổ sung, không thể tính toán các tín hiệu ambisonic tại Vị trí B cho các tín hiệu ambisonic tại Vị trí A.

Tài liệu tham khảo

- [1] Liitola, T., *Headphone sound externalization*, Ph.D. thesis, Helsinki University of Technology. Department of Electrical and Communications Engineering Laboratory of Acoustics and Audio Signal Processing., 2006.
- [2] Blauert, J., *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, The MIT Press, 1996, ISBN 0262024136.

[3] Zhang, W., Samarasinghe, P. N., Chen, H., and Abhayapala, T. D., “Surround by Sound: A Re-view of Spatial Audio Recording and Reproduction,” *Applied Sciences*, 7(5), p. 532, 2017.

[4] Bates, E. and Boland, F., “Spatial Music, Virtual Reality, and 360 Media,” in *Audio Eng. Soc. Int. Conf. on Audio for Virtual and Augmented Reality*, Los Angeles, CA, U.S.A., 2016.

[5] Anderson, R., Gallup, D., Barron, J. T., Kontkanen, J., Snavely, N., Esteban, C. H., Agarwal, S., and Seitz, S. M., “Jump: Virtual Reality Video,” *ACM Transactions on Graphics*, 35(6), p. 198, 2016.

[6] Merimaa, J., *Analysis, Synthesis, and Perception of Spatial Sound: Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*, Ph.D. thesis, Helsinki University of Technology, 2006.

[7] Kronlachner, M. and Zotter, F., “Spatial Transformations for the Enhancement of Ambisonic Recordings,” in *2nd International Conference on Spatial Audio*, Erlangen, Germany, 2014.

[8] Tsingos, N., Gallo, E., and Drettakis, G., “Perceptual Audio Rendering of Complex Virtual Environments,” *ACM Transactions on Graphics*, 23(3), pp. 249–258, 2004.

[9] Taylor, M., Chandak, A., Mo, Q., Lauterbach, C., Schissler, C., and Manocha, D., “Guided multi-view ray tracing for fast auralization,” *IEEE Trans. Visualization & Comp. Graphics*, 18, pp. 1797–1810, 2012.

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

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

[12] Khaddour, H., Schimmel, J., and Rund, F., “A Novel Combined System

of Direction Estimation and Sound Zooming of Multiple Speakers,” *Radioengineering*, 24(2), 2015.

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

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

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

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

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

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

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

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

[21] Rummukainen, O., Schlecht, S., Plinge, A., and Habets, E. A. P., “Evaluating Binaural Reproduction Systems from Behavioral Patterns in a Virtual Reality – A Case Study with Impaired Binaural Cues and Tracking Latency,” in *Audio Eng. Soc. Conv. 143*, New York, NY, USA, 2017.

[22] Engelke, U., Darcy, D. P., Mulliken, G. H., Bosse, S., Martini, M. G.,

Arndt, S., Antons, J.-N., Chan, K. Y., Ramzan, N., and Brunnström, K., "Psychophysiology-Based QoE Assessment: A Survey," *IEEE Selected Topics in Signal Processing*, 11(1), pp. 6–21, 2017.

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

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

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

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

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

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

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

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

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

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

Bản chất kỹ thuật của sáng chế

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

Mục đích này đạt được bởi thiết bị để tạo mô tả trường âm thanh nâng cao, thiết bị để tạo mô tả trường âm thanh biến đổi, phương pháp để tạo mô tả trường âm thanh nâng cao, phương pháp để tạo mô tả trường âm thanh biến đổi, chương trình máy tính hoặc mô tả trường âm thanh nâng cao.

Sáng chế dựa trên việc phát hiện rằng các mô tả trường âm thanh điển hình mà liên quan đến địa điểm tham chiếu cần thông tin bổ sung để các mô tả trường âm thanh này có thể được xử lý để mô tả trường âm thanh biến đổi mà không liên quan đến địa điểm tham chiếu ban đầu nhưng liên quan đến địa điểm tham chiếu khác có thể được tính toán. Để đạt được mục đích này, siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh được tạo và siêu dữ liệu cùng với mô tả trường âm thanh tương ứng với mô tả trường âm thanh nâng cao mà có thể, ví dụ, được truyền hoặc được lưu trữ. Để tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu và, cụ thể, siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, trường âm thanh biến đổi được tính toán sử dụng thông tin âm thanh này, mô tả trường âm thanh và thông tin tịnh tiến biểu thị phép tịnh tiến từ địa điểm tham chiếu đến địa điểm tham chiếu khác. Do đó, mô tả trường âm thanh nâng cao bao gồm mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh này nằm dưới mô tả trường âm thanh được xử lý để thu được mô tả trường âm thanh biến đổi liên quan đến địa điểm tham chiếu khác được định rõ bởi thông tin tịnh tiến bổ sung mà có thể, ví dụ, được cung cấp hoặc được sử dụng tại phía bộ giải mã.

Tuy nhiên, sáng chế không chỉ liên quan đến kịch bản bộ mã hóa/bộ giải mã, mà còn có thể được áp dụng trong ứng dụng mà tại đó cả hai việc tạo mô tả trường âm thanh nâng cao và tạo mô tả trường âm thanh biến đổi diễn ra trên cơ bản một và cùng một địa điểm. Mô tả trường âm thanh biến đổi có thể, ví dụ, là mô tả của chính trường âm thanh biến đổi hoặc thực tế trường âm thanh biến đổi trong các tín hiệu kênh, các

tín hiệu lập thể hoặc, một lần nữa, trường âm thanh liên quan đến địa điểm tham chiếu mà, tuy nhiên, liên quan đến địa điểm tham chiếu mới hoặc khác thay vì địa điểm tham chiếu ban đầu. Chẳng hạn, ứng dụng như vậy sẽ ở trong kịch bản thực tế ảo ở đó mô tả trường âm thanh cùng với siêu dữ liệu tồn tại và ở đó người nghe di chuyển ra khỏi địa điểm tham chiếu mà trường âm thanh được đưa ra và di chuyển đến địa điểm tham chiếu khác và ở đó, sau đó, trường âm thanh cho người nghe di chuyển xung quanh trong khu vực ảo được tính toán tương ứng với trường âm thanh nhưng bây giờ tại vị trí tham chiếu ở đó người dùng đã di chuyển tới.

Theo phương án cụ thể, mô tả trường âm thanh nâng cao có mô tả trường âm thanh thứ nhất liên quan đến địa điểm tham chiếu (thứ nhất) và mô tả trường âm thanh thứ hai liên quan đến địa điểm tham chiếu (thứ hai) tiếp theo mà khác với địa điểm tham chiếu (thứ nhất), và siêu dữ liệu có thông tin trên địa điểm tham chiếu và địa điểm tham chiếu tiếp theo, chẳng hạn như các vectơ chỉ từ điểm bắt đầu được định trước tới các địa điểm tham chiếu này. Ngoài ra, siêu dữ liệu có thể là vectơ đơn chỉ đến hoặc địa điểm tham chiếu hoặc địa điểm tham chiếu tiếp theo và vectơ mở rộng giữa hai địa điểm tham chiếu, trong đó hai mô tả trường âm thanh khác nhau có liên quan.

Các mô tả trường âm thanh có thể là các mô tả trường âm thanh không tham số, chẳng hạn như mô tả ambisonic cấp một hoặc mô tả ambisonic cấp cao. Thay thế hoặc thêm vào đó, các mô tả trường âm thanh có thể là các mô tả DirAC hoặc các mô tả trường âm thanh tham số khác, hoặc một mô tả trường âm thanh có thể là, ví dụ, mô tả trường âm thanh tham số và mô tả trường âm thanh khác có thể là, ví dụ, mô tả trường âm thanh không tham số.

Do đó, mô tả trường âm thanh có thể tạo, cho mỗi mô tả trường âm thanh, mô tả DirAC của trường âm thanh có một hoặc nhiều tín hiệu trộn giảm và dữ liệu hướng riêng và dữ liệu khuếch tán tùy chọn cho các ngăn thời gian - tần số khác nhau. Trong ngữ cảnh này, bộ tạo siêu dữ liệu được tạo cấu hình để tạo siêu dữ liệu hình học cho cả hai mô tả trường âm thanh để có thể nhận dạng địa điểm tham chiếu và địa điểm tham chiếu bổ sung từ siêu dữ liệu. Sau đó, có thể trích xuất các nguồn riêng từ cả hai mô tả

trường âm thanh và để thực hiện việc xử lý bổ sung cho mục đích tạo mô tả trường âm thanh nâng cao hoặc biến đổi.

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

Những bản ghi ambisonic cấp một (FOA) có thể được xử lý và được tái tạo qua tai nghe chụp đầu. Chúng có thể được xoay để tính đến sự định hướng đầu của người nghe. Tuy nhiên, các hệ thống thực tế ảo (VR) cho phép người nghe di chuyển trong sáu độ tự do (6DoF), nghĩa là, ba vòng xoay cộng với ba độ tịnh tiến tự do. Ở đây, các góc và khoảng cách rõ ràng của các nguồn âm thanh phụ thuộc vào vị trí của người nghe. Kỹ thuật để tạo điều kiện cho 6DoF được mô tả. Cụ thể, bản ghi FOA được mô

tả sử dụng mô hình tham số, mà biến đổi dựa trên vị trí người nghe và thông tin về các khoảng cách tới các nguồn. Phương pháp được đánh giá bằng bài kiểm tra nghe, so sánh các kết xuất lập thể khác nhau của cảnh âm thanh tổng hợp trong đó người nghe có thể di chuyển tự do.

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

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

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

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

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

Theo một phương án, thiết bị bao gồm bộ phân tích để phân tích các thành phần trường âm thanh để suy ra, đối với các ngăn thời gian hoặc tần số khác nhau, thông tin hướng đến. Thiết bị ngoài ra có bộ biến đổi tịnh tiến để tính toán thông tin DoA biến đổi trên ngăn tần số hoặc thời gian sử dụng thông tin DoA và siêu dữ liệu, ở đó siêu dữ liệu liên quan đến bản đồ chiều sâu kết hợp khoảng cách tới nguồn có trong cả hai mô tả trường âm thanh thu được bởi ví dụ xử lý phép đặc tam giác (triangulation) sử dụng hai góc đối với hai địa điểm tham chiếu khác nhau và các khoảng cách/vị trí hoặc các địa điểm tham chiếu. Điều này có thể áp dụng cho biểu diễn toàn băng hoặc các ngăn tần số khác nhau của khung thời gian.

Hơn nữa, bộ tính toán trường âm thanh có bộ bù khoảng cách để tính toán trường âm thanh biến đổi sử dụng thông tin bù khoảng cách tùy thuộc vào khoảng cách được tính toán sử dụng siêu dữ liệu giống nhau cho từng ngăn tần số hoặc thời gian

của nguồn khác nhau cho mỗi hoặc một số ngăn thời gian/tần số, và từ khoảng cách mới liên kết với ngăn thời gian hoặc tần số, khoảng cách mới có liên quan đến thông tin DoA biến đổi.

Theo một phương án, bộ tính trường âm thanh tính véc tơ thứ nhất chỉ từ địa điểm tham chiếu tới nguồn âm thanh thu được bằng cách phân tích trường âm thanh. Hơn nữa, bộ tính trường âm thanh tính véc tơ thứ hai chỉ từ địa điểm tham chiếu khác đến nguồn âm thanh và sự tính toán này được thực hiện sử dụng véc tơ thứ nhất và thông tin tịnh tiến, ở đó thông tin tịnh tiến định rõ véc tơ tịnh tiến từ địa điểm tham chiếu đến địa điểm tham chiếu khác. Và, sau đó, khoảng cách từ địa điểm tham chiếu khác tới nguồn âm thanh được tính toán sử dụng véc tơ thứ hai.

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

Theo một phương án, bộ tính trường âm thanh được tạo cấu hình để xác định các tín hiệu nguồn từ mô tả trường âm thanh và các hướng của các tín hiệu nguồn liên quan đến địa điểm tham chiếu bởi phép phân tích âm thanh.

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

Theo một phương án, việc tổng hợp trường âm thanh được thực hiện bằng cách quét các tín hiệu nguồn âm thanh sang hướng được cung cấp bởi thông tin hướng mới liên quan đến thiết lập tái tạo, và việc chia tỷ lệ của các tín hiệu nguồn âm thanh được thực hiện sử dụng thông tin khoảng cách trước khi thực hiện thao tác quét.

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

Đặc biệt, nên ưu tiên thực hiện tổng hợp nguồn âm thanh trong phép biểu diễn phổ ở đó thông tin hướng mới được tính toán cho mỗi ngăn tần số, ở đó thông tin khoảng cách được tính toán cho mỗi ngăn tần số, và ở đó sự tổng hợp hướng cho mỗi ngăn tần số sử dụng tín hiệu âm thanh cho ngăn tần số được thực hiện sử dụng tín hiệu âm thanh cho ngăn tần số, sự khuếch đại quét cho ngăn tần số thu nhận từ thông tin hướng mới và hệ số tỷ lệ cho ngăn tần số thu nhận từ thông tin khoảng cách cho ngăn tần số được thực hiện.

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

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

Theo các phương án tiếp theo, và, đặc biệt, để tạo mô tả trường âm thanh biến đổi từ mô tả trường âm thanh và siêu dữ liệu liên quan đến thông tin không gian của mô tả trường âm thanh, bộ tính toán trường âm thanh tính toán trường âm thanh biến đổi sử dụng thông tin không gian trên mô tả trường âm thanh thứ nhất, sử dụng thông

tin không gian trên mô tả trường âm thanh thứ hai, và sử dụng thông tin tịnh tiến biểu thị phép tịnh tiến của địa điểm tham chiếu tới địa điểm tham chiếu khác. Cụ thể, siêu dữ liệu có thể, ví dụ, là vectơ hướng đến địa điểm tham chiếu của mô tả trường âm thanh và vectơ khác hướng từ cùng một nguồn tới địa điểm tham chiếu tiếp theo của mô tả trường âm thanh thứ hai.

Để giải quyết thông tin tịnh tiến, các đối tượng được tạo bằng cách áp dụng phân tách nguồn, hoặc định dạng chùm, hoặc, nói chung, bất kỳ phép phân tích nguồn âm thanh nào cho mô tả trường âm thanh thứ nhất và thứ hai. Sau đó, hướng của thông tin đến của tất cả các đối tượng bất kể các đối tượng này là đối tượng băng thông rộng hay các đối tượng cho các ngăn thời gian/tần số riêng được tính toán. Sau đó, các đối tượng được trích xuất từ các mô tả trường âm thanh khác nhau được khớp với nhau để tìm ra ít nhất một đối tượng khớp, nghĩa là, đối tượng xuất hiện cả trong mô tả trường âm thanh thứ nhất và thứ hai. Bước khớp này được thực hiện, ví dụ, bằng phương pháp hiệu chỉnh hoặc tính toán hiệu chỉnh sử dụng các tín hiệu đối tượng và/hoặc thông tin hướng đến hoặc thông tin khác.

Do đó, kết quả của quá trình là có tồn tại, đối với đối tượng khớp, thông tin DoA thứ nhất liên quan đến địa điểm tham chiếu và thông tin DoA thứ hai liên quan đến địa điểm tham chiếu tiếp theo. Do đó, các vị trí của các đối tượng khớp và, đặc biệt là, khoảng cách của đối tượng khớp tới địa điểm tham chiếu hoặc địa điểm tham chiếu tiếp theo được tính toán dựa trên phép đặc tam giác sử dụng thông tin trên địa điểm tham chiếu hoặc địa điểm tham chiếu có trong siêu dữ liệu được liên kết.

Thông tin này, và, đặc biệt là, thông tin vị trí đối với đối tượng khớp sau đó được sử dụng để biến đổi mỗi đối tượng khớp dựa trên vị trí ước tính và vị trí mong muốn, nghĩa là, sau khi tịnh tiến, bằng cách xử lý bù khoảng cách. Để tính toán thông tin DoA mới cho vị trí người nghe mới, thông tin DoA cũ từ cả hai địa điểm tham chiếu và thông tin tịnh tiến được sử dụng. Về cơ bản, việc xử lý này có thể được thực hiện cho cả hai mô tả trường âm thanh riêng, do mỗi đối tượng khớp xuất hiện trong cả hai mô tả trường âm thanh. Tuy nhiên, theo các phương án được ưu tiên, mô tả trường âm thanh có địa điểm tham chiếu gần nhất với vị trí người nghe mới tiếp theo phép tịnh tiến được sử dụng.

Do đó, DoA mới được sử dụng để tính toán mô tả trường âm thanh mới cho đối tượng khớp liên quan đến địa điểm tham chiếu khác, nghĩa là, địa điểm mà người dùng vừa di chuyển đến. Do đó, và cũng để hợp nhất các đối tượng không phù hợp, các mô tả trường âm thanh đối với các đối tượng đó cũng được tính toán nhưng sử dụng thông tin DoA cũ. Và, cuối cùng, trường âm thanh biến đổi được tạo bằng cách bổ sung tất cả mô tả trường âm thanh riêng với nhau.

Bất kỳ thay đổi nào với định hướng đều có thể được nhận ra bằng cách áp dụng vòng xoay duy nhất cho tín hiệu ambisonic ảo.

Do đó, siêu dữ liệu không được sử dụng để cung cấp trực tiếp khoảng cách của đối tượng tới địa điểm tham chiếu. Thay vào đó, siêu dữ liệu được cung cấp để nhận dạng địa điểm tham chiếu của mỗi hai hoặc nhiều hơn hai mô tả trường âm thanh và khoảng cách giữa địa điểm tham chiếu và đối tượng khớp nhất định được tính toán dựa trên, ví dụ, các bước xử lý phép đạc tam giác.

Mô tả vắn tắt các hình vẽ

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

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

Fig.1b là minh họa giải thích vấn đề ví dụ trong sáng chế;

Fig.2 là phương án thực hiện ưu tiên của thiết bị tạo mô tả trường âm thanh nâng cao;

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

Fig.3b minh họa khác của trường âm thanh nâng cao bao gồm dữ liệu âm thanh và siêu dữ liệu liên quan đến thông tin không gian chẳng hạn như thông tin hình học cho mỗi mô tả trường âm thanh;

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

Fig.4b minh họa phương án thực hiện tiếp theo của thiết bị để tạo mô tả trường âm thanh biến đổi;

Fig.4c minh họa kịch bản với vị trí/địa điểm tham chiếu A, vị trí/địa điểm tham chiếu B tiếp theo, và địa điểm tham chiếu khác do tình tiến;

Fig.5 minh họa bước tái tạo 6DoF âm thanh không gian theo cảm giác chung;

Fig.6a minh họa phương án được ưu tiên thực hiện bộ tính toán trường âm thanh;

Fig.6b minh họa phương án thực hiện được ưu tiên để tính toán DoA mới và khoảng cách mới của nguồn âm thanh đối với địa điểm tham chiếu mới/khác;

Fig.6c minh họa phương án được ưu tiên của bước tái tạo 6DoF bao gồm thiết bị để tạo mô tả trường âm thanh nâng cao, ví dụ, cho mỗi mô tả trường âm thanh riêng lẻ và thiết bị để tạo mô tả trường âm thanh biến đổi cho các nguồn khớp;

Fig.7 minh họa phương án được ưu tiên để lựa chọn một trong số các mô tả trường âm thanh thứ nhất và thứ hai để tính toán trường âm thanh biến đổi cho đối tượng bằng thông rộng hoặc hẹp;

Fig.8 minh họa thiết bị ví dụ để tạo mô tả trường âm thanh từ tín hiệu âm thanh như tín hiệu đơn và dữ liệu hướng đến;

Fig.9 minh họa phương án được ưu tiên tiếp theo cho bộ tính toán trường âm thanh;

Fig.10 minh họa phương án thực hiện được ưu tiên của thiết bị để tạo mô tả trường âm thanh nâng cao;

Fig.11 minh họa phương án thực hiện được ưu tiên tiếp theo của thiết bị để tạo mô tả trường âm thanh nâng cao;

Fig.12a minh họa phương án thực hiện phân tích DirAC sáng chế trước đó; và

Fig.12b minh họa phương án thực hiện tổng hợp DirAC sáng chế trước đó.

Mô tả chi tiết sáng chế

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

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

- các phép biểu diễn được mở rộng có thể được tịnh tiến thành các phép biểu diễn không mở rộng hiện có (ví dụ, để kết xuất), và
- cho phép sử dụng lại các phương án thực hiện phần mềm và phần cứng hiện có khi làm việc với phép biểu diễn được mở rộng.

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

Cảnh âm thanh được mô tả sử dụng hai hoặc nhiều hơn hai tín hiệu ambisonic, mỗi tín hiệu mô tả cảnh âm thanh tại vị trí khác nhau, hay nói cách khác là từ một góc nhìn khác. Nó được giả định rằng các vị trí tương đối được biết đến. Tín hiệu ambisonic biến đổi tại vị trí mong muốn trong cảnh âm thanh được tạo ra từ các tín hiệu ambisonic đầu vào. Cách tiếp cận dựa trên tín hiệu hoặc dựa trên tham số có thể được sử dụng để tạo tín hiệu ambisonic ảo tại vị trí mong muốn.

Khái niệm về phép biểu diễn ambisonic đa điểm có thể áp dụng cho cả ambisonic (kiểu DirAC) truyền thống và tham số.

Tín hiệu ambisonic ảo tại vị trí mong muốn (nghĩa là, sau phép tịnh tiến) được tính toán sử dụng các bước sau trong phương án tịnh tiến dựa trên tín hiệu:

1. Các đối tượng được tạo bằng cách áp dụng tách nguồn cho từng tín hiệu ambisonic truyền thống.

2. DOA của tất cả các đối tượng được tính toán cho từng tín hiệu ambisonic truyền thống.

3. Các đối tượng được trích xuất từ một tín hiệu ambisonic truyền thống được khớp với các đối tượng được trích xuất từ các tín hiệu ambisonic truyền thống khác. Bước khớp được thực hiện dựa trên các DOA tương ứng và/hoặc các tín hiệu (ví dụ, bằng phương pháp hiệu chỉnh/nhất quán).

4. Các vị trí của các đối tượng khớp được ước lượng dựa trên phép đặc tam giác.

5. Mỗi đối tượng khớp (đầu vào kênh đơn) biến đổi dựa trên vị trí được ước lượng và vị trí mong muốn (nghĩa là, sau phép tịnh tiến) sử dụng bộ lọc bù khoảng cách.

6. DOA tại vị trí mong muốn (nghĩa là, sau phép tịnh tiến) được tính toán cho mỗi đối tượng khớp. DOA này được biểu diễn bởi DOA' .

7. Tín hiệu đối tượng ambisonic được tính cho mỗi đối tượng khớp. Tín hiệu đối tượng ambisonic được tạo sao cho đối tượng khớp có hướng đến DOA' .

8. Tín hiệu đối tượng ambisonic được tính cho mỗi đối tượng không khớp. Tín hiệu đối tượng ambisonic được tạo sao cho đối tượng không khớp có hướng đến DOA.

9. Tín hiệu ambisonic ảo thu được bằng cách bổ sung tất cả tín hiệu đối tượng ambisonic với nhau.

Tín hiệu ambisonic ảo tại vị trí mong muốn (nghĩa là, sau phép tịnh tiến) được tính toán sử dụng các bước sau trong phương án tịnh tiến dựa trên tham số phù hợp với phương án tiếp theo:

1. Mô hình trường âm thanh được giả định. Trường âm thanh có thể được khai triển thành một hoặc nhiều thành phần âm thanh có hướng và các thành phần âm thanh khuếch tán. Các thành phần âm thanh có hướng bao gồm thông tin tín hiệu và vị trí (ví dụ, ở tọa độ cực hoặc Đề-các). Ngoài ra, trường âm thanh có thể được khai triển thành

một hoặc nhiều thành phần âm thanh có hướng/cơ bản và thành phần âm thanh còn lại (đa kênh hoặc đơn kênh).

2. Các thành phần tín hiệu và các tham số của mô hình trường âm thanh giả định được ước lượng sử dụng các tín hiệu ambisonic đầu vào.

3. Các thành phần tín hiệu và/hoặc các tham số biến đổi phụ thuộc vào sự tịnh tiến mong muốn, hoặc vị trí mong muốn, trong cảnh âm thanh.

4. Tín hiệu ambisonic ảo được tạo sử dụng các thành phần tín hiệu biến đổi và các tham số biến đổi.

Việc tạo các tín hiệu ambisonic đa điểm rất đơn giản cho nội dung máy tính tạo và sản xuất cũng như trong bối cảnh ghi âm tự nhiên thông qua dây micro hoặc micro không gian (ví dụ, micro định dạng B). Theo một phương án, ưu tiên thực hiện khớp nguồn tiếp theo bước 2 hoặc tính toán phép đặc tam giác trước bước 3. Hơn nữa, một hoặc nhiều bước của cả hai phương án cũng có thể được sử dụng theo các phương án khác tương ứng.

Sự thay đổi trong định hướng có thể được nhận ra bằng cách áp dụng vòng xoay đơn cho tín hiệu ambisonic ảo.

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

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

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

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

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

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

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

Fig.2 minh họa thiết bị để tạo mô tả trường âm thanh nâng cao, trong đó bộ tạo trường âm thanh bao gồm bộ tạo trường âm thanh 250 cho trường âm thanh thứ nhất, bộ tạo trường âm thanh 260 cho trường âm thanh thứ hai và, số lượng bộ tạo trường âm thanh tùy ý cho một hoặc nhiều trường âm thanh chẳng hạn như trường âm thanh thứ ba, thứ tư, v.v. Ngoài ra, siêu dữ liệu được tạo cấu hình để tính toán và chuyển tiếp đến bộ kết hợp 120 thông tin trên trường âm thanh thứ nhất và trường âm thanh thứ

hai. Tất cả các thông tin này được sử dụng bởi bộ kết hợp 120 để tạo mô tả trường âm thanh nâng cao. Do đó, bộ kết hợp 120 cũng được tạo cấu hình thành giao diện đầu ra để tạo mô tả trường âm thanh nâng cao.

Fig.3a minh họa mô tả trường âm thanh nâng cao như là dòng dữ liệu bao gồm mô tả trường âm thanh thứ nhất 330, mô tả trường âm thanh thứ hai 340 và, kèm theo đó, siêu dữ liệu 350 bao gồm thông tin về mô tả trường âm thanh thứ nhất và mô tả trường âm thanh thứ hai. Mô tả trường âm thanh thứ nhất có thể, ví dụ, là mô tả định dạng B hoặc mô tả cấp cao hoặc bất kỳ mô tả nào khác mà cho phép xác định phân phối hướng của các nguồn âm thanh trong phép biểu diễn toàn dải hoặc trong phép biểu diễn tần số được chọn. Do đó, mô tả trường âm thanh thứ nhất 330 và mô tả trường âm thanh thứ hai 340 cũng có thể, ví dụ, là các mô tả trường âm thanh tham số cho các vị trí tham chiếu khác nhau có, ví dụ, tín hiệu trộn giảm và hướng đến của dữ liệu cho các ngăn thời gian/tần số khác nhau.

Tuy nhiên, thông tin hình học 350 cho các mô tả trường âm thanh thứ nhất và thứ hai là giống nhau cho tất cả các nguồn có trong mô tả trường âm thanh thứ nhất 330 hoặc, cho các nguồn trong mô tả trường âm thanh thứ hai 340, tương ứng. Do đó, ví dụ, khi tồn tại ba nguồn trong mô tả trường âm thanh thứ nhất 330 và thông tin hình học trên mô tả trường âm thanh thứ nhất thì thông tin hình học này giống nhau cho ba nguồn trong mô tả trường âm thanh thứ nhất. Tương tự, ví dụ, khi tồn tại năm nguồn trong mô tả trường âm thanh thứ hai, thì thông tin hình học cho trường âm thanh thứ hai có trong siêu dữ liệu 350 là giống nhau cho tất cả các nguồn trong mô tả trường âm thanh thứ hai.

Fig.3b minh họa cấu trúc kiểu mẫu của siêu dữ liệu 350 trên Fig.3a. Theo một phương án, địa điểm tham chiếu 351 có thể có trong siêu dữ liệu. Tuy nhiên, điều này không nhất thiết là trường hợp trong thông tin địa điểm tham chiếu 351 cũng có thể được bỏ qua.

Đối với trường âm thanh thứ nhất, thông tin hình học thứ nhất được đưa ra, ví dụ, mà có thể là thông tin về vectơ A được minh họa trên Fig.4c chỉ từ gốc đến vị trí/địa điểm tham chiếu A, mà liên quan đến trường âm thanh thứ nhất.

Thông tin hình học thứ hai có thể, ví dụ, là thông tin về vectơ B chỉ từ gốc đến vị trí/địa điểm tham chiếu thứ hai B, mà liên quan đến mô tả trường âm thanh thứ hai.

A và B là các địa điểm tham chiếu hoặc các vị trí ghi cho cả hai mô tả trường âm thanh.

Ví dụ, thông tin hình học thay thế có thể là thông tin về vectơ D kéo dài giữa địa điểm tham chiếu A và địa điểm tham chiếu B tiếp theo và/hoặc gốc và vectơ chỉ từ gốc đến một trong hai điểm. Do đó, thông tin hình học có trong siêu dữ liệu có thể bao gồm vectơ A và vectơ D hoặc có thể bao gồm vectơ B và vectơ D hoặc có thể bao gồm vectơ A và vectơ B mà không có vectơ D hoặc có thể bao gồm thông tin khác, mà từ đó địa điểm tham chiếu A và địa điểm tham chiếu B có thể được nhận dạng trong hệ tọa độ ba chiều nhất định. Tuy nhiên, sự cân nhắc tương tự cũng được áp dụng bổ sung cho mô tả âm thanh hai chiều cũng như được minh họa tỉ mỉ trên Fig.4c mà chỉ thể hiện trường hợp hai chiều.

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

Theo một phương án, bộ tính toán trường âm thanh 420 được kết nối với giao diện đầu vào 400 để nhận mô tả trường âm thanh nâng cao như, ví dụ, được thảo luận liên quan đến Fig.1a hoặc Fig.2 và giao diện đầu vào 400 sau đó phân tách mô tả trường âm thanh một mặt, nghĩa là, những gì đã được tạo ra bởi khối 100 trên Fig.1a hoặc khối 210 trên Fig.2. Hơn nữa, giao diện đầu ra 400 phân tách siêu dữ liệu từ mô tả trường âm thanh nâng cao, nghĩa là, mục 350 trên Fig.3a hoặc tùy chọn 351 và 352 đến 354 trên Fig.3b.

Hơn nữa, giao diện tịnh tiến 410 thu thông tin tịnh tiến và/hoặc thông tin xoay bổ sung hoặc riêng biệt từ người nghe. Việc thực hiện giao diện tịnh tiến 410 có thể là bộ phận theo dõi đầu không chỉ theo dõi sự xoay của đầu trong môi trường thực tế ảo,

mà còn là sự tịnh tiến của đầu từ một vị trí, nghĩa là, vị trí A trên Fig.1b tới vị trí khác, nghĩa là, vị trí B trên Fig.1b.

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

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

Ví dụ, Fig.6 được mô tả sau đó minh họa tình huống, trong đó bộ tổng hợp DirAC chỉ hoạt động trên thành phần trộn giảm chẳng hạn như thành phần đẳng hướng hoặc áp lực, trong khi, trong phương án thay thế khác được minh họa đối với Fig.12b,

bộ tổng hợp DirAC hoạt động trên toàn bộ dữ liệu trường âm thanh, nghĩa là, phép biểu diễn thành phần đầy đủ có trong phương án này trên Fig.12b, mô tả trường âm thanh với thành phần đẳng hướng w và ba thành phần định hướng x, y, z .

Fig.4c minh họa kịch bản dựa trên các phương án được ưu tiên của sáng chế. Hình vẽ minh họa vị trí/địa điểm tham chiếu A, vị trí/địa điểm tham chiếu thứ hai B và hai nguồn âm thanh khác nhau A và B, và vectơ tịnh tiến I .

Cả hai nguồn âm thanh A và B có trong mô tả trường âm thanh liên quan đến địa điểm tham chiếu A và mô tả trường âm thanh thứ hai liên quan đến vị trí tham chiếu B.

Để tính toán khoảng cách của nguồn A, ví dụ, tới vị trí tham chiếu thứ nhất hoặc vị trí tham chiếu thứ hai, cả hai mô tả trường âm thanh khác nhau liên quan đến A và B đều phải tuân theo quy trình tách nguồn và sau đó, kết hợp các nguồn thu được bởi các quy trình tách âm thanh khác nhau này được thu. Điều này sẽ dẫn đến nguồn A, ví dụ. Nguồn A được tìm thấy trong thuật toán tách nguồn cho mô tả trường âm thanh thứ nhất và cũng cho mô tả trường âm thanh thứ hai. Thông tin hướng đến cho nguồn A sẽ là, khi được nhận từ mô tả trường âm thanh thứ nhất liên quan đến vị trí tham chiếu A góc α . Ngoài ra, thông tin hướng đến cho cùng nguồn A nhưng hiện tại được thu từ mô tả trường âm thanh thứ hai liên quan đến vị trí tham chiếu B tiếp theo sẽ là góc β .

Hiện tại, dựa trên khoảng cách D có thể tính toán hoặc đã biết, ví dụ, có thể lấy được hoặc có thể tính toán từ siêu dữ liệu cho các mô tả trường âm thanh, và dựa trên hai góc α và β , tam giác được định rõ bởi nguồn A, vị trí tham chiếu A và vị trí tham chiếu B được định rõ đầy đủ. Do đó, khoảng cách từ nguồn A tới vị trí tham chiếu A hoặc khoảng cách từ nguồn A tới vị trí tham chiếu B hoặc vị trí chung của nguồn A, nghĩa là, vectơ chỉ từ gốc tới vị trí thực của nguồn A có thể được tính toán, ví dụ bởi các hoạt động xử lý phép đặc tam giác. Vị trí hoặc khoảng cách đều thể hiện thông tin vị trí trên khoảng cách hoặc trên vị trí.

Quy trình tương tự sau đó có thể được thực hiện cho mỗi nguồn phù hợp, tức là, cũng đối với nguồn B.

Do đó, thông tin khoảng cách/vị trí cho mỗi nguồn phù hợp được tính toán và, sau đó, mỗi nguồn phù hợp có thể được xử lý như thể khoảng cách/vị trí được biết đầy đủ hoặc, ví dụ, được đưa ra bởi siêu dữ liệu bổ sung. Tuy nhiên, chỉ thông tin hình học cho mô tả trường âm thanh thứ nhất và trường âm thanh thứ hai là cần thiết thay vì bất kỳ thông tin khoảng cách/chiều sâu cho từng nguồn riêng.

Fig.8 minh họa phương án thực hiện khác để thực hiện tổng hợp khác với bộ tổng hợp DirAC. Ví dụ, khi bộ phân tích trường âm thanh tạo, cho mỗi tín hiệu nguồn, tín hiệu đơn riêng biệt S và hướng đến gốc và khi, phụ thuộc vào thông tin tịnh tiến, hướng đến mới được tính toán, sau đó bộ tạo tín hiệu ambisonic 430 trên Fig.8, ví dụ, sẽ được sử dụng để tạo mô tả trường âm thanh cho tín hiệu nguồn âm thanh, nghĩa là, tín hiệu đơn S nhưng cho dữ liệu hướng đến mới (DoA) bao gồm góc ngang θ hoặc góc độ cao θ và góc phương vị ϕ . Sau đó, quy trình được thực hiện bởi bộ tính toán trường âm thanh 420 trên Fig.4b sẽ được tạo, ví dụ, mô tả trường âm thanh ambisonic cấp một cho mỗi nguồn âm thanh với hướng đến mới và, sau đó, biến đổi tiếp theo trên mỗi nguồn âm thanh có thể được thực hiện sử dụng hệ số tỷ lệ tùy thuộc vào khoảng cách của nguồn âm thanh tới địa điểm tham chiếu mới và, sau đó, tất cả trường âm thanh từ các nguồn riêng biệt có thể chồng lên nhau để cuối cùng thu được trường âm thanh biến đổi, một lần nữa, ví dụ, trong phép biểu diễn ambisonic liên quan đến địa điểm tham chiếu mới nhất định.

Khi một người giải thích rằng mỗi ngăn thời gian/tần số được xử lý bởi bộ phân tích DirAC 422, 422a, 422b trên Fig.6 biểu diễn nguồn âm thanh nhất định (giới hạn băng thông), thì có thể sử dụng bộ tạo tín hiệu ambisonic 430, thay vì bộ tổng hợp DirAC 425, 425a, 425b để tạo, đối với mỗi ngăn thời gian/tần số, phép biểu diễn ambisonic đầy đủ sử dụng tín hiệu trộn giảm hoặc tín hiệu áp lực hoặc thành phần đẳng hướng cho ngăn thời gian/tần số này như là "tín hiệu đơn S " trên Fig.8. Sau đó, sự biến đổi tần số-thời gian riêng biệt trong bộ biến đổi tần số-thời gian cho mỗi thành phần trong số các thành phần W, X, Y, Z sau đó sẽ dẫn đến mô tả trường âm thanh khác với những gì được minh họa trên Fig.4c.

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

nghe, người được theo dõi trong 6DoF, xem Fig.5. Nguồn âm thanh duy nhất được hiển thị ở đây để minh họa, mối quan hệ giữ cho mỗi ngăn tần số-thời gian.

Fig.5 minh họa bước tái tạo 6DoF của âm thanh không gian. Nguồn âm thanh được ghi lại bởi micrô với DoA $\mathbf{r}_r$ trong khoảng cách d_r tương đối với vị trí và định hướng của các micrô (đường màu đen và đường cung). Nó phải được tái tạo tương ứng với người nghe di chuyển với DoA $\mathbf{r}_l$ và khoảng cách d_l (nét đứt). Điều này phải xem xét phép tịnh tiến $\mathbf{I}$ và phép xoay $\mathbf{o}$ của người nghe (chấm). DOA được biểu diễn là véctơ với độ dài đơn vị hướng về nguồn.

Nguồn âm thanh tại tọa độ $\mathbf{d}_r \in \mathbb{R}^3$ được ghi từ hướng đến (DoA) được biểu thị bằng véctơ đơn vị $\mathbf{r}_r = \mathbf{d}_r / \|\mathbf{d}_r\|$. DoA này có thể được ước tính từ phân tích bản ghi. Nó đến từ khoảng cách $d_r = \|\mathbf{d}_r\|$. Giả định rằng thông tin này có thể được suy ra từ siêu dữ liệu cho mỗi nguồn hoặc, nói chung, từ mục 352, 353, 354 trên Fig.3b, và có thể được biểu diễn dưới dạng mô tả khoảng cách có bất kỳ hướng $\mathbf{r}$ nào từ vị trí ghi đến khoảng cách (ví dụ được tính bằng mét hoặc hơn) như được thu ví dụ bằng cách xử lý phép đặc tam giác sử dụng hai góc đối với hai địa điểm tham chiếu khác nhau và các vị trí/khoảng cách hoặc địa điểm tham chiếu.

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

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

Phương pháp để thu tín hiệu ảo từ phối cảnh người nghe bởi các phép biến đổi chuyên dụng dựa trên phép biểu diễn tham số, như được giải thích trong phần dưới đây, được phác thảo. Phương pháp được đề xuất dựa trên tiếp cận DirAC cơ bản để mã hóa âm thanh không gian tham số, xem [16]. Giả định rằng có một nguồn có hướng chỉ phối trên mỗi trường hợp thể hiện thời gian – tần số của phổ được phân tích và chúng có thể được xử lý độc lập. Bản ghi được biến đổi thành phép biểu diễn thời gian – tần số sử dụng biến đổi Fourier thời gian ngắn (short time Fourier transform - STFT). Chỉ số khung thời gian được ký hiệu là n và chỉ số tần số được ký hiệu là k . bản ghi được

biến đổi sau đó được phân tích, ước tính các hướng $r_r(k, n)$ và độ khuếch tán $\psi(k, n)$ cho mỗi ngăn thời gian – tần số của phổ phức tạp $P(k, n)$. Trong quá trình tổng hợp, tín hiệu được chia thành phần có hướng và phần khuếch tán. Ở đây, tín hiệu loa phát thanh được tính toán bằng cách quét phần có hướng tùy thuộc vào các vị trí loa phát thanh và bổ sung phần khuếch tán.

Phương pháp để biến đổi bản ghi tín hiệu FOA đến phối cảnh người nghe trong 6DoF có thể được chia thành năm bước, xem Fig.6c.

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

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

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

gradien được biến đổi.

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

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

Độ khuếch tán được ước tính từ hệ số biến thiên của vectơ này [18]:

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

trong đó E biểu thị toán tử kỳ vọng dọc theo các khung thời gian, được thực hiện dưới dạng số trung bình động.

Do dự định điều khiển âm thanh sử dụng bản đồ khoảng cách dựa trên hướng có khoảng cách cho mỗi nguồn (khớp) đến địa điểm tham chiếu, phương sai của ước tính hướng sẽ thấp trong phương án tùy chọn. Vì các khung thường ngắn, điều này không phải lúc nào cũng đúng. Do đó, số trung bình động được áp dụng để thu được ước tính hướng làm nhẵn $\bar{I}_a(k, n)$. DoA của phần trực tiếp của tín hiệu khi đó, trong phương án được tính toán là vectơ độ dài đơn vị theo hướng ngược lại:

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

Vì hướng được mã hóa dưới dạng vectơ ba chiều có độ dài đơn vị cho mỗi ngăn thời gian - tần số, nên việc tích hợp thông tin khoảng cách là đơn giản. Các vectơ chỉ hướng được nhân với mục nhập bản đồ tương ứng của chúng sao cho độ dài vectơ biểu diễn khoảng cách của nguồn âm thanh tương ứng $d_r(k, n)$:

$$d_r(k, n) = r_r(k, n)d_r(k, n) = r_r(k, n)m(r_r(k, n)) \quad (4)$$

trong đó $\mathbf{d}_r(k, n)$ là vectơ chỉ từ vị trí ghi của micrô đến nguồn âm thanh hoạt động tại thời điểm n và tần số k .

Vị trí người nghe được đưa ra bởi hệ thống theo dõi cho khung xử lý hiện tại là $l(n)$. Với phép biểu diễn vectơ của các vị trí nguồn, có thể trừ vectơ vị trí theo dõi $l(n)$ để tạo ra vectơ hướng được tịnh tiến mới $\mathbf{d}_1(k, n)$ với độ dài $d_1(k, n) = \|\mathbf{d}_1(k, n)\|$, xem Fig.6b. Khoảng cách từ PoV của người nghe đến các nguồn âm thanh được suy ra, và các DoA được điều chỉnh theo một bước duy nhất:

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

Khía cạnh quan trọng của bước tái tạo hiện thực là sự giảm dần khoảng cách. Sự giảm dần được coi là hàm của khoảng cách giữa nguồn âm thanh và người nghe. Độ dài của các vectơ chỉ hướng là để mã hóa sự giảm dần hoặc khuếch đại để tái tạo. Khoảng cách đến vị trí ghi được mã hóa theo $\mathbf{d}_r(k, n)$ theo bản đồ khoảng cách, và khoảng cách được tái tạo được mã hóa theo $d_1(k, n)$. Nếu bình thường hóa các vectơ thành chiều dài đơn vị và sau đó nhân với tỷ lệ khoảng cách mới và cũ, sẽ thấy rằng độ dài cần thiết được đưa ra bằng cách chia $d_1(k, n)$ cho chiều dài của vectơ ban đầu:

$$\mathbf{d}_v(k, n) = \frac{\mathbf{d}_1(k, n)}{\|\mathbf{d}_1(k, n)\|} \frac{\|\mathbf{d}_1(k, n)\|}{\|\mathbf{d}_r(k, n)\|} = \frac{\mathbf{d}_1(k, n)}{\|\mathbf{d}_r(k, n)\|} \quad (6)$$

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

$$\mathbf{d}_p(k, n) = \mathbf{R}_y(o_y(n))\mathbf{R}_z(o_z(n))\mathbf{R}_x(o_x(n))\mathbf{d}_v(k, n) \quad (7)$$

DoA kết quả cho người nghe sau đó được đưa ra bởi vectơ chuẩn hóa theo chiều dài đơn vị:

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

Véc tơ hướng được tịnh tiến, độ khuếch tán và phổ phức tạp được sử dụng để tổng hợp các tín hiệu cho thiết lập loa phát thanh ảo 8+4 được phân bố đồng đều. Tám ống nghe ảo được đặt trong các bước góc 45° trên mặt phẳng người nghe (độ cao 0°), và bốn loa phát thanh trong hình thành chéo 90° ở trên độ cao 45° . Quá trình tổng hợp được chia thành phần trực tiếp và phần khuếch tán cho mỗi kênh loa phát thanh $1 \leq i \leq I$, trong đó $I = 12$ là số lượng loa phát thanh [16]:

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

Đối với phần có hướng, quét biên độ cạnh biên mờ dần (edge fading amplitude panning - EFAP) được áp dụng để tái tạo âm thanh từ hướng đúng cho dạng hình loa phát thanh ảo [20]. Với véc tơ DoA $\mathbf{r}_p(k, n)$, điều này cung cấp độ khuếch đại quét $G_i(\mathbf{r})$ cho mỗi kênh loa phát thanh ảo i . Độ khuếch đại phụ thuộc vào khoảng cách cho mỗi DoA được suy ra từ chiều dài kết quả của véc tơ chỉ hướng, $\mathbf{d}_p(k, n)$. Tổng hợp hướng cho kênh i trở thành:

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

trong đó hàm mũ γ là thừa số điều chỉnh thường được đặt thành khoảng 1 [19]. Lưu ý rằng với $\gamma = 0$ độ khuếch đại phụ thuộc vào khoảng cách bị tắt.

Áp lực $P(k, n)$ được sử dụng để tạo I các tín hiệu giải tương quan $\tilde{P}_i(k, n)$. Các tín hiệu giải tương quan này được bổ sung vào các kênh loa phát thanh riêng lẻ dưới dạng thành phần khuếch tán. Điều này tuân theo phương pháp tiêu chuẩn [16]:

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

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

Fig.6a minh họa phương án được ưu tiên tiếp theo để tính toán trường âm thanh

biến đổi sử dụng thông tin không gian, và mô tả trường âm thanh thứ nhất và thứ hai và thông tin tịnh tiến biểu thị sự tịnh tiến của địa điểm tham chiếu tới địa điểm tham chiếu khác, ví dụ, như đã được thảo luận đối với vectơ I trên Fig.4c hoặc Fig.5.

Fig.6a minh họa khối 700 biểu thị ứng dụng phân tách âm thanh hoặc, nói chung, quy trình phân tích âm thanh cho mô tả trường âm thanh thứ nhất liên quan đến vị trí tham chiếu A trên Fig.4c và mô tả trường âm thanh thứ hai liên quan đến vị trí tham chiếu B trên Fig.4c.

Quy trình này sẽ dẫn đến nhóm thứ nhất của một hoặc nhiều đối tượng được trích xuất và, ngoài ra, nhóm thứ hai của một hoặc nhiều đối tượng được trích xuất.

Các nhóm này được sử dụng trong khối 702 để tính toán thông tin hướng đến cho tất cả các nguồn được phân tách, nghĩa là, cho nhóm thứ nhất của các nguồn được trích xuất và nhóm thứ hai của một hoặc nhiều nguồn được trích xuất.

Theo các phương án khác, các bước 700 và 702 được thực hiện trong một quy trình duy nhất cung cấp, một mặt, tín hiệu cho nguồn và mặt khác thông tin DoA cho nguồn. Điều này cũng đúng với các quy trình tham số chẳng hạn như các quy trình chọn thời gian/tần số như DirAC, trong đó tín hiệu nguồn là tín hiệu của phép biểu diễn định dạng B trong ngăn thời gian/tần số hoặc tín hiệu áp lực hoặc tín hiệu đẳng hướng của ngăn thời gian/tần số và thông tin DoA như tham số DoA cho ngăn cụ thể này.

Sau đó, trong bước 704, bước khớp nguồn được thực hiện giữa các nguồn của nhóm thứ nhất và các nguồn của nhóm thứ hai và kết quả của bước khớp nguồn là các nguồn khớp.

Các nguồn khớp này được sử dụng để tính toán trường âm thanh cho mỗi đối tượng khớp sử dụng DoA mới và khoảng cách mới như được minh họa trong khối 710. Hơn nữa, hướng thông tin đến của các đối tượng khớp, nghĩa là hai trên mỗi đối tượng chẳng hạn như α và β trên Fig.4c cho nguồn A được sử dụng trong khối 706 để tính toán các vị trí của các đối tượng khớp hoặc thay thế hoặc bổ sung, các khoảng cách của các đối tượng khớp sử dụng, ví dụ, các hoạt động phép đặc tam giác.

Kết quả của khối 706 là vị trí của mỗi đối tượng khớp hoặc thay thế hoặc bổ

sung, khoảng cách của đối tượng khớp đến một trong số các địa điểm tham chiếu thứ nhất hoặc thứ hai A, B, được minh họa, ví dụ như trên Fig.4c.

Ngoài ra, không chỉ ưu tiên sử dụng thông tin tịnh tiến trong khối 708 mà còn sử dụng thông tin xoay để tính toán thông tin hướng đến mới và các khoảng cách mới cho các đối tượng khớp.

Mặc dù đã được phác thảo rằng các vị trí của các đối tượng khớp được đưa vào bước 708, nhưng cần nhấn mạnh rằng, chỉ tính toán thông tin hướng đến mới cho đối tượng khớp, vị trí thực tế của đối tượng khớp hoặc, nói cách khác, khoảng cách của đối tượng khớp là không cần thiết để tính toán hướng đến mới đối với địa điểm tham chiếu (khác) mới, tới địa điểm mà người nghe đã di chuyển đến, ví dụ, khoảng cách là không cần thiết.

Tuy nhiên, khoảng cách sau đó là cần thiết để thích ứng tín hiệu nguồn với vị trí mới. Do đó, khi khoảng cách của nguồn hoặc đối tượng âm thanh tới vị trí tham chiếu mới đã trở nên ngắn hơn, thì hệ số tỷ lệ sẽ được tính toán thấp hơn một. Tuy nhiên, khi khoảng cách trở nên cao hơn, thì hệ số tỷ lệ được tính là cao hơn một, ví dụ, như được thảo luận đối với Fig.6b. Do đó, mặc dù được minh họa trên Fig.6a cho phương án, không nhất thiết phải xác định các vị trí rõ ràng của các đối tượng khớp và, sau đó, các khoảng cách của các đối tượng khớp được tính toán và, sau đó, trường âm thanh được tính toán cho mỗi đối tượng khớp sử dụng hướng đến mới và khoảng cách mới. Thay vào đó, chỉ khoảng cách của đối tượng khớp tới một địa điểm tham chiếu trong số hai địa điểm tham chiếu nói chung là đủ và sau đó, trường âm thanh cho mỗi đối tượng khớp được tính toán sử dụng DoA mới và khoảng cách mới.

Ngoài ra, khối 714 minh họa phép tính toán của các trường âm thanh cho các đối tượng không khớp sử dụng thông tin DoA thu được bởi khối 702. Hơn nữa, các trường âm thanh cho các đối tượng thu được trong khối 710 và các đối tượng không khớp thu được bởi khối 714 được kết hợp trong khối 712 để thu mô tả trường âm thanh biến đổi mà có thể, ví dụ, là mô tả ambisonic chẳng hạn như mô tả ambisonic cấp một, mô tả ambisonic cấp cao hoặc, ngoài ra, mô tả kênh loa phát thanh liên quan đến thiết lập loa phát thanh nhất định, tất nhiên, giống với khối 710 và khối 714 để có

thể thực hiện bổ sung từng kênh đơn giản trong khối 712.

Fig.6b minh họa phương án thực hiện được ưu tiên của bộ tính toán trường âm thanh 420. Trong khối 1102, việc phân tách nguồn và tính toán thông tin hướng đến hoặc hướng nói chung cho mỗi nguồn được thực hiện. Sau đó, trong khối 1104, vectơ hướng đến được nhân với vectơ thông tin khoảng cách, nghĩa là, vectơ từ địa điểm tham chiếu ban đầu đến nguồn âm thanh, nghĩa là, vectơ từ mục 520 đến mục 510 trên Fig.5, ví dụ. Sau đó, trong khối 1106, thông tin tịnh tiến, nghĩa là, vectơ từ mục 520 đến mục 500 trên Fig.5 được tính đến để tính toán vectơ hướng tịnh tiến mới mà là vectơ từ vị trí người nghe 500 đến vị trí nguồn âm thanh 510. Sau đó, vectơ hướng đến mới với độ dài chính xác được biểu thị bởi $\mathbf{d}_v$ được tính toán trong khối 1108. Vectơ này được định hướng cùng hướng với $\mathbf{d}_r$, nhưng có độ dài khác nhau, vì độ dài của vectơ này phản ánh thực tế rằng nguồn âm thanh 510 được ghi trong trường âm thanh ban đầu với âm lượng nhất định và do đó, độ dài của $\mathbf{d}_v$ nhiều hơn hoặc ít hơn biểu thị sự thay đổi cường độ âm. Điều này có được bằng cách chia vectơ $\mathbf{d}_i$ cho khoảng cách ghi $\mathbf{d}_r$, tức là, độ dài của vectơ $\mathbf{d}_r$ từ micrô 520 đến nguồn âm thanh 510. Như đã nêu, độ dài của vectơ $\mathbf{d}_r$ từ micrô 520 đến nguồn âm thanh 510 có thể được suy ra bằng tính toán phép đặc tam giác. Khi micrô ở địa điểm tham chiếu của mô tả trường âm thanh thứ nhất, thì khoảng cách từ địa điểm tham chiếu của mô tả trường âm thanh thứ nhất tới nguồn âm thanh được sử dụng. Tuy nhiên, khi micrô ở địa điểm tham chiếu tiếp theo của mô tả trường âm thanh thứ hai, thì khoảng cách từ địa điểm tham chiếu tiếp theo của mô tả trường âm thanh thứ hai tới nguồn âm thanh được sử dụng.

Như trên Fig.5, khi khoảng cách được tái tạo lớn hơn khoảng cách được ghi, thì độ dài của $\mathbf{d}_v$ sẽ thấp hơn đơn vị. Điều này sẽ dẫn đến sự giảm dần của nguồn âm thanh 510 cho việc tái tạo tại vị trí người nghe mới. Tuy nhiên, khi khoảng cách được tái tạo $\mathbf{d}_i$ nhỏ hơn khoảng cách được ghi, thì độ dài của $\mathbf{d}_v$ như được tính toán bởi khối 1108 sẽ lớn hơn 1 và hệ số tỷ lệ tương ứng sẽ dẫn đến khuếch đại nguồn âm thanh.

Trên Fig.6a, mục 710 biểu thị rằng trường âm thanh cho mỗi đối tượng khớp được tính toán sử dụng thông tin hướng đến mới và khoảng cách mới. Tuy nhiên, về cơ bản, để tính toán trường âm thanh cho mỗi đối tượng khớp, các tín hiệu đối tượng thu được hoặc từ nhóm thứ nhất của một hoặc nhiều nguồn được trích xuất hoặc từ

nhóm thứ hai của một hoặc nhiều nguồn được trích xuất có thể được sử dụng nói chung. Tuy nhiên, theo một phương án, lựa chọn cụ thể được minh họa trên Fig.7 được thực hiện để xác định nguồn âm thanh nào được sử dụng để thực hiện phép tính trường âm thanh trong khối 710. Trong khối 720, khoảng cách thứ nhất của vị trí người nghe đến địa điểm tham chiếu thứ nhất của mô tả trường âm thanh thứ nhất được xác định. Đối với Fig.4c, đây là khoảng cách giữa địa điểm tham chiếu chênh lệch và vị trí tham chiếu A.

Hơn nữa, trong bước 722, khoảng cách thứ hai của vị trí người nghe mới tới địa điểm tham chiếu thứ hai của mô tả trường âm thanh thứ hai được xác định. Trong phương án trên Fig.4c, đây sẽ là khoảng cách giữa địa điểm tham chiếu khác (do tịnh tiến) và vị trí tham chiếu B.

Có vẻ như khoảng cách từ địa điểm tham chiếu khác đến vị trí tham chiếu B nhỏ hơn chênh lệch từ địa điểm tham chiếu khác đến vị trí tham chiếu A. Do đó, điều này sẽ được xác định trong bước 724. Và, trong bước 726, tín hiệu đối tượng được chọn từ nhóm được suy ra từ mô tả trường âm thanh với khoảng cách nhỏ hơn. Do đó, để kết xuất các nguồn âm thanh A và B mà tương ứng với các nguồn khớp trên hình minh họa Fig.4c, các tín hiệu nguồn âm thanh được suy ra từ mô tả trường âm thanh thứ hai liên quan đến vị trí tham chiếu tiếp theo B sẽ được sử dụng.

Tuy nhiên, theo các phương án khác, trong đó phép tịnh tiến chỉ từ gốc đến địa điểm tham chiếu khác, ví dụ, đến bên trái trong minh họa Fig.4c, khoảng cách nhỏ hơn sẽ từ địa điểm tham chiếu khác này đến vị trí tham chiếu A và, sau đó, mô tả trường âm thanh thứ nhất sẽ được sử dụng để tính toán cuối cùng trường âm thanh cho mỗi đối tượng khớp trong khối 710 trên Fig.6b. Một lần nữa, lựa chọn sẽ được thực hiện theo quy trình được minh họa trên Fig.7.

Fig.9 minh họa phương án được ưu tiên tiếp theo. Trong bước 740, phép phân tích trường âm thanh trên mô tả trường âm thanh thứ nhất được thực hiện, ví dụ, phân tích trường âm thanh tham số dưới dạng phân tích DirAC được minh họa trong khối 422 trên Fig.6c.

Điều này dẫn đến bộ tham số thứ nhất, ví dụ, cho mỗi ngăn thời gian/tần số,

trong đó mỗi bộ tham số bao gồm tham số DoA và, tùy chọn, tham số khuếch tán.

Trong bước 741, phép phân tích trường âm thanh được thực hiện trên mô tả trường âm thanh thứ hai và, một lần nữa, phân tích DirAC được thực hiện như trong khối 740 và như được, ví dụ, thảo luận đối với khối 422 trên Fig.6c.

Điều này dẫn đến bộ tham số thứ hai, ví dụ, đối với các ngăn thời gian/tần số.

Sau đó, trong khối 746, vị trí cho mỗi cặp tham số có thể được xác định sử dụng tham số DoA tương ứng từ ngăn thời gian/tần số thứ nhất và tham số DoA từ ngăn thời gian/tần số tương tự từ bộ tham số thứ hai. Điều này sẽ dẫn đến vị trí cho mỗi cặp tham số. Tuy nhiên, vị trí sẽ hữu dụng hơn khi độ khuếch tán thấp hơn cho ngăn thời gian/tần số tương ứng nằm trong bộ tham số thứ nhất và/hoặc bộ tham số thứ hai.

Do đó, chỉ ưu tiên sử dụng thêm các vị trí từ các ngăn thời gian/tần số mà dẫn đến độ khuếch tán khá thấp ở cả bộ tham số thứ nhất và thứ hai.

Ngoài ra, cũng ưu tiên thực hiện hiệu chỉnh cho các tín hiệu tương ứng trong ngăn thời gian/tần số cũng như đầu ra của khối 740 và khối 741.

Do đó, ví dụ, "khớp nguồn" của bước 704 trên Fig.6a có thể tránh hoàn toàn và được thay thế bằng cách xác định các nguồn khớp/các ngăn thời gian/tần số khớp dựa trên các tham số khuếch tán hoặc bước khớp có thể được thực hiện bổ sung sử dụng tín hiệu tương ứng trong ngăn thời gian/tần số từ các thành phần định dạng B ví dụ, hoặc từ tín hiệu áp lực hoặc tín hiệu đối tượng đưa ra bởi khối 422 trên Fig.6c.

Trong bất kỳ trường hợp nào, khối 46 sẽ dẫn đến các vị trí nhất định cho các ngăn thời gian/tần số (được chọn) nhất định mà tương ứng với "các đối tượng khớp" được tìm thấy trong khối 704 trên Fig.6a.

Do đó, trong khối 748, các tham số biến đổi và/hoặc các tín hiệu cho các vị trí thu được bởi khối 746 và/hoặc phép tịnh tiến/xoay tương ứng như thu được, ví dụ, bởi bộ theo dõi mũ được tính toán và đầu ra của khối 748 đại diện cho các tham số biến đổi và/hoặc các tín hiệu biến đổi cho các ngăn thời gian/tần số.

Do đó, khối 748 có thể tương ứng với biến đổi tịnh tiến 423 và biến đổi xoay của khối 424 nhằm mục đích tính toán các tham số biến đổi và phép tính toán các tín

hiệu biến đổi sẽ, ví dụ, được thực hiện bởi khối 425 trên Fig.6c tốt hơn là dưới sự xem xét của hệ số tỷ lệ nhất định được suy ra từ các vị trí cho các ngăn thời gian/tần số tương ứng.

Cuối cùng, tổng hợp mô tả trường âm thanh được thực hiện trong khối 750 sử dụng dữ liệu biến đổi. Ví dụ, điều này có thể được thực hiện bằng tổng hợp DirAC sử dụng hoặc mô tả trường âm thanh thứ nhất hoặc thứ hai hoặc có thể được thực hiện bởi bộ tạo tín hiệu ambisonic như được minh họa trong khối 425 và kết quả sẽ là mô tả trường âm thanh mới để truyền/lưu trữ/kết xuất.

Fig.10 minh họa phương án thực hiện được ưu tiên tiếp theo của bộ tính toán trường âm thanh 420. Ít nhất các phần của quy trình được minh họa trên Fig.10 được thực hiện cho mỗi nguồn âm thanh khớp một cách riêng biệt. Khối 1120 xác định khoảng cách cho nguồn khớp, ví dụ, bằng tính toán phép đạc tam giác.

Dựa trên mô tả trường âm thanh, hướng đến băng đầy đủ hoặc hướng đến của mỗi băng được xác định trong 1100. Những thông tin hướng đến này biểu diễn dữ liệu hướng đến của trường âm thanh. Dựa trên dữ liệu hướng đến này, phép biến đổi tịnh tiến được thực hiện trong khối 1110. Để đạt được mục đích này, khối 1120 tính toán khoảng cách cho mỗi nguồn khớp. Dựa trên dữ liệu, khối 1110 tạo dữ liệu hướng đến mới cho trường âm thanh mà, trong phương án thực hiện này, chỉ phụ thuộc vào sự tịnh tiến từ địa điểm tham chiếu đến địa điểm tham chiếu khác. Để đạt được mục đích này, khối 1110 nhận thông tin tịnh tiến được tạo, ví dụ, bằng cách theo dõi trong bối cảnh phương án thực hiện tế ảo.

Tốt hơn là hoặc thay vào đó, dữ liệu xoay cũng được sử dụng. Để đạt được mục đích này, khối 1130 thực hiện biến đổi xoay sử dụng thông tin xoay. Khi cả phép tịnh tiến và xoay được thực hiện, thì nó ưu tiên thực hiện biến đổi xoay sau khi tính toán các DoA mới của trường âm thanh đã bao gồm thông tin từ phép tịnh tiến và khoảng cách nguồn từ khối 1120.

Sau đó, trong khối 1140, mô tả trường âm thanh mới được tạo. Để đạt được mục đích này, mô tả trường âm thanh ban đầu có thể được sử dụng hoặc, thay vào đó, các tín hiệu nguồn mà được tách ra từ mô tả trường âm thanh bởi thuật toán tách

nguồn có thể được sử dụng hoặc bất kỳ ứng dụng nào khác có thể được sử dụng. Về cơ bản, mô tả trường âm thanh mới có thể là, ví dụ, mô tả trường âm thanh định hướng như thu được bởi bộ tạo ambisonic 430 hoặc như được tạo bởi bộ tổng hợp DirAC 425 hoặc có thể là phép biểu diễn lập thể được tạo từ phép biểu diễn loa phát thanh ảo trong kết xuất lập thể tiếp theo.

Tốt hơn là, như được minh họa trên Fig.10, khoảng cách trên mỗi hướng đến cũng được sử dụng để tạo mô tả trường âm thanh mới để thích ứng âm lượng hoặc cường độ âm của nguồn âm thanh nhất định với địa điểm mới, nghĩa là, địa điểm tham chiếu mới hoặc khác.

Mặc dù Fig.10 minh họa tình huống, trong đó phép biến đổi xoay được thực hiện sau phép biến đổi tịnh tiến, cần lưu ý rằng thứ tự có thể khác nhau. Đặc biệt, phép biến đổi xoay có thể được áp dụng cho các DoA của trường âm thanh như được tạo bởi khối 1100 và, sau đó, phép biến đổi tịnh tiến bổ sung được áp dụng do tịnh tiến chủ thể từ địa điểm tham chiếu đến địa điểm tham chiếu khác.

Ngay sau khi các DoA của trường âm thanh được xác định bởi khối 1100, thông tin khoảng cách được lấy từ siêu dữ liệu sử dụng khối 1120 và thông tin khoảng cách này sau đó được sử dụng bằng cách tạo mô tả trường âm thanh mới trong khối 1140 để tính toán khoảng cách bị thay đổi và, do đó, cường độ âm bị thay đổi của nguồn nhất định đối với địa điểm tham chiếu nhất định. Về cơ bản, có thể nói rằng trong trường hợp khoảng cách trở nên lớn hơn, thì tín hiệu nguồn âm thanh cụ thể có thể được giảm dần trong khi, khi khoảng cách trở nên ngắn hơn, thì tín hiệu nguồn âm thanh được khuếch đại. Đương nhiên, sự giảm dần hoặc khuếch đại của nguồn âm thanh nhất định phụ thuộc vào khoảng cách được thực hiện theo tỷ lệ thay đổi khoảng cách, nhưng, theo các phương án khác, các hoạt động ít phức tạp hơn có thể được áp dụng cho khuếch đại hoặc giảm dần này của các tín hiệu nguồn âm thanh theo mức tăng khá thô. Ngay cả việc thực hiện ít phức tạp hơn như vậy mang lại kết quả vượt trội so với tình huống mà bất kỳ thay đổi khoảng cách nào đều bị bỏ qua hoàn toàn.

Fig.11 minh họa phương án thực hiện được ưu tiên tiếp theo của bộ tính toán trường âm thanh. Trong khối 1200, các nguồn âm thanh riêng lẻ từ trường âm thanh

được xác định, ví dụ, cho mỗi băng hoặc toàn băng. Khi phép xác định cho mỗi khung và băng được thực hiện, thì điều này có thể được hoàn thành bằng phép phân tích DirAC. Nếu phép xác định toàn băng hoặc băng phụ được thực hiện, thì điều này có thể được hoàn thành bằng bất kỳ loại thuật toán tách nguồn toàn băng hoặc băng phụ nào.

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

Trong khối 1220, khoảng cách cũ cho mỗi nguồn được xác định sử dụng siêu dữ liệu và, ví dụ, sử dụng siêu dữ liệu để tính toán phép đặc tam giác. Do đó, mỗi băng được coi là nguồn nhất định (với điều kiện là độ khuếch tán thấp hơn một ngưỡng nhất định), và sau đó, khoảng cách nhất định cho mỗi ngăn thời gian/tần số có giá trị độ khuếch tán thấp được xác định.

Sau đó, trong khối 1230, khoảng cách mới cho mỗi nguồn được thu, ví dụ, bởi phép tính vectơ cho mỗi băng mà, ví dụ, được thảo luận trong bối cảnh trên Fig.6b.

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

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

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

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

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

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

Fig.12a minh họa bộ phân tích DirAC như được bộc lộ ban đầu, ví dụ, trong tài liệu tham khảo được trích dẫn trước đó "Mã hóa âm thanh có hướng" ("Directional Audio Coding") từ IWPASH năm 2009.

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

Trong DirAC, cả phép phân tích và tổng hợp đều được thực hiện trong miền tần số. Có một số phương pháp để phân chia âm thanh thành các băng tần số, 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 băng lọc gương cầu phương (Quadrature mirror filter - QMF). Ngoài ra, có toàn quyền tự do để thiết kế băng lọc với các bộ lọc tùy ý được tối ưu hóa cho bất kỳ mục đích cụ thể nào. Mục tiêu của phân tích hướng là ước tính ở mỗi dải tần số hướng đến của âm thanh, cùng với ước tính nếu âm thanh đến từ một hoặc nhiều hướng cùng một lúc. Về nguyên tắc, điều này có thể được thực hiện với một số kỹ thuật, tuy nhiên, phân tích năng lượng của trường âm thanh đã được tìm thấy là phù hợp, được minh họa trên Fig.12a. Phân tích năng lượng có thể được thực hiện, khi tín hiệu áp lực và các tín hiệu vận tốc ở

một, hai hoặc ba chiều được thu ở một vị trí duy nhất. Trong các tín hiệu định dạng B cấp một, tín hiệu đẳng hướng được gọi là tín hiệu W, mà được thu nhỏ bởi căn bậc hai của hai. Áp lực âm thanh có thể được ước tính như $P = \sqrt{2} * W$ được biểu diễn trong miền STFT.

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

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

Fig.12b minh họa phép tổng hợp DirAC, một lần nữa có dàn bộ lọc băng 1370, khối micrô ảo 1400, khối tổng hợp hướng/khuếch tán 1450, và thiết lập loa phát thanh nhất định hoặc thiết lập loa phát thanh dự định ảo 1460. Ngoài ra, bộ biến đổi khuếch đại khuếch tán 1380, khối bảng khuếch đại quét biên độ dựa trên vectơ (vector based amplitude panning - VBAP) 1390, khối bù micrô 1420, khối trung bình khuếch đại loa phát thanh 1430 và bộ phân phối 1440 cho các kênh khác được sử dụng.

Trong phép tổng hợp DirAC này, với các loa phát thanh, phiên bản chất lượng cao của tổng hợp DirAC được thể hiện trên Fig.12b nhận tất cả các tín hiệu định dạng B, trong đó tín hiệu micrô ảo được tính toán cho mỗi hướng loa phát thanh của thiết lập loa phát thanh 1460. Mẫu định hướng được sử dụng thường là lưỡng cực. Các tín

hiệu micrô ảo sau đó biến đổi theo kiểu phi tuyến tính, tùy thuộc vào siêu dữ liệu. Phiên bản tốc độ bit thấp của DirAC không được thể hiện trên Fig.12b, tuy nhiên, trong tình huống này, chỉ một kênh của âm thanh được truyền như được minh họa trên Fig.6. Sự khác biệt trong bước xử lý là tất cả các tín hiệu micrô ảo sẽ được thay thế bằng một kênh âm thanh duy nhất được nhận. Các tín hiệu micrô ảo được chia thành hai dòng: dòng khuếch tán và dòng không khuếch tán, mà được xử lý riêng.

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

Trong nhiều trường hợp, siêu dữ liệu định hướng phải tuân theo các thay đổi đột ngột theo thời gian. Để tránh các giả âm, các hệ số khuếch đại cho các loa phát thanh được tính toán với VBAP được làm nhẵn bằng cách tích hợp tạm thời với các hằng số thời gian phụ thuộc vào tần số tương đương với khoảng 50 chu kỳ ở mỗi băng tần. Điều này có hiệu quả loại bỏ các giả âm, tuy nhiên, các thay đổi theo hướng không được cảm nhận là chậm hơn so với không lấy trung bình trong hầu hết các trường hợp.

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

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

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

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

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

Một số phương án theo sáng chế bao gồm vật mang dữ liệu không chuyển tiếp có các tín hiệu điều khiển có thể đọc được bằng điện tử, có khả năng kết hợp với hệ thống máy tính có thể lập trình, chẳng hạn như một trong số các phương pháp được mô tả trong sáng chế này được thực hiện.

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

Các phương án khác bao gồm chương trình máy tính để thực hiện một trong số các phương pháp được mô tả trong sáng chế này, được lưu trữ trên vật mang có thể đọc được bằng máy.

Nói cách khác, phương án theo phương pháp sáng chế, do đó, là chương trình máy tính có mã chương trình để thực hiện một trong số các phương pháp được mô tả trong sáng chế này, khi chương trình máy tính chạy trên máy tính.

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

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

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

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

Trong một số phương án, thiết bị logic lập trình (ví dụ mảng cổng lập trình trường) có thể được sử dụng để thực hiện một số hoặc tất cả các chức năng của các phương pháp được mô tả trong tài liệu này. Trong một số phương án, mảng cổng lập trình trường có thể kết hợp với bộ vi xử lý để thực hiện một trong số các phương pháp được mô tả trong tài liệu này. Nói chung, các phương pháp tốt hơn là được thực hiện bởi bất kỳ thiết bị phần cứng nào.

Các phương án được mô tả ở trên chỉ mang tính minh họa cho các nguyên tắc của sáng chế. Điều này được hiểu rằng các biến đổi và biến thể của các sắp xếp và các

chi tiết được mô tả trong tài liệu này sẽ rõ ràng đối với những người có hiểu biết trung bình trong lĩnh vực tương ứng. Do đó, sáng chế chỉ bị giới hạn bởi phạm vi của các điểm yêu cầu bảo hộ mà không phải bởi các chi tiết cụ thể được mô tả và giải thích theo các phương án theo sáng chế trong tài liệu này.

YÊU CẦU BẢO HỘ

1. Thiết bị tạo mô tả trường âm thanh nâng cao, thiết bị bao gồm:

bộ tạo trường âm thanh để tạo mô tả trường âm thanh thứ nhất biểu thị trường âm thanh đối với địa điểm tham chiếu thứ nhất; và mô tả trường âm thanh thứ hai biểu thị trường âm thanh đối với địa điểm tham chiếu thứ hai, địa điểm tham chiếu thứ hai khác với địa điểm tham chiếu thứ nhất; và

bộ tạo siêu dữ liệu để tạo siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh, trong đó bộ tạo siêu dữ liệu được tạo cấu hình để xác định ít nhất một trong số mô tả hình học thứ nhất của mô tả trường âm thanh thứ nhất và mô tả hình học thứ hai của trường âm thanh thứ hai dưới dạng siêu dữ liệu, và

trong đó mô tả trường âm thanh thứ nhất, mô tả trường âm thanh thứ hai, và siêu dữ liệu bao gồm ít nhất một trong số mô tả hình học thứ nhất và mô tả hình học thứ hai, tạo thành mô tả trường âm thanh nâng cao.

2. Thiết bị theo điểm 1,

trong đó bộ tạo siêu dữ liệu được tạo cấu hình để xác định, dưới dạng mô tả hình học thứ nhất, một trong số địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai hoặc khoảng cách giữa địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai hoặc vectơ địa điểm giữa địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai dưới dạng siêu dữ liệu.

3. Thiết bị theo điểm 1,

trong đó mô tả trường âm thanh thứ nhất là mô tả ambisonic thứ nhất, và trong đó mô tả trường âm thanh thứ hai là mô tả ambisonic thứ hai, hoặc

trong đó mô tả trường âm thanh thứ nhất là mô tả DirAC thứ nhất và mô tả trường âm thanh thứ hai là mô tả DirAC thứ hai.

4. Thiết bị theo điểm 1,

trong đó mô tả hình học thứ nhất là thông tin về véctor thứ nhất được hướng từ điểm gốc được xác định trước đến địa điểm tham chiếu thứ nhất của mô tả trường âm thanh thứ nhất, và

trong đó mô tả hình học thứ hai là thông tin trên véctor thứ hai được hướng từ gốc được xác định trước đến địa điểm tham chiếu thứ hai của mô tả trường âm thanh thứ hai.

5. Thiết bị theo điểm 1,

trong đó mô tả hình học thứ nhất là thông tin trên một trong số véctor thứ nhất được hướng từ gốc được xác định trước đến địa điểm tham chiếu thứ nhất của mô tả trường âm thanh thứ nhất và véctor thứ hai được hướng từ gốc được xác định trước đến địa điểm tham chiếu thứ hai của mô tả trường âm thanh thứ hai, và trong đó mô tả hình học thứ hai bao gồm thông tin trên véctor giữa địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai.

6. Thiết bị theo điểm 1,

trong đó một trong số địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai là gốc được xác định trước biểu diễn mô tả hình học thứ nhất, và trong đó dữ liệu mô tả hình học thứ hai bao gồm thông tin trên véctor giữa gốc được xác định trước và địa điểm khác của địa điểm tham chiếu thứ nhất và địa điểm tham chiếu thứ hai.

7. Thiết bị theo điểm 1, trong đó bộ tạo trường âm thanh được tạo cấu hình để tạo mô tả trường âm thanh thứ nhất hoặc mô tả trường âm thanh thứ hai sử dụng thiết bị micro thực hoặc bằng phép tổng hợp âm thanh sử dụng kỹ thuật micro ảo.

8. Phương pháp tạo mô tả trường âm nâng cao, phương pháp bao gồm các bước:

tạo mô tả trường âm thanh thứ nhất biểu thị trường âm thanh đối với địa điểm tham chiếu thứ nhất;

tạo mô tả trường âm thanh thứ hai biểu thị trường âm thanh đối với địa điểm tham chiếu thứ hai, địa điểm tham chiếu thứ hai khác với địa điểm tham chiếu thứ nhất; và

tạo siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh, bước tạo siêu dữ liệu bao gồm xác định ít nhất một trong số mô tả hình học thứ nhất của mô tả trường âm thanh thứ nhất và mô tả hình học thứ hai của mô tả trường âm thanh thứ hai dưới dạng siêu dữ liệu, và

trong đó mô tả trường âm thanh thứ nhất, mô tả trường âm thanh thứ hai và siêu dữ liệu bao gồm ít nhất một trong số mô tả hình học thứ nhất và mô tả hình học thứ hai, tạo thành mô tả trường âm thanh nâng cao.

9. Vật ghi lưu trữ số không tạm thời có chương trình máy tính được lưu trữ trên đó để thực hiện phương pháp tạo mô tả trường âm thanh nâng cao, bao gồm các bước:

tạo mô tả trường âm thanh thứ nhất biểu thị trường âm thanh đối với địa điểm tham chiếu thứ nhất;

tạo mô tả trường âm thanh thứ hai biểu thị trường âm thanh đối với địa điểm tham chiếu thứ hai, địa điểm tham chiếu thứ hai khác với địa điểm tham chiếu thứ nhất; và

tạo siêu dữ liệu liên quan đến thông tin không gian của trường âm thanh, bước tạo siêu dữ liệu bao gồm xác định ít nhất một trong số mô tả hình học thứ nhất của mô tả trường âm thanh thứ nhất và mô tả hình học thứ hai của mô tả trường âm thanh thứ hai dưới dạng siêu dữ liệu, và

trong đó mô tả trường âm thanh thứ nhất, mô tả trường âm thanh thứ hai và siêu dữ liệu bao gồm ít nhất một trong số mô tả hình học thứ nhất và mô tả hình học thứ hai, tạo thành mô tả trường âm thanh nâng cao,

khi chương trình máy tính này được chạy bằng máy tính.

1/16

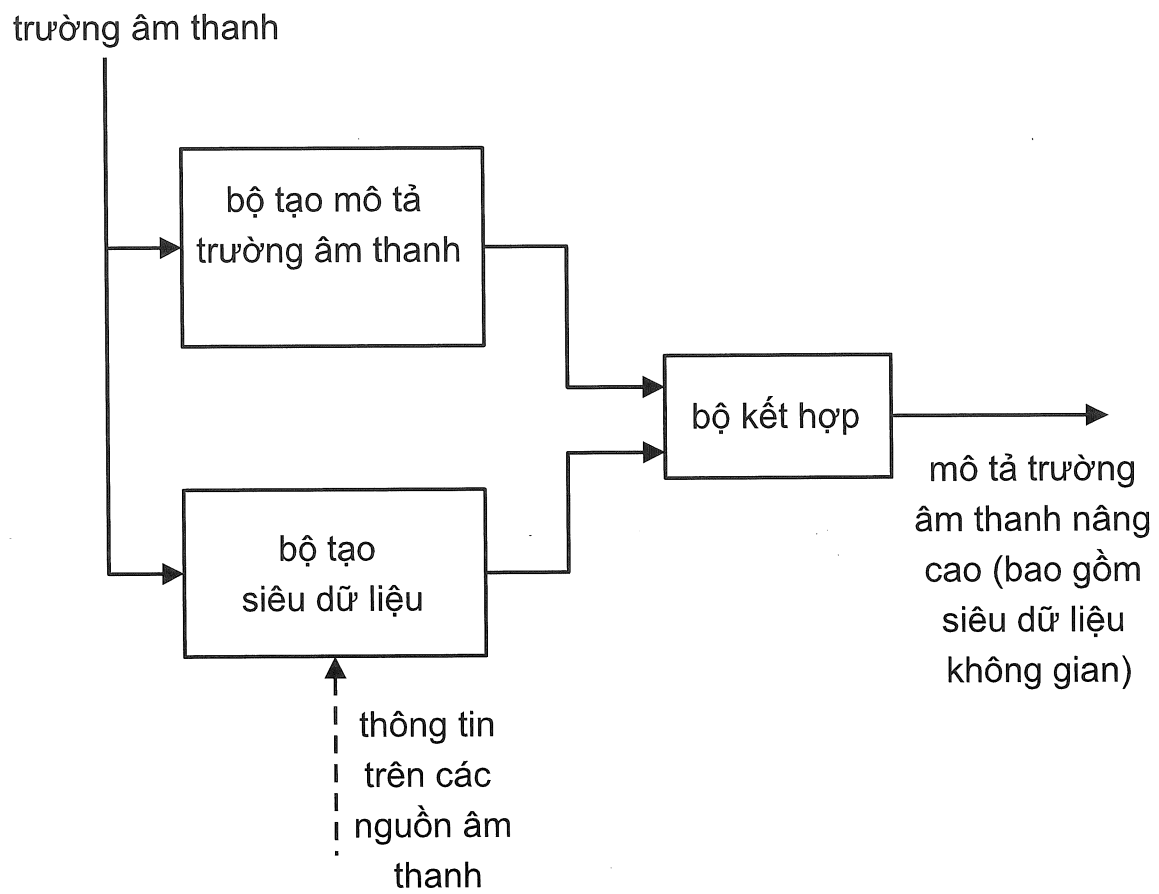


Fig. 1a

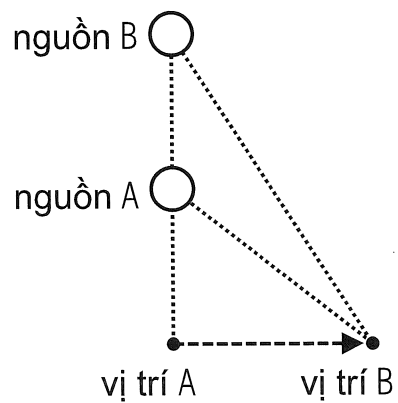


Fig. 1b

2/16

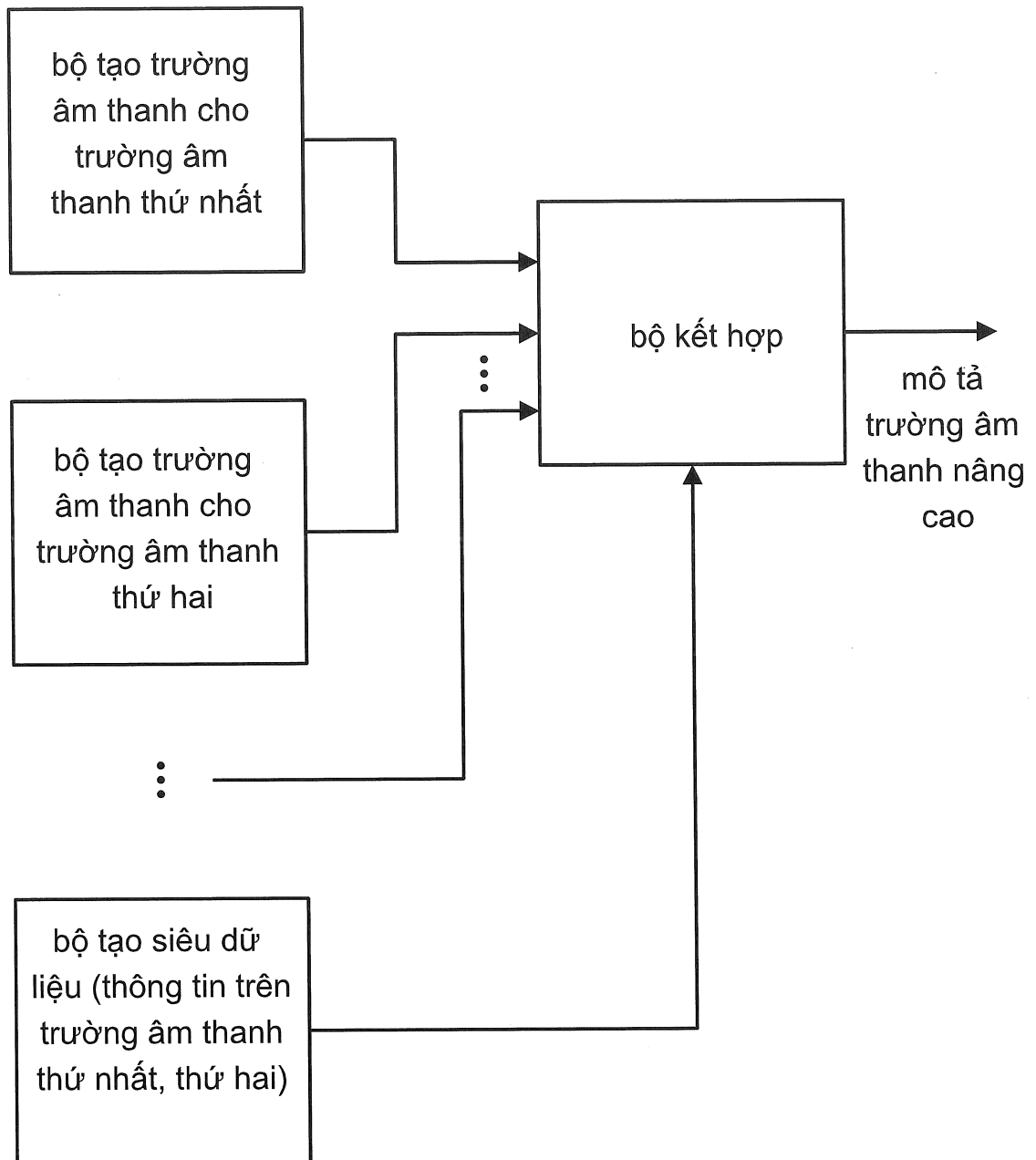


Fig. 2

3/16

	mô tả trường âm thanh thứ nhất	mô tả trường âm thanh thứ hai	thông tin trên trường âm thanh thứ nhất và thứ hai
--	-----------------------------------	----------------------------------	--

Fig. 3a

điểm tham chiếu	
trường âm thanh thứ nhất: thông tin hình học thứ nhất	thông tin trên vectơ $\vec{A}$
trường âm thanh thứ hai: thông tin hình học thứ hai	thông tin trên vectơ $\vec{B}$ hoặc
trường âm thanh thứ ba: thông tin hình học thứ ba (tùy chọn)	thông tin trên vectơ $\vec{D}$ và/hoặc gốc và/hoặc vectơ $\vec{A}$
⋮	hoặc thông tin trên vectơ $\vec{B}$ và trên vectơ $\vec{D}$

Fig. 3b

4/16

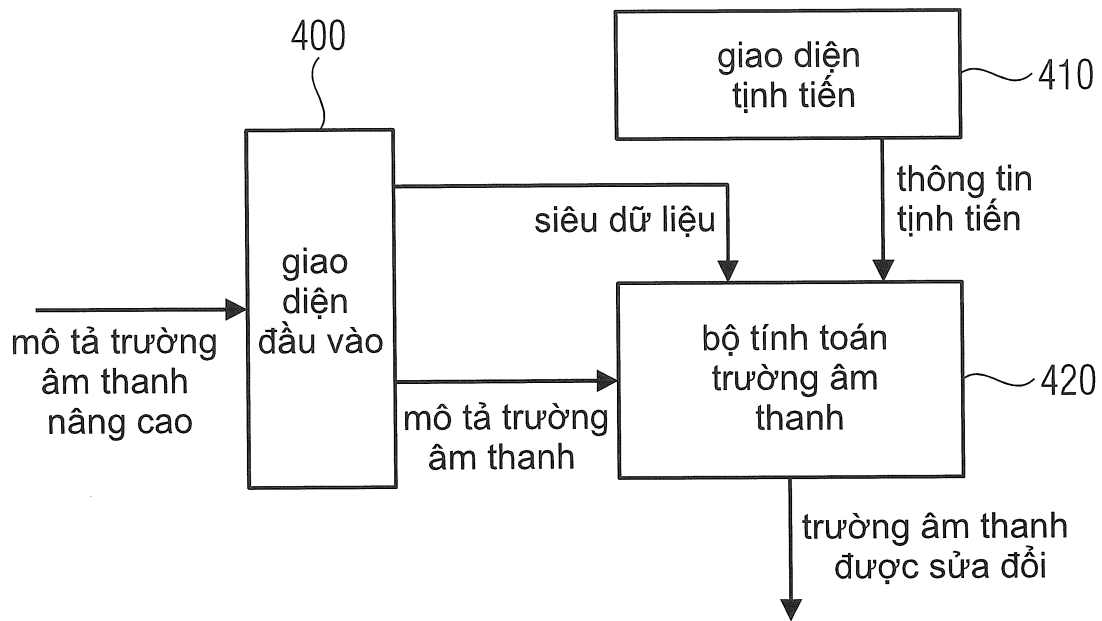


Fig. 4a

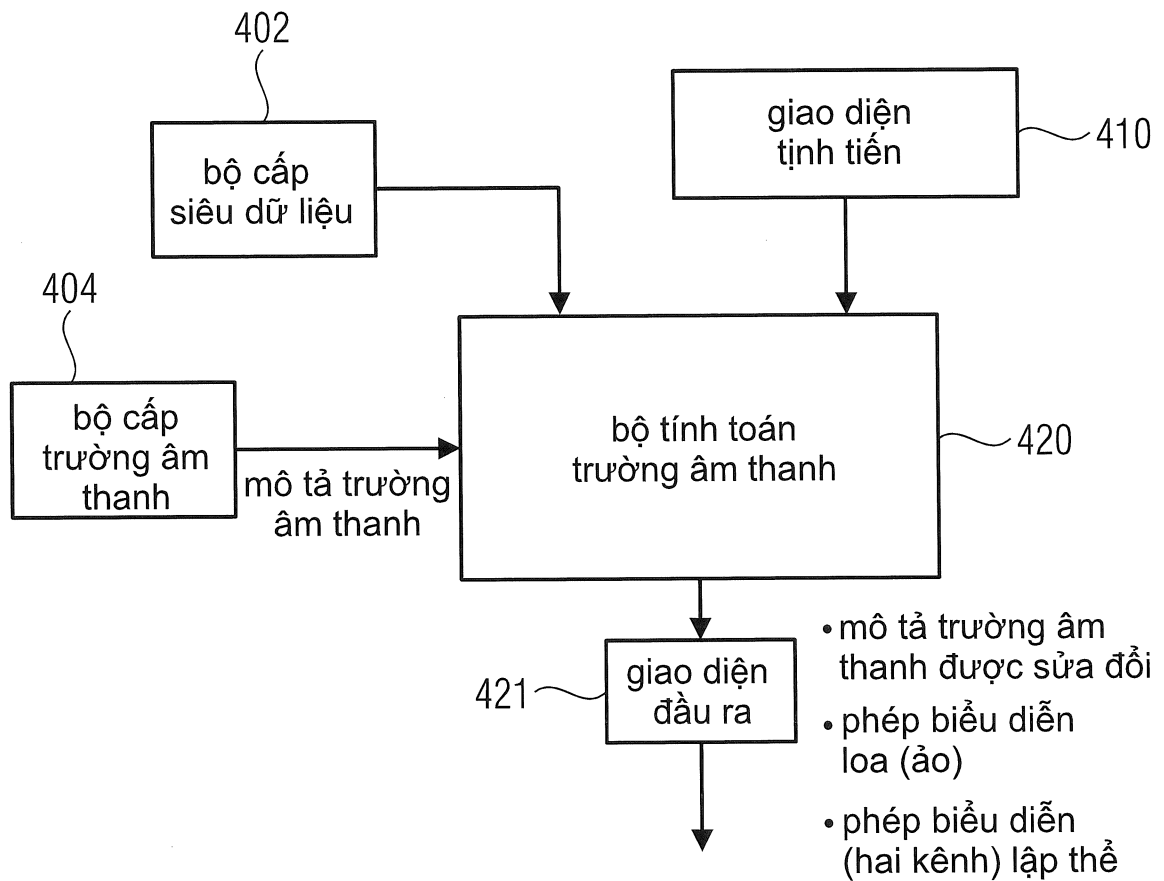


Fig. 4b

5/16

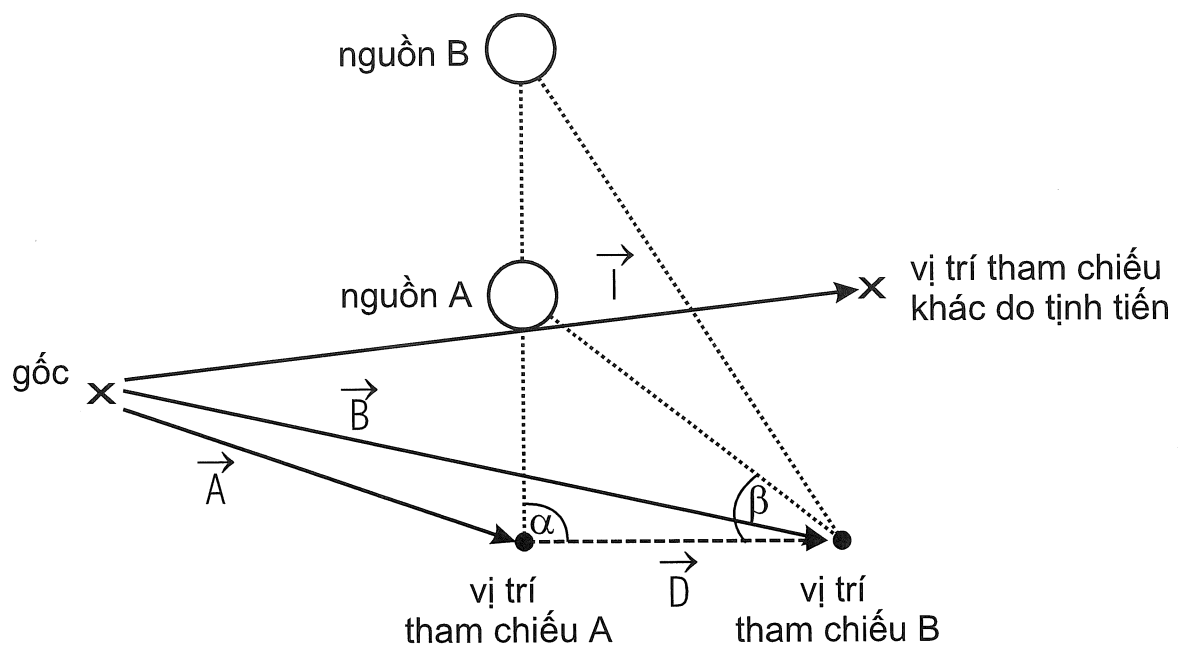


Fig. 4c

6/16

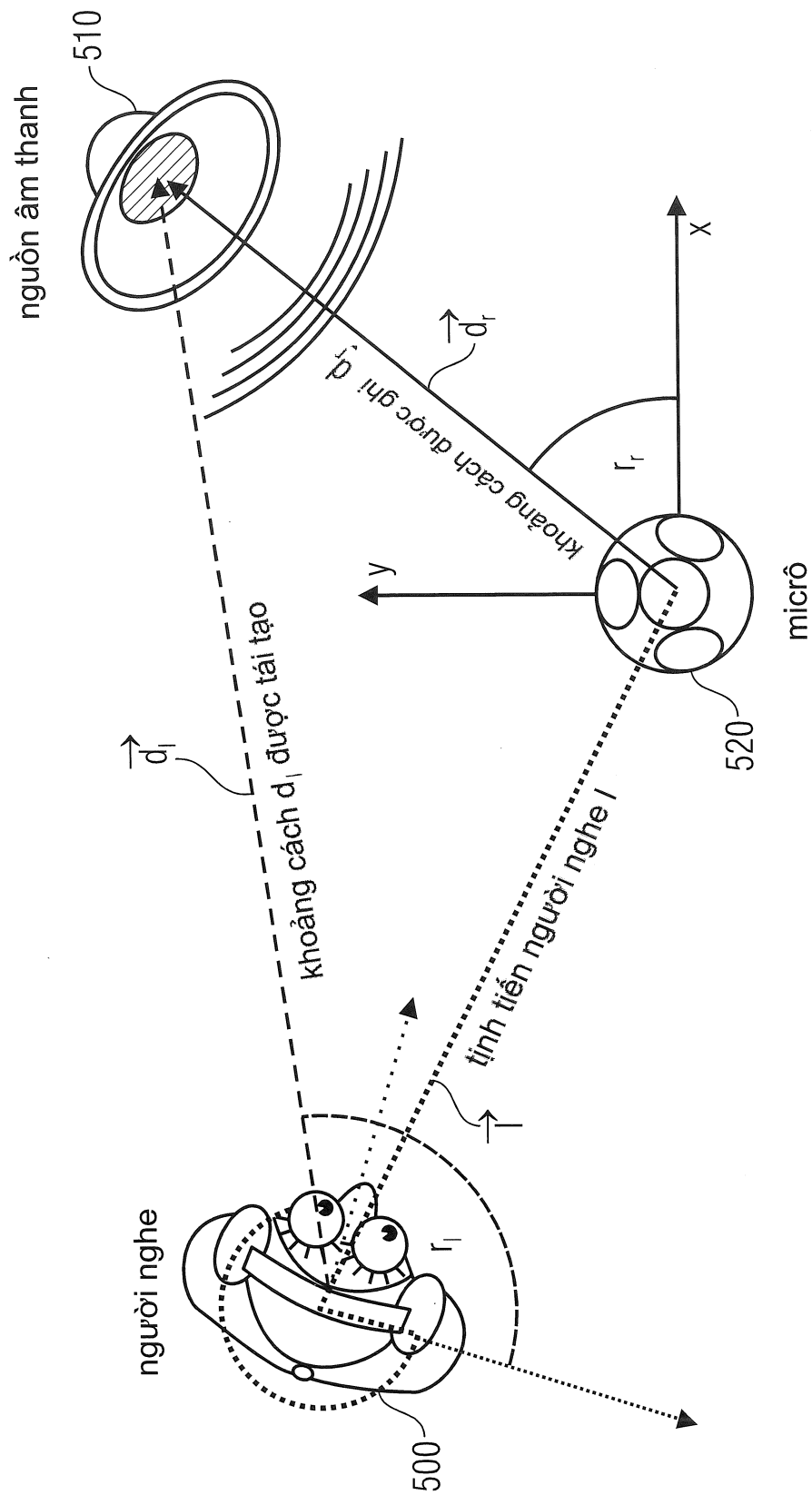


Fig. 5

7/16

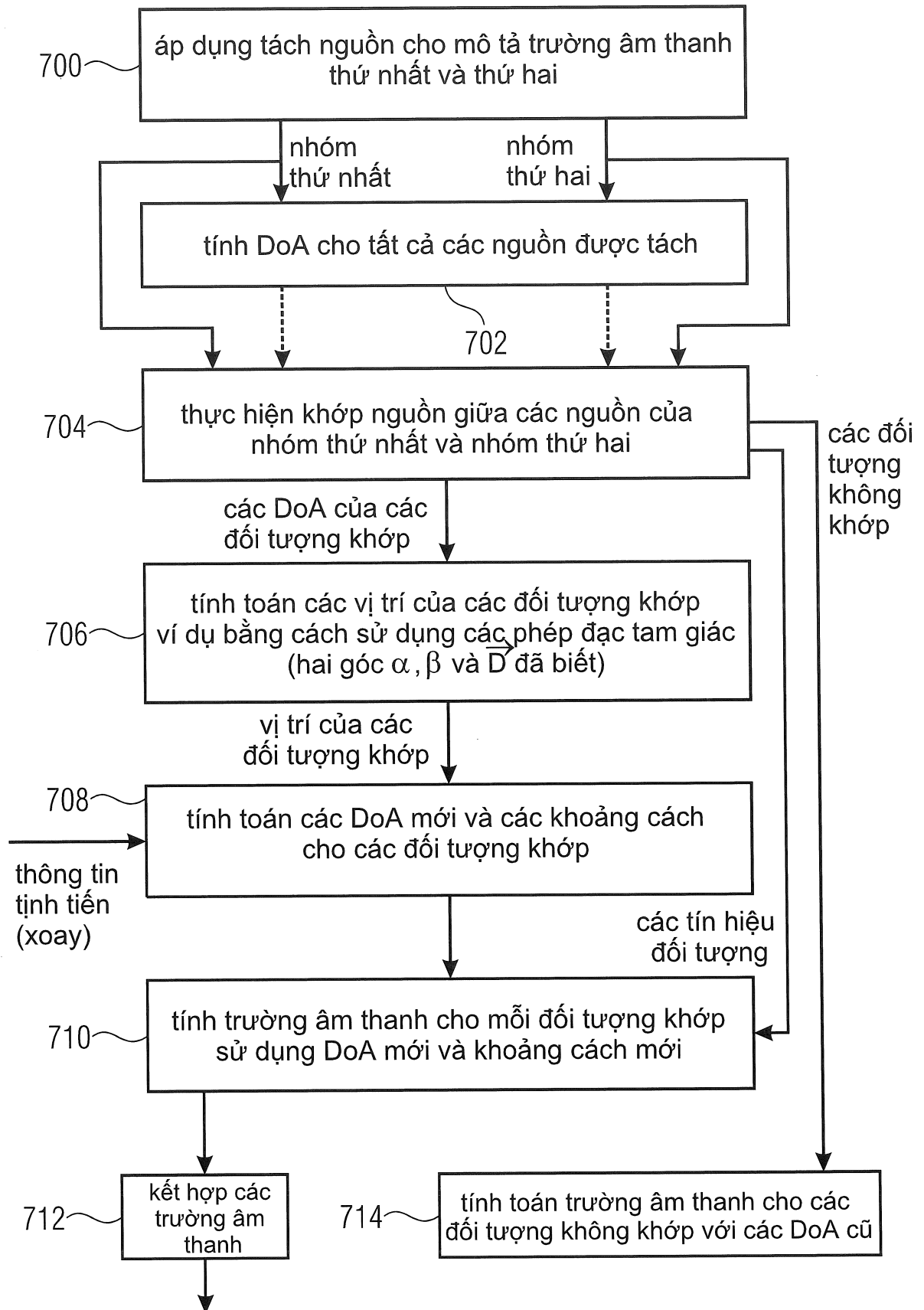


Fig. 6a

8/16

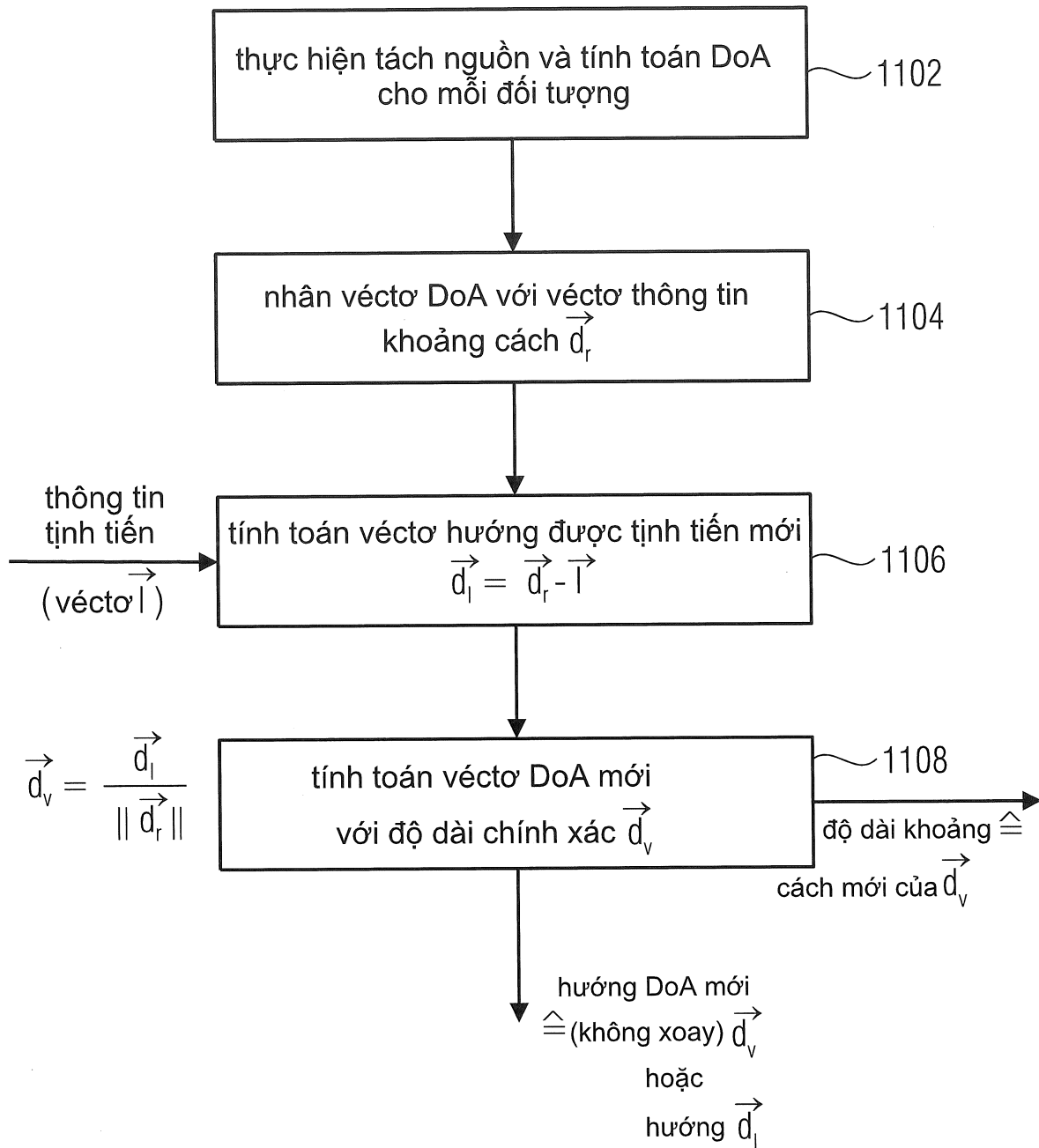


Fig. 6b

9/16

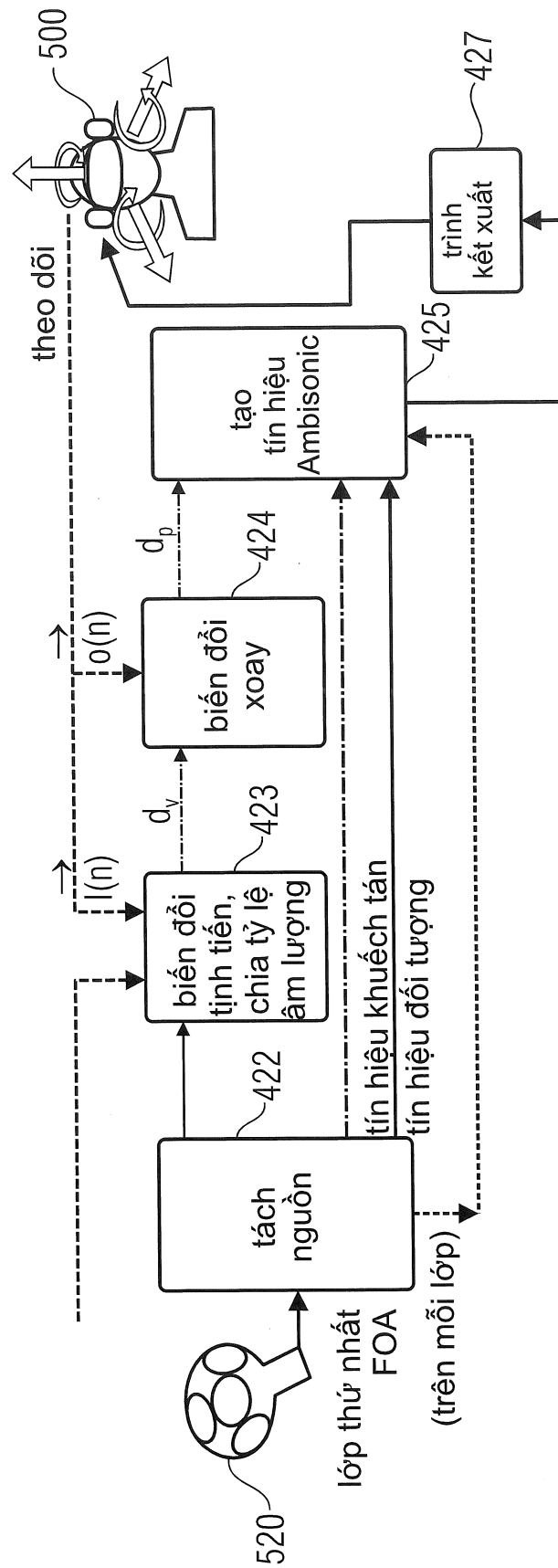


Fig. 6c

10/16

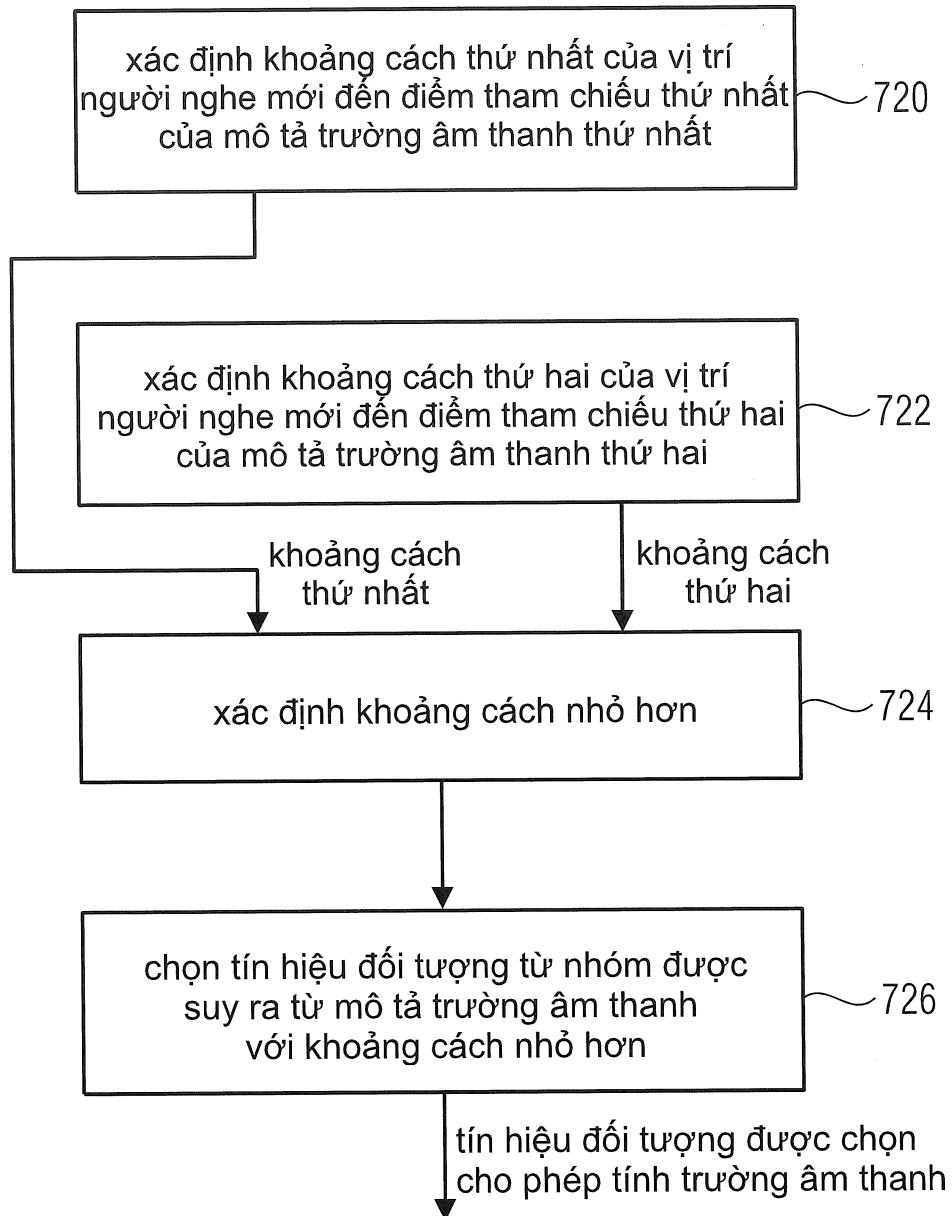
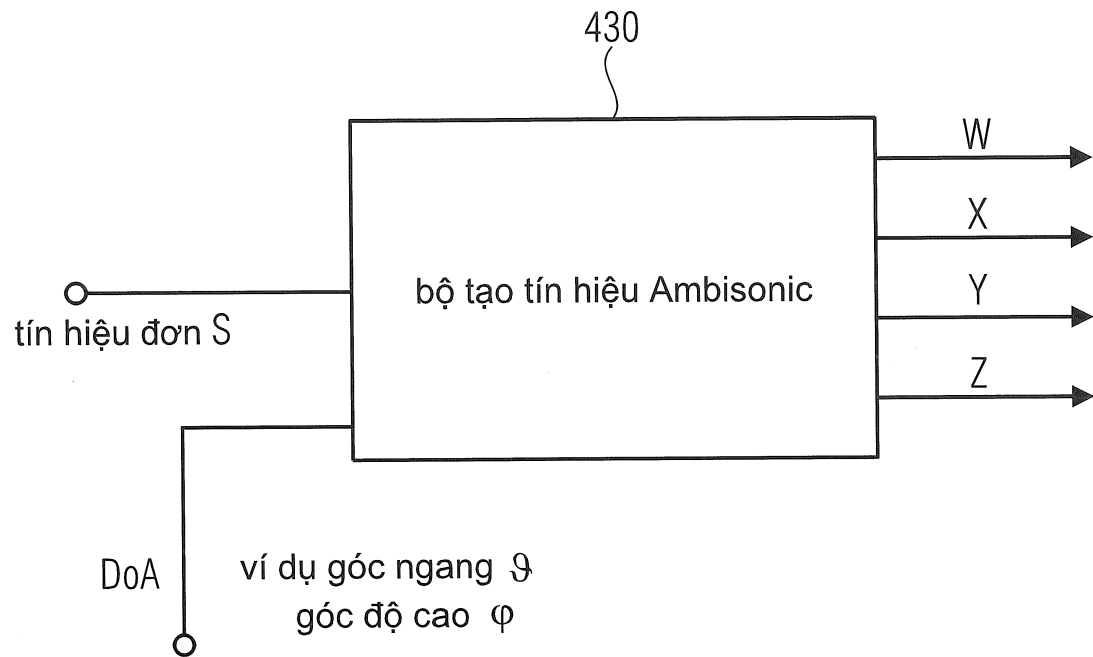


Fig. 7

11/16



phức hợp đẳng hướng $W = S \cdot \frac{1}{\sqrt{2}}$

$$\text{các thành phần định hướng} \begin{cases} X = S \cdot \cos \vartheta \cos \varphi \\ Y = S \cdot \sin \vartheta \cos \varphi \\ Z = S \cdot \sin \varphi \end{cases}$$

Fig. 8

12/16

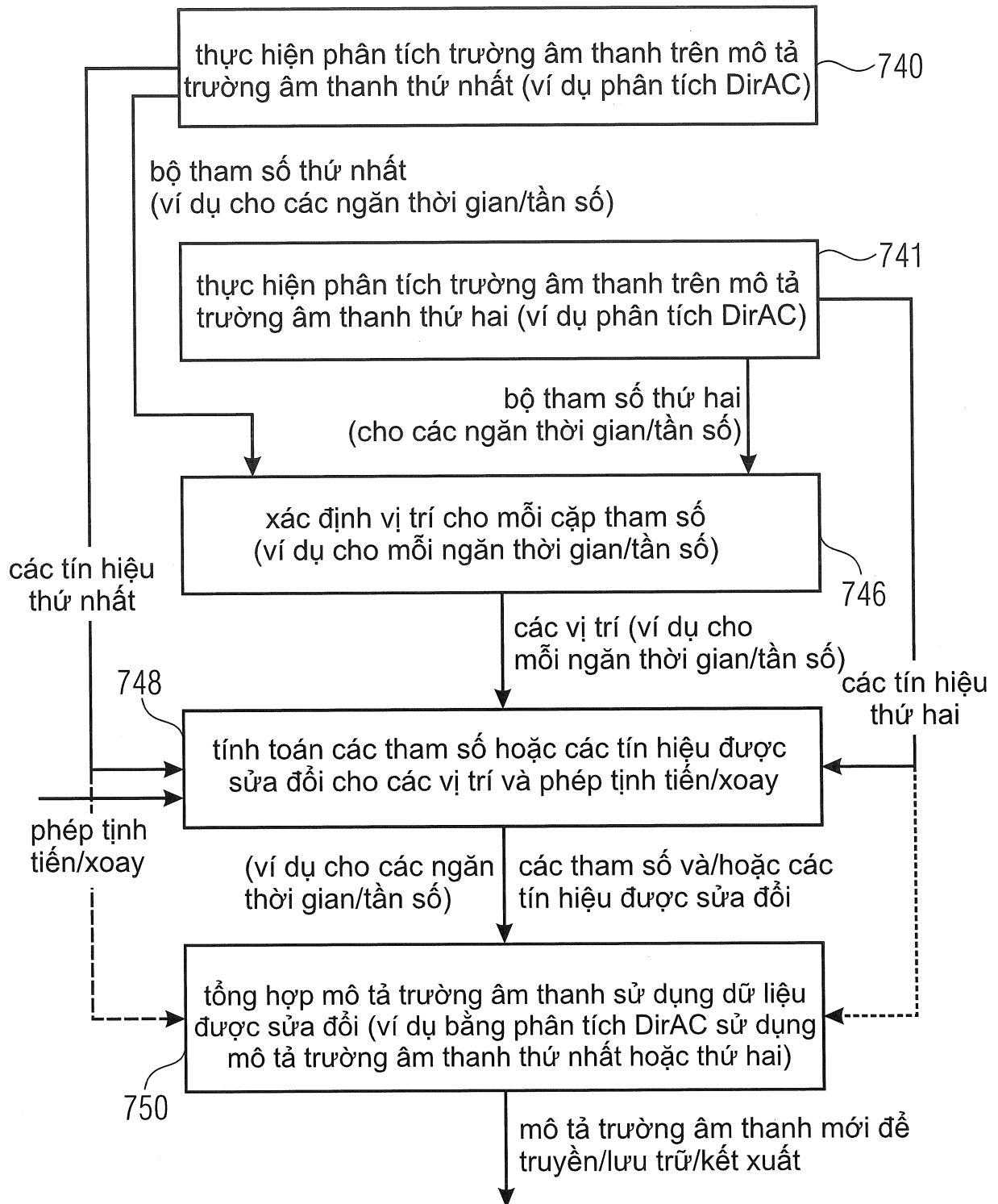


Fig. 9

13/16

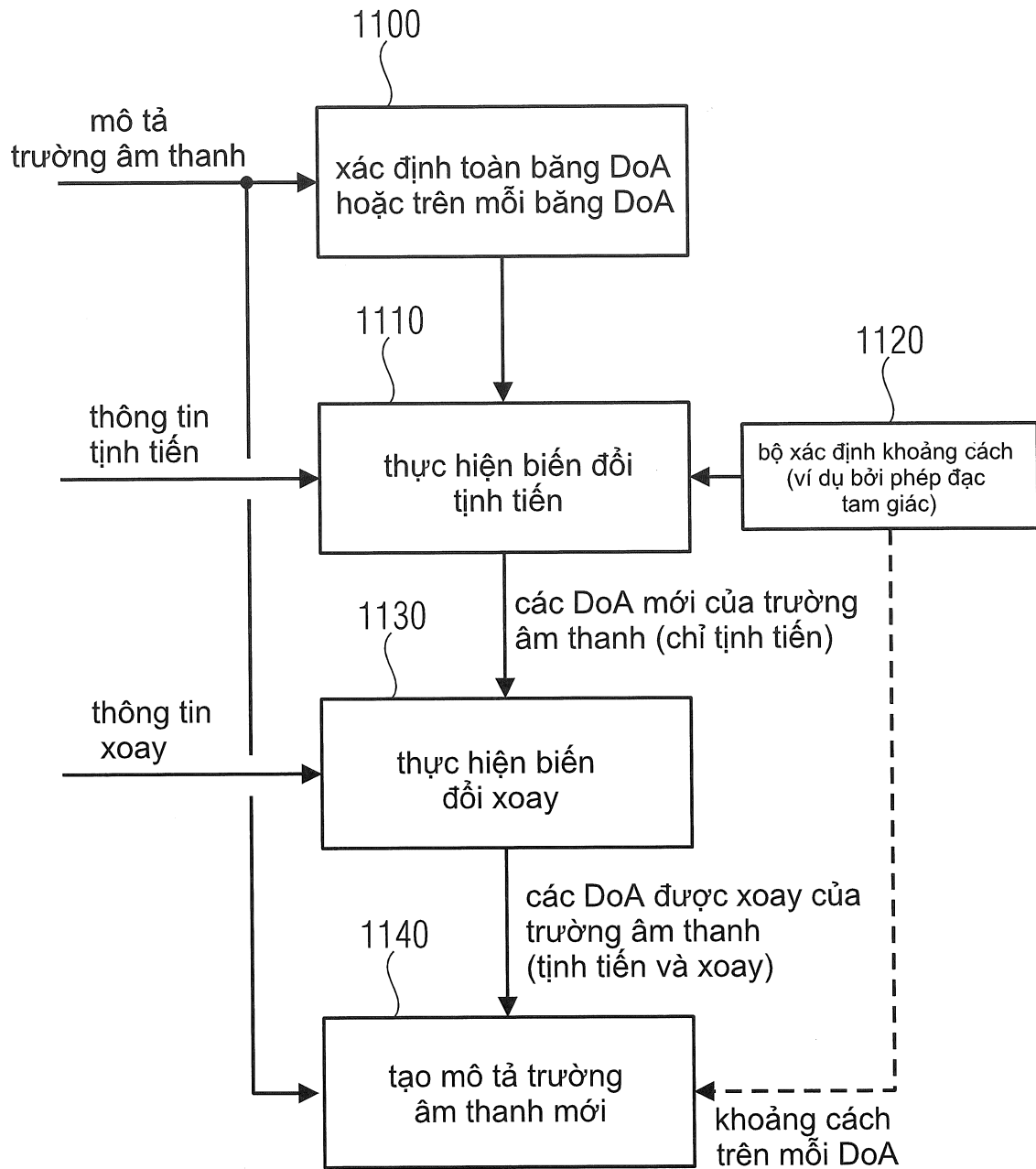


Fig. 10

14/16

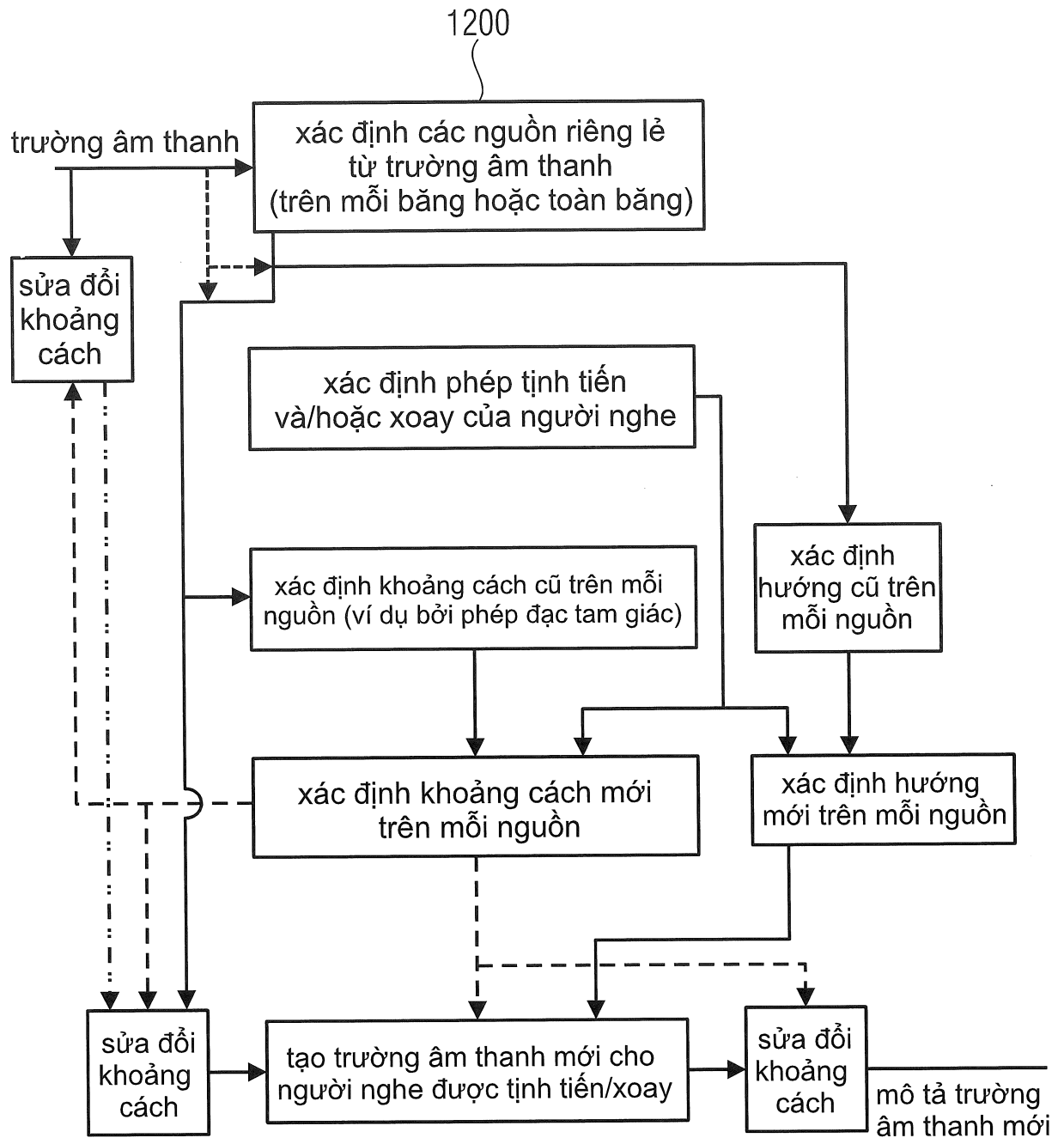
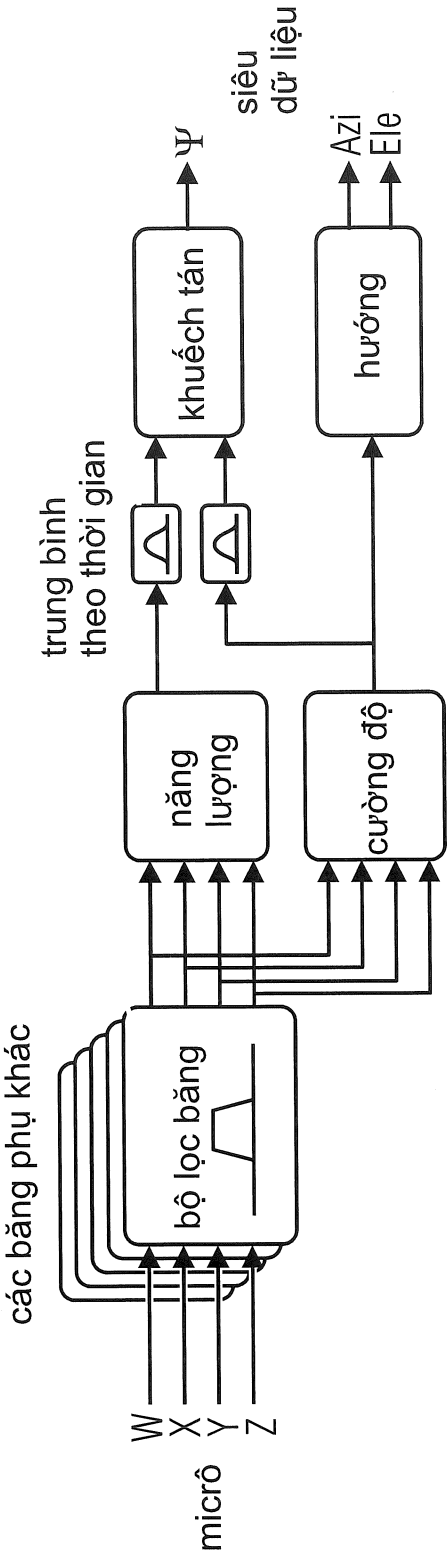
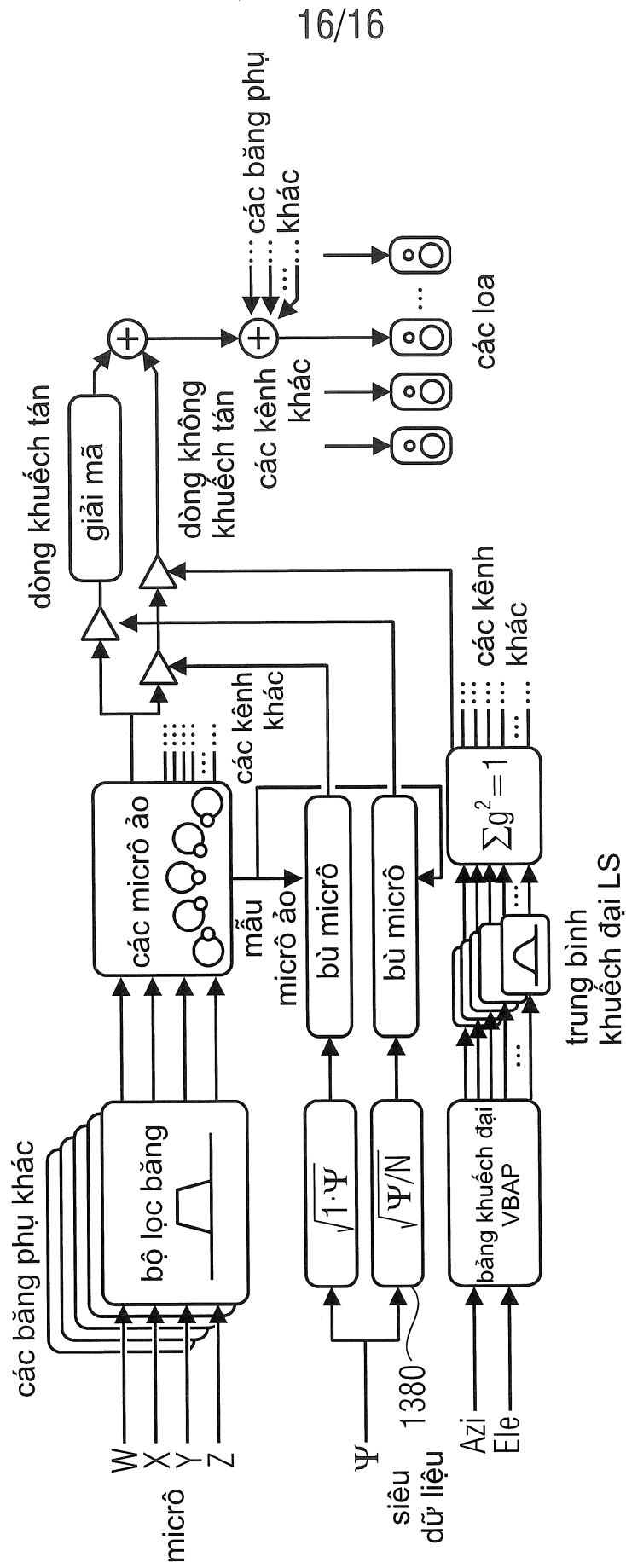


Fig. 11



Phân tích DirAC

Fig. 12a



DirAC synthesis

Fig. 12b