



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

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

(11)



1-0024725

(51)⁷ G10L 19/00

(13) B

(21) 1-2015-00720

(22) 05/08/2013

(86) PCT/EP2013/066405 05/08/2013

(87) WO2014/020182A2 06/02/2014

(30) 61/679,404 03/08/2012 US

(45) 27/07/2020 388

(43) 27/07/2015 328A

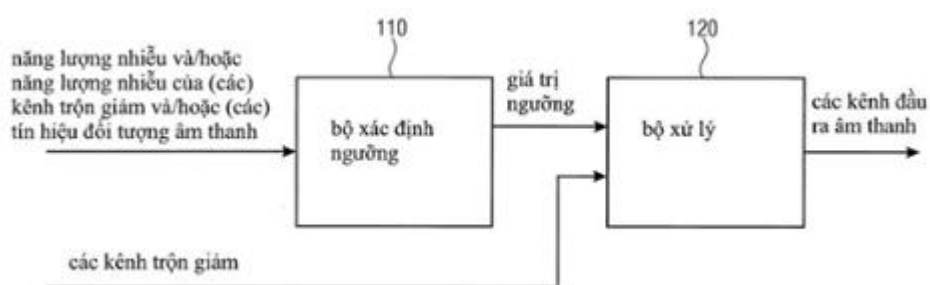
(73) Fraunhofer-Gesellschaft zur Foerderung angewadten Forschung e.V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany

(72) TERENTIV, Leon (DE); HELLMUTH, Oliver (DE); HERRE, Juergen (DE);
KASTNER, Thorsten (DE).

(74) Công ty Luật TNHH AMBYS Hà Nội (AMBYS HANOI)

(54) BỘ GIẢI MÃ VÀ PHƯƠNG PHÁP GIẢI MÃ TẠO RA TÍN HIỆU ĐẦU RA ÂM THANH

(57) Sáng chế đề cập đến bộ giải mã và phương pháp tạo ra tín hiệu đầu ra âm thanh. Bộ giải mã để tạo ra tín hiệu đầu ra âm thanh bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm được đề xuất. Tín hiệu trộn giảm mã hóa một hoặc nhiều tín hiệu đối tượng âm thanh. Bộ mã hóa bao gồm bộ xác định ngưỡng (110) để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiều của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh và/hoặc phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiều của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm. Hơn nữa, bộ giải mã bao gồm bộ xử lý (120) để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề xuất thiết bị và phương pháp mã hóa đối tượng âm thanh không gian khái quát hóa sử dụng khái niệm tham số cho các trường hợp trộn giảm/trộn tăng.

Tình trạng kỹ thuật của sáng chế

Trong các hệ thống âm thanh kỹ thuật số hiện đại, việc cho phép thay đổi liên quan đến đối tượng âm thanh của nội dung được truyền trên phía máy thu đang là xu hướng chính. Những việc thay đổi này bao gồm những việc thay đổi tăng thêm của các phần được chọn của tín hiệu âm thanh và/hoặc sự tái định vị không gian của các đối tượng âm thanh dành riêng trong trường hợp bộ phận phát lại đa kênh thông qua các loa được phân bố theo không gian. Điều này có thể đạt được bằng cách phân phối riêng biệt các phần khác nhau của nội dung âm thanh đến các loa khác nhau.

Nói cách khác, trong kỹ thuật xử lý âm thanh, truyền âm thanh, và lưu trữ âm thanh, có mong muốn ngày càng tăng để cho phép sự tương tác người dùng trên bộ phận phát lại nội dung âm thanh hướng đối tượng và cũng có nhu cầu ngày càng tăng để sử dụng các khả năng mở rộng của bộ phận phát lại đa kênh để biểu diễn riêng lẻ các nội dung âm thanh hoặc các phần của nó nhằm cải thiện cảm giác thính giác. Bằng cách này, việc sử dụng nội dung âm thanh đa kênh mang theo sự cải thiện đáng kể cho người sử dụng. Ví dụ, cảm giác thính giác ba chiều có thể đạt được, điều này mang lại sự hài lòng của người sử dụng được tăng lên trong các chương trình ứng dụng giải trí. Tuy nhiên, nội dung âm thanh đa kênh cũng hữu ích trong các môi trường chuyên nghiệp, ví dụ, trong các chương trình ứng dụng hội nghị qua điện thoại, bởi vì độ nghe rõ của người nói có thể được cải thiện bằng cách sử dụng bộ phận phát lại âm thanh đa kênh. Chương trình ứng dụng khả thi khác là để cung cấp đến người nghe bản nhạc để điều chỉnh riêng biệt mức độ phát lại và/hoặc vị trí không gian của các phần khác nhau (cũng được gọi là "các đối tượng âm thanh") hoặc các rãnh ghi âm, như phần xướng âm hoặc các nhạc cụ khác nhau. Người sử dụng có thể thực hiện như điều chỉnh vì các lý do sở thích cá nhân, để dễ dàng chuyển biên một hoặc nhiều phần từ các bản nhạc, các mục đích giáo dục, karaoke, diễn tập, v.v.

Sự truyền rời rạc thẳng của tất cả nội dung âm thanh đa kênh hoặc đa đối tượng kỹ thuật số, ví dụ, dưới dạng dữ liệu điều biến mã xung (pulse code modulation - PCM) hoặc thậm chí các định dạng âm thanh được nén, yêu cầu tốc độ truyền bit rất cao. Tuy nhiên, cũng có khả năng mong muốn truyền dẫn và lưu trữ dữ liệu âm thanh theo cách hiệu quả tốc độ truyền bit. Do đó, người ta sẵn sàng chấp nhận sự cân bằng hợp lý giữa chất lượng âm thanh và các yêu cầu tốc độ truyền bit để tránh lượng tải nguồn dư được gây ra bởi các chương trình ứng dụng đa kênh/đa đối tượng.

Gần đây, trong lĩnh vực mã hóa âm thanh, các kỹ thuật tham số để truyền dẫn/lưu trữ có hiệu quả tốc độ truyền bit của các tín hiệu âm thanh đa kênh/đa đối tượng đã được giới thiệu bởi, ví dụ, Nhóm chuyên gia về hình ảnh di động (the Moving Picture Experts Group - MPEG) và những người khác. Một ví dụ đó là MPEG Surround (MPS) như cách tiếp cận hướng kênh [MPS, BCC], hoặc Mã hóa đối tượng âm thanh không gian (Spatial Audio Object Coding - SAOC) MPEG như cách tiếp cận hướng đối tượng [JSC, SAOC, SAOC1, SAOC2]. Cách tiếp cận hướng đối tượng khác được gọi là "tách nguồn được thông tin" [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]. Những kỹ thuật này nhằm mục đích thiết lập lại lớp âm thanh đầu ra mong muốn hoặc đối tượng nguồn âm thanh mong muốn trên cơ sở của sự trộn giảm của các kênh/các đối tượng và thông tin phụ bổ sung mô tả cảnh âm thanh được truyền dẫn/được lưu trữ và/hoặc các đối tượng nguồn âm thanh trong cảnh âm thanh.

Sự đánh giá và ứng dụng của thông tin phụ liên quan đến kênh/đối tượng trong những hệ thống như vậy được hoàn thành theo cách có chọn lọc thời gian-tần số. Do đó, những hệ thống như vậy sử dụng các phép biến đổi như phép biến đổi Fourier rời rạc (Discrete Fourier Transform - DFT), phép biến đổi Fourier thời gian ngắn (Short Time Fourier Transform - STFT) hoặc các dàn lọc như dàn lọc gương cầu phương (Quadrature Mirror Filter - QMF), v.v. Nguyên tắc cơ bản của các hệ thống trên được mô tả trên Fig.2, sử dụng ví dụ của MPEG SAOC.

Trong trường hợp STFT, kích thước tạm thời được thể hiện bằng số khối thời gian và kích thước phổ được thu bởi số hệ số quang phổ ("bộ nhớ"). Trong trường hợp QMF, kích thước tạm thời được thể hiện bằng số khe thời gian và kích thước phổ được thu bởi số dải băng con ("bộ nhớ"). Nếu độ phân giải quang phổ của QMF được cải

thiện bởi chương trình ứng dụng tiếp theo của bước lọc thứ hai, toàn bộ dải lọc được gọi là QMF lai và các dải băng con độ phân giải tốt được gọi là các dải băng con lai.

Như được đề cập ở trên, trong SAOC, sự xử lý chung được thực hiện theo cách có chọn lọc thời gian-tần suất và có thể được mô tả như sau trong mỗi dải băng tần suất, như được thể hiện trên Fig.2:

- Các tín hiệu đối tượng âm thanh đầu vào N $s_1 \dots s_N$ được trộn giảm xuống các kênh P $x_1 \dots x_P$ như phần của sự xử lý mã hóa sử dụng ma trận trộn giảm chứa các thành phần $d_{1,1} \dots d_{N,P}$. Ngoài ra, bộ mã hóa tách các thông tin phụ mô tả các đặc điểm của các đối tượng âm thanh đầu vào (môđun ước lượng thông tin phụ (side-information-estimator - SIE)). Với MPEG, SAOC, các mối quan hệ của các năng lượng đối tượng đối với nhau là mẫu cơ bản nhất của thông tin phụ như vậy.

- (Các) tín hiệu trộn giảm và thông tin phụ được truyền dẫn/lưu trữ. Vì mục đích này, (các) tín hiệu âm thanh trộn giảm có thể được nén, ví dụ, sử dụng các bộ mã hóa âm thanh cảm nhận nổi tiếng như MPEG-1/2 Lớp II hoặc III (còn được biết đến là .mp3), MPEG-2/4 Mã hóa âm thanh tiên tiến (Advanced Audio Coding - AAC) v.v.

- Trên đầu nhận, bộ giải mã cố gắng một cách có khái niệm phục hồi các tín hiệu đối tượng gốc (“việc tách đối tượng”) từ các tín hiệu trộn giảm (được giải mã) sử dụng thông tin phụ được truyền dẫn. Các tín hiệu đối tượng dự đoán $\hat{s}_1 \dots \hat{s}_N$ sau đó được trộn vào cảnh đích được thể hiện bằng các kênh đầu ra âm thanh M $\hat{y}_1 \dots \hat{y}_M$ sử dụng ma trận biểu diễn được mô tả bằng các hệ số $r_{1,1} \dots r_{N,M}$ trên Fig.2. Cảnh đích mong muốn có thể, trong trường hợp cực hạn, là việc biểu diễn của một tín hiệu nguồn duy nhất trong số hỗn hợp (viễn cảnh việc tách nguồn), mà còn bất kỳ lớp âm thanh bất kỳ khác chứa các đối tượng đã truyền dẫn. Ví dụ, đầu ra có thể là kênh đơn, âm lập thể 2 kênh, hoặc cảnh đích đa kênh 5.1.

Việc tăng các cải thiện đang có sẵn và tiếp tục dải thông/bộ lưu trữ trong lĩnh vực mã hóa âm thanh cho phép người dùng chọn từ các lựa chọn đăng tăng lên một cách ổn định các sản phẩm âm thanh đa kênh. Các định dạng âm thanh 5.1 đa kênh đã là chuẩn mực trong đầu DVD và các sản phẩm Blu-Ray. Các định dạng âm thanh mới như Âm thanh 3D MPEG-H với thậm chí nhiều các kênh vận chuyển âm thanh hơn

xuất hiện trong phạm vi hiểu biết, điều này sẽ cung cấp cho người dùng cuối trải nghiệm âm thanh chìm đắm cao.

Các chương trình mã hóa đối tượng âm thanh tham số hiện đang bị giới hạn đến tối đa hai kênh trộn giảm. Chúng chỉ có thể được áp dụng ở một mức độ nào đó trên các hỗn hợp đa kênh, ví dụ, trên chỉ hai kênh trộn giảm được chọn. Tính linh hoạt các chương trình mã hóa này cung cấp cho người dùng để điều chỉnh cảnh âm thanh theo sở thích của riêng mình do đó bị giới hạn một cách khắt khe, v.v., đối với việc thay đổi mức âm thanh của bình luận viên thể thao và bầu không khí khi phát sóng chương trình thể thao.

Hơn nữa, các chương trình mã hóa đối tượng âm thanh hiện nay chỉ cung cấp sự biến thiên hạn chế trong quy trình trộn tại phía bộ mã hóa. Quy trình trộn được giới hạn trong việc trộn biến thể thời gian của các đối tượng âm thanh, và việc trộn biến thể tần suất là bất khả thi.

Do đó, nó sẽ được đánh giá cao nếu các khái niệm được cải thiện cho việc mã hóa đối tượng âm thanh được cung cấp.

Tài liệu tham khảo

[MPS] ISO/IEC 23003-1:2007, MPEG-D (MPEG audio technologies), Part 1: MPEG Surround, 2007.

[BCC] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003

[JSC] C. Faller, "Parametric Joint-Coding of Audio Sources", 120th AES Convention, Paris, 2006

[SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC - Recent Developments in Parametric Coding of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, UK, April 2007

[SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers and W. Oomen: " Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008

[SAOC] ISO/IEC, “MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC),” ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2.

[ISS1] M. Parvaix and L. Girin: “Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding”, IEEE ICASSP, 2010

[ISS2] M. Parvaix, L. Girin, J.-M. Brossier: “A watermarking-based method for informed source separation of audio signals with a single sensor”, IEEE Transactions on Audio, Speech and Language Processing, 2010

[ISS3] A. Liutkus and J. Pinel and R. Badeau and L. Girin and G. Richard: “Informed source separation through spectrogram coding and data embedding”, Signal Processing Journal, 2011

[ISS4] A. Ozerov, A. Liutkus, R. Badeau, G. Richard: “Informed source separation: source coding meets source separation”, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011

[ISS5] Shuhua Zhang and Laurent Girin: “An Informed Source Separation System for Speech Signals”, INTERSPEECH, 2011

[ISS6] L. Girin and J. Pinel: “Informed Audio Source Separation from Compressed Linear Stereo Mixtures”, AES 42nd International Conference: Semantic Audio, 2011

Bản chất kỹ thuật của sáng chế

Theo phương án của sáng chế, bộ giải mã để tạo ra tín hiệu đầu ra âm thanh có một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm có một hoặc nhiều kênh trộn giảm, trong đó tín hiệu trộn giảm mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh có thể có: bộ xác định ngưỡng để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh

trộn giảm, và bộ xử lý để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.

Theo phương án khác của sáng chế, phương pháp tạo ra tín hiệu đầu ra âm thanh có một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm có một hoặc nhiều kênh trộn giảm, trong đó tín hiệu trộn giảm mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh có thể có các bước bao gồm: xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiều âm của ít nhất một tín hiệu trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiều âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, và tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.

Phương pháp khác có thể có chương trình máy tính thực thi phương pháp theo điểm 13 khi chạy trên máy tính hoặc bộ xử lý tín hiệu.

Bộ giải mã tạo ra tín hiệu đầu ra âm bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm được đề xuất. Tín hiệu trộn giảm mã hóa một hoặc nhiều tín hiệu đối tượng âm thanh. Bộ giải mã bao gồm bộ xác định ngưỡng để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiều âm của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh và/hoặc phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiều âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm. Hơn nữa, bộ giải mã bao gồm bộ xử lý để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.

Theo phương án, tín hiệu trộn giảm có thể bao gồm hai hoặc nhiều hơn hai kênh trộn giảm, và bộ xác định ngưỡng có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng nhiều âm của mỗi kênh trong số hai hoặc nhiều hơn hai kênh trộn giảm.

Trong phương án, bộ xác định ngưỡng có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào tổng tất cả năng lượng nhiều âm trong hai hoặc nhiều hơn hai kênh trộn giảm.

Theo phương án, tín hiệu trộn giảm có thể mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh, và bộ xác định ngưỡng có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu của tín hiệu đối tượng âm thanh trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh mà có năng lượng tín hiệu lớn nhất trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh.

Theo phương án, tín hiệu trộn giảm có thể bao gồm hai hoặc nhiều hơn hai kênh trộn giảm, và bộ xác định ngưỡng có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm.

Theo phương án, tín hiệu trộn giảm có thể mã hóa một hoặc nhiều tín hiệu trộn giảm đối tượng âm thanh cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số. Bộ xác định ngưỡng có thể được tạo cấu hình để xác định giá trị ngưỡng cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, trong đó giá trị ngưỡng thứ nhất của ô thời gian - tần số thứ nhất trong số nhiều ô thời gian - tần số. Bộ xử lý có thể được tạo cấu hình để tạo ra cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số giá trị kênh của mỗi kênh trong số một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng của ô thời gian - tần số đã nêu.

Trong phương án, bộ giải mã có thể được tạo cấu hình để xác định giá trị ngưỡng T theo đêxiben theo công thức

$$T[dB] = E_{noise}[dB] - E_{ref}[dB] - Z \text{ hoặc theo công thức}$$

$$T[dB] = E_{noise}[dB] - E_{ref}[dB],$$

trong đó $T[dB]$ biểu thị giá trị ngưỡng theo đêxiben, trong đó $E_{noise}[dB]$ biểu thị tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo đêxiben, trong đó $E_{ref}[dB]$ biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối

tượng âm thanh, và trong đó Z biểu thị tham số bổ sung là số. Trong phương án thay thế, $E_{noise}[dB]$ biểu thị tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo decibel được chia cho số lượng của hai hoặc nhiều hơn hai kênh trộn giảm.

Theo phương án, bộ giải mã có thể được tạo cấu hình để xác định giá trị ngưỡng T theo công thức

$$T = \frac{E_{noise}}{E_{ref} \cdot Z} \text{ hoặc theo công thức}$$

$$T = \frac{E_{noise}}{E_{ref}},$$

trong đó T biểu thị giá trị ngưỡng, trong đó E_{noise} biểu thị tổng tất cả các năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm, trong đó E_{ref} biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối tượng âm thanh, và trong đó Z biểu thị tham số bổ sung là số. Trong phương án thay thế, E_{noise} theo decibel biểu thị tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo decibel được chia cho số lượng của hai hoặc nhiều hơn hai kênh trộn giảm.

Theo phương án, bộ xử lý có thể được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào ma trận hiệp phương sai đối tượng (E) của một hoặc nhiều tín hiệu đối tượng âm thanh, phụ thuộc vào ma trận trộn giảm (D) để trộn giảm hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được hai hoặc nhiều hơn hai kênh trộn giảm, và phụ thuộc vào giá trị ngưỡng.

Trong phương án, bộ xử lý được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm từ một hoặc nhiều kênh trộn giảm bằng cách áp dụng giá trị ngưỡng trong hàm để nghịch đảo ma trận tương quan chéo kênh trộn giảm Q, trong đó Q được định nghĩa là $Q = DED^*$, trong đó D là ma trận trộn giảm để trộn giảm hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được hai hoặc nhiều hơn hai kênh trộn giảm, và trong đó E là ma trận hiệp phương sai đối tượng của một hoặc nhiều tín hiệu đối tượng âm thanh.

Ví dụ, bộ xử lý có thể được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách tính toán trị riêng của ma trận tương quan chéo kênh trộn giảm Q hoặc bằng cách tính toán giá trị đơn của ma trận tương quan chéo kênh trộn giảm Q .

Ví dụ, bộ xử lý có thể được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách nhân trị riêng lớn nhất trong số các trị riêng của ma trận tương quan chéo kênh trộn giảm Q với giá trị ngưỡng để thu được ngưỡng tương ứng.

Ví dụ, bộ xử lý có thể được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách tạo ra ma trận được cải biên. Bộ xử lý có thể được tạo cấu hình để tạo ra ma trận được cải biên chỉ phụ thuộc vào các trị riêng này của ma trận tương quan chéo kênh trộn giảm Q , mà có trị riêng trong số các trị riêng của ma trận tương quan chéo kênh trộn giảm Q , mà lớn hơn hoặc bằng ngưỡng được cải biên. Hơn nữa, bộ xử lý có thể được tạo cấu hình để thực hiện phép nghịch đảo ma trận của ma trận được cải biên để thu được ma trận được nghịch đảo. Hơn nữa, bộ xử lý có thể được tạo cấu hình để áp dụng ma trận được nghịch đảo trên một hoặc nhiều kênh trộn giảm để tạo ra một hoặc nhiều kênh đầu ra âm thanh.

Hơn nữa, phương pháp tạo ra tín hiệu đầu ra âm bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm được đề xuất. Tín hiệu trộn giảm mã hóa một hoặc nhiều tín hiệu đối tượng âm thanh. Bộ giải mã bao gồm:

Xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm. Và:

Tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.

Hơn nữa, chương trình máy tính để thực hiện phương pháp được mô tả bên trên khi được chạy trên máy tính hoặc bộ xử lý tín hiệu được đề xuất.

Mô tả vắn tắt các hình vẽ.

Dưới đây, các phương án của sáng chế được mô tả chi tiết hơn với sự viện dẫn đến các hình, trong đó:

Fig.1 minh họa bộ giải mã để tạo ra tín hiệu đầu ra âm thanh bao gồm một hoặc nhiều kênh đầu ra âm thanh theo phương án,

Fig.2 là tổng quan hệ thống SAOC mô tả nguyên lý của các hệ thống như vậy sử dụng ví dụ MPEG SAOC,

Fig.3 minh họa tổng quan của khái niệm trộn tăng tham số G-SAOC, và

Fig.4 minh họa khái niệm trộn giảm/trộn tăng chung.

Mô tả chi tiết sáng chế

Trước khi mô tả các phương án của sáng chế, các tình trạng kỹ thuật của các hệ thống SAOC được đề xuất.

Fig.2 cho thấy cách sắp xếp khái quát bộ mã hóa SAOC 10 và bộ giải mã SAOC 12. Bộ mã hóa SAOC 10 nhận như các đối tượng N đầu vào, tức là, các tín hiệu âm thanh từ s_I đến s_N . Cụ thể, bộ mã hóa 10 bao gồm bộ trộn giảm 16 mà nhận các tín hiệu âm thanh từ s_I đến s_N và trộn giảm các tín hiệu đó thành tín hiệu trộn giảm 18. Ngoài ra, sự trộn giảm có thể được cung cấp bên ngoài ("trộn giảm nghệ thuật") và hệ thống ước tính thông tin phụ bổ sung để khiến sự trộn giảm được cung cấp phù hợp với sự trộn giảm được tính toán. Trong Fig.2, tín hiệu trộn giảm được thể hiện là tín hiệu kênh P. Do đó, bất kỳ cấu hình tín hiệu trộn giảm kênh đơn ($P=1$), kênh đôi ($P=2$) hoặc đa kênh ($P>2$) có thể nhận thức được.

Trong trường hợp trộn giảm đôi, các kênh của tín hiệu trộn giảm 18 được ký hiệu là $L0$ và $R0$, trong trường hợp trộn giảm đơn tương tự cũng chỉ ký hiệu đơn giản là $L0$. Để cho phép bộ giải mã SAOC 12 có thể phục hồi các đối tượng riêng lẻ từ s_I đến s_N , bộ ước tính thông tin phụ 17 cung cấp bộ giải mã SAOC 12 với thông tin phụ bao gồm các tham số SAOC. Ví dụ, trong trường hợp trộn giảm đôi, các tham số SAOC bao gồm các sự chênh lệch mức đối tượng (object level differences - OLD), sự tương quan đối tượng liên đới (inter-object correlations - IOC) (các tham số tương quan chéo đối tượng liên đới), giá trị độ lợi trộn giảm (downmix gain values - DMG) và các sự chênh

lệch mức kênh trộn giảm (downmix channel level differences - DCLD). Thông tin phụ 20, bao gồm các tham số SAOC, cùng với tín hiệu trộn giảm 18, tạo ra dòng dữ liệu đầu ra SAOC thu được bởi bộ giải mã SAOC 12.

Bộ giải mã SAOC 12 bao gồm bộ trộn tăng mà thu tín hiệu trộn giảm 18 cũng như thông tin phụ 20 để phục hồi và biểu diễn các tín hiệu âm thanh s_I và s_N lên trên bất kỳ bộ các kênh được người dùng lựa chọn từ y_I đến y_M , với việc biểu diễn được quy định bởi đầu vào thông tin biểu diễn 26 vào trong bộ giải mã SAOC 12.

Các tín hiệu âm thanh từ s_I đến s_N có thể được đưa vào bộ mã hóa 10 trong bất kỳ miền mã hóa nào, ví dụ, trong miền thời gian hoặc quang phổ. Trong trường hợp các tín hiệu âm thanh từ s_I đến s_N được cấp vào trong bộ giải mã 10 trong miền thời gian, như PCM được mã hóa, bộ mã hóa 10 có thể sử dụng dàn lọc, ví dụ dàn QMF lai, để truyền dẫn các tín hiệu vào trong miền quang phổ, trong đó các tín hiệu âm thanh được thể hiện trong một vài dải băng con với các phần quang phổ khác nhau, tại độ phân giải dàn lọc cụ thể. Nếu các tín hiệu âm thanh từ s_I đến s_N đã sẵn sàng trong sự thể hiện được trông đợi bởi bộ mã hóa 10, tín hiệu âm thanh được nhắc đến không cần phải biểu diễn sự khai triển quang phổ.

Tính linh hoạt hơn trong quy trình trộn cho phép sự khai thác tối ưu nhất các đặc tính đối tượng tín hiệu. Sự trộn giảm có thể được sản xuất mà tối ưu hóa cho sự tách tham số tại phía bộ giải mã đối với chất lượng nhận biết.

Các phương án mở rộng phần tham số của chương trình SAOC đến số bất kỳ của các kênh trộn giảm/trộn tăng. Hình vẽ dưới đây cung cấp khái niệm tổng quan của khái niệm trộn tăng tham số Mã hóa đối tượng âm thanh không gian khái quát hóa (Generalized Spatial Audio Object Coding - G-SAOC):

Fig.3 minh họa tổng quan của khái niệm trộn tăng tham số G-SAOC. Sự trộn sau (biểu diễn) linh hoạt hoàn toàn của các đối tượng âm thanh được cấu trúc lại theo tham số có thể được nhận ra.

Không kể những cái khác, Fig.3 minh họa bộ giải mã âm thanh 310, bộ tách đối tượng 320 và bộ biểu diễn 330.

Hãy cùng xem xét các ký hiệu chung sau:

x	- tín hiệu đối tượng âm thanh đầu vào	(kích thước N_{obj})
y	- tín hiệu âm thanh trộn giảm	(kích thước N_{dmx})
z	- tín hiệu lớp đầu ra được biểu diễn	(kích thước N_{upmix})
D	- ma trận trộn giảm	(kích thước $N_{obj} \times N_{dmx}$)
R	- ma trận biểu diễn	(kích thước $N_{obj} \times N_{upmix}$)
G	- ma trận trộn tăng tham số	(kích thước $N_{dmx} \times N_{upmix}$)
E	- ma trận phương sai đối tượng	(kích thước $N_{obj} \times N_{obj}$)

Tất cả các ma trận được giới thiệu (nhìn chung) là biến thể thời gian và tần suất.

Dưới đây, mối quan hệ cơ bản cho việc trộn tăng tham số được đề xuất.

Đầu tiên, các khái niệm trộn giảm/trộn tăng khái quát được đề xuất với sự viện dẫn đến Fig.4. Cụ thể, Fig.4 minh họa khái niệm trộn giảm/trộn tăng khái quát, trong đó Fig.4 minh họa các hệ thống mẫu (bên trái) và trộn tăng tham số (bên phải).

Cụ thể hơn, Fig.4 minh họa bộ biểu diễn 410, bộ trộn giảm 421 và bộ trộn tăng tham số 422.

Tín hiệu cảnh đầu ra được biểu diễn lý tưởng (mẫu) z được định nghĩa như, xem Fig (bên trái):

$$Rx = z \quad (1)$$

Tín hiệu âm thanh trộn giảm y được định nghĩa như, xem Fig.4 (bên phải):

$$Dx = y \quad (2)$$

Mối quan hệ cơ bản (được áp dụng cho tín hiệu âm thanh trộn giảm) cho sự cấu trúc lại tín hiệu lớp đầu ra tham số có thể được thể hiện như, xem Fig.4 (bên phải):

$$Gy = z \quad (3)$$

Ma trận trộn tăng tham số có thể được định nghĩa từ (1) và (2) như chức năng dưới đây của các ma trận trộn giảm và biểu diễn $G = G(D, R)$:

$$\mathbf{G} = \mathbf{RED}^* (\mathbf{DED}^*)^{-1} \quad (4)$$

Dưới đây, việc cải thiện tính ổn định của việc ước tính nguồn tham số dựa theo các phương án được xem xét.

Chương trình tách tham số trong MPEG SAOC được dựa trên sự ước tính bình phương trung bình tối thiểu (Least Mean Square - LMS) của các nguồn của hỗn hợp. Sự ước tính LMS bao gồm sự nghịch đảo của ma trận phương sai kênh trộn giảm được mô tả theo tham số $\mathbf{Q}=\mathbf{DED}^*$. Các thuật toán cho sự nghịch đảo ma trận nhìn chung nhạy với các ma trận có điều kiện xấu. Sự nghịch đảo của ma trận như vậy có thể dẫn tới âm thanh không tự nhiên, được gọi là thành phần lạ, trong cảnh đầu ra được biểu diễn. Ngưỡng cố định được định theo cách tự tìm ra $\mathbf{T}$ trong MPEG SAOC gần đây tránh được điều này. Mặc dù các thành phần lạ được tránh nhờ phương pháp này, sự biểu diễn tách khả thi đầy đủ tại phía bộ giải mã có thể nhờ đó không đạt được.

Fig.1 minh họa bộ giải mã để tạo ra tín hiệu đầu ra âm thanh bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm theo phương án. Tín hiệu trộn giảm mã hóa một hoặc nhiều tín hiệu đối tượng âm thanh.

Bộ mã hóa bao gồm bộ xác định ngưỡng 110 để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiễu của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh và/hoặc phụ thuộc vào năng lượng tín hiệu và/hoặc năng lượng nhiễu của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm.

Hơn nữa, bộ giải mã bao gồm bộ xử lý 120 để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng.

Đổi lập với trạng thái của kỹ thuật, giá trị ngưỡng được xác định bởi bộ xác định giá trị ngưỡng 110 phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu của một hoặc nhiều kênh trộn giảm hoặc của một hoặc nhiều các tín hiệu đối tượng âm thanh đã được giải mã. Trong các phương án, vì các năng lượng tín hiệu và nhiễu của một hoặc nhiều kênh trộn giảm và/hoặc của một hoặc nhiều các giá trị tín hiệu đối tượng

âm thanh biến đổi, vậy nên biến đổi giá trị ngưỡng, ví dụ, từ đối tượng thời gian đến đối tượng thời gian, hoặc từ ô thời gian-tần suất đến ô thời gian-tần suất.

Các phương án đề xuất phương pháp ngưỡng có khả năng thích nghi cho sự nghịch đảo ma trận để đạt được sự tách tham số được cải thiện của các đối tượng âm thanh tại phía bộ giải mã. Sự biểu diễn tách trung bình là tốt nhưng không bao giờ ít chương trình ngưỡng cố định được sử dụng hiện nay được dùng trong MPEG SAOC trong thuật toán để nghịch đảo ma trận $\mathbf{Q}$.

Ngưỡng $\mathbf{T}$ được thích ứng động với sự chính xác của dữ liệu cho mỗi ô thời gian-tần suất được xử lý. Sự thực hiện tách do đó được cải thiện và các thành phần lạ trong lớp đầu ra được biểu diễn được gây ra bởi sự nghịch đảo của các ma trận tình trạng kém được tránh.

Theo phương án, tín hiệu trộn giảm có thể bao gồm hai hoặc nhiều kênh trộn giảm, và bộ xác định ngưỡng 110 có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng nhiều của mỗi hai hay nhiều kênh trộn giảm.

Trong phương án, bộ xác định ngưỡng 110 có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào tổng của tất cả năng lượng nhiều trong hai hoặc nhiều kênh trộn giảm.

Theo phương án, tín hiệu trộn giảm có thể mã hóa hai hay nhiều tín hiệu đối tượng âm thanh, và bộ xác định ngưỡng 110 có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu của tín hiệu đối tượng âm thanh của hai hay nhiều tín hiệu đối tượng âm thanh mà có năng lượng tín hiệu lớn nhất của hai hay nhiều tín hiệu đối tượng âm thanh.

Trong phương án, tín hiệu trộn giảm có thể bao gồm hai hoặc nhiều kênh trộn giảm, và bộ xác định ngưỡng 110 có thể được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào tổng tất cả năng lượng nhiều trong hai hay nhiều kênh trộn giảm.

Theo phương án, tín hiệu trộn giảm có thể mã hóa một hoặc nhiều tín hiệu đối tượng âm thanh để mỗi ô thời gian-tần suất của nhiều ô thời gian-tần suất. Bộ xác định ngưỡng 110 có thể được tạo cấu hình để xác định giá trị ngưỡng cho mỗi ô thời gian-tần suất của nhiều ô thời gian-tần suất phụ thuộc vào năng lượng tín hiệu hoặc năng

lượng nhiễu của ít nhất một tín hiệu trong số một hoặc nhiều tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, trong đó giá trị ngưỡng thứ nhất của ô thời gian-tần suất thứ nhất của nhiều ô thời gian-tần suất có thể khác với ô thời gian-tần suất thứ hai của nhiều ô thời gian-tần suất. Bộ xử lý 120 có thể được tạo cấu hình để tạo ra cho mỗi ô thời gian-tần suất của nhiều ô thời gian-tần suất giá trị kênh của mỗi một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng của ô thời gian-tần suất nêu trên.

Theo phương án, bộ giải mã có thể được tạo cấu hình để xác định giá trị ngưỡng T dựa theo công thức

$$T = \frac{E_{noise}}{E_{ref} \cdot Z} \quad \text{hoặc dựa theo công thức}$$

$$T = \frac{E_{noise}}{E_{ref}},$$

trong đó T biểu thị giá trị ngưỡng, trong đó E_{noise} biểu thị tổng của tất cả năng lượng nhiễu trong hai hay nhiều kênh trộn giảm, trong đó E_{ref} biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối tượng, và trong đó Z biểu thị tham số bổ sung là số. Trong phương án thay thế, E_{noise} biểu thị tổng của tất cả năng lượng nhiễu trong hai hay nhiều kênh trộn giảm được chia bởi số các kênh trộn giảm.

Trong phương án, bộ giải mã có thể được tạo cấu hình để xác định giá trị ngưỡng T theo deciben dựa theo công thức

$$T[dB] = E_{noise}[dB] - E_{ref}[dB] - Z \quad \text{hoặc dựa theo công thức}$$

$$T[dB] = E_{noise}[dB] - E_{ref}[dB],$$

trong đó $T[dB]$ biểu thị giá trị ngưỡng theo deciben, trong đó $E_{noise}[dB]$ biểu thị tổng của tất cả năng lượng nhiễu trong hai hay nhiều kênh trộn giảm theo deciben, trong đó $E_{ref}[dB]$ biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối tượng

theo đexiben, và trong đó Z biểu thị tham số bổ sung là số. Trong phương án thay thế, $E_{noise}[dB]$ biểu thị tổng của tất cả năng lượng nhiễu trong hai hay nhiều kênh trộn giảm theo đexiben được chia bởi số các kênh trộn giảm.

Cụ thể, việc ước tính sơ bộ của ngưỡng có thể được đưa cho mỗi ô thời gian-tần suất bằng:

$$T[dB] = E_{noise}[dB] - E_{ref}[dB] - Z \quad (5)$$

E_{noise} có thể biểu thị mức sàn nhiễu, ví dụ, tổng của tất cả năng lượng ồn trong các kênh trộn giảm. Sàn nhiễu có thể được định nghĩa bởi độ phân giải của dữ liệu âm thanh, ví dụ, sàn nhiễu được gây ra bởi mã hóa PCM của các kênh. Khả năng khác là để giải thích việc mã hóa nhiễu nếu sự trộn giảm được nén lại. Với trường hợp như vậy, sàn nhiễu được gây ra bởi thuật toán mã hóa có thể được thêm vào, Trong phương án thay thế, $E_{noise}[dB]$ biểu thị tổng của tất cả năng lượng nhiễu trong hai hay nhiều kênh trộn giảm theo đexiben được chia bởi số các kênh trộn giảm.

E_{ref} có thể biểu thị năng lượng tín hiệu tham chiếu. Trong mẫu đơn giản nhất, điều này có thể là năng lượng của đối tượng âm thanh mạnh nhất:

$$E_{ref} = \max(E) \quad (6)$$

Z có thể biểu thị yếu tố phạt để đương đầu cho các tham số bổ sung mà ảnh hưởng đến độ phân giải tách, ví dụ, sự khác nhau của số các kênh trộn giảm và số các đối tượng nguồn. Sự thực hiện tách giảm với số đối tượng âm thanh giảm. Hơn nữa, ảnh hưởng của sự lượng tử hóa của thông tin phụ tham số trên sự tách có thể cũng được bao gồm.

Trong phương án, bộ xử lý 120 được tạo cấu hình để tạo ra một hoặc nhiều các kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào ma trận phương sai đối tượng E của một hoặc nhiều tín hiệu đối tượng âm thanh, phụ thuộc vào ma trận trộn giảm D để trộn giảm hai hoặc nhiều tín hiệu đối tượng âm thanh để thu được hai hay nhiều kênh trộn giảm, và phụ thuộc vào giá trị ngưỡng.

Theo phương án, để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng, bộ xử lý 120 có thể được tạo cấu hình để bắt đầu như sau:

Ngưỡng (mà có thể được đề cập đến như "ngưỡng tách-phân giải") được áp dụng tại phía bộ giải mã trong chức năng để nghịch đảo ma trận tương quan chéo kênh trộn giảm được ước tính tham số Q .

Các giá trị suy biến của Q hoặc các giá trị riêng của Q được tính toán.

Giá trị riêng lớn nhất được lấy và nhân với ngưỡng T .

Tất cả ngoại trừ giá trị riêng lớn nhất được so sánh với ngưỡng liên quan và được loại bỏ nếu chúng nhỏ hơn.

Sự nghịch đảo ma trận sau đó được thực hiện trên ma trận được thay đổi, trong đó ma trận được thay đổi có thể, ví dụ, là ma trận được định nghĩa bởi bộ được giảm các véc-tơ. Nên lưu ý rằng với trường hợp mà tất cả ngoại trừ giá trị riêng lớn nhất được loại bỏ, giá trị riêng lớn nhất nên được đặt ở mức sàn nhiều nếu giá trị riêng thấp.

Ví dụ, bộ xử lý 120 có thể được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách tạo ra ma trận đã được thay đổi. Ma trận được thay đổi có thể được tạo ra phụ thuộc vào chỉ những véc-tơ riêng của ma trận tương quan chéo kênh trộn giảm Q , mà có giá trị riêng của các giá trị riêng của ma trận tương quan chéo kênh trộn giảm Q , mà lớn hơn hoặc bằng ngưỡng được thay đổi. Bộ xử lý 120 có thể được tạo cấu hình để thực hiện sự nghịch đảo ma trận của ma trận đã được thay đổi để thu được ma trận đã nghịch đảo. Sau đó, bộ xử lý 120 có thể được tạo cấu hình để áp dụng ma trận đã nghịch đảo trên một hoặc nhiều kênh trộn giảm để tạo ra một hoặc nhiều kênh đầu ra âm thanh. Ví dụ, ma trận được nghịch đảo có thể được áp dụng trên một hoặc nhiều kênh trộn giảm theo một trong các cách như ma trận nghịch đảo của sản phẩm ma trận **DED*** được áp dụng trên các kênh trộn giảm (xem, ví dụ. [SAOC], xem, cụ thể, ví dụ: [SAOC] ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2:2010, cụ thể, xem,

chương “SAOC Processing”, cụ thể hơn, xem mục “Transcoding modes” và mục “Decoding modes”).

Các tham số mà có thể được dùng để ước tính ngưỡng **T** có thể được xác định tại bộ mã hóa và được nhúng trong thông tin phụ tham số hoặc được ước tính trực tiếp tại phía bên bộ giải mã.

Phiên bản được rút gọn của bộ ước tính ngưỡng có thể được sử dụng tại phía bên bộ mã hóa để chỉ ra các khả năng bất ổn định tiềm tàng trong sự ước tính nguồn tại phía bộ giải mã. Trong mẫu đơn giản nhất, bỏ qua tất cả các giới hạn nhiễu, quy tắc của ma trận trộn giảm có thể được tính chỉ ra rằng tiềm năng đầy đủ của các kênh trộn giảm có sẵn để ước tính tham số các tín hiệu nguồn tại phía bộ giải mã không thể được khai thác. Bộ chỉ báo như vậy có thể được dùng trong suốt quy trình trộn để tránh các ma trận trộn mà tới hạn để ước tính các tín hiệu nguồn.

Đối với sự tham số hóa của ma trận phương sai đối tượng, một người có thể nhìn thấy phương pháp trộn tăng tham số được mô tả dựa trên mối quan hệ cơ bản (4) là bất biến đối với ký hiệu của sự tồn tại ngoài đường chéo của ma trận tương quan đối tượng **E**. Các kết quả này trong khả năng của sự tham số hóa (sự lượng tử hóa và mã hóa) hiệu quả (khi so sánh với SAOC) của các giá trị thể hiện các mối tương liên đối tượng liên đới.

Liên quan đến việc vận chuyển thông tin thể hiện ma trận trộn giảm, các tín hiệu đầu vào và trộn giảm âm thanh x , y cùng với ma trận phương sai **E** được xác định tại phía bộ mã hóa. Sự thể hiện được mã hóa của tín hiệu trộn giảm âm thanh y và thông tin mô tả ma trận phương sai **E** được truyền dẫn đến phía bên bộ mã hóa (qua đoạn mã dòng bit). Ma trận biểu diễn **R** được đặt và có sẵn tại phía bộ giải mã.

Thông tin thể hiện ma trận trộn giảm **D** (được áp dụng tại bộ mã hóa và được sử dụng như bộ giải mã) có thể được xác định (tại bộ mã hóa) và thu được (tại bộ giải mã) sử dụng các phương pháp nguyên lý sau.

Ma trận trộn giảm **D** có thể:

- được đặt và áp dụng (tại bộ mã hóa) và việc chỉ định sự thể hiện được lượng tử hóa và mã hóa của nó được truyền (đến bộ giải mã) thông qua đoạn mã dòng bit.

- được gán giá trị và áp dụng (tại bộ mã hóa) và được phục hồi (tại bộ giải mã) sử dụng các bảng tìm kiếm được lưu trữ (tức là, bộ các ma trận trộn giảm được xác định trước).
- được gán giá trị và áp dụng (tại bộ mã hóa) và được áp dụng (tại bộ giải mã) dựa theo thuật toán hoặc phương pháp đặc biệt (ví dụ, vị trí cách đều nhau có trọng số và được sắp xếp đặc biệt của các đối tượng âm thanh đến các kênh trộn giảm có sẵn).
- được ước tính và áp dụng (tại bộ mã hóa) và được phục hồi (tại bộ giải mã) sử dụng tiêu chuẩn tối ưu hóa cụ thể cho phép "trộn linh hoạt" của các đối tượng âm thanh đầu vào (tức là sự tạo thành ma trận trộn giảm mà được tối ưu hóa để ước tính tham số của các đối tượng âm thanh tại phía bộ giải mã). Ví dụ, bộ mã hóa tạo ra ma trận trộn giảm theo cách để khiến sự trộn tăng tham số hiệu quả hơn, về việc cấu trúc lại thuộc tính tín hiệu đặc biệt, giống như phương sai, mối tương quan tín hiệu liên đới hoặc cải thiện/đảm bảo sự ổn định thuộc số của thuật toán trộn tăng tham số.

Các phương án được đề xuất có thể áp dụng trên số bất kỳ của các kênh trộn giảm/trộn tăng. Nó có thể được kết hợp với bất kỳ định dạng âm thanh gần đây và trong tương lai.

Tính linh hoạt của phương pháp sáng tạo cho phép đường rẽ của các kênh không thay đổi để giảm sự phức tạp tính toán, giảm đoạn mã dòng bit/ lượng dữ liệu được giảm.

Bộ mã hóa âm thanh, phương pháp hoặc chương trình máy tính để mã hóa được đề xuất. Hơn nữa, bộ giải mã âm thanh, phương pháp hoặc chương trình máy tính để giải mã được đề xuất. Hơn nữa, tín hiệu được mã hóa được đề xuất.

Mặc dù vài khía cạnh đã được mô tả trong bối cảnh của thiết bị, rõ ràng rằng các khía cạnh này cũng thể hiện sự mô tả của phương pháp tương ứng, nơi mà 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 thể hiện sự mô tả của khối hoặc mục tin hoặc đặc điểm tương ứng của thiết bị tương ứng

Tín hiệu khai triển sáng tạo có thể được lưu trữ trên môi trường lưu trữ số hoặc có thể được truyền dẫn trên môi trường truyền dẫn như môi trường truyền dẫn không dây hoặc môi trường truyền dẫn có dây như mạng Internet.

Phụ thuộc vào các yêu cầu thực hiện nhất định, các phương án của sáng chế có thể được thực hiện trong phần cứng hoặc trong phần mềm. Sự thực hiện có thể được thực hiện việc sử dụng môi trường lưu trữ số, ví dụ đĩa mềm, DVD, CD, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ FLASH, có các tín hiệu điều khiển có thể đọc được bằng điện tử được lưu trữ trên đó, môi trường này 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.

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

Nhìn chung, các phương án của sáng chế có thể thực hiện như sản phẩm chương trình máy tính với mã chương trình, mã chương trình đang sử dụng để thực hiện một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Ví dụ, mã chương trình có thể được lưu trữ trên vật mang có thể đọc được máy.

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

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

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

Phương án nữa của các phương pháp sáng tạo là, do đó, dòng dữ liệu hoặc chuỗi các tín hiệu thể hiện chương trình máy tính để thực hiện một trong số các phương pháp được mô tả trong này. Ví dụ, dòng dữ liệu hoặc chuỗi các tín hiệu có thể được tạo cấu hình để được truyền dẫn qua kết nối truyền dữ liệu, ví dụ, qua mạng Internet.

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

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

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

Các phương án được mô tả trên đây chỉ đơn thuần minh họa cho các nguyên lý của sáng chế. Điều này được hiểu rằng các sự cải biên và thay đổi của các chỉnh hợp và các chi tiết được mô tả trong đây sẽ rõ ràng với những người có hiểu biết trung bình trong lĩnh vực kỹ thuật. Do đó, nguyên lý được giới hạn chỉ bằng phạm vi của các yêu cầu bảo hộ của sáng chế và không phải bằng các chi tiết cụ thể được thể hiện bằng cách mô tả và giải thích của các phương án trong đây.

YÊU CẦU BẢO HỘ

1. Bộ giải mã tạo ra tín hiệu đầu ra âm thanh bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm, trong đó tín hiệu trộn giảm mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh, trong đó bộ giải mã bao gồm:

bộ xác định ngưỡng (110) để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, và

bộ xử lý (120) để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng,

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào ma trận hiệp phương sai đối tượng của một hoặc nhiều tín hiệu đối tượng âm thanh, phụ thuộc vào ma trận trộn giảm để trộn giảm hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được một hoặc nhiều kênh trộn giảm, và phụ thuộc vào giá trị ngưỡng,

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách áp dụng giá trị ngưỡng trong chức năng nghịch đảo ma trận tương quan chéo kênh trộn giảm Q ,

trong đó Q được định nghĩa là $Q=DED^*$,

trong đó D là ma trận trộn giảm để trộn giảm hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó E là ma trận hiệp phương sai đối tượng của một hoặc nhiều tín hiệu đối tượng âm thanh, và

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách tính toán các trị riêng của ma trận tương quan chéo kênh trộn giảm Q ;

trong đó mỗi trị riêng ngoại trừ trị riêng lớn nhất được so sánh với giá trị ngưỡng, và được loại bỏ nếu chúng nhỏ hơn.

2. Bộ giải mã theo điểm 1,

trong đó tín hiệu trộn giảm bao gồm hai hoặc nhiều hơn hai kênh trộn giảm, và

trong đó bộ xác định ngưỡng (110) được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng nhiễu âm của mỗi kênh trong số hai hoặc nhiều hơn hai kênh trộn giảm.

3. Bộ giải mã theo điểm 2, trong đó bộ xác định ngưỡng (110) được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm.

4. Bộ giải mã theo điểm 1, trong đó bộ xử lý (120) được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách nhân trị riêng lớn nhất trong số các trị riêng của ma trận tương quan chéo kênh trộn giảm **Q** với giá trị ngưỡng để thu được ngưỡng tương ứng.

5. Bộ giải mã theo điểm 4,

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm bằng cách tạo ra ma trận được cải biên,

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra ma trận được cải biên chỉ phụ thuộc vào các trị riêng của ma trận tương quan chéo kênh trộn giảm **Q**, mà bao gồm trị riêng trong số các trị riêng của ma trận tương quan chéo kênh trộn giảm **Q**, mà lớn hơn hoặc bằng ngưỡng tương ứng,

trong đó bộ xử lý (120) được tạo cấu hình để thực hiện phép nghịch đảo ma trận của ma trận được cải biên để thu được ma trận được nghịch đảo, và

trong đó bộ xử lý (120) được tạo cấu hình để áp dụng ma trận được nghịch đảo vào một hoặc nhiều kênh trộn giảm để tạo ra một hoặc nhiều kênh đầu ra âm thanh.

6. Bộ giải mã theo điểm 1,

trong đó bộ xác định ngưỡng (110) được tạo cấu hình để xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu của tín hiệu đối tượng âm thanh trong số hai hoặc

nhiều hơn hai tín hiệu đối tượng âm thanh mà bao gồm năng lượng tín hiệu lớn nhất trong số hai hoặc nhiều hơn hai tín hiệu đối tượng.

7. Bộ giải mã theo điểm 1,

trong đó tín hiệu trộn giảm mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số,

trong đó bộ xác định ngưỡng (110) được tạo cấu hình để xác định giá trị ngưỡng cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, và

trong đó bộ xử lý (120) được tạo cấu hình để tạo ra cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số giá trị kênh của mỗi kênh trong số một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng của ô thời gian - tần số nêu trên.

8. Bộ giải mã theo điểm 1,

trong đó tín hiệu trộn giảm bao gồm hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó bộ giải mã được tạo cấu hình để xác định giá trị ngưỡng T theo đêxiben theo công thức

$$T[dB] = E_{noise}[dB] - E_{ref}[dB] - Z \text{ hoặc theo công thức}$$

$$T[dB] = E_{noise}[dB] - E_{ref}[dB],$$

trong đó $T[dB]$ biểu thị giá trị ngưỡng theo đêxiben,

trong đó $E_{noise}[dB]$ biểu thị tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo đêxiben, hoặc $E_{noise}[dB]$ biểu thị tổng tất cả năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo đêxiben được chia cho số lượng của hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó $E_{ref}[dB]$ biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối tượng âm thanh, và

trong đó Z biểu thị tham số bổ sung là số.

9. Bộ giải mã theo điểm 1,

trong đó tín hiệu trộn giảm bao gồm hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó bộ giải mã được tạo cấu hình để xác định giá trị ngưỡng T theo công thức

$$T = \frac{E_{noise}}{E_{ref} \cdot Z} \text{ hoặc theo công thức}$$

$$T = \frac{E_{noise}}{E_{ref}},$$

trong đó T biểu thị giá trị ngưỡng,

trong đó E_{noise} biểu thị tổng tất cả các năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm, hoặc E_{noise} theo decibels biểu thị tổng tất cả các năng lượng nhiễu âm trong hai hoặc nhiều hơn hai kênh trộn giảm theo decibels được chia cho số lượng của hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó E_{ref} biểu thị năng lượng tín hiệu của một tín hiệu trong số các tín hiệu đối tượng âm thanh, và

trong đó Z biểu thị tham số bổ sung là số.

10. Phương pháp giải mã tạo ra tín hiệu đầu ra âm thanh bao gồm một hoặc nhiều kênh đầu ra âm thanh từ tín hiệu trộn giảm bao gồm một hoặc nhiều kênh trộn giảm, trong đó tín hiệu trộn giảm mã hóa hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh, trong đó phương pháp bao gồm:

xác định giá trị ngưỡng phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một tín hiệu trong số hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh hoặc phụ thuộc vào năng lượng tín hiệu hoặc năng lượng nhiễu âm của ít nhất một kênh trong số một hoặc nhiều kênh trộn giảm, và

tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào giá trị ngưỡng,

trong đó tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm phụ thuộc vào ma trận hiệp phương sai đối tượng (E) của một hoặc nhiều tín hiệu đối tượng âm thanh được dẫn ra, phụ thuộc vào ma trận trộn giảm để trộn giảm (D) hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được một hoặc nhiều kênh trộn giảm, và phụ thuộc vào giá trị ngưỡng,

trong tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm được dẫn ra bằng cách áp dụng giá trị ngưỡng trong chức năng nghịch đảo ma trận tương quan chéo kênh trộn giảm Q ,

trong đó Q được định nghĩa là $Q=DED^*$,

trong đó D là ma trận trộn giảm để trộn giảm hai hoặc nhiều hơn hai tín hiệu đối tượng âm thanh để thu được hai hoặc nhiều hơn hai kênh trộn giảm,

trong đó E là ma trận hiệp phương sai đối tượng của một hoặc nhiều tín hiệu đối tượng âm thanh, và

trong tạo ra một hoặc nhiều kênh đầu ra âm thanh từ một hoặc nhiều kênh trộn giảm được dẫn ra bằng cách tính toán các trị riêng của ma trận tương quan chéo kênh trộn giảm Q ;

trong đó mỗi trị riêng ngoại trừ trị riêng lớn nhất được so sánh với giá trị ngưỡng, và được loại bỏ nếu chúng nhỏ hơn.

11. Vật ghi lưu trữ kỹ thuật số không tạm thời bao gồm chương trình máy tính để thực hiện phương pháp theo điểm 10 khi chạy trên máy tính hoặc bộ xử lý tín hiệu.

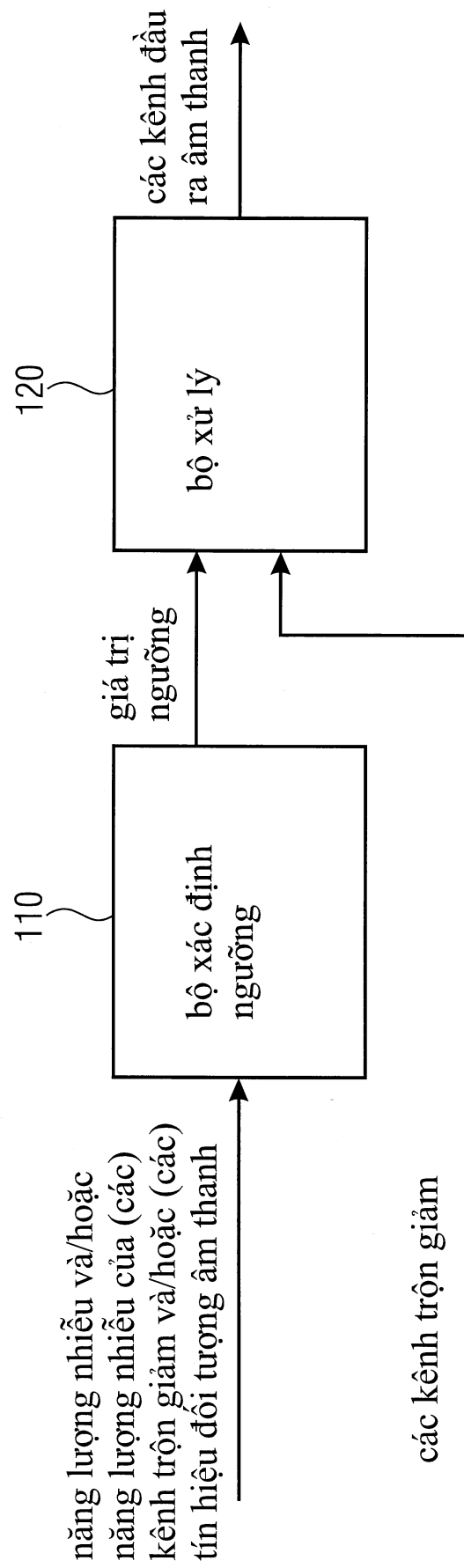


FIG 1

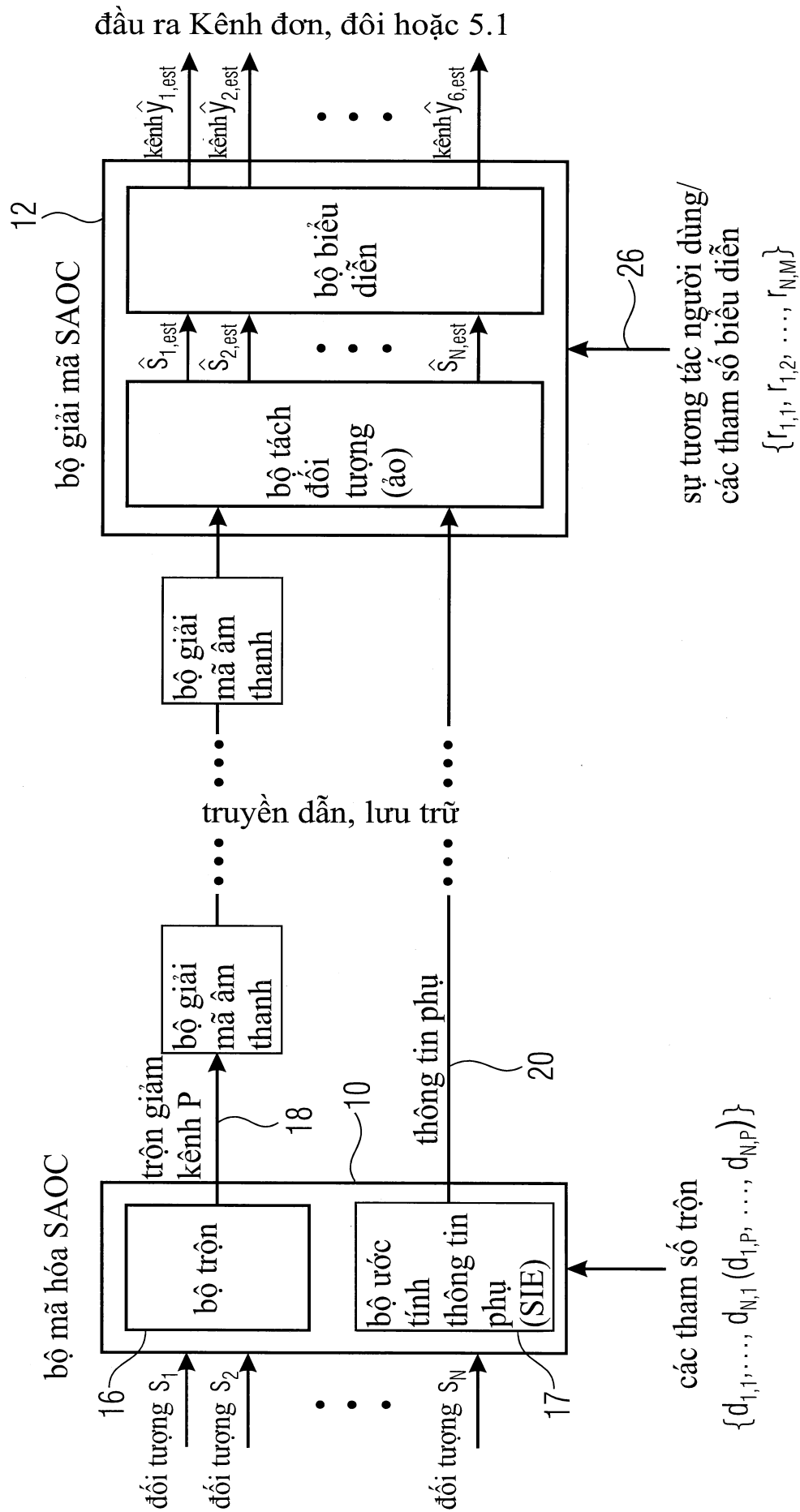


FIG 2

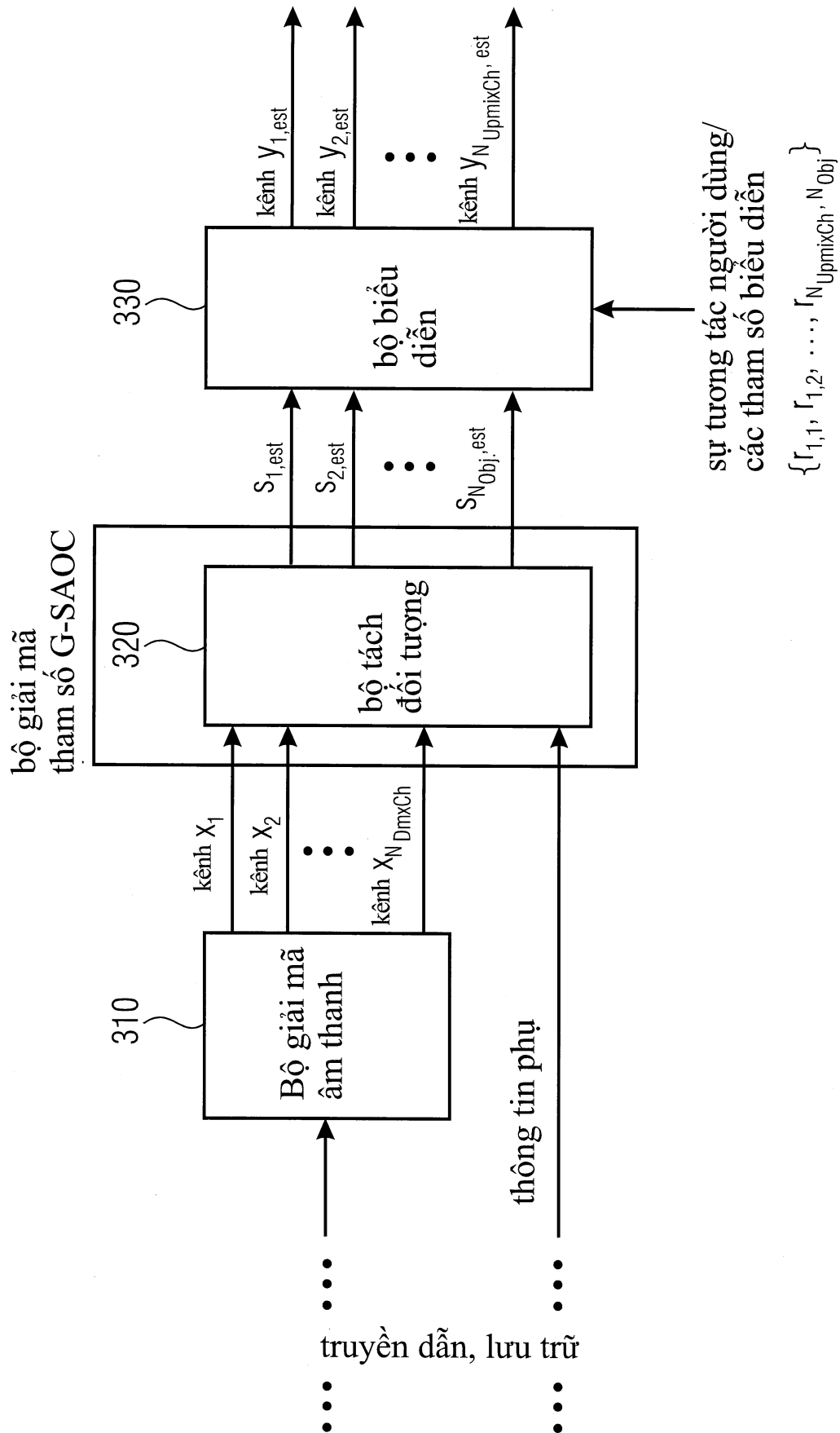


FIG 3

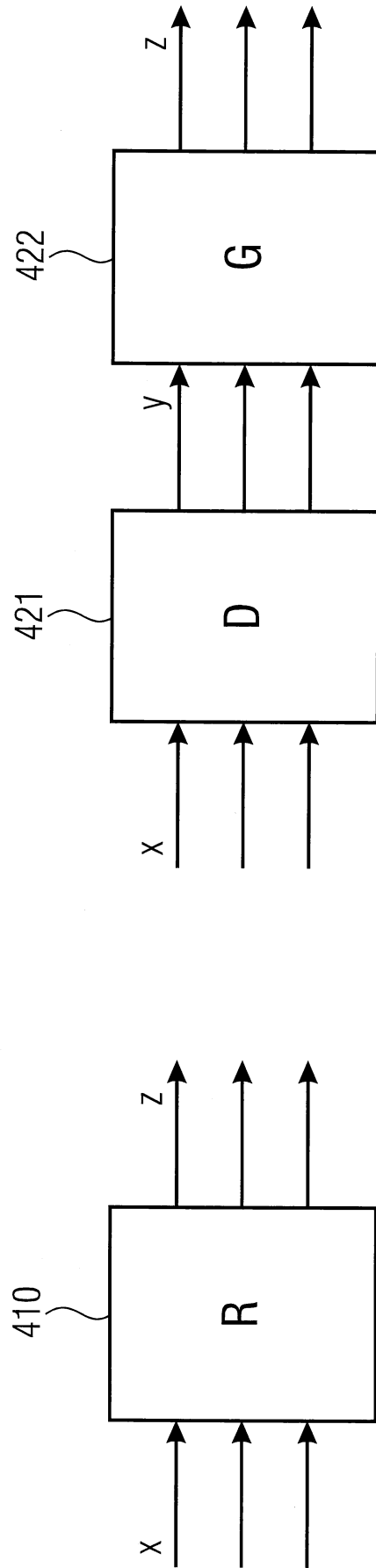


FIG 4