



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

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

(11)



1-0026244

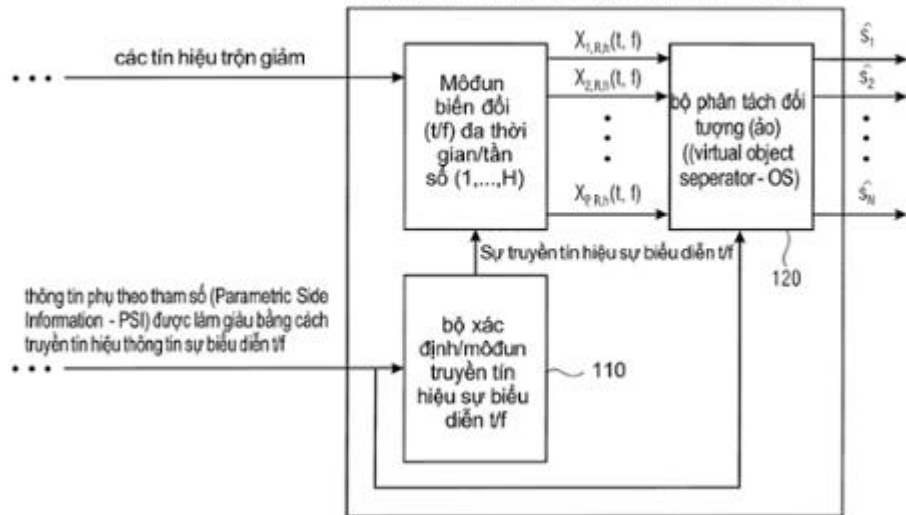
(51)⁷ G10L 19/008; G10L 25/18; G10L 19/20 (13) B

- (21) 1-2015-04368 (22) 09/05/2014
(86) PCT/EP2014/059570 09/05/2014 (87) WO 2014/184115 A1 20/11/2014
(30) 13167484.8 13/05/2013 EP
(45) 25/11/2020 392 (43) 25/01/2016 334A
(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) DISCH, Sascha (DE); PAULUS, Jouni (DE); KASTNER, Thorsten (DE).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)

(54) BỘ GIẢI MÃ ÂM THANH VÀ PHƯƠNG PHÁP GIẢI MÃ TÍN HIỆU ÂM THANH ĐA ĐỐI TƯỢNG, BỘ MÃ HÓA ÂM THANH VÀ PHƯƠNG PHÁP MÃ HÓA NHIỀU ĐỐI TƯỢNG ÂM THANH

(57) Sáng chế đề cập đến bộ giải mã âm thanh và phương pháp giải mã tín hiệu âm thanh đa đối tượng, bộ mã hóa âm thanh và phương pháp mã hóa nhiều đối tượng âm thanh. Bộ giải mã âm thanh được đề xuất để giải mã tín hiệu âm thanh đa đối tượng chứa tín hiệu trộn giảm (X) và thông tin phụ (PSI). Thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng (PSI_i) cho đối tượng âm thanh (s_i) trong vùng thời gian/tần số ($R(t_R, f_R)$), và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) của thông tin phụ chỉ rõ đối tượng cho đối tượng âm thanh (s_i) trong vùng thời gian/tần số ($R(t_R, f_R)$). Bộ giải mã âm thanh bao gồm bộ xác định độ phân giải thời gian/tần số (110) được tạo cấu hình để xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) từ thông tin phụ (PSI) cho đối tượng âm thanh (s_i). Bộ giải mã âm thanh còn bao gồm bộ phân tách đối tượng (120) được tạo cấu hình để phân tách đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$). Bộ mã hóa tương ứng và phương pháp tương ứng cho giải mã và mã hóa cũng được mô tả.

Bộ phân tách đối tượng được tăng cường
(ảo)((virtual) Enhanced Object Separator - E-OS)



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc xử lý tín hiệu âm thanh và, cụ thể, bộ giải mã, bộ mã hóa, hệ thống, các phương pháp và chương trình máy tính để mã hóa đối tượng âm thanh sử dụng độ phân giải thời gian-tần số riêng lẻ thích ứng đối tượng âm thanh.

Các phương án theo sáng chế đề cập đến bộ giải mã âm thanh để giải mã tín hiệu âm thanh đa đối tượng gồm tín hiệu trộn giảm và thông tin phụ theo tham số liên quan đối tượng (parametric side information - PSI). Phương án khác theo sáng chế đề cập đến bộ giải mã cung cấp phép biểu diễn tín hiệu trộn tăng phụ thuộc vào phép biểu diễn tín hiệu trộn giảm và PSI liên quan đối tượng. Phương án khác theo sáng chế đề cập đến phương pháp giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm và PSI liên quan. Phương án khác theo sáng chế đề cập đến phương pháp cung cấp phép biểu diễn tín hiệu trộn tăng phụ thuộc vào phép biểu diễn tín hiệu trộn giảm và PSI liên quan đối tượng.

Phương án khác của sáng chế đề cập đến bộ mã hóa âm thanh để mã hóa tín hiệu nhiều tín hiệu đối tượng âm thanh thành tín hiệu trộn giảm và PSI. Phương án khác của sáng chế đề cập đến phương pháp mã hóa nhiều tín hiệu đối tượng âm thanh thành tín hiệu trộn giảm và PSI.

Phương án khác theo sáng chế đề cập đến chương trình máy tính tương ứng với (các) phương pháp để giải mã, mã hóa, và/hoặc cung cấp tín hiệu trộn tăng.

Phương án khác của sáng chế đề cập đến độ phân giải thời gian-tần số riêng lẻ thích ứng đối tượng âm thanh chuyển mạch cho sự thao tác hỗn hợp tín hiệu.

Tình trạng kỹ thuật của sáng chế

Trong các hệ thống âm thanh số hiện đại, xu hướng cơ bản là cho phép các hiệu chỉnh liên quan đến đối tượng-âm thanh của nội dung được truyền trên phía bộ nhận. Những sửa đổi này gồm các sửa đổi độ khuếch đại của các phần được chọn của tín hiệu âm thanh và/hoặc xác định lại vị trí trong không gian các đối tượng âm thanh

được dành cho trường hợp phát lại đa kênh qua các loa được phân bố trong không gian. Điều này có thể đạt được bằng cách phân phối riêng lẻ 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, ngày càng có nhiều mong muốn cho phép sự tương tác người dùng lên việc phát lại nội dung âm thanh định hướng đối tượng cũng như yêu cầu sử dụng các khả năng mở rộng việc phát lại đa kênh để kết xuất riêng lẻ các nội dung âm thanh hoặc các phần của chúng để cải thiện được ấn tượng nghe. Bằng cách này, việc sử dụng nội dung âm thanh đa kênh mang đến những cải thiện đáng kể cho người dùng. Ví dụ, có thể thu được ấn tượng nghe ba chiều, điều này mang đến sự hài lòng được cải thiện về các ứng dụng giải trí cho người dùng. Tuy nhiên, nội dung âm thanh đa kênh cũng hữu dụng trong các môi trường chuyên nghiệp, ví dụ, trong các ứng dụng hội thoại bằng điện thoại, bởi vì khả năng nghe rõ người nói có thể được cải thiện bằng cách sử dụng sự phát lại âm thanh đa kênh. Ứng dụng khả thi khác là cung cấp cho người nghe bản nhạc để điều chỉnh riêng biệt mức phát lại và/hoặc vị trí không gian của các phần khác nhau (còn được gọi là “các đối tượng âm thanh”) hoặc đoạn nhạc, chẳng hạn như phần xướng âm hoặc các nhạc cụ khác nhau. Người dùng có thể thực hiện sự điều chỉnh này với lý do sở thích cá nhân, để chuyển biên dễ dàng hơn một hoặc nhiều hơn một phần từ một đoạn nhạc, cho các mục đích giáo dục, karaoke, diễn tập, v.v.

Việc truyền rời rạc thẳng tất cả nội dung âm thanh đa kênh hoặc đa đối tượng dạng 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 bit rất cao. Tuy nhiên, người ta cũng mong muốn truyền và lưu trữ dữ liệu âm thanh theo hướng hiệu quả về tốc độ bit. Do đó, người ta sẵn sàng chấp nhận sự đánh đổi hợp lý giữa chất lượng âm thanh và các yêu cầu tốc độ bit để tránh được việc tải nguồn quá mức gây ra bởi các ứng dụng đa kênh/đa đối tượng.

Gần đây, trong lĩnh vực mã hóa âm thanh, các kỹ thuật tham số cho việc truyền/lưu trữ hiệu quả tốc độ bit của các tín hiệu âm thanh đa đối tượng/đa kênh được đưa vào bởi, ví dụ, nhóm chuyên gia hình ảnh động (Moving Picture Experts Group (MPEG)) và khác nữa. Một ví dụ là MPEG Surround (MPS) là phương pháp định

hướng kênh [MPS, BCC], hoặc mã hóa đối tượng âm thanh không gian MPEG (MPEG Spatial Audio Object Coding - SAOC)) là phương pháp định hướng đối tượng [JSC, SAOC, SAOC1, SAOC2]. Phương pháp định hướng đối tượng khác được gọi là “phân tách nguồn được thông báo” [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]. Các kỹ thuật này nhằm tái tạo cảnh âm thanh đầu ra mong muốn hoặc đối tượng nguồn âm thanh mong muốn trên cơ sở của việc trộn giảm các kênh/các đối tượng và thông tin phụ bổ sung mô tả cảnh âm thanh được lưu trữ/truyền và/hoặc các đối tượng nguồn âm thanh trong cảnh âm thanh.

Việc ước lượng và áp dụng thông tin phụ liên quan đến kênh/đối tượng trong các hệ thống này được thực hiện theo cách thức chọn thời gian-tần số. Do đó, các hệ thống này sử dụng các biến đổi thời gian-tần số như biến đổi Fourier rời rạc (Discrete Fourier Transform - DFT), biến đổi Fourier thời gian ngắn (Short Time Fourier Transform - STFT) hoặc các giàn bộ lọc như các giàn bộ lọc gương cầu phương (Quadrature Mirror Filter - QMF), v.v. Nguyên lý cơ bản của các hệ thống này được mô tả trên Fig.1, sử dụng ví dụ MPEG SAOC.

Trong trường hợp của STFT, chiều thời gian được biểu diễn bởi số khối thời gian và chiều phổ được lấy bởi số lượng hệ số phổ (“ngăn”). Trong trường hợp của QMF, chiều thời gian được biểu diễn bởi số khe thời gian và chiều phổ được lấy bởi số băng phụ. Nếu độ phân giải phổ của QMF được cải thiện bởi việc áp dụng sau của giai đoạn bộ lọc thứ hai, toàn bộ giàn bộ lọc được gọi là QMF lai và các băng phụ độ phân giải mịn được gọi là các băng phụ lai.

Như đã nêu trên, trong SAOC, xử lý chung được tiến hành theo cách lựa chọn thời gian-tần số và có thể được mô tả như sau trong mỗi băng tần số:

- N tín hiệu đối tượng âm thanh đầu vào $s_1 \dots s_N$ được trộn giảm tới P kênh $x_1 \dots x_P$ như là một phần của việc xử lý bộ mã hóa sử dụng ma trận trộn giảm bao gồm các phần tử $d_{1,1} \dots d_{N,P}$. Ngoài ra, bộ mã hóa tách thông tin phụ mô tả các đặc tính của các đối tượng âm thanh đầu vào (môđun bộ ước lượng thông tin phụ (side information estimator - SEI)). Đối với MPEG SAOC, các mối quan hệ của các năng lượng đối tượng đối với nhau là dạng cơ bản nhất của thông tin phụ này.

- (Các) tín hiệu trộn giảm và thông tin phụ được truyền/lưu trữ. Để đạt được điều này, (các) tín hiệu âm thanh trộn giảm có thể được nén, ví dụ, sử dụng các bộ mã hóa âm thanh cảm giác đã biết như MPEG-1/2 lớp II hoặc III (cũng được biết đến như .mp3), MPEG-2/4 mã hóa âm thanh tiên tiến (Advanced Audio Coding - AAC) v.v.
- Ở đầu nhận, bộ giải mã cố gắng dựa trên khái niệm phục hồi lại các đối tượng âm thanh ban đầu (“phân tách đối tượng”) từ các tín hiệu trộn giảm (đã được giải mã) sử dụng thông tin phụ đã được truyền. Các tín hiệu đối tượng được lấy gần đúng $\hat{s}_1 \dots \hat{s}_N$ này tiếp đó được trộn trong cảnh mục tiêu được biểu diễn bởi M kênh đầu ra âm thanh $\hat{y}_1 \dots \hat{y}_M$ sử dụng ma trận biểu diễn được mô tả bởi các hệ số $r_{1,1} \dots r_{N,M}$ trên Fig.1. Cảnh mục tiêu mong muốn có thể là, trong trường hợp đặc biệt, sự kết xuất của duy nhất một tín hiệu nguồn khỏi hỗn hợp (kịch bản phân tách nguồn), ngoài ra còn cảnh âm thanh tùy ý bất kỳ gồm các đối tượng được truyền.

Các hệ thống trên cơ sở thời gian-tần số có thể sử dụng phép biến đổi thời gian-tần số (t/f) với độ phân giải theo thời gian và tần số tĩnh. Lựa chọn lưới độ phân giải t/f cố định điển hình bao gồm sự cân bằng giữa độ phân giải thời gian và tần số.

Hiệu quả của độ phân giải t/f cố định có thể được trình bày trong ví dụ về các đối tượng tín hiệu điển hình trong hỗn hợp tín hiệu âm thanh. Ví dụ, phổ âm thanh đưa ra cấu trúc liên quan hài hòa với tần số cơ sở và một số sóng hài bậc cao. Năng lượng của tín hiệu đó được tập trung tại khu vực tần số nhất định. Với những tín hiệu này, độ phân giải tần số cao của phép biểu diễn t/f được sử dụng có lợi cho việc phân tách các vùng phổ âm thanh băng hẹp từ hỗn hợp tín hiệu. Ngược lại, các tín hiệu nhất thời, như các âm thanh trống, thường có các cấu trúc thời gian khác biệt: năng lượng thực chất chỉ xuất hiện trong khoảng thời gian ngắn và được lan truyền trên phạm vi rộng của các tần số. Với những tín hiệu này, độ phân giải thời gian cao của phép biểu diễn t/f được sử dụng là thuận lợi cho việc phân tách các tín hiệu nhất thời từ hỗn hợp tín hiệu.

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. Holzer, 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. Brassier: "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. Pintel: "Informed Audio Source Separation from Compressed Linear stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011

Bản chất kỹ thuật của sáng chế

Cần phải tính đến các nhu cầu khác nhau của các loại đối tượng âm thanh khác nhau liên quan đến phép biểu diễn của chúng trong miền thời gian-tần số khi tạo ra và/hoặc đánh giá thông tin phụ đối tượng cụ thể lần lượt tại phía bộ mã hóa và phía bộ giải mã.

Mong muốn này và/hoặc các mong muốn khác nữa được giải quyết bằng bộ giải mã âm thanh để giải mã tín hiệu âm thanh đa đối tượng, bằng bộ mã hóa âm thanh để mã hóa nhiều tín hiệu đối tượng âm thanh thành tín hiệu trộn giảm và thông tin phụ, bằng phương pháp giải mã tín hiệu âm thanh đa đối tượng, bằng phương pháp mã hóa các tín hiệu đối tượng âm thanh, hoặc bằng chương trình máy tính tương ứng, được xác định bởi các điểm yêu cầu bảo hộ độc lập.

Theo ít nhất một số phương án, bộ giải mã âm thanh để giải mã tín hiệu đa đối tượng được đề xuất. Tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm và thông tin phụ. Thông tin phụ bao gồm các thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh trong ít nhất một vùng thời gian/tần số. Thông tin phụ còn bao gồm thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng biểu thị bởi độ phân giải thời gian/tần số chỉ rõ đối tượng của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh trong ít nhất một vùng thời gian/tần số. Bộ giải mã âm thanh bao gồm bộ xác định độ phân giải thời gian/tần số được tạo cấu hình để xác định thông tin phụ độ phân giải thời gian/tần số đối-tượng riêng biệt từ thông tin phụ cho ít nhất một đối tượng âm thanh. Bộ giải mã âm thanh còn bao gồm bộ phân tách đối tượng được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh từ tín hiệu trộn giảm sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian tần số chỉ rõ đối tượng.

Phương án khác của sáng chế đề xuất bộ mã hóa để mã hóa các tín hiệu đối tượng âm thanh thành tín hiệu trộn giảm và thông tin phụ. Bộ mã hóa âm thanh bao gồm bộ biến đổi thời gian thành tần số được tạo cấu hình để biến đổi nhiều đối tượng âm thanh ít nhất thành nhiều biến đổi tương ứng thứ nhất sử dụng độ phân giải thời gian/tần số thứ nhất và nhiều biến đổi tương ứng thứ hai sử dụng độ phân giải thời gian/tần số thứ hai. Bộ mã hóa âm thanh còn bao gồm bộ xác định thông tin phụ được tạo cấu hình để xác định ít nhất thông tin phụ thứ nhất cho nhiều biến đổi tương ứng thứ nhất và thông tin phụ thứ hai cho nhiều biến đổi tương ứng thứ hai. Thông tin phụ thứ nhất và thứ hai chỉ ra mối liên hệ của nhiều đối tượng âm thanh với nhau trong lần lượt các độ phân giải thời gian/tần số thứ nhất và thứ hai, trong vùng thời gian/tần số. Bộ mã hóa âm thanh còn bao gồm bộ chọn thông tin phụ được tạo cấu hình để chọn, cho ít nhất một đối tượng âm thanh của các đối tượng âm thanh, một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thứ hai trên cơ sở tiêu chí phù hợp. Tiêu chí phù hợp biểu thị sự phù hợp của ít nhất độ phân giải tần số thứ nhất và thứ hai để biểu diễn đối tượng âm thanh trong miền thời gian/tần số. Thông tin phụ chỉ rõ đối tượng được chọn được chèn vào đầu ra thông tin phụ bởi bộ mã hóa âm thanh.

Phương án khác của sáng chế đề xuất phương pháp giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm và thông tin phụ. Thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh trong ít nhất một vùng thời gian/tần số, và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng của thông tin phụ đối tượng-riêng biệt cho ít nhất một đối tượng âm thanh trong ít nhất một vùng thời gian/tần số. Phương pháp bao gồm việc xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng từ thông tin phụ của ít nhất một đối tượng âm thanh. Phương pháp còn bao gồm việc phân tách ít nhất một đối tượng âm thanh từ tín hiệu trộn giảm sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian/tần số chỉ rõ đối tượng.

Phương án khác của sáng chế này đề xuất phương pháp mã hóa nhiều đối tượng âm thanh thành tín hiệu trộn giảm và thông tin phụ. Phương pháp bao gồm việc biến đổi nhiều đối tượng âm thanh ít nhất thành nhiều biến đổi tương ứng thứ nhất sử dụng độ phân giải thời gian/tần số thứ nhất và nhiều biến đổi tương ứng thứ hai sử dụng độ

phân giải thời gian/tần số thứ hai. Phương pháp còn bao gồm việc xác định ít nhất một thông tin phụ thứ nhất cho các biến đổi tương ứng thứ nhất và một thông tin phụ thứ hai cho các biến đổi tương ứng thứ hai. Thông tin phụ thứ nhất và thứ hai chỉ ra mối liên hệ giữa nhiều đối tượng âm thanh với nhau lần lượt trong các độ phân giải thời gian/tần số thứ nhất và thứ hai, trong vùng thời gian/tần số. Phương pháp còn bao gồm việc lựa chọn, cho ít nhất một đối tượng âm thanh của các đối tượng âm thanh, một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thứ hai trên cơ sở tiêu chí phù hợp. Tiêu chí phù hợp biểu thị sự phù hợp của ít nhất độ phân giải thời gian/tần số thứ nhất hoặc thứ hai để biểu diễn đối tượng âm thanh trong miền thời gian/tần số. Thông tin phụ chỉ rõ đối tượng được chèn vào đầu ra thông tin phụ bởi bộ mã hóa âm thanh.

Hiệu suất phân tách đối tượng âm thanh thường giảm nếu phép biểu diễn t/f được sử dụng không phù hợp với đặc tính thời gian và/hoặc đặc tính phổ của đối tượng âm thanh được tách ra từ hỗn hợp. Hiệu suất không đủ có thể dẫn đến nhiễu xuyên âm giữa các đối tượng được phân tách. Nhiễu xuyên âm được hiểu là các tiếng vọng trước hoặc tiếng vọng sau, phép biến đổi âm sắc, trong trường hợp giọng nói con người, được gọi là vọng âm. Phương án của sáng chế đưa ra các phép biểu diễn t/f thay thế mà từ đó phép biểu diễn t/f phù hợp nhất có thể được lựa chọn cho đối tượng âm thanh và vùng thời gian/tần số đã cho khi xác định thông tin phụ tại phía bộ mã hóa, hoặc khi sử dụng thông tin phụ tại phía bộ giải mã. Điều này cung cấp hiệu suất phân tách được cải thiện cho việc phân tách các đối tượng âm thanh và chất lượng chủ quan được cải thiện cho các tín hiệu đầu ra được kết xuất so với tình trạng của lĩnh vực kỹ thuật.

So với các sơ đồ khác để mã hóa/giải mã đối tượng âm thanh trong không gian, lượng thông tin phụ có thể hoàn toàn giống hoặc cao hơn một chút. Theo các phương án của sáng chế, thông tin phụ được sử dụng một cách hiệu quả, vì nó được áp dụng theo cách chỉ rõ đối tượng xem xét đến thuộc tính chỉ rõ đối tượng của đối tượng âm thanh nhất định liên quan đến cấu trúc thời gian và phổ của nó. Nói cách khác, phép biểu diễn t/f của thông tin phụ được thiết kế riêng cho các đối tượng âm thanh khác nhau.

Mô tả vắn tắt các hình vẽ

Các phương án theo sáng chế sẽ được mô tả sau đây có tham chiếu đến các hình vẽ kèm theo, trong đó:

Fig.1 thể hiện sơ đồ khối dạng giản lược của tổng quan về khái niệm của hệ thống SAOC;

Fig.2 thể hiện sơ đồ giản lược và minh họa của phép biểu diễn thời gian-phổ của tín hiệu âm thanh đơn kênh;

Fig.3 thể hiện sơ đồ khối dạng giản lược của phép tính chọn lọc thời gian-tần số của thông tin phụ trong bộ mã hóa SAOC;

Fig.4 minh họa sơ lược về nguyên lý của bộ ước lượng thông tin phụ nâng cao theo một số phương án;

Fig.5 minh họa sơ lược về vùng t/f $R(t_R, f_R)$ được biểu diễn bởi nhiều phép biểu diễn t/f khác nhau;

Fig.6 là sơ đồ khối dạng giản lược của phép tính thông tin phụ và môđun lựa chọn theo các phương án;

Fig.7 minh họa sơ lược về việc giải mã SAOC bao gồm môđun phân tách đối tượng (ảo) nâng cao (Enhanced (virtual) Object Separation (EOS) module - EOS);

Fig.8 thể hiện sơ đồ khối dạng giản lược của môđun phân tách đối tượng nâng cao (enhanced object separation module - EOS-module);

Fig.9 là sơ đồ khối dạng giản lược của bộ giải mã âm thanh theo các phương án;

Fig.10 là sơ đồ khối dạng giản lược của bộ giải mã âm thanh mà giải mã các phép biểu diễn t/f thay thế H và sau đó lựa chọn các chỉ rõ đối tượng, theo phương án tương đối đơn giản;

Fig.11 minh họa sơ lược về vùng t/f $R(t_R, f_R)$ được biểu diễn trong các phép biểu diễn t/f khác nhau và kết quả thu được về việc xác định ma trận hiệp phương sai E được ước lượng trong vùng t/f ;

Fig.12 minh họa sơ lược về khái niệm cho việc phân tách các đối tượng âm thanh sử dụng phép biến đổi thu phóng để thể hiện sự phân tách đối tượng âm thanh trong phép biểu diễn thời gian/tần số được thu phóng;

Fig.13 thể hiện lưu đồ dạng giản lược của phương pháp để giải mã tín hiệu trộn giảm với thông tin phụ được kết hợp; và

Fig.14 thể hiện lưu đồ dạng giản lược của phương pháp để mã hóa các đối tượng âm thanh thành tín hiệu trộn giảm và thông tin phụ.

Mô tả chi tiết sáng chế

Fig.1 thể hiện sự sắp xếp chung của bộ mã hóa SAOC 10 và bộ giải mã SAOC 12. Bộ mã hóa SAOC 10 nhận N đối tượng đầu vào, tức là, tín hiệu âm thanh s_1 đến s_N . Cụ thể, bộ mã hóa 10 bao gồm bộ trộn giảm 16 mà nhận các tín hiệu âm thanh s_1 đến s_N và trộn giảm chúng thành tín hiệu trộn giảm 18. Ngoài ra, sự trộn giảm có thể được cung cấp bên ngoài ("sự trộn giảm khéo léo") và hệ thống ước lượng thông tin phụ bổ sung để khiến sự trộn giảm được đề xuất phù hợp với sự trộn giảm được tính toán. Trên Fig.1, tín hiệu trộn giảm được thể hiện là P -tín hiệu kênh. Do đó, bất kì cấu hình tín hiệu trộn giảm kênh đơn sắc ($P=1$), kênh lập thể ($P=2$) hoặc đa kênh ($P \geq 2$) đều có thể hiểu được.

Trong trường hợp trộn lập thể, các kênh của tín hiệu trộn giảm 18 được biểu thị là $L0$ và $R0$, trong trường hợp trộn đơn sắc, các kênh của tín hiệu trộn giảm được biểu thị đơn giản là $L0$. Để cho phép bộ giải mã SAOC 12 khôi phục các chỉ rõ đối tượng S_1 đến S_N , bộ ước lượng 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 lập thể, các tham số SAOC bao gồm các chênh lệch mức đối tượng (object level differences - OLD), các tham số tương quan liên đối tượng (inter-object correlations - IOC), các giá trị hệ số khuếch đại trộn giảm (downmix gain values - DMG) và các chênh lệch mức kênh trộn giảm (downmix channel level differences - DCLD). Thông tin phụ 20 gồm tham số SAOC, cùng với tín hiệu trộn giảm 18, hình thành dòng dữ liệu đầu ra được nhận bởi bộ giải mã SAOC 12.

Bộ giải mã SAOC 12 bao gồm bộ trộn tăng mà nhận các tín hiệu trộn giảm 18 cùng với thông tin phụ 20 để khôi phục và kết xuất các tín hiệu âm thanh s_1 và s_N trên bất kỳ tập hợp các kênh được lựa chọn bởi người dùng từ y_1 đến y_M , với sự kết xuất được quy định bởi thông tin kết xuất 26 được nhập vào trong bộ giải mã SAOC 12.

Các tín hiệu âm thanh s_1 đến s_N có thể được nhập vào bộ mã hóa 10 trong miền mã hóa bất kỳ, chẳng hạn như, trong miền thời gian hoặc phổ. Trong trường hợp các tín hiệu âm thanh s_1 đến s_N được đưa vào trong bộ mã hóa 10 trong miền thời gian, như PCM được mã hóa, bộ mã hóa 10 có thể sử dụng giàn bộ lọc, như giàn QMF lai, để truyền các tín hiệu vào trong miền phổ, trong đó các tín hiệu âm thanh được biểu diễn trong vài băng phụ được kết hợp với các phân phổ khác nhau, tại độ phân giải giàn bộ lọc cụ thể. Nếu các tín hiệu âm thanh s_1 đến s_N là sẵn sàng trong phép biểu diễn được mong đợi bởi bộ mã hóa 10, các tín hiệu âm thanh không phải thực hiện sự phân tích phổ.

Fig.2 thể hiện tín hiệu âm thanh trong miền phổ vừa được đề cập. Như có thể thấy, tín hiệu âm thanh được biểu diễn là các tín hiệu băng phụ. Mỗi tín hiệu băng phụ 30_1 đến 30_K bao gồm một chuỗi liên tiếp các trị số băng phụ được chỉ ra bởi các ô nhỏ 32. Như có thể thấy, các trị số băng phụ 32 của các tín hiệu băng phụ 30_1 đến 30_K được đồng bộ hóa với nhau theo thời gian sao cho đối với mỗi khe thời gian của giàn bộ lọc liên tiếp 34, mỗi băng phụ 30_1 đến 30_K bao gồm chính xác một trị số băng phụ 32. Như được minh họa bởi trục tần số 36, các tín hiệu băng phụ 30_1 đến 30_K được kết hợp với các vùng tần số khác nhau, và như được minh họa bởi trục thời gian 38, các khe thời gian giàn bộ lọc 34 được sắp đặt liên tiếp theo thời gian.

Như đã nêu ở trên, bộ tách thông tin phụ 17 tính toán các thông số SAOC từ các tín hiệu âm thanh đầu vào s_1 đến s_N . Theo tiêu chuẩn SAOC được cung cấp hiện tại, bộ mã hóa 10 thực hiện phép tính này trong độ phân giải thời gian/tần số mà độ phân giải này có thể bị giảm đi tương đối so với độ phân giải thời gian/tần số ban đầu được xác định bởi các khe thời gian giàn bộ lọc 34 và sự phân tích băng phụ, bởi một lượng nhất định, với lượng nhất định này được báo hiệu cho phía bộ giải mã trong thông tin phụ 20. Các nhóm khe thời gian giàn bộ lọc liên tiếp 34 có thể tạo ra khung SAOC 41. Ngoài ra, số lượng băng tham số trong khung SAOC 41 được truyền trong thông tin phụ 20. Do đó, các miền thời gian/tần số được chia thành các ô thời gian/tần số được minh họa trên Fig.2 bằng đường nét đứt 42. Trên Fig.2 các băng tham số được phân phối theo cách tương tự trong các khung SAOC 41 khác nhau được miêu tả, do đó thu được sự sắp xếp đều đặn của các ô thời gian/tần số. Tuy nhiên, nói chung, các băng

tham số có thể thay đổi từ một khung SAOC 41 đến khung tiếp theo, phụ thuộc vào các nhu cầu khác nhau cho độ phân giải phổ trong các khung SAOC 41 tương ứng. Hơn nữa, chiều dài của các khung SAOC 41 cũng có thể thay đổi. Kết quả là, sự sắp đặt các ô thời gian/tần số có thể là không đều. Tuy nhiên, các ô thời gian/tần số trong khung SAOC cụ thể 41 thường có cùng thời lượng và được căn chỉnh theo hướng thời gian, nghĩa là tất cả các ô t/f trong khung SAOC 41 đã nêu bắt đầu khi bắt đầu khung SAOC 41 đã cho và kết thúc khi kết thúc khung SAOC 41 đã nêu.

Bộ tách thông tin phụ 17 tính tham số SAOC theo các công thức sau. Cụ thể, bộ tách thông tin phụ 17 tính toán các chênh lệch mức đối tượng cho mỗi đối tượng i là:

$$OLD_i^{l,m} = \frac{\sum_{n \in l} \sum_{k \in m} x_i^{n,k} x_i^{n,k*}}{\max_j \left(\sum_{n \in l} \sum_{k \in m} x_j^{n,k} x_j^{n,k*} \right)}$$

trong đó các tổng và các chỉ số n và k lần lượt đi qua tất cả các chỉ số theo thời gian 34, và tất cả các chỉ số phổ 30 thuộc cùng ô thời gian/tần số 42 nhất định, được tham chiếu bởi các chỉ số l cho khung SAOC (hoặc khe thời gian xử lý) và m cho các băng tham số. Do đó, các năng lượng của tất cả các trị số băng phụ x_i của tín hiệu hoặc đối tượng âm thanh i được tính tổng và được chuẩn hóa đến trị số năng lượng cao nhất của ô đó trong số tất cả các đối tượng hoặc các tín hiệu âm thanh.

Hơn nữa, bộ tách thông tin phụ SAOC 17 có thể tính toán phép đo tương tự của các ô thời gian/tần số tương ứng của cặp đối tượng đầu vào khác nhau s_1 đến s_N . Mặc dù bộ trộn giảm SAOC có thể tính toán các phép đo tương tự giữa các cặp đối tượng đầu vào s_1 đến s_N , bộ trộn giảm 16 cũng có thể ngăn chặn việc truyền tín hiệu của các phép đo tương tự hoặc hạn chế sự tính toán của các phép đo tương tự với các đối tượng âm thanh s_1 đến s_N mà tạo nên kênh trái và kênh phải của kênh âm lập thể thông thường. Trong trường hợp bất kỳ, phép đo tương tự được gọi là tham số tương quan chéo liên đối tượng $IOC_{i,j}^{l,m}$. Phép tính là như sau:

$$IOC_{i,j}^{l,m} = IOC_{j,i}^{l,m} = \text{Re} \left\{ \frac{\sum_{n \in l} \sum_{k \in m} x_i^{n,k} x_j^{n,k*}}{\sqrt{\sum_{n \in l} \sum_{k \in m} x_i^{n,k} x_i^{n,k*} \sum_{n \in l} \sum_{k \in m} x_j^{n,k} x_j^{n,k*}}} \right\}$$

với chỉ số n và k đi qua tất cả các trị số bằng phụ thuộc ô thời gian/tần số 42 nhất định và i và j biểu thị cặp đối tượng âm thanh s_1 đến s_N nhất định.

Bộ trộn giảm 16 trộn giảm các đối tượng s_1 đến s_N bằng cách sử dụng thừa số khuếch đại được áp dụng vào mỗi đối tượng s_1 đến s_N . Đó là, thừa số khuếch đại D_i được áp dụng vào đối tượng i và sau đó tất cả các đối tượng s_1 đến s_N được gán trọng số được tính tổng để thu được tín hiệu trộn giảm đơn sắc, được lấy làm ví dụ trên Fig.1 nếu $P=1$. Trong một trường hợp ví dụ khác của tín hiệu trộn giảm hai kênh, được miêu tả trên Fig.1 nếu $P=2$, thừa số khuếch đại $D_{1,i}$ được áp dụng cho đối tượng i và sau đó tất cả các đối tượng được khuếch đại thu được này được tính tổng để thu được kênh trộn giảm trái $L0$, và các thừa số khuếch đại $D_{2,i}$ được áp dụng cho đối tượng i và sau đó các đối tượng được khuếch đại được tính tổng để thu được kênh trộn giảm phải $R0$. Quy trình tương tự như trên được áp dụng trong trường hợp trộn giảm đa kênh ($P \geq 2$).

Quy định trộn giảm này được phát tín hiệu đến phía bộ giải mã bằng các hệ số khuếch đại trộn giảm DMG_i và trong trường hợp tín hiệu trộn giảm lập thể, và các chênh lệch mức kênh trộn giảm $DCLD_i$.

Các hệ số khuếch đại trộn giảm được tính toán theo:

$$DMG_i = 20 \log_{10} (D_i + \varepsilon) \quad , \text{ (trộn giảm đơn sắc),}$$

$$DMG_i = 10 \log_{10} (D_{1,i}^2 + D_{2,i}^2 + \varepsilon) \quad , \text{ (trộn giảm lập thể),}$$

trong đó ε là số nhỏ chẳng hạn như 10^{-9} .

Đối với các $DCLD_s$ công thức sau đây áp dụng:

$$DCLD_i = 20 \log_{10} \left(\frac{D_{1,i}}{D_{2,i} + \varepsilon} \right)$$

Ở chế độ thông thường, bộ trộn giảm 16 tạo ra tín hiệu trộn giảm lần lượt theo:

$$(L0) = (D_i) \begin{pmatrix} Obj_1 \\ \vdots \\ Obj_N \end{pmatrix}$$

đối với trộn giảm đơn sắc, hoặc

$$\begin{pmatrix} L0 \\ R0 \end{pmatrix} = \begin{pmatrix} D_{1,i} \\ D_{2,i} \end{pmatrix} \begin{pmatrix} Obj_1 \\ \vdots \\ Obj_N \end{pmatrix}$$

đối với trộn giảm lập thể.

Do đó, trong các công thức nêu trên, thông số OLD và IOC là hàm của các tín hiệu âm thanh và các thông số DMG và DCLD là hàm của D. Bằng cách này, cần lưu ý rằng D có thể biến thiên theo thời gian.

Do đó, trong chế độ bình thường, bộ trộn giảm 16 trộn tất cả các đối tượng s_1 đến s_n không có sự ưu tiên, tức là, xử lý tất cả các đối tượng s_1 đến s_N một cách đều nhau.

Ở phía bộ giải mã, bộ trộn tăng thực hiện sự đảo ngược quy trình trộn giảm và sự thực hiện của "thông tin kết xuất" 26 được biểu diễn bởi ma trận **R** (trong tài liệu đôi khi còn được gọi là A) trong một bước tính toán, cụ thể, trong trường hợp trộn giảm hai kênh:

$$\begin{pmatrix} Ch_1 \\ \vdots \\ Ch_M \end{pmatrix} = RED^* (DED^*)^{-1} \begin{pmatrix} L0 \\ R0 \end{pmatrix},$$

trong đó ma trận **E** là hàm của các tham số OLD và IOC. Ma trận **E** là ma trận hiệp phương sai được ước lượng của các đối tượng âm thanh s_1 đến s_N . Trong các phương án SAOC hiện tại, sự tính toán ma trận hiệp phương sai được ước lượng **E** thường được thực hiện trong độ phân giải phổ/thời gian của các tham số SAOC, tức là, đối với mỗi (l,m) , do đó ma trận hiệp phương sai được ước lượng có thể được viết là $E^{l,m}$. Ma trận hiệp phương sai được ước lượng $E^{l,m}$ có kích thước $N \times N$ với các hệ số của nó được định nghĩa là:

$$e_{i,j}^{l,m} = \sqrt{OLD_i^{l,m} OLD_j^{l,m} IOC_{i,j}^{l,m}}.$$

Do đó, ma trận $E^{l,m}$ với

$$E^{l,m} = \begin{pmatrix} e_{1,1}^{l,m} & \cdots & e_{1,N}^{l,m} \\ \vdots & \ddots & \vdots \\ e_{N,1}^{l,m} & \cdots & e_{N,N}^{l,m} \end{pmatrix}$$

có các chênh lệch mức đối tượng dọc theo đường chéo của ma trận, cụ thể, $e_{i,j}^{l,m} = OLD_i^{l,m}$ đối với $i=j$, vì $OLD_i^{l,m} = OLD_j^{l,m}$ và $IOC_{i,j}^{l,m} = 1$ đối với $i=j$. Phía ngoài đường chéo của ma trận, ma trận hiệp phương sai được ước lượng E có các hệ số ma trận biểu diễn số trung bình nhân của các chênh lệch mức đối tượng của các đối tượng i và j , tương ứng, có trọng số với phép đo tương quan chéo giữa đối tượng $IOC_{i,j}^{l,m}$.

Fig.3 hiển thị nguyên tắc khả thi của việc thực hiện ví dụ của bộ ước lượng thông tin phụ (Side Information Estimator - SIE) như là một phần của bộ mã hóa SAOC 10. Bộ mã hóa SAOC 10 bao gồm bộ trộn 16 và bộ ước lượng thông tin phụ SIE. SIE theo khái niệm gồm có hai môđun: Một môđun để tính toán phép biểu diễn t/f dựa trên thời gian ngắn (ví dụ, STFT hoặc QMF) của mỗi tín hiệu. Phép biểu diễn t/f thời gian ngắn được tính toán được đưa vào môđun thứ hai, môđun ước lượng thông tin chọn lọc t/f (t/f-selective Side Information Estimation module - t/f-SIE). Môđun t/f-SIE tính toán thông tin phụ cho mỗi ô t/f. Trong các triển khai SAOC hiện tại, phép biến đổi thời gian/tần số được cố định và giống nhau đối với tất cả các đối tượng âm thanh s_1 đến s_N . Hơn nữa, các tham số SAOC được xác định trên các khung SAOC mà các khung này là giống nhau đối với tất cả các đối tượng âm thanh và có cùng độ phân giải thời gian/tần số cho tất cả các đối tượng âm thanh s_1 đến s_N , do đó bỏ qua các nhu cầu chỉ rõ đối tượng để phân giải thời gian mịn trong một số trường hợp hoặc độ phân giải phổ mịn trong các trường hợp khác.

Một vài hạn chế của khái niệm SAOC hiện tại được mô tả ngay đây: Để giữ lượng dữ liệu kết hợp với thông tin phụ tương đối nhỏ, thông tin phụ của các đối tượng âm thanh khác nhau được xác định bằng cách thô được ưu tiên cho các vùng thời gian/tần số mà mở rộng một vài khe thời gian và vài băng phụ (lai) của các tín hiệu đầu vào tương ứng với các đối tượng âm thanh. Như đã nói ở trên, hiệu quả phân tách được quan sát tại phía bộ giải mã có thể gần tối ưu nếu phép biểu diễn t/f được sử dụng không thích nghi với các đặc tính thời gian và phổ của tín hiệu đối tượng để phân tách

từ tín hiệu trộn (tín hiệu trộn giảm) trong mỗi khối xử lý (tức là, vùng t/f hoặc ô t/f). Các thông tin phụ cho các phần âm thanh của đối tượng âm thanh và phân nhất thời của đối tượng âm thanh được xác định và áp dụng vào sự chia ô thời gian/tần số tương tự, không liên quan đến các đặc tính của đối tượng hiện tại. Điều này thường dẫn tới thông tin phụ cho các phần âm chủ yếu của đối tượng âm thanh được xác định tại độ phân giải phổ mà có phần quá thô, và thông tin phụ này cũng cho các phần theo thời gian của đối tượng âm thanh được xác định tại độ phân giải thời gian mà có phần quá thô. Tương tự, việc áp dụng các thông tin phụ không thích ứng trong bộ giải mã dẫn tới kết quả phân tách đối tượng gần tối ưu mà bị suy yếu bởi nhiễu xuyên âm dưới dạng, ví dụ, độ thô phổ và/hoặc âm vọng trước và âm vọng sau.

Để cải thiện hiệu suất phân tách tại phía bộ giải mã, cần phải cho phép bộ giải mã hoặc phương pháp tương ứng để giải mã để thích ứng một cách riêng lẻ phép biểu diễn t/f được sử dụng để xử lý các tín hiệu đầu vào bộ giải mã ("thông tin phụ và trộn giảm") theo các đặc tính của tín hiệu mục tiêu mong muốn được phân tách. Với mỗi tín hiệu mục tiêu (đối tượng) phép biểu diễn phù hợp nhất được lựa chọn một cách riêng lẻ cho việc xử lý và phân tách, ví dụ, không thuộc tập hợp các phép biểu diễn sẵn có. Bộ giải mã qua đó được điều khiển bởi các thông tin phụ mà truyền tín hiệu của phép biểu diễn t/f để được sử dụng cho mỗi đối tượng riêng lẻ tại khoảng thời gian và vùng phổ cho sẵn. Thông tin này được tính toán ở bộ mã hóa và được truyền tải cùng với các thông tin phụ đã được truyền trong SAOC.

- Sáng chế đề cập đến bộ ước lượng thông tin phụ nâng cao (Enhanced Side Information Estimator - E-SIE) tại bộ mã hóa để tính toán các thông tin phụ được làm giàu bởi các thông tin mà chỉ ra phép biểu diễn t/f riêng lẻ phù hợp nhất cho mỗi tín hiệu đối tượng.
- Sáng chế còn đề cập đến bộ tách đối tượng nâng cao (ảo) ((virtual) Enhanced Object Separator - E-OS) tại đầu nhận cuối. Bộ tách đối tượng nâng cao khai thác các thông tin bổ sung mà truyền tín hiệu phép biểu diễn t/f thực mà sau đó được sử dụng cho việc ước lượng mỗi đối tượng.

E-SIE có thể bao gồm hai mô đun. Một mô đun tính toán mỗi tín hiệu đối tượng lên tới H phép biểu diễn t/f, mà khác nhau về độ phân giải theo thời gian và phổ và đáp

ứng các yêu cầu sau: các vùng thời gian/tần số $R(t_R, f_R)$ có thể được xác định mà nội dung tín hiệu trong các vùng trên có thể được mô tả bằng bất cứ phép H biểu diễn t/f nào. Fig.5 mô tả khái niệm về ví dụ của H phép biểu diễn t/f và thể hiện vùng t/f $R(t_R, f_R)$ biểu diễn bởi hai phép biểu diễn t/f khác nhau. Nội dung tín hiệu trong vùng t/f $R(t_R, f_R)$ có thể được biểu diễn với độ phân giải phổ cao, nhưng độ phân giải thời gian thấp (phép biểu diễn t/f #1), với độ phân giải thời gian cao, nhưng độ phân giải phổ thấp (phép biểu diễn t/f #2), hoặc với một vài tổ hợp khác của các độ phân giải thời gian và phổ (phép biểu diễn t/f # H). Số lượng các phép biểu diễn t/f khả thi là không giới hạn.

Theo đó, bộ mã hóa âm thanh để mã hóa các tín hiệu đối tượng âm thanh s_i thành tín hiệu trộn giảm X và thông tin phụ PSI được cung cấp. Bộ mã hóa âm thanh bao gồm bộ tách thông tin phụ được cải tiến E-SIE được mô tả sơ lược trên Fig.4. Bộ ước lượng thông tin phụ nâng cao E-SIE bao gồm bộ biến đổi thời gian/tần số 52 được tạo cấu hình để biến đổi các tín hiệu đối tượng âm thanh s_i ít nhất thành nhiều tín hiệu được biến đổi tương ứng thứ nhất $s_{1,1}(t,f) \dots s_{N,a}(t,f)$ sử dụng ít nhất độ phân giải thời gian/tần số thứ nhất TFR_1 (sự rời rạc thời gian/tần số thứ nhất) và nhiều biến đổi tương ứng thứ hai $s_{1,2}(t,f) \dots s_{N,2}(t,f)$ sử dụng độ phân giải thời gian/tần số thứ hai TFR_2 (sự rời rạc thời gian/tần số thứ hai). Trong một vài phương án, bộ biến đổi thời gian-tần số 52 có thể được tạo cấu hình để sử dụng nhiều hơn hai độ phân giải thời gian/tần số TFR_1 đến TFR_H . Bộ ước lượng thông tin phụ nâng cao (E-SIE) còn bao gồm phép tính toán thông tin phụ và môđun lựa chọn (selection module - SI-SC) 54. Phép tính toán thông tin phụ và môđun lựa chọn bao gồm (xem Fig.6) bộ xác định thông tin phụ (side information determiner - t/f-SIE) hoặc các bộ xác định thông tin phụ 55-1...55-H được tạo cấu hình để xác định ít nhất thông tin phụ thứ nhất cho nhiều biến đổi tương ứng thứ nhất $s_{1,1}(t,f) \dots s_{N,1}(t,f)$ và thông tin phụ thứ hai cho nhiều biến đổi tương ứng $s_{1,2}(t,f) \dots s_{N,2}(t,f)$, thông tin phụ thứ nhất và thông tin phụ thứ hai chỉ ra mối liên hệ của các tín hiệu đối tượng âm thanh s_i với nhau lần lượt trong độ phân giải thời gian/tần số TFR_1 , TFR_2 , trong vùng thời gian/tần số $R(t_R, f_R)$. Mối liên hệ giữa các tín hiệu âm thanh s_i với nhau có thể, ví dụ, liên quan tới các năng lượng tương đối của các tín hiệu âm thanh trong các băng tần khác nhau và/hoặc mức độ tương quan giữa các tín hiệu âm thanh. Phép tính toán thông tin phụ và môđun lựa chọn 54 còn bao gồm bộ chọn

thông tin phụ (side information selector - SI-AS) 56 được tạo cấu hình để lựa chọn, cho mỗi tín hiệu đối tượng âm thanh s_i , một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thông tin phụ thứ hai trên tiêu chí phù hợp biểu thị sự phù hợp cho ít nhất độ phân giải thời gian/tần số thứ nhất và thứ hai để biểu diễn tín hiệu đối tượng âm thanh s_i trong miền thời gian/tần số. Thông tin phụ chỉ rõ đối tượng sau đó được chèn vào đầu ra PSI thông tin phụ bởi bộ mã hóa âm thanh.

Lưu ý rằng việc nhóm các mặt phẳng t/f thành các vùng t/f $R(t_R, f_R)$ có thể không cần thiết phải đặt cách đều nhau, như Fig.5 biểu thị. Việc nhóm thành các vùng $R(t_R, f_R)$ có thể, ví dụ, không đồng nhất để được thích ứng về mặt cảm giác. Việc nhóm này có thể phù hợp với các sơ đồ mã hóa đối tượng âm thanh hiện có, như SAOC, để cho phép sơ đồ mã hóa tương thích theo chiều ngược với các khả năng ước lượng đối tượng nâng cao.

Sự thích ứng của độ phân giải t/f không chỉ giới hạn trong việc định rõ ô tham số khác nhau cho các đối tượng khác nhau, mà việc chuyển đổi sơ đồ SAOC dựa trên (tức là, thường được trình bày bởi độ phân giải thời gian/tần số thông thường được sử dụng trong tình trạng hệ thống kỹ thuật này cho việc xử lý SAOC) có thể được điều chỉnh để phù hợp hơn với các đối tượng mục tiêu riêng biệt. Điều này đặc biệt hữu ích, ví dụ, khi độ phân giải phổ cao hơn được cung cấp bằng biến đổi thông thường sơ đồ SAOC dựa trên là cần thiết. Trong trường hợp ví dụ của MPEG SAOC, độ phân giải thô bị giới hạn bởi độ phân giải (thường) hoặc giàn QMF (lai). Bằng việc xử lý theo sáng chế, có thể tăng độ phân giải phổ, nhưng như một sự đánh đổi, một vài độ phân giải thời gian bị mất đi trong quy trình. Điều này được thực hiện bằng cách sử dụng cái gọi là phép biến đổi thu phóng (phổ) được áp dụng vào đầu ra của giàn bộ lọc đầu tiên. Về mặt khái niệm, một số mẫu đầu ra của giàn bộ lọc liên tiếp được xử lý như một tín hiệu miền thời gian và phép biến đổi thứ hai được áp dụng vào chúng để thu được số lượng tương ứng của mẫu phổ (với chỉ một khe thời gian). Biến đổi thu phóng có thể dựa trên giàn bộ lọc (tương tự như giai đoạn bộ lọc lai trong MPEG SAOC), hoặc phép biến đổi trên khối như DFT hoặc phép biến đổi cosin rời rạc cải biên (Complex Modified Discrete Cosine Transform - CMDCT). Theo cách tương tự, có thể tăng độ phân giải thời gian bằng độ phân giải phổ (phép biến đổi thu phóng theo thời gian):

Một số đầu ra đồng thời của một vài bộ lọc của giàn QMF (lai) được lấy mẫu làm tín hiệu miền tần số và phép biến đổi thứ hai được áp dụng cho chúng để thu được số lượng tương ứng của các mẫu thử theo thời gian (chỉ với một băng phổ rộng bao gồm các dãy phổ của vài bộ lọc).

Với mỗi đối tượng, H phép biểu diễn t/f được nạp cùng với các tham số trộn trong môđun thứ hai, phép tính toán thông tin phụ và môđun lựa chọn SI-CS. Bộ xác định môđun SI-CS, cho mỗi tín hiệu đối tượng, mà các H phép biểu diễn t/f cần được sử dụng cho miền t/f $R(t_R, f_R)$ tại bộ giải mã để ước lượng tín hiệu đối tượng. Fig.6 trình bày chi tiết nguyên lý của môđun SI-CS.

Đối với mỗi H phép biểu diễn t/f khác nhau, thông tin phụ (side information - SI) tương ứng được tính toán. Ví dụ, môđun t/f-SIE trong SAOC có thể được sử dụng. H dữ liệu thông tin phụ được tính toán và nhập vào môđun đánh giá và lựa chọn thông tin phụ (Side Information Assessment and Selection module - SI-AS). Đối với mỗi tín hiệu đối tượng, môđun SI-AS xác định phép biểu diễn phù hợp nhất với mỗi vùng t/f để ước lượng tín hiệu đối tượng từ hỗn hợp tín hiệu.

Ngoài các tham số trộn cảnh thông thường, các đầu ra tín hiệu SI-AS, với mỗi tín hiệu đối tượng và mỗi vùng t/f, thông tin phụ liên quan tới phép biểu diễn t/f được lựa chọn một cách riêng biệt. Tham số bổ sung biểu thị phép biểu diễn t/f tương ứng, có thể được xuất ra.

Hai phương pháp để lựa chọn phép biểu diễn t/f phù hợp nhất cho mỗi tín hiệu đối tượng được đưa ra:

1. SI-AS dựa trên ước lượng nguồn: Mỗi tín hiệu đối tượng được ước lượng từ hỗn hợp tín hiệu sử dụng dữ liệu thông tin phụ được tính toán dựa trên H phép biểu diễn t/f mang lại sự H ước lượng nguồn cho mỗi tín hiệu đối tượng. Với mỗi đối tượng, chất lượng ước lượng trong mỗi vùng t/f $R(t_R, f_R)$ được đánh giá cho mỗi H phép biểu diễn t/f bằng cách đo hiệu suất ước lượng nguồn. Một ví dụ đơn giản của phép đo này là tỉ lệ tín hiệu trên méo (Signal to Distortion Ratio - SDR) đạt được. Phức tạp hơn, các phép đo cảm giác có thể được sử dụng. Lưu ý rằng SDR có thể thực hiện hiệu quả chỉ dựa trên các thông tin phụ tham số được xác định trong SAOC mà không biết về các tín hiệu đối tượng ban đầu hoặc hỗn hợp tín hiệu. Khái niệm về sự

ước lượng tham số của SDR đối với trường hợp sự ước lượng đối tượng dựa trên SAOC sẽ được mô tả dưới đây. Đối với mỗi vùng t/f $R(t_R, f_R)$, phép biểu diễn t/f mà sinh ra SDR cao nhất được chọn cho việc truyền dẫn và ước lượng thông tin phụ, và để ước lượng tín hiệu đối tượng bên phía bộ giải mã.

2. SI-AS dựa trên sự phân tích H phép biểu diễn t/f : Riêng đối với từng đối tượng, sự rải rác của mỗi H phép biểu diễn tín hiệu đối tượng được xác định. Nói cách khác, nó được đánh giá năng lượng của tín hiệu đối tượng tốt như thế nào trong mỗi phép biểu diễn khác nhau được tập trung vào một vài trị số hoặc trải rộng trên tất cả các trị số. Phép biểu diễn t/f , mà biểu diễn đối tượng một cách rải rác nhất, được lựa chọn. Sự rải rác của các phép biểu diễn tín hiệu có thể được đánh giá, ví dụ, với các phép đo biểu thị đặc điểm độ phẳng và điểm cực đại của các phép biểu diễn tín hiệu. Phép đo độ phẳng phổ (Spectral-Flatness Measure - SFM), hệ số đỉnh (Crest-Factor - CF) và quy tắc $L0$ là các ví dụ của các phép đo này. Theo phương án này, tiêu chí phù hợp có thể dựa trên sự rải rác của ít nhất phép biểu diễn thời gian/tần số thứ nhất và phép biểu diễn thời gian/tần số thứ hai (và có thể còn các phép biểu diễn thời gian/tần số) của đối tượng âm thanh đã cho. Bộ chọn lọc thông tin phụ (side information selector - SI-AS) được tạo cấu hình để chọn lọc thông tin phụ giữa ít nhất thông tin phụ thứ nhất và thông tin phụ thứ hai mà tương ứng với phép biểu diễn thời gian/tần số mà biểu diễn tín hiệu đối tượng âm thanh s_i rải rác nhất.

Sự ước lượng tham số của SDR cho trường hợp ước lượng đối tượng dựa trên SAOC được mô tả:

Các ký hiệu:

S	Ma trận của N tín hiệu đối tượng âm thanh ban đầu
X	Ma trận của M tín hiệu hỗn hợp
$D \in^{o} M \times N$	Ma trận trộn giảm
$X=DS$	Phép tính của cảnh trộn giảm
S_{est}	Ma trận của N tín hiệu đối tượng âm thanh được ước lượng

Trong SAOC, các tín hiệu đối tượng được ước lượng một cách khái niệm từ các tín hiệu hỗn hợp với công thức:

$$S_{est} = ED^*(DED^*)^{-1}X \quad \text{với } E=SS^*$$

Thay X bằng DS đưa ra:

$$S_{est} = ED^*(DED^*)^{-1}DS = TS$$

Năng lượng của các phần tín hiệu đối tượng ban đầu trong các tín hiệu đối tượng có thể được tính toán là:

$$E_{est} = S_{est}S_{est}^* = TSS^*T^* = TET^*$$

Các giới hạn méo trong tín hiệu được ước lượng có thể được tính toán bởi:

$E_{dist} = diag(E) - E_{est}$, với $diag(E)$ là ma trận đường chéo chứa năng lượng của các tín hiệu đối tượng ban đầu. SDR có thể được tính toán bằng liên kết $diag(E)$ với E_{dist} . Để ước lượng SDR theo cách tương đối với năng lượng nguồn mục tiêu trong vùng t/f $R(t_R, f_R)$ nhất định, phép tính toán năng lượng biến dạng phải được thực hiện trên mỗi ô t/f được xử lý trong vùng $R(t_R, f_R)$, và năng lượng mục tiêu và năng lượng biến dạng được tích lũy trên các ô t/f trong vùng t/f $R(t_R, f_R)$.

Do đó, các tiêu chí phù hợp có thể dựa trên sự ước lượng nguồn. Trong trường hợp này bộ chọn lọc thông tin phụ (side information selector - SI-AS) 56 có thể còn bao gồm bộ ước lượng nguồn được tạo cấu hình để ước lượng ít nhất một tín hiệu đối tượng âm thanh được chọn của các tín hiệu đối tượng âm thanh s_i sử dụng tín hiệu trộn giảm X và ít nhất một thông tin thứ nhất và thông tin thứ hai lần lượt tương ứng với độ phân giải thời gian/tần số thứ nhất và thứ hai TFR1, TFR2. Bộ ước lượng nguồn cung cấp ít nhất một tín hiệu đối tượng âm thanh được ước lượng thứ nhất $s_{i, estim1}$ và tín hiệu đối tượng âm thanh được ước lượng thứ hai $s_{i, estim2}$ (có thể lên tới H đối tượng âm thanh được ước lượng $s_{i, estim H}$). Bộ chọn lọc thông tin phụ 56 còn bao gồm bộ đánh giá chất lượng được tạo cấu hình để đánh giá chất lượng của ít nhất một tín hiệu đối tượng âm thanh được ước lượng thứ nhất $s_{i, estim1}$ và tín hiệu đối tượng âm thanh được ước lượng thứ hai $s_{i, estim2}$. Ngoài ra, bộ đánh giá chất lượng có thể được tạo cấu hình để đánh giá chất lượng của ít nhất một tín hiệu đối tượng âm thanh được ước lượng thứ nhất $s_{i, estim1}$ và tín hiệu đối tượng âm thanh được ước lượng thứ hai $s_{i, estim2}$ dựa trên nền tảng của tỉ lệ tín hiệu trên méo SDR như là phép đo hiệu suất ước lượng nguồn, tỉ

lệ tín hiệu trên méo SDR chỉ được quyết định dựa trên thông tin phụ PSI, cụ thể là ma trận hiệp phương sai được ước lượng $\mathbf{E}_{\text{est}}$.

Bộ mã hóa âm thanh theo một vài phương án còn có thể bao gồm một bộ xử lý tín hiệu trộn giảm được mà được tạo cấu hình để biến đổi tín hiệu trộn giảm X thành phép biểu diễn được lấy mẫu trong miền thời gian/tần số thành các khe thời gian và băng phụ (lai). Vùng thời gian/tần số $R(t_R, f_R)$ có thể mở rộng đến ít nhất hai mẫu của tín hiệu trộn giảm X . Độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h định rõ ít nhất một đối tượng âm thanh có thể mịn hơn vùng thời gian/tần số $R(t_R, f_R)$. Như đã nói ở trên, trong mỗi liên hệ với các nguyên lý bất định của phép biểu diễn thời gian/tần số độ phân giải phổ của tín hiệu có thể được tăng bởi của độ phân giải thời gian hoặc ngược lại. Mặc dù tín hiệu trộn giảm được gửi đi từ bộ mã hóa âm thanh tới bộ giải mã âm thanh thường được phân tích trong bộ giải mã âm thanh bằng phép biến đổi thời gian-tần số với độ phân giải thời gian/tần số được định trước cố định, bộ giải mã âm thanh có thể vẫn biến đổi tín hiệu trộn giảm được phân tích trong vùng thời gian/tần số dự tính $R(t_R, f_R)$ đối tượng riêng lẻ thành độ phân giải thời gian/tần số khác mà phù hợp hơn để tách đối tượng âm thanh được cho s_i từ hỗn hợp tín hiệu. Phép biến đổi của tín hiệu trộn giảm này ở bộ giải mã được gọi là biến đổi thu phóng trong tài liệu này. Phép biến đổi thu phóng có thể là phép biến đổi thu phóng thời gian hoặc phép biến đổi thu phóng phổ.

Giảm lượng thông tin phụ

Về nguyên tắc, trong các phương án đơn giản của hệ thống theo sáng chế, thông tin phụ cho đến H phép biểu diễn t/f phải được truyền cho từng đối tượng và từng vùng t/f $R(t_R, f_R)$ như là sự phân tách tại phía các bộ giải mã mà được thực hiện bằng cách chọn từ cho đến H phép biểu diễn t/f. Lượng dữ liệu lớn này có thể được cắt giảm rất nhiều mà không làm mất đi chất lượng cảm giác một cách đáng kể. Với mỗi đối tượng, nó là đủ để truyền cho mỗi vùng t/f $R(t_R, f_R)$ thông tin sau đây:

- Một tham số mô tả tổng thể/thô nội dung tín hiệu của đối tượng âm thanh trong vùng t/f $R(t_R, f_R)$, ví dụ, năng lượng tín hiệu trung bình của đối tượng trong vùng $R(t_R, f_R)$.

- Mô tả về cấu trúc mịn của đối tượng âm thanh. Mô tả này thu được từ phép biểu diễn t/f riêng lẻ đã được lựa chọn để ước lượng tối ưu đối tượng âm thanh từ hỗn hợp. Lưu ý rằng thông tin trên cấu trúc mịn có thể được mô tả hiệu quả bằng việc tham số hóa sự chênh lệch giữa các phép biểu diễn tín hiệu thô và cấu trúc mịn.
- Tín hiệu thông tin biểu thị phép biểu diễn t/f được sử dụng cho việc ước lượng đối tượng âm thanh.

Tại bộ giải mã, việc ước lượng của các đối tượng âm thanh mong muốn từ hỗn hợp tại bộ giải mã có thể được thực hiện như mô tả dưới đây trong vùng t/f $R(t_R, f_R)$.

- Phép biểu diễn t/f riêng lẻ như được chỉ ra bằng thông tin phụ bổ sung cho đối tượng âm thanh được tính toán.
- Để phân tách đối tượng âm thanh mong muốn, thông tin tín hiệu đối tượng (cấu trúc mịn) tương ứng được sử dụng.
- Với tất cả các đối tượng âm thanh còn lại, tức là, đối tượng âm thanh gây nhiễu mà phải được khử, thông tin tín hiệu đối tượng cấu trúc mịn được sử dụng nếu thông tin có sẵn cho phép biểu diễn t/f được chọn. Mặt khác, sự mô tả tín hiệu thô được sử dụng. Một tùy chọn khác là sử dụng thông tin tín hiệu đối tượng cấu trúc mịn sẵn có cho đối tượng âm thanh còn lại cụ thể và để làm gần đúng với phép biểu diễn t/f được lựa chọn, ví dụ, lấy trung bình thông tin tín hiệu đối tượng âm thanh cấu trúc mịn trong các vùng con của vùng t/f $R(t_R, f_R)$: Theo cách này, độ phân giải t/f không mịn bằng phép biểu diễn t/f được lựa chọn, nhưng vẫn mịn hơn phép biểu diễn t/f thô.

Bộ giải mã SAOC với phép ước lượng đối tượng âm thanh nâng cao

Fig.7 minh họa sơ lược dưới dạng dạng đồ sự giải mã SAOC bao gồm mô đun phân tách đối tượng (ảo) nâng cao (Enhanced (virtual) Object Separation - E-OS) và hiển thị nguyên lý trên ví dụ của bộ giải mã SAOC được cải thiện bao gồm bộ tách đối tượng nâng cao (ảo) (Enhanced (virtual) Object Separation - E-OS). Bộ giải mã SAOC được nhập cùng với hỗn hợp tín hiệu với thông tin phụ tham số nâng cao (Enhanced Parametric Side Information - E-PSI). E-PSI bao gồm thông tin trên các đối tượng âm thanh, các tham số hỗn hợp và thông tin bổ sung. Bằng thông tin phụ bổ sung này, nó

được tín hiệu hóa đến EOS ảo, trong đó phép biểu diễn t/f cần được sử dụng cho mỗi đối tượng $s_1 \dots s_N$ và mỗi vùng t/f $R(t_R, f_R)$. Đối với vùng t/f $R(t_R, f_R)$ cho sẵn, bộ phân tách đối tượng ước lượng mỗi đối tượng, sử dụng phép biểu diễn t/f riêng lẻ mà được truyền tín hiệu cho mỗi đối tượng trong thông tin phụ.

Fig.8 trình bày chi tiết khái niệm của mô đun E-OS. Với mỗi vùng t/f $R(t_R, f_R)$ cho sẵn, phép biểu diễn t/f riêng lẻ $\#h$ để tính toán P các tín hiệu trộn giảm được truyền tín hiệu bằng các mô đun truyền tín hiệu phép biểu diễn t/f 110 đến mô đun biến đổi t/f phức. Bộ phân tách đối tượng (ảo) 120 cố gắng một cách khái niệm để ước lượng thông tin nguồn s_n , dựa trên phép biến đổi t/f $\#h$ chỉ ra bởi thông tin phụ bổ sung. Bộ phân tách đối tượng (ảo) khai thác thông tin trên cấu trúc mịn của các đối tượng, nếu truyền cho phép biến đổi t/f $\#h$ được chỉ ra, và sử dụng mô tả thô được truyền của các tín hiệu nguồn khác. Lưu ý rằng số lớn nhất có thể của các phép biểu diễn t/f khác nhau được tính toán được cho mỗi vùng t/f $R(t_R, f_R)$ là H . Mô đun phép biến đổi thời gian/tần số phức có thể được tạo cấu hình để thực hiện các biến đổi thu phóng nói trên của P (các) tín hiệu trộn giảm.

Fig.9 thể hiện sơ đồ khối dạng giản lược của bộ giải mã âm thanh để giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm X và thông tin phụ PSI. Thông tin phụ PSI bao gồm thông tin phụ chỉ rõ đối tượng PSI_i với $i=1 \dots N$ cho ít nhất một đối tượng âm thanh s_i trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$. Thông tin phụ PSI cũng bao gồm thông tin độ phân giải thời gian/tần số $TFRI_i$ với $i=1 \dots NTF$. Biến số NTF biểu thị số đối tượng âm thanh mà thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng được cung cấp và $NTF \leq N$. Thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng $TFRI_i$ có thể được đề cập tới như thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng. Cụ thể, thuật ngữ "độ phân giải thời gian/tần số" không nên được hiểu nhất thiết là sự rời rạc đồng đều của miền thời gian/tần số, mà cũng có thể đề cập tới sự rời rạc không đồng đều trong ô t/f hoặc trên tất cả các ô t/f của phổ toàn băng. Thông thường và được ưu tiên hơn, độ phân giải thời gian/tần số được chọn sao cho một trong hai chiều của ô t/f cho sẵn có độ phân giải mịn và chiều khác có độ phân giải thấp, ví dụ, với các tín hiệu nhất thời của chiều thời gian có độ phân giải mịn và độ phân giải phổ là thô, ngược lại với các tín hiệu dừng độ phân giải phổ là mịn và

chiều thời gian có độ phân giải kém. Thông tin độ phân giải thời gian/tần số TFR_i biểu thị của độ phân giải thời gian/tần số của chỉ rõ đối tượng TFR_h ($h=1...H$) của thông tin phụ chỉ rõ đối tượng PSI_i cho ít nhất một đối tượng âm thanh s_i trong ít nhất một vùng thời gian tần số $R(t_R, f_R)$. Bộ giải mã âm thanh bao gồm bộ xác định độ phân giải thời gian/tần số 110, được tạo cấu hình để xác định thông tin phụ độ phân giải thời gian/tần số chỉ rõ đối tượng từ thông tin phụ TFR_i từ thông tin phụ PSI cho ít nhất một đối tượng âm thanh s_i . Bộ giải mã âm thanh bao gồm bộ phân tách đối tượng 120 được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh s_i từ tín hiệu trộn giảm X sử dụng thông tin phụ chỉ rõ đối tượng PSI_i tương ứng với độ phân giải thời gian tần số chỉ rõ đối tượng TFR_i . Nó có nghĩa là thông tin phụ chỉ rõ đối tượng PSI_i có độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i đặc trưng bởi thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i , và độ phân giải thời gian/tần số chỉ rõ đối tượng được chú ý tới khi thực hiện sự phân tách đối tượng bằng bộ phân tách đối tượng 120.

Thông tin phụ chỉ rõ đối tượng (PSI_i) có thể bao gồm thông tin phụ chỉ rõ đối tượng cấu trúc hạt mịn $fsl_i^{\eta, \kappa}$, $fsc_{i,j}^{\eta, \kappa}$ cho ít nhất một đối tượng âm thanh s_i trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$. Thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{\eta, \kappa}$ có thể là thông tin mức cấu trúc mịn mô tả (ví dụ, năng lượng tín hiệu, cường độ tín hiệu, biên độ, v.v của đối tượng âm thanh) biến thiên trong vùng thời gian/tần số $R(t_R, f_R)$. Thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsc_{i,j}^{\eta, \kappa}$ có thể là thông tin tương quan liên đối tượng của các đối tượng âm thanh lần lượt là i và j . Ở đây, thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{\eta, \kappa}$, $fsc_{i,j}^{\eta, \kappa}$ được xác định trên lưới thời gian/tần số theo độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i , với các khe thời gian cấu trúc mịn η và các băng phụ (lai) cấu trúc mịn κ . Vấn đề này sẽ được mô tả bên dưới trong ngữ cảnh của Fig.12. Hiện tại, ít nhất ba trường hợp có thể được phân biệt:

a) Độ phân giải thời gian/tần số TFR_i tương ứng với độ kết hạt của các khe thời gian QMF và các băng phụ (lai). Trong trường hợp này $\eta=n$ và $\kappa=k$.

b) Thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i chỉ ra phép thu phóng phổ được thực hiện trong vùng thời gian/tần số $R(t_R, f_R)$ hoặc một phần của nó. Trong trường hợp này, mỗi băng phụ (lai) k được chia nhỏ thành hai hoặc

nhiều băng phụ (lai) cấu trúc mịn $\kappa_k, \kappa_{k+1}, \dots$ để độ phân giải phổ được tăng lên. Nói cách khác, băng phụ (lai) cấu trúc mịn $\kappa_k, \kappa_{k+1}, \dots$ là các phân đoạn của băng phụ (lai) ban đầu. Trong chuyển đổi, độ phân giải thời gian bị giảm, do tính bất định thời gian/tần số. Do đó, khe thời gian cấu trúc mịn η bao gồm hai hoặc nhiều hơn hai khe thời gian $n, n+1, \dots$.

c) Thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFRI_i biểu thị phép thu phóng thời gian được thực hiện trong vùng thời gian/tần số $R(t_R, f_R)$ hoặc một phần của nó. Trong trường hợp này, mỗi khe thời gian n được chia nhỏ thành hai hoặc nhiều hơn hai khe thời gian cấu trúc mịn $\eta_n, \eta_{n+1}, \dots$ để độ phân giải thời gian được tăng lên. Nói cách khác, khe thời gian cấu trúc mịn $\eta_n, \eta_{n+1}, \dots$ là các phần của khe thời gian n . Trong chuyển đổi, độ phân giải phổ bị giảm, do tính bất định thời gian/tần số. Do đó, băng phụ (lai) cấu trúc mịn κ bao gồm hai hoặc nhiều hơn hai các băng phụ (lai) $k, k+1, \dots$.

Thông tin phụ còn bao gồm thông tin phụ chỉ rõ đối tượng thô OLD_i , $\text{IOC}_{i,j}$, và/hoặc mức năng lượng tuyệt đối NRG_i cho ít nhất một đối tượng âm thanh s_i trong vùng thời gian/tần số $R(t_R, f_R)$ được xem xét. Thông tin phụ chỉ rõ đối tượng thô OLD_i , $\text{IOC}_{i,j}$, và/hoặc NRG_i là bất biến trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$.

Fig.10 thể hiện sơ đồ khối dạng giản lược của bộ giải mã âm thanh mà được tạo cấu hình để nhận và xử lý thông tin phụ cho tất cả N đối tượng âm thanh trong tất cả H phép biểu diễn t/f trong một ô thời gian/tần số $R(t_R, f_R)$. Dựa vào số N của các đối tượng âm thanh và số H của các phép biểu diễn t/f , lượng thông tin phụ được truyền và lưu trữ cho mỗi vùng t/f $R(t_R, f_R)$ có thể trở nên tương đối lớn để các khái niệm được thể hiện trên Fig.10 có nhiều khả năng sẽ được sử dụng cho các tình huống với một số lượng nhỏ các đối tượng âm thanh và các phép biểu diễn t/f khác nhau. Tuy nhiên, các ví dụ minh họa trên Fig.10 đề xuất một quan niệm về một số các nguyên tắc của việc sử dụng các phép biểu diễn t/f chỉ rõ đối tượng khác nhau cho các đối tượng âm thanh khác nhau.

Tóm lại, theo phương án thể hiện trên Fig.10 toàn bộ tập hợp các tham số (cụ thể là OLD and IOC) được xác định và truyền/lưu trữ cho tất cả H phép biểu diễn t/f được quan tâm. Ngoài ra, thông tin phụ biểu thị cho mỗi đối tượng âm thanh mà cụ thể phép

biểu diễn t/f đối tượng âm thanh này nên được tách và tổng hợp. Trong bộ giải mã âm thanh, việc tái thiết đối tượng $\hat{S}_h$ trong tất cả các phép biểu diễn t/f h được thực hiện. Các đối tượng âm thanh cuối cùng sau đó được ghép lại, theo thời gian và tần số, từ các ô chỉ rõ đối tượng, hoặc các vùng t/f, mà được tạo ra sử dụng độ phân giải t/f cụ thể được truyền tín hiệu trong thông tin phụ cho đối tượng âm thanh và các ô được quan tâm.

Tín hiệu trộn giảm X được cung cấp cho các bộ phân tách đối tượng 120_1 đến 120_H . Mỗi bộ phân tách đối tượng 120_1 đến 120_H được tạo cấu hình để thực hiện các tác vụ phân tách cho mỗi phép biểu diễn t/f cụ thể. Nhằm mục đích này, mỗi bộ phân tách đối tượng 120_1 đến 120_H còn nhận thông tin phụ của N đối tượng âm thanh s_1 đến s_N khác nhau trong phép biểu diễn t/f cụ thể mà bộ phân tách được kết hợp cùng. Lưu ý rằng Fig.10 thể hiện nhiều H bộ phân tách đối tượng cho các mục đích minh họa. Trong một phương án thay thế, H tác vụ phân tách với mỗi vùng t/f $R(t_R, f_R)$ có thể được thực hiện bằng ít bộ phân tách đối tượng hơn, hoặc thậm chí bằng một bộ phân tách đối tượng. Theo các phương án khả thi khác, các tác vụ phân tách có thể được thể hiện trên bộ xử lý đa chức năng hoặc trên bộ xử lý đa lõi như các chuỗi khác nhau. Một số tác vụ phân tách có tính điện toán chuyên sâu hơn các tác vụ khác, tùy thuộc vào độ mịn của phép biểu diễn tương ứng t/f. Đối với mỗi vùng t/f $R(t_R, f_R)$ $N \times H$ bộ của thông tin phụ được cung cấp tới bộ giải mã âm thanh.

Các bộ phân tách đối tượng 120_1 đến 120_H cung cấp $N \times H$ đối tượng âm thanh được ước lượng $\hat{s}_{1,1} \dots \hat{s}_{N,H}$ mà có thể được nhập vào bộ chuyển đổi độ phân giải bổ sung 130 để mang đến các đối tượng âm thanh được phân tách ước lượng $\hat{s}_{1,1} \dots \hat{s}_{N,H}$ đến các phép biểu diễn t/f thông thường, nếu đây không phải trường hợp đã có. Thông thường, độ phân giải hoặc phép biểu diễn t/f thông thường có thể là độ phân giải t/f đúng của giàn bộ lọc hoặc phép biến đổi việc xử lý chung của các tín hiệu âm thanh được dựa trên, tức là, trong trường hợp MPEG SAOC độ phân giải thông thường là độ mịn của các khe thời gian QMF và các băng phụ (lai). Đối với mục đích minh họa có thể được giả định rằng các đối tượng âm thanh được ước lượng được lưu trữ tạm thời trong ma trận 140. Trong thực hiện thực tế, các đối tượng âm thanh được phân tách được ước lượng sẽ không được sử dụng sau này có thể được loại bỏ ngay lập tức hoặc

thậm chí không được tính toán ngay từ đầu. Mỗi hàng của ma trận 140 bao gồm H ước lượng khác nhau của cùng một đối tượng âm thanh, tức là, đối tượng âm thanh được phân tách được ước lượng xác định dựa trên H phép biểu diễn t/f khác nhau. Phần giữa của ma trận 140 được biểu thị sơ lược bằng lưới. Mỗi yếu tố ma trận $\hat{s}_{1,1} \dots \hat{s}_{N,H}$ tương ứng với tín hiệu âm thanh của đối tượng âm thanh được phân tách được ước lượng. Nói cách khác, mỗi phần tử ma trận bao gồm các mẫu khe thời gian/băng phụ trong vùng t/f đích $R(t_R, f_R)$ (ví dụ, 7 khe thời gian x 3 băng phụ = 21 mẫu thời gian/băng phụ trong mỗi ví dụ của Fig.11).

Bộ giải mã âm thanh còn được tạo cấu hình để nhận thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng $TFRI_1$ đến $TFRI_N$ cho mỗi đối tượng âm thanh khác nhau và cho vùng t/f hiện thời $R(t_R, f_R)$. Đối với mỗi đối tượng âm thanh i , thông tin độ phân giải thời gian/tần số $TFRI_i$ biểu thị các đối tượng âm thanh được phân tách $\hat{s}_{i,1} \dots \hat{s}_{i,H}$ cần được sử dụng để tái tạo gần như các đối tượng âm thanh ban đầu. Các thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng thường được xác định bởi bộ mã hóa và cung cấp cho các bộ giải mã như là một phần của thông tin phụ. Trên Fig.10, các hộp nét đứt và các đường gạch chéo trong ma trận 140 biểu thị phép biểu diễn nào đã được lựa chọn cho mỗi đối tượng âm thanh. Sự lựa chọn này được tạo bởi bộ lựa chọn 112 mà nhận thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng $TFRI_1 \dots TFRI_N$.

Bộ chọn lọc 112 xuất N tín hiệu âm thanh được lựa chọn có thể còn được xử lý. Ví dụ, N tín hiệu âm thanh có thể được cung cấp tới bộ kết xuất 150 được tạo cấu hình để kết xuất các tín hiệu đối tượng âm thanh được chọn đến bộ cài đặt loa phát thanh sẵn có, ví dụ lập thể hoặc bộ cài đặt âm thanh 5.1. Nhằm mục đích này, bộ kết xuất 150 có thể nhận được thông tin kết xuất cài sẵn và/hoặc thông tin kết xuất người dùng mà mô tả tín hiệu âm thanh của đối tượng âm thanh được phân tách cần được phân phối đến các loa phát thanh có sẵn như thế nào. Bộ kết xuất 150 là tùy chọn và các đối tượng âm thanh được phân tách $\hat{s}_{i,1} \dots \hat{s}_{i,H}$ tại đầu ra tín hiệu của bộ lựa chọn 112 có thể được sử dụng và xử lý trực tiếp. Trong các phương án thay thế, bộ kết xuất 150 có thể được đặt tới các cài đặt cực trị như "chế độ solo" hoặc "chế độ karaoke". Trong chế độ solo, một đối tượng âm thanh đơn được ước lượng được chọn để kết xuất thành tín hiệu đầu ra. Trong chế độ karaoke, chỉ một đối tượng âm thanh được ước lượng được

chọn để kết xuất thành tín hiệu đầu ra. Thông thường, phần âm chính không được kết xuất, còn phần âm đệm được kết xuất. Cả hai chế độ đều yêu cầu cao về hiệu suất phân tách, một ít nhiễu xuyên âm có thể cảm thấy được.

Fig.11 minh họa sơ lược thông tin phụ cấu trúc mịn $fsI_i^{m,k}$ và thông tin phụ thô cho mỗi đối tượng âm thanh i có thể được tổ chức. Phần phía trên của Fig.11 minh họa một phần của miền thời gian/tần số mà được lấy mẫu theo các khe thời gian (thông thường được chỉ ra bởi chỉ mục n trong tài liệu và cụ thể là các tiêu chuẩn ISO/IEC liên quan đến mã hóa âm thanh) và băng phụ (lai) (thông thường được nhận ra bởi chỉ mục k trong tài liệu). Miền thời gian/tần số cũng được chia nhỏ thành các vùng thời gian/tần số khác nhau (được chỉ ra bằng hình vẽ các đoạn thẳng dày đứt nét trên Fig.11). Thông thường một miền t/f bao gồm một vài mẫu khe thời gian/băng phụ. Một vùng t/f $R(t_R, f_R)$ sẽ đóng vai trò như một ví dụ đại diện cho các vùng t/f khác. Vùng t/f được minh họa xem xét $R(t_R, f_R)$ mở rộng đến bảy khe thời gian n đến $n+6$ và ba băng (lai) k đến $k+2$ và do đó bao gồm 21 mẫu khe thời gian/băng phụ. Giả sử hai đối tượng âm thanh khác nhau là i và j . Đối tượng âm thanh i có thể có đặc tính âm thanh trong vùng t/f $R(t_R, f_R)$, trong khi đó đối tượng âm thanh j có thể có đặc tính nhất thời đáng kể trong vùng t/f $R(t_R, f_R)$. Để biểu diễn một cách đầy đủ hơn các đặc tính khác nhau của các đối tượng âm thanh i và j , vùng t/f $R(t_R, f_R)$ có thể còn được chia nhỏ thành nhiều hướng phổ cho đối tượng âm thanh i và nhiều hướng thời gian cho đối tượng âm thanh j . Lưu ý rằng các vùng t/f không nhất thiết phải bằng hoặc phân bố đều trong miền t/f, nhưng có thể thích ứng kích thước, vị trí và phân bố theo nhu cầu của các đối tượng âm thanh. Nói cách khác, tín hiệu trộn giảm X được lấy mẫu trong miền thời gian/tần số thành các khe thời gian và các băng phụ (lai). Vùng thời gian/tần số $R(t_R, f_R)$ có thể mở rộng đến ít nhất hai mẫu của tín hiệu trộn giảm X . Độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h được làm mịn hơn vùng thời gian/tần số $R(t_R, f_R)$.

Khi xác định các thông tin phụ cho đối tượng âm thanh i tại phía bộ mã hóa âm thanh, bộ mã hóa âm thanh phân tích đối tượng âm thanh i trong vùng t/f $R(t_R, f_R)$ và xác định thông tin phụ thô và thông tin phụ cấu trúc mịn. Thông tin phụ thô có thể là chênh lệch mức đối tượng OLD_i , hiệp phương sai liên đối tượng $IOC_{i,j}$ và/hoặc mức

năng lượng tuyệt đối NRG_i , như được xác định trong, giữa các loại khác, tiêu chuẩn SAOC ISO/IEC 23003-2. Thông tin phụ thô được xác định trên cơ sở vùng t/f và thường cung cấp khả năng tương thích ngược như các bộ giải mã SAOC sử dụng loại thông tin phụ này. Thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{n,k}$ cho đối tượng i cung cấp ba trị số chỉ ra năng lượng của đối tượng âm thanh i được phân bố như thế nào giữa ba vùng con phổ. Trong trường hợp được minh họa, mỗi vùng con phổ tương ứng với một băng phụ (lai) nhưng những sự phân bố khác cũng khả thi. Nó thậm chí có thể được dự kiến để làm một vùng con phổ nhỏ hơn các vùng con phổ khác để có độ giải phổ mịn cụ thể sẵn có trong băng phụ phổ nhỏ. Trong cách tương tự, vùng t/f $R(t_R, f_R)$ như nhau có thể được chia nhỏ thành một vài vùng con thời gian để biểu diễn nhiều hơn nội dung của đối tượng âm thanh j trong vùng t/f $R(t_R, f_R)$.

Thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{n,k}$ có thể mô tả sự chênh lệch giữa thông tin phụ chỉ rõ đối tượng thô (ví dụ, OLD_i , $IOC_{i,j}$, and/or NRG_i) và ít nhất một đối tượng âm thanh s_i .

Phần dưới của Fig.11 minh họa ma trận hiệp phương sai được ước lượng E biến thiên trên vùng t/f $R(t_R, f_R)$ do thông tin phụ cấu trúc mịn cho đối tượng âm thanh i và j . Các ma trận hoặc các giá trị khác được sử dụng trong tác vụ phân tách đối tượng cũng có thể phải chịu sự biến thiên trong vùng t/f $R(t_R, f_R)$. Sự biến thiên của ma trận hiệp phương sai E (và sự khả thi của các ma trận và trị số) được xem xét bởi bộ phân tách đối tượng 120. Trong trường hợp được minh họa, ma trận hiệp phương sai E khác được xác định bởi mỗi mẫu khe thời gian/băng phụ của vùng t/f $R(t_R, f_R)$. Trong trường hợp chỉ một đối tượng âm thanh có cấu trúc phổ mịn kết hợp với nó, ví dụ, đối tượng i , ma trận hiệp phương sai E sẽ là hằng số trong mỗi ba vùng con phổ (ở đây, hằng số trong một trong ba băng phụ (lai), nhưng thường các vùng con phổ khác là khả thi).

Bộ phân tách đối tượng 120 có thể được tạo cấu hình để xác định ma trận hiệp phương sai được ước lượng $E^{n,k}$ với các phần tử $e_{i,j}^{n,k}$ của ít nhất một đối tượng âm thanh s_i và ít nhất một đối tượng âm thanh khác s_j theo

$$e_{i,j}^{n,k} = \sqrt{fsl_i^{n,k} fsl_j^{n,k} fsc_{i,j}^{n,k}},$$

trong đó

$e_{i,j}^{n,k}$ là hiệp phương sai được ước lượng của các đối tượng âm thanh i và j cho khe thời gian n và băng phụ (lai) k ;

$fsl_i^{n,k}$ và $fsl_j^{n,k}$ là thông tin phụ chỉ rõ đối tượng của các đối tượng âm thanh i và j cho khe thời gian n và băng phụ (lai) k ;

$fsc_{i,j}^{n,k}$ là thông tin tương quan liên đối tượng của các đối tượng âm thanh i và j cho khe thời gian n và băng phụ (lai) k ;

Ít nhất một trong $fsl_i^{n,k}$, $fsl_j^{n,k}$, và $fsc_{i,j}^{n,k}$ biến thiên trong vùng thời gian/tần số $R(t_R, f_R)$ theo độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h cho các đối tượng âm thanh i hoặc j chỉ ra lần lượt bởi thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_{I_i} , TFR_{I_j} . Bộ tách đối tượng 120 có thể còn được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh s_i từ tín hiệu trộn giảm X sử dụng ma trận hiệp phương sai $E^{n,k}$ trong cách được miêu tả ở trên.

Một sự thay thế cho phương án được miêu tả ở trên phải được thực hiện khi độ phân giải phổ hoặc thời gian được tăng từ độ phân giải của phép biến đổi cơ bản, ví dụ, với phép biến đổi thu phóng tiếp sau. Trong trường hợp này, sự ước lượng của ma trận hiệp phương sai đối tượng cần được thực hiện ở miền được thu phóng và sự khôi phục lại đối tượng cũng xảy ra ở miền thu phóng. Kết quả của sự khôi phục có thể được biến đổi nghịch đảo trở lại miền của phép biến đổi ban đầu, ví dụ, QMF(lai), sự đan xen của các ô thành sự khôi phục cuối cùng được xảy ra ở miền này. Theo nguyên lý, các tính toán vận hành theo cách giống nhau như chúng làm trong trường hợp sử dụng sự tạo ô thông số khác nhau ngoại trừ phép biến đổi bổ sung.

Fig.12 minh họa dưới dạng giản đồ phép biến đổi thu phóng qua ví dụ thu phóng trong trục phổ, việc xử lý trong miền được thu phóng và phép biến đổi thu phóng nghịch đảo. Xem xét sự trộn giảm trong vùng thời gian/tần số $R(t_R, f_R)$ tại độ phân giải t/f của tín hiệu trộn giảm được xác định bằng các khe thời gian n và các băng phụ lai k . Trong ví dụ được thể hiện trên Fig.12, vùng thời gian-tần số $R(t_R, f_R)$ mở rộng bốn khe thời gian từ n đến $n+3$ và một băng phụ k . Phép thu phóng có thể được biến đổi bởi bộ biến đổi tín hiệu tín hiệu 115. Phép biến đổi thu phóng có thể là phép biến đổi thu phóng thời gian hoặc phép biến đổi thu phóng phổ, như trên Fig.12, phép biến đổi thu

phóng phổ. Phép biến đổi thu phóng phổ có thể được thực hiện bằng các phương pháp DFT, STFT, QMF dựa trên giàn lọc phân tích, v.v. Phép biến đổi thu phóng phổ có thể được thực hiện bằng các phương pháp DFT nghịch đảo, STFT nghịch đảo, QMF nghịch đảo dựa trên giàn lọc phân tích, v.v. Trong ví dụ của Fig.12, tín hiệu trộn giảm X được chuyển đổi từ phép biểu diễn thời gian/tần số tín hiệu trộn giảm được xác định bằng các khe thời gian n và các băng phụ (lai) k thành phép biểu diễn thu phóng phổ trải rộng chỉ một khe thời gian chỉ rõ đối tượng η , nhưng bốn băng phụ (lai) chỉ rõ đối tượng κ đến $\kappa+3$. Do đó, độ phân giải phổ của tín hiệu trộn giảm trong vùng thời gian/tần số $R(t_R, f_R)$ được tăng hệ số 4 bằng phí tổn của độ phân giải theo thời gian.

Việc xử lý được thực hiện ở độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h bằng bộ phân tách đối tượng 121 mà cũng nhận được thông tin phụ của ít nhất một trong các đối tượng âm thanh trong độ phân giải thời gian/tần số TFR_h . Trong ví dụ của Fig.12, đối tượng âm thanh i được xác định bằng thông tin phụ trong vùng thời gian/tần số $R(t_R, f_R)$ mà phù hợp với độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h , tức là, một khe thời gian chỉ rõ đối tượng η và bốn băng phụ lai η đến $\eta+3$. Cho các mục đích minh họa, thông tin phụ cho hai đối tượng âm thanh khác $i+1$ và $i+2$ cũng được minh họa dưới dạng giản đồ trên Fig.12. Đối tượng âm thanh $i+1$ được xác định bằng thông tin phụ có độ phân giải thời gian/tần số của tín hiệu trộn giảm. Đối tượng âm thanh $i+2$ được xác định bằng thông tin phụ có độ phân giải của hai khe thời gian/tần số chỉ rõ đối tượng và hai băng phụ (lai) chỉ rõ đối tượng trong vùng thời gian/tần số $R(t_R, f_R)$. Đối với đối tượng âm thanh $i+1$, bộ phân tách đối tượng 121 có thể xem xét thông tin phụ thô trong vùng thời gian/tần số $R(t_R, f_R)$. Đối với đối tượng âm thanh $i+2$, bộ phân tách đối tượng 121 có thể xem xét trị số trung bình phổ trong vùng thời gian/tần số $R(t_R, f_R)$, như được chỉ ra bằng hai đường gạch sọc khác nhau. Trong trường hợp chung, các trị số phổ trung bình và/hoặc các trị số thời gian trung bình có thể được xem xét bằng bộ phân tách đối tượng 121, nếu thông tin phụ cho đối tượng âm thanh tương ứng không sẵn có trong độ phân giải thời gian/tần số chỉ rõ đối tượng chính xác TFR_h mà được xử lý hiện tại bằng bộ phân tách đối tượng 121 nhưng bị rời rạc mịn hơn trong chiều thời gian và/hoặc phổ so với vùng thời gian/tần số $R(t_R, f_R)$. Theo cách này, bộ phân tách đối tượng 121 được lợi từ việc sẵn có thông tin phụ chỉ rõ đối tượng mà làm mịn rời rạc hơn thông tin phụ thô (ví dụ, OLD, IOC,

and/or NRG), mặc dù không nhất thiết phải mịn như độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h hiện tại được xử lý bằng bộ phân tách đối tượng 121.

Bộ phân tách đối tượng 121 xuất ra ít nhất một đối tượng âm thanh bị tách s_i cho vùng thời gian/tần số $R(t_R, f_R)$ tại độ phân giải thời gian/tần số chỉ rõ đối tượng (độ phân giải thời gian/tần số thu phóng). Ít nhất một đối tượng âm thanh được tách s_i sau đó được biến đổi thu phóng nghịch đảo bằng bộ biến đổi thu phóng nghịch đảo 132 để thu đối tượng âm thanh được tách s_i trong $R(t_R, f_R)$ tại độ phân giải thời gian/tần số của tín hiệu trộn giảm hoặc tại độ phân giải tần số/thời gian mong muốn khác. Đối tượng âm thanh được tách s_i trong $R(t_R, f_R)$ sau đó được kết hợp với đối tượng âm thanh được tách s_i trong vùng thời gian/tần số khác, ví dụ, $R(t_{R-1}, f_{R-1})$, $R(t_{R-1}, f_R)$, ... $R(t_{R+1}, f_{R+1})$, để kết hợp các đối tượng âm thanh được tách s_i .

Theo các phương án tương ứng, bộ giải mã âm thanh bao gồm bộ biến đổi thời gian/tần số tín hiệu trộn giảm 115 được tạo cấu hình để biến đổi tín hiệu trộn giảm X trong vùng thời gian/tần số $R(t_R, f_R)$ từ độ phân giải thời gian/tần số tín hiệu trộn giảm đến ít nhất độ phân giải thời gian tần số chỉ rõ đối tượng TFR_h của ít nhất một đối tượng âm thanh s_i để thu được tín hiệu trộn giảm được biến đổi lại $X^{\eta, \kappa}$. Độ phân giải thời gian/tần số tín hiệu trộn giảm liên quan tới các khe thời gian trộn giảm n và các băng phụ (lai) k . Độ phân giải thời gian/tần số đối tượng riêng biệt TFR_h liên quan đến các khe thời gian chỉ rõ đối tượng η và băng phụ (lai) chỉ rõ đối tượng κ . Các khe thời gian chỉ rõ đối tượng η có thể mịn hơn hoặc thô hơn các khe thời gian trộn giảm n của độ phân giải thời gian/tần số trộn giảm. Cũng như vậy, các băng phụ (lai) chỉ rõ đối tượng η có thể mịn hơn hoặc thô hơn các băng phụ (lai) trộn giảm của độ phân giải thời gian/tần số trộn giảm. Như đã giải thích ở trên, trong mối liên hệ với các nguyên lý bất định của phép biểu diễn thời gian/tần số, độ phân giải phổ của tín hiệu có thể được tăng bằng phí tổn của độ phân giải thời gian hoặc ngược lại. Bộ giải mã âm thanh còn bao gồm bộ biến đổi thời gian/tần số nghịch đảo 132 được tạo cấu hình để biến đổi thời gian/tần số của ít nhất một đối tượng âm thanh s_i trong vùng thời gian tần số $R(t_R, f_R)$ từ độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h trở lại độ phân giải thời gian/tần số tín hiệu trộn giảm. Bộ phân tách 121 được tạo cấu hình để phân tách ít nhất

một đối tượng âm thanh s_i từ tín hiệu trộn giảm X tại độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h .

Trong miền được thu phóng, ma trận hiệp phương sai được ước lượng $\mathbf{E}^{\eta,\kappa}$ được xác định cho các khe thời gian chỉ rõ đối tượng η và các băng phụ (lai) chỉ rõ đối tượng κ . Công thức nêu trên cho các phần tử của ma trận hiệp phương sai được ước lượng của ít nhất một đối tượng âm thanh s_i và ít nhất đối tượng âm thanh còn lại s_j có thể được thể hiện trong miền được thu phóng như sau:

$$e_{i,j}^{\eta,\kappa} = \sqrt{fsl_i^{\eta,\kappa} fsl_j^{\eta,\kappa}} fsc_{i,j}^{\eta,\kappa},$$

trong đó

$e_{i,j}^{\eta,\kappa}$ là hiệp phương sai được ước lượng của các đối tượng âm thanh i và j cho khe thời gian chỉ rõ đối tượng η và băng phụ (lai) chỉ rõ đối tượng κ ;

$fsl_i^{\eta,\kappa}$ và $fsl_j^{\eta,\kappa}$ là thông tin phụ chỉ rõ đối tượng của các đối tượng âm thanh i và j cho khe thời gian chỉ rõ đối tượng η và băng phụ (lai) chỉ rõ đối tượng κ ;

$fsc_{i,j}^{\eta,\kappa}$ là hiệp phương sai được ước lượng của các đối tượng âm thanh i và j , lần lượt cho khe thời gian chỉ rõ đối tượng η và băng phụ (lai) chỉ rõ đối tượng κ ;

Như đã giải thích ở trên, đối tượng âm thanh j khác có thể được xác định bằng thông tin phụ có độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h của đối tượng âm thanh i để các tham số $fsl_j^{\eta,\kappa}$ và $fsc_{i,j}^{\eta,\kappa}$ có thể sẵn có hoặc có thể được xác định tại độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h . Trong trường hợp này, thông tin phụ của của đối tượng âm thanh j trong $R(\text{tr}, \text{fr})$ hoặc các trị số trung bình thời gian hoặc các trị số trung gian phổ có thể được sử dụng để làm gần đến các tham số $fsl_j^{\eta,\kappa}$ và $fsc_{i,j}^{\eta,\kappa}$ trong vùng thời gian/tần số $R(\text{tr}, \text{fr})$ hoặc trong vùng con của nó.

Cũng ở phía bộ mã hóa, thông tin phụ cấu trúc mịn nên thường được xem xét. Trong bộ mã hóa theo sáng chế bộ xác định thông tin (t/f-SIE) 55-1...55-H còn được tạo cấu hình để cung cấp thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{\eta,\kappa}$ hoặc $fsl_j^{\eta,\kappa}$ và thông tin phụ chỉ rõ đối tượng thô OLD_i như một phần của ít nhất một thông tin phụ thứ nhất và thông tin phụ thứ hai. Thông tin phụ chỉ rõ đối tượng thô OLD_i là hằng số

trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$. Thông tin phụ chỉ rõ đối tượng cấu trúc mịn $fsl_i^{n,k}$, $fsl_i^{n,k}$ có thể mô tả sự chênh lệch giữa thông tin phụ chỉ rõ đối tượng thô OLD_i và ít nhất một đối tượng âm thanh s_i . Tương quan liên đối tượng $IOC_{i,j}$ và $fsC_{i,j}^{n,k}$, $fsC_{i,j}^{n,k}$ có thể được xử lý theo cách tương tự, cũng như các thông tin phụ tham số khác.

Fig.13 thể hiện lưu đồ dạng giản lược của phương pháp giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm X và thông tin phụ PSI. Thông tin bao gồm thông tin phụ chỉ rõ đối tượng PSI_i cho ít nhất một đối tượng âm thanh s_i trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$, và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_h của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh s_i trong ít nhất một vùng thời gian/tần số $R(t_R, f_R)$. Phương pháp bao gồm bước 1302 của việc xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i từ thông tin phụ PSI cho ít nhất một đối tượng âm thanh s_i . Phương pháp còn bao gồm bước 1304 của việc phân tách ít nhất một đối tượng âm thanh s_i từ tín hiệu trộn giảm X sử dụng thông tin phụ chỉ rõ đối tượng trong mối liên hệ với độ phân giải thời gian/tần số chỉ rõ đối tượng TFR_i .

Fig.14 thể hiện lưu đồ dạng giản lược của phương pháp mã hóa các tín hiệu đối tượng âm thanh s_i thành các tín hiệu trộn giảm X và thông tin phụ PSI theo các phương án khác. Bộ mã hóa âm thanh bao gồm phép biến đổi các tín hiệu đối tượng âm thanh s_i thành ít nhất các biến đổi tương ứng thứ nhất $s_{1,1}(t,f) \dots s_{N,1}(t,f)$ tại bước 1402. Độ phân giải thời gian/tần số thứ nhất TFR_1 được sử dụng cho mục đích này. Các tín hiệu đối tượng âm thanh s_i cũng được biến đổi ít nhất thành phép biến đổi tương ứng thứ hai $s_{1,2}(t,f) \dots s_{N,2}(t,f)$ sử dụng sự rời rạc thời gian/tần số thứ hai TFR_2 . Tại bước 1404 ít nhất thông tin phụ thứ nhất cho các biến đổi tương ứng thứ nhất $s_{1,1}(t,f) \dots s_{N,1}(t,f)$ và thông tin phụ thứ hai cho các biến đổi tương ứng thứ hai $s_{1,2}(t,f) \dots s_{N,2}(t,f)$ được xác định. Thông tin phụ thứ nhất và thứ hai biểu thị sự liên hệ giữa các đối tượng âm thanh s_i với nhau trong độ phân giải thời gian/tần số lần lượt là thứ nhất và thứ hai TFR_1 , TFR_2 , trong vùng thời gian/tần số $R(t_R, f_R)$. Phương pháp cũng bao gồm bước 1406 để lựa chọn, cho mỗi tín hiệu đối tượng âm thanh s_i , một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thông tin phụ thứ hai dựa trên các tiêu chí phù hợp

biểu thị sự phù hợp của ít nhất độ phân giải thời gian/tần số thứ nhất hoặc thứ hai để biểu diễn tín hiệu đối tượng âm thanh s_i trong miền thời gian/tần số, thông tin phụ chỉ rõ đối tượng được chèn vào thông tin phụ PSI được xuất ra nhờ bộ mã hóa âm thanh.

Tương thích theo chiều ngược với SAOC

Các giải pháp được đề xuất thuận lợi cải thiện chất lượng âm thanh cảm nhận, thậm chí có thể theo cách tương thích với bộ mã hóa một cách hoàn toàn. Bằng cách xác định các miền t/f $R(t_r, f_r)$ để tương đẳng với việc nhóm t/f trong tình trạng kỹ thuật SAOC, các bộ giải mã SAOC tiêu chuẩn hiện hành có thể giải mã phần tương thích theo chiều ngược của PSI và tạo ra sự tái tạo của các đối tượng ở mức độ độ phân giải t/f thô. Nếu thông tin được thêm vào được sử dụng nhằm nâng cao bộ giải mã SAOC, chất lượng cảm nhận của sự tái tạo sẽ được cải thiện đáng kể. Đối với mỗi đối tượng âm thanh, thông tin phụ được bổ sung này bao gồm thông tin, mà phép biểu diễn t/f riêng lẻ nên được sử dụng để ước lượng đối tượng, cùng với sự mô tả cấu trúc mịn đối tượng dựa trên phép biểu diễn t/f được chọn lọc.

Ngoài ra, nếu bộ giải mã SAOC nâng cao chạy trên nguồn tài nguyên hạn chế, sự nâng cao có thể bị bỏ qua, và sự khôi phục chất lượng cơ bản có thể thu được chỉ yêu cầu độ phức tạp tính toán thấp.

Lĩnh vực ứng dụng cho quy trình sáng tạo

Các khái niệm về các phép biểu diễn t/f chỉ rõ đối tượng và sự truyền tín hiệu được kết hợp của nó đến các bộ giải mã có thể được áp dụng trên bất kỳ sơ đồ SAOC nào. Nó có thể được kết hợp với bất kỳ định dạng âm thanh hiện tại nào cũng như trong tương lai. Khái niệm này cho phép sự ước lượng đối tượng âm thanh cảm giác nâng cao trong ứng dụng SAOC nhờ sự lựa chọn thích ứng đối tượng âm thanh của độ phân giải t/f riêng lẻ cho sự ước lượng tham số của các đối tượng âm thanh.

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

của thiết bị tương ứng. Một số hoặc tất cả các bước phương pháp có thể được thực hiện bởi (hoặc sử dụng) thiết bị phần cứng, ví dụ, bộ vi xử lý, máy tính có thể lập trình hoặc mạch điện. Theo một số phương án, một hoặc nhiều bước phương pháp có thể được thực hiện bởi thiết bị này.

Tín hiệu đã được mã hóa có thể được cất trữ trên môi trường lưu trữ số hoặc có thể được truyền trên môi trường truyền như môi trường truyền không dây hoặc môi trường truyền có dây như 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. Phương án có thể được thực hiện sử dụng môi trường lưu trữ số, ví dụ, đĩa mềm, DVD, đĩa Blue-Ray, 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, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính khả trình sao cho phương pháp tương ứng được thực hiện. Do đó, môi trường lưu trữ số có thể đọc được bằng máy tính.

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

Thông thường, các phương án của sáng chế có thể được thực hiện như một sản phẩm chương trình máy tính với mã chương trình, mã chương trình hoạt động để thực hiện một trong các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Mã chương trình có thể ví dụ được lưu trữ trên vật mang đọc được bằng máy tính.

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

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

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

đây. Vật mang dữ liệu, môi trường lưu trữ số hoặc môi trường được ghi thường là hữu hình và/hoặc không truyền sóng.

Phương án khác của các phương pháp theo sáng chế là, dòng dữ liệu hoặc chuỗi các tín hiệu thể hiện chương trình máy tính để thực hiện một trong các phương pháp đã được mô tả ở đây. Ví dụ, có thể cấu hình dòng dữ liệu hoặc chuỗi các tín hiệu để được truyền thông qua sự kết nối truyền thông dữ liệu, ví dụ thông qua Internet.

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

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

Theo một số phương án, thiết bị logic lập trình được (ví dụ, mảng cổng lập trình được dạng trường) có thể được sử dụng để thực hiện một số hoặc tất cả các chức năng của các phương pháp được mô tả ở đây. Theo một số phương án, mảng cổng lập trình được dạng trường có thể kết hợp với bộ vi xử lý để thực hiện một trong các phương pháp được mô tả ở đây. Thông thường, các phương pháp ưu tiên được thực hiện bởi thiết bị phần cứng bất kỳ.

Các phương án được mô tả bên trên chỉ mang tính minh họa cho các nguyên lý của sáng chế. Cần hiểu rằng các biến thể và biến đổi của các phương án và các chi tiết được mô tả ở đây sẽ là rõ ràng đối với người có trình độ trung bình trong lĩnh vực kỹ thuật tương ứng. Do đó, mục đích của nó chỉ bị giới hạn chỉ bởi phạm vi của các điểm yêu cầu bảo hộ và không bởi các chi tiết cụ thể được thể hiện bởi bản mô tả và sự giải thích các phương án ở đây.

YÊU CẦU BẢO HỘ

1. Bộ giải mã âm thanh để giải mã các tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm (X) và thông tin phụ (PSI), thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng (PSI_i) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_i) biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), bộ giải mã âm thanh bao gồm:

bộ xác định độ phân giải thời gian/tần số chỉ rõ đối tượng (110) được tạo cấu hình để xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_i) từ thông tin phụ (PSI) cho ít nhất một đối tượng âm thanh (s_i); và

bộ phân tách đối tượng (120) được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian tần số chỉ rõ đối tượng (TFR_i); và

trong đó thông tin phụ chỉ rõ đối tượng là thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsI_i^{n,k}$, $fsI_j^{n,k}$) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và trong đó thông tin phụ (PSI) còn bao gồm thông tin phụ chỉ rõ đối tượng thô cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), thông tin phụ chỉ rõ đối tượng thô không đổi trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$); hoặc

trong đó thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsI_i^{n,k}$) mô tả sự chênh lệch giữa thông tin phụ chỉ rõ đối tượng thô và ít nhất một đối tượng âm thanh (s_i).

2. Bộ giải mã âm thanh theo điểm 1, trong đó tín hiệu trộn giảm (X) được lấy mẫu trong miền thời gian/tần số thành nhiều khe thời gian và nhiều băng phụ (lai), trong đó miền thời gian/tần số ($R(t_R, f_R)$) mở rộng tới ít nhất hai mẫu của tín hiệu trộn giảm (X), và trong đó độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) mịn hơn trong ít nhất một trong cả hai chiều so với vùng thời gian/tần số ($R(t_R, f_R)$).

3. Bộ giải mã theo điểm 1, trong đó bộ phân tách đối tượng (120) được tạo cấu hình để xác định ma trận hiệp phương sai được ước lượng ($\mathbf{E}^{\eta,\kappa}$) với các phần tử $e_{i,j}^{\eta,\kappa}$ của ít nhất một đối tượng âm thanh (s_i) và ít nhất một đối tượng âm thanh khác (s_j) theo

$$e_{i,j}^{\eta,\kappa} = \sqrt{fsl_i^{\eta,\kappa} fsl_j^{\eta,\kappa} fsc_{i,j}^{\eta,\kappa}},$$

trong đó

$e_{i,j}^{\eta,\kappa}$ là hiệp phương sai được ước lượng của các đối tượng âm thanh i và j cho khe thời gian cấu trúc mịn η và băng phụ (lai) cấu trúc mịn κ ;

$fsl_i^{\eta,\kappa}$ và $fsl_j^{\eta,\kappa}$ là thông tin phụ chỉ rõ đối tượng của các đối tượng âm thanh i và j cho khe thời gian cấu trúc mịn η và băng phụ (lai) cấu trúc mịn κ ;

$fsc_{i,j}^{\eta,\kappa}$ là thông tin tương quan liên đối tượng của các đối tượng âm thanh i và j , tương ứng, khe thời gian cấu trúc mịn η và băng phụ (lai) cấu trúc mịn κ ;

trong đó ít nhất một trong số $fsl_i^{\eta,\kappa}$, $fsl_j^{\eta,\kappa}$, và $fsc_{i,j}^{\eta,\kappa}$ biến thiên trong vùng thời gian/tần số ($R(t_R, f_R)$) theo độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) cho các đối tượng âm thanh i và j được chỉ ra bởi thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$, $TFRI_j$), và

trong đó bộ phân tách đối tượng (120) còn được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng ma trận hiệp phương sai được ước lượng ($\mathbf{E}^{\eta,\kappa}$).

4. Bộ giải mã âm thanh theo điểm 1, bộ giải mã âm thanh còn bao gồm:

bộ biến đổi thời gian/tần số tín hiệu trộn giảm được tạo cấu hình để biến đổi tín hiệu trộn giảm (X) trong vùng thời gian/tần số ($R(t_R, f_R)$) từ độ phân giải thời gian/tần số tín hiệu trộn giảm đến ít nhất độ phân giải thời gian tần số chỉ rõ đối tượng (TFR_h) của ít nhất một đối tượng âm thanh (s_i) để thu được tín hiệu trộn giảm được biến đổi ($X^{\eta,\kappa}$);

bộ biến đổi thời gian/tần số nghịch đảo được tạo cấu hình để biến đổi thời gian/tần số ít nhất một đối tượng âm thanh (s_i) trong vùng thời gian tần số ($R(t_R, f_R)$) từ

độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) trở lại độ phân giải t/f thông thường hoặc độ phân giải thời gian/tần số tín hiệu trộn giảm;

trong đó bộ phân tách (120) được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) tại độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h).

5. Bộ mã hóa âm thanh để mã hóa nhiều đối tượng âm thanh (s_i) thành tín hiệu trộn giảm (X) và thông tin phụ (PSI), bộ mã hóa âm thanh bao gồm:

bộ biến đổi thời gian thành tần số được tạo cấu hình để biến đổi nhiều đối tượng âm thanh (s_i) ít nhất thành nhiều biến đổi tương ứng thứ nhất ($s_{1,1}(t,f) \dots s_{N,a}(t,f)$) sử dụng độ phân giải thời gian/tần số thứ nhất (TFR_1) và nhiều biến đổi tương ứng thứ hai ($s_{1,2}(t,f) \dots s_{N,2}(t,f)$) sử dụng độ phân giải thời gian/tần số thứ hai (TFR_2);

bộ xác định thông tin phụ (t/f-SIE) được tạo cấu hình để xác định ít nhất thông tin phụ thứ nhất cho nhiều biến đổi tương ứng thứ nhất ($s_{1,1}(t,f) \dots s_{N,1}(t,f)$) và thông tin phụ thứ hai cho nhiều biến đổi tương ứng thứ hai ($s_{1,2}(t,f) \dots s_{N,2}(t,f)$), thông tin phụ thứ nhất và thông tin phụ thứ hai chỉ ra mối liên hệ của nhiều tín hiệu đối tượng âm thanh (s_i) với nhau lần lượt trong độ phân giải thời gian/tần số thứ nhất và thứ hai (TFR_1 , TFR_2), trong vùng thời gian/tần số ($R(t_R, f_R)$); và

bộ chọn lọc thông tin phụ (SI-AS) được tạo cấu hình để lựa chọn, cho ít nhất một đối tượng âm thanh (s_i) của nhiều đối tượng âm thanh, một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thông tin phụ thứ hai dựa trên các tiêu chí phù hợp biểu thị sự phù hợp của ít nhất độ phân giải thời gian/tần số thứ nhất hoặc thứ hai để biểu diễn đối tượng âm thanh (s_i) trong miền thời gian/tần số, thông tin phụ chỉ rõ đối tượng được chèn vào thông tin phụ (PSI) được xuất ra nhờ bộ mã hóa âm thanh.

6. Bộ mã hóa âm thanh theo điểm 5, trong đó các tiêu chí phù hợp được dựa trên sự ước lượng nguồn và trong đó bộ chọn lọc thông tin phụ (SI-AS) bao gồm:

bộ ước lượng nguồn được tạo cấu hình để ước lượng ít nhất đối tượng âm thanh được chọn của nhiều đối tượng âm thanh (s_i) sử dụng tín hiệu trộn giảm (X) và ít nhất thông tin thứ nhất và thông tin thứ hai lần lượt tương ứng với các độ phân giải thời gian/tần số thứ nhất và thứ hai (TFR_1 , TFR_2), bộ ước lượng nguồn do đó cung cấp ít

nhất đối tượng âm thanh được ước lượng thứ nhất ($s_{i, \text{estim1}}$) và đối tượng âm thanh được ước lượng thứ hai ($s_{i, \text{estim2}}$);

bộ đánh giá chất lượng được tạo cấu hình để đánh giá chất lượng của ít nhất đối tượng âm thanh được ước lượng thứ nhất ($s_{i, \text{estim1}}$) và đối tượng âm thanh được ước lượng thứ hai ($s_{i, \text{estim2}}$).

7. Bộ mã hóa âm thanh theo điểm 6, trong đó bộ đánh giá chất lượng có thể được tạo cấu hình để đánh giá chất lượng của ít nhất đối tượng âm thanh được ước lượng thứ nhất ($s_{i, \text{estim1}}$) và đối tượng âm thanh được ước lượng thứ hai ($s_{i, \text{estim2}}$) dựa trên cơ sở tỉ lệ tín hiệu trên méo (signal-to-distortion ratio - SDR) như là phép đo hiệu suất ước lượng nguồn, tỉ lệ tín hiệu trên méo (SDR) được xác định chỉ dựa trên thông tin phụ (PSI).

8. Bộ mã hóa âm thanh theo điểm 5, trong đó các tiêu chí phù hợp cho ít nhất một đối tượng âm thanh (s_i) trong nhiều đối tượng âm thanh dựa trên độ rải rác của nhiều hơn một phép biểu diễn độ phân giải t/f của ít nhất một đối tượng âm thanh theo ít nhất độ phân giải thời gian/tần số thứ nhất (TFR_1) và độ phân giải thời gian/tần số thứ hai (TFR_2), và trong đó bộ chọn lọc thông tin phụ (side information selector - SI-AS) được tạo cấu hình để chọn lọc thông tin phụ trong ít nhất thông tin phụ thứ nhất và thứ hai mà được kết hợp với phép biểu diễn t/f rời rạc nhất của ít nhất một đối tượng âm thanh (s_i).

9. Bộ mã hóa âm thanh theo điểm 5, trong đó bộ xác định thông tin phụ (t/f-SIE) còn được tạo cấu hình để cung cấp thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsI_i^{n,k}$) và thông tin phụ chỉ rõ đối tượng thô như là một phần của ít nhất một thông tin phụ thứ nhất và thông tin phụ thứ hai, thông tin phụ chỉ rõ đối tượng thô không đổi trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$).

10. Bộ mã hóa âm thanh theo điểm 9, trong đó thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsI_i^{n,k}$) mô tả sự chênh lệch giữa thông tin phụ chỉ rõ đối tượng thô và ít nhất một đối tượng âm thanh (s_i).

11. Bộ mã hóa âm thanh theo điểm 5, còn bao gồm bộ xử lý tín hiệu trộn giảm được tạo cấu hình để biến đổi tín hiệu trộn giảm (X) thành phép biểu diễn mà được lấy mẫu

trong miền thời gian/tần số thành nhiều khe thời gian và nhiều băng phụ (lai), trong đó vùng thời gian/tần số $(R(t_R, f_R))$ mở rộng tới ít nhất hai mẫu tín hiệu trộn giảm (X) , và trong đó độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) đặc trưng cho ít nhất một đối tượng âm thanh là mịn hơn trong ít nhất một trong cả hai chiều so với vùng thời gian/tần số $(R(t_R, f_R))$.

12. Phương pháp giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm (X) và thông tin phụ (PSI), thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng (PSI_i) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số $(R(t_R, f_R))$, và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng $(TFRI_i)$ biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số $(R(t_R, f_R))$, phương pháp bao gồm:

xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng $(TFRI_i)$ từ thông tin phụ (PSI) cho ít nhất một đối tượng âm thanh (s_i) ; và

phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian/tần số chỉ rõ đối tượng $(TFRI_i)$;

trong đó thông tin phụ chỉ rõ đối tượng là thông tin phụ chỉ rõ đối tượng cấu trúc mịn $(fsl_i^{n,k}, fsl_j^{n,k})$ cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số $(R(t_R, f_R))$, và trong đó thông tin phụ (PSI) còn bao gồm thông tin phụ chỉ rõ đối tượng thô cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số $(R(t_R, f_R))$, thông tin phụ chỉ rõ đối tượng thô là không đổi trong ít nhất một vùng thời gian/tần số $(R(t_R, f_R))$; hoặc

trong đó thông tin phụ chỉ rõ đối tượng cấu trúc thô $(fsl_i^{n,k})$ mô tả sự khác nhau giữa thông tin phụ chỉ rõ đối tượng thô và ít nhất một đối tượng âm thanh (s_i) .

13. Phương pháp mã hóa nhiều đối tượng âm thanh (s_i) thành tín hiệu trộn giảm (X) và thông tin phụ (PSI), phương pháp bao gồm:

biến đổi nhiều đối tượng âm thanh (s_i) ít nhất thành nhiều biến đổi tương ứng thứ nhất $(s_{1,1}(t,f) \dots s_{N,a}(t,f))$ sử dụng độ phân giải thời gian/tần số thứ nhất (TFR_1) và thành

nhiều biến đổi tương ứng thứ hai $(s_{1,2}(t,f)...s_{N,2}(t,f))$ sử dụng độ phân giải thời gian/tần số thứ hai (TFR_2);

xác định ít nhất thông tin phụ thứ nhất cho nhiều biến đổi tương ứng thứ nhất $(s_{1,1}(t,f)...s_{N,1}(t,f))$ và thông tin phụ thứ hai cho nhiều biến đổi tương ứng thứ hai $(s_{1,2}(t,f)...s_{N,2}(t,f))$, thông tin phụ thứ nhất và thông tin phụ thứ hai chỉ ra mối liên hệ của nhiều tín hiệu đối tượng âm thanh (s_i) với nhau lần lượt trong độ phân giải thời gian/tần số (TFR_1, TFR_2), trong vùng thời gian/tần số ($R(t_R, f_R)$); và

lựa chọn, cho ít nhất một đối tượng âm thanh (s_i) của nhiều đối tượng âm thanh, một thông tin phụ chỉ rõ đối tượng từ ít nhất thông tin phụ thứ nhất và thứ hai dựa trên các tiêu chí phù hợp biểu thị sự phù hợp của ít nhất độ phân giải thời gian/tần số thứ nhất hoặc thứ hai để biểu diễn đối tượng âm thanh (s_i) trong miền thời gian/tần số, thông tin phụ chỉ rõ đối tượng được chèn vào đầu ra thông tin phụ (PSI) bởi bộ mã hóa âm thanh.

14. Bộ giải mã âm thanh để giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm (X) và thông tin phụ (PSI), thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng (PSI_i) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), bộ giải mã âm thanh bao gồm:

bộ xác định độ phân giải thời gian/tần số chỉ rõ đối tượng (110), được tạo cấu hình để xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) từ thông tin phụ (PSI) cho ít nhất một đối tượng âm thanh (s_i); và

bộ phân tách đối tượng (120) được tạo cấu hình để phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$), trong đó thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_j) khác trong tín hiệu trộn giảm có độ phân giải thời gian/tần số chỉ rõ đối tượng khác (TFR);

trong đó thông tin phụ chỉ rõ đối tượng là thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsl_i^{n,k}, fsl_j^{n,k}$) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và trong đó thông tin phụ (PSI) còn bao gồm thông tin phụ chỉ rõ đối tượng thô cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), thông tin phụ chỉ rõ đối tượng thô là không đổi trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$); hoặc

trong đó thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsl_i^{n,k}$) mô tả sự khác biệt giữa thông tin phụ chỉ rõ đối tượng thô và ít nhất một đối tượng âm thanh (s_i).

15. Phương pháp giải mã tín hiệu âm thanh đa đối tượng bao gồm tín hiệu trộn giảm (X) và thông tin phụ (PSI), thông tin phụ bao gồm thông tin phụ chỉ rõ đối tượng (PSI_i) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) biểu thị độ phân giải thời gian/tần số chỉ rõ đối tượng (TFR_h) của thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), phương pháp này bao gồm:

xác định thông tin độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$) từ thông tin phụ (PSI) cho ít nhất một đối tượng âm thanh (s_i); và

phân tách ít nhất một đối tượng âm thanh (s_i) từ tín hiệu trộn giảm (X) sử dụng thông tin phụ chỉ rõ đối tượng tương ứng với độ phân giải thời gian/tần số chỉ rõ đối tượng ($TFRI_i$), trong đó thông tin phụ chỉ rõ đối tượng cho ít nhất một đối tượng âm thanh (s_i) khác trong tín hiệu trộn giảm có độ phân giải thời gian/tần số chỉ rõ đối tượng khác (TFR);

trong đó thông tin phụ chỉ rõ đối tượng là thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsl_i^{n,k}, fsl_j^{n,k}$) cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), và trong đó thông tin phụ (PSI) còn bao gồm thông tin phụ chỉ rõ đối tượng thô cho ít nhất một đối tượng âm thanh (s_i) trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$), thông tin phụ chỉ rõ đối tượng thô là không đổi trong ít nhất một vùng thời gian/tần số ($R(t_R, f_R)$); hoặc

trong đó thông tin phụ chỉ rõ đối tượng cấu trúc mịn ($fsl_i^{n,k}$) mô tả sự khác biệt giữa thông tin phụ chỉ rõ đối tượng thô và ít nhất một đối tượng âm thanh (s_i).

16. Vật ghi đọc được bằng máy tính chứa chương trình máy tính để thực hiện phương pháp theo điểm 12 khi chương trình máy tính chạy trên máy tính.

17. Vật ghi đọc được bằng máy tính chứa chương trình máy tính để thực hiện phương pháp theo điểm 13 khi chương trình máy tính chạy trên máy tính.

18. Vật ghi đọc được bằng máy tính chứa chương trình máy tính để thực hiện phương pháp theo điểm 15 khi chương trình máy tính chạy trên máy tính.

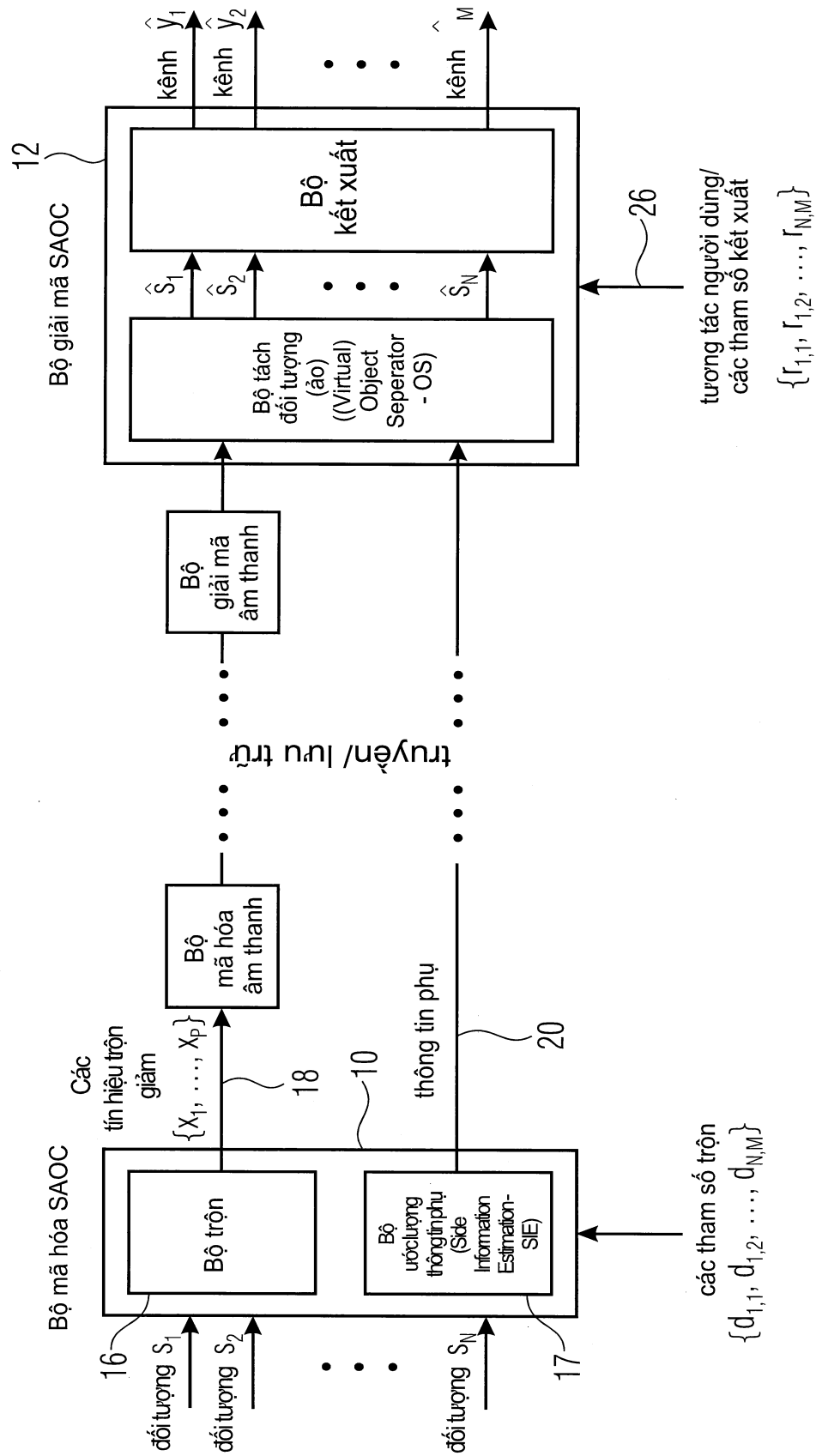


FIG 1

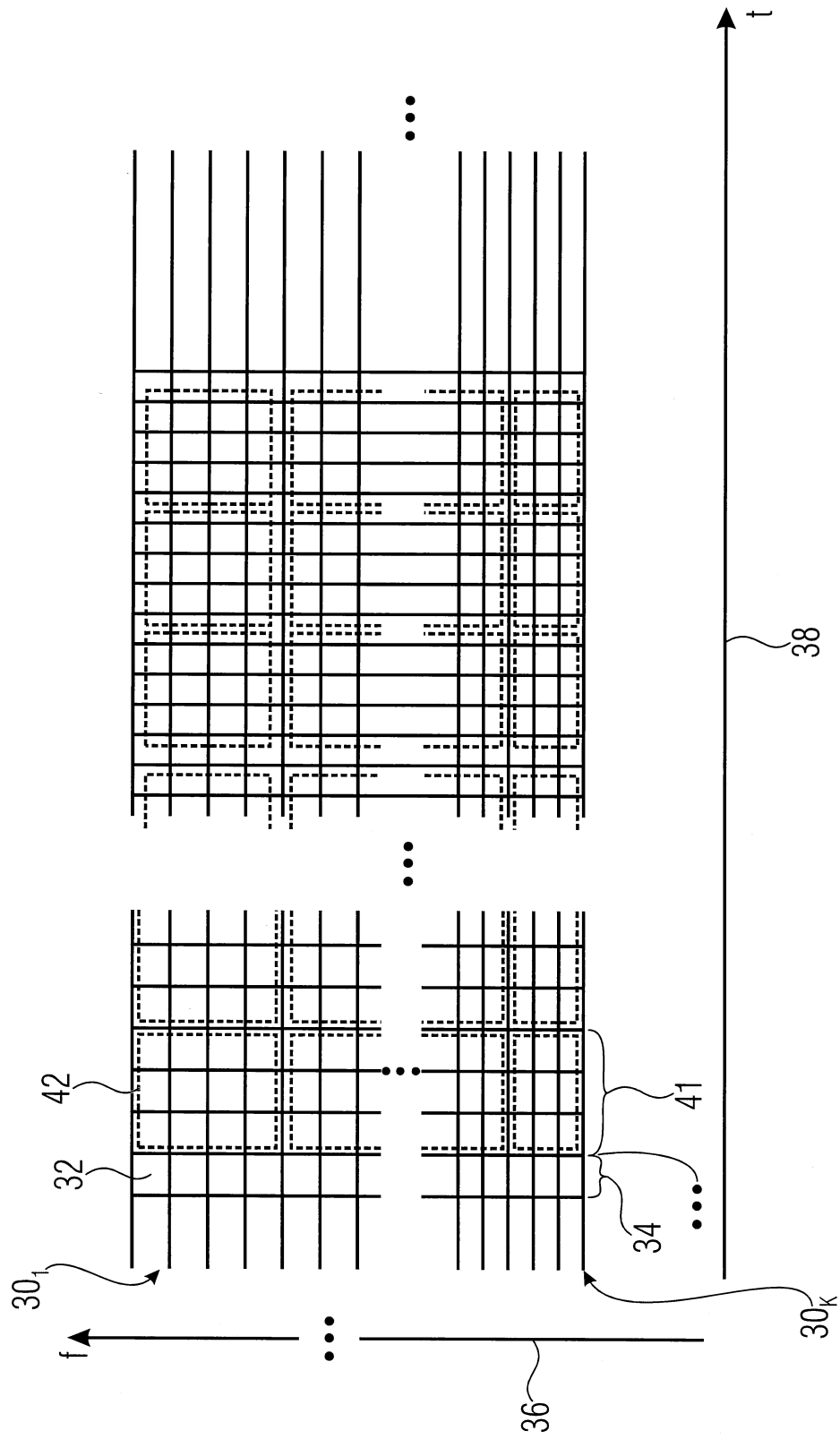


FIG 2

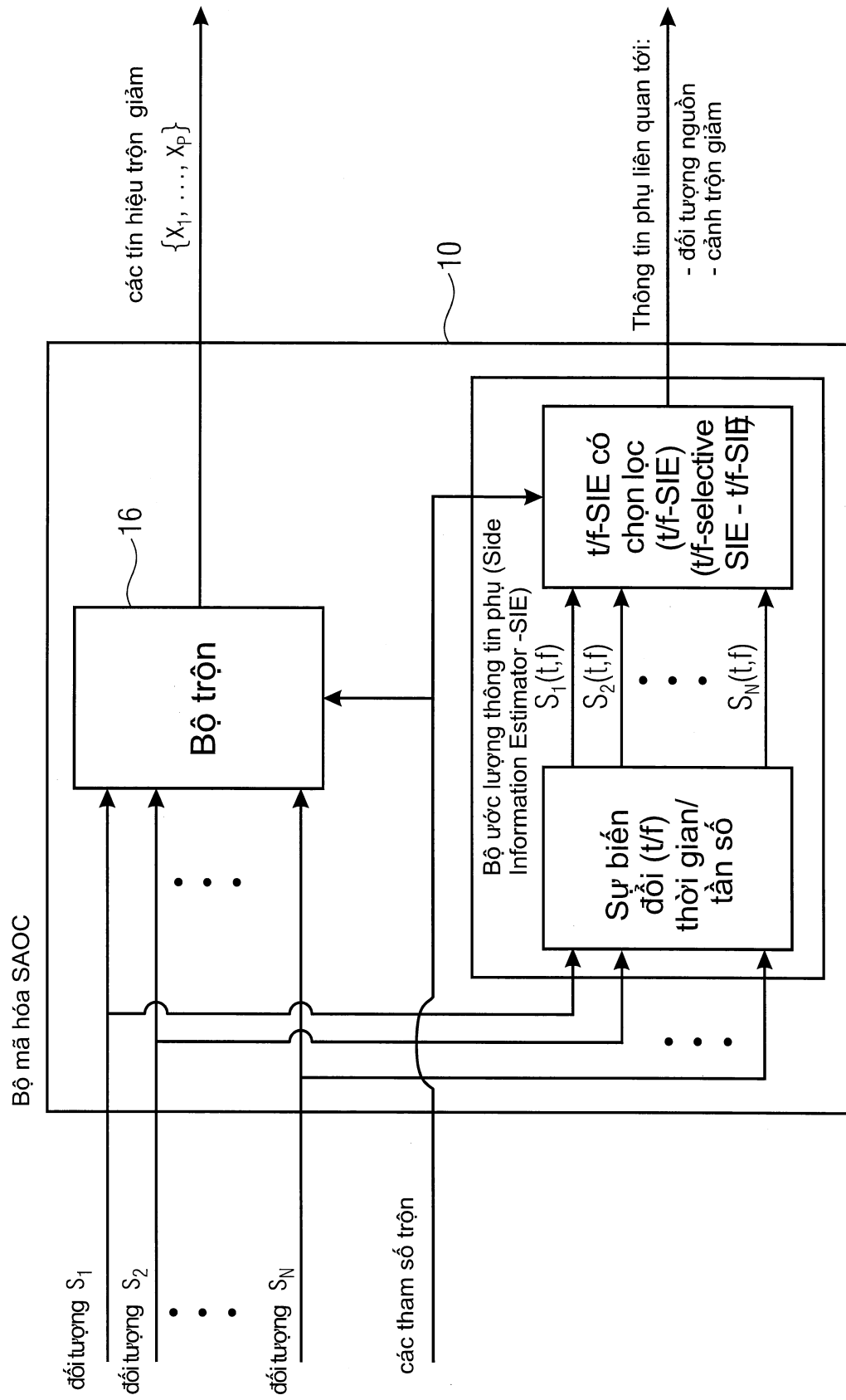


FIG 3

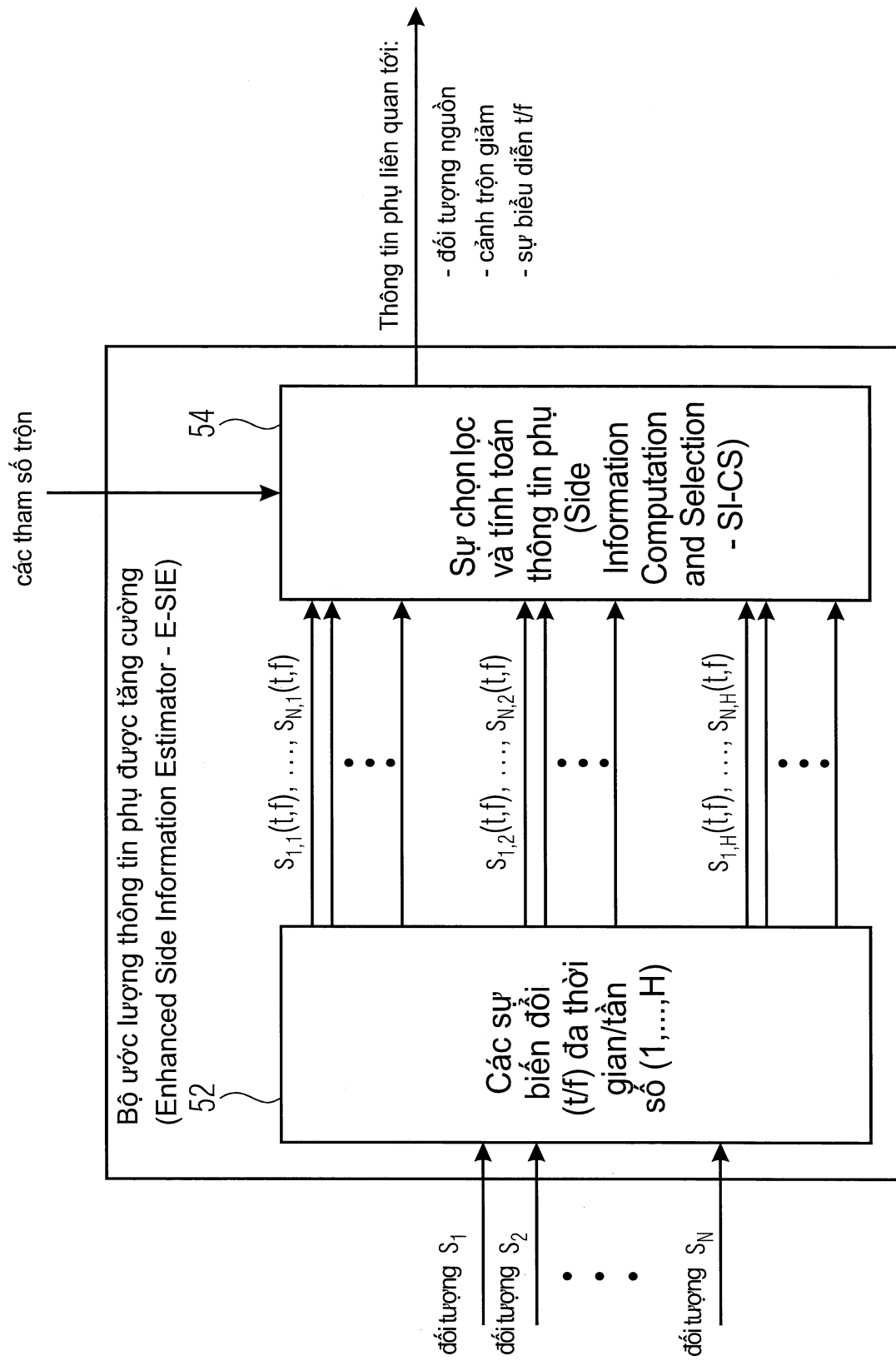


FIG 4

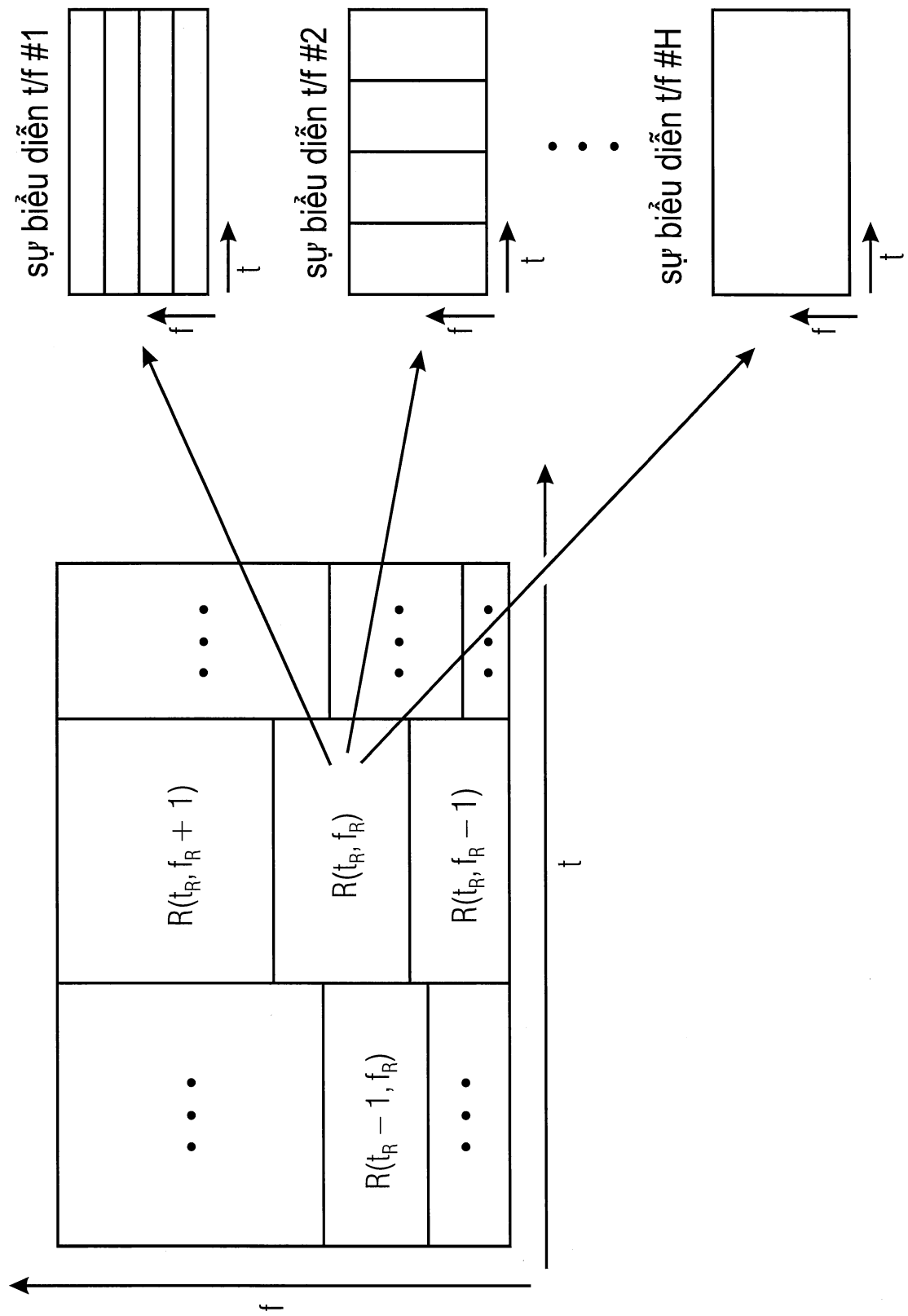


FIG 5

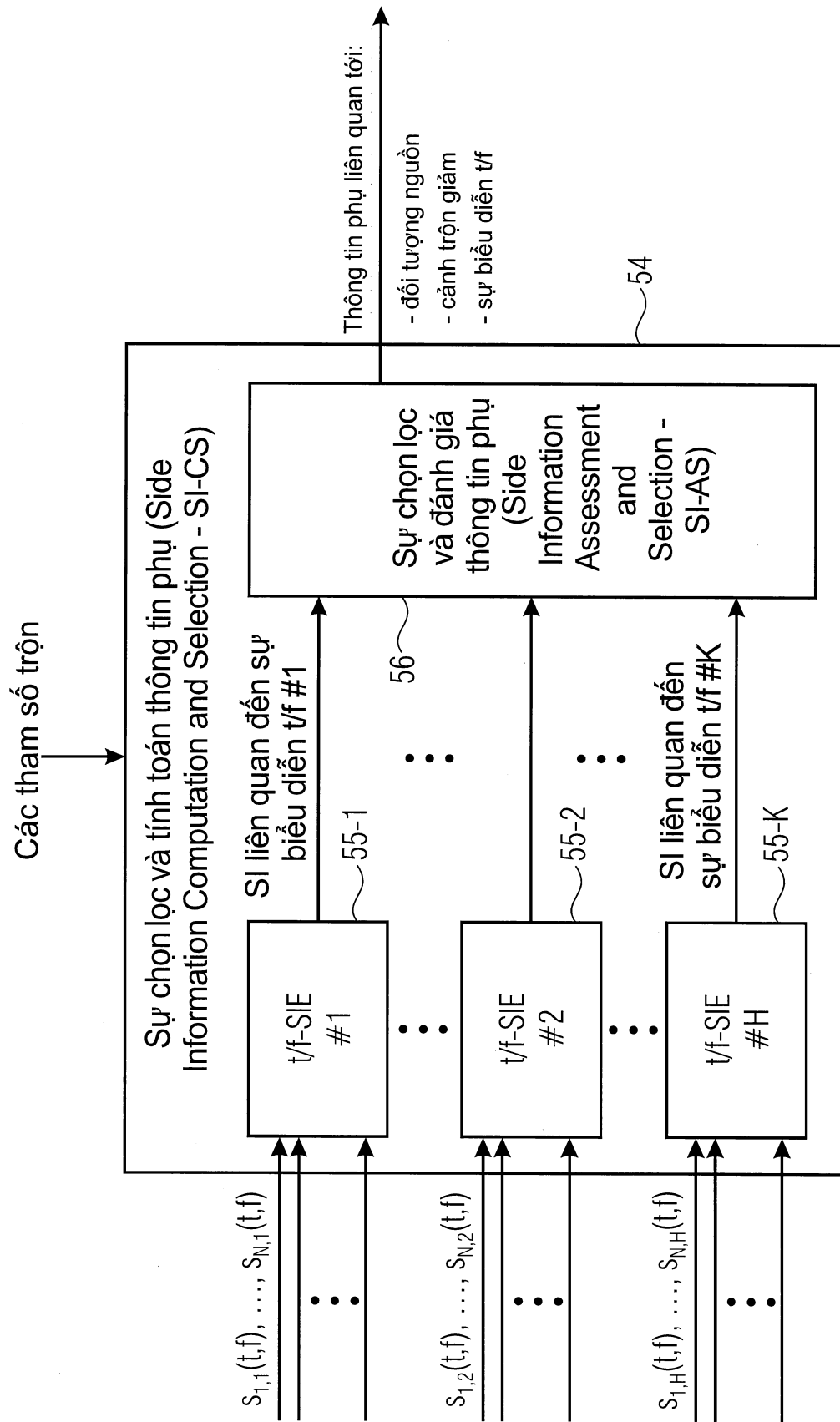


FIG 6

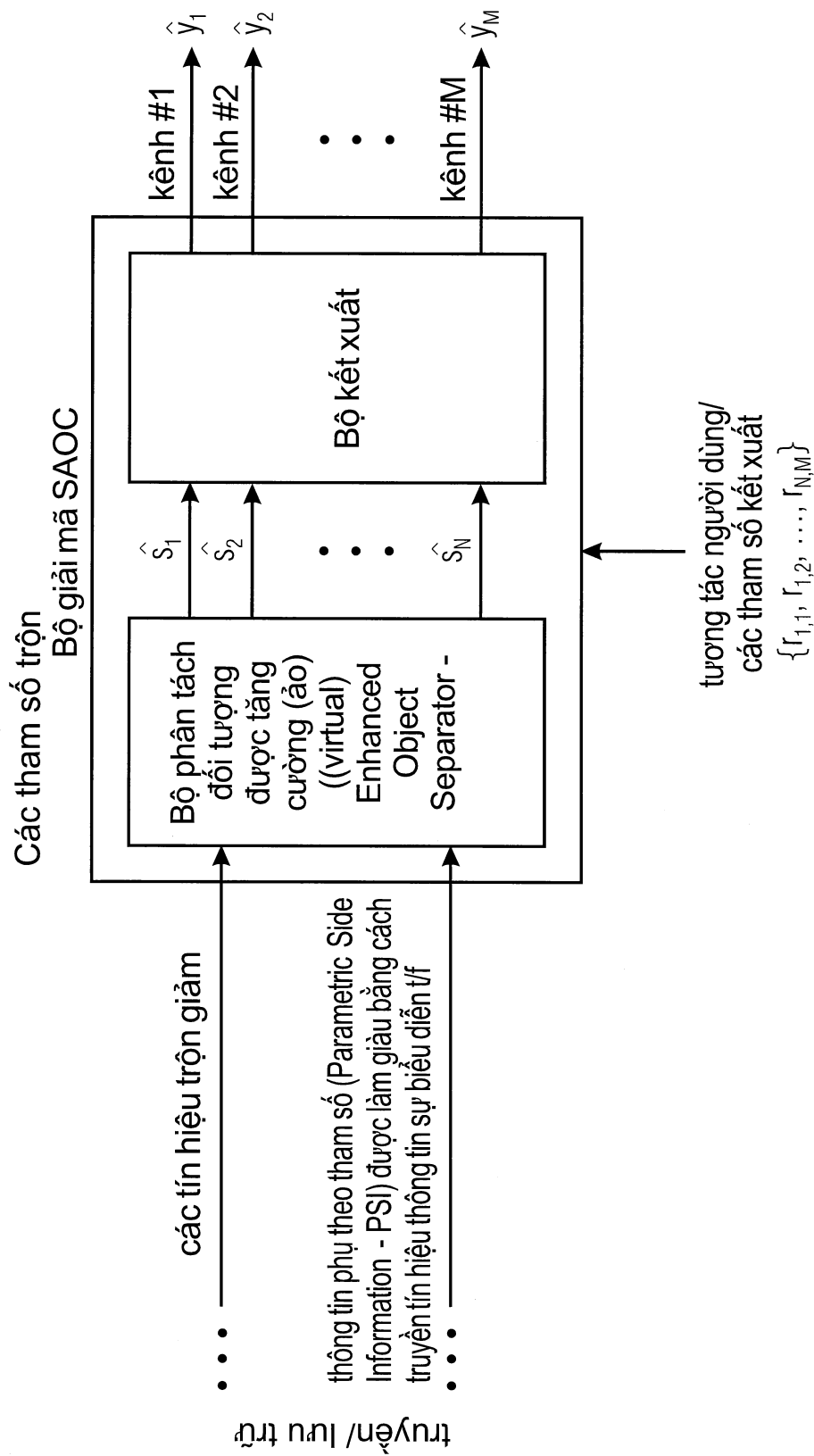


FIG 7

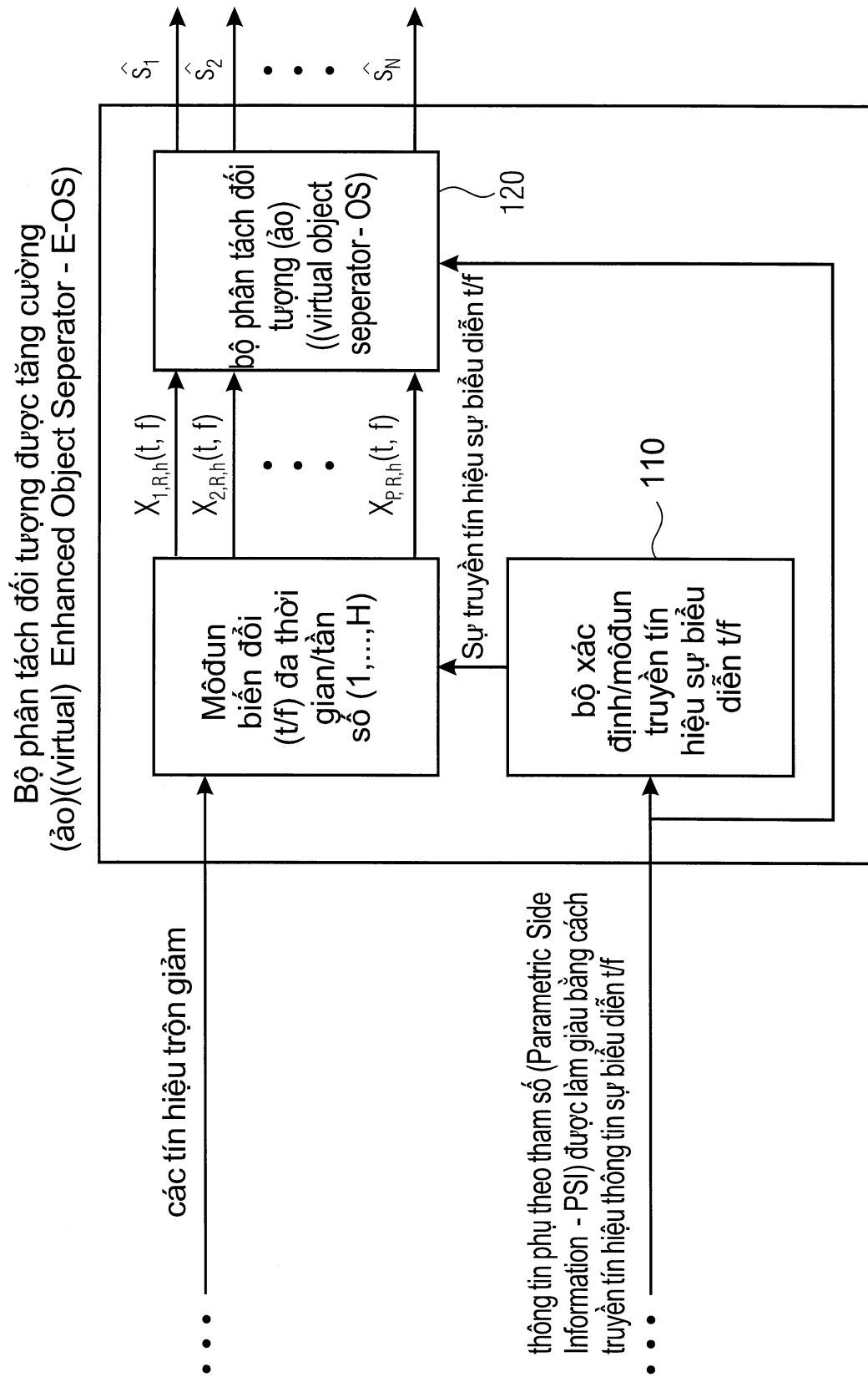


FIG 8

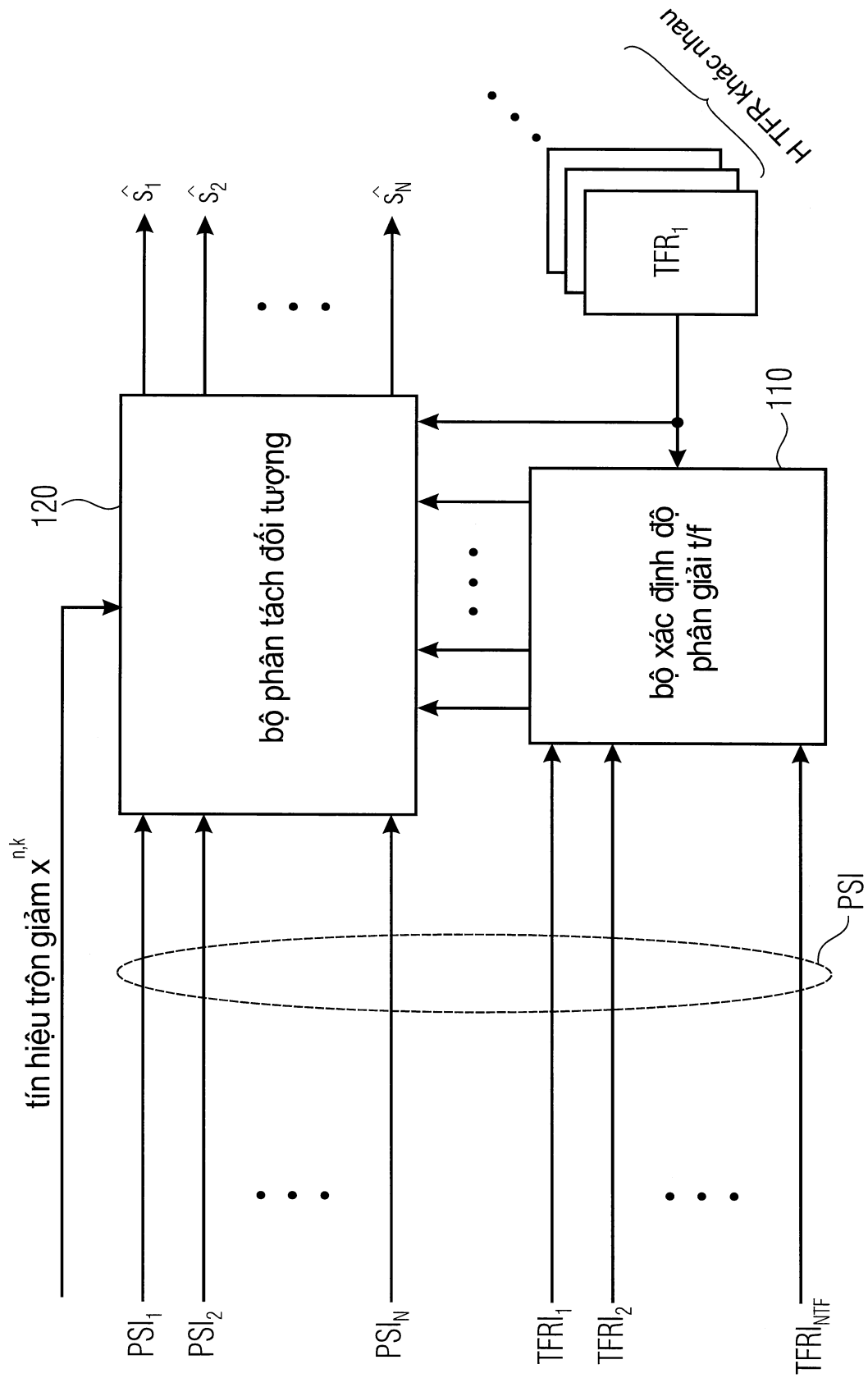


FIG 9

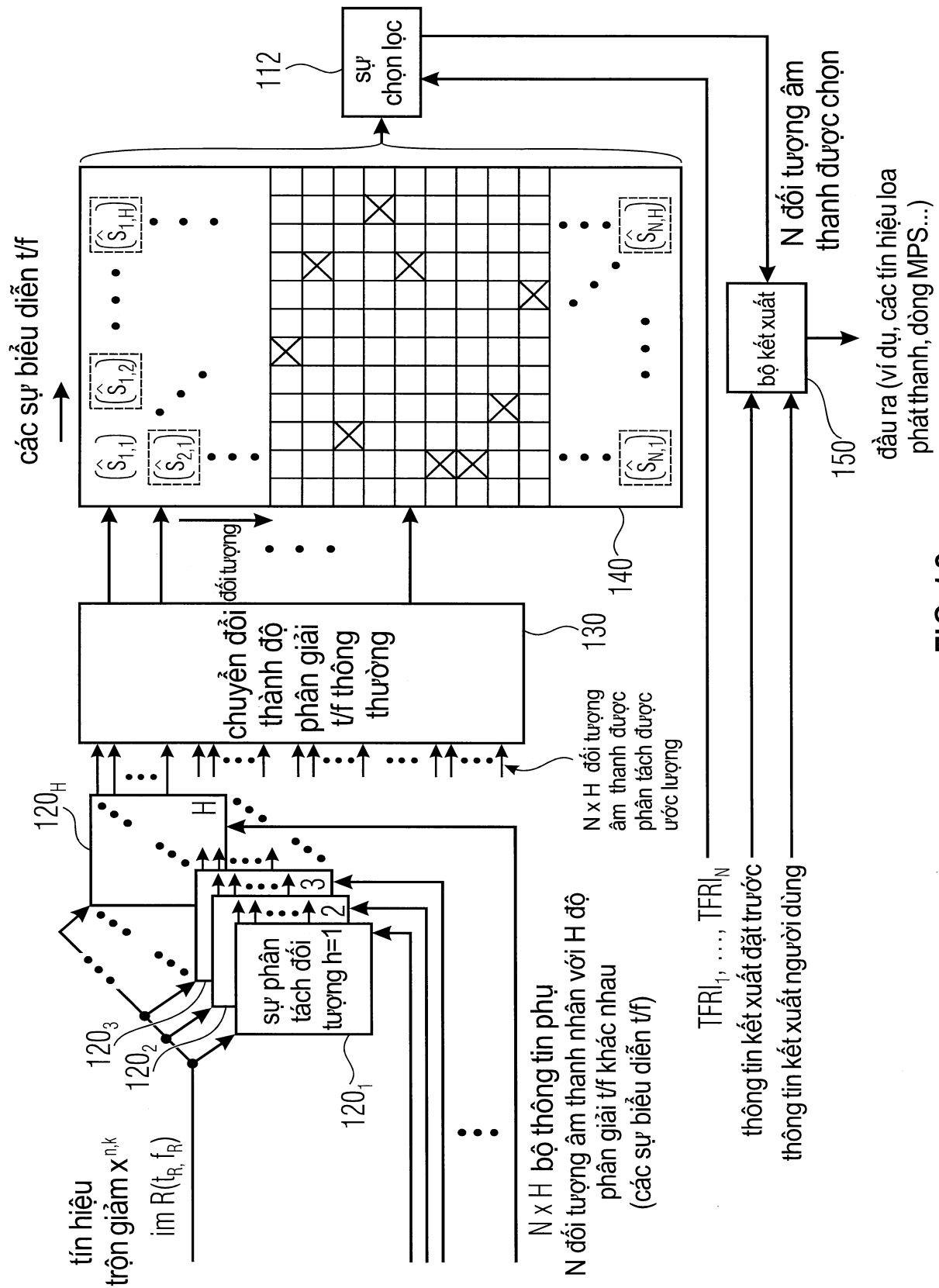


FIG 10

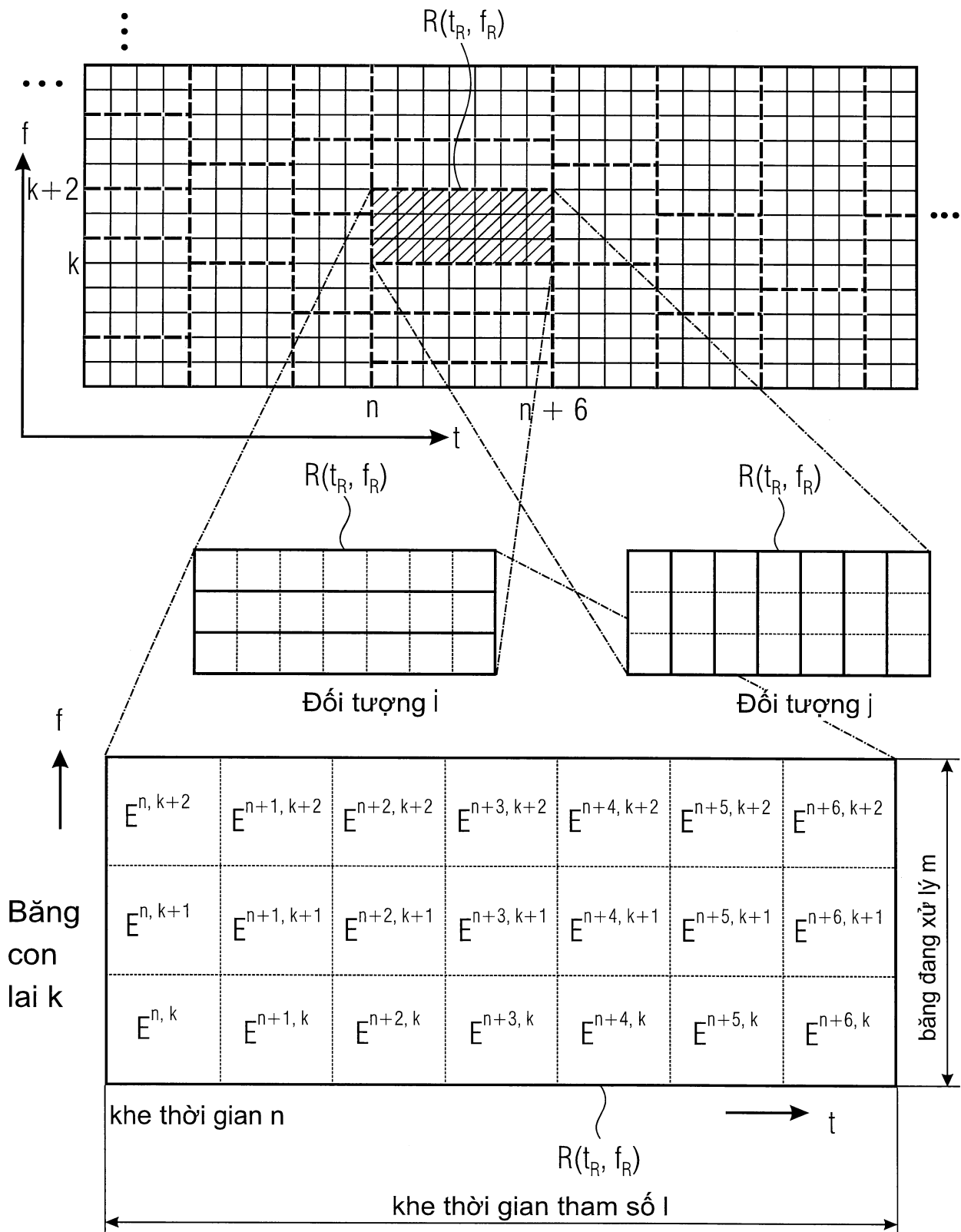


FIG 11

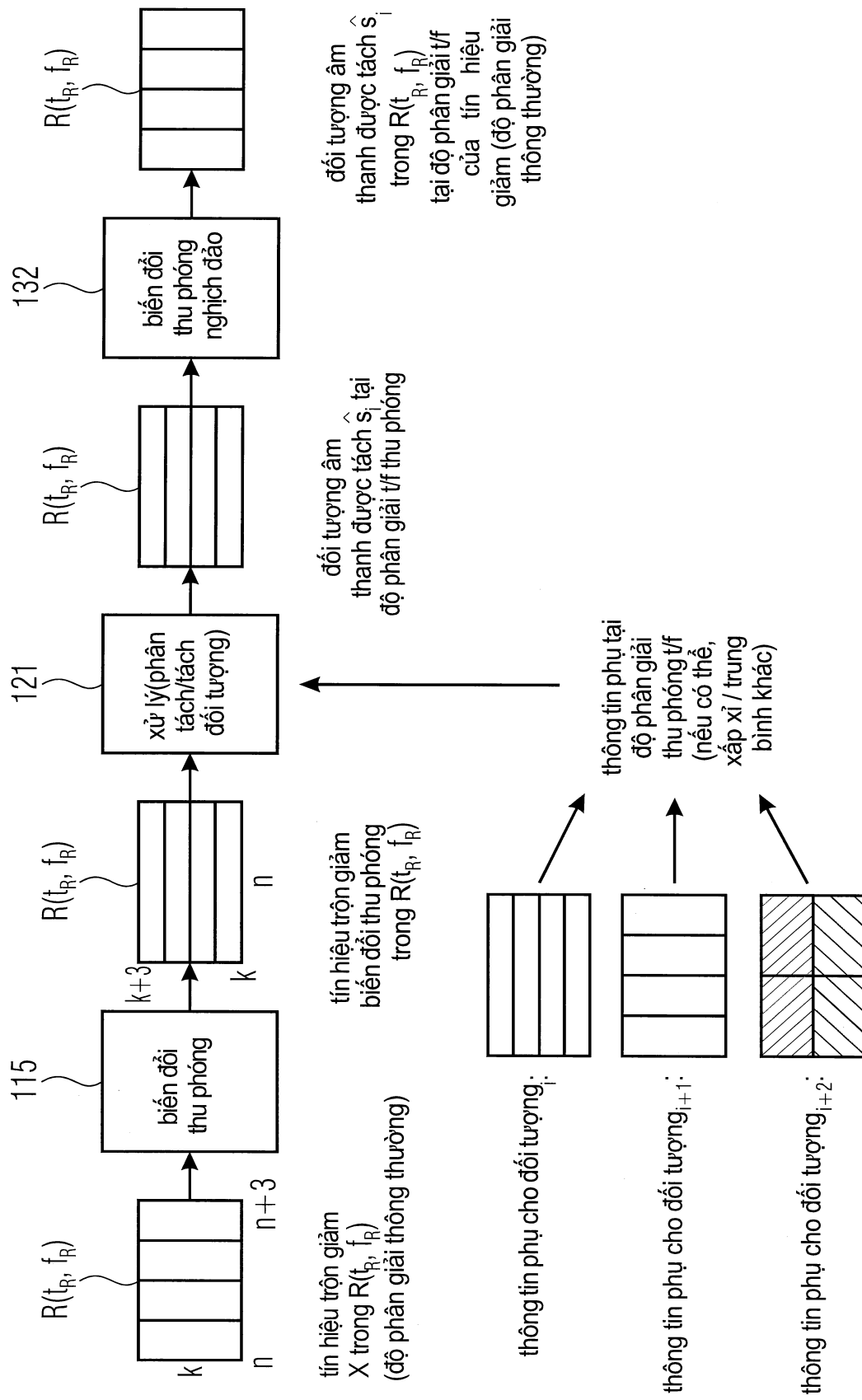


FIG 12

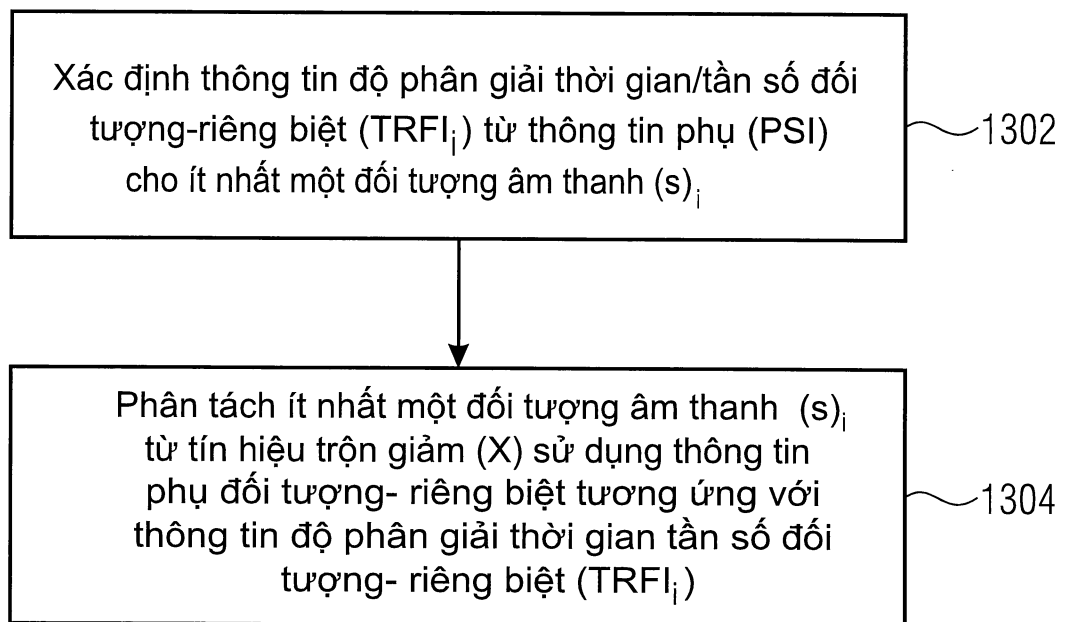


FIG 13

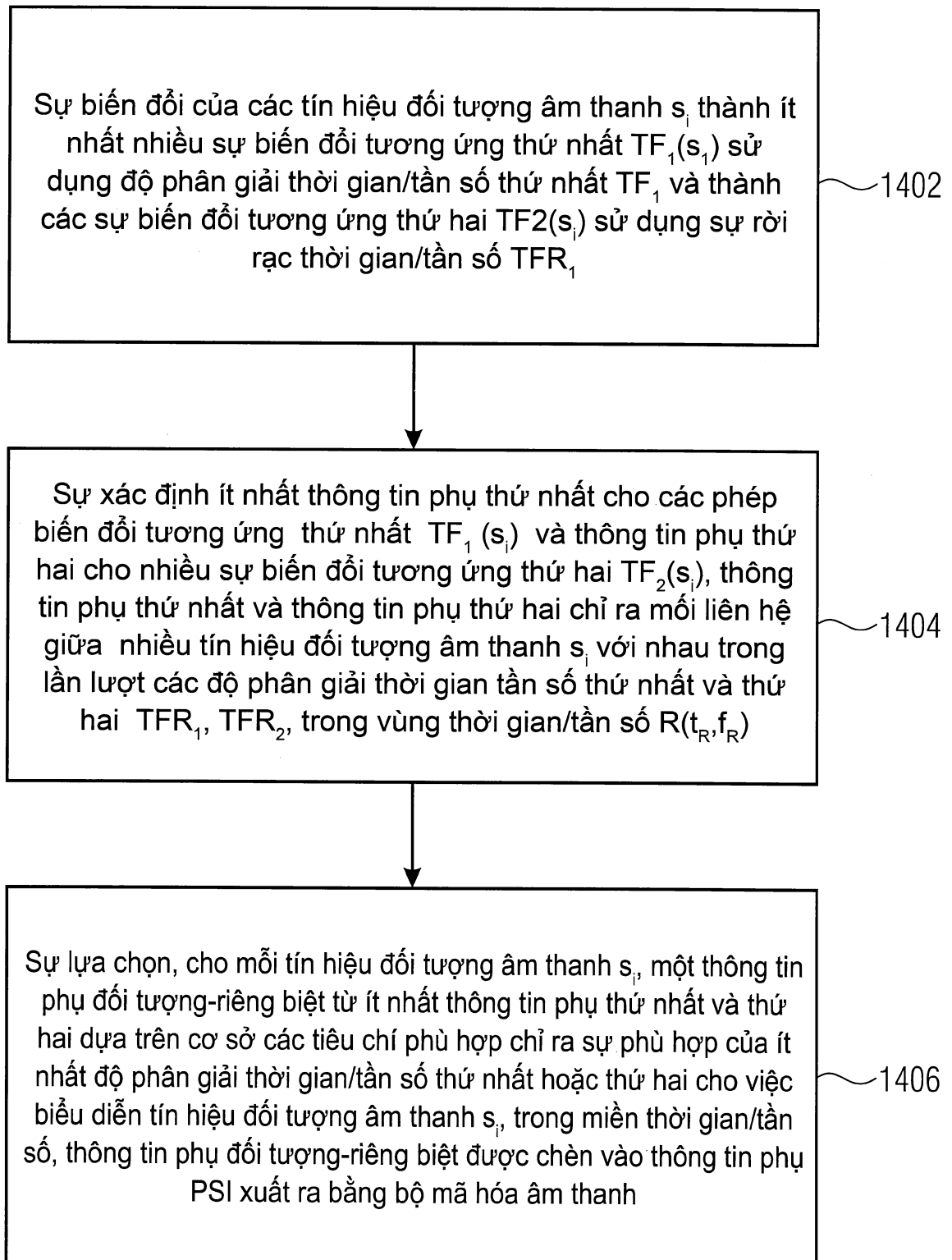


FIG 14