



(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)



1-0034067

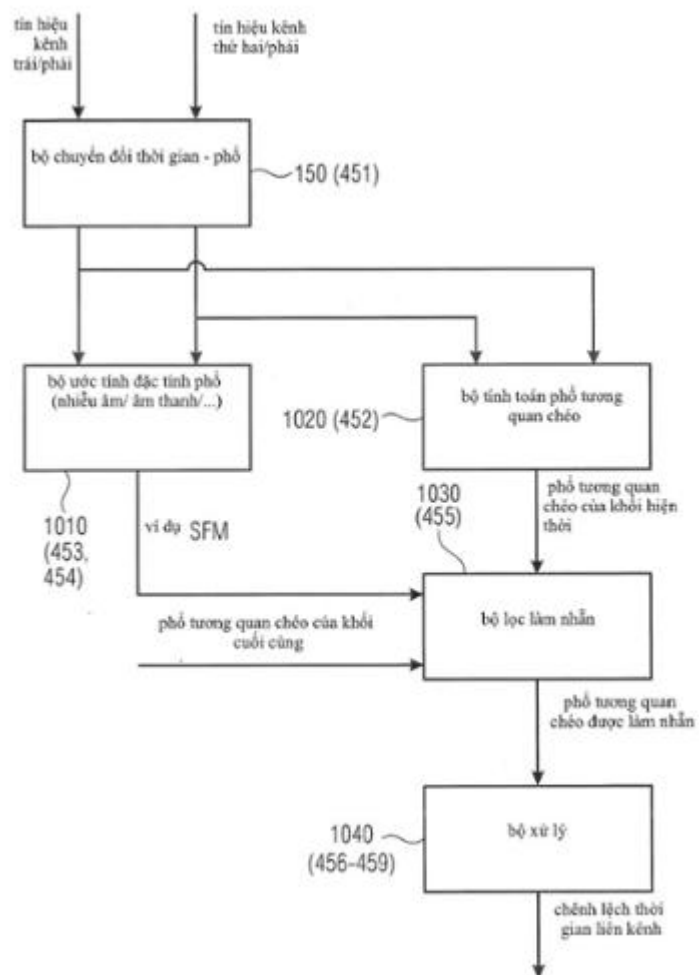
(51)⁷ G10L 19/02; G10L 19/008

(13) B

-
- (21) 1-2018-03664 (22) 20/01/2017
(86) PCT/EP2017/051214 20/01/2017 (87) WO 2017/125563 27/07/2017
(30) 16152453.3 22/01/2016 EP; 16152450.9 22/01/2016 EP
(45) 25/11/2022 416 (43) 26/11/2018 368A
(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) BAYER, Stefan (AT); FOTOPOULOU, Eleni (GR); MULTRUS, Markus (DE);
FUCHS, Guillaume (FR); RAVELLI, Emmanuel (FR); SCHNELL, Markus (DE);
DOEHLA, Stefan (DE); JAEGER, Wolfgang (DE); DIETZ, Martin (DE);
MARKOVIC, Goran (RS).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)
-

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP ƯỚC LƯỢNG CHÊNH LỆCH THỜI GIAN LIÊN KÊNH

(57) Sáng chế đề cập đến thiết bị và phương pháp ước lượng chênh lệch thời gian liên kênh. Thiết bị ước lượng chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai, bao gồm: bộ tính toán (1020) để tính toán phổ tương quan chéo cho khối thời gian từ tín hiệu kênh thứ nhất trong khối thời gian và tín hiệu kênh thứ hai trong khối thời gian; bộ ước lượng đặc tính phổ (1010) để ước lượng đặc tính của phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai cho khối thời gian; bộ lọc làm nhẵn (1030) để làm nhẵn phổ tương quan chéo theo thời gian sử dụng đặc tính phổ để thu phổ tương quan chéo được làm nhẵn; và bộ xử lý (1040) để xử lý phổ tương quan chéo được làm nhẵn để thu được chênh lệch thời gian liên kênh.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến xử lý âm lập thể hoặc nói chung, xử lý đa kênh, trong đó tín hiệu đa kênh có hai kênh chẳng hạn như kênh bên trái và kênh bên phải trong trường hợp tín hiệu âm lập thể hoặc nhiều hơn hai kênh, chẳng hạn như ba, bốn, năm kênh hoặc bất kỳ số lượng kênh nào khác.

Tình trạng kỹ thuật của sáng chế

Tiếng nói âm lập thể và cụ thể hơn là tiếng nói âm lập thể dạng hội thoại ít nhận được sự quan tâm của giới khoa học hơn nhiều so với việc lưu trữ và phát sóng âm nhạc âm thanh nổi. Thực vậy, ngày nay, trong việc truyền thông tiếng nói, truyền dẫn đơn âm vẫn chủ yếu được sử dụng. Tuy nhiên với sự gia tăng của dung lượng và băng thông mạng, có thể hình dung rằng truyền thông dựa trên các kỹ thuật âm lập thể sẽ trở nên phổ biến và mang đến trải nghiệm nghe tốt hơn.

Việc mã hóa hiệu quả vật liệu âm lập thể đã được nghiên cứu trong thời gian dài trong việc mã hóa âm thanh cảm quan của âm nhạc để lưu trữ và phát sóng hiệu quả. Ở tốc độ bit cao, khi việc bảo toàn dạng sóng là rất quan trọng, âm lập thể chênh lệch tổng, được biết đến như là âm lập thể giữa/bên (mid/side - M/S), đã được sử dụng trong thời gian dài. Ở các tốc độ bit thấp, âm lập thể cường độ và mã hóa âm lập thể theo tham số gần đây đã được đưa ra. Kỹ thuật mới nhất đã được áp dụng trong các tiêu chuẩn khác nhau như HeAACv2 và Mpeg USAC. Nó tạo ra sự trộn giảm của tín hiệu hai kênh và kết hợp với thông tin phụ theo không gian gọn nhẹ.

Mã hóa âm lập thể kết hợp thường được xây dựng trên độ phân giải tần số cao, tức là độ phân giải thời gian thấp, sự biến đổi thời gian-tần số của tín hiệu và sau đó không tương thích với độ trễ thấp và việc xử lý miền thời gian được thực hiện trong hầu hết các bộ giải mã tiếng nói. Hơn nữa, tốc độ bit được tạo ra thường cao.

Mặt khác, âm lập thể theo tham số sử dụng giàn lọc phụ được đặt ở phía trước của bộ mã hóa như là bộ xử lý trước và phía sau của bộ giải mã như là bộ xử lý sau. Do đó, âm lập thể theo tham số có thể được sử dụng với các bộ mã hóa tiếng nói thông

thường như ACELP do nó được thực hiện trong MPEG USAC. Hơn nữa, việc tham số hóa các cảnh thính giác có thể đạt được với lượng thông tin phụ tối thiểu mà thích hợp với các tốc độ bit thấp. Tuy nhiên, âm lập thể theo tham số chẳng hạn như trong MPEG USAC không được thiết kế cụ thể cho độ trễ thấp và không mang đến chất lượng đồng đều cho các kịch bản hội thoại khác nhau. Trong sự biểu diễn theo tham số thông thường của cảnh không gian, chiều rộng của hình ảnh âm lập thể được tái tạo theo cách nhân tạo bằng bộ giải tương quan được áp dụng lên hai kênh được tổng hợp và được điều khiển bởi các tham số kết hợp liên kênh (Inter-channel Coherence - IC) được tính toán và truyền dẫn bởi bộ mã hóa. Đối với hầu hết tiếng nói âm lập thể, cách mở rộng hình ảnh âm lập thể này không phù hợp để tạo lại môi trường xung quanh tự nhiên của tiếng nói mà là âm thanh tương đối trực tiếp do nó được sinh ra bởi nguồn đơn được định vị tại vị trí cụ thể trong không gian (đôi khi với một số âm vang trong phòng). Ngược lại, các nhạc cụ âm nhạc cho chiều rộng tự nhiên lớn hơn nhiều so với tiếng nói, mà có thể được mô phỏng tốt hơn bằng cách giải tương quan các kênh.

Các vấn đề cũng xảy ra khi tiếng nói được ghi âm lại với các ống phóng thanh không trùng nhau, như trong cấu hình A-B khi các ống phóng thanh ở xa nhau hoặc để ghi âm hoặc kết xuất lập thể. Các kịch bản này có thể được hình dung khi thu tiếng nói trong các hội nghị từ xa hoặc để tạo kịch bản âm thanh ảo với người nói từ xa trong bộ phận điều khiển đa điểm (multipoint control unit - MCU). Thời gian đến của tín hiệu sau đó khác với một kênh này đến kênh khác không giống việc ghi âm được thực hiện trên các ống phóng thanh trùng nhau như X-Y (ghi âm cường độ) hoặc M-S (ghi âm giữa-bên). Việc tính toán sự kết hợp của hai kênh không được căn chỉnh theo thời gian sau đó có thể bị ước lượng sai mà dẫn đến thất bại trong việc tổng hợp môi trường xung quanh nhân tạo.

Tài liệu tham khảo về tình trạng kỹ thuật liên quan để xử lý âm lập thể là patent Mỹ số 5,434,948 hoặc patent Mỹ số 8,811,621.

Công bố đơn quốc tế số WO 2006/089570 A1 bộc lộ sơ đồ bộ mã hóa/bộ giải mã đa kênh rõ ràng hoặc gần như rõ ràng. Các sơ đồ bộ mã hóa/bộ giải mã đa kênh còn tạo tín hiệu dư dạng sóng. Tín hiệu dư này được truyền dẫn cùng với một hoặc nhiều tham số đa kênh đến bộ giải mã. Trái với bộ giải mã đa kênh theo tham số hoàn

toàn, bộ giải mã được nâng cao tạo ra tín hiệu đầu ra đa kênh có chất lượng đầu ra được nâng cao do tín hiệu dư bổ sung. Ở phía bộ mã hóa, kênh bên trái và kênh bên phải đều được lọc bởi giàn lọc phân tích. Sau đó, đối với mỗi tín hiệu băng con, giá trị căn chỉnh và giá trị độ khuếch đại được tính toán cho băng con. Việc căn chỉnh này sau đó được thực hiện trước khi tiếp tục xử lý. Ở phía bộ giải mã, việc hủy căn chỉnh và xử lý độ khuếch đại được thực hiện và các tín hiệu tương ứng sau đó được tổng hợp bởi giàn lọc tổng hợp để tạo ra tín hiệu bên trái được giải mã và tín hiệu bên phải được giải mã.

Trong các ứng dụng xử lý âm lập thể này, việc tính toán liên kênh hoặc chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai là có lợi để thực hiện diễn hình quy trình căn chỉnh thời gian băng rộng. Tuy nhiên, các ứng dụng khác cũng tồn tại cho việc sử dụng chênh lệch thời gian liên kênh giữa kênh thứ nhất và kênh thứ hai, trong đó các ứng dụng này thuộc lưu trữ và truyền dẫn dữ liệu theo tham số, xử lý âm lập thể/đa kênh bao gồm việc căn chỉnh thời gian của hai kênh, chênh lệch thời gian của việc ước lượng thời gian tới, ví dụ, cho việc xác định vị trí người nói trong phòng, lọc theo không gian bằng điều hướng chùm sóng, khai triển tiền cảnh/hậu cảnh hoặc vị trí của nguồn âm thanh, ví dụ bằng phép đặc tam giác âm thanh để chỉ gọi tên một số.

Đối với tất cả các ứng dụng này, việc xác định hiệu quả, chính xác và chắc chắn chênh lệch thời gian liên kênh giữa tín hiệu thứ nhất và tín hiệu thứ hai là cần thiết.

Việc xác định này đã tồn tại được biết đến bởi thuật ngữ “GCC-PHAT” hoặc nói cách khác, biến đổi pha tương quan chéo được tổng quát hóa. Thông thường, phổ tương quan chéo được tính toán giữa hai tín hiệu kênh và sau đó, hàm gán trọng số được áp dụng cho phổ tương quan chéo để thu được cái được gọi là phổ tương quan chéo được tổng quát hóa trước khi thực hiện biến đổi phổ nghịch đảo chẳng hạn như DFT nghịch đảo đối với phổ tương quan chéo được tổng quát hóa để tìm sự biểu diễn miền thời gian. Sự biểu diễn miền thời gian này biểu diễn các giá trị đối với các độ trễ thời gian nhất định và sau đó đỉnh cao nhất của sự biểu diễn miền thời gian thường tương ứng với độ trễ thời gian hoặc chênh lệch thời gian, tức là, độ trễ thời gian liên kênh của chênh lệch giữa hai tín hiệu kênh.

Tuy nhiên, nó đã được thể hiện rằng, cụ thể trong các tín hiệu mà, ví dụ, khác với tiếng nói sạch không có âm vang hoặc tạp âm nền, độ chắc chắn của kỹ thuật thông thường này không phải tối ưu.

Bản chất kỹ thuật của sáng chế

Do đó, mục đích của sáng chế là đề xuất giải pháp được cải thiện cho việc ước lượng chênh lệch thời gian liên kênh giữa hai tín hiệu kênh.

Mục đích này đạt được bởi thiết bị ước lượng chênh lệch thời gian liên kênh theo điểm 1 yêu cầu bảo hộ, hoặc phương pháp ước lượng chênh lệch thời gian liên kênh theo điểm 15 yêu cầu bảo hộ hoặc chương trình máy tính theo điểm 16 yêu cầu bảo hộ.

Sáng chế dựa trên việc phát hiện ra rằng việc làm nhấn phổ tương quan chéo theo thời gian mà được điều khiển bởi đặc tính phổ của phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai cải thiện đáng kể độ chắc chắn và độ chính xác của việc xác định chênh lệch thời gian liên kênh.

Trong phương án được ưu tiên, đặc tính âm điệu/nhiều âm của phổ được xác định, và trong trường hợp tín hiệu giống âm điệu, việc làm nhấn là mạnh hơn, trong khi đó, trong trường hợp tín hiệu nhiều âm, việc làm nhấn được thực hiện ít mạnh hơn.

Ưu tiên hơn, số đo độ phẳng phổ được sử dụng và, trong trường hợp tín hiệu giống âm điệu, số đo độ phẳng phổ sẽ thấp và việc làm nhấn sẽ trở nên mạnh hơn, và trong trường hợp tín hiệu giống nhiều âm, số đo độ phẳng phổ sẽ cao ví dụ bằng 1 hoặc gần bằng 1 và việc làm nhấn là yếu.

Do đó, theo sáng chế, thiết bị ước lượng chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai bao gồm bộ tính toán để tính toán phổ tương quan chéo cho khối thời gian đối với tín hiệu kênh thứ nhất trong khối thời gian và tín hiệu kênh thứ hai trong khối thời gian. Thiết bị còn bao gồm bộ ước lượng đặc tính phổ để ước lượng đặc tính của phổ của tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai cho khối thời gian và, ngoài ra, bộ lọc làm nhấn để làm nhấn phổ tương quan chéo theo thời gian sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm nhấn. Sau đó, phổ tương quan chéo được làm nhấn được tiếp tục xử lý bởi bộ xử lý để thu

được tham số chênh lệch thời gian liên kênh.

Đối với các phương án được ưu tiên liên quan đến việc xử lý tiếp theo của phổ tương quan chéo được làm nhẵn, phép toán tạo ngưỡng thích ứng được thực hiện, trong đó sự biểu diễn miền thời gian của phổ tương quan chéo được tổng quát hóa được làm nhẵn được phân tích để xác định ngưỡng biến số, mà phụ thuộc vào sự biểu diễn miền thời gian và đỉnh của sự biểu diễn miền thời gian được so sánh với ngưỡng biến số, trong đó chênh lệch thời gian liên kênh được xác định như là độ trễ thời gian được kết hợp với đỉnh nằm trong mối quan hệ được xác định trước với ngưỡng chẳng hạn như lớn hơn ngưỡng.

Trong một phương án, ngưỡng biến số được xác định là giá trị bằng bội số nguyên của một giá trị trong số các giá trị lớn nhất, ví dụ mười phần trăm của các giá trị của sự biểu diễn miền thời gian hoặc, cách khác, trong phương án khác đối với sự xác định biến số, ngưỡng biến số được tính toán bởi phép nhân của ngưỡng biến số và giá trị, trong đó giá trị này phụ thuộc vào đặc tính tỉ số tín hiệu trên nhiễu của các tín hiệu kênh thứ nhất và thứ hai, trong đó giá trị trở nên cao hơn đối với tỉ số tín hiệu trên nhiễu cao hơn và trở nên thấp hơn đối với tỉ số tín hiệu trên nhiễu thấp hơn.

Như đã nêu trước đó, việc tính toán chênh lệch thời gian liên kênh có thể được sử dụng trong nhiều ứng dụng khác nhau chẳng hạn như lưu trữ và truyền dẫn dữ liệu theo tham số, mã hóa/xử lý âm lập thể/đa kênh, căn chỉnh thời gian của hai kênh, chênh lệch thời gian của ước lượng thời gian đến để xác định vị trí người nói trong phòng với hai ống phóng thanh và thiết lập ống phóng thanh đã biết, đối với mục đích điều hướng chùm sóng, lọc theo không gian, khai triển tiền cảnh/hậu cảnh hoặc xác định vị trí của nguồn âm thanh, ví dụ bằng phép đặc tam giác âm thanh dựa trên các chênh lệch thời gian của hai hoặc ba tín hiệu.

Tuy nhiên, sau đây, sự thực hiện được ưu tiên và việc sử dụng tính toán chênh lệch thời gian liên kênh được mô tả nhằm mục đích cho việc căn chỉnh thời gian bằng rộng của hai tín hiệu âm lập thể trong quy trình mã hóa tín hiệu đa kênh có ít nhất hai kênh.

Thiết bị mã hóa tín hiệu đa kênh có ít nhất hai kênh bao gồm bộ xác định tham số để xác định một mặt là tham số căn chỉnh bằng rộng và mặt khác là nhiều tham số

căn chỉnh băng hẹp. Các tham số này được sử dụng bởi bộ căn chỉnh tín hiệu để căn chỉnh ít nhất hai kênh sử dụng các tham số này để thu được các kênh được căn chỉnh. Sau đó, bộ xử lý tín hiệu tính toán tín hiệu giữa và tín hiệu bên sử dụng các kênh được căn chỉnh và tín hiệu giữa và tín hiệu bên sau đó được mã hóa và được chuyển tiếp vào tín hiệu đầu ra được mã hóa mà còn có tham số căn chỉnh băng rộng và các tham số căn chỉnh băng hẹp như là thông tin phụ.

Ở phía bộ giải mã, bộ giải mã tín hiệu giải mã tín hiệu giữa được mã hóa và tín hiệu bên được mã hóa để thu các tín hiệu giữa và tín hiệu bên được giải mã. Các tín hiệu này sau đó được xử lý bởi bộ xử lý tín hiệu để tính toán kênh thứ nhất được giải mã và kênh thứ hai được giải mã. Các kênh được giải mã sau đó được hủy căn chỉnh sử dụng thông tin trên tham số căn chỉnh băng rộng và thông tin trên nhiều tham số băng hẹp được bao gồm trong tín hiệu đa kênh được mã hóa để thu tín hiệu đa kênh được giải mã.

Trong việc thực hiện cụ thể, tham số căn chỉnh băng rộng là tham số căn chỉnh thời gian liên kênh và nhiều tham số căn chỉnh băng hẹp là các chênh lệch pha liên kênh.

Sáng chế dựa trên việc phát hiện ra rằng cụ thể đối với các tín hiệu tiếng nói trong đó có nhiều hơn một người nói, mà còn đối với các tín hiệu âm thanh khác trong đó có một vài nguồn âm thanh, các địa điểm khác nhau của nguồn âm thanh được ánh xạ vào hai kênh của tín hiệu đa kênh có thể được xem xét cho việc sử dụng tham số căn chỉnh băng rộng chẳng hạn như tham số chênh lệch thời gian liên kênh được áp dụng cho toàn bộ phổ của một hoặc nhiều kênh. Ngoài tham số căn chỉnh băng rộng, các tác giả sáng chế còn tìm ra rằng một số tham số căn chỉnh băng hẹp mà khác với từng băng con còn dẫn đến sự căn chỉnh tín hiệu tốt hơn ở cả hai kênh.

Do đó, việc căn chỉnh băng rộng tương ứng với độ trễ thời gian tương tự trong mỗi băng con cùng với căn chỉnh pha tương ứng với các sự luân phiên pha khác nhau đối với các băng con khác nhau dẫn đến sự căn chỉnh tối ưu của cả hai kênh trước khi hai kênh này sau đó được chuyển đổi thành sự biểu diễn giữa/bên mà sau đó được tiếp tục mã hóa. Do việc căn chỉnh tối ưu thu được, một mặt năng lượng trong tín hiệu giữa là cao nhất có thể và mặt khác năng lượng trong tín hiệu bên là nhỏ nhất có thể sao cho

kết quả mã hóa tối ưu với tốc độ bit thấp nhất có thể hoặc chất lượng âm thanh cao nhất có thể cho tốc độ bit nhất định có thể thu được.

Cụ thể đối với vật liệu tiếng nói chuyển đổi, có vẻ như người nói thường hoạt động ở hai địa điểm khác nhau. Ngoài ra, thông thường trường hợp chỉ có một người nói mà nói từ địa điểm thứ nhất và sau đó người nói thứ hai nói từ địa điểm hoặc vị trí thứ hai. Ảnh hưởng của các vị trí khác nhau lên hai kênh chẳng hạn như kênh thứ nhất hoặc bên trái và kênh thứ hai hoặc bên phải được phản ánh bởi thời gian tới khác nhau và do đó, độ trễ thời gian nhất định giữa hai kênh do các địa điểm khác nhau, và độ trễ thời gian này thay đổi theo thời gian. Thông thường, sự ảnh hưởng này được phản ánh trong hai tín hiệu kênh như là sự hủy căn chỉnh băng rộng mà có thể được giải quyết bởi tham số căn chỉnh băng rộng.

Mặt khác, các ảnh hưởng khác, cụ thể từ âm vang hoặc các nguồn nhiễu âm khác có thể được xem xét bởi các tham số căn chỉnh pha riêng biệt cho các băng riêng biệt mà chồng lên các thời gian tới khác nhau của băng rộng hoặc việc hủy căn chỉnh băng rộng của cả hai kênh.

Theo khía cạnh sử dụng cả tham số căn chỉnh băng rộng và nhiều tham số căn chỉnh băng hẹp trên đỉnh của tham số căn chỉnh băng rộng dẫn đến việc căn chỉnh kênh tối ưu bên phía bộ mã hóa để thu được sự biểu diễn giữa/bên tốt và rất nhỏ gọn trong khi mặt khác, việc hủy căn chỉnh tương ứng sau khi giải mã phía bộ giải mã mang đến chất lượng âm thanh tốt cho tốc độ bit nhất định hoặc trong tốc độ bit thấp đối với chất lượng âm thanh được yêu cầu nhất định.

Hiệu quả có lợi của sáng chế là đề xuất sơ đồ mã hóa âm lập thể mới mà phù hợp hơn nhiều cho việc chuyển đổi tiếng nói âm lập thể so với các sơ đồ mã hóa âm lập thể đã biết. Theo sáng chế, các kỹ thuật âm lập thể theo tham số và các kỹ thuật mã hóa âm lập thể kết hợp được kết hợp một cách cụ thể bằng cách khai thác sự chênh lệch thời gian liên kênh xảy ra trong các kênh của tín hiệu đa kênh, đặc biệt không chỉ trong trường hợp các nguồn tiếng nói mà còn trong trường hợp các nguồn âm thanh khác.

Một số phương án đề xuất hiệu quả có lợi được thảo luận sau.

Phương pháp mới là phương thức lai hòa trộn các yếu tố từ âm lập thể theo tham số và âm lập thể M/S thông thường. Trong M/S thông thường, các kênh được trộn giảm một cách bị động để tạo ra tín hiệu giữa và tín hiệu bên. Quy trình có thể còn được mở rộng bằng cách luân phiên kênh sử dụng biến đổi Karhunen-Loeve (Karhunen-Loeve transform - KLT), cũng được biết đến như là phương pháp phân tích thành phần chính (Principal Component Analysis - PCA) trước khi lấy tổng và lấy vi phân các kênh. Tín hiệu giữa được mã hóa trong mã hóa mã sơ cấp trong khi tín hiệu bên được truyền đến bộ mã hóa thứ cấp. Âm lập thể M/S được phát triển có thể còn sử dụng sự dự báo của tín hiệu bên bằng kênh giữa được mã hóa trong khung hiện thời hoặc khung trước đó. Mục đích chính của việc luân phiên và dự báo là tối đa hóa năng lượng của tín hiệu giữa trong khi tối thiểu hóa năng lượng của tín hiệu bên. Âm lập thể M/S là bảo toàn dạng sóng và trong khía cạnh này là chắc chắn đối với bất kỳ tình huống âm lập thể nào, nhưng có thể rất tốn kém về mặt tiêu thụ bit.

Để có hiệu suất cao nhất ở các tốc độ bit thấp, âm lập thể theo tham số tính toán và mã hóa các tham số, giống như các chênh lệch mức liên kênh (Inter-channel Level difference - ILD), các chênh lệch pha liên kênh (Inter-channel Phase difference - IPD), các chênh lệch thời gian liên kênh (Inter-channel Time difference - ITD) và kết hợp liên kênh (Inter-channel Coherence - IC). Chúng biểu diễn hình ảnh âm lập thể một cách nhỏ gọn và là các tín hiệu điều khiển của cảnh âm thanh (định vị nguồn, ghép nhĩ, chiều rộng của âm lập thể...). Mục đích sau đó là tham số hóa cảnh âm lập thể và chỉ mã hóa tín hiệu trộn giảm mà có thể nằm tại bộ giải mã với sự trợ giúp của các tín hiệu điều khiển âm lập thể được truyền dẫn có thể được không gian hóa một lần nữa.

Phương pháp này kết hợp hai giải pháp. Thứ nhất, các tín hiệu điều khiển ITD và IPD được tính toán và áp dụng lên hai kênh. Mục đích là để biểu diễn chênh lệch thời gian trong băng rộng và pha trong các băng tần số khác nhau. Hai kênh sau đó được căn chỉnh theo thời gian và theo pha và việc mã hóa M/S sau đó được thực hiện. ITD và IPD được phát hiện ra là có ích trong việc tạo mô hình tiếng nói âm lập thể và là sự thay thế tốt cho sự luân phiên dựa trên KLT trong M/S. Không giống việc mã hóa theo tham số đơn thuần, môi trường xung quanh không được tạo mô hình thêm bởi các IC mà được tạo mô hình một cách trực tiếp bằng tín hiệu bên mà được mã hóa và/hoặc dự báo. Các tác giả sáng chế đã phát hiện ra rằng phương pháp này chắc chắn hơn đặc

biệt khi xử lý các tín hiệu tiếng nói.

Việc tính toán và xử lý các ITD là phần chủ yếu của sáng chế. Các ITD đã được khai thác trong lĩnh vực kỹ thuật mã hóa tín hiệu điều khiển lập thể (Binaural Cue Coding - BCC) nhưng theo cách không hiệu quả khi các ITD thay đổi theo thời gian. Để tránh các thiếu sót này, việc tạo cửa sổ cụ thể được thiết kế để làm nhấn sự chuyển tiếp giữa hai ITD khác nhau và có thể chuyển từ người nói này đến người nói được định vị tại địa điểm khác một cách trơn tru.

Các phương án khác đề cập đến quy trình mà, ở phía bộ mã hóa, việc xác định tham số để xác định nhiều tham số căn chỉnh băng hẹp được thực hiện sử dụng các kênh mà đã được căn chỉnh với tham số căn chỉnh băng rộng được xác định trước đó.

Tương ứng, việc hủy căn chỉnh băng hẹp trên phía bộ giải mã được thực hiện trước khi việc hủy căn chỉnh băng rộng được thực hiện sử dụng tham số căn chỉnh băng rộng đơn lẻ điển hình.

Trong các phương án khác, ưu tiên rằng, ở phía bộ mã hóa nhưng quan trọng hơn là ở phía bộ giải mã, một số loại tạo cửa sổ và thao tác chồng lắp-cộng của bất kỳ loại nào của nhiều xuyên kênh từ một khối sang khối tiếp theo được thực hiện sau tất cả các căn chỉnh và cụ thể, sau việc căn chỉnh theo thời gian sử dụng tham số căn chỉnh băng rộng. Việc này tránh các thành phần lạ nghe được chẳng hạn như tiếng lách cách khi tham số căn chỉnh băng rộng hoặc theo thời gian thay đổi từ khối này sang khối kia.

Trong các phương án khác, các độ phân giải phổ khác nhau được áp dụng. Cụ thể, các tín hiệu kênh phải tuân theo sự chuyển đổi thời gian-phổ có độ phân giải tần số cao chẳng hạn như phổ DFT trong khi các tham số chẳng hạn như các tham số căn chỉnh băng hẹp được xác định cho các băng tham số có độ phân giải phổ thấp. Thông thường, băng tham số có nhiều hơn một vạch phổ so với phổ tín hiệu và thường có tập hợp các vạch phổ từ phổ DFT. Ngoài ra, các băng tham số tăng từ tần số thấp đến tần số cao để xem xét các vấn đề tâm thính học.

Các phương án khác đề cập đến việc sử dụng bổ sung tham số mức chẳng hạn như chênh lệch liên mức hoặc các quy trình khác để xử lý tín hiệu bên chẳng hạn như

các tham số điền đầy âm lập thể, v.v.. Tín hiệu bên được mã hóa có thể được biểu diễn bởi chỉnh bản thân tín hiệu bên thực, hoặc bởi tín hiệu dự báo được thực hiện sử dụng tín hiệu bên của khung hiện thời hoặc bất kỳ khung nào khác, hoặc bởi tín hiệu bên hoặc tín hiệu dự báo bên chỉ nằm trong tập con gồm các băng và các tham số dự báo cho các băng còn lại, hoặc thậm chí bởi các tham số dự báo cho tất cả các băng mà không có thông tin tín hiệu bên về độ phân giải tần số cao. Do đó, theo cách thay thế cuối cùng ở trên tín hiệu bên được mã hóa chỉ được biểu diễn bởi tham số dự báo cho mỗi băng tham số hoặc chỉ cho tập con của các băng tham số sao cho đối với các băng tham số còn lại, không tồn tại bất kỳ thông tin nào trên tín hiệu bên ban đầu.

Ngoài ra, ưu tiên có nhiều các tham số căn chỉnh băng hẹp không phải cho tất cả các băng tham số phản ánh toàn bộ băng thông của tín hiệu băng rộng mà chỉ cho tập hợp các băng thấp hơn chẳng hạn như thấp hơn 50% băng tham số. Mặt khác, các tham số điền đầy âm lập thể không được sử dụng để ghép các băng thấp hơn, vì đối với các băng này, bản thân tín hiệu bên hoặc tín hiệu dự báo được truyền dẫn để đảm bảo rằng ít nhất đối với các băng thấp hơn, sự biểu diễn hiệu chỉnh dạng sóng là có sẵn. Mặt khác, tín hiệu bên không được truyền dẫn trong sự biểu diễn dạng sóng chính xác cho các băng cao hơn để tăng thêm tốc độ bit, nhưng tín hiệu bên thường được biểu diễn bởi các tham số điền đầy âm lập thể.

Ngoài ra, ưu tiên thực hiện toàn bộ phân tích tham số và căn chỉnh trong một và miền tần số tương tự dựa trên phổ DFT tương tự. Để đạt được mục đích này, ưu tiên hơn là sử dụng tương quan chéo được tổng quát hóa với kỹ thuật biến đổi pha (GCC-PHAT) nhằm mục đích xác định chênh lệch thời gian liên kênh. Trong phương án được ưu tiên của quy trình này, sự làm nhẵn phổ tương quan dựa trên thông tin trên hình dạng phổ, thông tin này ưu tiên là số đo độ phẳng phổ được thực hiện theo cách mà sự làm nhẵn sẽ yếu trong trường hợp các tín hiệu giống nhiễu âm và sự làm nhẵn trở nên mạnh trong trường hợp tín hiệu giống âm điệu.

Hơn nữa, ưu tiên hơn là thực hiện sự luân phiên theo pha đặc biệt trong đó biên độ kênh được xem xét. Cụ thể, sự luân phiên theo pha được phân bố giữa hai kênh nhằm mục đích căn chỉnh trên phía bộ mã hóa và đương nhiên, để nhằm mục đích hủy căn chỉnh trên phía bộ giải mã trong đó kênh có biên độ cao hơn được coi như là kênh

dẫn trước và sẽ ít chịu ảnh hưởng bởi sự luân phiên theo pha, tức là sẽ được luân phiên ít hơn so với kênh có biên độ thấp hơn.

Hơn nữa, việc tính toán chênh lệch tổng được thực hiện sử dụng việc định tỉ lệ năng lượng với hệ số định tỉ lệ được suy ra từ các năng lượng của cả hai kênh và ngoài ra được giới hạn trong khoảng nhất định để đảm bảo rằng việc tính toán bên/giữa không ảnh hưởng quá nhiều đến năng lượng. Tuy nhiên, mặt khác, cần lưu ý rằng, để đạt được mục đích của sáng chế, kiểu chuyển đổi năng lượng này là không quan trọng như trong các quy trình trong tình trạng kỹ thuật, do thời gian và pha đã được căn chỉnh trước đó. Do đó, các dao động năng lượng do việc tính toán tín hiệu giữa và tín hiệu bên từ bên trái và bên phải (trên phía bộ mã hóa) hoặc do việc tính toán tín hiệu bên trái và bên phải từ giữa và bên (trên phía bộ giải mã) là không đáng kể như trong tình trạng kỹ thuật.

Mô tả vắn tắt các hình vẽ

Tiếp theo, các phương án ưu tiên của sáng chế được thảo luận với việc tham chiếu đến các hình vẽ đi kèm trong đó:

Fig.1 là sơ đồ khối của sự thực hiện được ưu tiên của thiết bị mã hóa tín hiệu đa kênh;

Fig.2 là phương án được ưu tiên của thiết bị giải mã tín hiệu đa kênh được mã hóa;

Fig.3 minh họa các độ phân giải tần số khác nhau và các khía cạnh liên quan đến tần số đối với các phương án nhất định;

Fig.4 minh họa lưu đồ của các quy trình được thực hiện trong thiết bị mã hóa nhằm mục đích căn chỉnh các kênh;

Fig.4b minh họa phương án được ưu tiên của các quy trình được thực hiện trong miền tần số;

Fig.4c minh họa phương án được ưu tiên của các quy trình được thực hiện trong thiết bị mã hóa sử dụng cửa sổ phân tích và các phần đệm số không và các khoảng chồng lấp;

Fig.4d minh họa lưu đồ của các quy trình khác được thực hiện trong thiết bị mã hóa;

Fig.4e minh họa lưu đồ thể hiện sự thực hiện được ưu tiên của việc ước lượng chênh lệch thời gian liên kênh;

Fig.5 minh họa lưu đồ thể hiện phương án khác của các quy trình được thực hiện trong thiết bị mã hóa;

Fig.6 minh họa sơ đồ khối của phương án của bộ mã hóa;

Fig.6b minh họa lưu đồ của phương án tương ứng của bộ giải mã;

Fig.7 minh họa kịch bản của sổ được ưu tiên với các cửa sổ hình sin chồng lấp thấp với đệm số không cho việc phân tích và tổng hợp thời gian - tần số âm lập thể;

Fig.8 minh họa bảng thể hiện việc tiêu thụ bit của các giá trị tham số khác nhau;

Fig.9a minh họa các quy trình được thực hiện bởi thiết bị giải mã tín hiệu đa kênh được mã hóa trong phương án được ưu tiên;

Fig.9b minh họa sự thực hiện được ưu tiên của thiết bị giải mã tín hiệu đa kênh được mã hóa;

Fig.9c minh họa quy trình được thực hiện trong ngữ cảnh việc hủy căn chỉnh băng rộng trong ngữ cảnh giải mã tín hiệu đa kênh được mã hóa;

Fig.10a minh họa phương án của thiết bị ước lượng chênh lệch thời gian liên kênh;

Fig.10b minh họa sự biểu diễn theo giản đồ việc xử lý tiếp tục tín hiệu trong đó chênh lệch thời gian liên kênh được áp dụng;

Fig.11a minh họa các quy trình được thực hiện bởi bộ xử lý của Fig.10a;

Fig.11b minh họa các quy trình khác được thực hiện bởi bộ xử lý của Fig.10a;

Fig.11c minh họa sự thực hiện khác của việc tính toán ngưỡng biến số và việc sử dụng ngưỡng biến số trong phân tích sự biểu diễn miền thời gian;

Fig.11d minh họa phương án thứ nhất để xác định ngưỡng biến số;

Fig.11e minh họa sự thực hiện khác để xác định ngưỡng;

Fig.12 minh họa sự biểu diễn miền thời gian cho phổ tương quan chéo được làm nhẵn đối với tín hiệu tiếng nói sạch;

Fig.13 minh họa sự biểu diễn miền thời gian của phổ tương quan chéo được làm nhẵn đối với tín hiệu tiếng nói có nhiễu âm và môi trường xung quanh.

Mô tả chi tiết sáng chế

Fig.10a minh họa phương án của thiết bị ước lượng chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất chẳng hạn như kênh bên trái và tín hiệu kênh thứ hai chẳng hạn như kênh bên phải. Các kênh này được đưa vào bộ chuyển đổi thời gian-phổ 150 mà được minh họa bổ sung ở Fig.4e, mục 451.

Hơn nữa, các sự biểu diễn miền thời gian của các tín hiệu kênh bên trái và bên phải được đưa vào bộ tính toán 1020 để tính toán phổ tương quan chéo cho khối thời gian từ tín hiệu kênh thứ nhất trong khối thời gian và tín hiệu kênh thứ hai trong khối thời gian. Hơn nữa, thiết bị bao gồm bộ ước lượng đặc tính phổ 1010 để ước lượng đặc tính của phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai đối với khối thời gian. Thiết bị còn bao gồm bộ lọc làm nhẵn 1030 để làm nhẵn phổ tương quan chéo theo thời gian sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm nhẵn. Thiết bị còn bao gồm bộ xử lý 1040 để xử lý phổ tương quan chéo được làm nhẵn để thu được chênh lệch thời gian liên kênh.

Cụ thể, các chức năng của bộ ước lượng đặc tính phổ cũng được phản ánh bởi Fig.4e, mục 453, 454 trong phương án được ưu tiên.

Hơn nữa, các chức năng của bộ tính toán phổ tương quan chéo 1020 cũng được phản ánh bởi mục 452 trên Fig.4e được mô tả sau trong phương án được ưu tiên.

Tương ứng, các chức năng của bộ lọc làm nhẵn 1030 cũng được phản ánh bởi mục 453 trong ngữ cảnh của Fig.4e được mô tả sau. Ngoài ra, các chức năng của bộ xử lý 1040 cũng được mô tả trong ngữ cảnh của Fig.4e trong phương án được ưu tiên như các mục 456 đến 459.

Tốt hơn là, việc ước lượng đặc tính phổ tính toán độ nhiễu âm và âm điệu của phổ trong đó sự thực hiện được ưu tiên là sự tính toán số đo độ phẳng phổ gần bằng 0 trong trường các tín hiệu âm điệu hoặc không phải nhiễu âm và gần bằng 1 trong

trường hợp các tín hiệu nhiều âm hoặc giống nhiều âm.

Cụ thể, bộ lọc làm nhấn sau đó được tạo cấu hình để áp dụng việc làm nhấn mạnh hơn với mức làm nhấn thứ nhất theo thời gian trong trường hoặc đặc tính ít nhiều âm thứ nhất hoặc đặc tính nhiều âm điệu thứ nhất, hoặc để áp dụng việc làm nhấn yếu hơn với mức làm nhấn thứ hai theo thời gian trong trường hợp đặc tính nhiều nhiều âm thứ hai hoặc đặc tính ít âm điệu thứ hai.

Cụ thể, mức làm nhấn thứ nhất lớn hơn mức làm nhấn thứ hai, trong đó đặc tính nhiều âm thứ nhất ít nhiều hơn so với đặc tính nhiều âm thứ hai hoặc đặc tính âm điệu thứ nhất nhiều âm điệu hơn đặc tính âm điệu thứ hai. Sự thực hiện được ưu tiên là số đo độ phẳng phổ.

Hơn nữa, như được minh họa trên Fig.11a, bộ xử lý được ưu tiên thực hiện để chuẩn hóa phổ tương quan chéo được làm nhấn như được minh họa tại 456 trên Fig.4e và Fig.11a trước khi thực hiện tính toán sự biểu diễn miền thời gian trong bước 1031 tương ứng với các bước 457 và 458 trong phương án của Fig.4e. Tuy nhiên, như cũng được thể hiện trên Fig..11a, bộ xử lý có thể hoạt động mà không cần chuẩn hóa trong bước 456 trên Fig.4e. Sau đó, bộ xử lý được tạo cấu hình để phân tích sự biểu diễn miền thời gian như được minh họa trong khối 1032 của Fig.11a để phát hiện chênh lệch thời gian liên kênh. Phân tích này có thể được thực hiện theo bất kỳ cách nào đã biết và mang đến độ chắc chắn được cải thiện, do phân tích này được thực hiện dựa trên phổ tương quan chéo được làm nhấn tương ứng với đặc tính phổ.

Như được minh họa trên Fig.11b, sự thực hiện được ưu tiên của phân tích miền thời gian 1032 là lọc thông thấp của sự biểu diễn miền thời gian như được minh họa tại 458 trên Fig.11b tương ứng với mục 458 của Fig.4e và tiếp tục xử lý sau đó 1033 sử dụng phép toán tìm kiếm đỉnh/lựa chọn đỉnh trong sự biểu diễn miền thời gian được lọc thông thấp.

Như được minh họa trên Fig.11c, sự thực hiện được ưu tiên của phép toán lựa chọn đỉnh hoặc tìm kiếm đỉnh là để thực hiện phép toán này sử dụng ngưỡng biến số. Cụ thể, bộ xử lý được tạo cấu hình để thực hiện phép toán tìm kiếm đỉnh/lựa chọn đỉnh trong sự biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm nhấn bằng cách xác định 1034 ngưỡng biến số từ sự biểu diễn miền thời gian và bằng cách

so sánh đỉnh hoặc một số đỉnh của sự biểu diễn miền thời gian (thu được cùng hoặc không cùng với việc chuẩn hóa phổ) với ngưỡng biến số, trong đó chênh lệch thời gian liên kênh được xác định là độ trễ thời gian được liên kết với đỉnh trong mối quan hệ được xác định trước với ngưỡng chẳng hạn như lớn hơn ngưỡng biến số.

Như được minh họa trên Fig.11d, một phương án ưu tiên được minh họa trong mã giả được đề cập trên các hình vẽ từ Fig.4e đến Fig.4b được mô tả sau bao gồm trong việc phân loại 1034a các giá trị tương ứng với cường độ của chúng. Sau đó, như được minh họa trong mục 1034b trên Fig.11d, ví dụ về giá trị cao nhất bằng 10% hoặc 5% giá trị được xác định.

Sau đó, như được minh họa trong bước 1034c, số chẳng hạn như số 3 được nhân với giá trị thấp nhất của giá trị cao nhất bằng 10% hoặc 5% để thu được ngưỡng biến số.

Như đã đề cập ở trên, tốt hơn là, giá trị cao nhất bằng 10% hoặc 5% được xác định nhưng nó có thể cũng có ích để xác định số thấp nhất của giá trị cao nhất bằng 50% của các giá trị và để sử dụng số nhân cao hơn chẳng hạn như 10. Về bản chất, thậm chí số lượng nhỏ hơn chẳng hạn như giá trị cao nhất bằng 3% của các giá trị sau đó được nhân với số mà ví dụ bằng 2,5 hoặc 2, tức là nhỏ hơn 3. Do đó, các tổ hợp của các số và tỷ lệ phần trăm khác nhau có thể được sử dụng trong phương án được minh họa trên Fig.11d. Khác với các tỷ lệ phần trăm, các số có thể cũng thay đổi, và các số lớn hơn 1,5 được ưu tiên.

Trong phương án khác được minh họa trên Fig.11e, sự biểu diễn miền thời gian được chia thành các khối con như được minh họa bởi khối 1101, và các khối con này được biểu thị bằng 1300 trên Fig.13. Ở đây, khoảng 16 khối con được sử dụng cho khoảng hợp lệ sao cho mỗi khối có khoảng trễ thời gian bằng 20. Tuy nhiên, số lượng khối con có thể lớn hơn giá trị này hoặc nhỏ hơn và tốt hơn là lớn hơn 3 và nhỏ hơn 50.

Trong bước 1102 của Fig.11e, đỉnh trong mỗi khối con được xác định và trong bước 1103, đỉnh trung bình trong tất cả các khối con được xác định. Sau đó, trong bước 1104, giá trị nhân được xác định rằng một mặt phụ thuộc vào tỉ số tín hiệu trên nhiễu và trong phương án khác, phụ thuộc vào chênh lệch giữa ngưỡng và đỉnh cực đại

như được biểu thị ở bên trái của khối 1104. Phụ thuộc và các giá trị đầu vào này, một trong ba giá trị nhân khác nhau ưu tiên hơn được xác định trong đó giá trị nhân bằng a_{low} ($a_{thấp}$), a_{high} (a_{cao}) và a_{lowest} ($a_{thấp nhất}$).

Sau đó, trong bước 1105, giá trị nhân a được xác định trong khối 1104 được nhân bởi ngưỡng trung bình để thu được ngưỡng biến số mà sau đó được sử dụng trong phép toán so sánh trong khối 1106. Đối với phép toán so sánh, một lần nữa sự biểu diễn miền thời gian được đưa vào khối 1101 có thể được sử dụng hoặc các đỉnh đã được xác định trong mỗi khối con như là được trình bày sơ lược trong khối 1102 có thể được sử dụng.

Sau đó, các phương án khác về việc đánh giá và phát hiện đỉnh trong hàm tương quan chéo miền thời gian được trình bày sơ lược.

Việc đánh giá và phát hiện đỉnh trong hàm tương quan chéo miền thời gian từ phương pháp tương quan chéo được tổng quát hóa (generalized cross-correlation - GCC-PHAT) để ước lượng chênh lệch thời gian liên kênh (Inter-channel Time Difference - ITD) không phải lúc nào cũng đơn giản do các kịch bản đầu vào khác nhau. Đầu vào tiếng nói sạch có thể dẫn đến hàm tương quan chéo lệch thấp với đỉnh mạnh, trong khi tiếng nói trong môi trường âm vang nhiều âm có thể sinh ra vectơ với độ lệch cao và các đỉnh với giá trị thấp nhưng vẫn có cường độ nổi bật biểu thị sự tồn tại của ITD. Thuật toán phát hiện đỉnh mà thích hợp và linh hoạt để chứa các kịch bản đầu vào khác nhau được mô tả.

Do sự hạn chế về độ trễ, toàn bộ hệ thống có thể xử lý căn chỉnh thời gian kênh tới giới hạn nhất định, cụ thể là ITD_MAX . Thuật toán được đề xuất được thiết kế để phát hiện liệu ITD hợp lệ có tồn tại trong các trường hợp:

- ITD hợp lệ do đỉnh nổi bật. Đỉnh nổi bật nằm trong các giới hạn $[-ITD_MAX, ITD_MAX]$ của hàm tương quan chéo được thể hiện.
- Không tương quan. Khi không có tương quan giữa hai kênh, không có đỉnh nổi bật. Ngưỡng phải được xác định, mà trên đó đỉnh đủ mạnh để được coi là giá trị ITD hợp lệ. Mặt khác, không có xử lý ITD cần được truyền tín hiệu, tức là ITD được thiết lập bằng không và không có căn chỉnh thời gian được thực hiện.
- Nằm ngoài các giới hạn ITD. Các đỉnh mạnh của hàm tương quan chéo nằm

ngoài vùng $[-ITD_MAX, ITD_MAX]$ cần được đánh giá để xác định liệu các ITD nằm ngoài khả năng xử lý của hệ thống tồn tại hay không. Trong trường hợp này không có xử lý ITD nào cần được truyền tín hiệu và do đó không có căn chỉnh thời gian nào được thực hiện.

Để xác định liệu cường độ đỉnh có đủ cao để được coi như là giá trị chênh lệch thời gian, ngưỡng thích hợp cần được định rõ. Đối với các kích bản đầu vào khác nhau, đầu ra hàm tương quan chéo biến thiên phụ thuộc và các tham số khác nhau, ví dụ môi trường (nhiều âm, âm vang, v.v.), thiết lập ống phóng thanh (AB, M/S, v.v.). Do đó, để xác định một cách thích ứng ngưỡng là cần thiết.

Trong thuật toán được đề xuất, ngưỡng được xác định bằng cách trước hết tính toán giá trị trung bình của tính toán thô đường bao của cường độ của hàm tương quan chéo trong vùng $[-ITD_MAX, ITD_MAX]$ (Fig.13), sau đó trung bình này được gán trọng số phụ thuộc vào ước lượng SNR.

Mô tả từng bước thuật toán được mô tả dưới đây.

Đầu ra của DFT nghịch đảo của GCC-PHAT, mà biểu diễn tương quan chéo miền thời gian, được sắp xếp lại theo các độ trễ thời gian từ âm đến dương (Fig.12).

Vector tương quan chéo được chia thành ba vùng chính: vùng quan tâm cụ thể là $[-ITD_MAX, ITD_MAX]$ và vùng ngoài giới hạn ITD_MAX , cụ thể là các độ trễ thời gian nhỏ hơn so với $-ITD_MAX$ (max_low) và cao hơn so với ITD_MAX (max_high). Các đỉnh tối đa của các vùng "ngoài giới hạn" được phát hiện và được lưu lại để so sánh với đỉnh cực đại được phát hiện trong vùng quan tâm.

Để xác định liệu ITD hợp lệ có mặt hay không, vùng vector con $[-ITD_MAX, ITD_MAX]$ của hàm tương quan chéo được xem xét. Vector con được chia thành N khối con (Fig.13).

Đối với mỗi khối con, cường độ đỉnh cực đại $peak_sub$ và vị trí độ trễ thời gian tương đương $index_sub$ được tìm kiếm và được lưu.

Cực đại của đỉnh cực đại cục bộ $peak_max$ được xác định và sẽ được so sánh với ngưỡng để xác định sự tồn tại của giá trị ITD hợp lệ.

Đỉnh cực đại $peak_max$ được so sánh với max_low và max_high . Nếu đỉnh $peak_max$ thấp hơn cả hai đỉnh, không có xử lý ITD được truyền tín hiệu và không có

căn chỉnh thời gian được thực hiện. Do hạn chế xử lý ITD của hệ thống, các cường độ của các đỉnh nằm ngoài giới hạn không cần được đánh giá.

Trung bình cường độ của các đỉnh được tính toán:

$$peak_{mean} = \frac{\sum_N peak_{sub}}{N}$$

Ngưỡng $thres$ sau đó được tính toán bằng cách gán trọng số $peak_{mean}$ với hệ số gán trọng số phụ thuộc SNR a_w :

$$thres = a_w peak_{mean}, \text{ where } a_w = \begin{cases} a_{low}, & SNR \leq SNR_{threshold} \\ a_{high}, & SNR > SNR_{threshold} \end{cases}$$

Trong trường hợp trong đó $SNR \ll SNR_{threshold}$ và $|thres - peak_{max}| < \varepsilon$, cường độ đỉnh cũng được so sánh với ngưỡng giảm hơn một chút ($a_w = a_{lowest}$), để tránh việc không chấp nhận đỉnh nổi bật với các đỉnh liền kề cao. Các hệ số gán trọng số ví dụ có thể là $a_{high} = 3$, $a_{low} = 2.5$ và $a_{lowest} = 2$, trong khi $SNR_{threshold}$ ví dụ có thể là 20dB và giới hạn $\varepsilon = 0,05$.

Các khoảng ưu tiên là từ 2,5 đến 5 đối với a_{high} ; 1,5 đến 4 đối với a_{low} ; 1,0 đến 3 đối với a_{lowest} ; 10 đến 30dB đối với $SNR_{threshold}$; và 0,01 đến 0,5 đối với ε , trong đó a_{high} lớn hơn a_{low} mà lớn hơn a_{lowest} .

Nếu $peak_{max} > thres$ độ trễ thời gian tương đương được trả về là ITD được ước lượng, nếu không thì không có xử lý itd nào được truyền tín hiệu (ITD=0).

Các phương án khác được mô tả với sự tham chiếu đến Fig.4e.

Tiếp theo, sự thực hiện được ưu tiên của sáng chế trong khối 1050 của Fig.10b cho mục đích của bộ xử lý tín hiệu khác được thảo luận với sự tham chiếu đến các hình vẽ từ Fig.1 đến Fig.9e, tức là, trong ngữ cảnh của xử lý/mã hóa âm lập thể/đa kênh và căn chỉnh thời gian hai kênh.

Tuy nhiên, như được đề cập và được minh họa trên Fig.10b, nhiều lĩnh vực khác tồn tại, trong đó xử lý tín hiệu khác sử dụng chênh lệch thời gian liên kênh được xác định có thể cũng được thực hiện.

Fig.1 minh họa thiết bị mã hóa tín hiệu đa kênh có ít nhất hai kênh. Tín hiệu đa

kênh 10 một mặt được đưa vào bộ xác định tham số 100 và mặt khác là bộ căn chỉnh tín hiệu 200. Bộ xác định tham số 100 một mặt xác định tham số căn chỉnh băng rộng và một mặt là nhiều tham số căn chỉnh băng hẹp từ tín hiệu đa kênh. Các tham số được xuất ra qua đường tham số 12. Hơn nữa, các tham số cũng có thể được xuất ra qua đường tham số khác 14 đến giao diện đầu ra 500 như được minh họa. Trên đường tham số 14, các tham số bổ sung chẳng hạn như các tham số mức được chuyển tiếp từ bộ xác định tham số 100 đến giao diện đầu ra 500. Bộ căn chỉnh tín hiệu 200 được tạo cấu hình để căn chỉnh ít nhất hai kênh của tín hiệu đa kênh 10 sử dụng tham số căn chỉnh băng rộng và nhiều tham số căn chỉnh băng hẹp được nhận qua đường tham số 10 để thu được các kênh được căn chỉnh 20 tại đầu ra của bộ căn chỉnh tín hiệu 200. Các kênh được căn chỉnh 20 được chuyển tiếp đến bộ xử lý tín hiệu 300 được tạo cấu hình để tính toán tín hiệu giữa 31 và tín hiệu bên 32 từ các kênh được căn chỉnh được nhận qua đường 20. Thiết bị mã hóa còn bao gồm bộ mã hóa tín hiệu 400 để mã hóa tín hiệu giữa từ đường 31 và tín hiệu bên từ đường 32 để thu được tín hiệu giữa được mã hóa trên đường 41 và tín hiệu bên được mã hóa trên đường 42. Cả hai tín hiệu này được chuyển tiếp đến giao diện đầu ra 500 để tạo tín hiệu đa kênh được mã hóa tại đường đầu ra 50. Tín hiệu được mã hóa tại đường đầu ra 50 bao gồm tín hiệu giữa được mã hóa từ đường 41, tín hiệu bên được mã hóa từ đường 42, các tham số căn chỉnh băng hẹp và các tham số căn chỉnh băng rộng từ đường 14, và tùy chọn, tham số mức từ đường 14 và tùy chọn bổ sung, tham số điền đầy âm lập thể được tạo ra bởi bộ mã hóa tín hiệu 400 và được chuyển tiếp tới giao diện đầu ra 500 qua đường tham số 43.

Tốt hơn là, bộ căn chỉnh tín hiệu được tạo cấu hình để căn chỉnh các kênh từ tín hiệu đa kênh sử dụng tham số căn chỉnh băng rộng, trước khi bộ xác định tham số 100 thực sự tính toán các tham số băng hẹp. Do đó, trong phương án này, bộ căn chỉnh tín hiệu 200 gửi các kênh được căn chỉnh băng rộng trở lại bộ xác định tham số 100 qua đường kết nối 15. Sau đó, bộ xác định tham số 100 xác định nhiều tham số căn chỉnh băng hẹp từ tham số đã có đối với đến tín hiệu đa kênh được căn chỉnh có đặc tính băng rộng. Tuy nhiên, trong các phương án khác, các tham số được xác định mà không có chuỗi cụ thể gồm các quy trình.

Fig.4a minh họa sự thực hiện được ưu tiên, trong đó chuỗi các bước xảy ra

đường kết nối 15 được thực hiện. Trong bước 16, tham số căn chỉnh băng rộng được xác định sử dụng hai kênh và tham số căn chỉnh băng rộng chẳng hạn như chênh lệch thời gian liên kênh hoặc tham số ITD được thu. Sau đó, trong bước 21, hai kênh được căn chỉnh bởi bộ căn chỉnh tín hiệu 200 của Fig.1 sử dụng tham số căn chỉnh băng rộng. Sau đó, trong bước 17, các tham số băng hẹp được xác định sử dụng các kênh được căn chỉnh trong bộ xác định tham số 100 để xác định nhiều tham số căn chỉnh băng hẹp chẳng hạn như nhiều tham số chênh lệch pha liên kênh đối với các băng khác nhau của tín hiệu đa kênh. Sau đó, trong bước 22, các giá trị phổ trong mỗi băng tham số được căn chỉnh sử dụng tham số căn chỉnh băng hẹp tương ứng cho băng cụ thể này. Khi quy trình trong bước 22 được thực hiện cho mỗi băng, mà tham số căn chỉnh băng hẹp có sẵn, thì các kênh thứ nhất và thứ hai hoặc trái/phải được căn chỉnh sẽ có sẵn cho việc xử lý tín hiệu sau đó bởi bộ xử lý tín hiệu 300 của Fig.1.

Fig.4b minh họa sự thực hiện khác của bộ mã hóa đa kênh của Fig.1 trong đó một vài quy trình được thực hiện trong miền tần số.

Cụ thể, bộ mã hóa đa kênh còn bao gồm bộ chuyển đổi thời gian-tần số 150 để chuyển đổi tín hiệu đa kênh miền thời gian thành sự biểu diễn phổ của ít nhất hai kênh trong miền tần số.

Hơn nữa, như được minh họa tại 152, bộ xác định tham số, bộ căn chỉnh tín hiệu và bộ xử lý tín hiệu được minh họa tại 100, 200 và 300 trên Fig.1 đều hoạt động trong miền tần số.

Hơn nữa, bộ mã hóa đa kênh và cụ thể là bộ xử lý tín hiệu còn bao gồm bộ chuyển đổi phổ-thời gian để tạo ra ít nhất sự biểu diễn miền thời gian của tín hiệu giữa.

Tốt hơn là, bộ chuyển đổi thời phổ ngoài ra còn chuyển đổi sự biểu diễn phổ của tín hiệu bên cũng được xác định bởi các quy trình được biểu diễn bởi khối 152 thành sự biểu diễn miền thời gian, và bộ mã hóa tín hiệu 400 của Fig.1 sau đó được tạo cấu hình để tiếp tục mã hóa tín hiệu giữa và/hoặc tín hiệu bên như là các tín hiệu miền thời gian phụ thuộc vào sự thực hiện cụ thể của bộ mã hóa tín hiệu 400 của Fig.1.

Tốt hơn là, bộ chuyển đổi thời gian-phổ 150 của Fig.4b được tạo cấu hình để thực hiện các bước 155, 156 và 157 của Fig.4c. Cụ thể, bước 155 bao gồm cung cấp

cửa sổ phân tích với ít nhất một phần đệm số không tại một đầu của nó và cụ thể, phần đệm số không tại phần cửa sổ khởi tạo và phần đệm số không tại phần cửa sổ kết thúc như được minh họa, ví dụ, trên Fig.7 sau đó. Hơn nữa, cửa sổ phân tích còn có các khoảng chồng lấp hoặc các phần chồng lấp tại nửa thứ nhất của cửa sổ và tại nửa thứ hai của cửa sổ và, ngoài ra, tốt hơn là, phần giữa có thể là khoảng không chồng lấp tùy trường hợp.

Trong bước 156, mỗi kênh được tạo cửa sổ sử dụng cửa sổ phân tích với các khoảng chồng lấp. Cụ thể, mỗi kênh được tạo cửa sổ sử dụng cửa sổ phân tích theo cách mà khối thứ nhất của kênh được thu được. Sau đó, khối thứ hai của kênh tương tự được thu mà có khoảng chồng lấp nhất định với khối thứ nhất, v.v., sao cho ví dụ, sau năm phép toán tạo cửa sổ, năm khối của các mẫu được tạo cửa sổ của mỗi kênh là có sẵn mà sau đó được biến đổi một cách riêng biệt thành sự biểu diễn phổ như được minh họa tại 157 trên Fig.4c. Quy trình tương tự cũng được thực hiện với kênh còn lại sao cho, tại cuối bước 157, chuỗi gồm các khối của các giá trị phổ, và cụ thể, các giá trị phổ phức hợp chẳng hạn như các giá trị phổ DFT hoặc các mẫu băng con phức hợp là có sẵn.

Trong bước 158 mà được thực hiện bởi bộ xác định tham số 100 của Fig.1, tham số căn chỉnh băng rộng được xác định và trong bước 159, mà được thực hiện bởi căn chỉnh tín hiệu 200 của Fig.1, dịch chuyển vòng tròn được thực hiện sử dụng tham số căn chỉnh băng rộng. Trong bước 160, một lần nữa được thực hiện bởi bộ xác định tham số 100 của Fig.1, các tham số căn chỉnh băng hẹp được xác định đối với các băng/bản con riêng biệt và trong bước 161, các giá trị phổ được căn chỉnh được luân phiên cho mỗi băng sử dụng các tham số căn chỉnh băng hẹp tương ứng được xác định cho các băng cụ thể.

Fig.4d minh họa các quy trình khác được thực hiện bởi bộ xử lý tín hiệu 300. Cụ thể, bộ xử lý tín hiệu 300 được tạo cấu hình để tính toán tín hiệu bên và tín hiệu giữa như được minh họa tại bước 301. Trong bước 302, một số dạng xử lý tiếp theo của tín hiệu bên có thể được thực hiện và sau đó, trong bước 303, mỗi khối của tín hiệu giữa và tín hiệu bên được biến đổi trở lại vào miền thời gian, và trong bước 304, cửa sổ tổng hợp được áp dụng cho mỗi khối được thu được bởi bước 303 và, trong

bước 305, một mặt phép toán cộng chồng lấp cho tín hiệu giữa và mặt khác phép toán cộng chồng lấp cho tín hiệu bên được thực hiện để cuối cùng thu được các tín hiệu giữa/bên miền thời gian.

Cụ thể, các phép toán của các bước 304 và 305 dẫn đến dạng nhiễu xuyên kênh từ một khối của tín hiệu giữa hoặc tín hiệu bên trong khối kế tiếp của tín hiệu giữa và tín hiệu bên được thực hiện sao cho, thậm chí khi bất kỳ sự thay đổi tham số nào xảy ra chẳng hạn như tham số chênh lệch thời gian liên kênh hoặc tham số chênh lệch pha liên kênh, tuy nhiên, điều này sẽ dẫn đến việc không thể nghe trong các tín hiệu giữa/bên miền thời gian được thu bởi bước 305 trên Fig.4d.

Mã hóa âm lập thể độ trễ thấp mới là mã hóa âm lập thể giữa/bên (M/S) kết hợp khai thác một số tín hiệu điều khiển không gian, trong đó kênh giữa được mã hóa bởi bộ mã hóa lõi đơn sắc sơ cấp, và kênh bên được mã hóa trong bộ mã hóa lõi thứ cấp. Các nguyên lý của bộ mã hóa và bộ giải mã được mô tả trên Fig.6a, Fig.6b.

Xử lý âm lập thể được thực hiện chủ yếu trong miền tần số (Frequency Domain - FD). Một cách tùy ý, một số xử lý âm lập thể có thể được thực hiện trong miền thời gian (Time Domain - TD) trước khi phân tích tần số. Trong trường hợp đối với việc tính toán ITD, mà có thể được tính toán và áp dụng trước khi phân tích tần số để căn chỉnh các kênh trong thời gian trước khi tiếp tục phân tích và xử lý âm lập thể. Cách khác, xử lý ITD có thể được thực hiện một cách trực tiếp trong miền tần số. Do các bộ mã hóa tín hiệu tiếng nói thông thường như ACELP không chứa bất kỳ khai triển thời gian - tần số nội bộ nào, mã hóa âm lập thể thêm giàn lọc điều biến phức hợp bổ sung bằng giàn lọc phân tích và tổng hợp trước bộ mã hóa lõi và tầng khác của giàn lọc phân tích-tổng hợp sau bộ giải mã lõi. Trong phương án được ưu tiên, DFT được lấy mẫu quá mức với vùng chồng lấp được sử dụng. Tuy nhiên, trong các phương án khác, việc khai triển thời gian-tần số được tạo giá trị phức hợp với độ phân giải theo thời gian tương tự có thể được sử dụng.

Xử lý âm lập thể bao gồm tính toán tín hiệu điều khiển theo không gian: các chênh lệch thời gian liên kênh (inter-channel Time Difference - ITD) và các chênh lệch mức liên kênh (inter-channel Level Difference - ILD). Các ITD và IPD được sử dụng trên tín hiệu âm lập thể đầu vào để căn chỉnh hai kênh L và R theo thời gian và

theo pha. ITD được tính toán trong băng rộng hoặc trong miền thời gian trong khi các IPD và ILD được tính toán cho từng hoặc một phần của các băng tham số, tương ứng với việc khai triển không đồng bộ của không gian tần số. Khi hai kênh được căn chỉnh, âm lập thể M/S kết hợp được áp dụng, trong đó tín hiệu bên sau đó còn được dự báo từ tín hiệu giữa. Độ khuếch đại dự báo được suy ra từ các ILD.

Tín hiệu giữa còn được mã hóa bởi bộ mã hóa lõi sơ cấp. Trong phương án được ưu tiên, bộ mã hóa lõi sơ cấp là tiêu chuẩn #GPP EVS, hoặc sự mã hóa được suy ra từ nó mà có thể chuyển giữa chế độ mã hóa tiếng nói, ACELP, và chế độ âm nhạc dựa trên biến đổi MDCT. Tốt hơn là, ACELP và bộ mã hóa dựa trên MDCT được hỗ trợ bởi mở rộng băng thông miền thời gian (Time Domain BandWidth Extension - TD-BWE) và các môđun điền đầy khoảng trống thông minh (Intelligent Gap Filling - IGF) tương ứng.

Tín hiệu bên trước hết được dự báo bởi tín hiệu giữa sử dụng độ khuếch đại dự báo được suy ra từ các ILD. Phần dư có thể còn được dự báo bởi phiên bản trễ của tín hiệu giữa hoặc được mã hóa một cách trực tiếp bởi bộ mã hóa lõi thứ cấp, được thực hiện trong phương án được ưu tiên trong miền MDCT. Xử lý âm lập thể tại bộ mã hóa có thể được tóm tắt trên Fig.5 như sẽ được giải thích sau.

Fig.2 minh họa sơ đồ khối của phương án của thiết bị giải mã tín hiệu đa kênh được mã hóa được nhận tại đường đầu vào 50.

Cụ thể, tín hiệu được nhận bởi giao diện đầu vào 600. Được kết nối với giao diện đầu vào 600 là bộ giải mã tín hiệu 700 và bộ hủy căn chỉnh tín hiệu 900. Ngoài ra, bộ xử lý tín hiệu 800 một mặt được kết nối với bộ giải mã tín hiệu và mặt khác được kết nối với bộ hủy căn chỉnh tín hiệu.

Cụ thể, tín hiệu đa kênh được mã hóa bao gồm tín hiệu giữa được mã hóa, tín hiệu bên được mã hóa, thông tin trên tham số căn chỉnh băng rộng và thông tin trên nhiều tham số băng hẹp. Do đó, tín hiệu đa kênh được mã hóa trên đường 50 có thể giống chính xác với tín hiệu như được xuất ra bởi giao diện đầu ra 500 của Fig.1.

Tuy nhiên, quan trọng là, cần lưu ý rằng, trái ngược với những gì được minh họa trên Fig.1, tham số căn chỉnh băng rộng và nhiều tham số căn chỉnh băng hẹp

được bao gồm trong tín hiệu được mã hóa ở dạng nhất định có thể là chính xác là các tham số căn chỉnh như được sử dụng bởi bộ căn chỉnh tín hiệu 200 trên Fig.1, cách khác, có thể cũng là các giá trị nghịch đảo của nó, tức là, các tham số mà có thể được sử dụng bởi chính xác các phép toán tương tự được thực hiện bởi bộ căn chỉnh tín hiệu 200 nhưng với các giá trị nghịch đảo sao cho việc hủy căn chỉnh được thu được.

Do đó, thông tin trên các tham số căn chỉnh có thể là các tham số căn chỉnh được sử dụng bởi bộ căn chỉnh tín hiệu 200 trên Fig.1 hoặc có thể là các giá trị nghịch đảo, tức là "các tham số hủy căn chỉnh" thực tế. Ngoài ra, các tham số này thường được lượng tử hóa dưới dạng nhất định như sẽ được thảo luận sau ở Fig.8.

Giao diện đầu vào 600 của Fig.2 phân tách thông tin trên tham số căn chỉnh băng rộng và nhiều tham số căn chỉnh băng hẹp từ các tín hiệu giữa/bên được mã hóa và chuyển tiếp thông tin này qua đường tham số 610 đến bộ hủy căn chỉnh tín hiệu 900. Mặt khác, tín hiệu giữa được mã hóa được chuyển tiếp tới bộ giải mã tín hiệu 700 qua đường 601 và tín hiệu bên được mã hóa được chuyển tiếp đến bộ giải mã tín hiệu 700 qua đường tín hiệu 602.

Bộ giải mã tín hiệu được tạo cấu hình để giải mã tín hiệu giữa được mã hóa và giải mã tín hiệu bên được mã hóa để thu được tín hiệu giữa được giải mã trên đường 701 và tín hiệu bên được giải mã trên đường 702. Các tín hiệu này được sử dụng bởi bộ xử lý tín hiệu 800 để tính toán tín hiệu kênh thứ nhất được giải mã hoặc tín hiệu bên trái được giải mã và để tính toán kênh thứ hai được giải mã hoặc tín hiệu kênh bên phải được giải mã từ tín hiệu giữa được giải mã và tín hiệu bên được giải mã, và kênh thứ nhất được giải mã và kênh thứ hai được giải mã được xuất ra lần lượt trên các đường 801, 802. Bộ hủy căn chỉnh tín hiệu 900 được tạo cấu hình để hủy căn chỉnh kênh thứ nhất được giải mã trên đường 801 và kênh bên phải được giải mã 802 sử dụng thông tin trên tham số căn chỉnh băng rộng và ngoài ra sử dụng thông tin trên nhiều tham số căn chỉnh băng hẹp để thu tín hiệu đa kênh được giải mã, tức là tín hiệu được giải mã có ít nhất hai kênh được giải mã và được hủy căn chỉnh trên các đường 901 và 902.

Fig.9a minh họa chuỗi được ưu tiên gồm các bước được thực hiện bởi bộ hủy căn chỉnh tín hiệu 900 từ Fig.2. Cụ thể, bước 910 nhận các kênh bên trái và bên phải

được căn chỉnh như có sẵn trên các đường 801, 802 từ Fig.2. Trong bước 901, bộ hủy căn chỉnh tín hiệu 900 hủy căn chỉnh các băng con riêng lẻ sử dụng thông tin trên các tham số căn chỉnh băng hẹp để thu được kênh thứ nhất và kênh thứ hai hoặc kênh bên trái hoặc kênh bên phải được giải mã được hủy căn chỉnh theo pha tại 911a và 911b. Trong bước 912, các kênh được hủy căn chỉnh sử dụng tham số căn chỉnh băng rộng sao cho, tại 913a và 913b, các kênh được hủy căn chỉnh theo pha và theo thời gian được thu.

Trong bước 914, bất kỳ sự xử lý nào khác được thực hiện mà bao gồm việc sử dụng phép toán tạo cửa sổ hoặc chồng lắp-cộng nào, nói chung, bất kỳ phép toán nhiều xuyên kênh nào để thu được, tại 915a hoặc 915b, tín hiệu được giải mã có thành phần lạ được giảm thiểu hoặc không có thành phần lạ, tức là đối với các kênh được giải mã mà không có bất kỳ thành phần lạ nào mặt dù, thông thường, có các tham số hủy căn chỉnh biến đổi theo thời gian một mặt đối với băng rộng và mặt khác đối với nhiều băng hẹp.

Fig.9 minh họa sự thực hiện của bộ giải mã đa kênh được minh họa trên Fig.2.

Cụ thể, bộ xử lý tín hiệu 800 từ Fig.2 bao gồm bộ chuyển đổi thời gian-phổ 810.

Bộ xử lý tín hiệu còn bao gồm bộ chuyển đổi giữa/bên thành trái/phải 802 để tính toán tín hiệu giữa M và tín hiệu bên S, tín hiệu bên trái L và tín hiệu bên phải R.

Tuy nhiên, quan trọng là để tính toán L và R bằng cách chuyển đổi giữa/bên-trái/phải trong khối 802, tín hiệu bên S không cần thiết được sử dụng. Thay vào đó, như được thảo luận sau này, các tín hiệu trái/phải được tính toán ban đầu chỉ sử dụng tham số độ khuếch đại được suy ra từ tham số chênh lệch mức liên kênh ILD. Thông thường, độ khuếch đại dự báo có thể cũng được xem như là một dạng của ILD. Độ khuếch đại có thể được suy ra từ ILD nhưng cũng có thể được tính toán một cách trực tiếp. Ưu tiên không tính toán ILD nữa mà tính toán độ khuếch đại dự báo một cách trực tiếp và để truyền và sử dụng độ khuếch đại dự báo trong bộ giải mã hơn là tham số ILD.

Do đó, trong sự thực hiện này, tín hiệu bên S chỉ được sử dụng trong bộ cập

nhật kênh 830 mà hoạt động để cung cấp tín hiệu trái/phải tốt hơn bằng cách sử dụng tín hiệu bên S được truyền dẫn như được minh họa bởi đường vòng 821.

Do đó, bộ chuyển đổi 820 hoạt động sử dụng tham số mức được thu qua đầu vào tham số 822 và không thực sự sử dụng tín hiệu bên S nhưng bộ cập nhật kênh 830 sau đó hoạt động sử dụng bên 821, và phụ thuộc vào sự thực hiện cụ thể, việc sử dụng tham số điền đầy âm lập thể được nhận qua đường 831. Bộ căn chỉnh tín hiệu 900 sau đó bao gồm bộ hủy căn chỉnh theo pha và bộ định tỉ lệ năng lượng 910. Việc định tỉ lệ năng lượng được điều khiển bởi hệ số định tỉ lệ được suy ra bởi bộ tính toán hệ số định tỉ lệ 940. Bộ tính toán hệ số định tỉ lệ 940 được nạp vào đầu ra của bộ cập nhật kênh 830. Dựa vào các tham số căn chỉnh băng hẹp được nhận qua đầu vào 911, việc hủy căn chỉnh theo pha được thực hiện, và trong khối 920, dựa trên tham số căn chỉnh băng rộng được nhận qua đường 921, việc hủy căn chỉnh theo thời gian được thực hiện. Cuối cùng, việc chuyển đổi phổ-thời gian 930 được thực hiện để cuối cùng thu được tín hiệu được giải mã.

Fig.9c minh họa chuỗi khác gồm các bước thường được thực hiện trong các khối 920 và 930 của Fig.9b trong phương án được ưu tiên.

Cụ thể, các kênh được hủy căn chỉnh băng hẹp được đưa vào chức năng hủy căn chỉnh băng rộng tương ứng với khối 920 của Fig.9b. DFT hoặc bất kỳ biến đổi nào được thực hiện trong khối 931. Sau việc tính toán thực sự của các mẫu miền thời gian, việc tạo cửa sổ tổng hợp tùy chọn sử dụng cửa sổ tổng hợp được thực hiện. Cửa sổ tổng hợp tốt hơn là giống hoàn toàn cửa sổ phân tích hoặc được suy ra từ cửa sổ phân tích, ví dụ, phép nội suy hoặc lấy một phần mười nhưng phụ thuộc vào cách nhất định từ cửa sổ phân tích. Sự phụ thuộc này tốt hơn là sao cho các hệ số nhân được xác định bởi hai cửa sổ chồng lấp cộng vào một cho mỗi điểm trong khoảng chồng lấp. Do đó, sau cửa sổ tổng hợp trong khối 932, phép toán chồng lấp và phép toán cộng sau đó được thực hiện. Cách khác, thay cho việc tạo cửa sổ tổng hợp và phép toán chồng lấp/cộng, bất kỳ nhiều xuyên âm nào giữa các khối tiếp sau đối với mỗi kênh được thực hiện để thu được, như đã thảo luận trong ngữ cảnh của Fig.9a, tín hiệu được mã hóa được giảm thành thành phần lạ.

Khi Fig.6b được xem xét, rõ ràng rằng các hoạt động giải mã thực sự đối với tín

hiệu-giữa, tức là, một mặt “bộ giải mã EVS”, đối với tín hiệu bên, việc lượng tử hóa vector nghịch đảo VQ^{-1} và hoạt động MDCT nghịch đảo (IMDCT) tương ứng với bộ giải mã tín hiệu 700 của Fig.2.

Hơn nữa, các phép toán DFT 810 trong các khối 810 tương ứng với phần tử 810 trên Fig.9b và các chức năng của sự xử lý âm lập thể nghịch đảo và dịch chuyển thời gian nghịch đảo tương ứng với các khối 800, 900 của Fig.2 và các phép toán DFT nghịch đảo 930 trên Fig.6b tương ứng với phép toán tương ứng trong khối 930 trên Fig.9b.

Sau đó, Fig.3 được thảo luận chi tiết hơn. Cụ thể, Fig.3 minh họa phổ DFT có các vạch phổ riêng biệt. Ưu tiên hơn, phổ DFT hoặc bất kỳ phổ nào khác được minh họa trên Fig.3 là phổ phức hợp và mỗi vạch là vạch phổ phức hợp có cường độ và pha hoặc có phần thực và phần ảo.

Ngoài ra, phổ cũng được chia thành các băng tham số khác nhau. Mỗi băng tham số có ít nhất một và tốt hơn là nhiều hơn một vạch phổ. Ngoài ra, các băng tham số tăng từ tần số thấp hơn đến tần số cao hơn. Thông thường, tham số căn chỉnh băng rộng là tham số căn chỉnh băng rộng đơn cho toàn bộ phổ, tức là, đối với phổ bao gồm tất cả các băng từ 1 đến 6 trong phương án ví dụ trên Fig.3.

Hơn nữa, nhiều tham số căn chỉnh băng hẹp được cung cấp sao cho có tham số căn chỉnh đơn cho mỗi băng tham số. Điều này có nghĩa là tham số căn chỉnh cho băng luôn áp dụng cho tất cả các giá trị phổ trong băng tương ứng.

Hơn nữa, ngoài các tham số căn chỉnh băng hẹp, các tham số mức cũng được cung cấp cho mỗi băng tham số.

Trái ngược với các tham số mức mà được cung cấp cho mỗi băng tham số và từng băng tham số từ băng 1 đến băng 6, ưu tiên cung cấp nhiều tham số căn chỉnh băng hẹp chỉ cho số lượng giới hạn các băng thấp hơn chẳng hạn như các băng 1,2,3 và 4.

Ngoài ra, các tham số điền đầy âm lập thể được cung cấp cho số lượng nhất định của các băng ngoại trừ các băng thấp hơn, ví dụ, trong phương án ví dụ, đối với các băng 4, 5 và 6 trong khi các giá trị phổ tín hiệu bên đối với các băng tham số thấp

hơn 1, 2 và 3 và do đó, không có tham số điền đầy âm lập thể nào tồn tại cho các băng thấp hơn này mà trong đó việc khớp dạng sóng được thu sử dụng bản thân tín hiệu bên hoặc tín hiệu dự dự báo biểu diễn tín hiệu bên.

Như được đề cập, tồn tại nhiều vạch phổ trong các băng cao hơn chẳng hạn như, trong phương án trên Fig.3, bảy vạch phổ trong băng tham số 6 so với chỉ ba vạch phổ trong băng tham số 2. Tuy nhiên, về bản chất, số lượng các băng tham số, số lượng các vạch phổ và số lượng các vạch phổ trong băng tham số và các giới hạn khác nhau đối với các tham số nhất định sẽ khác nhau.

Tuy nhiên, Fig.8 minh họa sự phân bố của tham số và số lượng các băng mà các tham số được cung cấp trong phương án nhất định mà trong đó, trái ngược với Fig.3, thực sự có 12 băng.

Như được minh họa, tham số mức ILD được cung cấp cho mỗi băng trong số 12 băng và được lượng tử hóa đến độ chính xác lượng tử hóa được biểu diễn bằng năm bit mỗi băng.

Hơn nữa, các tham số căn chỉnh băng hẹp IPD chỉ được cung cấp cho các băng thấp hơn đến tần số rộng hơn bằng 2,5 kHz. Ngoài ra, tham số chênh lệch thời gian liên kênh hoặc căn chỉnh băng rộng chỉ được cung cấp như tham số đơn lẻ cho toàn bộ phổ nhưng với độ chính xác lượng tử hóa rất cao bằng tám bit cho toàn bộ băng.

Hơn nữa, các tham số điền đầy âm lập thể được lượng tử hóa tương đối thô được cung cấp được biểu diễn bởi ba bit mỗi băng và không dành cho các băng thấp hơn 1kHz do, đối với các băng thấp hơn, thực tế tín hiệu bên được mã hóa hoặc các giá trị phổ dư của tín hiệu bên được bao gồm.

Tiếp theo, sự xử lý được ưu tiên trên phía bộ mã hóa được tóm tắt với sự viện dẫn đến Fig.5. Trong bước thứ nhất, phân tích DFT của kênh bên trái và bên phải được thực hiện. Quy trình này tương ứng với các bước 155 đến 157 của Fig.4c. Trong bước 158, tham số căn chỉnh băng rộng được tính toán và cụ thể, chênh lệch thời gian liên kênh (inter-channel time difference - ITD) của tham số căn chỉnh băng rộng được ưu tiên. Như được minh họa tại 170, sự dịch chuyển thời gian của L và R trong miền tần số được thực hiện. Cách khác, sự dịch chuyển thời gian này có thể cũng được thực

hiện trong miền thời gian. Sau đó DFT nghịch đảo được thực hiện, sự dịch chuyển thời gian được thực hiện trong miền thời gian và việc chuyển tiếp bổ sung DFT được thực hiện để một lần nữa có các sự biểu diễn phổ sau việc căn chỉnh sử dụng tham số căn chỉnh băng rộng.

Các tham số ILD, tức là, các tham số mức và các tham số pha (các tham số IPD), được tính toán cho mỗi băng tham số trên các sự biểu diễn L và R được dịch chuyển như được minh họa tại bước 171. Ví dụ, bước này tương ứng với bước 160 của Fig.4c. Các sự biểu diễn L và R được dịch chuyển thời gian được luân phiên như hàm của các tham số chênh lệch pha liên kênh như được minh họa trong bước 161 của Fig.4c hoặc Fig.5. Sau đó, các tín hiệu giữa và bên được tính toán như được minh họa trên bước 301 và, tốt hơn là, bổ sung với phép toán chuyển đổi năng lượng như được thảo luận sau đây. Trong bước tiếp theo 174, việc dự báo của S với M như hàm của ILD và một cách tùy ý với tín hiệu M trước đó, tức là tín hiệu giữa của khung sớm hơn được thực hiện. Sau đó, DFT nghịch đảo của tín hiệu giữa và tín hiệu bên được thực hiện tương ứng với các bước 303, 304, 305 của Fig.4d trong phương án được ưu tiên.

Trong bước cuối cùng 174, tín hiệu giữa miền thời gian m và một cách tùy ý, tín hiệu dư được mã hóa như được minh họa trong bước 174. Quy trình này tương ứng với những gì được thực hiện bởi bộ mã hóa tín hiệu 400 trên Fig.1.

Tại bộ giải mã trong xử lý âm lập thể nghịch đảo, tín hiệu *bên* (*Side*) được tạo ra trong miền DFT và được dự báo trước hết từ tín hiệu *giữa* (*Mid*) như là:

$$\widehat{Side} = g \cdot Mid$$

trong đó g là độ khuếch đại được tính toán cho mỗi băng tham số và là hàm của các chênh lệch mức liên kênh được truyền dẫn (Inter-channel Level Difference - ILD).

Phần dư của dự báo $Side - g \cdot Mid$ sau đó có thể được tinh chỉnh theo hai cách:

- Bằng mã hóa thứ cấp của tín hiệu dư:

$$\widehat{Side} = g \cdot Mid + g_{cod} \cdot \widehat{(Side - g \cdot Mid)}$$

trong đó g_{cod} là độ khuếch đại toàn cục được truyền dẫn cho toàn bộ phổ

- Bằng cách dự báo phần dư, được biết đến như là điền đầy âm lập thể,

việc dự báo phổ phía phần dư với phổ tín hiệu *Mid* được giải mã trước đó từ khung DFT trước đó:

$$\widehat{Side} = g \cdot Mid + g_{pred} \cdot Mid \cdot z^{-1}$$

trong đó g_{pred} là độ khuếch đại có thể dự báo được truyền dẫn cho mỗi băng tham số.

Hai loại tinh chỉnh mã hóa có thể được trộn trong cùng một phổ DFT. Trong phương án được ưu tiên, mã hóa phần dư được áp dụng trên các băng tham số thấp hơn, trong khi dự báo phần dư được áp dụng trên các băng còn lại. Mã hóa phần dư trong phương án được ưu tiên như được minh họa trên Fig.1 thực hiện trong miền MDCT sau khi tổng hợp tín hiệu bên dư trong miền thời gian và biến đổi nó bởi MDCT. Không giống DFT, MDCT được lấy mẫu tới hạn và thích hợp hơn cho việc mã hóa âm thanh. Các hệ số MDCT là các vector được lượng tử hóa một cách trực tiếp bởi lượng tử hóa vector theo dàn nhưng có thể được mã hóa theo cách khác bởi bộ lượng tử hóa vô hướng theo sau bởi bộ mã hóa entropi. Cách khác, tín hiệu dư có thể cũng được mã hóa trong miền thời gian bởi kỹ thuật mã hóa tiếng nói hoặc một cách trực tiếp trong miền DFT.

1. Phân tích thời gian - tần số: DFT

Điều quan trọng là khai triển thời gian-tần số bổ sung từ xử lý âm lập thể bằng các DFT cho phép phân tích cảnh âm thanh tốt trong khi không tăng đáng kể toàn bộ độ trễ của hệ thống mã hóa. Mặc định, độ phân giải thời gian là 10 mili giây (gấp đôi khung 20 mili giây của bộ mã hóa lõi) được sử dụng. Các cửa sổ phân tích và tổng hợp là giống nhau và đối xứng. Cửa sổ được biểu diễn tại tốc độ lấy mẫu bằng 16kHz trên Fig.7. Có thể thấy rằng vùng chồng lấp được giới hạn để giảm độ trễ được tạo ra và đệm số không cũng được bổ sung vào để cân bằng dịch chuyển vòng tròn khi áp dụng ITD trong miền tần số như được giải thích sau đây.

2. Các tham số âm lập thể

Các tham số âm lập thể có thể được truyền dẫn tại độ phân giải cực đại của âm lập thể DFT. Ở độ phân giải cực tiểu, nó có thể được giảm đến độ phân giải khung của bộ giải mã lõi, tức là 20 mili giây. Mặc định, khi không có sự chuyển tiếp được phát hiện, các tham số được tính toán mỗi 20 mili giây trên 2 cửa sổ DFT. Các băng tham

số cấu thành sự khai triển không đồng bộ và không chồng lấp của phổ sau khoảng 2 lần hoặc 4 lần băng rộng hình chữ nhật tương đương (Equivalent Rectangular Bandwidths - ERB). Mặc định, 4 lần tỉ lệ ERB được sử dụng cho tổng cộng 12 băng đối với băng thông tần số bằng 16 kHz (tốc độ lấy mẫu 32kbps, âm lập thể băng siêu rộng). Fig.8 tóm tắt ví dụ về cấu hình mà trong đó thông tin phụ âm lập thể được truyền dẫn ở tốc độ khoảng 5kbps.

3. Tính toán ITD và căn chỉnh thời gian kênh

ITD được tính toán bằng cách ước lượng độ trễ thời gian đến (Time Delay of Arrival - TDOA) sử dụng tương quan chéo được tổng quát hóa với biến đổi pha (Generalized Cross Correlation with Phase Transform - GCC-PHAT):

$$ITD = \underset{\tau}{\operatorname{argmax}} \left(\operatorname{IDFT} \left(\frac{L_i(f) R_i^*(k)}{|L_i(f) R_i^*(k)|} \right) \right)$$

trong đó L và R là phổ tần số của lần lượt các kênh bên trái và bên phải. Phân tích tần số có thể được thực hiện độc lập với DFT được sử dụng cho việc xử lý âm lập thể tiếp theo hoặc có thể được chia sẻ. Mã giả cho việc tính toán ITD như sau:

```

L =fft(window(l));
R =fft(window(r));
tmp = L .* conj( R );
sfm_L = prod(abs(L).^(1/length(L)))/(mean(abs(L))+eps);
sfm_R = prod(abs(R).^(1/length(R)))/(mean(abs(R))+eps);
sfm = max(sfm_L,sfm_R);
h.cross_corr_smooth = (1-sfm)*h.cross_corr_smooth+sfm*tmp;
tmp = h.cross_corr_smooth ./ abs( h.cross_corr_smooth+eps );
tmp = ifft( tmp );
tmp = tmp([length(tmp)/2+1:length(tmp) 1:length(tmp)/2+1]);
tmp_sort = sort( abs(tmp) );
thresh = 3 * tmp_sort( round(0.95*length(tmp_sort)) );
xcorr_time=abs(tmp(- ( h.stereo_itd_q_max - (length(tmp)-1)/2 - 1 ):- (
h.stereo_itd_q_min - (length(tmp)-1)/2 - 1 )));
%smooth output for better detection
xcorr_time=[xcorr_time 0];
xcorr_time2=filter([0.25 0.5 0.25],1,xcorr_time);
[m,i] = max(xcorr_time2(2:end));
if m > thresh
    itd = h.stereo_itd_q_max - i + 1;
else
    itd = 0;
end

```

Fig.4 minh họa lưu đồ để thực hiện mã giả được minh họa trước đó để thu được việc tính toán chắc chắn và hiệu quả chênh lệch thời gian liên kênh như được ví dụ

trên tham số căn chỉnh băng rộng.

Trong khối 451, phân tích DFT của các tín hiệu miền thời gian cho kênh thứ nhất (l) và kênh thứ hai (r) được thực hiện. Phân tích DFT thường sẽ giống phân tích DFT như được thảo luận, ví dụ, trong ngữ cảnh của các bước 155 đến 157 trên Fig.5 hoặc Fig.4c.

Tương quan chéo sau đó được thực hiện cho mỗi ngăn tần số như được minh họa trong khối 452.

Do đó, phổ tương quan chéo được thu cho toàn bộ phạm vi phổ của các kênh bên trái và bên phải.

Trong bước 453, số đo độ phẳng phổ sau đó được tính toán từ phổ cường độ của L và R và, trong bước 454, số đo độ phẳng lớn hơn được lựa chọn. Tuy nhiên, việc lựa chọn trong bước 454 không cần thiết phải là sự lựa chọn số đo lớn hơn mà việc xác định SFM đơn lẻ này từ cả hai kênh có thể cũng là sự lựa chọn và tính toán chỉ kênh bên trái hoặc chỉ kênh bên phải hoặc có thể là việc tính toán của giá trị trung bình được gán trọng số của cả hai giá trị SFM.

Trong bước 455, phổ tương quan chéo sau đó được làm nhẵn theo thời gian phụ thuộc vào số đo độ phẳng phổ.

Tốt hơn là, số đo độ phẳng phổ được tính toán bằng cách chia giá trị trung bình hình học của phổ cường độ bởi giá trị trung bình số học của phổ cường độ. Do đó, các giá trị đối với SFM được giới hạn giữa không và một.

Trong bước 45, phổ tương quan chéo được làm nhẵn sau đó được chuẩn hóa bởi cường độ của nó và trong bước 457 DFT nghịch đảo của phổ tương quan chéo được chuẩn hóa và được làm nhẵn được tính toán. Trong bước 458, bộ lọc miền thời gian nhất định tốt hơn là được thực hiện nhưng việc lọc miền thời gian này có thể cũng nằm bên trái phụ thuộc vào việc thực hiện nhưng được ưu tiên sẽ được trình bày sau này.

Trong bước 459, ước lượng ITD được thực hiện bằng cách lựa chọn đỉnh của hàm tương quan chéo được tổng quát hóa bởi bộ lọc và bằng cách thực hiện phép toán ngưỡng nhất định.

Nếu không đỉnh nào lớn hơn ngưỡng được thu, thì ITD được thiết lập bằng

không và không có căn chỉnh thời gian nào được thực hiện cho khối tương ứng.

Tính toán ITD có thể cũng được tóm tắt như sau. Tương quan chéo được tính toán trong miền tần số trước khi được làm nhẵn phụ thuộc vào số đo độ phẳng phổ (Spectral Flatness Measurement – SFM). SFM được giới hạn trong khoảng 0 và 1. Trong trường hợp các tín hiệu giống nhau, SFM sẽ cao (tức là khoảng 1) và việc làm nhẵn sẽ yếu. Trong trường hợp tín hiệu giống âm điệu, SFM sẽ thấp và việc làm nhẵn trở nên mạnh hơn. Tương quan chéo được làm nhẵn sau đó được chuẩn hóa bởi biên độ của nó trước khi được biến đổi trở lại miền thời gian. Chuẩn hóa tương ứng với biến đổi pha của tương quan chéo, và được biến đến để thể hiện hiệu suất tốt hơn tương quan chéo thông thường trong nhiễu âm thấp và các môi trường âm vang tương đối cao. Hàm miền thời gian được thu trước hết được lọc để đạt được sự khuếch đại đỉnh chắc chắn. Chỉ số tương ứng với biên độ cực đại tương ứng với ước lượng của chênh lệch thời gian giữa kênh bên trái và bên phải (ITD). Nếu biên độ của giá trị cực đại thấp hơn ngưỡng đã cho, thì việc ước lượng ITD không được coi là đáng tin cậy và được thiết lập bằng không.

nếu căn chỉnh thời gian được áp dụng trong miền thời gian, ITD được tính toán trong phân tích DFT riêng biệt. Việc dịch chuyển được thực hiện như sau:

$$\begin{cases} r(n) = r(n + ITD) \text{ if } ITD > 0 \\ l(n) = l(n - ITD) \text{ if } ITD < 0 \end{cases}$$

Độ trễ bổ sung được yêu cầu tại bộ mã hóa, mà bằng cực đại đến ITD tuyệt đối cực đại mà có thể được xử lý. Sự biến thiên của ITD theo thời gian được làm nhẵn bởi việc tạo cửa sổ phân tích DFT.

Cách khác, căn chỉnh thời gian có thể được thực hiện trong miền tần số. Trong trường hợp này, việc tính toán ITD và dịch chuyển vòng tròn nằm trong cùng một miền DFT, miền được chia sẻ với việc xử lý âm lập thể khác này. Dịch chuyển vòng tròn được đưa ra bởi:

$$\begin{cases} L(f) = L(f) e^{-j2\pi f \frac{ITD}{2}} \\ R(f) = R(f) e^{+j2\pi f \frac{ITD}{2}} \end{cases}$$

Đệm số không của các cửa sổ DFT là cần thiết cho việc mô phỏng việc dịch

chuyển theo thời gian với dịch chuyển vòng tròn. Kích thước của đệm số không tương ứng với ITD tuyệt đối cực đại mà có thể được xử lý. Trong phương án được ưu tiên, đệm số không được tách không đồng đều trên cả hai phía của các cửa sổ phân tích bằng cách cộng 3,125 mili giây của các số không vào cả hai đầu. Sau đó ITD khả thi tuyệt đối cực đại bằng 6,25 mili giây. Trong thiết lập ống phóng thanh A-B, nó tương ứng với trường hợp xấu nhất tới khoảng cách tối đa bằng khoảng 2,15 mét giữa hai ống phóng thanh. Sự biến thiên trong ITD theo thời gian được làm nhẵn bởi việc tạo cửa sổ tổng hợp và chồng lấp cộng DFT.

Quan trọng là việc dịch chuyển thời gian sau theo sau bởi việc tạo cửa sổ của tín hiệu được dịch chuyển. Đó là điểm khác biệt chính với tình trạng kỹ thuật mã hóa tín hiệu điều khiển lập thể (Binaural Cue Coding - BCC), trong đó việc dịch chuyển thời gian được áp dụng trên tín hiệu được tạo cửa sổ như không được tạo cửa sổ thêm tại tầng tổng hợp. Kết quả là, bất kỳ thay đổi trong ITD theo thời gian tạo ra sự chuyển tiếp/tiếng lạch cách nhân tạo trong tín hiệu được giải mã.

4. Tính toán các IPD và luân phiên kênh

Các IPD được tính toán sau khi căn chỉnh thời gian hai kênh và điều này đối với mỗi băng tham số hoặc ít nhất bằng *ipd_max_band* đã cho, phụ thuộc vào cấu hình âm lập thể.

$$IPD[b] = \text{angle} \left(\sum_{k=\text{band_limits}[b]}^{\text{band_limits}[b+1]} L[k] R^*[k] \right)$$

Các IPD sau đó được áp dụng cho hai kênh để căn chỉnh các pha của chúng:

$$\begin{cases} L'(k) = L(k)e^{-j\beta} \\ R'(k) = R(k)e^{j(IPD[b]-\beta)} \end{cases}$$

Trong đó $\beta = \text{atan2}(\sin(IPD_i[b]), \cos(IPD_i[b])) + c$, $c = 10^{ILD_i[b]/20}$ và b là các chỉ số băng tham số mà thuộc về chỉ số tần số k . Tham số β tương ứng với việc phân phối lượng luân phiên theo pha giữa hai kênh trong khi căn chỉnh pha của chúng. β phụ thuộc vào IPD nhưng cũng phụ thuộc vào mức biên độ tương ứng của các kênh, ILD. Nếu kênh có biên độ cao hơn, nó sẽ được coi là kênh dẫn trước và sẽ ít bị ảnh hưởng bởi sự luân phiên theo pha hơn so với kênh với biên độ thấp hơn.

5. Chênh lệch tổng và mã hóa tín hiệu bên

Biến đổi chênh lệch tổng được thực hiện trên phổ được căn chỉnh theo thời gian và theo pha của hai kênh theo cách mà năng lượng được bảo tồn trong tín hiệu giữa.

$$\begin{cases} M(f) = (L'(f) + R'(f)) \cdot a \cdot \sqrt{\frac{1}{2}} \\ S(f) = (L'(f) - R'(f)) \cdot a \cdot \sqrt{\frac{1}{2}} \end{cases}$$

trong đó $a = \sqrt{\frac{L'^2 + R'^2}{(L' + R')^2}}$ được giới hạn trong khoảng từ 1/1,2 đến 1,2, tức là -1,58 dB và +1,58 dB. Các giới hạn này tránh sự giả tạo khi điều chỉnh năng lượng của M và S. Đáng lưu ý rằng sự bảo toàn năng lượng này ít quan trọng hơn khi thời gian và pha được căn chỉnh từ trước. Ngoài ra, các giới hạn này có thể được tăng lên hoặc giảm xuống.

Tín hiệu bên S còn được dự báo với M:

$$S'(f) = S(f) - g(ILD)M(f)$$

trong đó $g(ILD) = \frac{c-1}{c+1}$, trong đó $c = 10^{ILD_i[b]/20}$. Ngoài ra, độ khuếch đại dự báo tối ưu g có thể được phát hiện bằng cách tối thiểu hóa sai số bình phương trung bình (Mean Square Error - MSE) của phần dư và các ILD được suy ra bởi phương trình trước đó.

Tín hiệu dư $S'(f)$ có thể được mô hình hóa bằng hai cách: dự báo nó với phổ được làm trễ của M hoặc mã hóa nó một cách trực tiếp trong miền MDCT.

6. Mã hóa âm lập thể

Tín hiệu giữa X và tín hiệu bên S trước hết được chuyển đổi thành các kênh bên trái và phải L và R như sau:

$$L_i[k] = M_i[k] + gM_i[k], \text{ for } band_limits[b] \leq k < band_limits[b+1],$$

$$R_i[k] = M_i[k] - gM_i[k], \text{ for } band_limits[b] \leq k < band_limits[b+1],$$

trong đó độ khuếch đại g cho mỗi băng tham số được suy ra từ tham số ILD:

$$g = \frac{c-1}{c+1}, \text{ trong đó } c = 10^{ILD_i[b]/20}.$$

Đối với các băng tham số dưới cod_max_band , hai kênh được cập nhật với tín hiệu bên được giải mã:

$$L_i[k] = L_i[k] + cod_gain_i \cdot S_i[k], \text{ for } 0 \leq k < band_limits[cod_max_band],$$

$$R_i[k] = R_i[k] - cod_gain_i \cdot S_i[k], \text{ for } 0 \leq k < band_limits[cod_max_band],$$

Đối với các băng tham số cao hơn, tín hiệu bên được dự báo và các kênh được cập nhật như:

$$L_i[k] = L_i[k] + cod_pred_i[b] \cdot M_{i-1}[k], \text{ for } band_limits[b] \leq k < band_limits[b+1],$$

$$R_i[k] = R_i[k] - cod_pred_i[b] \cdot M_{i-1}[k], \text{ for } band_limits[b] \leq k < band_limits[b+1],$$

Cuối cùng, các kênh được nhân bởi giá trị phức nhằm khôi phục năng lượng ban đầu và pha liên kênh của tín hiệu âm lập thể:

$$L_i[k] = a \cdot e^{j2\pi\beta} \cdot L_i[k]$$

$$R_i[k] = a \cdot e^{j2\pi\beta - IPD_i[b]} \cdot R_i[k]$$

trong đó:

$$a = \sqrt{2 \cdot \frac{\sum_{k=band_limits[b]}^{band_limits[b+1]} M_i^2[k]}{\sum_{k=band_limits[b]}^{band_limits[b+1]-1} L_i^2[k] + \sum_{k=band_limits[b]}^{band_limits[b+1]-1} R_i^2[k]}}$$

trong đó a được xác định và được giới hạn như được xác định trước đó, và trong đó $\beta = \text{atan2}(\sin(IPD_i[b]), \cos(IPD_i[b]) + c)$, và trong đó $\text{atan2}(x,y)$ là tiếp tuyến nghịch đảo góc phân tư của x trên y .

Cuối cùng, các kênh được dịch chuyển thời gian trong miền thời gian hoặc trong miền tần số phụ thuộc vào các ITD được truyền dẫn. Các kênh miền thời gian được tổng hợp bởi các DFT nghịch đảo và chồng lắp cộng.

Các dấu hiệu cụ thể của sáng chế liên quan đến tổ hợp các tín hiệu điều khiển

theo không gian và mã hóa âm lập thể kết hợp chênh lệch tổng. Cụ thể, các tín hiệu điều khiển theo không gian IDT và IPD được tính toán và được áp dụng lên các kênh âm lập thể (trái và phải). Ngoài ra, chênh lệch tổng (các tín hiệu M/S) được tính toán và tốt hơn là dự báo được áp dụng của S với M.

Ở phía bộ giải mã, các tín hiệu điều khiển theo không gian băng rộng và băng hẹp được tổ hợp với mã hóa âm lập thể kết hợp chênh lệch tổng. Cụ thể, tín hiệu bên được dự báo với tín hiệu giữa sử dụng ít nhất một tín hiệu điều khiển theo không gian chẳng hạn như ILD và chênh lệch tổng nghịch đảo được tính toán để có các kênh bên trái và bên phải, ngoài ra, các tín hiệu điều khiển theo không gian băng rộng và băng hẹp được áp dụng lên các kênh bên trái và bên phải.

Ưu tiên hơn, bộ mã hóa có cửa sổ và chồng lấp cộng liên quan đến các kênh được căn chỉnh thời gian sau khi xử lý sử dụng ITD. Hơn nữa, bộ giải mã còn có việc tạo cửa sổ và phép toán chồng lấp-cộng của các phiên bản được dịch chuyển hoặc được hủy căn chỉnh của các kênh sau khi áp dụng chênh lệch thời gian liên kênh.

Việc tính toán chênh lệch thời gian liên kênh với phương pháp GCC-PHAT là phương pháp đặc biệt chắc chắn.

Quy trình mới này là tình trạng kỹ thuật có lợi do nó đạt được mã hóa tốc độ bit của âm thanh lập thể hoặc âm thanh đa kênh tại độ trễ thấp. Nó được thiết kế cụ thể để tăng mạnh các bản chất khác nhau của các tín hiệu đầu và các thiết lập khác nhau của việc ghi âm lập thể hoặc đa kênh. Cụ thể, sáng chế đề xuất chất lượng tốt cho mã hóa tiếng nói âm lập thể tại tốc độ bit.

Quy trình được ưu tiên hơn phát hiện việc sử dụng trong phân phối phát sóng tất cả các loại âm lập thể hoặc nội dung âm thanh đa kênh chẳng hạn như tiếng nói và âm nhạc với chất lượng cảm quan không đổi tại tốc độ bit thấp đã cho. Các lĩnh vực ứng dụng là radio số, các ứng dụng trực tuyến Internet hoặc truyền thông âm thanh.

Các tín hiệu âm thanh được mã hóa theo sáng chế có thể được lưu trữ trên vật ghi lưu trữ số hoặc môi trường lưu trữ không chuyển tiếp hoặc có thể được truyền trong môi trường truyền dẫn như môi trường truyền không dây hoặc môi trường truyền có dây như Internet.

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

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 vật ghi lưu trữ số, ví dụ, đĩa mềm, DVD, CD, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ FLASH, có các tín hiệu điều khiển có thể đọc được bằng điện tử được lưu trữ trên đó, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính có thể lập trình được sao cho phương pháp tương ứng được thực hiện.

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

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

Các phương án khác 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 có thể đọc được bằng máy hoặc vật ghi lưu trữ không chuyển tiếp.

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

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

Phương án khác của các phương pháp theo sáng chế là, do đó, dòng dữ liệu

hoặc chuỗi các tín hiệu biểu diễn chương trình máy tính để thực hiện một trong các phương pháp đã được mô tả ở đây. Ví dụ, dòng dữ liệu hoặc chuỗi các tín hiệu có thể được tạo cấu hình để được truyền thông qua kết nối truyền thông dữ liệu, ví dụ thông qua Internet.

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

Phương án khác gồm có máy tính có chương trình máy tính được cài đặt 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. Trong một số phương án, mảng cổng lập trình được dạng trường có thể kết hợp với bộ vi xử lý để thực hiện một trong số các phương pháp được mô tả ở đây. Nói chung, các phương pháp tốt hơn là được thực hiện bởi thiết bị phần cứng bất kỳ.

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

YÊU CẦU BẢO HỘ

1. Thiết bị ước lượng chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai, thiết bị này bao gồm:

bộ tính toán (1020) để tính toán phổ tương quan chéo cho khối thời gian từ tín hiệu kênh thứ nhất trong khối thời gian và tín hiệu kênh thứ hai trong khối thời gian;

bộ ước lượng đặc tính phổ (1010) để ước lượng đặc tính của phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai cho khối thời gian;

bộ lọc làm nhẵn (1030) để làm nhẵn phổ tương quan chéo theo thời gian sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm nhẵn; và

bộ xử lý (1040) để xử lý phổ tương quan chéo được làm nhẵn để thu được chênh lệch thời gian liên kênh.

2. Thiết bị theo điểm 1,

trong đó bộ xử lý (1040) được tạo cấu hình để chuẩn hóa (456) phổ tương quan chéo được làm nhẵn sử dụng cường độ của phổ tương quan chéo được làm nhẵn.

3. Thiết bị theo điểm 1 hoặc điểm 2,

trong đó bộ xử lý (1040) được tạo cấu hình:

để tính toán (1031) sự biểu diễn miền thời gian của phổ tương quan chéo được làm nhẵn hoặc phổ tương quan chéo được làm nhẵn được chuẩn hóa; và

để phân tích (1032) sự biểu diễn miền thời gian để xác định chênh lệch thời gian liên kênh.

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

trong đó bộ xử lý (1040) được tạo cấu hình để lọc thông thấp (458) sự biểu diễn miền thời gian và để tiếp tục xử lý (1033) kết quả của việc lọc thông thấp.

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

trong đó bộ xử lý được tạo cấu hình để thực hiện sự xác định chênh lệch thời gian liên kênh bằng cách thực hiện phép toán tìm kiếm đỉnh hoặc chọn lọc đỉnh trong sự biểu diễn miền thời gian được xác định từ phổ tương quan chéo được làm nhẵn.

6. Thiết bị theo điểm bất kỳ trong số các điểm nêu trên,

trong đó bộ ước lượng đặc tính phổ (1010) được tạo cấu hình để xác định, như là đặc tính phổ, nhiều âm hoặc âm điệu của phổ; và

trong đó bộ lọc làm nhẵn (1030) được tạo cấu hình để áp dụng việc làm nhẵn mạnh hơn theo thời gian với mức độ làm nhẵn thứ nhất trong trường hợp đặc tính ít nhiều âm thứ nhất hoặc đặc tính nhiều âm điệu thứ nhất, hoặc để áp dụng việc làm nhẵn yếu hơn theo thời gian với mức độ làm nhẵn thứ hai trong trường hợp đặc tính nhiều nhiều âm thứ hai hoặc đặc tính ít âm điệu thứ hai,

trong đó mức độ làm nhẵn thứ nhất lớn hơn mức độ làm nhẵn thứ hai, và trong đó đặc tính nhiều âm thứ nhất ít nhiều âm hơn đặc tính nhiều âm thứ hai, hoặc đặc tính âm điệu thứ nhất nhiều âm điệu hơn đặc tính âm điệu thứ hai.

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

trong đó bộ ước lượng đặc tính phổ (101) được tạo cấu hình để tính toán, như đặc tính, số đo độ phẳng phổ thứ nhất của phổ của tín hiệu kênh thứ nhất và số đo độ phẳng phổ thứ hai của phổ thứ hai của tín hiệu kênh thứ hai, và để xác định đặc tính của phổ từ số đo độ phẳng phổ thứ nhất và thứ hai bằng cách lựa chọn giá trị cực đại, bằng cách xác định trung bình được gán trọng số hoặc trung bình không được gán trọng số giữa các số đo độ phẳng phổ, hoặc bằng cách lựa chọn giá trị cực tiểu.

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

trong đó bộ lọc làm nhẵn (1030) được tạo cấu hình để tính toán giá trị phổ tương quan chéo được làm nhẵn cho tần số bằng tổ hợp được gán trọng số của giá trị phổ tương quan chéo cho tần số từ khối thời gian và giá trị phổ tương quan chéo cho tần số từ ít nhất một khối thời gian trước đó, trong đó các hệ số gán trọng số cho tổ hợp được gán trọng số được xác định bởi đặc tính của phổ.

9. Thiết bị theo điểm bất kỳ trong số các điểm nêu trên,

trong đó bộ xử lý (1040) được tạo cấu hình để xác định khoảng hợp lệ và khoảng không hợp lệ trong sự biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm nhẵn,

trong đó ít nhất một đỉnh cực đại trong khoảng không hợp lệ được phát hiện và được so sánh với đỉnh cực đại trong khoảng hợp lệ, trong đó chênh lệch thời gian liên kênh chỉ được xác định, khi đỉnh cực đại trong khoảng hợp lệ lớn hơn ít nhất một đỉnh cực đại trong khoảng không hợp lệ.

10. Thiết bị theo điểm bất kỳ trong số các điểm nêu trên,

trong đó bộ xử lý (1040) được tạo cấu hình:

để thực hiện phép toán tìm kiếm đỉnh trong sự biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm nhẵn,

để xác định (1034) ngưỡng biến số từ sự biểu diễn miền thời gian; và

để so sánh (1035) đỉnh với ngưỡng biến số, trong đó chênh lệch thời gian liên kênh được xác định là độ trễ thời gian liên quan đến đỉnh nằm trong mối quan hệ được xác định trước với ngưỡng biến số.

11. Thiết bị theo điểm 10,

trong đó bộ xử lý được tạo cấu hình để xác định ngưỡng biến số (1334c) là giá trị bằng với bội số nguyên của giá trị trong số 10% giá trị lớn nhất của sự biểu diễn miền thời gian.

12. Thiết bị theo một trong số các điểm từ 1 đến 9,

trong đó bộ xử lý (1040) được tạo cấu hình để xác định biên độ đỉnh cực đại (1102) trong mỗi khối con của nhiều khối con của sự biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm nhẵn,

trong đó bộ xử lý (1040) được tạo cấu hình để tính toán (1104, 1105) ngưỡng biến số dựa trên cường độ đỉnh trung bình được suy ra từ các cường độ đỉnh cực đại của nhiều khối con, và

trong đó bộ xử lý được tạo cấu hình để xác định chênh lệch thời gian liên kênh như là giá trị độ trễ thời gian tương ứng với đỉnh cực đại của nhiều khối con lớn hơn ngưỡng biến số.

13. Thiết bị theo điểm 12,

trong đó bộ xử lý (1040) được tạo cấu hình để tính toán ngưỡng biến số bằng

phép nhân (1105) của ngưỡng trung bình được xác định là đỉnh trung bình trong số các đỉnh trong các khối con và giá trị,

trong đó giá trị được xác định (1104) bởi đặc tính tỉ số tín hiệu trên nhiễu (signal to noise ratio - SNR) của tín hiệu kênh thứ nhất và thứ hai, trong đó giá trị thứ nhất được liên kết với giá trị SNR thứ nhất và giá trị thứ hai được liên kết với giá trị SNR thứ hai, trong đó giá trị thứ nhất lớn hơn giá trị thứ hai, và trong đó giá trị SNR thứ nhất lớn hơn giá trị SNR thứ hai.

14. Thiết bị theo điểm 13,

trong đó bộ xử lý (1040) được tạo cấu hình để sử dụng (1104) giá trị thứ ba (a_{lowest}) thấp hơn giá trị thứ hai (a_{low}) trong trường hợp giá trị SNR thứ ba thấp hơn giá trị SNR thứ hai và khi chênh lệch giữa ngưỡng và đỉnh cực đại thấp hơn giá trị được định trước (ϵ).

15. Phương pháp ước lượng chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai, bao gồm:

tính toán (1020) phổ tương quan chéo cho khối thời gian từ tín hiệu kênh thứ nhất trong khối thời gian và tín hiệu kênh thứ hai trong khối thời gian;

ước lượng (1010) đặc tính của phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai cho khối thời gian;

làm nhẵn (1030) phổ tương quan chéo theo thời gian sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm nhẵn; và

xử lý (1040) phổ tương quan chéo được làm nhẵn để thu được chênh lệch thời gian liên kênh.

16. Vật ghi đọc được bằng máy tính bao gồm chương trình máy tính để thực hiện, khi chạy trên máy tính hoặc bộ xử lý, phương pháp theo điểm 15.

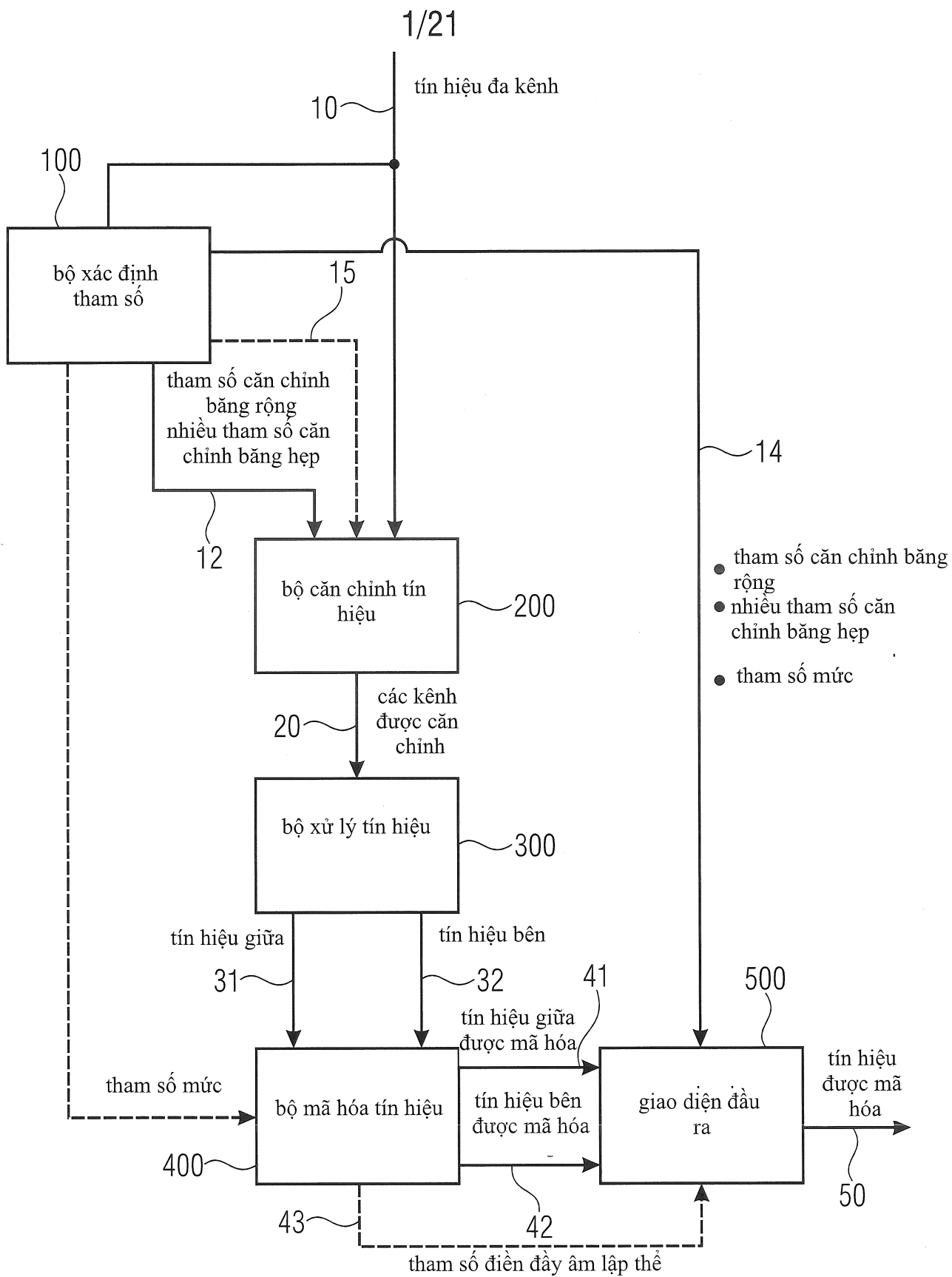


Fig. 1

2/21

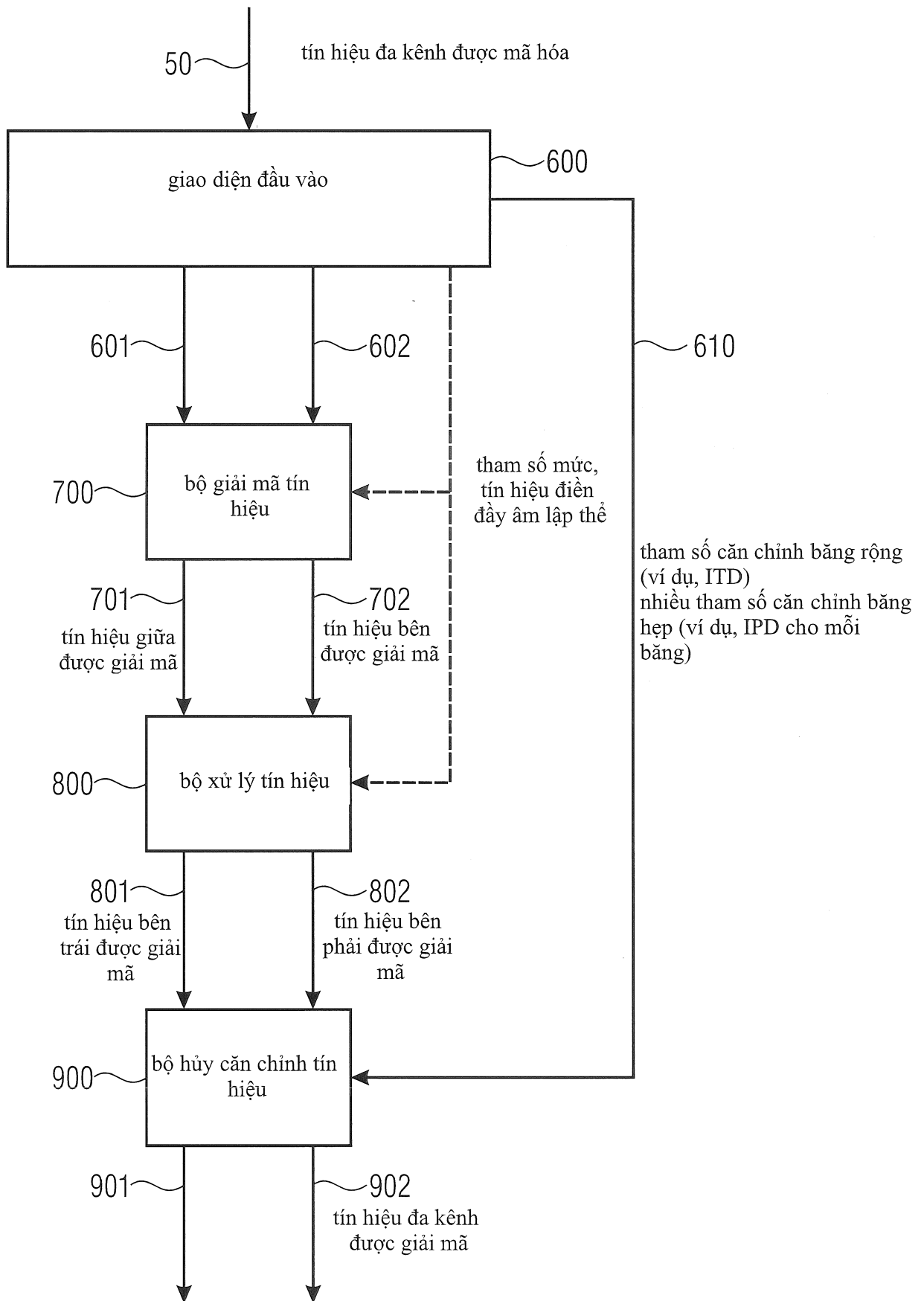
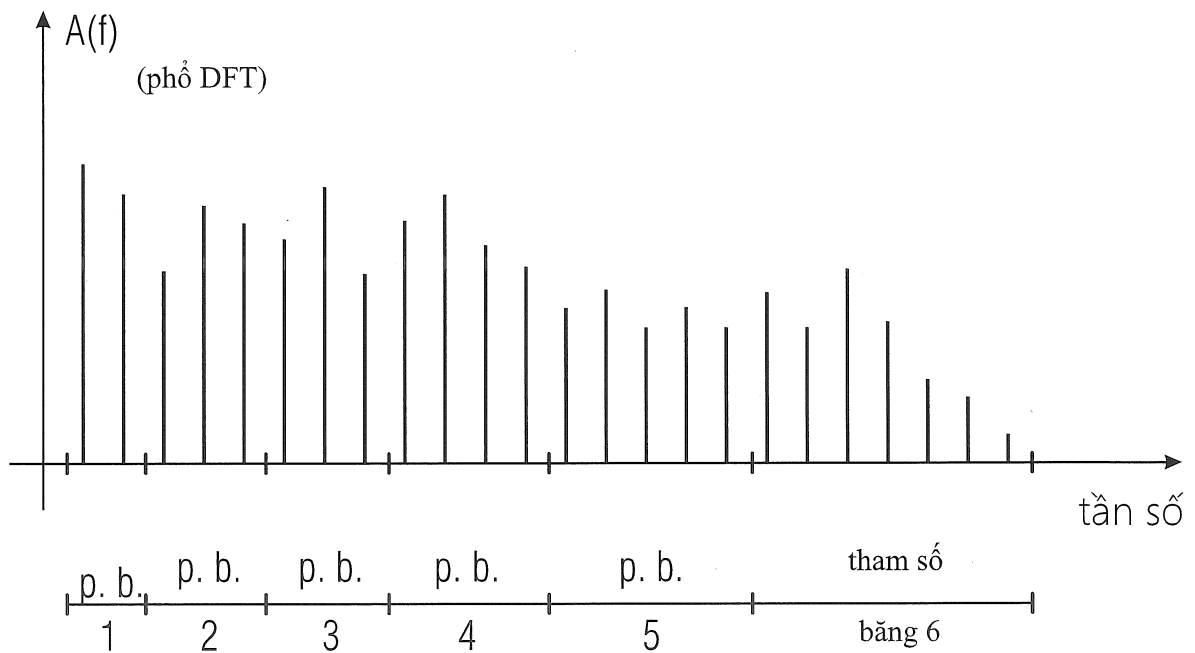


Fig. 2

3/21



- tham số căn chỉnh băng rộng đơn lẻ cho toàn bộ phổ (ví dụ băng tham số 1 đến băng tham số 6)
- nhiều tham số căn chỉnh băng hẹp cho các băng tham số 1,2,3,4 tức là 4 tham số băng hẹp;
- các tham số mức cho mỗi băng tham số, ví dụ, 6 tham số mức
- các tham số điền đầy âm lập thể cho các băng tham số 4, 5, 6, ví dụ, ba tham số điền đầy âm lập thể
- tín hiệu bên (dư) cho các băng tham số 1, 2, 3
- nhiều vạch phổ trong băng cao hơn, ví dụ, bảy vạch phổ trong băng tham số 6 so với ba vạch phổ trong băng tham số 2

Fig. 3

4/21

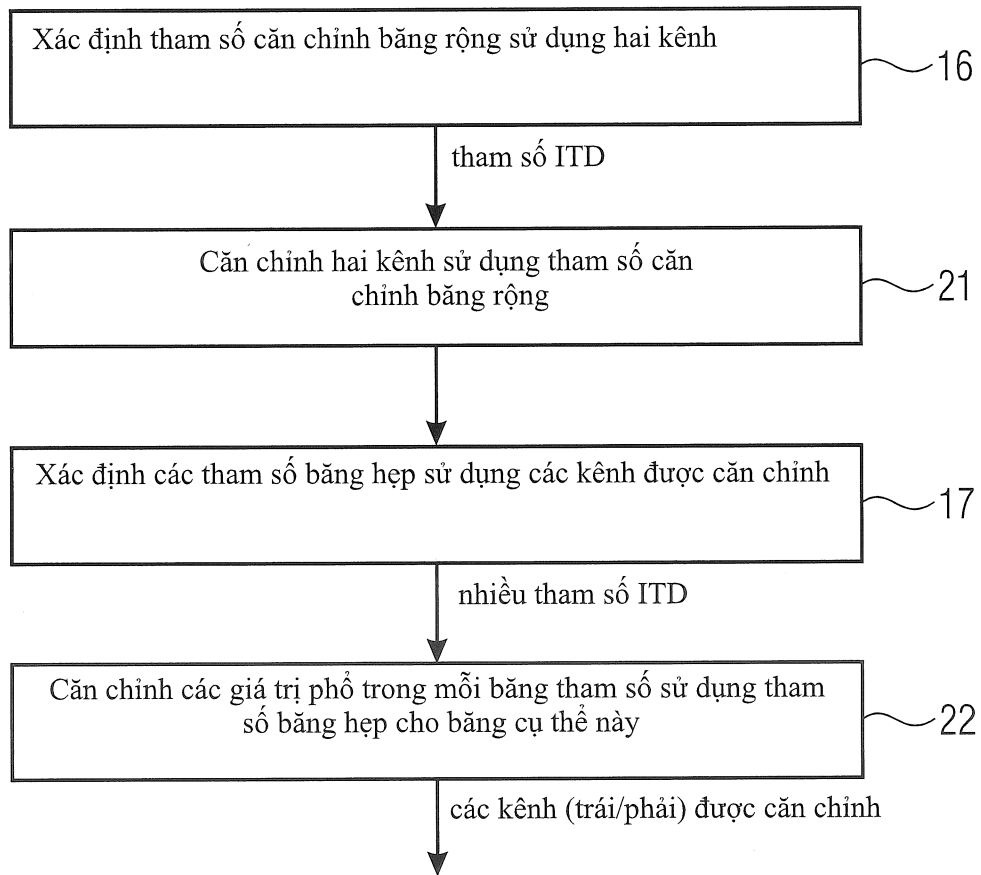


Fig. 4a

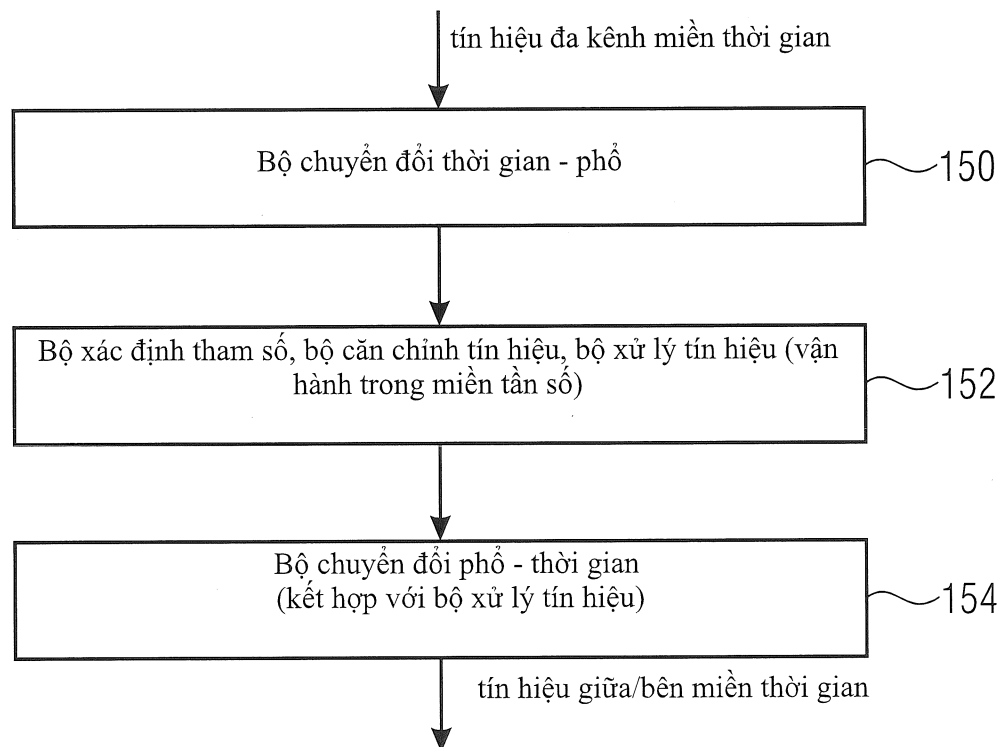


Fig. 4b

5/21

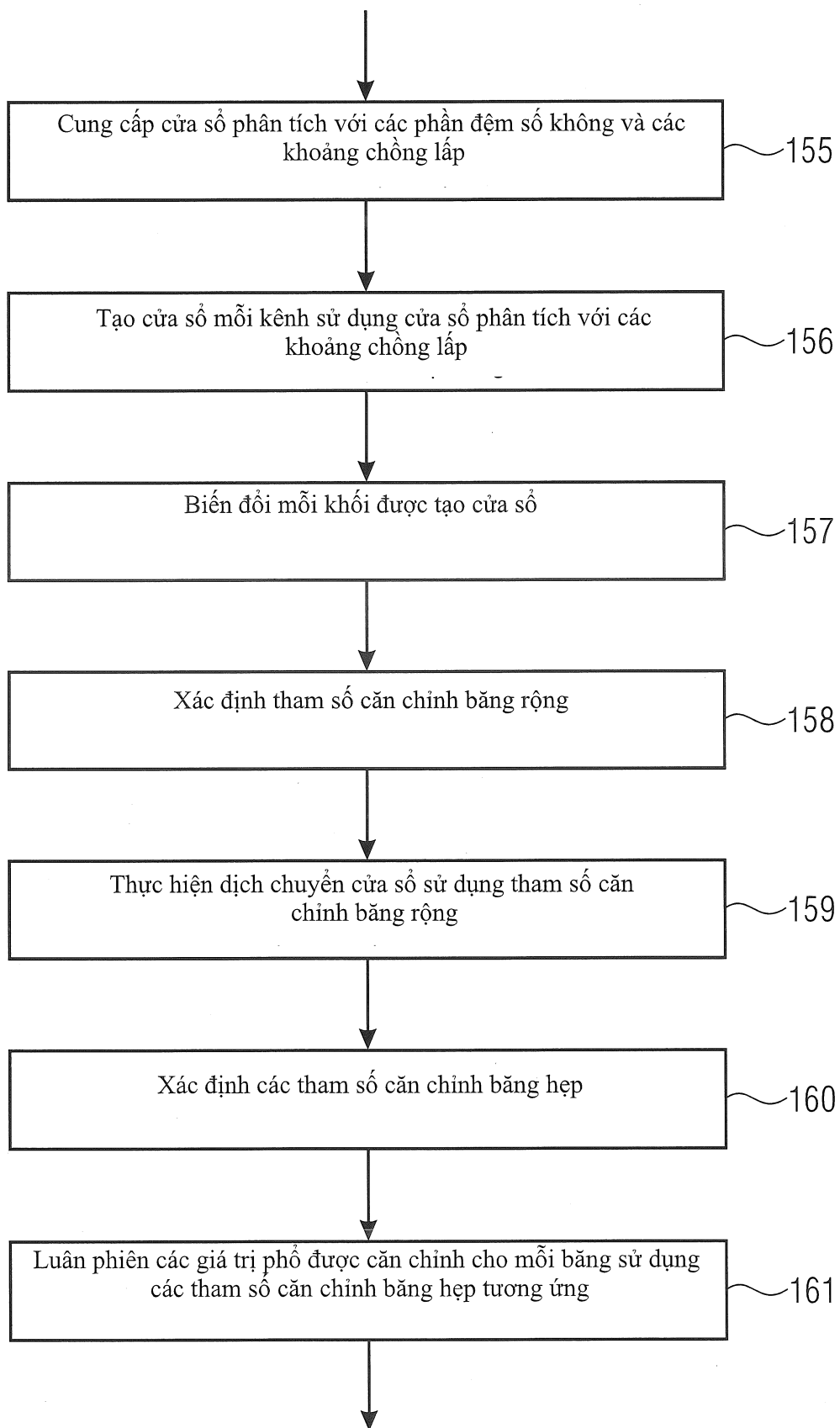


Fig. 4c

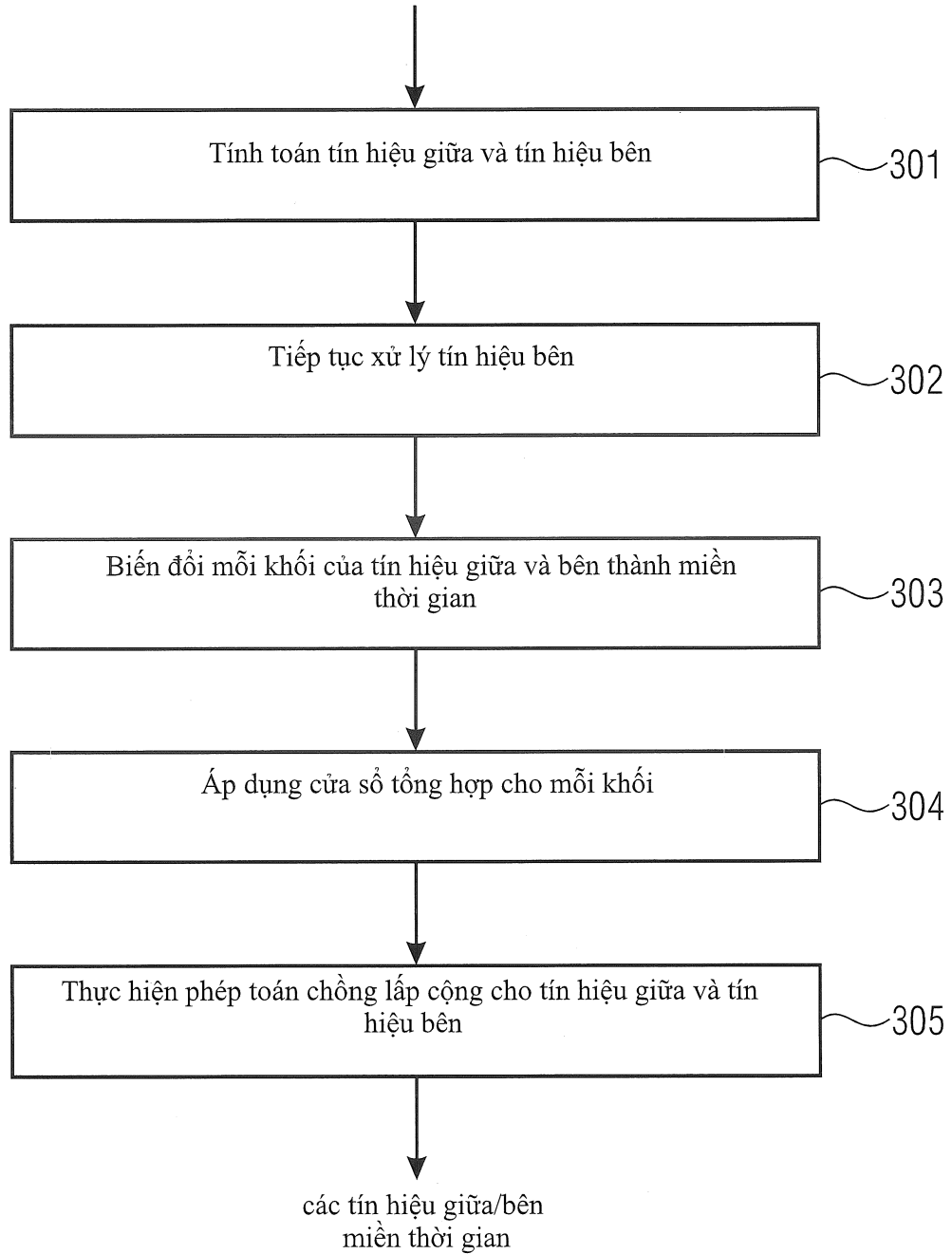


Fig. 4d

7/21

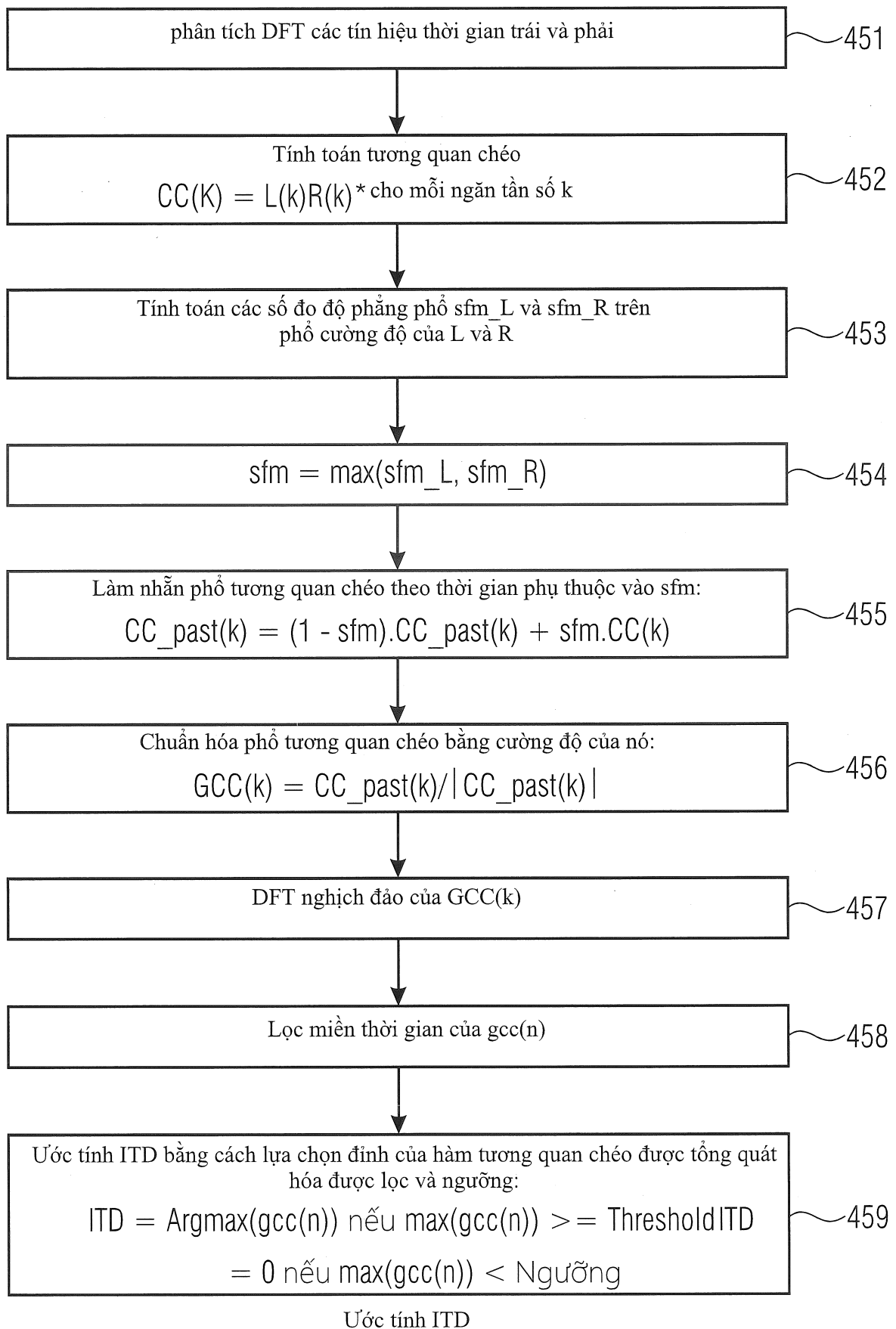


Fig. 4e

8/21

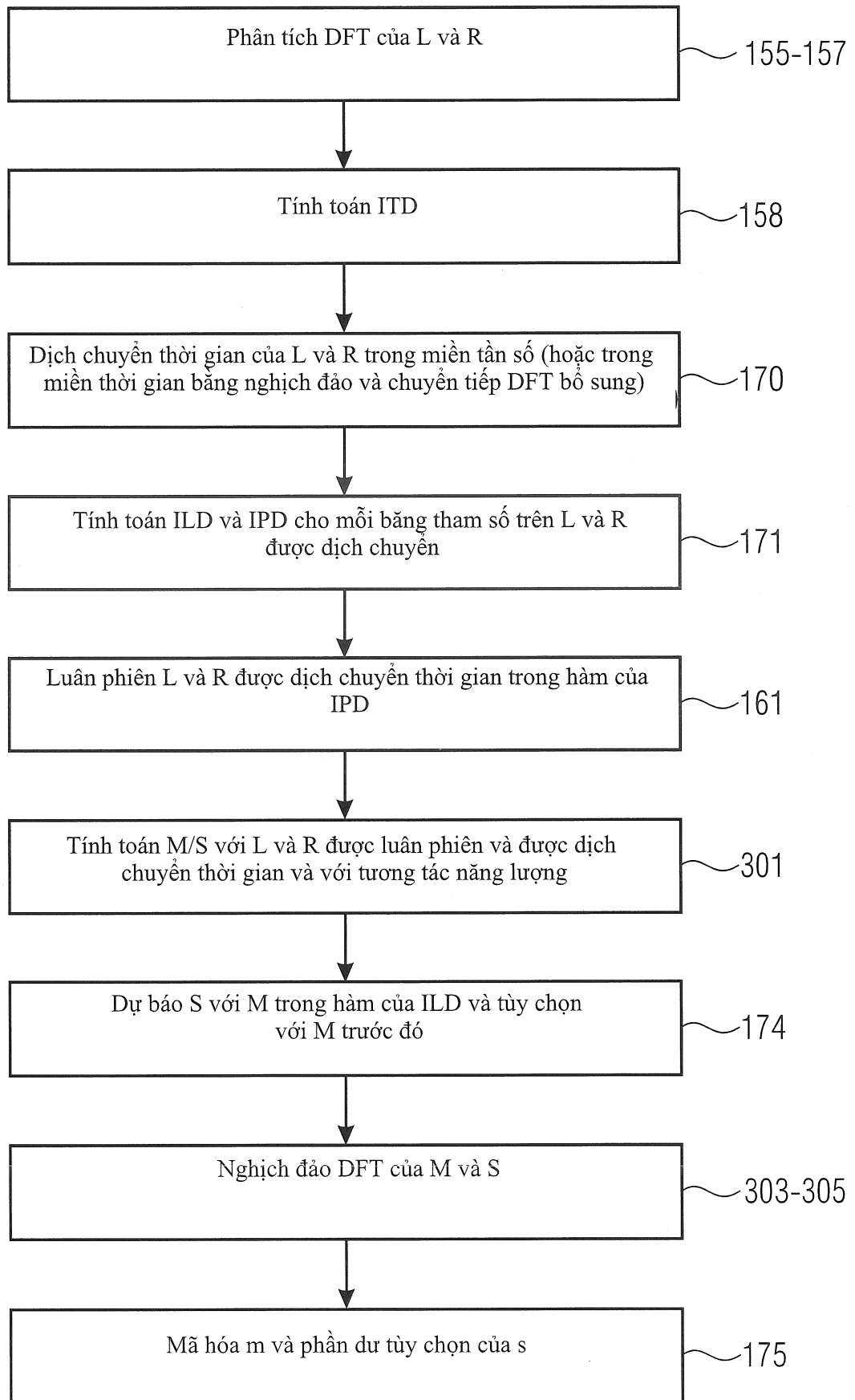


Fig. 5

9/21

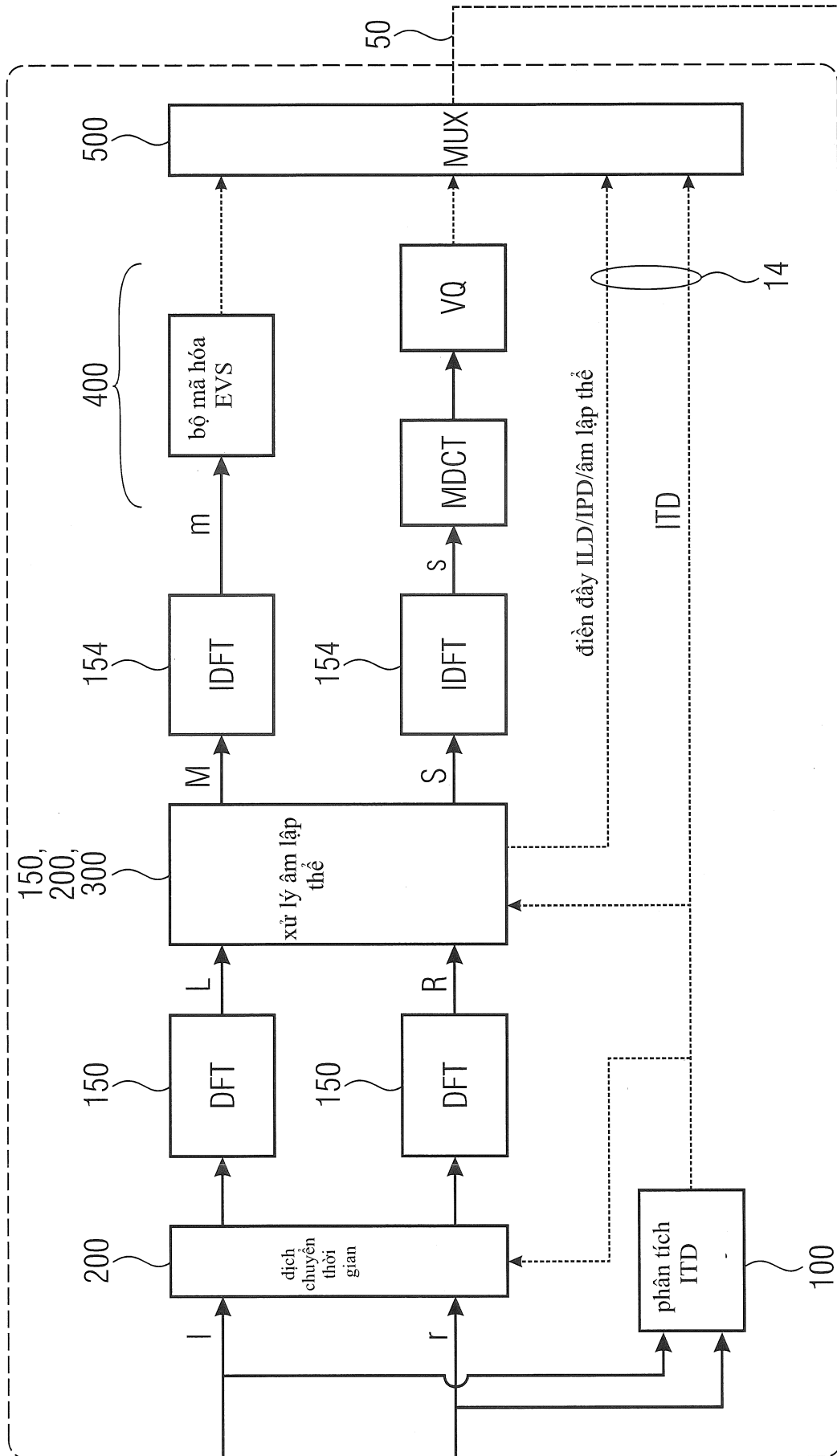


Fig. 6a

10/21

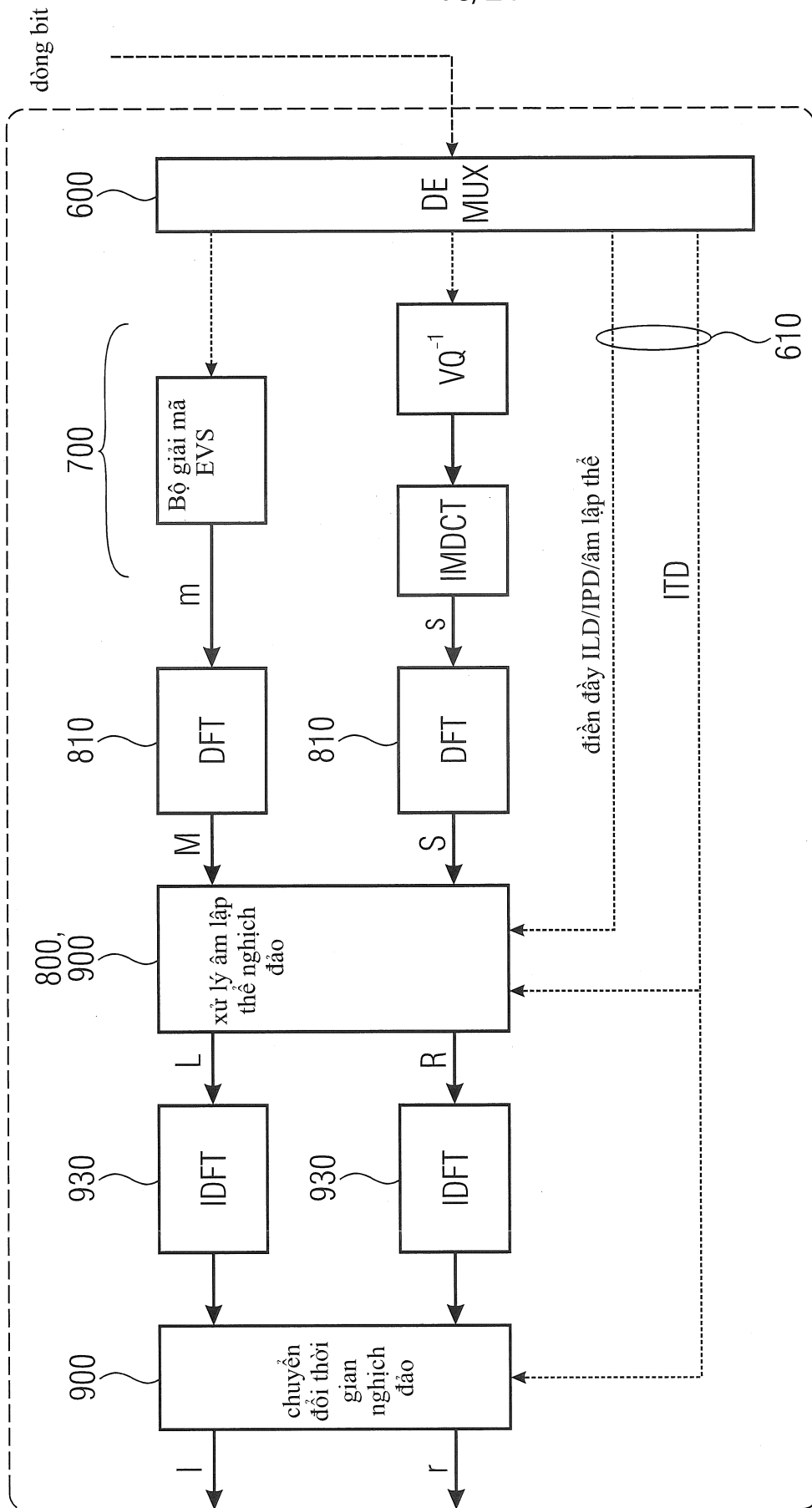


Fig. 6b

11/21

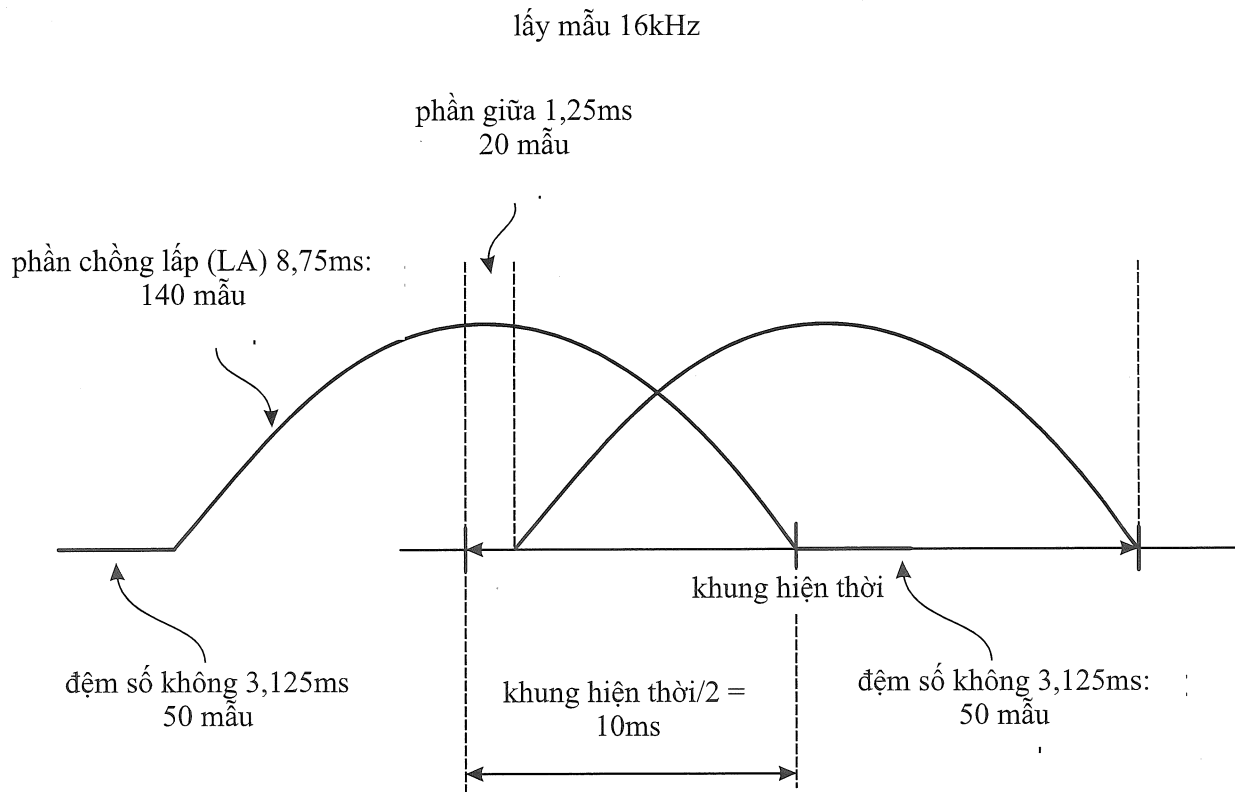


Fig. 7

Các tham số	các bit/băng	số lượng các băng	tốc độ bit
ILD	5	tất cả 12 băng	3,00 kbps
IPD	3	đến 2,5kHz	1,05 kbps
ITD	8	toàn bộ phổ	0,6 kbps
điền đầy âm lập thể	3	từ 1kHz	0,9 kbps
tổng			-5 kbps

Fig. 8

12/21

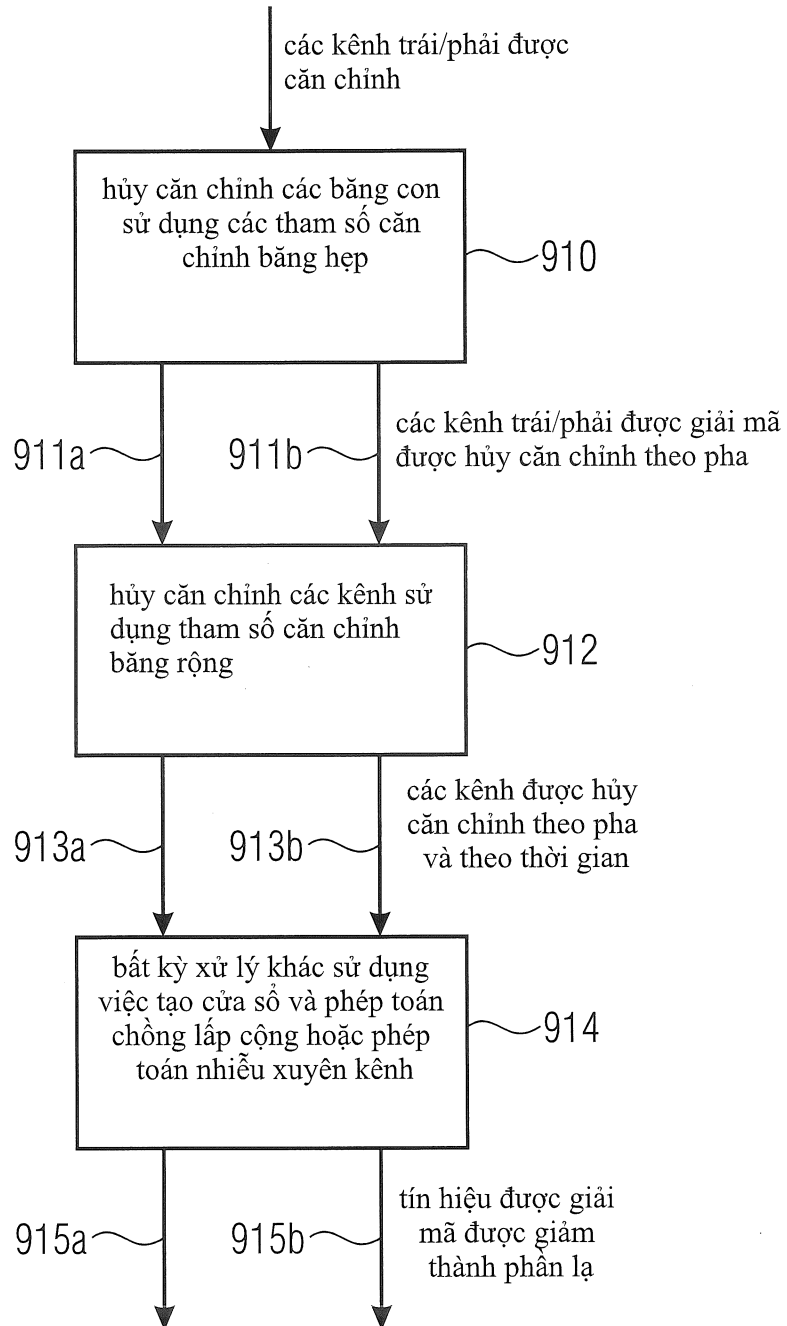


Fig. 9a

13/21

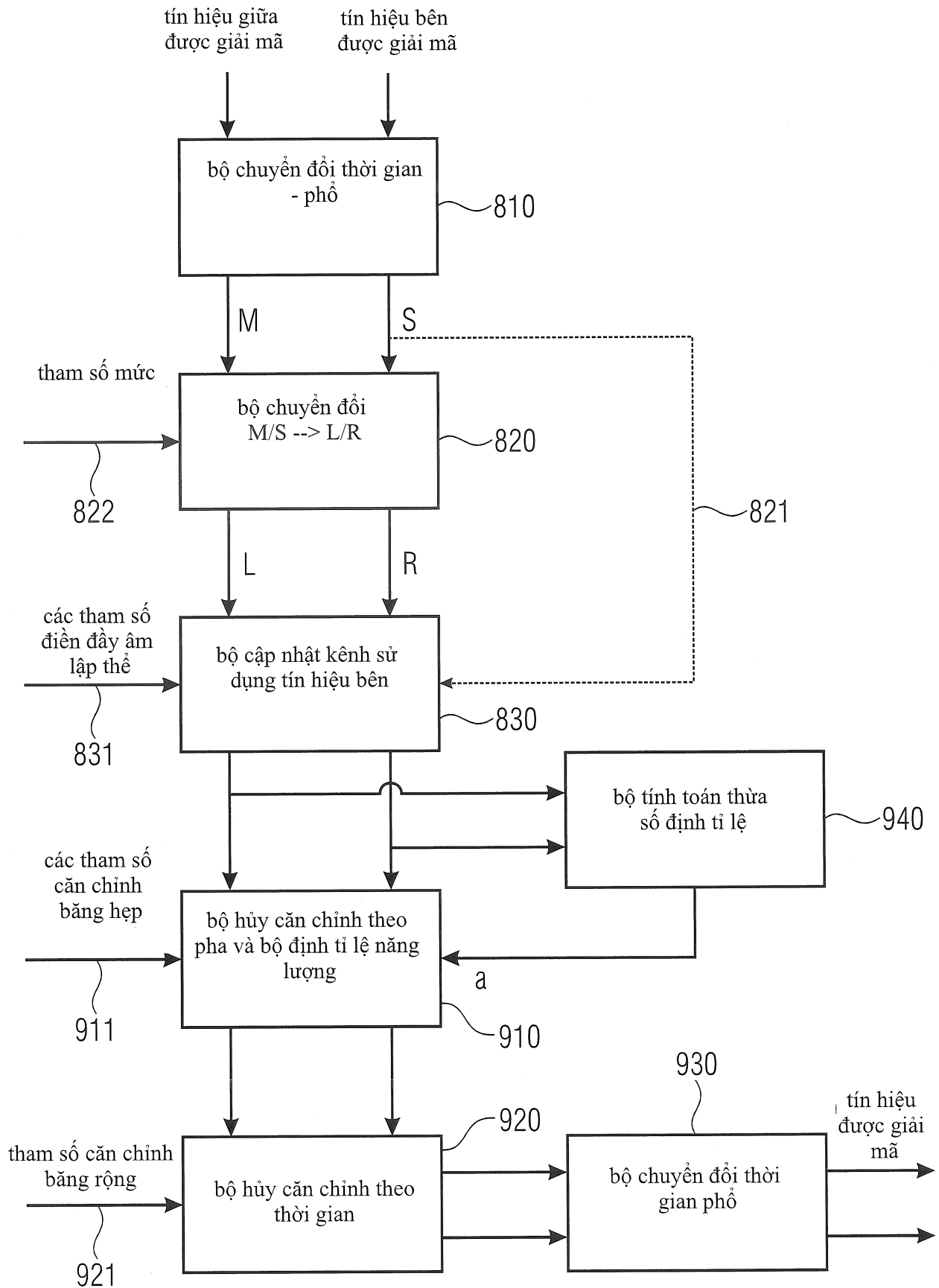


Fig. 9b

14/21

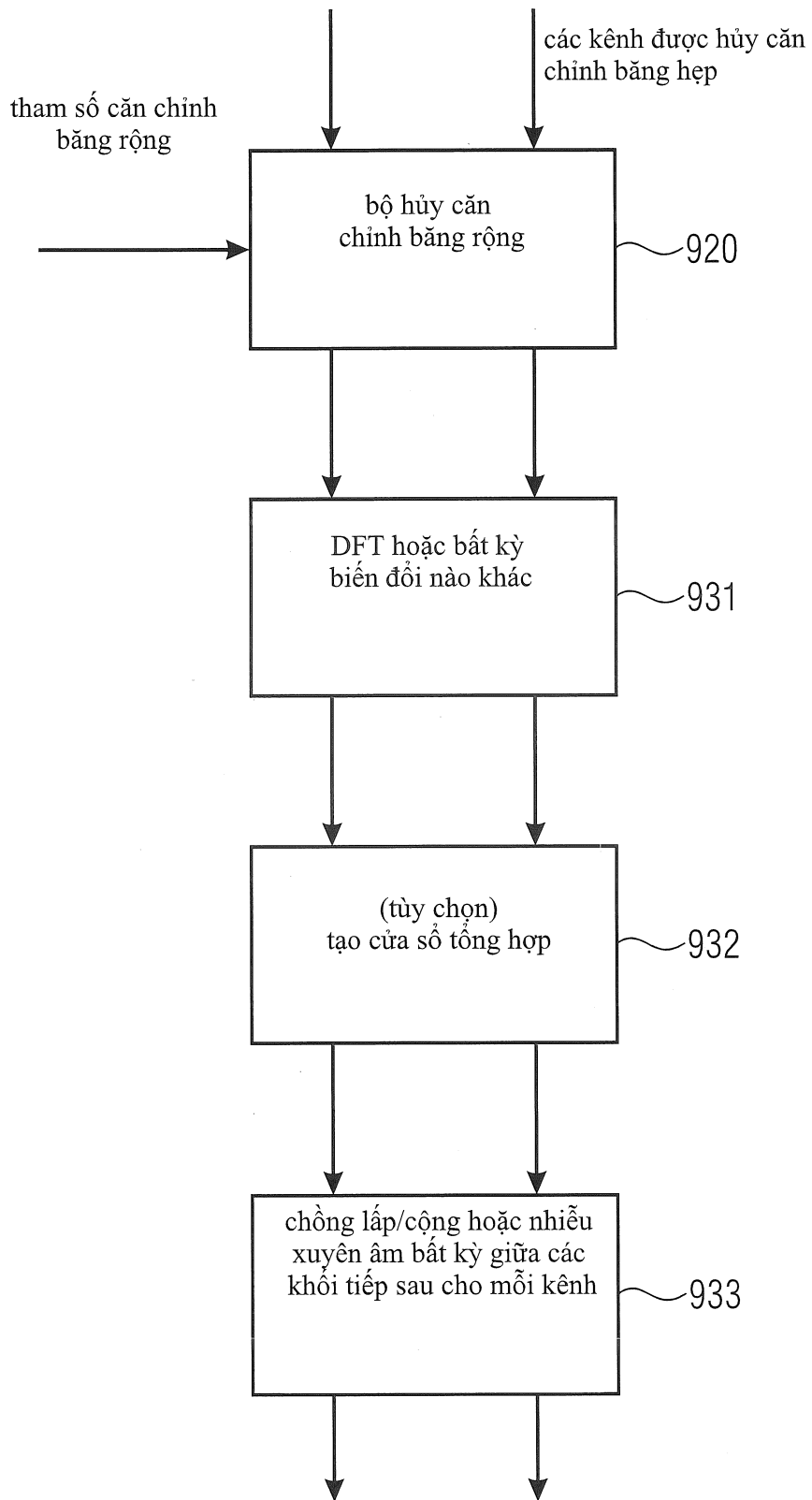


Fig. 9c

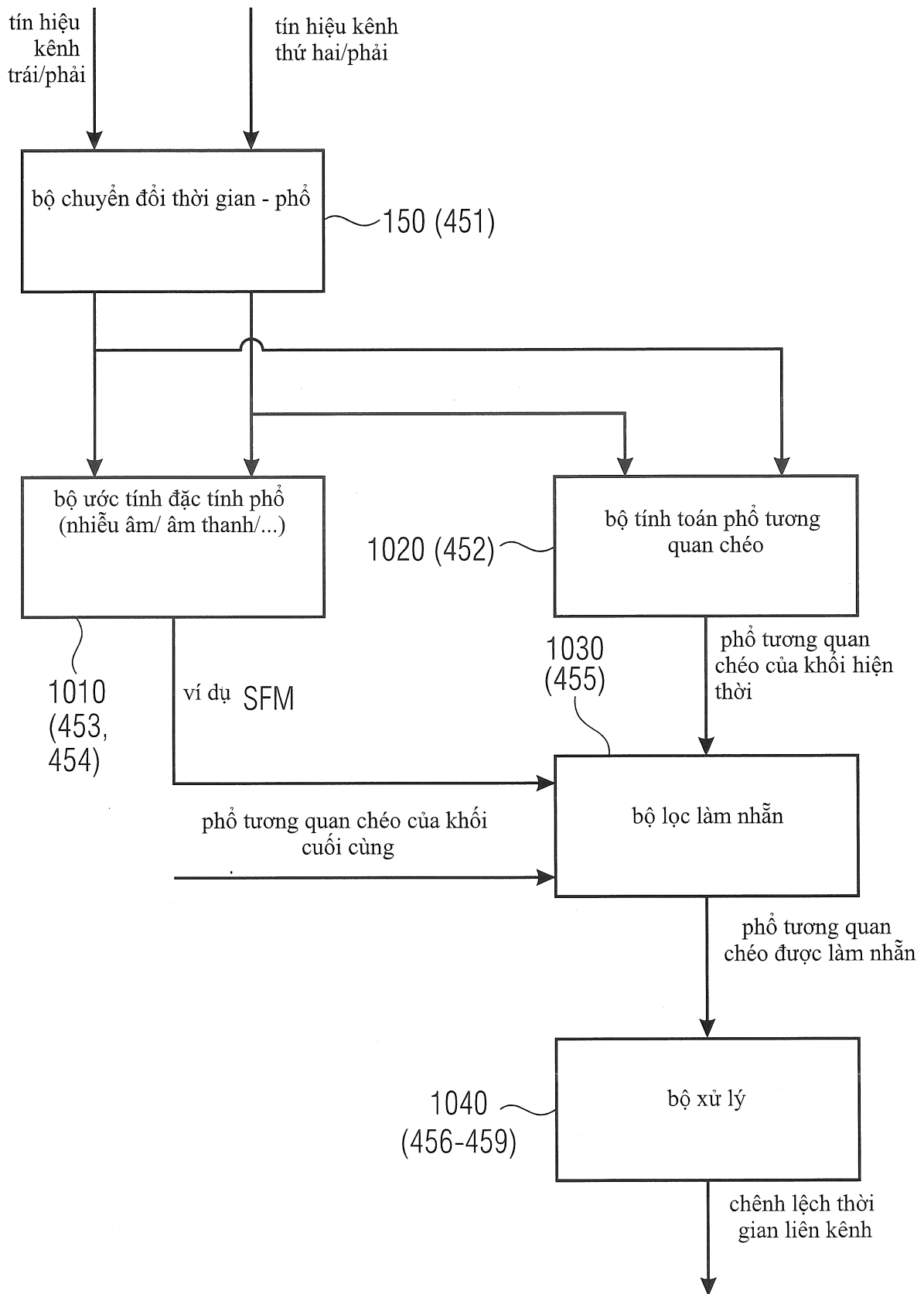
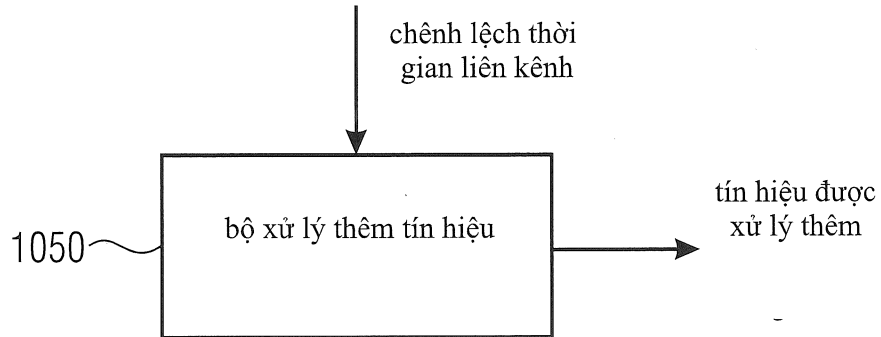


Fig. 10a



- lưu trữ/ truyền dẫn dữ liệu theo tham số
- xử lý/mã hóa âm lập thể/đa kênh
- căn chỉnh thời gian hai kênh
- ước tính chênh lệch thời gian tới để xác định vị trí người nói trong phòng với hai ống phóng thanh và thiết lập ống phóng thanh đã biết
- điều hướng chùm sóng
- lọc không gian
- hủy căn chỉnh tiền cảnh/hậu cảnh
- vị trí của nguồn âm thanh
ví dụ bằng phép đặc tam giác âm thanh dựa trên chênh lệch thời gian của hai/ba tín hiệu

Fig. 10b

17/21

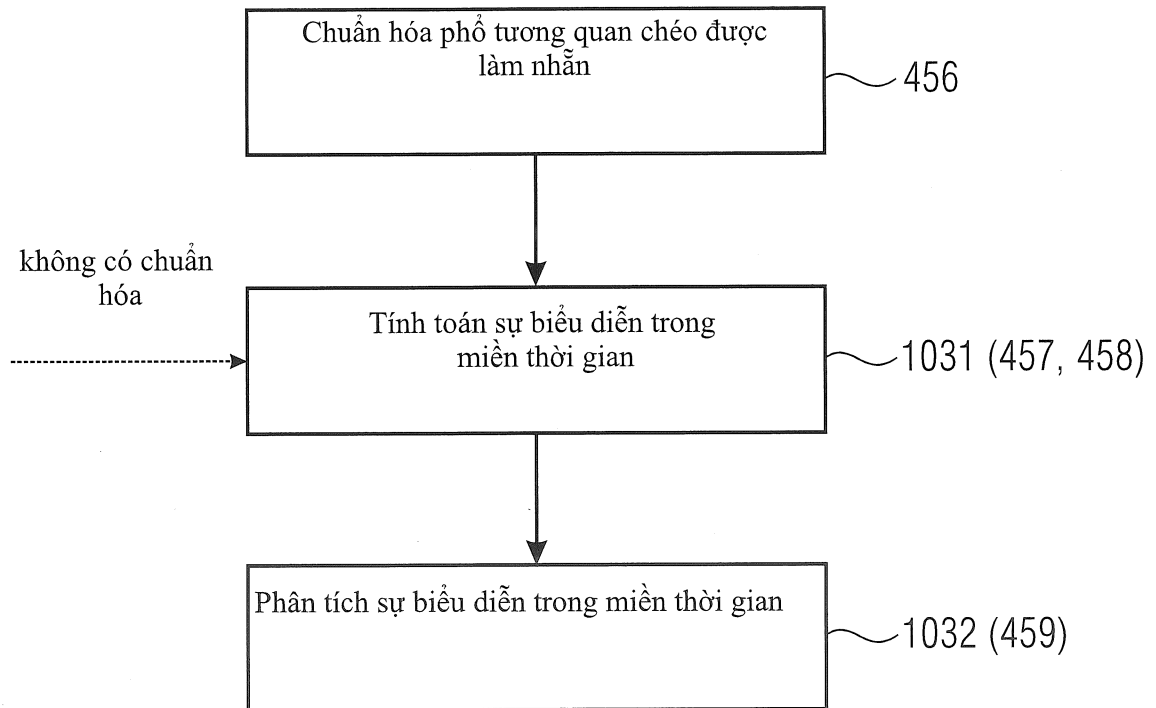


Fig. 11a

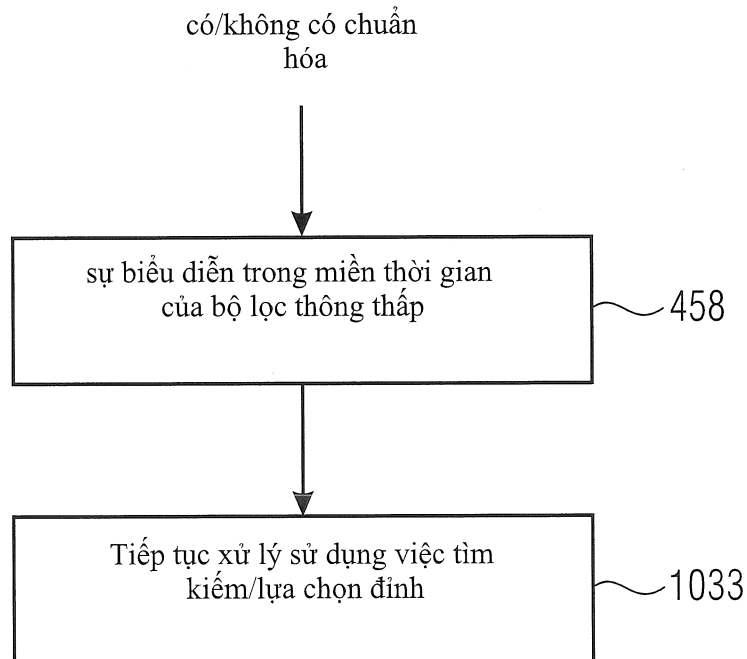


Fig. 11b

18/21

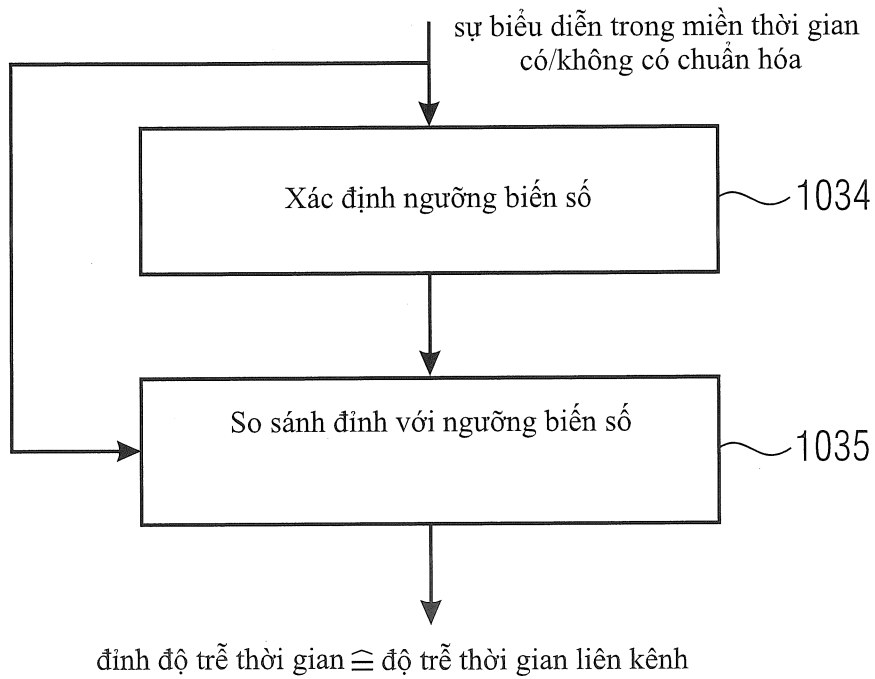


Fig. 11c

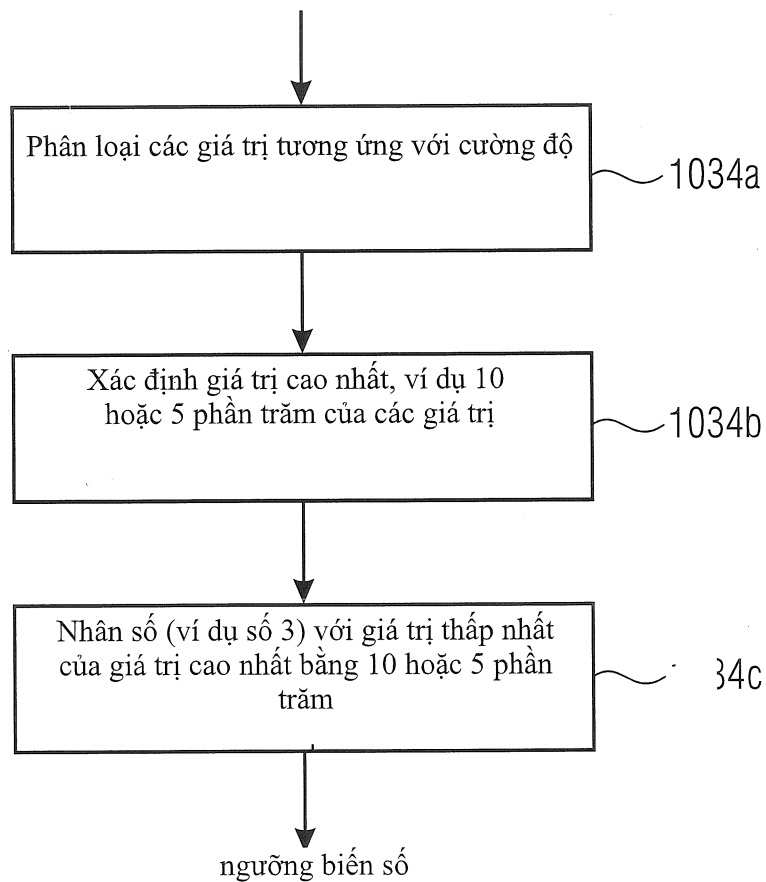


Fig. 11d

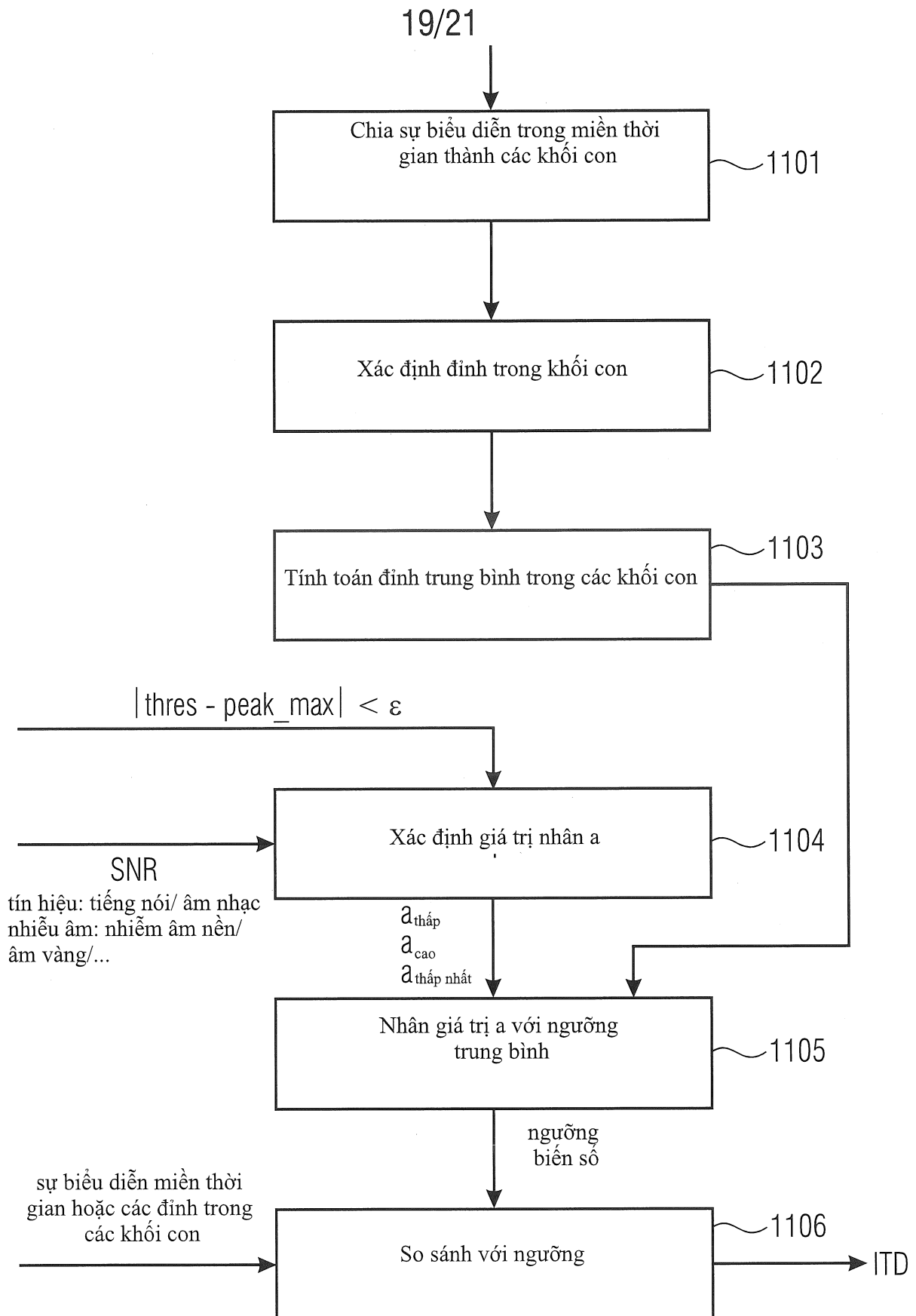


Fig. 11e

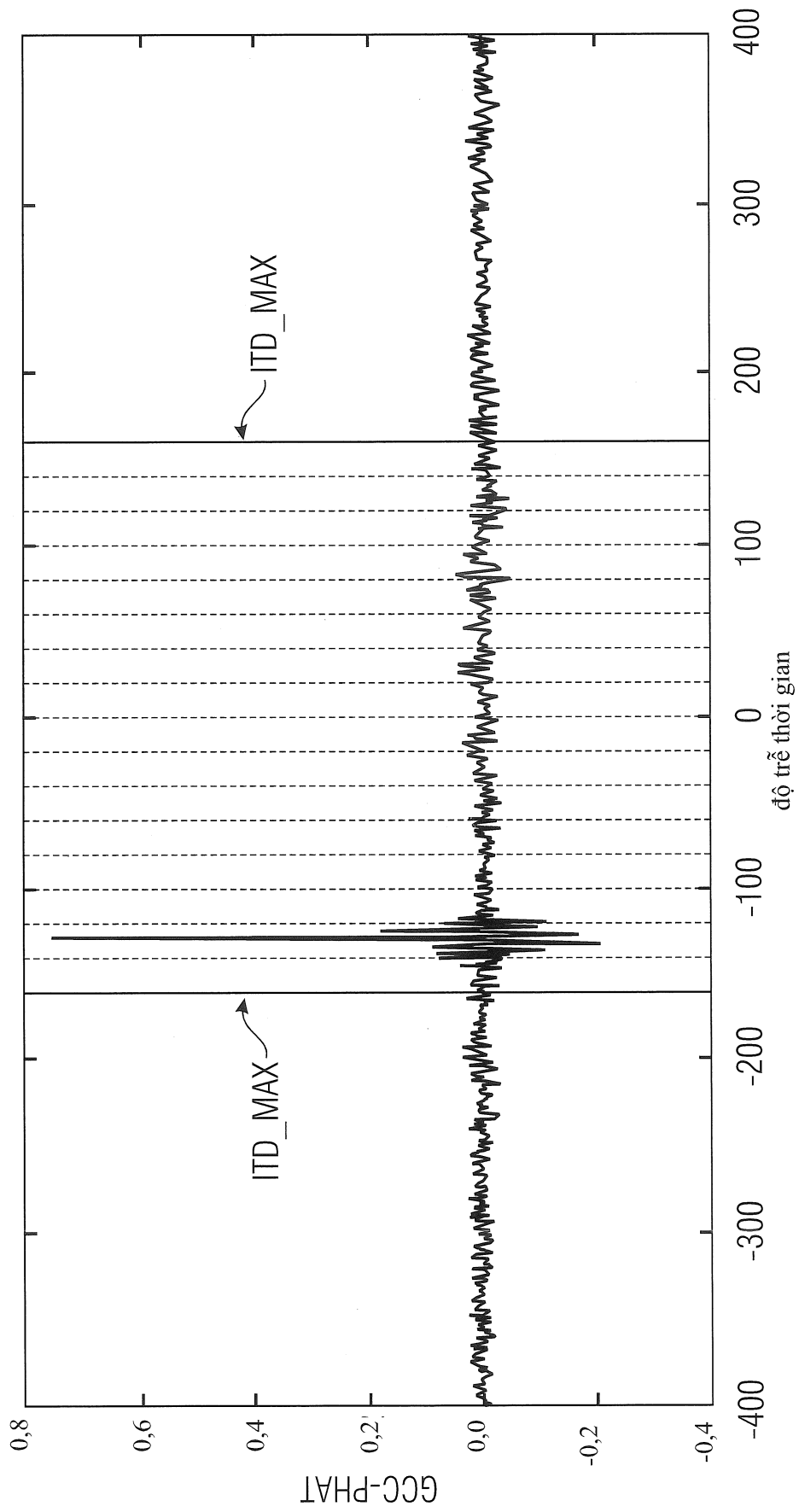


Fig. 12

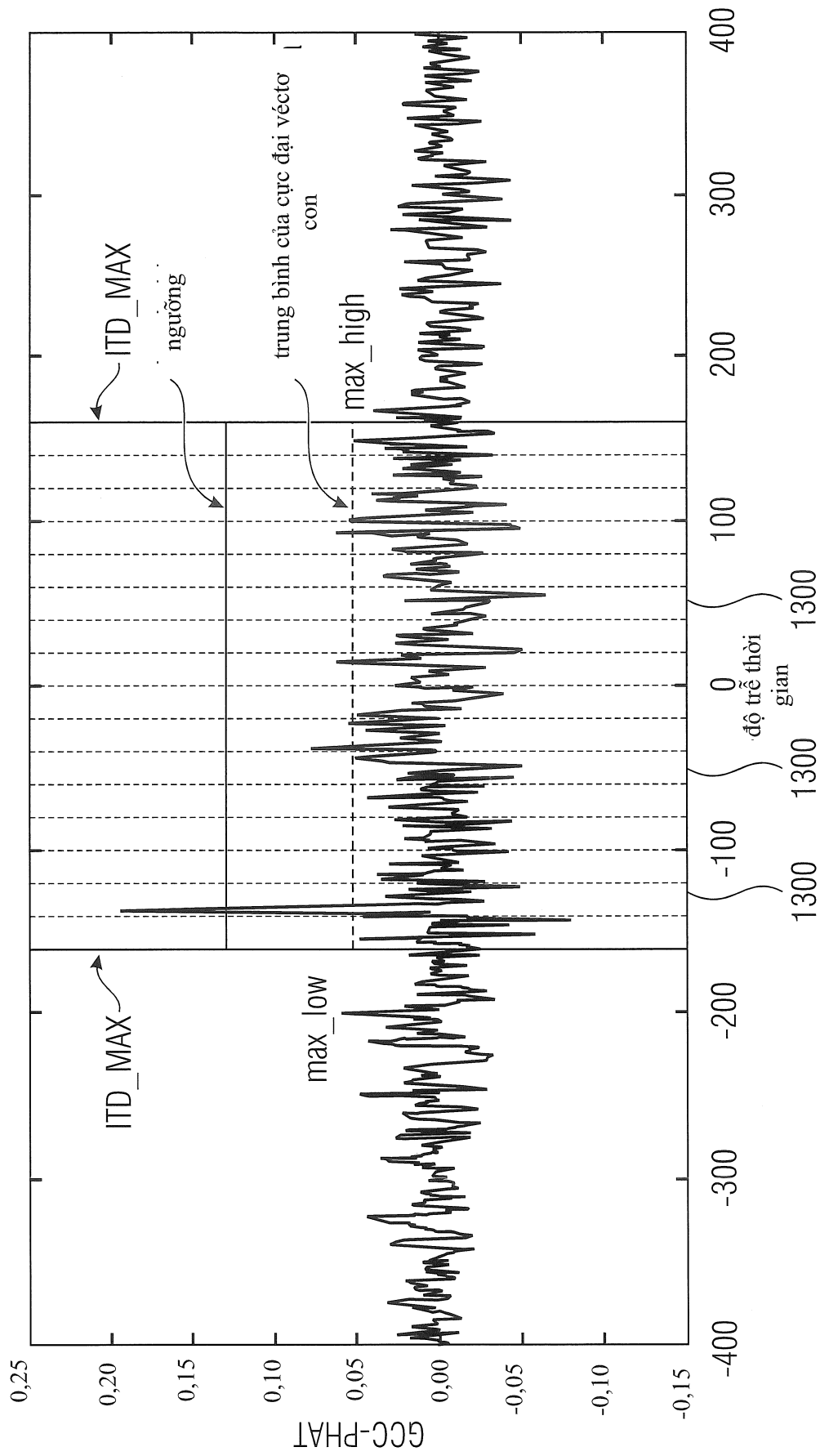


Fig. 13