



(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)



1-0049122

(51)^{2020.01} G10L 19/008; G10L 25/06

(13) B

-
- (21) 1-2020-06058 (22) 03/04/2019
(86) PCT/EP2019/058434 03/04/2019 (87) WO2019/193070 10/10/2019
(30) 18165882.4 05/04/2018 EP
(45) 25/07/2025 448 (43) 25/12/2020 393A
(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) FOTOPOULOU, Eleni (GR); BUETHE, Jan (DE); RAVELLI, Emmanuel (FR);
MABEN, Pallavi (IN); DIETZ, Martin (DE); REUTELHUBER, Franz (DE);
DOEHLA, Stefan (DE); KORSE, Srikanth (IN).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)
-

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

(21) 1-2020-06058

(57) Sáng chế đề cập đến thiết bị và phương pháp ước tính chênh lệch thời gian liên kênh. Thiết bị ước tính chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai, bao gồm bộ phân tích tín hiệu (1037) để ước tính đặc tính tín hiệu (1038) của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai hoặc cả hai tín hiệu hoặc tín hiệu được suy ra từ tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai; 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 gán trọng số (1036) để gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn để thu được phổ tương quan chéo được gán trọng số bằng cách sử dụng quy trình gán trọng số thứ nhất (1036a) hoặc sử dụng quy trình gán trọng số thứ hai (1036b) tùy thuộc vào đặc tính tín hiệu được ước tính bởi bộ phân tích tín hiệu (1037), trong đó quy trình gán trọng số thứ nhất khác với quy trình gán trọng số thứ hai; và bộ xử lý (1040) để xử lý phổ tương quan chéo được gán trọng số để thu được chênh lệch thời gian liên kênh.

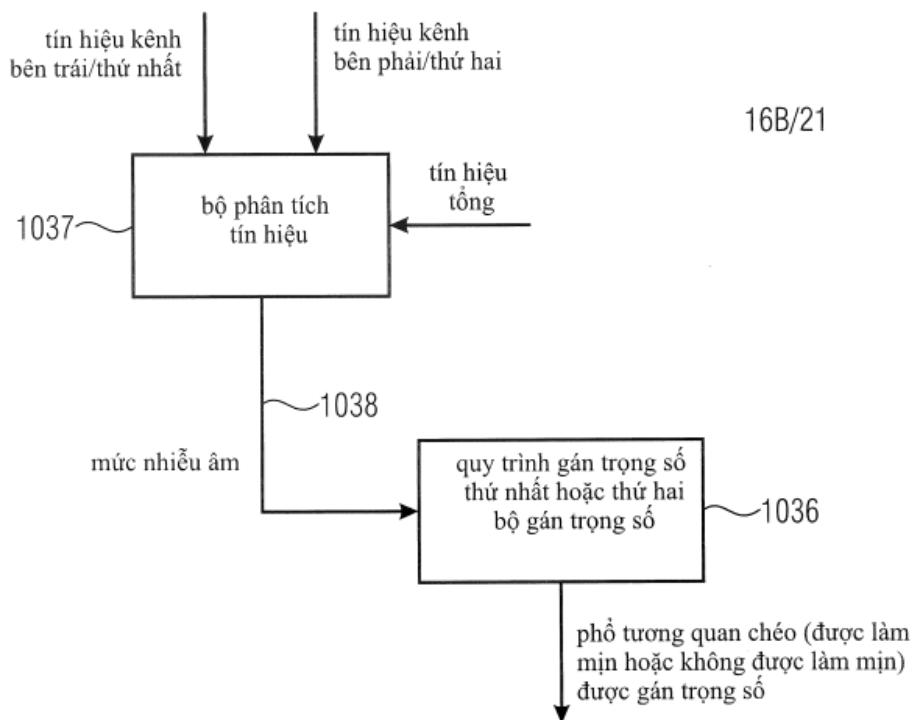


Fig. 10c

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc xử lý âm lập thể hoặc, nói chung là, 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 hoặc bất kỳ số lượng kênh nào khác.

Tình trạng kỹ thuật của sáng chế

Tiếng nói âm lập thể và đặc biệt là tiếng nói âm lập thể hội thoại ít nhận được sự quan tâm của giới khoa học hơn nhiều so với việc lưu trữ và phát thanh âm nhạc lập thể. Thật vậy, trong giao tiếp bằng tiếng nói, truyền đơn âm ngày nay vẫn được sử dụng chủ yếu. Tuy nhiên, với sự gia tăng của băng thông và dung lượng mạng, có thể hình dung rằng truyền thông dựa trên các công nghệ âm lập thể sẽ trở nên phổ biến hơn và mang lại trải nghiệm nghe tốt hơn.

Mã hóa hiệu quả tài liệu âm lập thể đã được nghiên cứu từ lâu trong việc mã hóa âm thanh cảm nhận của âm nhạc để lưu trữ hoặc phát thanh hiệu quả. Ở các tốc độ bit cao, nơi mà 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 một thời gian dài. Đối với các tốc độ bit thấp, âm lập thể cường độ và mã hóa âm lập thể tham số gần đây đã được giới thiệu. Kỹ thuật mới nhất đã được áp dụng trong các tiêu chuẩn khác nhau là HeAACv2 và Mpeg USAC. Nó tạo ra sự trộn giảm của tín hiệu hai kênh và kết hợp với thông tin phụ không gian thu gọn.

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, biến đổi thời gian – tần số của tín hiệu và sau đó không tương thích với xử lý miền thời gian và độ trễ thấp được thực hiện trong hầu hết các bộ mã hóa tiếng nói. Hơn nữa, tốc độ bit được tạo ra thường cao.

Mặt khác, âm lập thể tham số sử dụng giàn bộ lọc bổ sung được đặt ở đầu trước của bộ mã hóa như bộ xử lý trước và ở đầu cuối của bộ giải mã như bộ xử lý sau. Do đó, âm lập thể tham số có thể được sử dụng với các bộ mã hóa tiếng nói thông thường

như ACELP vì nó được thực hiện trong MPEG USAC. Hơn nữa, việc tham số hóa cảnh thính giác có thể đạt được với lượng thông tin phụ tối thiểu, phù hợp với các tốc độ bit thấp. Tuy nhiên, âm lập thể tham số như ví dụ trong MPEG USAC không được thiết kế đặc biệt cho độ trễ thấp và không mang lại chất lượng nhất quán cho các kịch bản hội thoại khác nhau. Trong biểu diễn tham số thông thường của cảnh không gian, chiều rộng của hình ảnh âm lập thể được tái tạo nhân tạo bởi bộ giải tương quan áp dụng trên hai kênh tổng hợp và được điều khiển bởi các tham số nhất quán liên kênh (Inter-channel Coherence - ICs) được tính toán và truyền đi bởi bộ mã hóa. Đối với hầu hết các tiếng nói âm lập thể, cách mở rộng hình ảnh âm lập thể này không thích hợp để tái tạo bầu không khí tự nhiên của tiếng nói vốn là âm thanh khá trực tiếp vì nó được tạo ra bởi nguồn đơn đặt tại vị trí cụ thể trong không gian (đôi khi có một số âm vang từ căn phòng). Ngược lại, nhạc cụ có chiều rộng tự nhiên hơn nhiều so với tiếng nói, có thể được bắt chước tốt hơn bằng cách giải tương quan các kênh.

Các vấn đề cũng xảy ra khi tiếng nói được ghi âm với các micrô không trùng nhau, như trong cấu hình A-B khi các micrô ở xa nhau hoặc để ghi âm hoặc kết xuất lập thể. Các kịch bản đó có thể được hình dung để ghi lại tiếng nói trong các hội nghị từ xa hoặc để tạo ra cảnh thính giác ảo với những loa ở xa trong thiết bị điều khiển đa điểm (multipoint control unit - MCU). Thời gian đến của tín hiệu sau đó sẽ khác nhau giữa kênh này với kênh khác không giống như các bản ghi được thực hiện trên các micrô trùng nhau như X-Y (ghi cường độ) hoặc M-S (ghi giữa-bên). Sau đó, việc tính toán sự nhất quán của hai kênh không được căn chỉnh theo thời gian như vậy có thể bị ước tính sai, điều này làm cho quá trình tổng hợp môi trường nhân tạo không thành công.

Các tài liệu tham khảo của tình trạng kỹ thuật trước đó liên quan đến xử lý âm lập thể là bằng sáng chế Mỹ số 5,434,948 hoặc bằng sáng chế Mỹ số 8,811,621.

Tài liệu WO 2006/089570 A1 bộc lộ sơ đồ bộ mã hóa/bộ giải mã đa kênh gần như trong suốt hoặc trong suốt. Sơ đồ bộ mã hóa/bộ giải mã đa kênh cũng tạo bổ sung tín hiệu dư dạng sóng. Tín hiệu dư này được truyền cùng với một hoặc nhiều tham số đa kênh đến bộ giải mã. Trái ngược với bộ giải mã đa kênh tham số thuần túy, bộ giải mã nâng cao tạo ra tín hiệu đầu ra đa kênh có chất lượng đầu ra được cải thiện do có thêm tín hiệu dư bổ sung. Ở phía bộ mã hóa, kênh bên trái và kênh bên phải đều được

lọc bởi giàn bộ lọc phân tích. Sau đó, đối với mỗi tín hiệu băng con, giá trị căn chỉnh và giá trị độ khuếch đại được tính cho băng con. Việc căn chỉnh như vậy sau đó sẽ được thực hiện trước khi xử lý tiếp. Ở phía bộ giải mã, quá trình khử căn chỉnh và xử lý độ khuếch đại được thực hiện và các tín hiệu tương ứng sau đó được tổng hợp bởi giàn bộ 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 như vậy, việc tính toán chênh lệch thời gian liên kênh hoặc giữa các kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai là hữu ích để thường thực hiện quy trình căn chỉnh theo thời gian băng rộng. Tuy nhiên, các ứng dụng khác vẫn tồn tại để sử dụng chênh lệch thời gian liên kênh giữa kênh thứ nhất và kênh thứ hai, trong đó các ứng dụng này là ở trong lưu trữ hoặc truyền dữ liệu tham số, xử lý âm lập thể/đa kênh bao gồm căn chỉnh theo thời của gian hai kênh, chênh lệch thời gian của ước tính đến để xác định vị trí của loa trong phòng, lọc không gian điều hướng chùm sóng (beamforming), phân tách mặt nổi/nền hoặc vị trí của nguồn âm thanh, ví dụ, bởi phép đặc tam giác âm để chỉ đặt tên một số.

Đối với tất cả các ứng dụng như vậy, sự xác định hiệu quả, mạnh và chính xác chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và thứ hai là cần thiết.

Hiện đã tồn tại những sự xác định như vậy được biết đến dưới thuật ngữ “GCC-PHAT” hoặc, được phát biểu khác đi, là biến đổi pha tương quan chéo tổng quát. Thông thường, phổ tương quan chéo được tính toán giữa hai tín hiệu kênh và sau đó, hàm gan trọng số được áp dụng cho phổ tương quan chéo để thu được cái gọi là phổ tương quan chéo tổng quát trước khi thực hiện sự biến đổi phổ nghịch đảo, chẳng hạn như DFT nghịch đảo đến phổ tương quan chéo tổng quát để tìm biểu diễn miền thời gian. Biểu diễn miền thời gian này biểu thị các giá trị cho độ trễ thời gian nhất định và đỉnh cao nhất của biểu diễn miền thời gian khi đó 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 thể hiện rằng, đặc biệt là trong các tín hiệu mà khác với, ví dụ, tiếng nói sạch mà không có bất kỳ tiếng vang hoặc nhiễu âm xung quanh nào, thì độ mạnh của kỹ thuật chung này không phải là tối ưu.

Tài liệu tham khảo

[1] Patent application. "Apparatus and Method for Estimating an Inter-Channel Time Difference." International Application Number PCT/EP2017/051214.

[2] Knapp, Charles, and Glifford Carter. "The generalized correlation method for estimation of time delay." IEEE Transactions on Acoustics, Speech, and Signal Processing 24.4 (1976): 320-327.

[3] Zhang, Cha, Dinei Florêncio, and Zhengyou Zhang. "Why does PHAT work well in low-noise, reverberative environments?" Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008.

[4] Rabinkin, Daniel V., et al. "DSP implementation of source location using microphone ar-rays." Advanced signal processing algorithms, architectures, and implementations VI. Vol. 2846. International Society for Optics and Photonics, 1996.

[5] Shen, Miao, and Hong Liu. "A modified cross power-spectrum phase method based on microphone array for acoustic source localization." Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on. IEEE, 2009.

[6] 3GPP TS 26.445; Codec for Enhanced Voice Services (EVS); Detailed algorithmic de-scription.

Bản chất kỹ thuật của sáng chế

Do đó, mục đích của sáng chế là cung cấp khái niệm cải tiến để ước tính chênh lệch thời gian liên kênh giữa hai tín hiệu kênh.

Mục đích này đạt được nhờ thiết bị ước tính chênh lệch thời gian liên kênh theo điểm 1 yêu cầu bảo hộ, hoặc phương pháp ước tính chênh lệch thời gian liên kênh theo điểm 28 yêu cầu bảo hộ hoặc chương trình máy tính theo điểm 30 yêu cầu bảo hộ.

Sáng chế dựa trên phát hiện ra rằng việc gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn để thu được phổ tương quan chéo được gán trọng số phải được thực hiện bằng cách sử dụng quy trình gán trọng số thứ nhất hoặc sử dụng quy trình gán trọng số thứ hai tùy thuộc vào đặc tính tín hiệu được ước tính bởi bộ phân tích tín hiệu, trong đó quy trình gán trọng số thứ nhất khác với các quy trình gán trọng số thứ hai.

Theo phương án khác, việc làm mịn phổ tương quan chéo theo thời gian 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 mạnh và độ chính xác khi xác định chênh lệch thời gian liên kênh.

Theo các phương án được ưu tiên, đặc tính âm sắc/nhiều âm của phổ được xác định, và trong trường hợp tín hiệu giống âm sắc, độ làm mịn mạnh hơn trong khi, trong trường hợp tín hiệu nhiều âm, độ làm mịn được thực hiện ít mạnh hơn.

Tốt hơn là, số đo độ phẳng phổ được sử dụng, và trong trường hợp các tín hiệu giống âm sắc, số đo độ phẳng phổ sẽ thấp và độ mịn sẽ trở nên mạnh hơn, và trong trường hợp tín hiệu giống như nhiều âm, số đo độ phẳng phổ sẽ cao, chẳng hạn như khoảng 1 hoặc gần bằng 1 và độ mịn sẽ yếu.

Do đó, theo sáng chế, thiết bị để ước tính 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 cho 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 tính đặc tính phổ để ước tính đặc tính phổ của tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai cho khối thời gian và ngoài ra, bộ lọc làm mịn để làm mịn phổ tương quan chéo theo thời gian bằng cách sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm mịn. Sau đó, phổ tương quan chéo được làm mịn được xử lý tiếp 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 mịn, hoạt động tạo ngưỡng thích ứng được thực hiện, trong đó biểu diễn miền thời gian của phổ tương quan chéo tổng quát được làm mịn được phân tích để xác định ngưỡng biến thiên, điều đó phụ thuộc vào biểu diễn miền thời gian và đỉnh của biểu diễn miền thời gian được so sánh với ngưỡng biến thiên, 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 xác định trước với ngưỡng chẳng hạn như lớn hơn ngưỡng.

Theo một phương án, ngưỡng biến thiên được xác định là giá trị bằng bội số nguyên của giá trị trong số giá trị lớn nhất, ví dụ mười phần trăm giá trị của biểu diễn miền thời gian hoặc, thay vào đó, theo phương án khác cho sự xác định biến thiên,

ngưỡng biến thiên được tính bằng phép nhân của ngưỡng biến thiên và giá trị, trong đó giá trị phụ thuộc vào đặc tính tỷ lệ tín hiệu trên nhiễu âm của các tín hiệu kênh thứ nhất và thứ hai, trong đó giá trị trở nên cao hơn đối với tỷ lệ tín hiệu trên nhiễu âm cao hơn và trở nên thấp hơn đối với tỷ lệ tín hiệu trên nhiễu âm thấp hơn.

Như đã nêu trước đó, phép tính chênh lệch thời gian liên kênh có thể được sử dụng trong nhiều ứng dụng khác nhau như lưu trữ hoặc truyền dữ liệu tham số, xử lý/mã hóa âm lập thể/đa kênh, căn chỉnh theo thời gian của hai kênh, chênh lệch thời gian của ước tính đến để xác định vị trí của loa trong phòng có hai micro và thiết lập micro đã biết, nhằm mục đích điều hướng chùm sóng, lọc không gian, phân tách mặt nổi/nền hoặc xác định vị trí của nguồn âm thanh, ví dụ bằng phép đặc tam giác âm dựa trên các chênh lệch thời gian của hai hoặc ba tín hiệu.

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

Thiết bị để mã hóa tín hiệu đa kênh có ít nhất hai kênh bao gồm bộ xác định tham số một mặt để xác định tham số căn chỉnh bằng rộng và mặt khác xác định nhiều tham số căn chỉnh bằng hẹp. Các tham số này được sử dụng bởi bộ căn chỉnh tín hiệu để căn chỉnh ít nhất hai kênh bằng cách 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 bằng cách sử dụng các kênh được căn chỉnh và tín hiệu giữa và tín hiệu bên sau đó được mã hóa và chuyển tiếp thành tín hiệu đầu ra được mã hóa, ngoài ra còn có, như thông tin phụ tham số, tham số căn chỉnh bằng rộng và nhiều tham số căn chỉnh bằng hẹp.

Ở phía bộ giải mã, bộ giải mã tín hiệu sẽ giải mã tín hiệu giữa được mã hóa và tín hiệu bên được mã hóa để thu được các tín hiệu giữa và 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 này sau đó được khử căn chỉnh bằng cách sử dụng thông tin về tham số căn chỉnh bằng rộng và thông tin về nhiều tham số bằng hẹp được bao gồm trong tín hiệu đa kênh được mã hóa để thu được tín hiệu đa kênh được giải mã.

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

Sáng chế dựa trên phát hiện đặc biệt cho các tín hiệu tiếng nói trong đó có nhiều hơn một loa, nhưng cũng cho các tín hiệu âm thanh khác trong đó có một số nguồn âm thanh, các vị trí khác nhau của các nguồn âm thanh mà cả hai đều ánh xạ thành hai kênh của tín hiệu đa kênh có thể được tính bằng cách 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 cả hai kênh. Ngoài tham số căn chỉnh băng rộng này, các tác giả sáng chế còn phát hiện ra rằng một số tham số căn chỉnh băng hẹp mà khác từ băng con này đến băng con khác cũng dẫn đến sự căn chỉnh tốt hơn của tín hiệu trong cả hai kênh.

Do đó, căn chỉnh băng rộng tương ứng với cùng độ trễ thời gian trong mỗi băng con cùng với căn chỉnh theo pha tương ứng với các chuyển động xoay pha khác nhau cho 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 biểu diễn giữa/bên mà sau đó được tiếp tục mã hóa. Do thực tế là đã đạt được sự căn chỉnh tối ưu, nên một mặt năng lượng trong tín hiệu giữa càng lớn càng tốt và mặt khác năng lượng trong tín hiệu bên càng nhỏ càng tốt sao cho có thể thu được kết quả mã hóa tối ưu với tốc độ bit thấp nhất có thể hoặc chất lượng âm thanh cao nhất có thể cho một tốc độ bit nhất định.

Cụ thể đối với tài liệu tiếng nói hội thoại, có vẻ như thường có loa đang hoạt động ở hai địa điểm khác nhau. Ngoài ra, tình huống như vậy, thông thường, chỉ có một loa đang nói từ địa điểm thứ nhất và sau đó loa thứ hai đang nói từ địa điểm hoặc vị trí thứ hai. Ảnh hưởng của các vị trí khác nhau trên hai kênh như kênh thứ nhất hoặc bên trái và kênh thứ hai hoặc bên phải được phản ánh bởi các thời gian đến khác nhau và do đó, độ trễ thời gian nhất định giữa cả hai kênh do các vị trí khác nhau, và độ trễ thời gian này luôn thay đổi theo thời gian. Nói chung, ảnh hưởng này được phản ánh trong hai tín hiệu kênh như sự giải căn chỉnh băng rộng có thể được giải quyết bởi tham số căn chỉnh băng rộng.

Mặt khác, các hiệu quả khác, đặc biệt đến từ tiếng vang hoặc các nguồn nhiễu âm khác có thể được tính bởi các tham số căn chỉnh theo pha riêng lẻ đối với các băng

riêng lẻ được chồng lên trên các thời gian đến khác nhau bằng rộng hoặc giải căn chỉnh bằng rộng của cả hai kênh.

Theo quan điểm đó, việc sử dụng cả hai, tham số căn chỉnh bằng rộng và nhiều tham số căn chỉnh bằng hẹp ở trên cùng của tham số căn chỉnh bằng rộng dẫn đến việc căn chỉnh kênh tối ưu ở phía bộ mã hóa để thu được sự biểu diễn giữa/bên tốt và rất nhỏ gọn trong khi, mặt khác, việc giải căn chỉnh tương ứng sau quá trình giải mã ở phía bộ giải mã dẫn đến chất lượng âm thanh tốt cho tốc độ bit nhất định hoặc tốc độ bit nhỏ cho chất lượng âm thanh bắt buộc nhất định.

Ưu điểm của sáng chế này là cung cấp sơ đồ mã hóa âm lập thể mới phù hợp hơn nhiều cho việc chuyển đổi tiếng nói âm lập thể so với các sơ đồ mã hóa âm lập thể hiện có. Theo sáng chế, các công nghệ âm lập thể tham số và các công nghệ mã hóa âm lập thể kết hợp được tổ hợp đặc biệt bằng cách khai thác chênh lệch thời gian liên kênh xảy ra trong các kênh của tín hiệu đa kênh cụ thể trong trường hợp các nguồn tiếng nói và cả trong trường hợp các nguồn âm thanh khác.

Một số phương án đưa ra các ưu điểm hữu ích như được thảo luận sau đây.

Phương pháp mới là phương thức lai trộn các yếu tố từ âm lập thể M/S thông thường và âm lập thể tham số. Trong M/S thông thường, các kênh được trộn giảm một cách thụ động để tạo ra tín hiệu giữa và bên. Quá trình này có thể được mở rộng hơn nữa bằng cách xoay kênh bằng cách sử dụng biến đổi Karhunen-Loeve (KLT), còn được biết đến là Phân tích thành phần chính (Principal Component Analysis - PCA) trước khi tổng hợp và phân biệt các kênh. Tín hiệu giữa được mã hóa theo mã hóa mã sơ cấp trong khi tín hiệu bên được chuyển đến bộ mã hóa thứ cấp. Âm lập thể M/S cải tiến có thể sử dụng thêm dự báo của tín hiệu bên bởi kênh giữa được mã hóa trong khung hiện thời hoặc trước đó. Mục tiêu chính của việc xoay 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 bảo tồn dạng sóng và ở khía cạnh này rất mạnh đối với mọi tình huống âm lập thể, nhưng có thể rất tổn kém về tiêu thụ bit.

Để đạt hiệu quả cao nhất ở các tốc độ bit thấp, tính toán âm lập thể tham số và các tham số mã, như các chênh lệch mức liên kênh (Inter-channel Level difference - ILD), chênh lệch pha liên kênh (Inter-channel Phase difference - IPD), chênh lệch thời

gian liên kênh (Inter-channel Time difference - ITD) và tính nhất quán liên kênh (Inter-channel Coherence - IC). Chúng thể hiện một cách nhỏ gọn hình ảnh âm lập thể và là những dấu hiệu của cảnh thính giác (khoanh vùng nguồn, ghép nhẵn, độ rộng của âm lập thể...). Mục đích sau đó là để tham số hóa cảnh âm lập thể và chỉ mã hóa tín hiệu trộn giảm có thể có ở bộ giải mã và với sự trợ giúp của các tín hiệu âm lập thể được truyền một lần nữa lại được không gian hóa.

Cách tiếp cận của các tác giả sáng chế trộn hai khái niệm. Thứ nhất, các tín hiệu âm lập thể ITD và IPD được tính và áp dụng trên 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 pha và việc mã hóa M/S sau đó được thực hiện. ITD và IPD được cho là hữu ích cho việc lập mô hình tiếng nói âm lập thể và là sự thay thế tốt của KTL dựa trên sự xoay trong M/S. Không giống như mã hóa tham số thuần túy, môi trường không được mô hình hóa bởi các IC mà trực tiếp bởi tín hiệu bên được mã hóa và/hoặc dự báo. Thấy rằng cách tiếp cận này mạnh hơn đặc biệt là khi xử lý các tín hiệu tiếng nói.

Việc tính toán và xử lý các ITD là phần quan trọng của sáng chế. Các ITD đã được khai thác trong mã hóa tín hiệu lập thể (Binaural Cue Coding - BCC) trong tình trạng kỹ thuật trước đó, nhưng theo cách mà nó không hiệu quả khi các ITD thay đổi theo thời gian. Để tránh thiếu sót này, việc tạo cửa sổ cụ thể được thiết kế để làm mượt quá trình chuyển tiếp giữa hai ITD khác nhau và có thể chuyển đổi liên mạch từ loa này sang loa khác được đặt ở các địa điểm khác nhau.

Các phương án khác liên quan 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 bằng cách sử dụng các kênh đã được căn chỉnh với tham số căn chỉnh băng rộng được xác định trước đó.

Tương ứng, việc khử căn chỉnh băng hẹp ở phía bộ giải mã được thực hiện trước khi thực hiện khử căn chỉnh băng rộng bằng cách sử dụng tham số căn chỉnh băng rộng đơn thông thường.

Trong các phương án khác, ưu tiên rằng, ở phía bộ mã hóa nhưng thậm chí còn quan trọng hơn ở phía bộ giải mã, một số loại hoạt động tạo cửa sổ và chồng lấp-cộng

hoặc bất kỳ loại giao nhau nào từ khối này sang khối tiếp theo được thực hiện sau tất cả các căn chỉnh và cụ thể là sau căn chỉnh theo thời gian bằng cách sử dụng tham số căn chỉnh băng rộng. Điều này tránh mọi thành phần lạ có thể nghe được, chẳng hạn như các tiếng lách cách khi tham số căn chỉnh theo thời gian hoặc băng rộng thay đổi từ khối này sang khối khác.

Theo các phương án khác, các độ phân giải phổ khác nhau được áp dụng. Cụ thể, các tín hiệu kênh chịu sự chuyển đổi thời gian-phổ có độ phân giải tần số cao như phổ DFT trong khi các tham số như tham số căn chỉnh băng hẹp được xác định cho các băng tham số có độ phân giải phổ thấp hơn. 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. Hơn nữa, các băng tham số tăng từ các tần số thấp đến các tần số cao để tính đến các vấn đề âm học thần kinh.

Các phương án khác liên quan đến việc sử dụng bổ sung tham số mức như chênh lệch liên mức hoặc các quy trình khác để xử lý tín hiệu bên như các tham số lấp đầy âm lập thể, v.v.. Tín hiệu bên được mã hóa có thể được biểu diễn bởi chính tín hiệu bên thực tế, hoặc bởi tín hiệu dự báo đang được thực hiện bằng cách sử dụng tín hiệu giữa của khung hiện thời hoặc bất kỳ khung nào khác, hoặc bởi tín hiệu bên hoặc tín hiệu dự báo bên chỉ trong tập con của các băng và các tham số dự báo chỉ cho các băng còn lại, hoặc thậm chí bởi các tham số dự báo cho tất cả các băng mà không có bất kỳ thông tin tín hiệu bên độ phân giải tần số cao nào. Do đó, trong sự lựa chọn 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 tập con của các băng tham số để đối với các băng tham số còn lại không tồn tại bất kỳ thông tin nào về tín hiệu bên ban đầu.

Hơn nữa, ưu tiên có nhiều 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 50 phần trăm thấp hơn của các băng tham số. Mặt khác, các tham số lấp đầy âm lập thể không được sử dụng cho một vài băng tần thấp hơn, vì đối với các băng này, bản thân tín hiệu bên hoặc tín hiệu dự báo được truyền để đảm bảo rằng, ít nhất là đối với các băng thấp hơn, sự biểu diễn đúng dạng sóng là có sẵn. Mặt khác, tín hiệu bên không được truyền dưới dạng biểu diễn chính xác dạng

sóng cho các băng cao hơn để giảm tốc độ bit hơn nữa, nhưng tín hiệu bên thường được biểu diễn bằng các tham số lấp đầy âm lập thể.

Hơn nữa, ưu tiên thực hiện toàn bộ phân tích và căn chỉnh tham số trong một và cùng một miền tần số dựa trên cùng một phổ DFT. Để đạt được điều này, ưu tiên hơn nữa cho việc sử dụng công nghệ tương quan chéo tổng quát với biến đổi pha (generalized cross correlation with phase transform - GCC-PHAT) cho mục đích xác định chênh lệch thời gian liên kênh. Trong phương án được ưu tiên của quy trình này, việc làm mịn phổ tương quan dựa trên thông tin về hình dạng phổ, thông tin tốt hơn là số đo độ phẳng phổ được thực hiện theo cách mà việc làm mịn sẽ yếu trong trường hợp các tín hiệu giống như nhiễu âm và sự làm mịn sẽ trở nên mạnh hơn trong trường hợp tín hiệu giống như âm sắc.

Hơn nữa, ưu tiên thực hiện xoay pha cụ thể, trong đó các biên độ kênh được tính đến. Cụ thể, sự xoay pha được phân phối giữa hai kênh nhằm mục đích căn chỉnh ở phía bộ mã hóa và tất nhiên, nhằm mục đích khử căn chỉnh ở phía bộ giải mã trong đó kênh có biên độ cao hơn được coi là kênh đầu và sẽ ít bị ảnh hưởng bởi sự xoay pha, tức là, sẽ ít bị xoay hơn so với kênh có biên độ thấp hơn.

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

Mô tả vắn tắt các hình vẽ

Sau đây, các phương án được ưu tiên của sáng chế được thảo luận liên quan đến các hình vẽ kèm theo, trong đó:

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

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

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

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

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

Fig.4c minh họa phương án của các quy trình được thực hiện trong thiết bị để mã hóa bằng cách sử dụng cửa sổ phân tích với các phần đệm 0 và các phạm vi chồng lấp;

Fig.4d minh họa sơ đồ cho các quy trình tiếp theo được thực hiện trong thiết bị để mã hóa;

Fig.4e minh họa sơ đồ thể hiện sự thực hiện ước tính chênh lệch thời gian liên kênh;

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

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

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

Fig.7 minh họa kịch bản cửa sổ được ưu tiên với các cửa sổ hình sin chồng lấp thấp với việc đệm 0 để phân tích và tổng hợp thời gian – tần số âm lập thể;

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

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

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

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

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

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

Fig.10c minh họa biểu diễn sơ đồ của bộ phân tích tín hiệu được thực hiện như bộ ước tính nhiễu âm trong phương án và bộ gán trọng số phù hợp với các phương án của sáng chế;

Fig.10d minh họa biểu diễn sơ đồ của bộ gán trọng số phù hợp với các phương án của sáng chế;

Fig.10e minh họa biểu diễn sơ đồ của bộ xử lý phù hợp với các phương án của sáng chế;

Fig.10f minh họa biểu diễn sơ đồ của bộ ước tính nhiễu âm phù hợp với các phương án của sáng chế;

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

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

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

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

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

Fig.11f minh họa biểu diễn sơ đồ của bộ xử lý phù hợp với các phương án của sáng chế;

Fig.12 minh họa biểu diễn miền thời gian cho phổ tương quan chéo được làm mịn để có tín hiệu tiếng nói sạch;

Fig.13 minh họa biểu diễn miền thời gian của phổ tương quan chéo được làm mịn cho 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 để ước tính 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 được minh họa bổ sung, liên quan đến Fig.4e như mục 451.

Hơn nữa, các 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 còn bao gồm bộ ước tính đặc tính phổ 1010 để ước tính đặc tính phổ của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai cho khối thời gian. Thiết bị còn bao gồm một bộ lọc làm mịn 1030 để làm mịn phổ tương quan chéo theo thời gian bằng cách sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm mịn. Thiết bị còn bao gồm một bộ xử lý 1040 để xử lý phổ tương quan được làm mịn để thu được chênh lệch thời gian liên kênh.

Ngoài ra, theo phương án khác, phần tử 1030 không có mặt, và do đó, phần tử 1010 cũng không cần thiết, như được chỉ ra bởi vạch đứt 1035. Thiết bị còn bao gồm bộ phân tích tín hiệu 1037 tính toán ước tính đặc tính tín hiệu chẳng hạn như ước tính nhiễu âm 1038. Ước tính này được chuyển tiếp đến bộ gán trọng số 1036 được tạo cấu hình để thực hiện các hoạt động gán trọng số khác nhau tùy thuộc vào ước tính đặc tính tín hiệu. Ước tính đặc tính tín hiệu tốt hơn là cũng được sử dụng để điều khiển bộ xử lý 1040, ví dụ khi bộ xử lý 1040 thực hiện phép tính chọn đỉnh. Fig.10c còn minh họa cho bộ phân tích tín hiệu 1037 và bộ gán trọng số có thể điều khiển 1036.

Cụ thể, thiết bị phù hợp với các phương án của sáng chế hướng đến việc ước tính chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai. Thiết bị này bao gồm bộ phân tích tín hiệu 1037 trên Fig.10a, bộ tính toán phổ tương quan chéo 1020 trên Fig.10a, bộ gán trọng số 1036 để gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn trên Fig.10a và bộ xử lý được kết nối sau đó 1040 để xử lý phổ tương quan chéo được gán trọng số.

Bộ chuyển đổi thời gian - phổ các phần tử 150, bộ ước tính đặc tính phổ 1010, bộ lọc làm mịn 1030 không cần thiết cho sự thực hiện cơ bản của sáng chế, nhưng được

ưu tiên cho các phương án được ưu tiên của sáng chế. Bộ phân tích tín hiệu 1037 được tạo cấu hình để ước tính đặc tính tín hiệu như mức nhiễu âm 1038 của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai hoặc cả hai tín hiệu hoặc tín hiệu thu được từ tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai. Do đó, đặc tính tín hiệu hoặc ước tính đặc tính tín hiệu chẳng hạn như ước tính nhiễu âm sẽ được sử dụng sau này bởi bộ gán trọng số 1036 và, tốt hơn là, cũng được sử dụng bởi bộ xử lý 1040 chỉ có thể suy ra từ tín hiệu kênh bên trái hoặc thứ nhất, chỉ từ tín hiệu thứ hai hoặc bên phải, hoặc có thể suy ra từ cả hai tín hiệu. Ví dụ, việc suy ra đặc tính tín hiệu từ cả hai tín hiệu có thể là việc suy ra đặc tính tín hiệu riêng lẻ của tín hiệu kênh thứ nhất, đặc tính tín hiệu riêng lẻ bổ sung từ tín hiệu kênh thứ hai hoặc bên phải và, sau đó, đặc tính tín hiệu cuối cùng 1038 sẽ, ví dụ, là trung bình hoặc trung bình được gán trọng số giữa cả hai kênh. Ở đây, ví dụ việc gán trọng số có thể được thực hiện phù hợp với biên độ sao cho các biên độ khác nhau trong, ví dụ, các khung của các kênh dẫn đến các ảnh hưởng khác nhau của ước tính nhiễu âm riêng lẻ tương ứng thành mức nhiễu âm cuối cùng 1038. Hơn nữa, tín hiệu suy ra từ tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai có thể, ví dụ, là tín hiệu kết hợp thu được bằng cách thêm tín hiệu kênh bên trái hoặc thứ nhất và tín hiệu kênh thứ hai hoặc bên phải với nhau để thu được tín hiệu được kết hợp và, sau đó, đặc tính tín hiệu 1038 được tính toán từ tín hiệu kết hợp.

Theo phương án được ưu tiên, bộ phân tích tín hiệu 1036 được thực hiện như bộ phân tích hoặc bộ ước tính nhiễu âm. Tuy nhiên, có thể thực hiện các cách phân tích tín hiệu khác, chẳng hạn như phân tích âm sắc, phát hiện hoạt động tiếng nói, phân tích nốt đệm, phân tích lập thể, phân tích tiếng nói/âm nhạc, phân tích loa giao thoa, phân tích nhạc nền, phân tích tiếng nói sạch hoặc bất kỳ phân tích tín hiệu nào khác để xác định xem tín hiệu có đặc tính thứ nhất hay đặc tính thứ hai để lựa chọn quy trình gán trọng số phù hợp.

Sự kết hợp có thể là sự kết hợp với các hệ số gán trọng số bằng nhau, tức là sự kết hợp giữa kênh bên trái mà không có bất kỳ sự gán trọng số nào và kênh bên phải không có bất kỳ sự gán trọng số nào, tương ứng với các hệ số gán trọng số bằng 1,0 hoặc, thay vào đó, có thể áp dụng các hệ số gán trọng số khác nhau. Hơn nữa, tín hiệu được suy ra từ kênh thứ nhất hoặc tín hiệu được suy ra từ kênh thứ hai có thể thu được bằng cách thực hiện lọc thông cao hoặc lọc thông thấp hoặc có thể thu được bằng cách

thực hiện xử lý bằng cách sử dụng sự nén biên độ hoặc chức năng nén nghịch đảo biên độ. Hàm nén biên độ sẽ là hàm log hoặc hàm có giá trị lũy thừa nhỏ hơn 1. Hàm nén nghịch đảo sẽ là hàm mũ hoặc phép toán lũy thừa với số mũ lớn hơn 1. Do đó, tùy thuộc vào các cách thực hiện nhất định, các hoạt động xử lý khác nhau có thể được áp dụng cho các tín hiệu kênh bên trái và bên phải khác nhau và cả hai kênh có thể được kết hợp hoặc không. Trong phương án được ưu tiên, tốt hơn là cộng các kênh bên trái và bên phải lại với nhau ngay cả khi không có bất kỳ sự gán trọng số cụ thể nào và ước tính đặc tính tín hiệu sau đó được tính toán từ kết quả của sự tính toán kết hợp.

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 có thể được thực hiện theo một số cách. Một cách là tương quan chéo được tính toán từ các tín hiệu miền thời gian trong các khung miền thời gian và kết quả sau đó được chuyển đổi từ miền thời gian sang miền phổ. Cách thực hiện khác là, ví dụ, bằng cách sử dụng DFT hoặc bất kỳ chuyển đổi thời gian thành phổ nào khác, các khung tiếp theo của tín hiệu kênh thứ nhất và các khung tiếp theo của tín hiệu kênh thứ hai được chuyển đổi thành biểu diễn phổ trong đó các khung tiếp theo có thể chồng lấp hoặc không chồng lấp. Do đó, đối với mỗi khối thời gian của tín hiệu kênh thứ nhất, sự biểu diễn phổ được thu và, tương ứng, đối với mỗi khối thời gian của tín hiệu kênh thứ hai, sự biểu diễn phổ được thu. Sự tính toán tương quan chéo được thực hiện bằng cách nhân giá trị phổ của ngân tần số nhất định k và khối thời gian nhất định hoặc chỉ số mẫu thời gian s với giá trị phức liên hợp của giá trị phổ có cùng chỉ số k và cùng chỉ số s từ sự biểu diễn phổ của cùng khối thời gian của kênh thứ hai. Cũng có thể sử dụng các quy trình tính toán tương quan chéo khác với quy trình được mô tả ở trên để tính toán phổ tương quan chéo cho khối thời gian.

Bộ gán trọng số 1036 được tạo cấu hình để gán trọng số phổ tương quan chéo mà bộ tính toán thu được. Trong việc thực hiện, phổ tương quan chéo là phổ tương quan chéo không được làm mịn, nhưng theo các phương án khác, phổ tương quan chéo được làm mịn trong đó việc làm mịn này là làm mịn theo thời gian. Do đó, với mục đích tính toán phổ tương quan chéo được làm mịn, phổ tương quan chéo của khối cuối cùng có thể được sử dụng cùng với phổ tương quan chéo (thô) của khối hiện thời, và tùy thuộc vào sự thực hiện, thông tin điều khiển làm mịn có thể được sử dụng như, ví dụ, được

cung cấp bởi bộ ước tính đặc tính phổ 1010 trên Fig.10a. Tuy nhiên, việc làm mịn cũng có thể được thực hiện bằng cách sử dụng cài đặt làm mịn được xác định trước, tức là không đổi hoặc bất biến theo thời gian. Theo các phương án của sáng chế, phổ tương quan chéo được gán trọng số được tính bằng cách sử dụng quy trình gán trọng số thứ nhất 1036a hoặc sử dụng quy trình được gán trọng số thứ hai 1036b mà, ví dụ, được minh họa trên Fig.10d. Cụ thể, việc lựa chọn, xem phổ tương quan chéo được gán trọng số được suy ra bằng cách sử dụng quy trình thứ nhất hay thứ hai được thực hiện tùy thuộc vào đặc tính tín hiệu được ước tính bởi bộ phân tích tín hiệu 1037. Do đó, theo sáng chế, việc gán trọng số với đặc tính gán trọng số thứ nhất được sử dụng cho đặc tính tín hiệu nhất định của kênh thứ nhất hoặc kênh thứ hai hoặc tín hiệu kết hợp trong khi quy trình gán trọng số thứ hai được áp dụng tùy thuộc vào đặc tính tín hiệu khác như được xác định bởi bộ phân tích tín hiệu 1037. Kết quả của bộ gán trọng số 1036 là phổ tương quan chéo được gán trọng số và được làm mịn hoặc không được làm mịn mà sau đó còn được xử lý bộ xử lý 1040 để thu được chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai.

Fig.10d minh họa sự thực hiện của bộ phân tích tín hiệu như bộ ước tính nhiễu âm và bộ gán trọng số khi kết nối với bộ xử lý 1040 theo phương án của sáng chế. Cụ thể, bộ ước tính nhiễu âm 1037 bao gồm bộ tính toán ước tính nhiễu âm 1037a và bộ phân loại ước tính nhiễu âm 1037b. Bộ phân loại ước tính nhiễu âm 1037b xuất ra tín hiệu điều khiển 1050 tương ứng với đầu ra ước tính nhiễu âm 1038 được tạo bởi khối 1037 trên Fig.10a. Tín hiệu điều khiển này có thể được áp dụng cho chuyển đổi thứ nhất 1036c hoặc chuyển đổi thứ hai 1036d. Trong sự thực hiện này, các nhân xử lý 1036a thực hiện quy trình gán trọng số thứ nhất và nhân tính toán khác để thực hiện quy trình gán trọng số thứ hai 1036b được cung cấp. Tùy thuộc vào sự thực hiện, chỉ chuyển đổi 1036c được cung cấp và, tùy thuộc vào tín hiệu điều khiển 1050, chỉ quy trình gán trọng số như được xác định bởi chuyển đổi 1036c được chọn, tức là phổ tương quan chéo được xác định bởi bộ tính toán 1020 được đưa vào chuyển đổi 1036c và tùy thuộc vào cài đặt chuyển đổi, chuyển tiếp đến nhân 1036a hoặc nhân 1036b. Trong cách thực hiện khác, chuyển đổi 1036c không ở đó bởi phổ tương quan chéo như được xác định bởi khối 1020 được đưa vào cả hai nhân xử lý 1036a và 1036b và, tùy thuộc vào điều khiển của chuyển đổi đầu ra 1036d, đầu ra của khối 1036a hoặc đầu ra của khối 1036b được chọn

và chuyển tiếp đến bộ xử lý 1040. Do đó, tùy thuộc vào sự thực hiện, chỉ phổ tương quan chéo được gán trọng số đơn được tính toán trong đó việc lựa chọn khối nào được tính toán được thực hiện bởi tín hiệu điều khiển 1050 và chuyển đổi đầu vào. Ngoài ra, cả hai phổ tương quan chéo được gán trọng số đều được tính toán và chỉ phổ tương quan chéo được chọn bởi chuyển đổi đầu ra 1036d mới được chuyển tiếp tới bộ xử lý 1040. Hơn nữa, chỉ có thể có nhân xử lý đơn mà không có bất kỳ chuyển đổi đầu vào/đầu ra nào và tùy thuộc vào tín hiệu điều khiển, quy trình gán trọng số chính xác được thiết lập cho khối thời gian tương ứng. Do đó, đối với mỗi khối thời gian, ước tính nhiễu âm hoặc tín hiệu điều khiển 1050 có thể được tính toán và đối với mỗi khối thời gian, việc gán trọng số có thể được chuyển đổi từ quy trình gán trọng số này sang quy trình gán trọng số khác. Trong ngữ cảnh này, cần lưu ý rằng cũng có thể thực hiện ba hoặc nhiều hơn ba quy trình gán trọng số khác nhau tùy thuộc vào ba hoặc nhiều hơn ba ước tính nhiễu âm khác nhau tùy từng trường hợp. Do đó, sáng chế không chỉ đề cập đến việc lựa chọn giữa hai quy trình gán trọng số khác nhau, mà còn bao gồm việc lựa chọn giữa ba hoặc nhiều hơn ba quy trình gán trọng số tùy thuộc vào tín hiệu điều khiển được suy ra từ đặc tính nhiễu âm của các tín hiệu kênh thứ nhất và thứ hai.

Trong cách thực hiện được ưu tiên, quy trình gán trọng số thứ nhất bao gồm việc gán trọng số sao cho biên độ được chuẩn hóa và pha được duy trì và quy trình gán trọng số thứ hai bao gồm hệ số gán trọng số được suy ra từ phổ tương quan chéo được làm mịn hoặc không được làm mịn bằng cách sử dụng hàm lũy thừa có lũy thừa thấp hơn 1 hoặc lớn hơn 0. Hơn nữa, quy trình gán trọng số thứ nhất có thể giống hầu hết với quy trình gán trọng số thứ hai ngoại trừ việc quy trình gán trọng số thứ hai sử dụng lũy thừa từ 0 đến 1, tức là lũy thừa lớn hơn 0 và nhỏ hơn 1, trong khi quy trình gán trọng số thứ nhất không áp dụng bất kỳ lũy thừa nào hoặc, nói cách khác, áp dụng lũy thừa 1. Do đó, việc chuẩn hóa được thực hiện bởi quy trình gán trọng số thứ hai được nén, tức là, hệ số chuẩn hóa đó được áp dụng bởi quy trình gán trọng số thứ nhất có một số giá trị và hệ số chuẩn hóa được áp dụng thông qua quy trình gán trọng số thứ hai cho cùng một giá trị tương quan chéo phổ có cường độ nhỏ hơn. Điều này áp dụng cho các giá trị phổ cao hơn của phổ tương quan chéo. Tuy nhiên, đối với các giá trị phổ tương quan chéo nhỏ, giá trị chuẩn hóa của quy trình gán trọng số thứ hai lớn hơn giá trị chuẩn hóa của quy trình gán trọng số thứ nhất đối với cùng một giá trị phổ của phổ tương quan chéo. Điều

này là do thực tế là phép toán lũy thừa có lũy thừa thấp hơn 1 chẳng hạn như phép toán căn bậc hai có lũy thừa $1/2$ làm tăng giá trị nhỏ nhưng làm giảm giá trị lớn. Do đó, các tính toán hệ số gán trọng số bổ sung cho quy trình gán trọng số thứ hai cũng có thể bao gồm bất kỳ hàm nén nào chẳng hạn như hàm log. Theo phương án được ưu tiên, quy trình gán trọng số thứ nhất hoạt động dựa trên việc gán trọng số được áp dụng cho biến đổi pha (phase transform - PHAT) và quy trình gán trọng số thứ hai hoạt động dựa trên các tính toán được áp dụng cho quy trình pha phổ lũy thừa chéo được cải biên (modified cross-power spectrum phase procedure - MCSP).

Hơn nữa, quy trình gán trọng số thứ hai tốt hơn là được thực hiện để bao gồm sự chuẩn hóa sao cho phạm vi đầu ra của quy trình chuẩn hóa thứ hai nằm trong phạm vi, trong đó phạm vi đầu ra của quy trình chuẩn hóa thứ nhất được định vị hoặc sao cho phạm vi đầu ra của quy trình chuẩn hóa thứ hai giống với phạm vi đầu ra của quy trình chuẩn hóa thứ nhất. Ví dụ, điều này có thể được thực hiện bằng cách tính toán các giá trị tuyệt đối của tất cả các giá trị phổ của phổ tương quan chéo được gán trọng số MCSP, bằng cách cộng tất cả các cường độ của một biểu diễn phổ tương ứng với một khối thời gian và sau đó chia kết quả cho số lượng các giá trị phổ trong khối thời gian.

Nói chung, bộ xử lý 1040 trên Fig.10a được tạo cấu hình để thực hiện một số bước xử lý liên quan đến phổ tương quan chéo được gán trọng số, trong đó, cụ thể, phép tính chọn đỉnh nhất định được thực hiện để cuối cùng thu được chênh lệch thời gian liên kênh. Tốt hơn là phép tính chọn đỉnh này diễn ra trong miền thời gian, tức là phổ tương quan chéo được gán trọng số và được làm mịn hoặc không được làm mịn được chuyển đổi từ biểu diễn phổ trong biểu diễn miền thời gian và sau đó, biểu diễn miền thời gian này được phân tích và, cụ thể, một đỉnh hoặc một số đỉnh được chọn dựa trên ngưỡng. Tùy thuộc vào việc cài đặt ước tính nhiễu âm, phép tính chọn đỉnh thứ nhất hoặc phép tính chọn đỉnh thứ hai được thực hiện, trong đó, tốt hơn là cả hai phép tính chọn đỉnh khác nhau về ngưỡng được sử dụng bởi phép tính chọn đỉnh.

Fig.10e minh họa tình huống tương tự, đối với chuyển đổi đầu vào 1040 và chuyển đổi đầu ra 1043, theo quy trình trên Fig.10d. Trong cách thực hiện được minh họa trên Fig.10e, cả hai phép tính chọn đỉnh đều có thể được áp dụng và kết quả của phép tính chọn đỉnh “đúng” có thể được chọn bởi chuyển đổi đầu ra 1043. Ngoài ra,

chuyển đổi đầu vào ở đó và tùy thuộc vào tín hiệu điều khiển 1050, chỉ có quy trình chọn đỉnh chính xác được chọn, tức là 1041 hoặc 1042. Do đó, trong sự thực hiện, sẽ không có cả hai chuyển đổi, nhưng trong sự thực hiện sẽ có chuyển đổi đầu vào 1040 hoặc chuyển đổi đầu ra 1043 tương tự như những gì đã được suy ra trước đó đối với Fig.10d. Trong cách thực hiện bổ sung, chỉ tồn tại nhân xử lý đơn áp dụng phép tính chọn đỉnh với ngưỡng biến thiên và tín hiệu điều khiển 1050 được sử dụng để đặt ngưỡng chính xác trong nhân xử lý đơn. Theo phương án được ưu tiên, việc cài đặt ngưỡng được thực hiện theo cách sao cho ngưỡng thứ hai cao hơn ngưỡng thứ nhất, trong đó ngưỡng thứ hai, do đó, được sử dụng khi quy trình gán trọng số thứ hai trong khối 1036b đã được áp dụng và trong đó ngưỡng thứ nhất ngưỡng được sử dụng, khi quy trình gán trọng số thứ nhất trong khối 1036a đã được áp dụng. Do đó, khi phát hiện mức nhiễu âm nền cao, thì quy trình gán trọng số thứ hai với lũy thừa từ 0 đến 1 hoặc phép tính log, tức là quy trình nén được áp dụng và khi đó, ngưỡng chọn đỉnh nên thấp hơn so với ngưỡng chọn đỉnh sẽ được sử dụng khi phát hiện mức nhiễu âm nền thấp, tức là khi quy trình gán trọng số thứ nhất được áp dụng mà thực hiện chuẩn hóa với hệ số chuẩn hóa không dựa vào hàm nén chẳng hạn như hàm log hoặc hàm lũy thừa với lũy thừa nhỏ hơn 1.

Sau đó, cách thực hiện được ưu tiên của bộ phân tích tín hiệu như bộ ước tính nhiễu âm 1037 được minh họa trên Fig.10f. Về cơ bản, bộ ước tính nhiễu âm 1037 bao gồm bộ tính toán ước tính nhiễu âm 1037a và một bộ phân loại ước tính nhiễu âm 1037b như được minh họa trên Fig.10d và cũng được chỉ ra trên Fig.10f. Bộ tính toán ước tính nhiễu âm 1037a bao gồm bộ ước tính nhiễu âm nền 1060 và sau đó là bộ làm mịn (thời gian) được kết nối 1061 mà, ví dụ, có thể được thực hiện như bộ lọc IIR.

Đầu vào vào bộ tính toán ước tính nhiễu âm 1037a hoặc, cụ thể là, bộ ước tính nhiễu âm nền 1060 là khung của tín hiệu kênh bên trái hoặc thứ nhất, khung của tín hiệu kênh thứ hai hoặc bên phải hoặc tín hiệu được suy ra từ tín hiệu kênh đó hoặc tín hiệu kết hợp thu được bằng cách thêm, ví dụ, biểu diễn miền thời gian của tín hiệu kênh thứ nhất và biểu diễn miền thời gian của tín hiệu kênh thứ hai trong cùng một khối thời gian.

Đối với bộ phân loại ước tính nhiễu âm 1037b, tín hiệu đầu vào được phân phối đến bộ phát hiện hoạt động tín hiệu 1070 điều khiển bộ chọn 1071. Dựa trên kết quả của

bộ phát hiện hoạt động tín hiệu 1070, bộ chọn 1071 chỉ chọn các khung hoạt động. Hơn nữa, bộ tính toán mức tín hiệu 1072 được kết nối sau bộ chọn 1071. Mức tín hiệu được tính toán sau đó được chuyển tiếp đến bộ làm mịn (thời gian) 1073 mà, ví dụ, được thực hiện như bộ lọc IIR. Sau đó, trong các khối 1074, sự tính toán tỷ số tín hiệu trên nhiễu âm sẽ diễn ra và kết quả được so sánh, trong bộ so sánh 1075 ưu tiên với ngưỡng được xác định trước, ví dụ, từ 45 dB đến 25 dB và tốt hơn là nằm trong phạm vi từ 30 đến 40 dB và tốt hơn nữa là ở 35 dB.

Đầu ra của bộ so sánh 1075 là kết quả phát hiện chỉ ra mức nhiễu âm cao hoặc mức nhiễu âm thấp hoặc chỉ ra rằng việc thiết lập ngưỡng theo cách nhất định được thực hiện bởi bộ xử lý quy trình gán trọng số đơn hoặc, khi có hai bộ xử lý quy trình gán trọng số như được minh họa trên Fig.10d, sau đó kết quả quyết định từ bộ so sánh 1075, tức là tín hiệu 1050 điều khiển chuyển đổi đầu vào 1036c hoặc chuyển đổi đầu ra 1036d để chuyển tiếp phổ tương quan chéo được gán trọng số đến bộ xử lý 1040 một cách chính xác.

Tốt hơn là kết quả phát hiện 1050 được tính toán cho từng khối thời gian hoặc khung. Do đó, khi, ví dụ, đối với khung nhất định, bộ phát hiện hoạt động tín hiệu 1070 chỉ ra rằng đây là khung không hoạt động, khi đó, sự tính toán mức tín hiệu cũng như làm mịn thời gian không được thực hiện cho khung này, vì bộ chọn 1071 chỉ chọn khung hoạt động. Do đó, đối với khung không hoạt động, sự tính toán tỷ lệ SNR không được thực hiện trong phương án và do đó, trong phương án này, đối với khung không hoạt động này, kết quả phát hiện hoàn toàn không được cung cấp. Do đó, trong sự thực hiện, quy trình gán trọng số tương tự như đã được xác định trước đó đối với khung hoạt động cuối cùng được sử dụng hoặc, thay vào đó, đối với khung không hoạt động, hoặc quy trình gán trọng số thứ nhất hoặc quy trình gán trọng số thứ hai hoặc thậm chí là quy trình trọng số thứ ba là được áp dụng như giải pháp dự phòng. Ngoài ra, bộ tính toán tỷ lệ SNR 1074 có thể được thực hiện để sử dụng, đối với khung không hoạt động, mức tín hiệu được làm mịn theo thời gian của khung hoạt động cuối cùng hoặc gần đây nhất. Do đó, kết quả phát hiện có thể thu được ngay cả đối với các khung không hoạt động, hoặc đối với các khung không hoạt động, quy trình gán trọng số (dự phòng) nhất định được sử dụng hoặc, đối với các khung không hoạt động, quy trình gán trọng số tương tự

nếu đã được xác định cho khung hoạt động cuối cùng trước khung không hoạt động được tiếp tục sử dụng tùy trường hợp.

Trong đơn xin cấp bằng sáng chế trước đây [1], bộ ước tính chênh lệch thời gian liên kênh (Inter-channel Time Difference - ITD) đã được giới thiệu. Bộ ước tính này dựa trên tương quan chéo tổng quát với biến đổi pha (Generalized Cross-Correlation with PHase Transform - GCC-PHAT), kỹ thuật được sử dụng rộng rãi trong tài liệu TDOA (bài báo ban đầu là [2], tài liệu tham khảo tốt khác là [3]). Chênh lệch thời gian giữa hai kênh được tìm thấy bằng cách chọn đỉnh đầu ra của GCC. Độ mạnh tốt hơn có thể được thu bằng cách sử dụng độ dài cửa sổ phân tích lớn hoặc bằng cách làm mịn phổ tương quan chéo theo thời gian. Đóng góp chính của [1] là làm cho việc làm mịn này trở nên thích ứng với hệ số làm mịn phụ thuộc vào số đo độ phẳng phổ.

Các bước của bộ ước tính ITD của [1] có thể được mô tả như sau:

1. Biến đổi Fourier rời rạc: tín hiệu của kênh bên trái $x_L(n)$ và tín hiệu của kênh bên phải $x_R(n)$ được đóng khung, được tạo cửa sổ và được chuyển đổi sang miền tần số bằng cách sử dụng DFT:

$$X_L(k, s) = \sum_{n=0}^{N_{DFT}-1} x_L(n + sN)w(n)e^{-i2\pi\frac{kn}{N_{DFT}}}$$

$$X_R(k, s) = \sum_{n=0}^{N_{DFT}-1} x_R(n + sN)w(n)e^{-i2\pi\frac{kn}{N_{DFT}}}$$

với n là chỉ số mẫu thời gian, s là chỉ số khung, k là chỉ số tần số, N là độ dài khung, N_{DFT} là độ dài DFT và $w(n)$ là cửa sổ phân tích.

2. Phổ tương quan chéo: sự tương quan giữa hai kênh được tính trong miền tần số:

$$C(k, s) = X_L(k, s)X_R^*(k, s)$$

3. Làm mịn: phổ tương quan chéo được làm mịn theo thời gian với hệ số làm mịn tùy thuộc vào số đo độ phẳng phổ. Việc làm mịn mạnh hơn được sử dụng khi độ phẳng phổ thấp để làm cho bộ ước tính ITD mạnh hơn trên các tín hiệu âm tĩnh. Việc làm mịn

yếu hơn được sử dụng khi độ phẳng phổ cao để làm cho bộ ước tính ITD thích ứng nhanh hơn trên các tín hiệu nốt đậm, tức là khi tín hiệu thay đổi nhanh chóng.

Việc làm mịn được thực hiện bằng cách sử dụng:

$$\tilde{C}(k, s) = (1 - sfm(s))\tilde{C}(k, s - 1) + sfm(s)C(k, s)$$

với:

$$sfm(s) = \max(sfm_chan(X_L), sfm_chan(X_R))$$

và:

$$sfm_chan(X) = \frac{\prod_{k=0}^{N_{sfm}-1} X(k, s)^{\frac{1}{N_{sfm}}}}{\sum_{k=0}^{N_{sfm}-1} \frac{X(k, s)}{N_{sfm}}}$$

4. Gán trọng số: phổ tương quan chéo được làm mịn được gán trọng số bằng cách nghịch đảo độ lớn của nó. Việc gán trọng số này chuẩn hóa biên độ và chỉ giữ lại pha, đây là lý do tại sao nó được gọi là biến đổi pha (Phase Transform - PHAT).

$$\tilde{C}_{PHAT}(k, s) = \frac{\tilde{C}(k, s)}{|\tilde{C}(k, s)|}$$

5. Biến đổi nghịch đảo: GCC cuối cùng thu được bằng cách biến đổi phổ tương quan chéo $\tilde{C}_{PHAT}(k, s)$ trở lại miền thời gian:

$$GCC(n) = \frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} \tilde{C}_{PHAT}(k, s) e^{i2\pi \frac{kn}{N_{DFT}}}$$

6. Chọn đỉnh: phương thức đơn giản nhất là tìm kiếm giá trị tuyệt đối tối đa toàn diện của GCC được tìm thấy ở bước 5. Nếu giá trị tối đa này có giá trị trên một số ngưỡng, ITD được ước tính như độ trễ n tương ứng với giá trị tối đa này. Các phương thức nâng cao hơn sử dụng thêm các cơ chế dựa trên độ trễ và/hoặc vết tích để có được ước tính ITD mượt mà hơn theo thời gian.

GCC-PHAT thực hiện rất tốt trong môi trường nhiễu âm, dội âm thấp (xem ví dụ [3]). Tuy nhiên, khi mức nhiễu âm nền cao hoặc khi có sự có mặt của các thành phần tín hiệu khác (chẳng hạn như âm nhạc, các nốt đậm, các cảnh âm lập thể phức tạp, các

khung được phân loại là không hoạt động, các loa giao thoa), hiệu suất GCC-PHAT giảm đáng kể. Đầu ra GCC sau đó bị nhiễu âm và không chứa một đỉnh mạnh đơn. Do đó, việc chọn đỉnh thường không tìm được ITD chính xác. Điều này là do sự biến đổi pha xử lý tất cả các tần số như nhau, bất kể tỷ lệ tín hiệu trên nhiễu âm. GCC sau đó bị ô nhiễm bởi pha của các ngăn có tỷ lệ tín hiệu trên nhiễu âm thấp.

Để tránh vấn đề này, nhiều sự gán trọng số GCC khác đã được đề xuất trong tài liệu. Một trong số chúng được phát hiện là rất hiệu quả đối với các tín hiệu kiểm tra có vấn đề của các tác giả sáng chế. Nó được đề xuất lần thứ nhất trong [4] và lúc đó được gọi là “pha phổ lũy thừa chéo được cải biên” (modified cross-power spectrum phase - MCSP). Hiệu suất tốt của nó trong các môi trường nhiễu âm cao sau đó đã được xác nhận trong một số bài báo khác (xem ví dụ [5]). Việc gán trọng số (bước 4 của tình trạng kỹ thuật trước đó) được cải biên như sau:

$$\tilde{C}_{MCSP}(k, s) = \frac{\tilde{C}(k, s)}{|\tilde{C}(k, s)|^\rho} = \tilde{C}_{PHAT}(k, s) |\tilde{C}(k, s)|^{1-\rho}$$

với ρ là tham số từ 0 đến 1. $\rho = 0$ tương ứng với trường hợp tương quan chéo thông thường và $\rho = 1$ tương ứng với trường hợp GCC-PHAT. Giá trị dưới đây nhưng gần bằng 1 thường được sử dụng, mà cho phép cải biên GCC-PHAT bằng cách nhấn mạnh hơn vào các ngăn có độ tương quan cao, những giá trị thường tương ứng với tín hiệu trong khi các ngăn có độ tương quan thấp tương ứng với nhiễu âm. Chính xác hơn, các tác giả sáng chế đã phát hiện ra rằng giá trị $\rho = 0.8$ cho hiệu suất tốt nhất (đó là 0,75 trong [4] và 0,78 trong [5]).

Thật không may, việc gán trọng số mới này chỉ thực hiện tốt hơn GCC-PHAT khi có nhiễu âm nền ở mức cao. Các kịch bản thay thế trong đó việc gán trọng số mới có thể thực hiện tốt hơn GCC-PHAT là các khung không hoạt động (tức là sự phát hiện hoạt động tiếng nói phát hiện không hoạt động, có thể chỉ ra mức tiếng nói thấp), sự có mặt của các nốt đệm, kịch bản âm lập thể phức tạp, âm nhạc, các loa giao thoa, sự có mặt của âm nhạc nền, tiếng nói không sạch, trong các môi trường sạch, chẳng hạn như tiếng nói không có hoặc chỉ có mức độ thấp nhiễu âm nền hoặc âm nhạc hoặc các thành phần tín hiệu khác lệch khỏi tiếng nói sạch, GCC-PHAT vẫn thực hiện tốt hơn. Để luôn

đạt được kết quả tốt nhất, khi đó cần phải chuyển đổi giữa hai phương thức tùy thuộc vào nội dung tín hiệu.

Để phát hiện sự có mặt của nhiễu âm nền ở mức cao trong tín hiệu, bộ ước tính nhiễu âm cùng với bộ phát hiện hoạt động tín hiệu (signal activity detector - SAD) được sử dụng. Mức độ của tín hiệu l_s có thể được ước tính trên các khung trong đó SAD phát hiện tín hiệu, trong khi mức nhiễu âm l_N được ước tính bởi bộ ước tính nhiễu âm. Khi đó, sự có mặt của nhiễu âm nền mức cao được phát hiện đơn giản bằng cách so sánh tỷ số tín hiệu trên nhiễu âm $snr = l_s - l_N$ (tính bằng dB) với một ngưỡng, ví dụ nếu $snr < 35$ sau đó mức nhiễu âm cao được phát hiện.

Khi đã biết liệu tín hiệu có chứa mức nhiễu âm nền cao hay không, sự quyết định sẽ được thực hiện để chọn việc gán trọng số PHAT hoặc việc gán trọng số MCSP để tính GCC (bước 4 trong tình trạng kỹ thuật trước đó). Việc chọn đỉnh (bước 6 trong tình trạng kỹ thuật trước đó) cũng có thể được cải biên tùy thuộc vào việc có phát hiện được mức nhiễu âm nền cao hay không, để lấy mẫu bằng cách hạ thấp ngưỡng.

Sau đó, phương án được ưu tiên được mô tả theo cách từng bước.

0. Phát hiện mức nhiễu âm nền cao:

a. bộ ước tính nhiễu âm (ví dụ từ [6]) được sử dụng để ước tính mức nhiễu âm nền l_N . Bộ lọc làm mịn IIR được sử dụng để làm mịn mức nhiễu âm theo thời gian.

b. bộ phát hiện hoạt động tín hiệu (ví dụ từ [6]) được sử dụng để phân loại khung là hoạt động hay không hoạt động. Các khung hoạt động sau đó được sử dụng để tính toán mức tín hiệu l_s , đơn giản bằng cách tính toán năng lượng tín hiệu và làm mịn nó theo thời gian bằng cách sử dụng bộ lọc làm mịn IIR.

c. Nếu tỷ lệ tín hiệu trên nhiễu âm $snr = l_s - l_N$ (tính bằng dB) dưới ngưỡng (ví dụ 35 dB), thì mức nhiễu âm nền cao sẽ được phát hiện.

1. Biến đổi Fourier rời rạc: giống như trong bất kỳ tình trạng kỹ thuật nào trước đó.

2. Phổ tương quan chéo: giống như trong bất kỳ tình trạng kỹ thuật nào trước đó.

3. Làm mịn: giống như trong bất kỳ tình trạng kỹ thuật nào trước đó hoặc như được mô tả trong tài liệu này dựa trên đặc tính phổ.

4. Gán trọng số:

Nếu phát hiện thấy mức nhiễu âm nền thấp, thì sử dụng việc gán trọng số tương tự như trong tình trạng kỹ thuật trước đó (GCC-PHAT).

Nếu phát hiện thấy mức nhiễu âm nền cao, thì việc gán trọng số MCSP được sử dụng:

$$\tilde{C}_{MCSP}(k, s) = \frac{\tilde{C}(k, s)}{|\tilde{C}(k, s)|^\rho}$$

với $0 < \rho < 1$ (ví dụ $\rho = 0.8$). Để giữ đầu ra GCC-MCSP trong cùng một phạm vi với đầu ra GCC-PHAT, bước chuẩn hóa bổ sung được thực hiện:

$$\tilde{C}_{MCSP}(k, s) = \frac{\tilde{C}_{MCSP}(k, s)}{\frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} |\tilde{C}_{MCSP}(k, s)|}$$

5. Biến đổi nghịch đảo: giống như trong bất kỳ tình trạng kỹ thuật nào trước đó.

6. Chọn đỉnh: việc chọn đỉnh có thể được thích ứng trong trường hợp phát hiện thấy mức nhiễu âm nền cao và việc gán trọng số MCSP được sử dụng. Cụ thể, thấy rằng ngưỡng thấp hơn là có lợi.

Hơn nữa, Fig.10a minh họa cách thực hiện khác với cách thực hiện Fig.10c. Trong bộ gán trọng số 1036 của Fig.10c, bộ gán trọng số thực hiện quy trình gán trọng số thứ nhất hoặc thứ hai. Tuy nhiên, trong bộ gán trọng số 1036 như được minh họa trên Fig.10a, bộ gán trọng số chỉ thực hiện quy trình gán trọng số thứ hai đối với ký hiệu trên Fig.10d hoặc 10c. Cách thực hiện này rất hữu ích, khi bộ lọc làm mịn như được minh họa trong khối 1030 được sử dụng mà thực hiện quy trình gán trọng số thứ nhất sau khi làm mịn hoặc cùng với việc làm mịn trong ví dụ phép tính toán học hoặc phần cứng đơn lẻ. Do đó, trong trường hợp thực hiện quy trình gán trọng số thứ nhất mà là phép tính chuẩn hóa mà không có bất kỳ sự nén nào trong bộ lọc làm mịn, thì cả hai, một mặt, bộ lọc làm mịn 1030 và mặt khác bộ gán trọng số thực tế 1036 tương ứng với bộ gán trọng

số thực tế để gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn. Do đó, trong cách thực hiện trên Fig.10a, ước tính nhiễu âm 1038 chỉ được cung cấp cho bộ gán trọng số riêng biệt 1036 và sự lựa chọn giữa đầu ra của bộ lọc làm mịn 1030 mà đã được gán trọng số phù hợp với quy trình gán trọng số và sự lựa chọn giữa đầu ra của bộ gán trọng số thực tế 1036 trên Fig.10a được thực hiện bởi cài đặt bộ xử lý nhất định 1040 mà tự động sử dụng đầu ra từ bộ lọc làm mịn 1030, khi bộ gán trọng số 1036 không cung cấp bất kỳ tín hiệu đầu ra nào nhưng tự động ưu tiên đầu ra của bộ gán trọng số 1036 hơn đầu ra của bộ lọc làm mịn 1030, khi bộ gán trọng số 1036 cung cấp và xuất ra. Sau đó, ước tính nhiễu âm 1038 hoặc, như đã thảo luận trong các hình vẽ khác, tín hiệu điều khiển 1050 sau đó được sử dụng để kích hoạt hoặc hủy kích hoạt bộ gán trọng số 1036. Do đó, bộ gán trọng số thực tế để gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn bằng cách sử dụng quy trình gán trọng số thứ tự thứ nhất có thể được thực hiện theo nhiều cách khác nhau chẳng hạn như trong chế độ kích hoạt/hủy kích hoạt cụ thể trên Fig.10a hoặc chế độ hai nhân trên Fig.10d với chuyển đổi đầu vào hoặc đầu ra hoặc phù hợp với nhân quy trình gán trọng số đơn lẻ mà, tùy thuộc vào tín hiệu điều khiển chọn một hoặc quy trình gán trọng số khác hoặc thích ứng bộ xử lý gán trọng số chung để thực hiện quy trình gán trọng số thứ nhất hoặc thứ hai.

Sau đó, phương án được ưu tiên, trong đó việc làm mịn được thực hiện trước khi việc gán trọng số được mô tả. Trong ngữ cảnh này, các chức năng của bộ ước tính đặc tính phổ cũng được phản ánh trên Fig.4e, các 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 đó trong phương án được ưu tiên.

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

Các phương án được ưu tiên của bộ xử lý 1040 cũng được mô tả trên Fig.10c.

Tốt hơn là, ước tính đặc tính phổ tính toán độ ồn hoặc âm sắc của phổ trong đó cách thực hiện được ưu tiên là tính toán số đo độ phẳng phổ gần bằng 0 trong trường hợp các tín hiệu âm sắc hoặc không nhiều âm và gần bằng 1 trong trường hợp các tín hiệu nhiều âm hoặc giống như nhiều âm.

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

Cụ thể, mức độ làm mịn thứ nhất lớn hơn mức độ làm mịn thứ hai, trong đó đặc tính nhiều âm thứ nhất ít nhiều âm hơn đặc tính nhiều âm thứ hai hoặc đặc tính âm sắc thứ nhất nhiều âm sắc hơn đặc tính âm sắc thứ hai. Cách thực hiện được ưu tiên là đo độ phẳng phổ.

Hơn nữa, như được minh họa trên Fig.11a, bộ xử lý tốt hơn là được thực hiện để chuẩn hóa phổ tương quan chéo được làm mịn như được minh họa ở 456 trên Fig.4e và 11a trước khi thực hiện tính toán biểu diễn miền thời gian trong bước 1031 tương ứng với các bước 457 và 458 theo phương án của Fig.4e. Tuy nhiên, như được phác thảo trên Fig.11a, bộ xử lý cũng có thể hoạt động mà không cần chuẩn hóa ở bước 456 trên Fig.4e. Sau đó, bộ xử lý được tạo cấu hình để phân tích biểu diễn miền thời gian như được minh họa trong khối 1032 của Fig.11a để tìm ra chênh lệch thời gian liên kênh. Sự phân tích này có thể được thực hiện theo bất kỳ cách nào đã biết và sẽ dẫn đến độ mạnh được cải thiện, vì sự phân tích được thực hiện dựa trên phổ tương quan chéo được làm mịn phù hợp với đặc tính phổ.

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

Như được minh họa trên Fig.11c, cách thực hiện được ưu tiên của phép tính chọn đỉnh hoặc tìm kiếm đỉnh là để thực hiện phép tính này bằng cách sử dụng ngưỡng biến

thiên. Cụ thể, bộ xử lý được tạo cấu hình để thực hiện phép tính tìm kiếm đỉnh/chọn đỉnh trong biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm mịn bằng cách xác định 1034 ngưỡng biến thiên từ biểu diễn miền thời gian và bằng cách so sánh đỉnh hoặc một số đỉnh của biểu diễn miền thời gian (thu được với hoặc không với sự chuẩn hóa phổ) đến ngưỡng biến thiên, trong đó chênh lệch thời gian liên kênh được xác định là độ trễ thời gian liên quan đến đỉnh trong mỗi quan hệ được xác định trước với ngưỡng, chẳng hạn như lớn hơn ngưỡng biến thiên.

Như được minh họa trên Fig.11d, một phương án được ưu tiên được minh họa trong mã giả liên quan đến Fig.4e-b được mô tả sau đó bao gồm việc phân loại 1034a các giá trị phù hợp với độ lớn của chúng. Sau đó, như được minh họa trong mục 1034b trên Fig.11d, các giá trị cao nhất, ví dụ 10 hoặc 5% các 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 10 hoặc 5% giá trị cao nhất để thu được ngưỡng biến thiên.

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

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

Trong bước 1102 của Fig.11e, đỉnh trong mỗi khối con được xác định và ở bước 1103, đỉnh trung bình trong tất cả các khối con được xác định. Sau đó, trong bước 1104, giá trị nhân a được xác định phụ thuộc vào tỷ lệ tín hiệu trên nhiễu âm một mặt, và theo phương án khác, phụ thuộc vào chênh lệch giữa ngưỡng và đỉnh tối đa như được chỉ ra

ở bên trái của khối 1104. Tùy thuộc vào các giá trị đầu vào này, một trong số tốt hơn là ba giá trị nhân khác nhau được xác định trong đó giá trị nhân có thể bằng a_{low} , a_{high} và a_{lowest} .

Sau đó, ở bước 1105, giá trị nhân a được xác định trong khối 1104 được nhân với ngưỡng trung bình để thu được ngưỡng biến thiên mà sau đó được sử dụng trong phép tính so sánh trong khối 1106. Đối với phép tính so sánh, một lần nữa biểu diễn miền thời gian được nhập vào khối 1101 có thể được sử dụng hoặc các đỉnh đã được xác định trong mỗi khối con như được phác thảo trong khối 1102 có thể được sử dụng.

Sau đó, các phương án khác liên quan đến việc đánh giá và phát hiện đỉnh trong hàm tương quan chéo miền thời gian được phác thảo.

Việc đánh giá và phát hiện đỉnh trong hàm tương quan chéo miền thời gian do phương pháp tương quan chéo tổng quát (generalized cross-correlation - GCC-PHAT) để ước tính 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 độ lệch thấp với đỉnh mạnh, trong khi tiếng nói trong môi trường dội âm nhiều âm có thể tạo ra vector có độ lệch cao và các đỉnh có cường độ thấp hơn nhưng vẫn nổi bật cho thấy sự tồn tại của ITD. Thuật toán phát hiện đỉnh mà thích ứng và linh hoạt để phù hợp với các kịch bản đầu vào khác nhau được mô tả.

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

- ITD hợp lệ do đỉnh nổi bật. Có đỉnh vượt trội trong giới hạn $[-ITD_MAX, ITD_MAX]$ của hàm tương quan chéo.

- Không tương quan. Khi không có sự tương quan giữa hai kênh, không có đỉnh nổi bật. Ngưỡng phải được xác định, trên đó đỉnh đủ mạnh để được coi là giá trị ITD hợp lệ. Nếu không, không có xử lý ITD nào được tạo tín hiệu, có nghĩa là ITD được đặt thành 0 và không có căn chỉnh theo thời gian nào được thực hiện.

- Ngoài giới hạn ITD. Các đỉnh mạnh của hàm tương quan chéo bên ngoài vùng $[-ITD_MAX, ITD_MAX]$ phải được đánh giá để xác định liệu các ITD nằm ngoài khả năng xử lý của hệ thống có tồn tại hay không. Trong trường hợp này, không có xử lý ITD nào được báo hiệu và do đó không có sự căn chỉnh theo thời gian nào được thực hiện.

Để xác định liệu cường độ của đỉnh có đủ cao để được coi là giá trị chênh lệch thời gian hay không, cần phải xác định ngưỡng phù hợp. Đối với các kịch bản đầu vào khác nhau, đầu ra của hàm tương quan chéo thay đổi tùy thuộc vào các tham số khác nhau, ví dụ môi trường (nhiều âm, độ vang, v.v.), thiết lập micrô (AB, M/S, v.v.). Do đó, để xác định một cách thích ứng ngưỡng là điều cần thiết.

Trong thuật toán được đề xuất, ngưỡng được xác định trước tiên bằng cách tính giá trị trung bình của phép tính sơ bộ về độ lớn của hàm tương quan chéo trong vùng $[-ITD_MAX, ITD_MAX]$ (Fig.13), giá trị trung bình sau đó được gán trọng số theo đó tùy thuộc vào ước tính SNR.

Mô tả từng bước của thuật toán được mô tả bên dưới.

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

Vector tương quan chéo được chia thành ba khu vực chính: khu vực quan tâm cụ thể là $[-ITD_MAX, ITD_MAX]$ và khu vực bên ngoài giới hạn ITD_MAX , cụ thể là các độ trễ thời gian nhỏ hơn $-ITD_MAX$ (max_low) và cao hơn 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 tối đa được phát hiện trong vùng quan tâm.

Để xác định xem 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 cao nhất $peak_sub$ và vị trí trễ thời gian tương đương $index_sub$ được tìm thấy và được lưu.

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

Giá trị tối đa $peak_max$ được so sánh với max_low và max_high . Nếu $peak_max$ thấp hơn một trong hai thì không có sự xử lý ITD nào được tạo tín hiệu và không có căn chỉnh theo thời gian nào được thực hiện. Do giới hạn xử lý ITD của hệ thống, không cần đánh giá các cường độ của các đỉnh nằm ngoài giới hạn.

Giá trị trung bình của các cường độ của các đỉnh được tính toán:

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

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

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

Trong trường hợp mà $SNR \ll SNR_{threshold}$ và $|thres - peak_max| < \varepsilon$, cường độ của đỉnh cũng được so sánh với ngưỡng thoải mái hơn một chút ($a_w = a_{lowest}$), để tránh từ chối đỉnh nổi bật với các đỉnh cao lân cận. Các hệ số gán trọng số có thể là ví dụ $a_{high} = 3$, $a_{low} = 2,5$ và $a_{lowest} = 2$, trong khi $SNR_{threshold}$ có thể là ví dụ 20dB và giới hạn $\varepsilon = 0,05$.

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

Nếu $peak_max > thres$ thì độ trễ thời gian tương đương được trả về là ITD được ước tính, nếu không thì không có sự xử lý ITD nào được tạo tín hiệu ($ITD = 0$). Các phương án khác được mô tả ở phần sau so với Fig.4e.

Fig.11f minh họa cách thực hiện được ưu tiên của việc xác định đầu ra ITD (inter-channel time difference - chênh lệch thời gian liên kênh) hợp lệ.

Các khối con của biểu diễn miền thời gian của phổ tương quan chéo được gán trọng số và được làm mịn hoặc không được làm mịn được đưa vào bước xác định trong bộ xử lý 1040. Bước xác định 1120 này xác định phạm vi hợp lệ và phạm vi không hợp lệ trong biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được gán trọng số và làm mịn hoặc không được làm mịn. Trong bước 1121, đỉnh tối đa được xác định

trong phạm vi không hợp lệ, và trong bước 1122, đỉnh tối đa được xác định trong phạm vi hợp lệ. Cụ thể, ít nhất một đỉnh cực đại được xác định trong phạm vi không hợp lệ và ít nhất một đỉnh cực đại được xác định trong phạm vi hợp lệ. Trong khối 1123, các đỉnh tối đa của phạm vi hợp lệ và phạm vi không hợp lệ được so sánh. Trong trường hợp đỉnh hợp lệ, tức là đỉnh tối đa trong phạm vi hợp lệ lớn hơn "đỉnh không hợp lệ", đỉnh tối đa trong phạm vi không hợp lệ, thì sự xác định ITD 1124 thực sự được thực hiện và đầu ra ITD hợp lệ được cung cấp. Tuy nhiên, khi phát hiện thấy "đỉnh không hợp lệ" lớn hơn "đỉnh hợp lệ" hoặc đỉnh không hợp lệ có cùng kích thước với đỉnh hợp lệ, thì đầu ra hợp lệ không được cung cấp và, tốt hơn là thông báo lỗi hoặc bất kỳ hành động có thể so sánh nào được thực hiện để khiến bộ xử lý chú ý.

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

Tuy nhiên, như đã nêu và như được minh họa trên Fig.10b, tồn tại nhiều trường khác, trong đó tín hiệu cũng có thể được tiếp tục xử lý bằng cách sử dụng chênh lệch thời gian liên kênh được xác định.

Fig.1 minh họa thiết bị để 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 bộ xác định tín hiệu 200. Bộ xác định tham số 100 xác định, một mặt, là tham số căn chỉnh băng rộng và, mặt khác, nhiều tham số căn chỉnh băng hẹp từ tín hiệu đa kênh. Các tham số này được xuất ra thông qua đường tham số 12. Hơn nữa, các tham số này cũng được xuất ra qua đường tham số 14 khác 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 bằng cách 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 ở đầu ra của bộ căn chỉnh tín hiệu 200. Các kênh được căn chỉnh 20 này được chuyển tiếp đến bộ xử lý tín hiệu 300 mà 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 dòng 20. Thiết bị để mã hóa còn bao gồm bộ mã hóa tín hiệu 400 để mã hóa tín hiệu giữa từ dòng 31 và tín hiệu bên từ dòng 32 để thu được tín hiệu giữa được mã hóa trên dòng 41 và tín hiệu bên được mã hóa trên dòng 42. Cả hai tín hiệu này đều được chuyển tiếp đến giao diện đầu ra 500 để tạo ra tín hiệu đa kênh được mã hóa ở dòng đầu ra 50. Tín hiệu được mã hóa ở dòng đầu ra 50 bao gồm tín hiệu giữa được mã hóa từ dòng 41, tín hiệu bên được mã hóa từ dòng 42, các tham số căn chỉnh băng hẹp và các tham số căn chỉnh băng rộng từ dòng 14 và, tùy chọn, tham số mức từ dòng 14 và, tùy chọn bổ sung, tham số lấp đầy âm lập thể được tạo bởi bộ mã hóa tín hiệu 400 và được chuyển tiếp đến giao diện đầu ra 500 qua dòng 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 bằng cách 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, theo 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 tín hiệu đa kênh được căn chỉnh đặc tính băng rộng. Tuy nhiên, theo các phương án khác, các tham số được xác định mà không có trình tự các quy trình cụ thể này.

Fig.4a minh họa cách thực hiện được ưu tiên, trong đó trình tự các bước cụ thể bao gồm đường kết nối 15 được thực hiện. Trong bước 16, tham số căn chỉnh băng rộng được xác định bằng cách sử dụng hai kênh và tham số căn chỉnh băng rộng như chênh lệch thời gian liên kênh hoặc tham số ITD được thu. Sau đó, ở bước 21, hai kênh được căn chỉnh bởi bộ căn chỉnh tín hiệu 200 của Fig.1 bằng cách sử dụng tham số căn chỉnh băng rộng. Sau đó, trong bước 17, các tham số băng hẹp được xác định bằng cách 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 cho các băng khác nhau của tín hiệu đa kênh. Sau đó, ở bước 22, các giá trị phổ trong mỗi băng tham số được căn chỉnh bằng cách sử dụng tham số căn chỉnh băng hẹp tương ứng cho băng cụ thể này. Khi quy trình này ở 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 bên trái/bên phải được căn chỉnh sẽ có sẵn để bộ xử lý tín hiệu 300 của Fig.1 xử lý tín hiệu thêm.

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

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

Hơn nữa, như được minh họa ở 152, bộ xác định tham số, bộ căn chỉnh tín hiệu và bộ xử lý tín hiệu được minh họa ở 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 154 để tạo ra ít nhất là biểu diễn miền thời gian của tín hiệu giữa.

Tốt hơn là, bộ chuyển đổi thời gian phổ còn chuyển đổi thêm biểu diễn phổ của tín hiệu bên cũng được xác định bởi các quy trình được biểu diễn bởi khối 152 thành biểu diễn miền thời gian, và bộ mã hóa tín hiệu 400 của Fig.1 sau đó được tạo cấu hình để mã hóa thêm tín hiệu giữa và/hoặc tín hiệu bên dưới dạng các tín hiệu miền thời gian tùy 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 việc cung cấp cửa sổ phân tích có ít nhất một phần đệm bằng 0 ở một đầu của nó và cụ thể là phần đệm bằng 0 ở phần cửa sổ ban đầu và phần đệm bằng 0 ở phần cửa sổ kết thúc như được minh họa, ví dụ, trên Fig.7 sau đó. Hơn nữa, cửa sổ phân tích cũng có các phạm vi chồng lấp hoặc các phần chồng lấp ở nửa thứ nhất của cửa sổ và ở nửa thứ hai của cửa sổ và, ngoài ra, tốt hơn là phần giữa là phạm vi không chồng lấp tùy trường hợp.

Trong bước 156, mỗi kênh được tạo cửa sổ bằng cách sử dụng cửa sổ phân tích với các phạm vi chồng lấp. Cụ thể, mỗi kênh được tạo cửa sổ bằng cách sử dụng cửa sổ phân tích theo cách mà khối thứ nhất của kênh được thu. Sau đó, khối thứ hai của cùng một kênh được thu mà có phạm vi chồng lấp nhất định với khối thứ nhất, v.v., sao cho tiếp theo, ví dụ, năm hoạt động cửa sổ, năm khối của các mẫu được tạo cửa sổ của mỗi kênh có sẵn sau đó được biến đổi riêng lẻ thành biểu diễn phổ như được minh họa ở 157 trên Fig.4c. Quy trình tương tự cũng được thực hiện cho kênh khác để ở cuối bước 157,

chuỗi các khối của các giá trị phổ và đặc biệt là các giá trị phổ phức tạp như các giá trị phổ DFT hoặc các mẫu băng con phức tạp là có sẵn.

Trong bước 158, được thực hiện bởi bộ xác định tham số 100 của Fig.1, tham số căn chỉnh băng rộng được xác định và ở bước 159, được thực hiện bởi căn chỉnh tín hiệu 200 của Fig.1, sự dịch chuyển tròn được thực hiện bằng cách 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 cho các băng /băng con riêng lẻ và trong bước 161, các giá trị phổ được căn chỉnh được xoay cho từng băng bằng cách 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 giữa và tín hiệu bên như được minh họa ở bước 301. Trong bước 302, một số loại xử lý khác 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 chuyển đổi trở lại miền thời gian và ở bước 304, cửa sổ tổng hợp được áp dụng cho mỗi khối được thu bởi bước 303 và, trong bước 305, phép tính cộng chồng lắp cho tín hiệu giữa một mặt và mặt khác là phép tính cộng chồng lắp cho tín hiệu bên được thực hiện để cuối cùng thu được các tín hiệu giữa/bên miền thời gian.

Cụ thể, các phép tính của các bước 304 và 305 dẫn đến loại tăng giảm cường độ chéo từ một khối của tín hiệu giữa hoặc tín hiệu bên trong khối tiếp theo của tín hiệu giữa và tín hiệu bên được thực hiện sao cho, ngay cả khi bất kỳ tham số nào xảy ra các thay đổi chẳng hạn như tham số chênh lệch thời gian liên kênh hoặc tham số chênh lệch pha liên kênh, tuy nhiên, điều này sẽ không thể nghe được 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 (Mid/Side - M/S) kết hợp khai thác một số tín hiệu không gian, trong đó kênh giữa được mã hóa bởi bộ mã hóa lõi đơn sơ cấp, và kênh bên được mã hóa trong bộ mã hóa lõi thứ cấp. Các nguyên tắc của bộ mã hóa và bộ giải mã được mô tả trên Fig.6a và Fig.6b.

Quá trình 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. Đó là trường hợp của

tính toán ITD, mà có thể được tính toán và áp dụng trước khi phân tích tần số để căn chỉnh các kênh kịp thời trước khi theo đuổi phân tích và xử lý âm lập thể. Ngoài ra, xử lý ITD có thể được thực hiện trực tiếp trong miền tần số. Vì các bộ mã hóa tiếng nói thông thường như ACELP không chứa bất kỳ phân tách thời gian – tần số nội bộ nào, nên mã hóa âm lập thể bổ sung thêm giàn bộ lọc được điều biến phức tạp thông qua giàn bộ lọc phân tích và tổng hợp trước bộ mã hóa lỗi và tầng khác của giàn bộ lọc phân tích - tổng hợp sau bộ giải mã lỗi. Trong phương án được ưu tiên, DFT được lấy mẫu quá mức với vùng chồng lấp thấp được sử dụng. Tuy nhiên, theo các phương án khác, có thể sử dụng bất kỳ sự phân tách thời gian – tần số được tạo giá trị phức tạp nào với độ phân giải theo thời gian tương tự.

Quá trình xử lý âm lập thể bao gồm tính toán các tín hiệu không gian: các chênh lệch thời gian liên kênh (inter-channel Time Difference - ITD), các chênh lệch pha liên kênh (inter-channel Phase Differences - IPD) và các chênh lệch mức liên kênh (inter-channel Level Differences - ILD). 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 trong băng rộng hoặc trong miền thời gian trong khi các IPD và các ILD được tính cho mỗi hoặc một phần của các băng tham số, tương ứng với sự phân tách không đồng nhất của không gian tần số. Khi hai kênh được căn chỉnh, âm lập thể M/S kết hợp sẽ được áp dụng, trong đó tín hiệu bên được tiếp tục dự báo từ tín hiệu giữa. Độ khuếch đại dự báo được suy ra từ các ILD.

Tín hiệu giữa được mã hóa thêm bởi một bộ mã lỗi sơ cấp. Trong phương án được ưu tiên, bộ mã hóa lỗi sơ cấp là tiêu chuẩn 3GPP EVS, hoặc mã hóa được suy ra từ nó mà có thể chuyển đổi giữa chế độ mã hóa tiếng nói, ACELP và chế độ âm nhạc dựa trên chuyển đổi MDCT. Tốt hơn là, ACELP và bộ mã hóa dựa trên MDCT được hỗ trợ bởi mở rộng băng rộng miền thời gian (Time Domain BandWidth Extension - TD-BWE) và/hoặc lấp đầy khoảng trống thông minh (Intelligent Gap Filling - IGF).

Tín hiệu Bên được dự báo lần thứ nhất bởi kênh giữa bằng cách sử dụng các độ khuếch đại dự báo được thu từ các ILD. Phần dư có thể tiếp tục được dự báo bởi phiên bản được tạo độ trễ của tín hiệu giữa hoặc được mã hóa trực tiếp bởi bộ mã hóa lỗi thứ cấp, được thực hiện theo phương án được ưu tiên trong miền MDCT. Quá trình xử lý

âm lập thể tại bộ mã hóa có thể được tóm tắt bằng Fig.5 như sẽ được giải thích ở phần sau.

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

Cụ thể, tín hiệu được nhận bởi giao diện đầu vào 600. Kết nối với giao diện đầu vào 600 là bộ giải mã tín hiệu 700, và bộ khử căn chỉnh tín hiệu 900. Hơn nữa, bộ xử lý tín hiệu 800 được kết nối với bộ giải mã tín hiệu 700 một mặt và mặt khác được kết nối với bộ khử 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 về tham số căn chỉnh băng rộng và thông tin về 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 đầu ra bởi giao diện đầu ra của 500 trên Fig.1.

Tuy nhiên, điều quan trọng cần lưu ý ở đây là, 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 có trong tín hiệu được mã hóa ở dạng nhất định có thể chính xác là các tham số căn chỉnh như được sử dụng bởi bộ căn chỉnh hiệu tín hiệu 200 trên Fig.1 nhưng, thay vào đó, cũng có thể là các giá trị nghịch đảo của chúng, tức là, các tham số có thể được sử dụng bởi chính các phép tính tương tự được thực hiện bởi bộ căn chỉnh tín hiệu 200 nhưng với các giá trị nghịch đảo để thu được sự khử căn chỉnh.

Do đó, thông tin về các tham số căn chỉnh có thể là các tham số căn chỉnh như 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 căn chỉnh” thực tế. Ngoài ra, các tham số này thường sẽ được lượng tử hóa ở dạng nhất định như sẽ được thảo luận ở phần sau đối với Fig.8.

Giao diện đầu vào 600 của Fig.2 chia thông tin về 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 tới bộ khử căn chỉnh tín hiệu 900. Mặt khác, tín hiệu giữa được mã hóa được chuyển tiếp đến 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 602.

Bộ giải mã tín hiệu được tạo cấu hình để giải mã tín hiệu giữa được mã hóa và để 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. Những 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 trên các đường 801, 802, tương ứng. Bộ khử căn chỉnh tín hiệu 900 được tạo cấu hình để khử căn chỉnh kênh thứ nhất được giải mã trên đường 801 và kênh bên phải được giải mã 802 bằng cách sử dụng thông tin về tham số căn chỉnh băng rộng và sử dụng thêm thông tin về nhiều tham số căn chỉnh băng hẹp để thu được tín hiệu đa kênh được giải mã, tức là tín hiệu được giải mã có ít nhất hai kênh được giải mã và được khử căn chỉnh trên các đường 901 và 902.

Fig.9a minh họa trình tự các bước được ưu tiên được thực hiện bởi bộ khử căn chỉnh tín hiệu 900 từ Fig.2. Cụ thể, bước 910 nhận các kênh bên trái và phải được căn chỉnh như có sẵn trên các đường 801, 802 từ Fig.2. Trong bước 910, bộ khử căn chỉnh tín hiệu 900 khử căn chỉnh các băng con riêng lẻ bằng cách sử dụng thông tin về các tham số căn chỉnh băng hẹp để thu được các kênh thứ nhất và thứ hai hoặc bên trái và bên phải được giải mã được khử căn chỉnh theo pha tại 911a và 911b. Trong bước 912, các kênh được khử căn chỉnh bằng cách sử dụng tham số căn chỉnh băng rộng để thu được các kênh được khử căn chỉnh theo pha và thời gian tại 913a và 913b.

Trong bước 914, bất kỳ quá trình xử lý nào được thực hiện bao gồm sử dụng việc tạo cửa sổ hoặc bất kỳ phép tính chồng lấp – cộng nào hoặc nói chung là bất kỳ phép tính tăng giảm cường độ chéo nào để thu được, ở 915a hoặc 915b, tín hiệu được giải mã giảm thiểu hoặc không có thành phần lạ, tức là, đối với các kênh được giải mã không có bất kỳ thành phần lạ nào mặc dù thường đã có các tham số khử căn chỉnh thay đổi theo thời gian một mặt cho băng rộng và mặt khác cho nhiều băng hẹp.

Fig.9b minh họa cách thực hiện được ưu tiên 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.

Ngoài ra, bộ xử lý tín hiệu còn bao gồm bộ chuyển đổi giữa/bên sang bên trái/bên phải 820 để tính toán từ tín hiệu giữa M và tín hiệu bên S sang 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 – bên trái/bên phải trong khối 820, tín hiệu bên S không nhất thiết phải được sử dụng. Thay vào đó, như được thảo luận sau đó, các tín hiệu bên trái/bên phải được tính toán ban đầu bằng cách chỉ sử dụng tham số độ khuếch đại được suy ra từ tham số chênh lệch mức liên kênh ILD. Nói chung, độ khuếch đại dự báo cũng có thể được coi là dạng của ILD. Độ khuếch đại có thể được suy ra từ ILD nhưng cũng có thể được tính toán trực tiếp. Ưu tiên không tính ILD nữa, mà tính độ khuếch đại dự báo trực tiếp và truyền và sử dụng độ khuếch đại dự báo trong bộ giải mã hơn là tham số ILD.

Do đó, trong cách 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 bên trái/bên phải tốt hơn bằng cách sử dụng tín hiệu bên được truyền S như được minh họa bằng đường vòng 821.

Do đó, bộ chuyển đổi 820 hoạt động bằng cách sử dụng tham số mức được thu thông qua đầu vào tham số mức 822 và không thực sự sử dụng tín hiệu bên S mà sử dụng bộ cập nhật kênh 830 sau đó hoạt động bằng cách sử dụng phía 821 và, tùy thuộc vào việc thực hiện cụ thể, sử dụng tham số lấp đầy âm lập thể nhận được qua đường 831. Sau đó, bộ căn chỉnh tín hiệu 900 bao gồm bộ khử căn chỉnh theo pha và bộ định tỷ lệ năng lượng 910. Việc định tỷ lệ năng lượng được kiểm soát bởi hệ số tỷ lệ được suy ra bởi bộ tính toán hệ số tỷ lệ 940. Bộ tính toán hệ số tỷ lệ 940 được cung cấp bởi đầu ra của bộ cập nhật kênh 830. Dựa trên các tham số căn chỉnh bằng hợp nhận được qua đầu vào 911, việc khử căn chỉnh theo pha được thực hiện và, trong khối 920, dựa trên tham số căn chỉnh bằng rộng nhận được qua đường 921, việc khử căn chỉnh theo thời gian được thực hiện. Cuối cùng, chuyển đổi phổ - thời gian 930 được thực hiện để cuối cùng thu được tín hiệu được giải mã.

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

Cụ thể, các kênh được khử căn chỉnh bằng hợp được đưa vào chức năng khử căn chỉnh bằng rộng tương ứng với khối 920 của Fig.9b. DFT hoặc bất kỳ sự biến đổi nào

khác được thực hiện trong khối 931. Sau khi tính toán thực tế 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 với cửa sổ phân tích hoặc được suy ra từ cửa sổ phân tích, ví dụ như nội suy hoặc số thập phân nhưng phụ thuộc theo cách nhất định từ kết quả 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 lại thành một cho mỗi điểm trong phạm vi chồng lấp. Do đó, tiếp theo cửa sổ tổng hợp trong khối 932, phép tính chồng lấp và phép tính cộng tiếp theo được thực hiện. Ngoài ra, thay vì tạo cửa sổ tổng hợp và phép tính chồng lấp/cộng, bất kỳ sự tăng giảm cường độ chéo nào giữa các khối tiếp theo cho mỗi kênh được thực hiện để thu được, như đã được thảo luận trong ngữ cảnh của Fig.9a, tín hiệu được giải mã được giảm thành phần lạ.

Khi Fig.6b được xem xét, rõ ràng là các hoạt động giải mã thực tế cho tín hiệu giữa, tức là “bộ giải mã EVS” một mặt và, đối với tín hiệu bên, sự lượng tử hóa vector nghịch đảo VQ^{-1} và hoạt động MDCT nghịch đảo (inverse MDCT operation - IMDCT) tương ứng với bộ giải mã tín hiệu 700 trên Fig.2.

Hơn nữa, các hoạt động DFT trong các khối 810 tương ứng với phần tử 810 trên Fig.9b và các chức năng của việc xử lý âm lập thể nghịch đảo và dịch chuyển thời gian nghịch đảo tương ứng với các khối 800, 900 trên Fig.2 và các hoạt động DFT nghịch đảo 930 trên Fig.6b tương ứng với hoạt động 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 lẻ. Tốt hơn là, phổ DFT hoặc bất kỳ phổ nào khác được minh họa trên Fig.3 là phổ phức tạp và mỗi vạch là vạch phổ phức tạp 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 minh họa trên Fig.3.

Hơn nữa, nhiều tham số căn chỉnh băng hẹp được cung cấp để 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.

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

Ngoài ra, các tham số lấp đầy âm lập thể được cung cấp cho số lượng băng nhất định không bao gồm các băng thấp hơn, chẳng hạn như, theo phương án minh họa, cho các băng 4, 5 và 6, trong khi có các giá trị phổ tín hiệu bên cho các băng tham số thấp hơn 1, 2 và 3 và, do đó, không tồn tại các tham số lấp đầy âm lập thể cho các băng thấp hơn này trong đó sự phù hợp dạng sóng thu được bằng cách sử dụng chính tín hiệu bên hoặc tín hiệu dự đoán biểu diễn tín hiệu bên.

Như đã nêu, tồn tại nhiều vạch phổ hơn ở các băng cao hơn, chẳng hạn như, theo 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, đương nhiên là, số lượng băng tham số, số lượng vạch phổ và số lượng 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 các tham số và số lượng băng mà các tham số được cung cấp trong phương án nhất định, trong đó, trái ngược với Fig.3, thực tế có 12 băng.

Như được minh họa, tham số mức ILD được cung cấp cho mỗi 12 băng và được lượng tử hóa với độ chính xác lượng tử hóa được biểu diễn bằng 5 bit trên mỗi băng.

Hơn nữa, các tham số căn chỉnh băng hẹp IPD chỉ được cung cấp cho các băng thấp hơn có tần số biên là 2,5 kHz. Ngoài ra, chênh lệch thời gian liên kênh hoặc tham số căn chỉnh băng rộng chỉ được cung cấp dưới dạng tham số đơn cho toàn phổ nhưng với độ chính xác lượng tử hóa rất cao được biểu diễn bằng 8 bit cho toàn băng.

Hơn nữa, các tham số lấp đầy âm lập thể được lượng tử hóa khá gần được cung cấp được biểu diễn bởi 3 bit trên mỗi băng và không phải cho các băng thấp hơn dưới 1 kHz vì đối với các băng thấp hơn, tín hiệu bên được mã hóa thực sự hoặc các giá trị pho dư của tín hiệu bên được bao gồm.

Sau đó, quy trình xử lý được ưu tiên ở phía bộ mã hóa được tóm tắt theo 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 từ 155 đến 157 của Fig.4c. Trong bước 158, tham số căn chỉnh băng rộng được tính toán và, đặc biệt là, chênh lệch thời gian liên kênh (ITD) tham số căn chỉnh băng rộng được ưu tiên. Như được minh họa trong 170, sự dịch chuyển thời gian của L và R trong miền tần số được thực hiện. Ngoài ra, sự dịch chuyển thời gian này cũng có thể được thực hiện trong miền thời gian. Sau đó, DFT nghịch đảo được thực hiện, dịch chuyển thời gian được thực hiện trong miền thời gian và DFT chuyển tiếp bổ sung được thực hiện để một lần nữa có các biểu diễn phổ tiếp theo việc căn chỉnh bằng cách 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 biểu diễn L và R được dịch chuyển như được minh họa ở bước 171. Ví dụ, bước này tương ứng với bước 160 của Fig.4c. Các biểu diễn L và R được dịch chuyển theo thời gian được xoay 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 trong bước 301, và tốt hơn là, bổ sung với hoạt động hội thoại năng lượng như được thảo luận ở phần sau. Trong bước tiếp theo 174, dự báo của S với M là hàm của ILD và tùy chọn với tín hiệu M trong quá khứ, tức là tín hiệu giữa của khung trước đó 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 175, tín hiệu giữa miền thời gian m và, tùy chọn, tín hiệu dư được mã hóa như được minh họa trong bước 175. 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 được tạo ra trong miền DFT và đầu tiên được dự báo từ tín hiệu giữa (Mid) như:

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

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

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

- Bằng cách mã hóa thứ cấp 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 diện được truyền cho toàn bộ phổ.

- Bằng sự dự báo dư, được gọi là lấp đầy âm lập thể, dự báo phổ bên dư với phổ tín hiệu giữa (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 dự báo được truyền trên mỗi băng tham số.

Hai loại tinh chỉnh mã hóa có thể được trộn trong cùng một phổ DFT. Theo phương án được ưu tiên, mã hóa dư được áp dụng trên các băng tham số thấp hơn, trong khi dự báo dư được áp dụng trên các băng còn lại. Mã hóa phần dư nằm trong phương án được ưu tiên như mô tả 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 như DFT, MDCT được lấy mẫu quan trọng và phù hợp hơn để mã hóa âm thanh. Các hệ số MDCT được lượng tử hóa vector trực tiếp bằng lượng tử hóa vector mạng nhưng có thể được mã hóa theo cách khác bởi bộ lượng tử hóa vô hướng theo sau bởi bộ mã hóa entropy. Ngoài ra, tín hiệu bên dư cũng có thể được mã hóa trong miền thời gian bằng kỹ thuật mã hóa tiếng nói hoặc trực tiếp trong miền DFT.

1. Phân tích thời gian – tần số: DFT

Điều quan trọng là sự phân tách thời gian – tần số bổ sung từ quá trình xử lý âm lập thể được thực hiện bởi các DFT cho phép phân tích cảnh thính giác tốt trong khi không làm tăng đáng kể độ trễ tổng thể của hệ thống mã hóa. Theo mặc định, độ phân giải thời gian 10 ms (gấp đôi khung 20 ms của bộ mã hóa lõi) được sử dụng. Các cửa sổ

phân tích và tổng hợp giống nhau và đối xứng. Cửa sổ được biểu diễn ở tốc độ lấy mẫu 16 kHz trên Fig.7. Có thể quan sát thấy rằng vùng chồng lấp được giới hạn để giảm độ trễ được tạo ra và phần đệm 0 cũng được thêm vào để chống lại sự cân bằng chuyển dịch vòng khi áp dụng ITD trong miền tần số như sẽ được giải thích sau đây.

2. Các tham số âm lập thể

Các tham số âm lập thể có thể được truyền tối đa tại độ phân giải thời gian của DFT âm lập thể. Ở mức tối thiểu, nó có thể được giảm xuống độ phân giải khung của bộ mã hóa lõi, tức là 20 ms. Theo mặc định, khi không phát hiện thấy nốt đệm, các tham số được tính sau mỗi 20 ms trên 2 cửa sổ DFT. Các băng tham số tạo thành sự phân tách không đồng nhất và không chồng lấp của phổ theo sau khoảng 2 lần hoặc 4 lần các băng thông hình chữ nhật tương đương (Equivalent Rectangular Bandwidth - ERB). Theo mặc định, thang ERB 4 lần được sử dụng cho tổng số 12 băng cho băng thông tần số là 16 kHz (tốc độ lấy mẫu 32 kbps, âm lập thể băng siêu rộng). Fig.8 tóm tắt ví dụ về cấu hình, trong đó thông tin bên âm lập thể được truyền với khoảng 5 kbps.

3. Tính toán ITD và căn chỉnh thời gian kênh

ITD được tính toán bằng cách ước tính độ trễ thời gian đến (Time Delay of Arrival - TDOA) bằng cách sử dụng tương quan chéo tổng quát 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 lần lượt là phổ tần số của các kênh bên trái và bên phải. Phân tích tần số có thể được thực hiện độc lập với DFT được sử dụng cho quá trình xử lý âm lập thể tiếp theo hoặc có thể được chia sẻ. Mã giả để 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.4e minh họa sơ đồ để thực hiện mã giả được minh họa trước đó để thu được tính toán mạnh và hiệu quả về chênh lệch thời gian liên kênh làm ví dụ cho 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 này thường sẽ là phân tích DFT giống như đã được thảo luận trong ngữ cảnh của các bước 155 đến 157 trên Fig.5 hoặc Fig.4c, chẳng hạn.

Sau đó, tương quan chéo 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, ở bước 454, số đo độ phẳng phổ rộng hơn được chọn. Tuy nhiên, lựa chọn trong bước 454 không nhất thiết phải là lựa chọn số đo độ phẳng phổ rộng hơn nhưng việc xác định này của SFM đơn từ cả hai kênh này cũng có thể là lựa chọn và tính toán

của chỉ kênh bên trái hoặc chỉ kênh bên phải hoặc có thể là tính toán trung bình được gán trọng số của cả hai giá trị SFM.

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

Tốt hơn là, số đo độ phẳng phổ được tính toán bằng cách chia trung bình hình học của phổ cường độ cho trung bình cộng của phổ cường độ. Do đó, các giá trị cho SFM được giới hạn từ 0 đến 1.

Trong bước 456, phổ tương quan chéo được làm mịn sau đó được chuẩn hóa theo cường độ của nó và ở bước 457, DFT nghịch đảo của phổ tương quan chéo được chuẩn hóa và được làm mịn được tính toán. Trong bước 458, bộ lọc miền thời gian nhất định được ưu tiên thực hiện nhưng bộ lọc miền thời gian này cũng có thể được gạt sang một bên tùy thuộc vào sự thực hiện nhưng được ưu tiên hơn như sẽ được trình bày ở phần sau.

Trong bước 459, ước tính ITD được thực hiện bằng cách chọn đỉnh của hàm tương quan chéo tổng hợp của bộ lọc và bằng cách thực hiện phép tính ngưỡng nhất định.

Nếu không thu được đỉnh cao hơn ngưỡng, thì ITD được đặt thành 0 và không có căn chỉnh theo thời gian nào được thực hiện cho khối tương ứng này.

Tính toán ITD cũng có thể được tóm tắt như sau. Tương quan chéo được tính toán trong miền tần số trước khi được làm mịn tùy thuộc vào số đo độ phẳng phổ. SFM được giới hạn giữa 0 và 1. Trong trường hợp tín hiệu giống như nhiễu âm, SFM sẽ cao (tức là khoảng 1) và làm mịn sẽ yếu. Trong trường hợp tín hiệu giống như âm sắc, SFM sẽ thấp và làm mịn sẽ trở nên mạnh hơn. Tương quan chéo được làm mịn sau đó được chuẩn hóa theo biên độ của nó trước khi được chuyển đổi trở lại miền thời gian. Sự chuẩn hóa tương ứng với dạng chuyển đổi pha của tương quan chéo, và được biết đến là thể hiện hiệu suất tốt hơn so với tương quan chéo thông thường trong môi trường nhiễu âm thấp và độ vang tương đối cao. Hàm miền thời gian thu được như vậy đầu tiên được lọc để đạt được đỉnh mạnh nhất. Chỉ số tương ứng với biên độ tối đa tương ứng với ước tính chênh lệch thời gian giữa kênh bên trái và bên phải (ITD). Nếu biên độ của

mức tối đa thấp hơn ngưỡng đã cho, thì ước tính của ITD không được coi là đáng tin cậy và được đặt bằng 0.

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

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

Nó yêu cầu thêm độ trễ ở bộ mã hóa, mà tối đa bằng với ITD tuyệt đối tối đa mà có thể được xử lý. Sự thay đổi của ITD theo thời gian được làm mịn bằng cửa sổ phân tích của DFT.

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

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

Cần có phần đệm 0 của các cửa sổ DFT để mô phỏng sự dịch chuyển thời gian với sự dịch chuyển tròn. Kích thước của phần đệm 0 tương ứng với ITD tuyệt đối tối đa có thể được xử lý. Trong phương án được ưu tiên, phần đệm 0 được phân chia đồng đều trên cả hai phía của cửa sổ phân tích, bằng cách thêm 3,125 ms số 0 ở cả hai đầu. ITD tối đa tuyệt đối có thể có sau đó là 6,25 ms. Trong thiết lập micrô A-B, nó tương ứng với trường hợp xấu nhất là khoảng cách tối đa giữa hai micrô khoảng 2,15 mét. Sự thay đổi trong ITD theo thời gian được làm mịn bằng cách tạo cửa sổ tổng hợp và chồng lấp cộng DFT.

Điều quan trọng là sự dịch chuyển thời gian được theo sau bởi việc tạo cửa sổ của tín hiệu được dịch chuyển. Đây là điểm khác biệt chính với mã hóa tín hiệu âm lập thể (Binaural Cue Coding - BCC) của tình trạng kỹ thuật trước đó, trong đó dịch chuyển thời gian được áp dụng trên tín hiệu được tạo cửa sổ nhưng không được tạo cửa sổ thêm

ở tầng tổng hợp. Do đó, bất kỳ thay đổi nào trong ITD theo thời gian đều tạo ra nốt đệm/tiếng lách cách nhân tạo trong tín hiệu được giải mã.

4. Tính toán các IPD và xoay kênh

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

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

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 $\beta = atan2(\sin(IPD_i[b]), \cos(IPD_i[b]) + c)$, $c = 10^{ILD_i[b]/20}$, và b là chỉ số băng tham số thuộc chỉ số tần số k . Tham số β chịu trách nhiệm phân phối lượng xoay pha giữa hai kênh trong khi làm cho pha của chúng được căn chỉnh. β phụ thuộc vào IPD mà còn là mức biên độ tương đối của các kênh, ILD. Nếu kênh có biên độ cao hơn, nó sẽ được coi là kênh dẫn và sẽ ít bị ảnh hưởng bởi sự xoay pha hơn so với kênh có biên độ thấp hơn.

5. Chênh lệch tổng và mã hóa tín hiệu bên

Sự biến đổi chênh lệch tổng được thực hiện trên phổ căn chỉnh theo thời gian và pha của hai kênh theo cách bảo toàn năng lượng 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 giữa 1/1,2 và 1,2, tức là giữa -1,58 và +1,58

dB. Giới hạn tránh giả âm khi điều chỉnh năng lượng của M và S. Cần lưu ý rằng sự bảo

toàn năng lượng này ít quan trọng hơn khi thời gian và pha đã được căn chỉnh trước. Ngoài ra, các giới hạn có thể được tăng hoặc giảm.

Tín hiệu bên S được dự báo tiếp tục 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 tìm thấy bằng cách giảm thiểu sai số bình phương trung bình (Mean Square Error - MSE) của phần dư và các ILD được luận ra bởi phương trình trước đó.

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

6. Giải mã âm lập thể

Tín hiệu giữa X và tín hiệu bên S đầu tiên được chuyển đổi sang các kênh bên trái và bên phải L và R như sau:

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

trong đó độ khuếch đại g trên 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:

$$\begin{aligned} L_i[k] &= L_i[k] + cod_gain_i \cdot S_i[k], \text{ cho } 0 \leq k < band_limits[cod_max_band], \\ R_i[k] &= R_i[k] - cod_gain_i \cdot S_i[k], \text{ cho } 0 \leq k < band_limits[cod_max_band], \end{aligned}$$

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

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

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

$$\begin{aligned} L_i[k] &= a \cdot e^{j2\pi\beta} \cdot L_i[k] \\ R_i[k] &= a \cdot e^{j2\pi\beta - \text{IPD}_i[b]} \cdot R_i[k] \end{aligned}$$

trong đó:

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

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

Cuối cùng, các kênh được dịch chuyển theo thời gian hoặc trong miền tần số tùy thuộc vào các ITD được truyền. Các kênh miền thời gian được tổng hợp bằng các DFT nghịch đảo và chồng lấp – cộng.

Các dấu hiệu cụ thể của sáng chế liên quan đến tổ hợp của các tín hiệu không gian và mã hóa âm lập thể kết hợp chênh lệch tổng. Cụ thể, các tín hiệu không gian IDT và IPD được tính và áp dụng trên các kênh âm lập thể (bên trái và bên phải). Hơn nữa, 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 không gian băng rộng và băng hẹp được tổ hợp với nhau bằng mã hóa âm lập thể kết hợp chênh lệch tổng. Cụ thể, tín hiệu bên được dự báo với tín hiệu giữa bằng cách sử dụng ít nhất một tín hiệu không gian như ILD và chênh lệch tổng nghịch đảo được tính toán để nhận được các kênh bên trái và phải, ngoài ra, các tín hiệu không gian băng rộng và băng hẹp được áp dụng trên các kênh bên trái và bên phải.

Tốt hơn là bộ mã hóa có cửa sổ và chồng lấp – cộng đối với các kênh được căn chỉnh theo thời gian sau khi xử lý bằng cách sử dụng ITD. Hơn nữa, bộ giải mã cũng có hoạt động tạo cửa sổ và chồng lấp – cộng của các phiên bản được dịch chuyển hoặc được khử căn chỉnh của các kênh sau khi áp dụng chênh lệch thời gian liên kênh.

Tính toán chênh lệch thời gian liên kênh với phương pháp GCC-Phat là phương pháp đặc biệt hiệu quả.

Quy trình mới là tình trạng kỹ thuật ưu việt vì đạt được mã hóa tốc độ bit của âm lập thể hoặc âm thanh đa kênh ở độ trễ thấp. Nó được thiết kế đặc biệt để phù hợp với các bản chất khác nhau của tín hiệu đầu vào và các thiết lập khác nhau của ghi âm đa kênh hoặc âm lập thể. Cụ thể, sáng chế này cung cấp chất lượng tốt để mã hóa tiếng nói âm lập thể tốc độ bit thấp.

Các quy trình được ưu tiên được sử dụng trong việc phân phối phát thanh tất cả các loại nội dung âm thanh âm lập thể hoặc đa kênh như tiếng nói và âm nhạc với chất lượng cảm nhận không đổi ở tốc độ bit thấp đã cho. Các lĩnh vực ứng dụng như vậy là đài kỹ thuật số, truyền trực tuyến qua internet hoặc các ứng dụng truyền thông âm thanh.

Mặc dù sáng chế này đã được mô tả theo một số phương án, nhưng có những thay đổi, hoán vị và tương đương nằm trong phạm vi của sáng chế này. Cũng cần lưu ý rằng có nhiều cách thay thế để thực hiện các phương pháp và thành phần của sáng chế. Do đó, các yêu cầu bảo hộ đi kèm sau đây được hiểu là bao gồm tất cả các thay đổi, hoán vị và tương đương như vậy là nằm trong nguyên lý và phạm vi thực sự của sáng chế.

Mặc dù một số khía cạnh đã được mô tả trong ngữ cảnh của thiết bị, nhưng rõ ràng những khía cạnh này cũng thể hiện mô tả của phương pháp tương ứng, trong đó 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 thể hiện mô tả về khối hoặc mục hoặc dấu hiệu tương ứng của thiết bị tương ứng. Một số hoặc tất cả các bước của phương pháp có thể được thực hiện bằng (hoặc sử dụng) thiết bị phần cứng, chẳng hạn như bộ vi xử lý, máy tính có thể lập trình hoặc mạch điện tử. Theo một số phương án, một số hoặc nhiều bước phương pháp quan trọng nhất có thể được thực hiện bởi thiết bị như vậy.

Tín hiệu hình ảnh được mã hóa theo sáng chế có thể được lưu trữ trên vật ghi lưu trữ số hoặc có thể được truyền trên vật ghi truyền dẫn như vật ghi truyền dẫn không dây hoặc vật ghi truyền dẫn có dây như Internet.

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

Một số phương án theo sáng chế bao gồm vật mang dữ liệu có 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 được, sao cho một trong các phương pháp được mô tả trong tài liệu này được thực hiện.

Nói chung, các phương án theo sáng chế có thể được thực hiện dưới dạng sản phẩm chương trình máy tính có mã chương trình, mã chương trình có thể hoạt động để thực hiện một trong các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Ví dụ, mã chương trình có thể được lưu trữ trên vật mang có thể đọc được bằng máy.

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

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

Do đó, phương án khác của 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) bao gồm, được ghi lại trên đó, chương trình máy tính để thực hiện một trong các phương pháp được mô tả trong tài liệu này. Vật mang dữ liệu, vật ghi lưu trữ số hoặc vật ghi được ghi lại thường là hữu hình và/hoặc không tạm thời.

Do đó, phương án khác của phương pháp theo sáng chế là dòng dữ liệu hoặc chuỗi các tín hiệu biểu diễn chương trình máy tính để thực hiện một trong các phương

pháp được mô tả trong tài liệu này. Ví dụ, dòng dữ liệu hoặc chuỗi tín hiệu có thể được tạo cấu hình để truyền qua kết nối truyền dữ liệu, ví dụ qua Internet.

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

Phương án khác bao gồm máy tính đã được cài đặt trên đó chương trình máy tính để thực hiện một trong các phương pháp được mô tả trong tài liệu này.

Phương án khác theo sáng chế bao gồm thiết bị hoặc hệ thống được tạo cấu hình để truyền (ví dụ, bằng điện tử hoặc quang học) chương trình máy tính để thực hiện một trong các phương pháp được mô tả trong tài liệu này tới bộ nhận. Ví dụ, bộ nhận có thể là máy tính, thiết bị di động, thiết bị ghi nhớ hoặc những thứ tương tự. Ví dụ, thiết bị hoặc hệ thống có thể bao gồm máy chủ tệp để truyền chương trình máy tính đến máy nhận.

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

Thiết bị được mô tả trong tài liệu này có thể được thực hiện bằng cách sử dụng thiết bị phần cứng, hoặc sử dụng máy tính, hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

Các phương pháp được mô tả trong tài liệu này có thể được thực hiện bằng cách sử dụng thiết bị phần cứng hoặc sử dụng máy tính, hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

Các phương án được mô tả ở trên chỉ minh họa cho các nguyên lý của sáng chế. Điều này được hiểu rằng các sửa đổi và biến thể của cách sắp xếp cũng như các chi tiết được mô tả trong tài liệu này sẽ rõ ràng đối với những người có trình độ trung bình trong lĩnh vực kỹ thuật tương ứng. Do đó, mục đích chỉ bị giới hạn bởi phạm vi của các yêu

cầu bảo hộ sáng chế kèm theo đây và không phải bởi các chi tiết cụ thể được trình bày bằng cách mô tả và giải thích của các phương án trong tài liệu này.

YÊU CẦU BẢO HỘ

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

bộ phân tích tín hiệu (1037) để ước tính đặc tính tín hiệu (1038) của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai hoặc cả hai tín hiệu hoặc tín hiệu được suy ra từ tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai;

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 gán trọng số (1036) để gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn để thu được phổ tương quan chéo được gán trọng số bằng cách sử dụng quy trình gán trọng số thứ nhất (1036a) hoặc sử dụng quy trình gán trọng số thứ hai (1036b) tùy thuộc vào đặc tính tín hiệu được ước tính bởi bộ phân tích tín hiệu (1037), trong đó quy trình gán trọng số thứ nhất khác với quy trình gán trọng số thứ hai; và

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

2. Thiết bị theo điểm 1, trong đó bộ phân tích tín hiệu (1037) được tạo cấu hình như bộ ước tính nhiễu âm (1037) để ước tính mức nhiễu âm (1038) của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai hoặc cả hai tín hiệu hoặc tín hiệu được suy ra từ tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai, và trong đó đặc tính tín hiệu thứ nhất là mức nhiễu âm thứ nhất và đặc tính tín hiệu thứ hai là mức nhiễu âm thứ hai, hoặc trong đó bộ phân tích tín hiệu (1037) được tạo cấu hình để thực hiện phân tích tiếng nói/âm nhạc, phân tích loa giao thoa, phân tích âm nhạc nền, phân tích tiếng nói sạch hoặc bất kỳ phân tích tín hiệu nào khác để xác định, liệu tín hiệu có đặc tính thứ nhất hay đặc tính thứ hai.

3. Thiết bị theo điểm 1, trong đó quy trình gán trọng số thứ nhất (1036a) được chọn cho đặc tính tín hiệu thứ nhất và quy trình gán trọng số thứ hai (1036b) được chọn cho đặc tính tín hiệu thứ hai, và trong đó đặc tính tín hiệu thứ nhất khác với đặc tính tín hiệu thứ hai.

4. Thiết bị theo một trong số các điểm nêu trên, trong đó quy trình gán trọng số thứ nhất (1036a) bao gồm việc gán trọng số sao cho biên độ được chuẩn hóa và pha được duy trì, hoặc trong đó quy trình gán trọng số thứ hai (1036b) bao gồm hệ số gán trọng số được suy ra từ phổ tương quan chéo được làm mịn hoặc không được làm mịn bằng cách sử dụng phép tính lũy thừa có lũy thừa thấp hơn 1 hoặc lớn hơn 0 hoặc hàm log.

5. Thiết bị theo một trong số các điểm nêu trên, trong đó quy trình gán trọng số thứ hai (1036b) bao gồm việc gán trọng số sao cho biên độ được chuẩn hóa và pha được duy trì và ngoài ra còn bao gồm hệ số gán trọng số được suy ra từ phổ tương quan chéo được làm mịn hoặc không được làm mịn bằng cách sử dụng phép tính lũy thừa có lũy thừa thấp hơn 1 hoặc lớn hơn 0 hoặc giữa 0,79 và 0,82.

6. Thiết bị theo một trong số các điểm nêu trên, trong đó quy trình gán trọng số thứ nhất (1036a) hoạt động theo phương trình sau:

$$\tilde{C}_{PHAT}(k, s) = \frac{\tilde{C}(k, s)}{|\tilde{C}(k, s)|}, \text{ hoặc}$$

trong đó quy trình gán trọng số thứ hai (1036b) hoạt động theo phương trình sau:

$$\tilde{C}_{MCSP}(k, s) = \frac{\tilde{C}(k, s)}{|\tilde{C}(k, s)|^\rho},$$

trong đó $\tilde{C}_{PHAT}(k, s)$ là giá trị phổ tương quan chéo được làm mịn hoặc không được làm mịn được gán trọng số cho chỉ số tần số k và chỉ số thời gian s thu được bằng cách áp dụng quy trình gán trọng số thứ nhất,

trong đó $\tilde{C}_{MCSP}(k, s)$ là giá trị phổ tương quan chéo được làm mịn hoặc không được làm mịn được gán trọng số cho chỉ số tần số k và chỉ số thời gian s thu được bằng cách áp dụng quy trình gán trọng số thứ hai,

trong đó $\tilde{C}(k, s)$ là giá trị phổ tương quan chéo được làm mịn hoặc không được làm mịn cho chỉ số tần số k và chỉ số thời gian s , và

trong đó ρ là giá trị lũy thừa khác 1.

7. Thiết bị theo một trong số các điểm nêu trên, trong đó quy trình gán trọng số thứ hai (1036b) bao gồm việc chuẩn hóa sao cho phạm vi đầu ra của quy trình chuẩn hóa thứ hai nằm trong phạm vi trong đó phạm vi đầu ra của quy trình chuẩn hóa thứ nhất được

xác định vị trí, hoặc sao cho phạm vi đầu ra của quy trình chuẩn hóa thứ hai giống với phạm vi đầu ra của quy trình chuẩn hóa thứ nhất.

8. Thiết bị theo một trong số các điểm nêu trên, trong đó quy trình gán trọng số thứ hai (1036b) bao gồm việc chuẩn hóa dựa trên phương trình sau:

$$\tilde{C}_{MCSP}(k, s) = \frac{\tilde{C}_{MCSP}(k, s)}{\frac{1}{N_{DFT}} \sum_{k=0}^{N_{DFT}-1} |\tilde{C}_{MCSP}(k, s)|}.$$

9. Thiết bị theo một trong số các điểm nêu trên, trong đó bộ xử lý (1040) được tạo cấu hình để thực hiện phép tính chọn đỉnh thứ nhất (1041) hoặc phép tính chọn đỉnh thứ hai (1042) tùy thuộc vào liệu quy trình gán trọng số thứ nhất (1036a) hay quy trình gán trọng số thứ hai (1036b) đã được sử dụng, trong đó phép tính chọn đỉnh thứ nhất khác với phép tính chọn đỉnh thứ hai.

10. Thiết bị theo một trong số các điểm nêu trên, trong đó phép tính chọn đỉnh thứ hai (1042) được sử dụng khi quy trình gán trọng số thứ hai được sử dụng, và trong đó phép tính chọn đỉnh thứ hai (1042) được tạo cấu hình để áp dụng ngưỡng thứ hai thấp hơn ngưỡng thứ nhất được sử dụng bởi phép tính chọn đỉnh thứ nhất (1041).

11. Thiết bị theo một trong số các điểm từ 2 đến 10,

trong đó bộ ước tính nhiễu âm (1037) được tạo cấu hình để ước tính (1060) mức nhiễu âm nền hoặc được tạo cấu hình để làm mịn (1061) mức nhiễu âm được ước tính theo thời gian hoặc được tạo cấu hình để sử dụng bộ lọc làm mịn IIR.

12. Thiết bị theo một trong số các điểm từ 2 đến 11,

trong đó bộ ước tính nhiễu âm (1037) còn bao gồm bộ phát hiện hoạt động tín hiệu (1070) để phân loại khối thời gian là hoạt động hay không hoạt động, trong đó bộ ước tính nhiễu âm (1037) được tạo cấu hình để tính (1072) mức tín hiệu bằng cách sử dụng một hoặc nhiều khối thời gian hoạt động, hoặc trong đó bộ ước tính nhiễu âm (1037) được tạo cấu hình để tạo tín hiệu (1050) mức nhiễu âm nền cao, khi tỷ lệ tín hiệu trên nhiễu âm dưới ngưỡng, ngưỡng này nằm trong phạm vi giữa 45 đến 25 dB.

13. Thiết bị theo một trong số các điểm từ 1 đến 12, thiết bị này còn bao gồm:

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

bộ lọc làm mịn (1030) để làm mịn phổ tương quan chéo theo thời gian bằng cách sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm mịn, và trong đó bộ gán trọng số (1036) được tạo cấu hình để gán trọng số phổ tương quan chéo được làm mịn.

14. Thiết bị theo một trong số các điểm từ 1 đến 13,

trong đó bộ xử lý (1040) được tạo cấu hình để chuẩn hóa (1036a) phổ tương quan chéo được làm mịn bằng cách sử dụng cường độ của phổ tương quan chéo được làm mịn.

15. Thiết bị theo một trong số các điểm từ 1 đến 14,

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

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

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

16. Thiết bị theo một trong số các điểm nêu trên,

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

17. Thiết bị theo một trong số các điểm nêu trên,

trong đó bộ xử lý (1040) được tạo cấu hình để thực hiện xác định chênh lệch thời gian liên kênh bằng cách thực hiện phép tính tìm kiếm đỉnh hoặc chọn đỉnh (1041, 1042) trong biểu diễn miền thời gian được xác định từ phổ tương quan chéo được làm mịn.

18. Thiết bị theo một trong số các điểm từ 13 đến 17,

trong đó bộ ước tính đặc tính phổ (1010) được tạo cấu hình để xác định, như đặc tính phổ, độ nhiễu âm hoặc âm sắc của phổ; và

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

việc làm mịn yếu hơn theo thời gian với mức độ làm mịn thứ hai trong trường hợp đặc tính nhiều âm thứ hai nhiều nhiều âm hơn hoặc đặc tính âm sắc thứ hai ít âm sắc hơn,

trong đó mức độ làm mịn thứ nhất lớn hơn mức độ làm mịn thứ hai, và trong đó đặc tính nhiều âm thứ nhất ít nhiều âm hơn so với đặc tính nhiều âm thứ hai, hoặc đặc tính âm sắc thứ nhất nhiều âm sắc hơn so với đặc tính âm sắc thứ hai.

19. Thiết bị theo một trong số các điểm từ 13 đến 18,

trong đó bộ ước tính đặc tính phổ (1010) 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 chọn giá trị lớn nhất, bằng cách xác định giá trị trung bình được gán trọng số hoặc giá trị trung bình không được gán trọng số giữa các số đo độ phẳng phổ, hoặc bằng cách chọn giá trị nhỏ nhất.

20. Thiết bị theo một trong số các điểm từ 13 đến 19,

trong đó bộ lọc làm mịn (1030) được tạo cấu hình để tính toán giá trị phổ tương quan chéo được làm mịn 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 trong quá khứ, trong đó các hệ số gán trọng số cho tổ hợp được gán trọng số được xác định bởi đặc tính của phổ.

21. Thiết bị theo một trong số các điểm nêu trên,

trong đó bộ xử lý (1040) được tạo cấu hình để xác định (1120) phạm vi hợp lệ và phạm vi không hợp lệ trong biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm mịn hoặc không được làm mịn được gán trọng số,

trong đó ít nhất một đỉnh tối đa trong phạm vi không hợp lệ được phát hiện (1121) và được so sánh (1123) với đỉnh tối đa trong phạm vi hợp lệ, trong đó chênh lệch thời gian liên kênh chỉ được xác định (1124), khi đỉnh tối đa trong phạm vi hợp lệ lớn hơn ít nhất một đỉnh tối đa trong phạm vi không hợp lệ.

22. Thiết bị theo một trong số các điểm nêu trên,

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

để thực hiện (1102) phép tính tìm kiếm đỉnh trong biểu diễn miền thời gian được suy ra từ phổ tương quan chéo được làm mịn,

để xác định (1105) biến thiên của ngưỡng cố định từ biểu diễn miền thời gian; và

để so sánh (1106, 1035) đỉnh với ngưỡng biến thiên, trong đó chênh lệch thời gian liên kênh được xác định là độ trễ thời gian liên quan đến đỉnh trong mỗi quan hệ xác định trước với ngưỡng biến thiên.

23. Thiết bị theo điểm 22,

trong đó bộ xử lý (1040) được tạo cấu hình để xác định (1105) ngưỡng biến thiên dưới dạng giá trị bằng bội số nguyên của giá trị trong phần rộng nhất chẳng hạn như 10% của các giá trị của biểu diễn miền thời gian.

24. Thiết bị theo một trong số các điểm từ 1 đến 23,

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

trong đó bộ xử lý (1040) được tạo cấu hình để tính toán (1105, 1034) ngưỡng biến thiên dựa trên cường độ đỉnh trung bình được suy ra (1103) từ các cường độ đỉnh tối đa của nhiều khối con, và

trong đó bộ xử lý (1140) được tạo cấu hình để xác định chênh lệch thời gian liên kênh dưới dạng giá trị độ trễ thời gian tương ứng với đỉnh tối đa của nhiều khối con lớn hơn ngưỡng biến thiên.

25. Thiết bị theo điểm 24,

trong đó bộ xử lý (1140) được tạo cấu hình để tính toán (1105) ngưỡng biến thiên bằng phép nhân 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 bởi đặc tính tỷ lệ tín hiệu trên nhiễu âm (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.

26. Thiết bị theo điểm 25,

trong đó bộ xử lý 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 tối đa thấp hơn giá trị được xác định trước.

27. Thiết bị theo một trong số các điểm từ 2 đến 26, trong đó bộ ước tính nhiễu âm (1037) bao gồm bộ ước tính nhiễu âm nền (1060) và bộ làm mịn thời gian (1061) để cung cấp ước tính nhiễu âm nền, hoặc

trong đó bộ ước tính nhiễu âm (1037) bao gồm bộ phát hiện hoạt động tín hiệu (1070), bộ chọn khung (1071) để chọn khung hoạt động chỉ dưới sự điều khiển của bộ phát hiện hoạt động tín hiệu (1070) và bộ tính toán mức tín hiệu (1072) để tính toán mức tín hiệu trong khung hoạt động và bộ làm mịn thời gian (1073) để làm mịn kết quả của bộ tính toán mức tín hiệu (1072) theo thời gian để cung cấp ước tính mức tín hiệu, hoặc

trong đó bộ ước tính nhiễu âm (1037) được tạo cấu hình để tính toán (1074) tỷ lệ tín hiệu trên nhiễu âm từ mức tín hiệu được làm mịn hoặc không được làm mịn và mức nhiễu âm nền được làm mịn hoặc không được làm mịn cho khung, và bộ so sánh (1075) để so sánh giá trị tỷ lệ tín hiệu trên nhiễu âm với ngưỡng cho khung để cung cấp mức nhiễu âm (1038, 1050) cho khung.

28. Thiết bị theo một trong số các điểm nêu trên, trong đó thiết bị được tạo cấu hình

để thực hiện lưu trữ hoặc truyền chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện xử lý âm lập thể hoặc đa kênh hoặc mã hóa các tín hiệu kênh thứ nhất và thứ hai bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện căn chỉnh thời gian của hai tín hiệu kênh bằng cách sử dụng chênh lệch thời gian liên kênh, hoặc

để thực hiện chênh lệch thời gian của ước tính đến bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện chênh lệch thời gian của ước tính đến bằng cách sử dụng chênh lệch thời gian liên kênh để xác định vị trí loa trong phòng có hai micrô và thiết lập micrô đã biết, hoặc

để thực hiện điều hướng chùm sóng bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện lọc không gian bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện phân tách mặt nổi hoặc nền bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện hoạt động định vị nguồn âm thanh sử dụng chênh lệch thời gian liên kênh được ước tính, hoặc

để thực hiện sự định vị của nguồn âm thanh bằng cách sử dụng chênh lệch thời gian liên kênh được ước tính bằng cách thực hiện phép đặc tam giác âm dựa trên các chênh lệch thời gian giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai hoặc tín hiệu kênh thứ nhất, tín hiệu kênh thứ hai và ít nhất một tín hiệu bổ sung.

29. Phương pháp ước tính chênh lệch thời gian liên kênh giữa tín hiệu kênh thứ nhất và tín hiệu kênh thứ hai, phương pháp này bao gồm:

ước tính đặc tính tín hiệu của tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai hoặc cả hai tín hiệu hoặc tín hiệu được suy ra từ tín hiệu kênh thứ nhất hoặc tín hiệu kênh thứ hai;

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;

gán trọng số phổ tương quan chéo được làm mịn hoặc không được làm mịn để thu được phổ tương quan chéo được gán trọng số bằng cách sử dụng quy trình gán trọng số thứ nhất hoặc sử dụng quy trình gán trọng số thứ hai tùy thuộc vào đặc tính tín hiệu được ước tính, trong đó quy trình gán trọng số thứ nhất khác với quy trình gán trọng số thứ hai; và

xử lý phổ tương quan chéo được gán trọng số để thu được chênh lệch thời gian liên kênh.

30. Phương pháp theo điểm 29, phương pháp này còn bao gồm:

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

làm mịn phổ tương quan chéo theo thời gian bằng cách sử dụng đặc tính phổ để thu được phổ tương quan chéo được làm mịn, và trong đó việc gán trọng số được gán trọng số cho phổ tương quan chéo được làm mịn.

31. 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 29 hoặc 30.

1/25

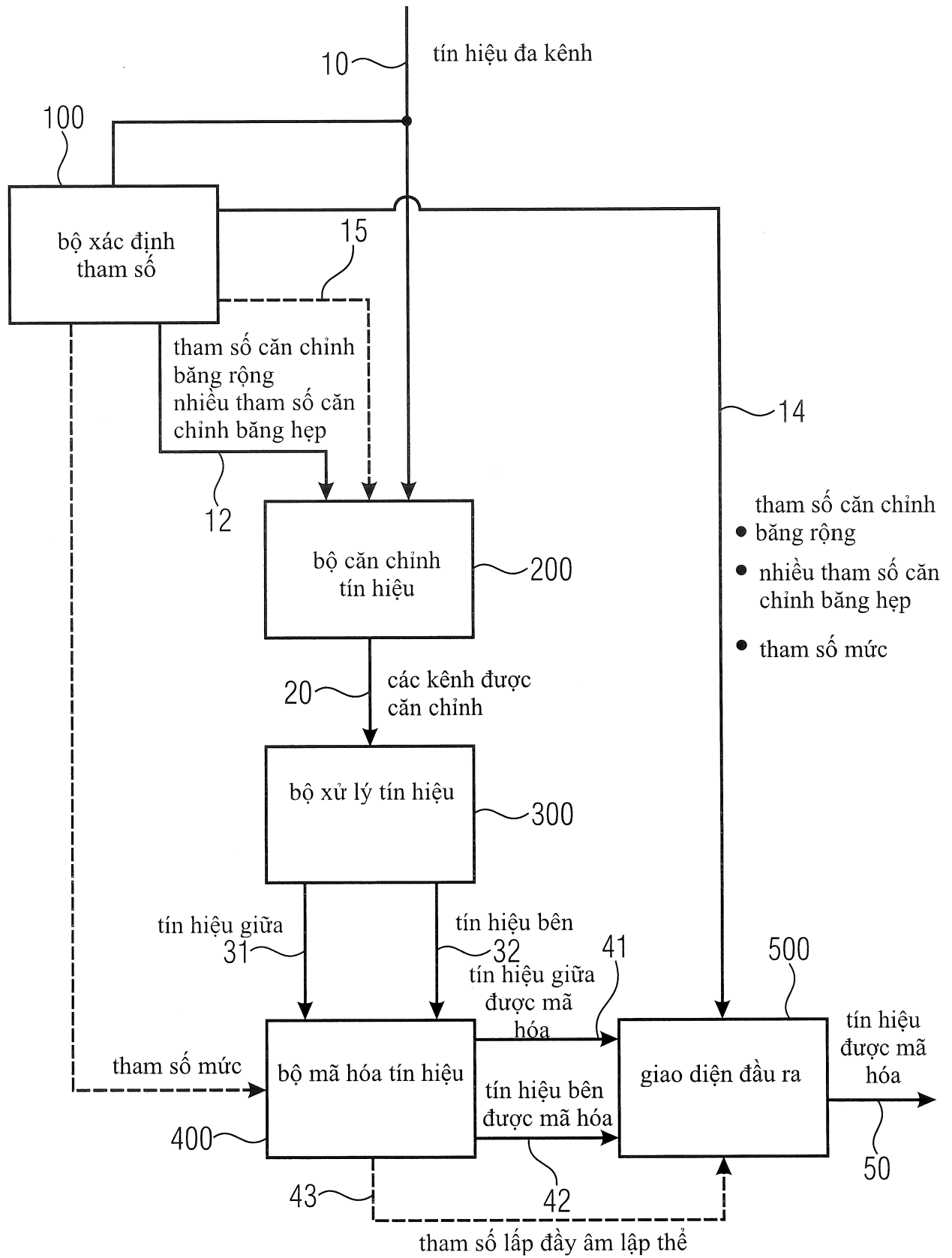


Fig. 1

2/25

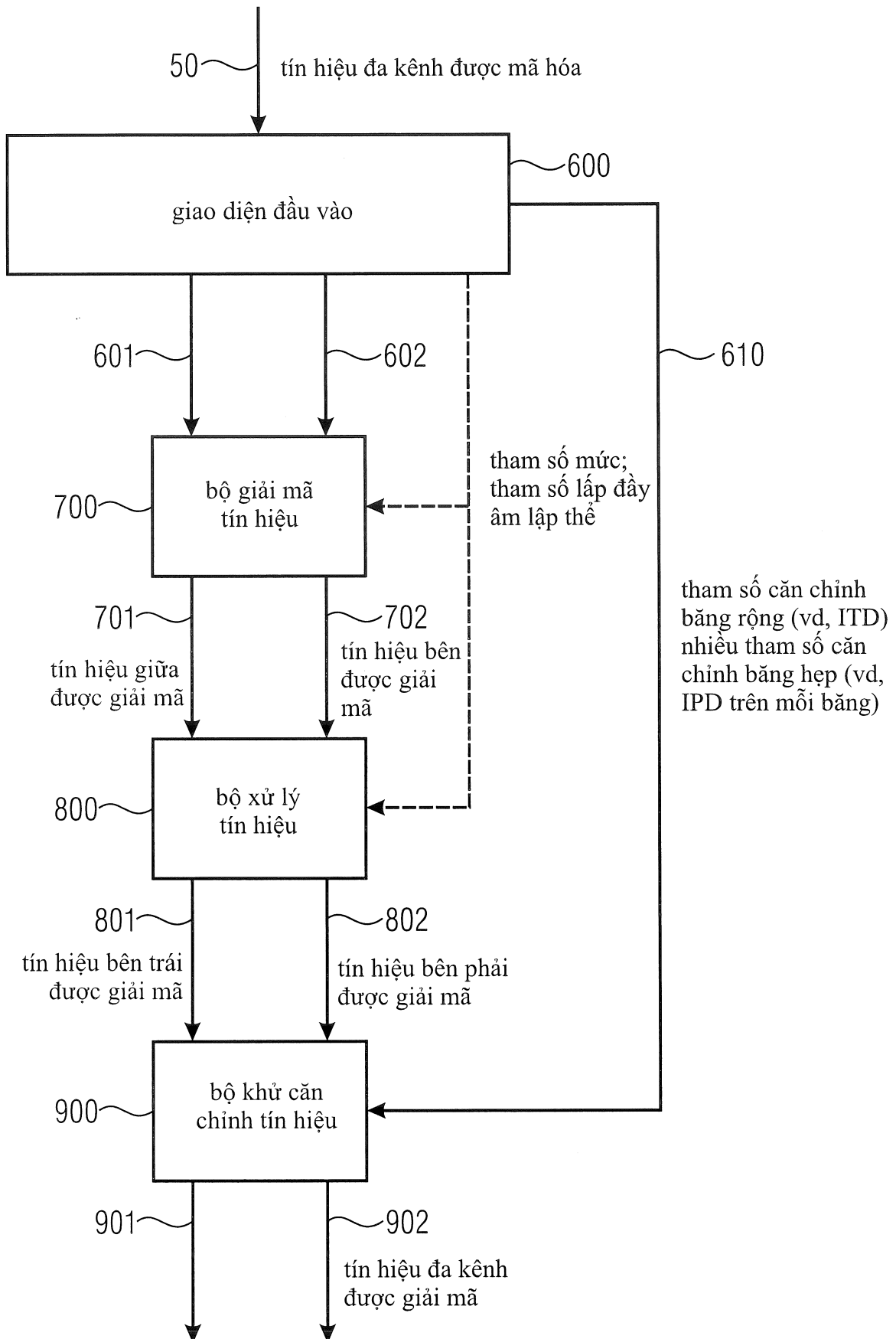
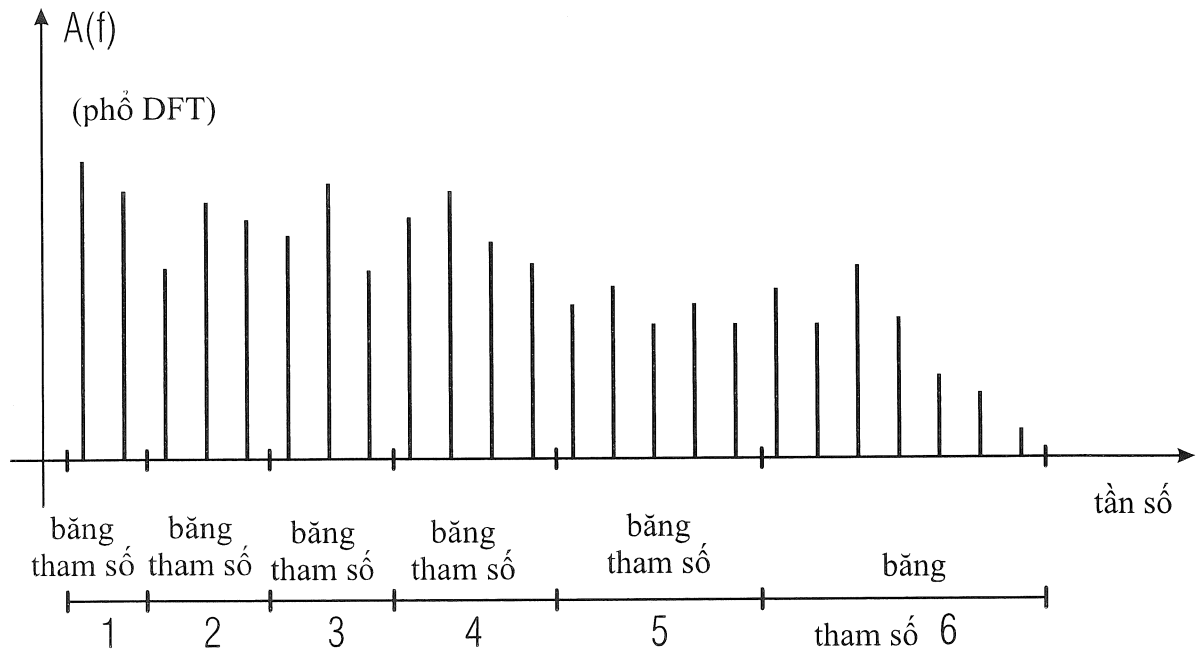


Fig. 2

3/25



- tham số căn chỉnh băng rộng đơn cho toàn bộ phổ (vd, 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, bốn tham số băng hẹp;
- các tham số mức cho mỗi băng tham số, vd, 6 tham số mức;
- các tham số lấp đầy âm lập thể cho các băng tham số 4, 5, 6, vd, ba tham số lấp đầy âm lập thể;
- tín hiệu (dư) bên cho các băng tham số 1, 2, 3;
- nhiều vạch phổ hơn 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

4/25

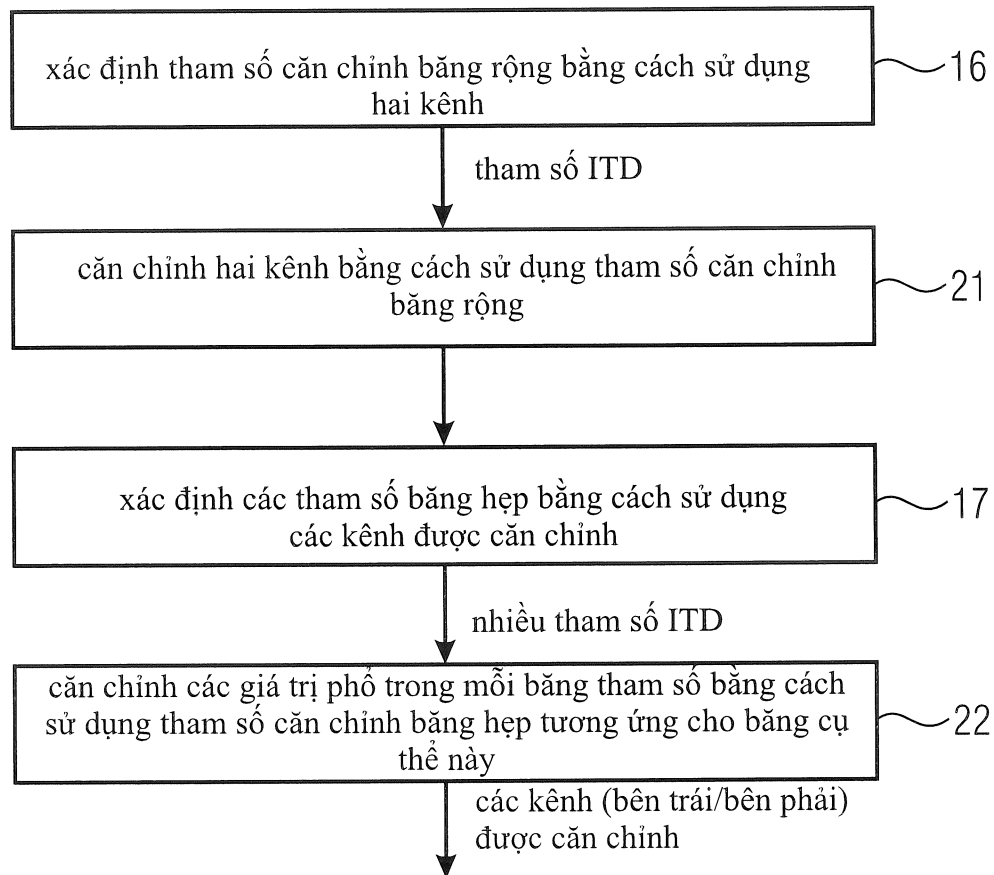


Fig. 4a

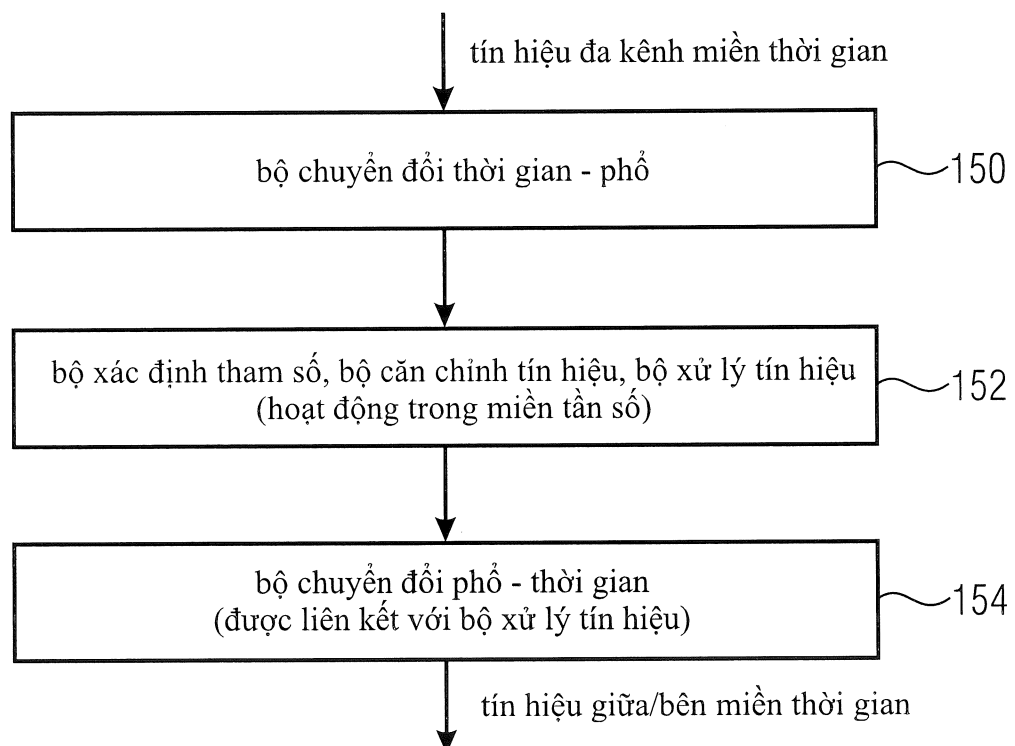


Fig. 4b

5/25

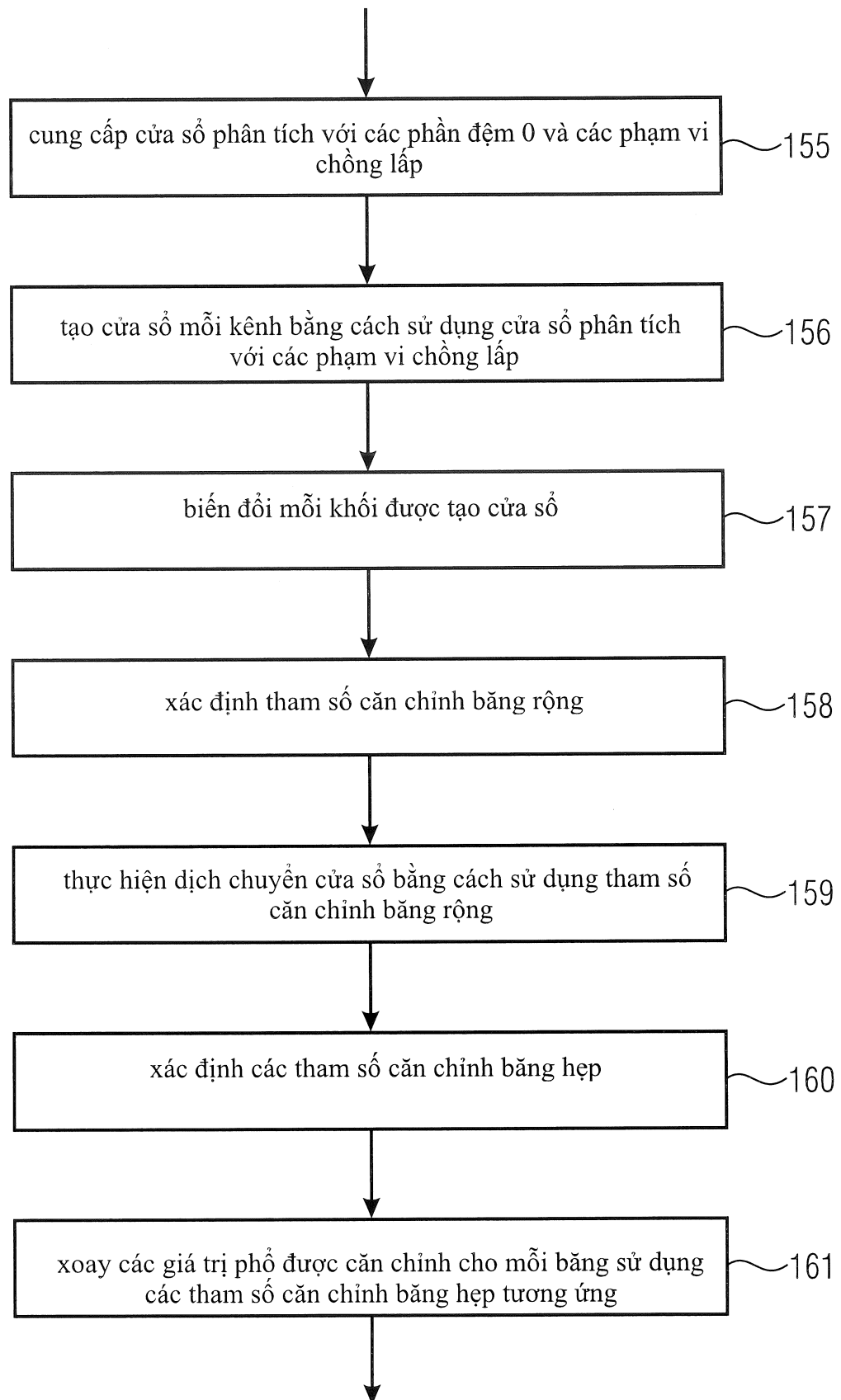


Fig. 4c

6/25

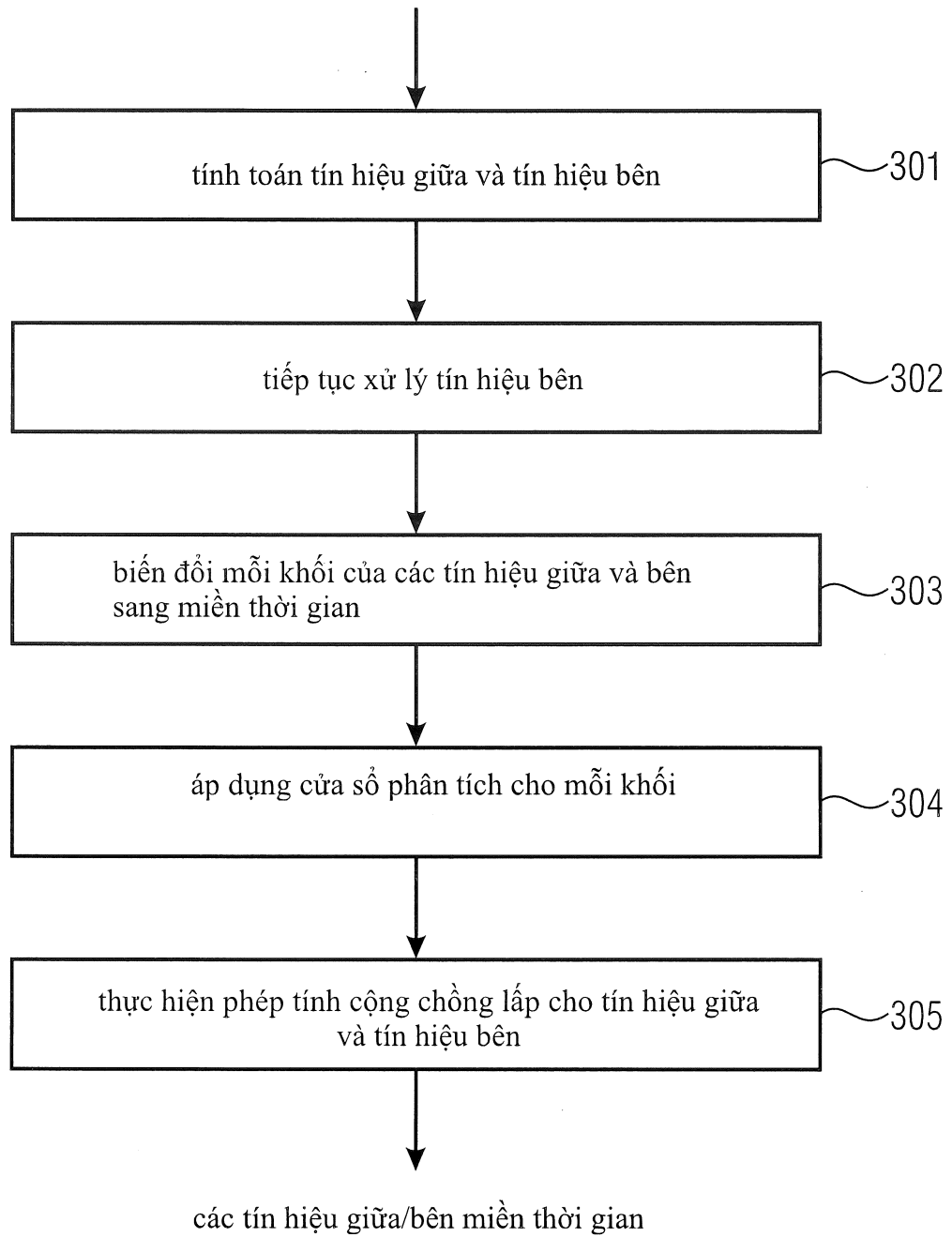
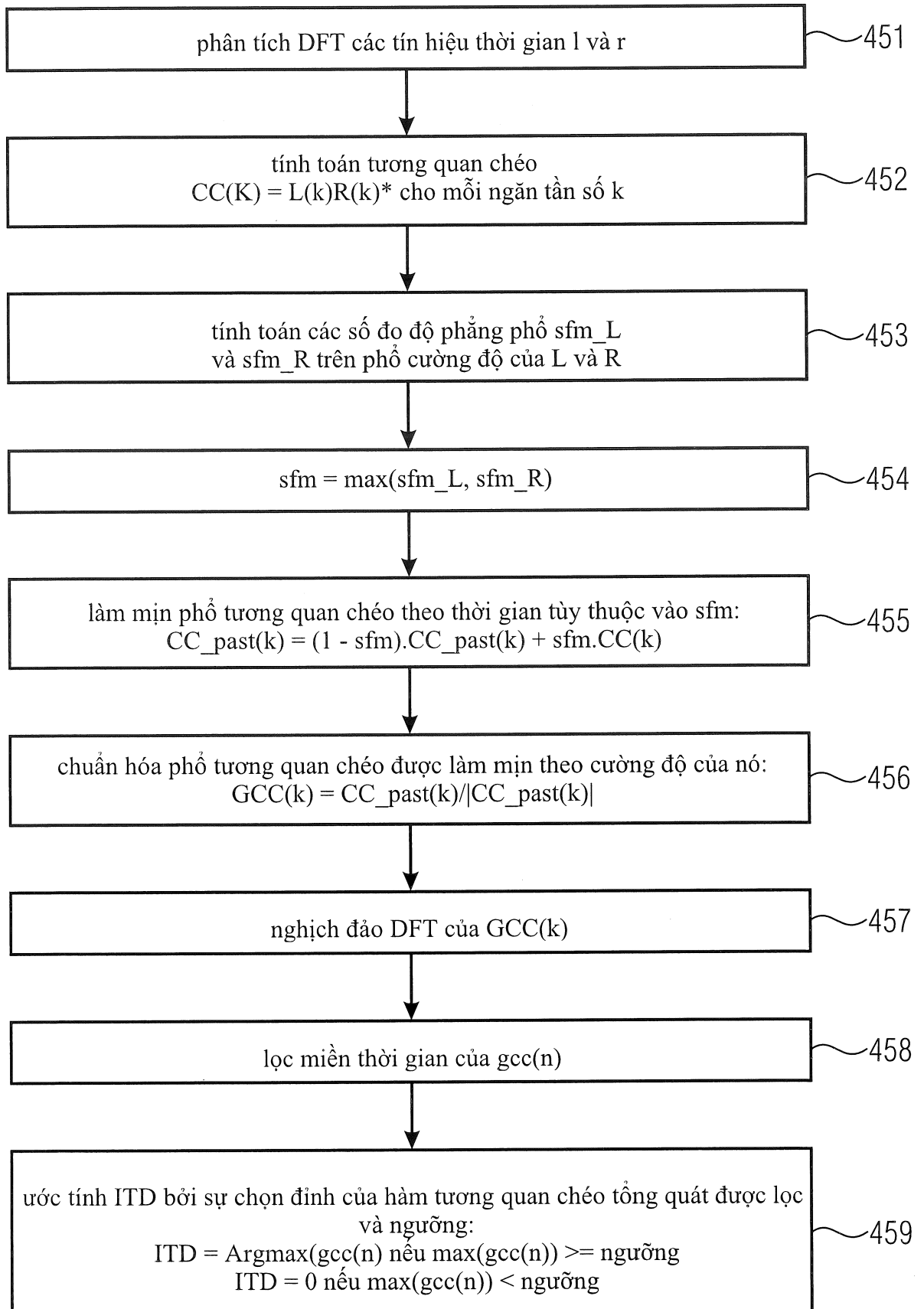


Fig. 4d

7/25



ước tính ITD

Fig. 4e

8/25

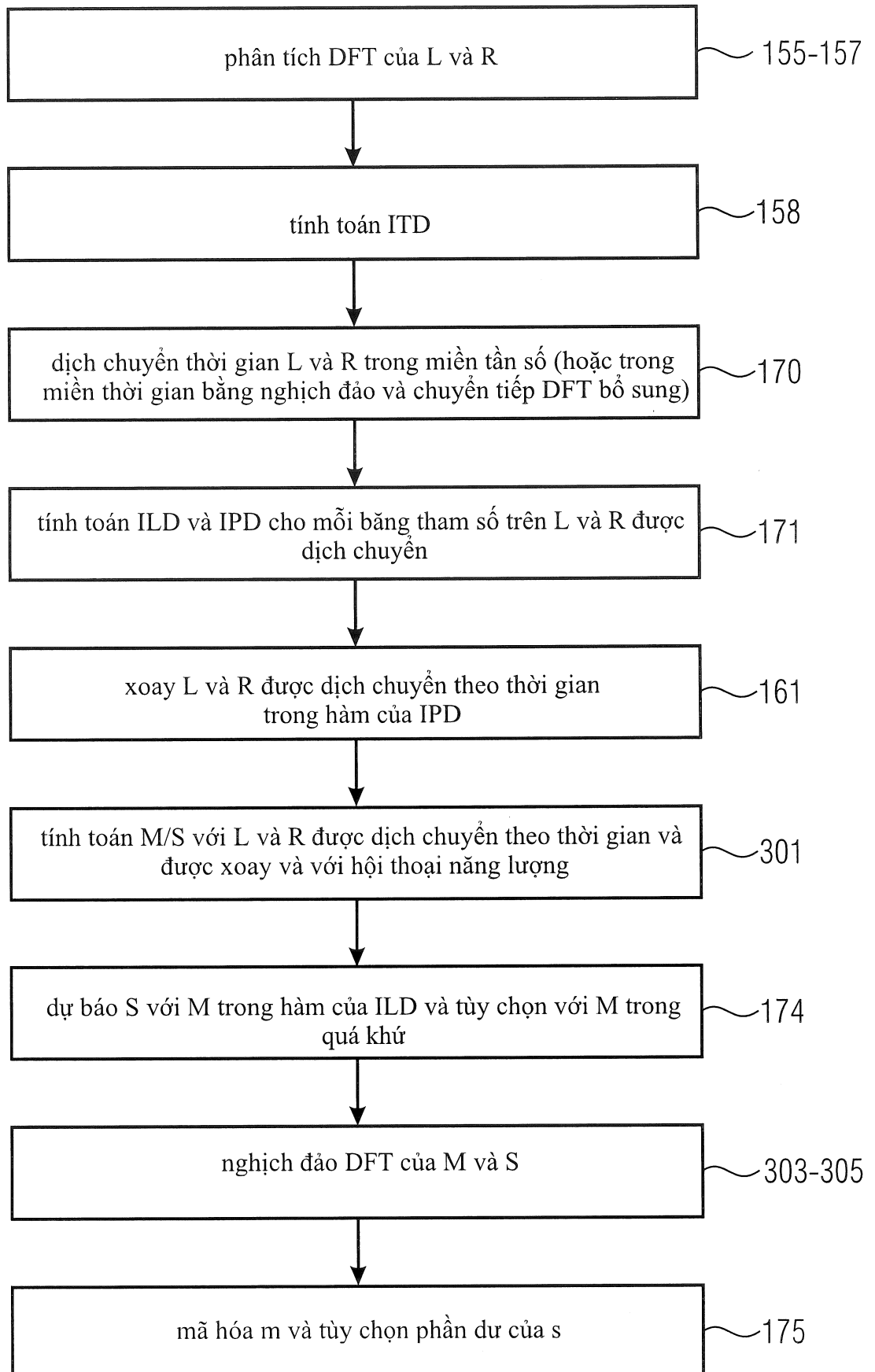


Fig. 5

9/25

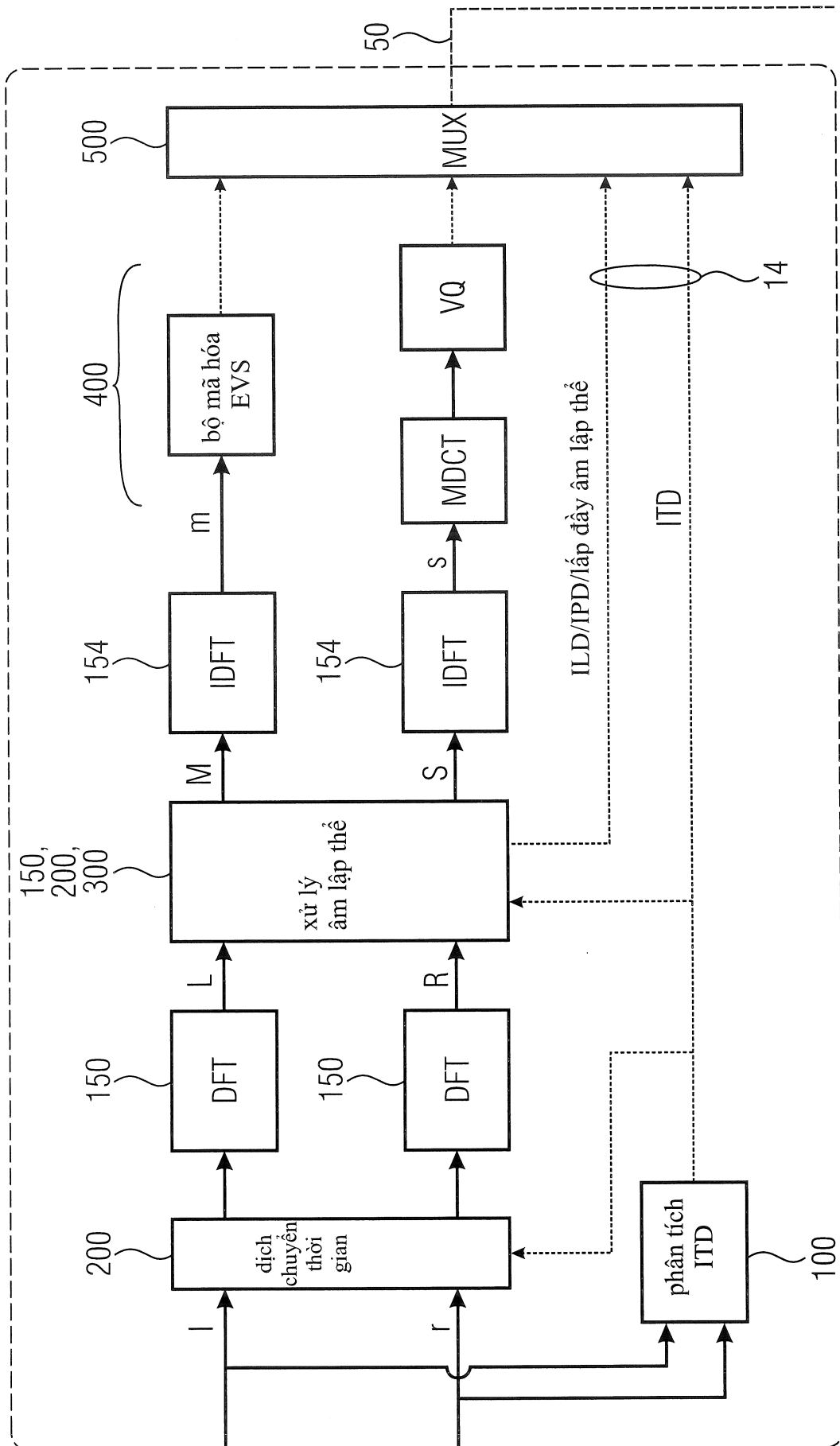


Fig. 6a

10/25

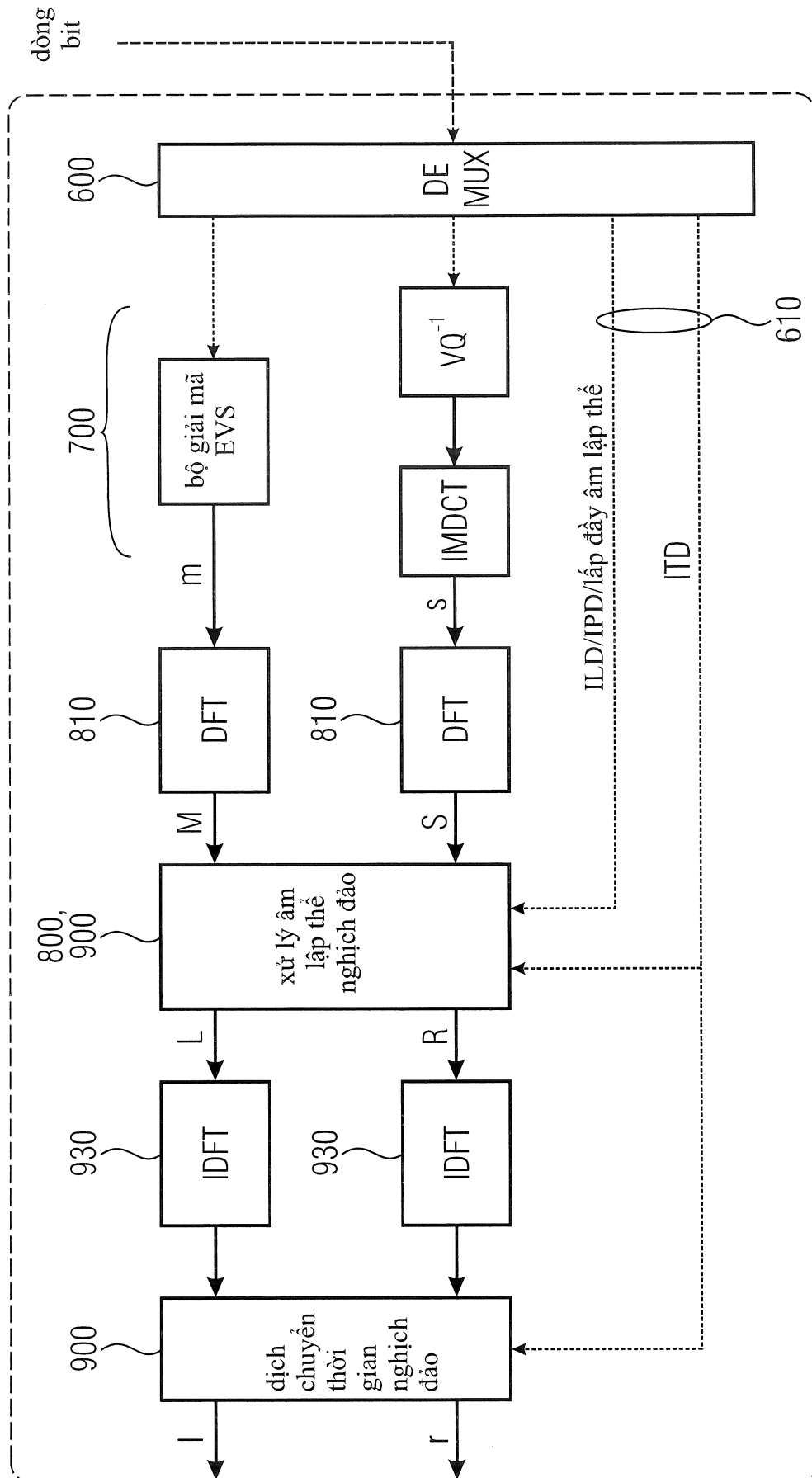


Fig. 6b

11/25

lấy mẫu 16 kHz

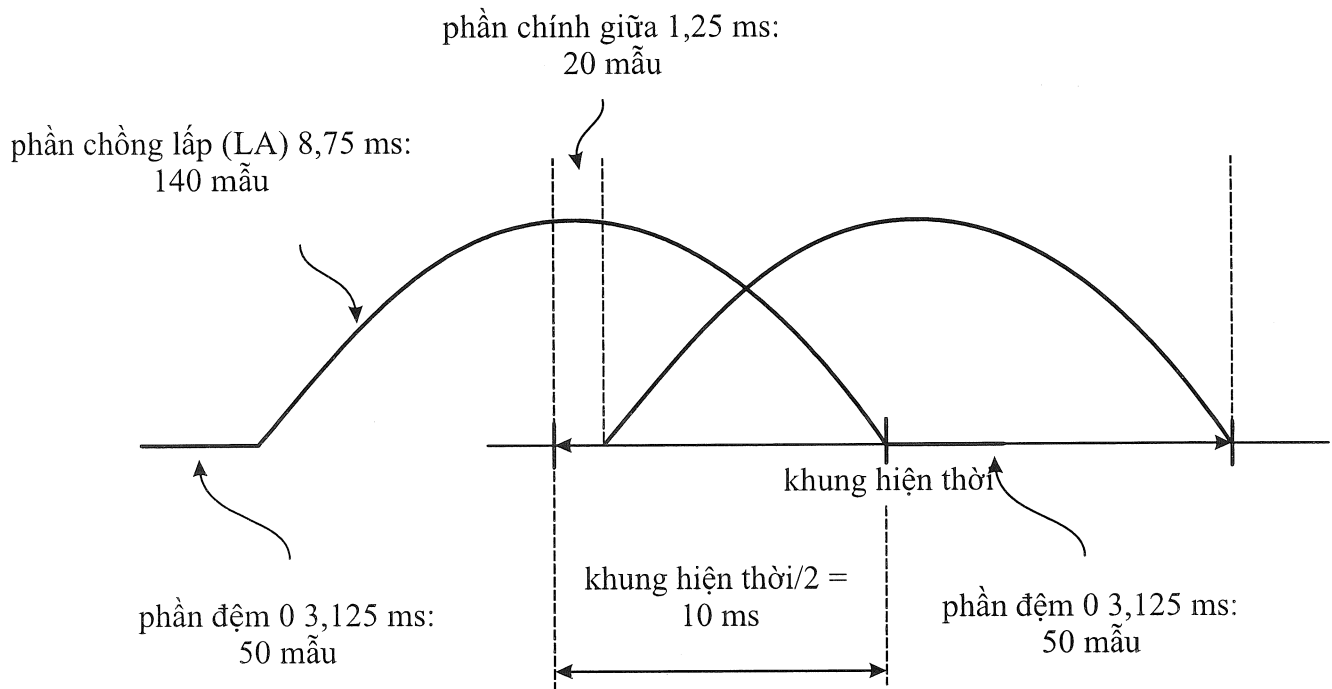


Fig. 7

các tham số	các bit/băng	số lượng băng	tốc độ bit
ILD	5	tất cả 12 bands	3.00 kbps
IPD	3	đến 2.5 kHz	1.05 kbps
ITD	8	toàn bộ phổ	0.6 kbps
lấp đầy âm lập thể	3	từ 1 kHz	0.9 kbps
tổng			-5 kbps

Fig. 8

12/25

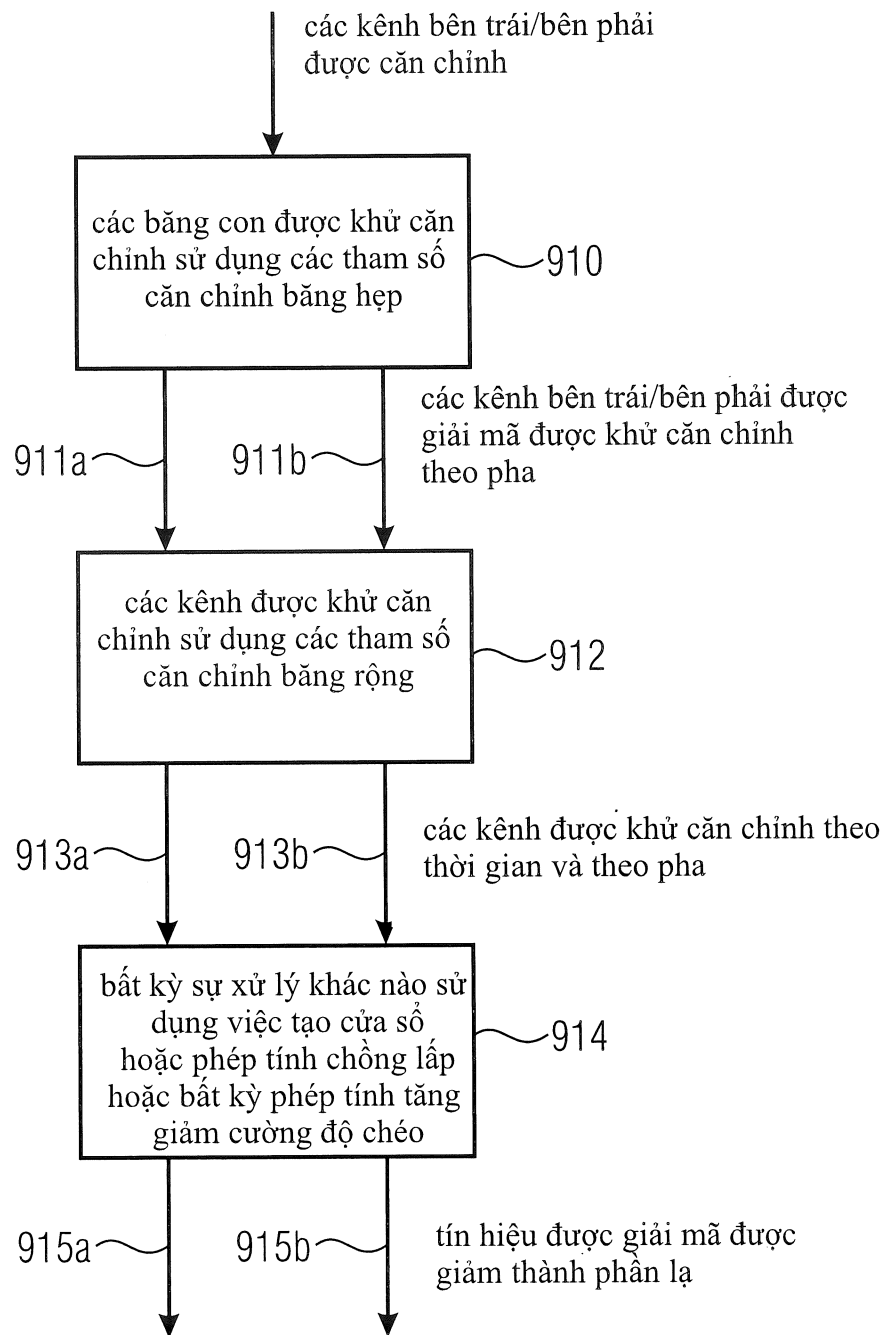


Fig. 9a

13/25

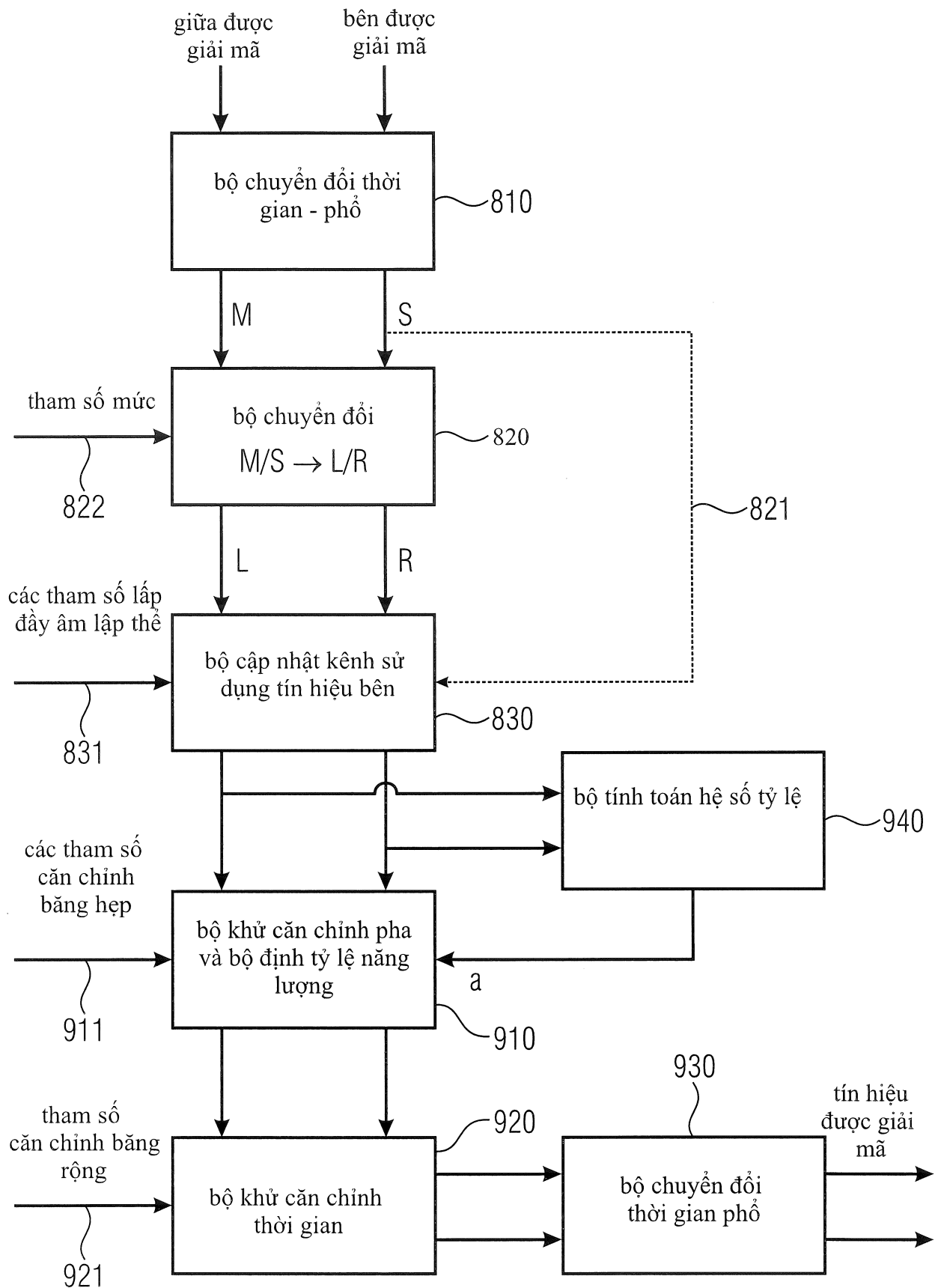


Fig. 9b

14/25

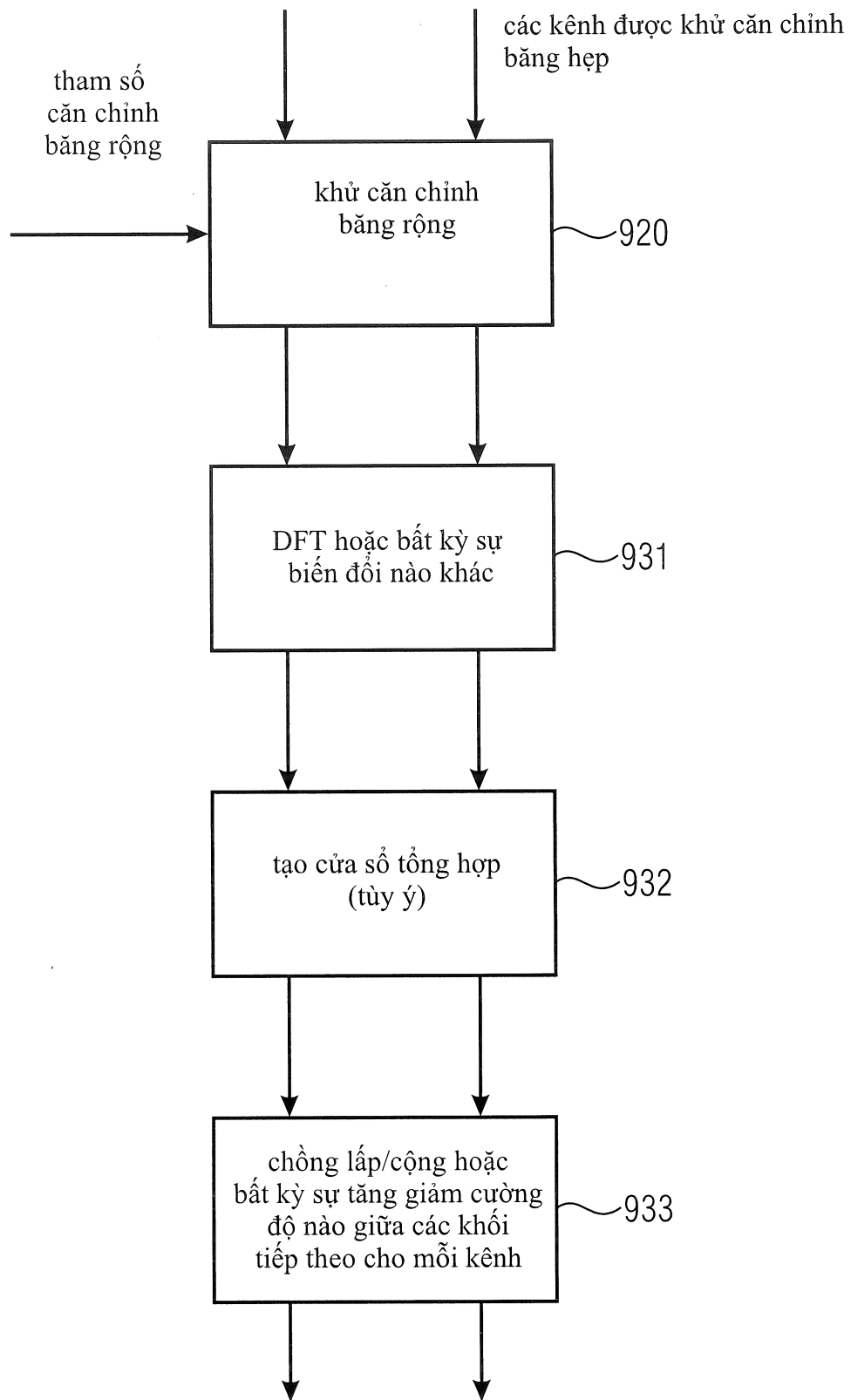


Fig. 9c

15/25

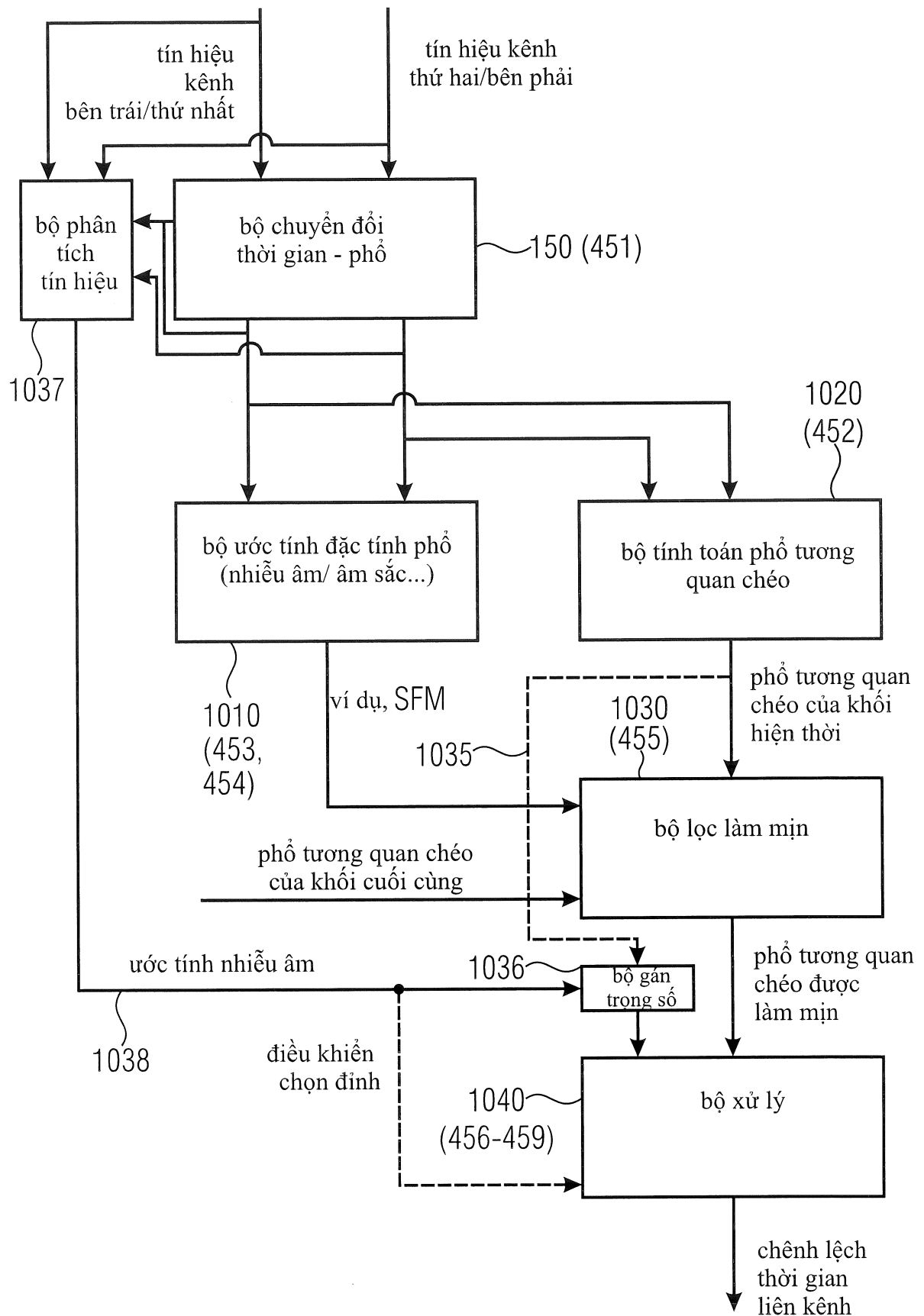
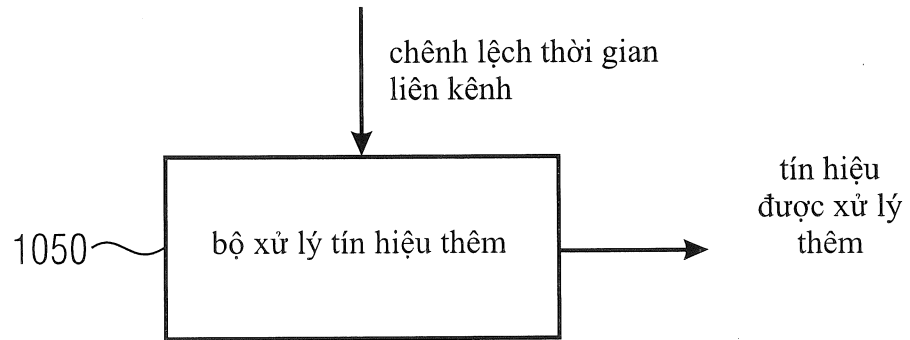


Fig. 10a

16/25



- lưu trữ/truyền dữ liệu tham số
- xử lý/mã hóa âm lập thể/đa kênh
- căn chỉnh thời gian của hai kênh
- chênh lệch thời gian của ước tính đến cho việc xác định vị trí loa trong phòng với hai micrô và thiết lập micrô đã biết
- điều hướng chùm sóng
- lọc không gian
- phân tách mặt nổi/nền
- xác định vị trí của nguồn âm thanh
ví dụ bởi phép đặc tam giác âm dựa trên các
chênh lệch thời gian của hai/ba tín hiệu

Fig. 10b

17/25

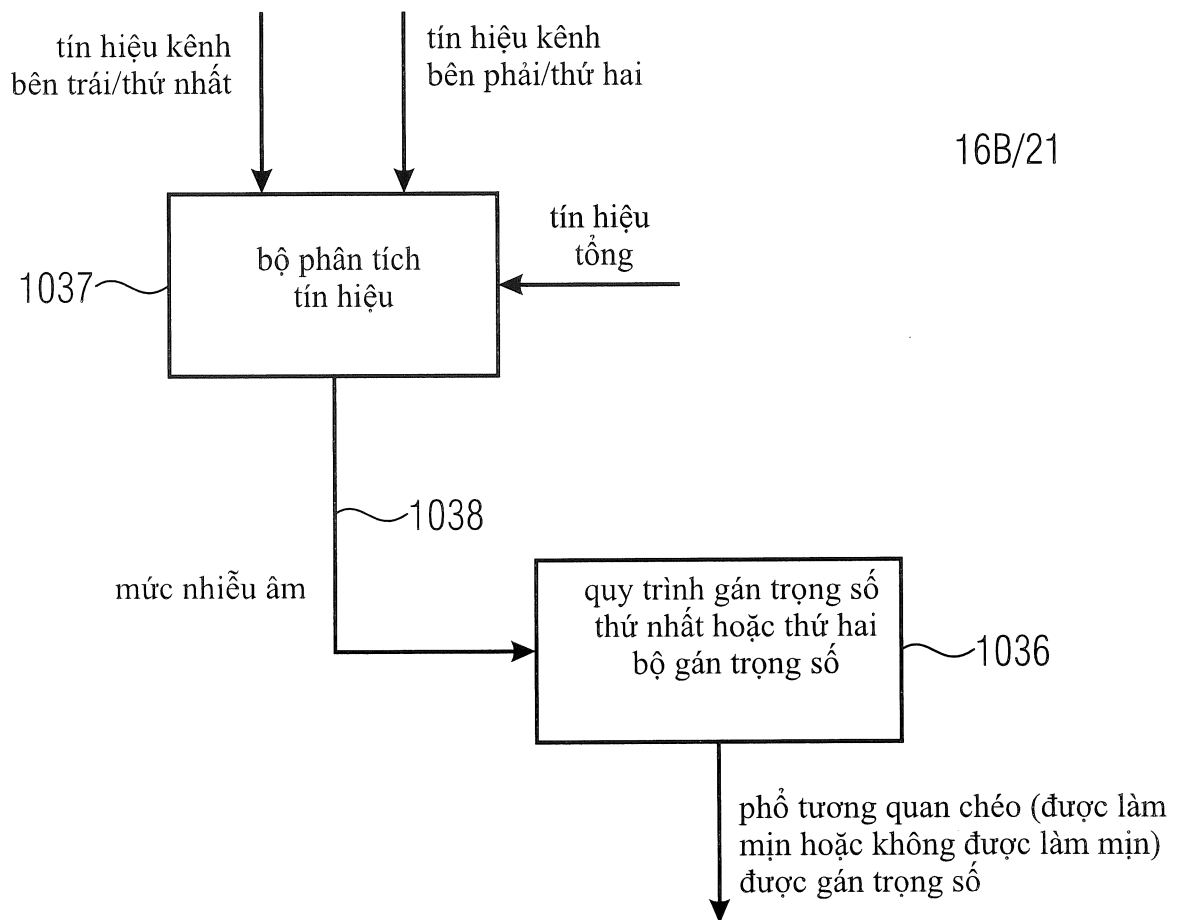


Fig. 10c

18/25

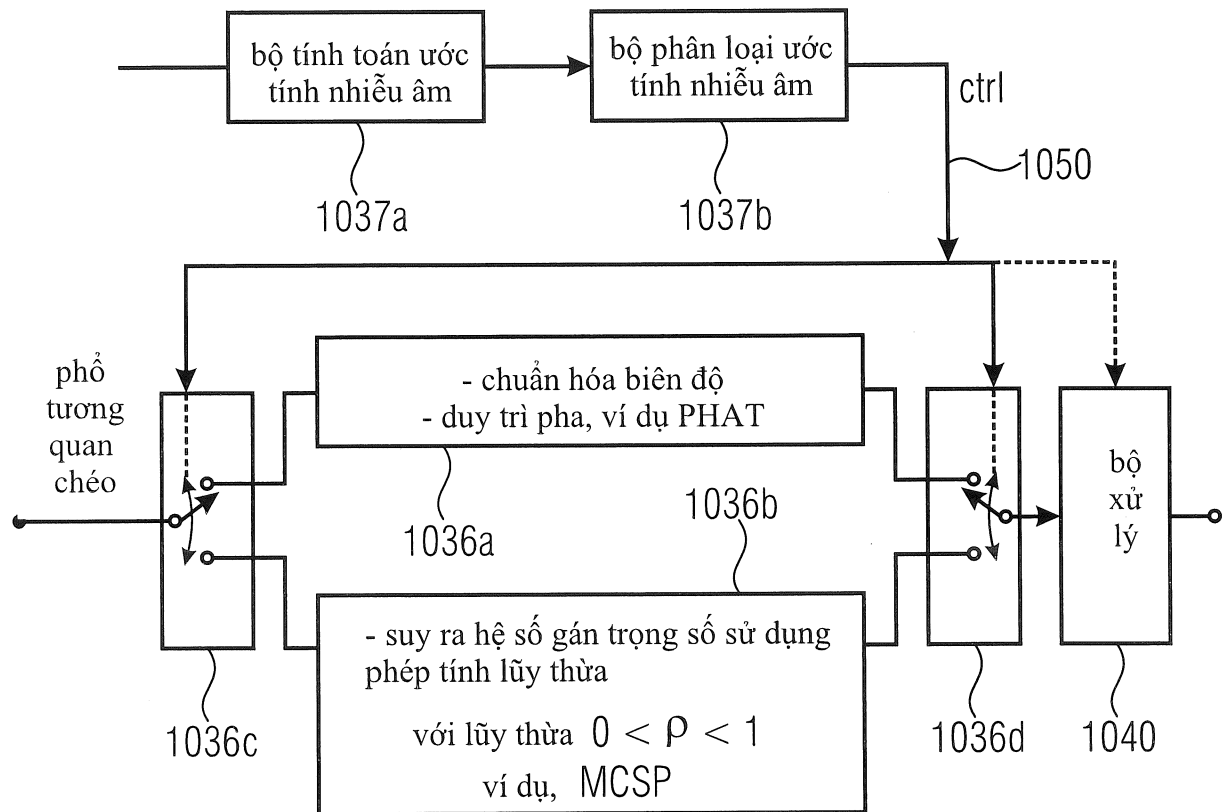
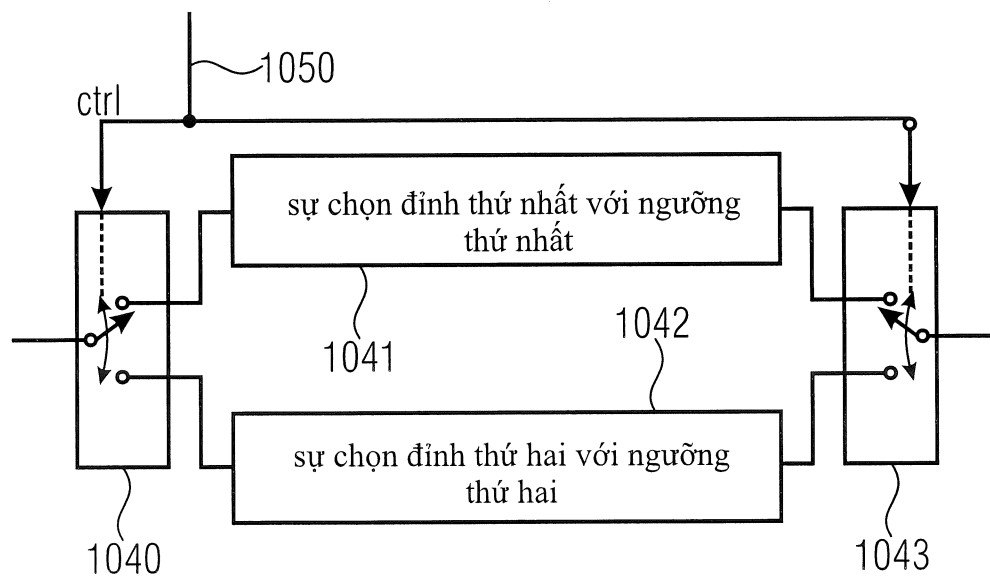


Fig. 10d



- ngưỡng thứ hai thấp hơn ngưỡng thứ nhất

Fig. 10e

19/25

tín hiệu đầu vào (từng khung)

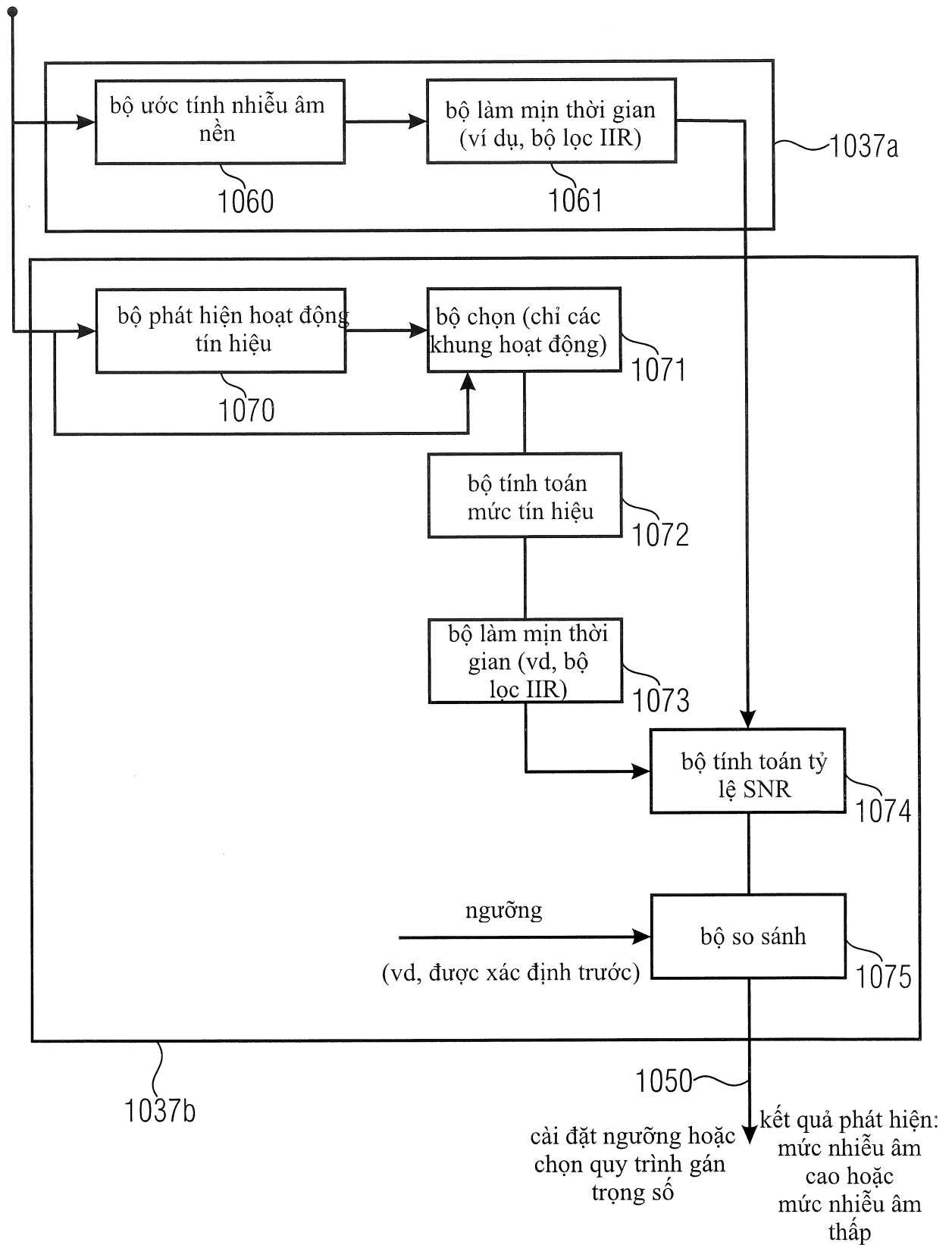


Fig. 10f

20/25

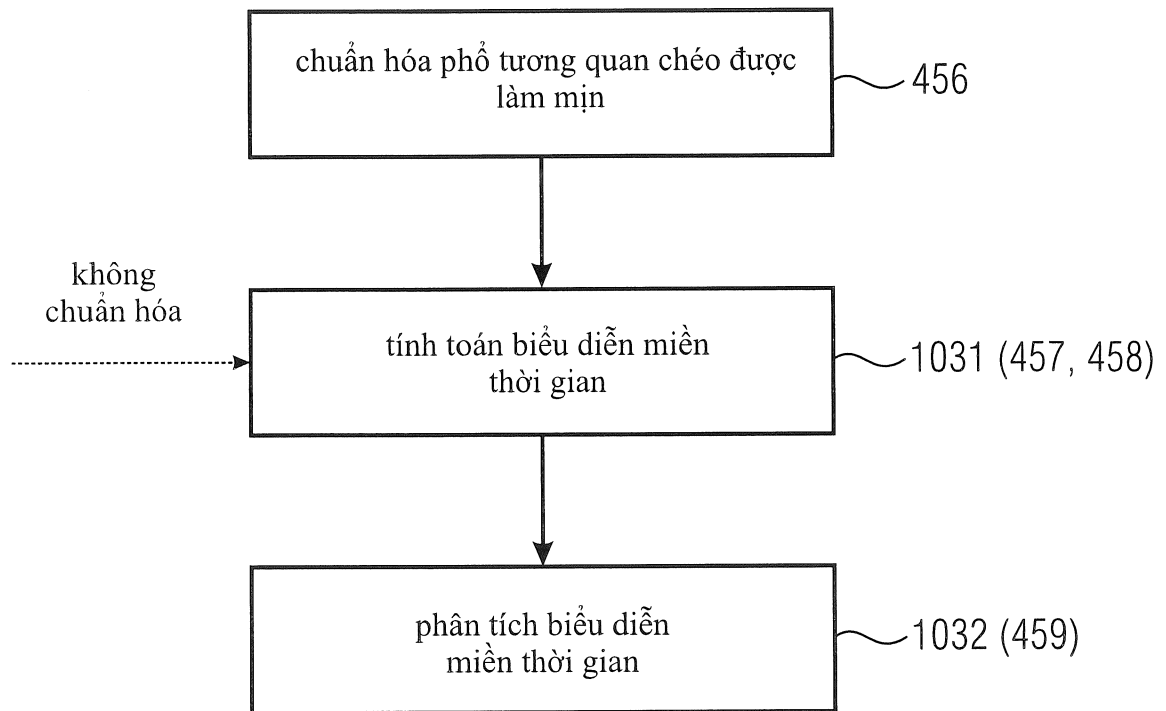


Fig. 11a

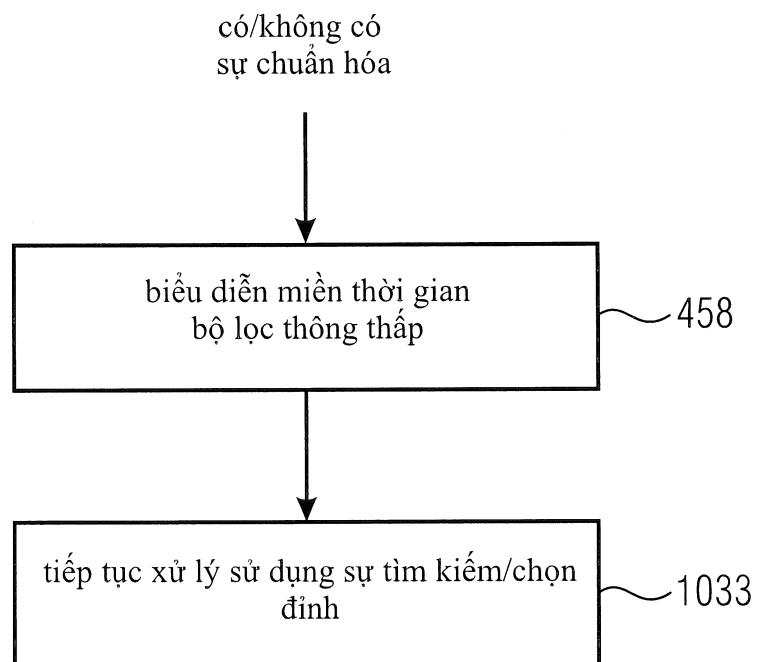


Fig. 11b

21/25

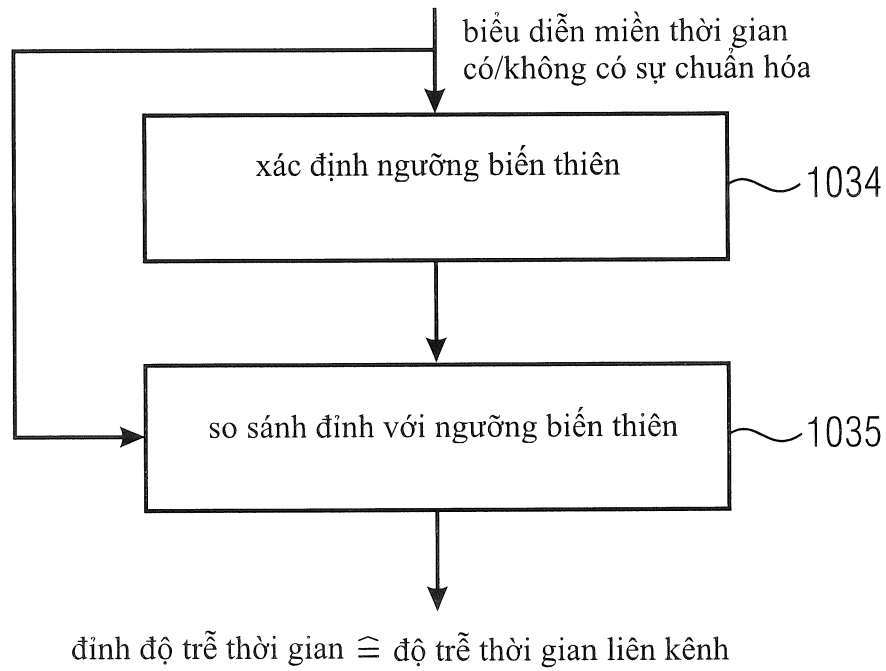


Fig. 11c

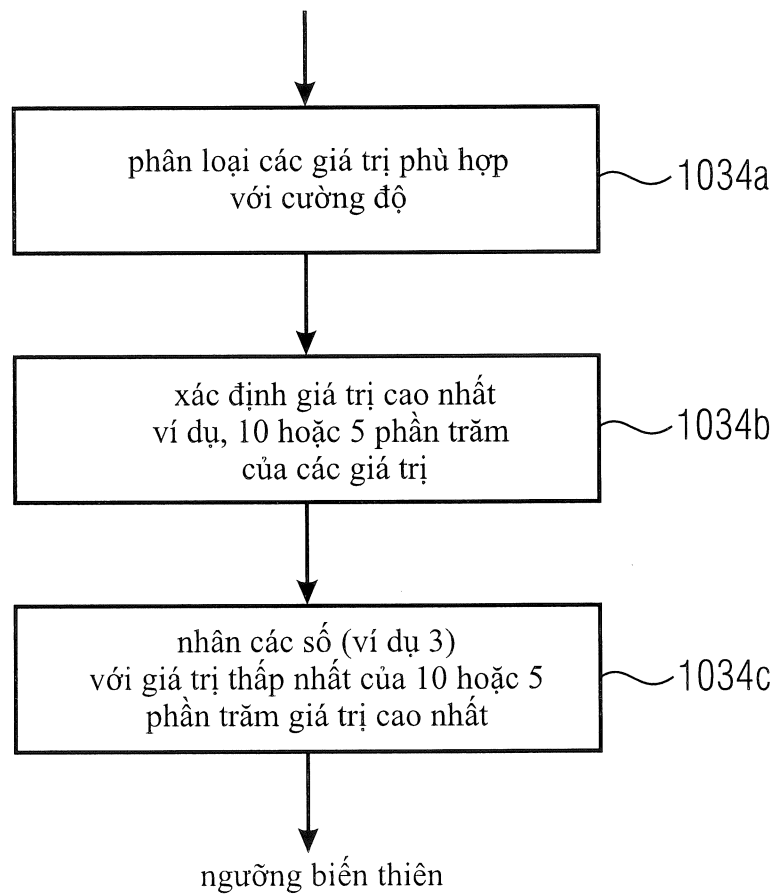


Fig. 11d

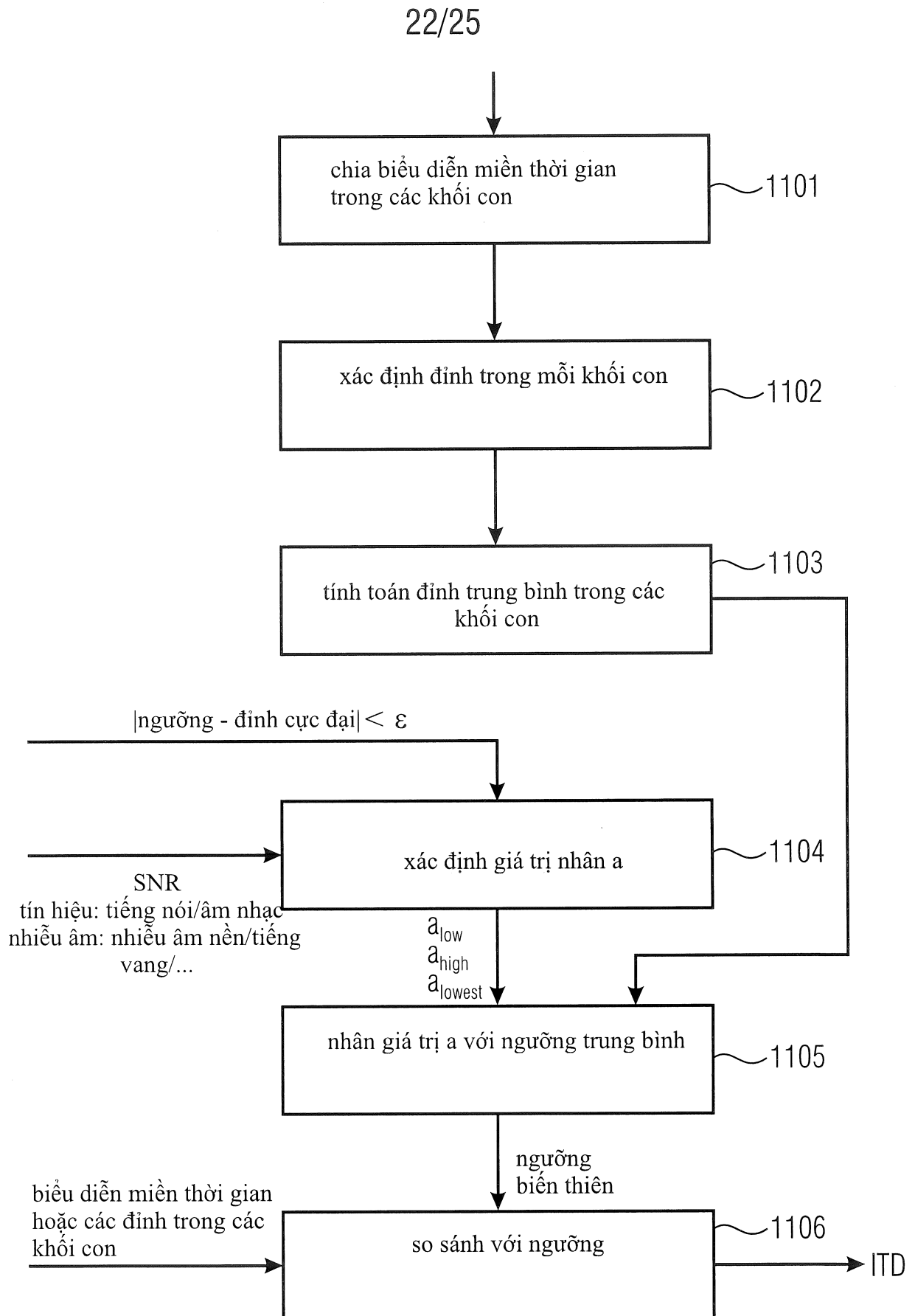


Fig. 11e

23/25

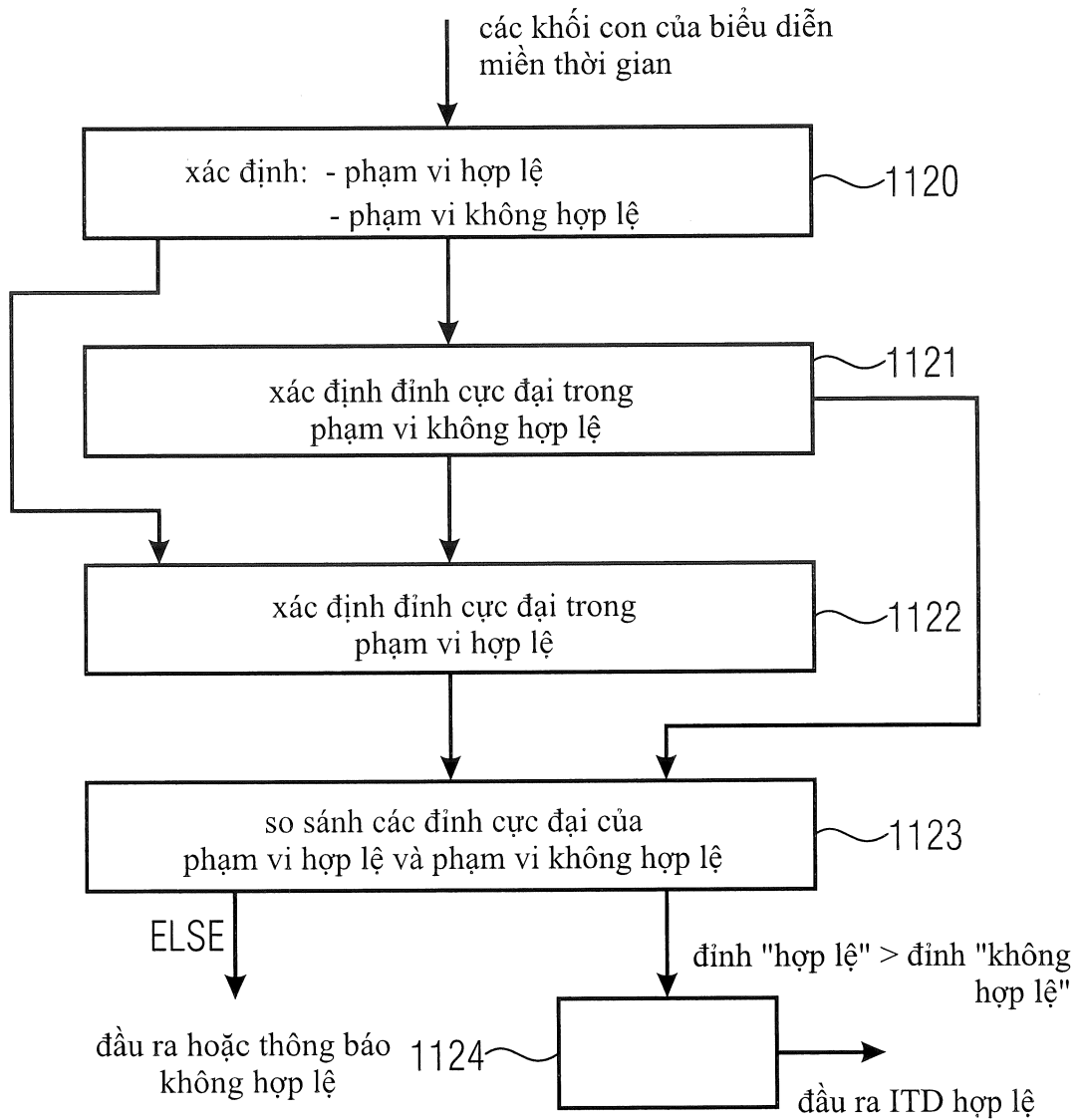


Fig. 11f

24/25

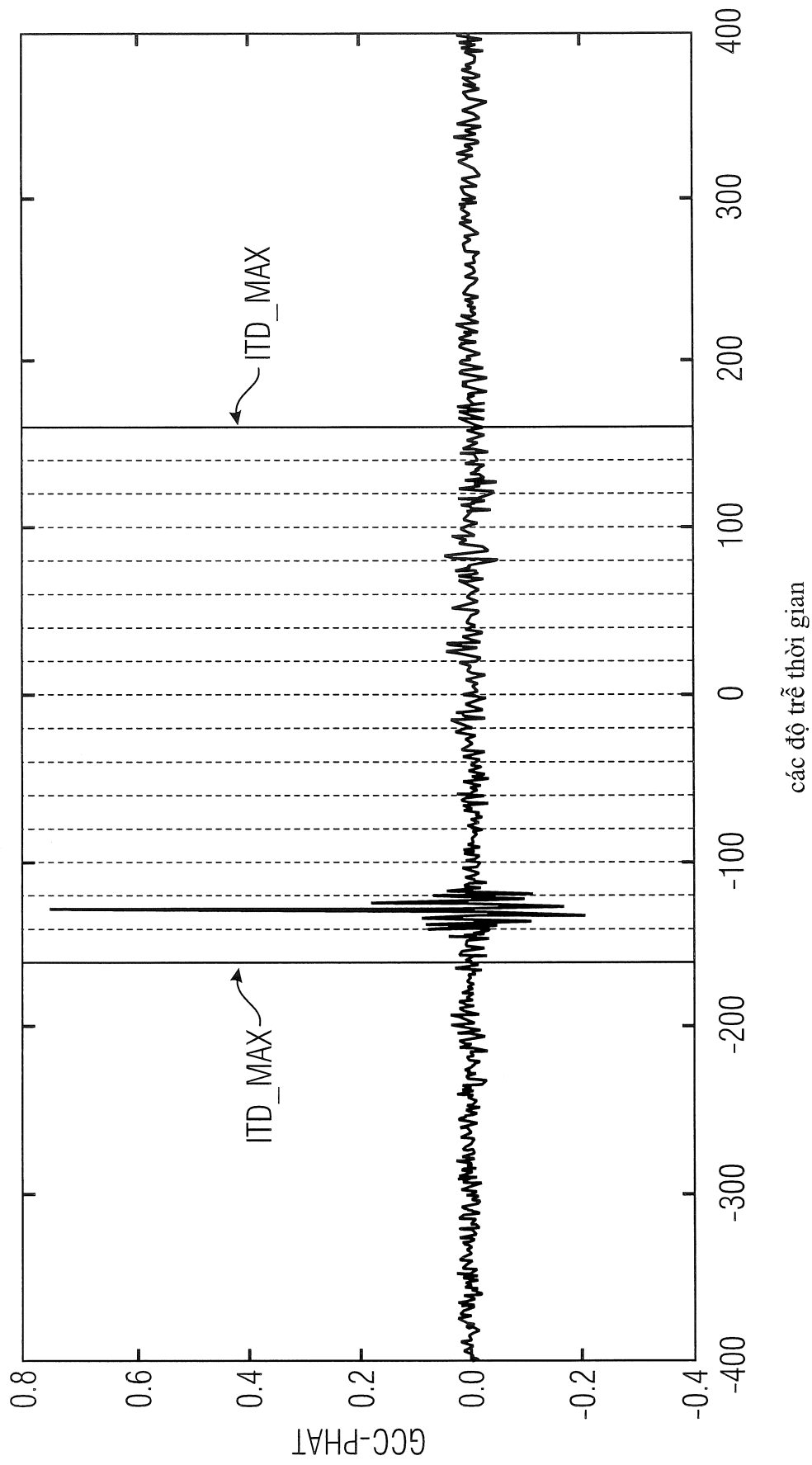


Fig. 12

25/25

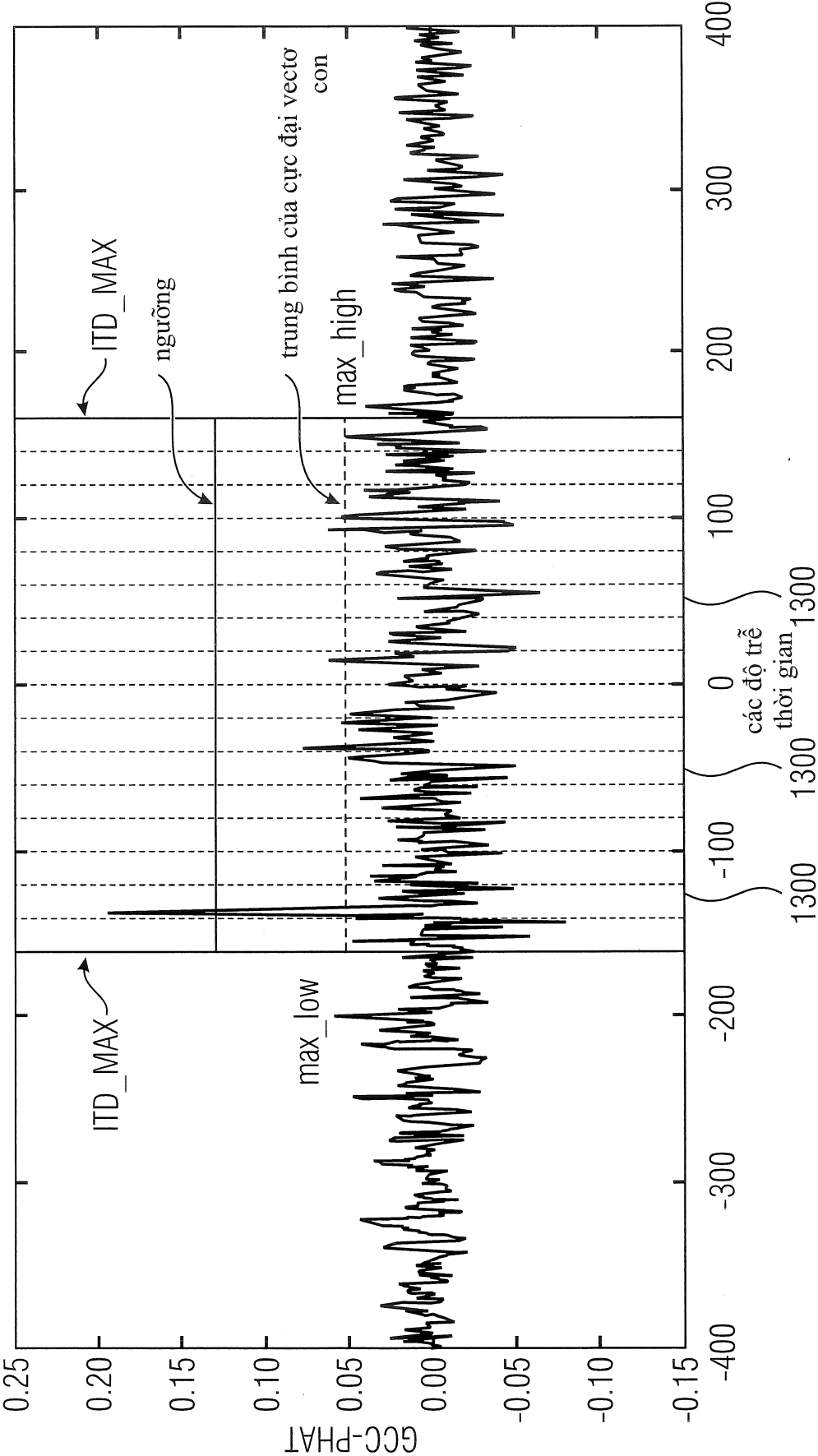


Fig. 13