



(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)



1-0031780

(51)⁷ G10L 21/0316; G10L 21/0272

(13) B

(21) 1-2017-00636

(22) 12/12/2014

(86) PCT/EP2014/077620 12/12/2014

(87) WO2016/091332 16/06/2016

(45) 25/05/2022 410

(43) 26/06/2017 351A

(73) HUAWEI TECHNOLOGIES CO., LTD. (CN)

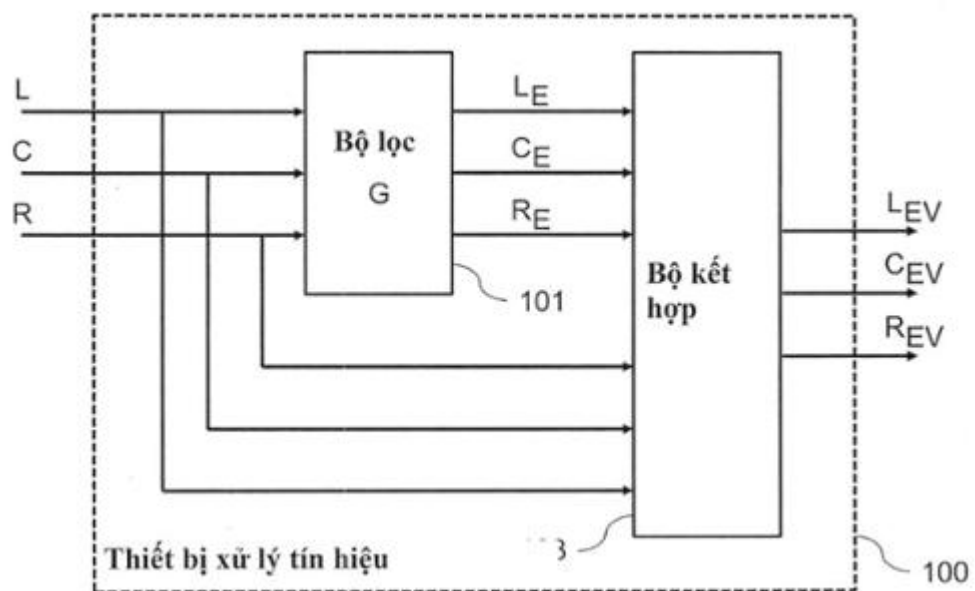
Huawei Administration Building Bantian, Longgang, Shenzhen, Guangdong 518129, China

(72) GEIGER, Jürgen (DE); GROSCHE, Peter (DE).

(74) Công ty Luật TNHH T&G (TGVN)

(54) THIẾT BỊ XỬ LÝ TÍN HIỆU, VẬT GHI ĐỌC ĐƯỢC BẰNG MÁY TÍNH, VÀ PHƯƠNG PHÁP XỬ LÝ TÍN HIỆU ĐỂ TĂNG CƯỜNG THÀNH PHẦN GIỌNG NÓI NẪM TRONG TÍN HIỆU AUDIO ĐA KÊNH

(57) Sáng chế đề cập tới thiết bị xử lý tín hiệu (100) để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh chứa tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), thiết bị xử lý tín hiệu (100) bao gồm bộ lọc (101) và bộ phận kết hợp (103); trong đó, bộ lọc (101) được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên cơ sở của tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), để thu hàm tăng ích ăng ten (G) dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, và để định trọng số tín hiệu audio kênh trái (L) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trái được định trọng số (L_E), để định trọng số tín hiệu audio kênh trung tâm (C) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trung tâm được định trọng số (C_E), và để định trọng số tín hiệu audio kênh phải (R) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh phải được định trọng số (R_E); và trong đó, bộ phận kết hợp (103) được đặt cấu hình để kết hợp tín hiệu audio kênh trái (L) với tín hiệu audio kênh trái được định trọng số (L_E) để thu tín hiệu audio kênh trái được kết hợp (L_{EV}), để kết hợp tín hiệu audio kênh trung tâm (C) với tín hiệu audio kênh trung tâm được định trọng số (C_E) để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và để kết hợp tín hiệu audio kênh phải (R) với tín hiệu audio kênh phải được định trọng số (R_E) để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực xử lý tín hiệu audio, cụ thể là đề cập đến việc tăng cường giọng nói nằm trong các tín hiệu audio đa kênh.

Tình trạng kỹ thuật của sáng chế

Để tăng cường thành phần giọng nói nằm trong các tín hiệu audio đa kênh, tức là các tín hiệu audio giải trí, các giải pháp khác nhau hiện đã được sử dụng.

Giải pháp đơn giản nhất để tăng cường thành phần giọng nói là gia tăng tín hiệu audio kênh trung tâm được chứa bởi tín hiệu audio đa kênh, hoặc một cách tương ứng là làm suy giảm tất cả các tín hiệu audio của các kênh khác. Giải pháp này khai thác việc giả sử là giọng nói thường bị dịch tới tín hiệu audio kênh trung tâm. Tuy nhiên, giải pháp này thường bị ảnh hưởng bởi hiệu quả làm việc kém của việc tăng cường giọng nói.

Một giải pháp phức tạp hơn cố để phân tích các tín hiệu audio của các kênh tách biệt. Liên quan tới vấn đề này, thông tin về quan hệ giữa tín hiệu audio kênh trung tâm và các tín hiệu audio của các kênh khác có thể được tạo ra cùng với việc trộn giảm âm thanh nổi để cho phép thực hiện việc tăng cường giọng nói. Tuy nhiên, giải pháp này không thể được áp dụng cho các tín hiệu audio âm thanh nổi và yêu cầu kênh audio giọng nói tách biệt.

Giải pháp khác là để cải thiện mức của các thành phần giọng nói nhỏ và làm suy giảm các thành phần không phải là giọng nói, lớn, nằm trong tín hiệu audio đa kênh là việc nén khoảng giới hạn động (dynamic range compression - DRC). Đầu tiên, giải pháp này chứa việc làm suy giảm các thành phần âm thanh lớn. Sau đó, mức độ lớn tổng thể được tăng, tạo thành việc gia tăng giọng nói hoặc hội thoại. Tuy nhiên, các giải pháp này không xét tới bản chất tự nhiên của tín hiệu audio đa kênh và việc biến đổi chỉ là phù hợp khi xét tới mức độ to của âm thanh.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất cách hiệu quả để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh.

Có thể đạt được mục đích này nhờ các đặc điểm của các yêu cầu bảo hộ độc lập. Các dạng áp dụng khác là rõ ràng từ các yêu cầu bảo hộ phụ thuộc, phần mô tả và các hình vẽ.

Sáng chế dựa trên việc tìm ra rằng tín hiệu audio đa kênh có thể được lọc dựa trên cơ sở của hàm tăng ích ăng ten, vốn có thể được xác định từ tất cả các kênh của tín hiệu audio đa kênh. Việc lọc có thể dựa trên giải pháp lọc Wiener, trong đó, tín hiệu audio kênh trung tâm của tín hiệu audio đa kênh có thể được xem xét là chứa thành phần giọng nói, và trong đó, các kênh khác của tín hiệu audio đa kênh có thể được xem xét là chứa các thành phần không phải là giọng nói. Để xem xét thay đổi của thành phần giọng nói nằm trong tín hiệu audio đa kênh qua thời gian, việc phát hiện hoạt động giọng nói có thể còn được thực hiện, trong đó, tất cả các kênh của tín hiệu audio đa kênh có thể được xử lý để tạo ra bộ chỉ thị hoạt động giọng nói. Tín hiệu audio đa kênh có thể là kết quả của việc xử lý trộn tăng âm thanh nổi của tín hiệu audio âm thanh nổi đầu vào. Do đó, có thể đạt được cải thiện một cách hiệu quả cho thành phần giọng nói nằm trong tín hiệu audio đa kênh.

Theo khía cạnh thứ nhất, sáng chế đề xuất thiết bị xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh chứa tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, thiết bị xử lý tín hiệu bao gồm bộ lọc và bộ phận kết hợp; trong đó, bộ lọc được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên cơ sở của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, để thu hàm tăng ích ăng ten dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, và để định trọng số tín hiệu audio kênh trái bởi hàm tăng ích ăng ten để thu tín hiệu audio kênh trái được định trọng số, để định trọng số tín hiệu audio kênh trung tâm bởi hàm tăng ích ăng ten để thu tín hiệu audio kênh trung tâm được định trọng số, và để định trọng số tín hiệu audio kênh phải bởi hàm tăng ích ăng ten để thu tín hiệu audio

kênh phải được định trọng số; và trong đó, bộ phận kết hợp được đặt cấu hình để kết hợp tín hiệu audio kênh trái với tín hiệu audio kênh trái được định trọng số để thu tín hiệu audio kênh trái được kết hợp, để kết hợp tín hiệu audio kênh trung tâm với tín hiệu audio kênh trung tâm được định trọng số để thu tín hiệu audio kênh trung tâm được kết hợp, và để kết hợp tín hiệu audio kênh phải với tín hiệu audio kênh phải được định trọng số để thu tín hiệu audio kênh phải được kết hợp. Do đó, có thể đạt được cách hiệu quả để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh.

Tín hiệu audio đa kênh chứa tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải. Tín hiệu audio đa kênh còn có thể chứa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Tín hiệu audio đa kênh có thể là tín hiệu audio âm thanh nổi LCR/3.0 hoặc tín hiệu audio bao quanh 5.1. Việc xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số chứa xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh trong miền tần số.

Hàm tăng ích ăng ten có thể chỉ thị tỉ lệ của độ lớn của thành phần giọng nói và độ lớn tổng thể của tín hiệu audio đa kênh, trong đó, nó được giả sử rằng thành phần giọng nói được chứa bởi tín hiệu audio kênh trung tâm. Độ lớn tổng thể của tín hiệu audio đa kênh có thể được xác định sử dụng việc cộng của thành phần giọng nói và các thành phần không phải là giọng nói nằm trong tín hiệu audio đa kênh qua tần số. Hàm tăng ích ăng ten có thể phụ thuộc tần số.

Theo dạng áp dụng thứ nhất của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy, bộ lọc được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm và số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải. Do đó, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh được xác định một cách hiệu quả và theo cách thích hợp hơn để được sử dụng để thu bộ lọc hàm tăng ích ăng ten, do khác biệt của tín hiệu audio kênh trái và tín hiệu audio kênh phải thể hiện tín hiệu dư không chứa các thành phần của tín hiệu audio kênh trung tâm.

Theo dạng áp dụng thứ hai của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ nêu trên của khía cạnh thứ nhất, bộ lọc được đặt cấu hình để xác định hàm tăng ích ăng ten theo các phương trình sau:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

trong đó, G biểu thị hàm tăng ích ăng ten, L biểu thị tín hiệu audio kênh trái, C biểu thị tín hiệu audio kênh trung tâm, R biểu thị tín hiệu audio kênh phải, P_C biểu thị công suất của tín hiệu audio kênh trung tâm làm số đo thể hiện độ lớn của tín hiệu audio kênh trung tâm, P_S biểu thị công suất của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và tổng của P_C và P_S biểu thị số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngăn tần số (frequency bin). Do đó, hàm tăng ích ăng ten được xác định theo cách hiệu quả và mạnh mẽ.

Hàm tăng ích ăng ten được xác định theo giải pháp lọc Wiener. Tín hiệu audio kênh trung tâm được coi là chứa thành phần giọng nói. Khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải được coi như là chứa thành phần không phải là giọng nói, dựa trên giả sử là các thành phần giọng nói được quét cho tín hiệu audio kênh trung tâm. Bằng cách xác định các thành phần của bộ lọc Wiener theo cách này, nó được loại bỏ khỏi việc sử dụng các phương pháp đắt tiền để ước lượng tỉ lệ tín hiệu so với nhiễu hoặc mật độ phổ công suất nhiễu của tín hiệu.

Thay cho việc sử dụng công suất nằm trong các phương trình, độ lớn hoặc công suất logarit có thể được sử dụng để xác định hàm tăng ích ăng ten. Khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải có thể đề cập tới tín hiệu audio còn dư chứa tổ hợp của các tín hiệu audio không phải là kênh trung tâm, trong đó, tất cả các tín hiệu audio trừ tín hiệu audio kênh trung tâm cũng có thể được đề cập tới như là các tín hiệu audio không phải là kênh trung tâm. Tín hiệu audio còn dư có thể là khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải.

Tổng của độ lớn của tín hiệu audio kênh trái và tín hiệu audio kênh phải tương ứng với việc tạo thành chùm là dạng cụ thể của việc tách kênh trung tâm, và cũng có

thể được sử dụng trong các phương án thực hiện của sáng chế. Tuy nhiên, khác biệt của độ lớn của tín hiệu audio kênh trái và audio kênh phải tương ứng với việc loại bỏ của thành phần của kênh trung tâm. Do đó, tín hiệu audio còn dư được định nghĩa như là khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải tạo thành trong ước lượng được cải tiến của khuếch đại bộ lọc.

Theo dạng áp dụng thứ ba của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, tín hiệu audio đa kênh còn chứa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, trong đó, bộ lọc được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số còn dựa trên cơ sở của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, và để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm, của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Do đó, các kênh bao quanh nằm trong tín hiệu audio đa kênh được xử lý một cách hiệu quả, bằng cách thu độ lớn từ khác biệt của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Tín hiệu khác biệt cho sự phân biệt tốt hơn cho tín hiệu audio kênh trung tâm.

Theo dạng áp dụng thứ tư của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, bộ lọc được đặt cấu hình để định trọng số các ngăn tần số (frequency bin) của tín hiệu audio kênh trái bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh trái được định trọng số, để định trọng số các ngăn tần số (frequency bin) của tín hiệu audio kênh trung tâm bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh trung tâm được định trọng số, và để định trọng số các ngăn tần số (frequency bin) của tín hiệu audio kênh phải bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh phải được định trọng số. Do đó, tín hiệu audio đa kênh được xử lý một cách hiệu quả trong miền tần số. Việc định trọng số tất cả các tín hiệu với cùng một bộ lọc có ưu điểm là không làm xuất hiện việc dịch chuyển của các vị trí nguồn tài nguyên audio

trong hình ảnh âm thanh nổi. Hơn nữa, theo cách này, thành phần giọng nói được tách từ tất cả các tín hiệu.

Bộ lọc còn có thể tiếp tục được đặt cấu hình để nhóm các ngăn tần số (frequency bin) theo thang tần số Mel để thu các băng tần số. Chỉ số k do đó có thể tương ứng với chỉ số băng tần số. Bộ lọc còn có thể được tiếp tục đặt cấu hình để chỉ xử lý các ngăn tần số (frequency bin) hoặc các băng tần số được sắp xếp nằm trong khoảng giới hạn tần số định trước, tức là từ 100Hz tới 8kHz. Theo cách này, chỉ các tần số chứa giọng nói của người được xử lý.

Theo dạng áp dụng thứ năm của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, thiết bị xử lý tín hiệu còn chứa bộ phát hiện hoạt động giọng nói đang được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói dựa trên cơ sở của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, bộ chỉ thị hoạt động giọng nói chỉ thị độ lớn của thành phần giọng nói nằm trong tín hiệu audio đa kênh qua thời gian, trong đó, bộ phân kết hợp còn được đặt cấu hình để kết hợp tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trái được kết hợp, để kết hợp tín hiệu audio kênh trung tâm được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trung tâm được kết hợp, và để kết hợp tín hiệu audio kênh phải được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh phải được kết hợp. Do đó, có thể đạt được cải thiện hiệu quả của thành phần giọng nói thay đổi theo thời gian nằm trong tín hiệu audio đa kênh, và các tín hiệu không phải là lời nói bị chặn.

Bộ chỉ thị hoạt động giọng nói chỉ thị độ lớn của thành phần giọng nói nằm trong tín hiệu audio đa kênh trong miền thời gian. Bộ chỉ thị hoạt động giọng nói là, ví dụ, bằng không khi không có thành phần giọng nói có mặt trong tín hiệu, và bằng một khi giọng nói có mặt. Các trị số giữa không và một có thể được giải thích như là khả năng mà giọng nói có mặt, và giúp thu tín hiệu đầu ra tron tru.

Theo dạng áp dụng thứ sáu của thiết bị xử lý tín hiệu theo dạng áp dụng thứ năm của khía cạnh thứ nhất, bộ phát hiện hoạt động giọng nói được đặt cấu hình để xác định số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh dựa trên cơ sở

của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, và để thu bộ chỉ thị hoạt động giọng nói dựa trên tỉ lệ giữa số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm và số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh. Do đó, bộ chỉ thị hoạt động giọng nói được xác định một cách hiệu quả bởi việc khai thác quan hệ giữa các số đo của sự biến đổi phổ.

Số đo thể hiện sự biến đổi phổ tổng thể có thể là thông lượng phổ hoặc đạo hàm theo thời gian. Thông lượng phổ có thể được xác định sử dụng các giải pháp khác cho việc chuẩn hóa. Thông lượng phổ có thể được tính toán như là khác biệt của phổ công suất giữa hai hoặc hơn hai khung phụ tín hiệu audio. Số đo thể hiện sự biến đổi phổ tổng thể có thể là tổng của F_C và F_S , trong đó, F_C biểu thị số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm, và trong đó, F_S biểu thị số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải.

Theo dạng áp dụng thứ bảy của thiết bị xử lý tín hiệu theo dạng áp dụng thứ sáu của khía cạnh thứ nhất, bộ phát hiện hoạt động giọng nói được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói theo phương trình sau:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

trong đó, V biểu thị bộ chỉ thị hoạt động giọng nói, F_C biểu thị số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm, F_S biểu thị số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và tổng của F_C và F_S biểu thị số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh, và a biểu thị thông số định cỡ được xác định từ trước. Do đó, bộ chỉ thị hoạt động giọng nói được xác định một cách hiệu quả. Các tín hiệu với cùng các trị số F_C và F_S tạo thành trong bộ chỉ thị hoạt động giọng nói với trị số bằng không. Các trị số lớn hơn của F_C dẫn tới các trị số lớn hơn của bộ chỉ thị hoạt động giọng nói. Thông số định cỡ a có thể điều khiển độ lớn của bộ chỉ thị hoạt động giọng nói.

Các trị số của bộ chỉ thị hoạt động giọng nói có thể độc lập với việc chuẩn hóa trước đó của các số đo. Các trị số của bộ chỉ thị hoạt động giọng nói có thể bị hạn chế vào khoảng gián đoạn $[0; 1]$.

Theo dạng áp dụng thứ tám của thiết bị xử lý tín hiệu theo dạng áp dụng thứ bảy của khía cạnh thứ nhất, bộ phát hiện hoạt động giọng nói được đặt cấu hình để xác định số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm làm thông lượng phổ và số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải làm thông lượng phổ theo các phương trình sau:

$$F_C(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_S(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

trong đó, F_C biểu thị thông lượng phổ của tín hiệu audio kênh trung tâm, F_S biểu thị thông lượng phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, C biểu thị tín hiệu audio kênh trung tâm, S biểu thị khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngăn tần số (frequency bin). Do đó, thông lượng phổ được xác định một cách hiệu quả.

Theo dạng áp dụng thứ chín của thiết bị xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ tám của khía cạnh thứ nhất, bộ phát hiện hoạt động giọng nói được đặt cấu hình để lọc bộ chỉ thị hoạt động giọng nói theo thời gian dựa trên cơ sở của hàm lọc thông thấp được xác định từ trước. Do đó, có thể đạt được việc làm giảm nhẹ một cách hiệu quả của các thành phần lạ nằm trong tín hiệu audio đa kênh và/hoặc việc làm trơn tru theo thời gian một cách hiệu quả của bộ chỉ thị hoạt động giọng nói.

Hàm lọc thông thấp được xác định từ trước có thể được nhận biết bởi bộ lọc thông thấp đáp ứng xung hữu hạn (finite impulse response - FIR) một chạm.

Theo dạng áp dụng thứ mười của thiết bị xử lý tín hiệu theo dạng áp dụng thứ năm đến dạng áp dụng thứ chín của khía cạnh thứ nhất, bộ phận kết hợp còn được đặt cấu hình để định trọng số tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải bởi thông số khuếch đại đầu vào được xác định từ trước, và để định trọng số bộ chỉ thị hoạt động giọng nói bởi thông số khuếch đại lời nói được xác định từ trước. Do đó, có thể đạt được việc điều khiển một cách hiệu quả độ lớn của thành phần giọng nói liên quan tới độ lớn của thành phần không phải là giọng nói.

Theo dạng áp dụng thứ mười một của thiết bị xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ mười của khía cạnh thứ nhất, bộ phận kết hợp được đặt cấu hình để cộng tín hiệu audio kênh trái vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trái được kết hợp, để cộng tín hiệu audio kênh trung tâm vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trung tâm được kết hợp, và để cộng tín hiệu audio kênh phải vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh phải được kết hợp. Do đó, bộ phận kết hợp được áp dụng một cách hiệu quả. Các thành phần giọng nói được tách được kết hợp với các tín hiệu gốc để tăng cường thành phần giọng nói trong các tín hiệu đầu ra.

Theo dạng áp dụng thứ mười hai của thiết bị xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ mười một của khía cạnh thứ nhất, tín hiệu audio đa kênh còn chứa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, trong đó, bộ phát hiện hoạt động giọng nói được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói còn dựa trên cơ sở của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Do đó, các kênh bao quanh nằm trong tín hiệu audio đa kênh cũng được xem xét để xác định bộ chỉ thị hoạt động giọng nói, tạo thành ước lượng tốt hơn của bộ chỉ thị hoạt động giọng nói.

Theo dạng áp dụng thứ mười ba của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, thiết bị xử lý tín hiệu còn chứa bộ biến đổi đang được đặt cấu hình để biến đổi tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải từ miền thời gian vào miền tần số. Do đó, có thể đạt được việc biến đổi một cách hiệu quả của các tín hiệu audio vào miền tần số. Điều này có thể được yêu cầu trong trường hợp mà việc tăng cường lời nói và việc phát hiện hoạt động giọng nói được thực hiện trong miền tần số.

Bộ biến đổi có thể được đặt cấu hình để thực hiện biến đổi Fourier rời rạc thời gian ngắn (short-time discrete Fourier transform - STFT) của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải.

Theo dạng áp dụng thứ mười bốn của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, thiết bị xử lý tín hiệu còn chứa bộ biến đổi ngược đang được đặt cấu hình để biến đổi ngược tín hiệu audio kênh trái được kết hợp, tín hiệu audio kênh trung tâm được kết hợp, và tín hiệu audio kênh phải được kết hợp từ miền tần số vào miền thời gian. Do đó, có thể đạt được việc biến đổi ngược một cách hiệu quả cho các tín hiệu audio vào miền thời gian, và thu được các tín hiệu đầu ra trong miền thời gian.

Bộ biến đổi ngược có thể được đặt cấu hình để thực hiện việc biến đổi ngược Fourier rời rạc thời gian ngắn (inverse short-time discrete Fourier transform - ISTFT) của tín hiệu audio kênh trái được kết hợp, tín hiệu audio kênh trung tâm được kết hợp, và tín hiệu audio kênh phải được kết hợp.

Theo dạng áp dụng thứ mười lăm của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, thiết bị xử lý tín hiệu còn chứa bộ trộn tăng đang được đặt cấu hình để xác định tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải dựa trên cơ sở của tín hiệu audio âm thanh nổi kênh trái đầu vào và tín hiệu audio âm thanh nổi kênh phải đầu vào. Theo cách này, thiết bị xử lý tín hiệu có thể được áp dụng cho việc xử lý hai kênh, tức là kênh trái và phải, tín hiệu audio âm thanh nổi đầu vào.

Theo dạng áp dụng thứ mười sáu của thiết bị xử lý tín hiệu theo dạng áp dụng thứ mười lăm của khía cạnh thứ nhất, bộ trộn tăng được đặt cấu hình để xác định tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải theo các phương trình sau:

$$C = \alpha \times (L_m + R_m)$$

$$L = L_m - C$$

$$R = R_m - C$$

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right)$$

trong đó, L_r biểu thị phần thực của tín hiệu audio âm thanh nổi kênh trái đầu vào, R_r biểu thị phần thực của tín hiệu audio âm thanh nổi kênh phải đầu vào, L_i biểu thị

phần ảo của tín hiệu audio âm thanh nổi kênh trái đầu vào, R_i biểu thị phần ảo của tín hiệu audio âm thanh nổi kênh phải đầu vào, α biểu thị tham số trực giao, L_{in} biểu thị tín hiệu audio âm thanh nổi kênh trái đầu vào, R_{in} biểu thị tín hiệu audio âm thanh nổi kênh phải đầu vào, L biểu thị tín hiệu audio kênh trái, C biểu thị tín hiệu audio kênh trung tâm, và R biểu thị tín hiệu audio kênh phải. Do đó, có thể đạt được việc tách một cách hiệu quả cho kênh trung tâm của tín hiệu audio âm thanh nổi đầu vào sử dụng việc phân tách trực giao. Tín hiệu audio kênh trái và tín hiệu audio kênh phải tạo thành là trực giao với nhau.

Theo dạng áp dụng thứ mười bảy của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, thiết bị xử lý tín hiệu còn chứa bộ trộn giảm đang được đặt cấu hình để xác định tín hiệu audio âm thanh nổi kênh trái đầu ra và tín hiệu audio âm thanh nổi kênh phải đầu ra dựa trên cơ sở của tín hiệu audio kênh trái được kết hợp, tín hiệu audio kênh trung tâm được kết hợp, và tín hiệu audio kênh phải được kết hợp. Do đó, hai kênh, tức là kênh trái và kênh phải của tín hiệu audio âm thanh nổi đầu ra sẽ được tạo ra một cách hiệu quả.

Theo dạng áp dụng thứ mười tám của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, số đo độ lớn chứa công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu. Do đó, số đo độ lớn có thể chỉ thị các trị số khác nhau tại các thang khác nhau.

Độ lớn của tín hiệu audio đa kênh chứa công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu audio đa kênh. Số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải chứa công suất, công suất logarit, độ lớn hoặc độ lớn logarit của khác biệt của tín hiệu audio kênh trái và tín hiệu audio kênh phải. Độ lớn của tín hiệu audio kênh trung tâm chứa công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu audio kênh trung tâm. Tín hiệu này có thể đề cập tới tín hiệu bất kỳ được xử lý bởi thiết bị xử lý tín hiệu.

Theo dạng áp dụng thứ mười chín của thiết bị xử lý tín hiệu theo khía cạnh thứ nhất như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ nhất, bộ phận kết hợp còn được đặt cấu hình để định trọng số tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải bởi thông số khuếch đại đầu vào được xác định

từ trước, và để định trọng số tín hiệu audio kênh trái được định trọng số, tín hiệu audio kênh trung tâm được định trọng số, và tín hiệu audio kênh phải được định trọng số bởi thông số khuếch đại lời nói được xác định từ trước. Do đó, có thể đạt được việc điều khiển một cách hiệu quả độ lớn của thành phần giọng nói liên quan tới độ lớn của thành phần không phải là giọng nói.

Các tín hiệu audio được định trọng số C_E , L_E , và R_E , có thể được định trọng số bởi thông số khuếch đại lời nói được xác định từ trước G_S . Việc định trọng số có thể được thực hiện mà không sử dụng bộ phát hiện hoạt động giọng nói.

Theo khía cạnh thứ hai, sáng chế đề xuất phương pháp xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh chứa tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, phương pháp xử lý tín hiệu chứa bước xác định, bởi bộ lọc, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên cơ sở của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, thu, bởi bộ lọc, hàm tăng ích ăng ten dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, định trọng số, bởi bộ lọc, tín hiệu audio kênh trái bởi hàm tăng ích ăng ten để thu tín hiệu audio kênh trái được định trọng số, định trọng số, bởi bộ lọc, tín hiệu audio kênh trung tâm bởi hàm tăng ích ăng ten để thu tín hiệu audio kênh trung tâm được định trọng số, định trọng số, bởi bộ lọc, tín hiệu audio kênh phải bởi hàm tăng ích ăng ten để thu tín hiệu audio kênh phải được định trọng số, kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh trái với tín hiệu audio kênh trái được định trọng số để thu tín hiệu audio kênh trái được kết hợp, kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh trung tâm với tín hiệu audio kênh trung tâm được định trọng số để thu tín hiệu audio kênh trung tâm được kết hợp, và kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh phải với tín hiệu audio kênh phải được định trọng số để thu tín hiệu audio kênh phải được kết hợp. Do đó, có thể đạt được cách hiệu quả để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh.

Phương pháp xử lý tín hiệu có thể được thực hiện bởi thiết bị xử lý tín hiệu. Các đặc điểm khác của phương pháp xử lý tín hiệu được tạo thành trực tiếp từ chức năng của thiết bị xử lý tín hiệu.

Theo dạng áp dụng thứ nhất của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy, phương pháp chứa bước xác định, bởi bộ lọc, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm và số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải. Do đó, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh được xác định một cách hiệu quả và theo cách thích hợp hơn để được sử dụng để thu bộ lọc hàm tăng ích ăng ten, do khác biệt của tín hiệu audio kênh trái và tín hiệu audio kênh phải thể hiện tín hiệu dư không chứa các thành phần của tín hiệu audio kênh trung tâm.

Theo dạng áp dụng thứ hai của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ lọc, hàm tăng ích ăng ten theo các phương trình sau:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

trong đó, G biểu thị hàm tăng ích ăng ten, L biểu thị tín hiệu audio kênh trái, C biểu thị tín hiệu audio kênh trung tâm, R biểu thị tín hiệu audio kênh phải, P_C biểu thị công suất của tín hiệu audio kênh trung tâm làm số đo thể hiện độ lớn của tín hiệu audio kênh trung tâm, P_S biểu thị công suất của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và tổng của P_C và P_S biểu thị số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngắn tần số (frequency bin). Do đó, hàm tăng ích ăng ten được xác định theo cách hiệu quả và mạnh mẽ.

Theo dạng áp dụng thứ ba của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, tín hiệu audio đa kênh còn chứa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, trong đó, phương pháp chứa bước xác định, bởi bộ lọc, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số còn dựa trên cơ sở của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, và xác định, bởi bộ lọc, số đo thể

hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm, của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Do đó, các kênh bao quanh nằm trong tín hiệu audio đa kênh được xử lý một cách hiệu quả, bằng cách thu độ lớn từ khác biệt của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Tín hiệu khác biệt cho sự phân biệt tốt hơn cho tín hiệu audio kênh trung tâm.

Theo dạng áp dụng thứ tư của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước định trọng số, bởi bộ lọc, các ngăn tần số (frequency bin) của tín hiệu audio kênh trái bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh trái được định trọng số, định trọng số, bởi bộ lọc, các ngăn tần số (frequency bin) của tín hiệu audio kênh trung tâm bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh trung tâm được định trọng số, và định trọng số, bởi bộ lọc, các ngăn tần số (frequency bin) của tín hiệu audio kênh phải bởi các ngăn tần số (frequency bin) của hàm tăng ích ăng ten để thu các ngăn tần số (frequency bin) của tín hiệu audio kênh phải được định trọng số. Do đó, tín hiệu audio đa kênh được xử lý một cách hiệu quả trong miền tần số. Việc định trọng số tất cả các tín hiệu với cùng một bộ lọc có ưu điểm là không tạo thành việc dịch chuyển của các vị trí nguồn tài nguyên audio trong hình ảnh âm thanh nổi. Hơn nữa, theo cách này, thành phần giọng nói được tách từ tất cả các tín hiệu.

Theo dạng áp dụng thứ năm của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ phát hiện hoạt động giọng nói, bộ chỉ thị hoạt động giọng nói dựa trên cơ sở của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, bộ chỉ thị hoạt động giọng nói chỉ thị độ lớn của thành phần giọng nói nằm trong tín hiệu audio đa kênh qua thời gian, kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trái được kết hợp, kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh trung tâm được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh

trung tâm được kết hợp, và kết hợp, bởi bộ phận kết hợp, tín hiệu audio kênh phải được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh phải được kết hợp. Do đó, có thể đạt được cải thiện hiệu quả của thành phần giọng nói thay đổi theo thời gian nằm trong tín hiệu audio đa kênh, và các tín hiệu không phải là lời nói sẽ bị chặn.

Theo dạng áp dụng thứ sáu của phương pháp xử lý tín hiệu theo dạng áp dụng thứ năm của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ phát hiện hoạt động giọng nói, số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh dựa trên cơ sở của tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải, và thu, bởi bộ phát hiện hoạt động giọng nói, bộ chỉ thị hoạt động giọng nói dựa trên tỉ lệ giữa số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm và số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh. Do đó, bộ chỉ thị hoạt động giọng nói được xác định một cách hiệu quả bởi việc khai thác quan hệ giữa các số đo của sự biến đổi phổ.

Theo dạng áp dụng thứ bảy của phương pháp xử lý tín hiệu theo dạng áp dụng thứ sáu của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ phát hiện hoạt động giọng nói, bộ chỉ thị hoạt động giọng nói theo phương trình sau:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

trong đó, V biểu thị bộ chỉ thị hoạt động giọng nói, F_c biểu thị số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm, F_s biểu thị số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, và tổng của F_c và F_s biểu thị số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh, và a biểu thị thông số định cỡ được xác định từ trước. Do đó, bộ chỉ thị hoạt động giọng nói được xác định một cách hiệu quả. Các tín hiệu với cùng các trị số F_c và F_s tạo thành trong bộ chỉ thị hoạt động giọng nói với trị số bằng không. Các trị số lớn hơn của F_c dẫn tới các trị số lớn hơn của bộ chỉ thị hoạt động giọng nói. Thông số định cỡ a có thể điều khiển độ lớn của bộ chỉ thị hoạt động giọng nói.

Theo dạng áp dụng thứ tám của phương pháp xử lý tín hiệu theo dạng áp dụng thứ bảy của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ phát hiện hoạt

động giọng nói, số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm làm thông lượng phổ và số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải làm thông lượng phổ theo các phương trình sau:

$$F_C(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_S(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

trong đó, F_C biểu thị thông lượng phổ của tín hiệu audio kênh trung tâm, F_S biểu thị thông lượng phổ của khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, C biểu thị tín hiệu audio kênh trung tâm, S biểu thị khác biệt giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngăn tần số (frequency bin). Do đó, thông lượng phổ được xác định một cách hiệu quả.

Theo dạng áp dụng thứ chín của phương pháp xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ tám của khía cạnh thứ hai, phương pháp chứa bước lọc, bởi bộ phát hiện hoạt động giọng nói, bộ chỉ thị hoạt động giọng nói theo thời gian dựa trên cơ sở của hàm lọc thông thấp được xác định từ trước. Do đó, có thể đạt được việc làm giảm nhẹ một cách hiệu quả của các thành phần lạ nằm trong tín hiệu audio đa kênh và/hoặc việc làm trơn tru theo thời gian một cách hiệu quả của bộ chỉ thị hoạt động giọng nói.

Theo dạng áp dụng thứ mười của phương pháp xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ chín của khía cạnh thứ hai, phương pháp chứa bước định trọng số, bởi bộ phận kết hợp, tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải bởi thông số khuếch đại đầu vào được xác định từ trước, và định trọng số, bởi bộ phận kết hợp, bộ chỉ thị hoạt động giọng nói bởi thông số khuếch đại lời nói được xác định từ trước. Do đó, có thể đạt được việc điều khiển một cách hiệu quả độ lớn của thành phần giọng nói liên quan tới độ lớn của thành phần không phải là giọng nói.

Theo dạng áp dụng thứ mười một của phương pháp xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ mười của khía cạnh thứ hai, phương pháp chứa

bước cộng, bởi bộ phận kết hợp, tín hiệu audio kênh trái vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trái được kết hợp, cộng, bởi bộ phận kết hợp, tín hiệu audio kênh trung tâm vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh trung tâm được kết hợp, và cộng, bởi bộ phận kết hợp, tín hiệu audio kênh phải vào tổ hợp của tín hiệu audio kênh trái được định trọng số với bộ chỉ thị hoạt động giọng nói để thu tín hiệu audio kênh phải được kết hợp. Do đó, việc kết hợp được thực hiện một cách hiệu quả. Các thành phần giọng nói được tách được kết hợp với các tín hiệu gốc để tăng cường thành phần giọng nói trong các tín hiệu đầu ra.

Theo dạng áp dụng thứ mười hai của phương pháp xử lý tín hiệu theo dạng áp dụng thứ năm tới dạng áp dụng thứ mười một của khía cạnh thứ hai, tín hiệu audio đa kênh còn chứa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, trong đó, phương pháp chứa bước xác định, bởi bộ phát hiện hoạt động giọng nói, bộ chỉ thị hoạt động giọng nói còn dựa trên cơ sở của tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải. Do đó, các kênh bao quanh nằm trong tín hiệu audio đa kênh cũng được xem xét để xác định bộ chỉ thị hoạt động giọng nói, tạo thành ước lượng tốt hơn của bộ chỉ thị hoạt động giọng nói.

Theo dạng áp dụng thứ mười ba của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước biến đổi, bởi bộ biến đổi, tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải từ miền thời gian vào miền tần số. Do đó, có thể đạt được việc biến đổi một cách hiệu quả của các tín hiệu audio vào miền tần số. Điều này được yêu cầu, ví dụ, nếu việc tăng cường lời nói và việc phát hiện hoạt động giọng nói được thực hiện trong miền tần số.

Theo dạng áp dụng thứ mười bốn của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước biến đổi ngược, bởi bộ biến đổi ngược, tín hiệu audio kênh trái được kết hợp, tín hiệu audio kênh trung tâm được kết hợp, và tín hiệu audio kênh phải được kết hợp từ miền tần số vào miền thời gian. Do đó, có thể đạt được việc biến đổi ngược một

cách hiệu quả cho các tín hiệu audio vào miền thời gian, và thu được các tín hiệu đầu ra trong miền thời gian.

Theo dạng áp dụng thứ mười lăm của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ trộn tăng, tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải dựa trên cơ sở của tín hiệu audio âm thanh nổi kênh trái đầu vào và tín hiệu audio âm thanh nổi kênh phải đầu vào. Theo cách này, phương pháp xử lý tín hiệu có thể được áp dụng cho việc xử lý tín hiệu audio âm thanh nổi đầu vào.

Theo dạng áp dụng thứ mười sáu của phương pháp xử lý tín hiệu theo dạng áp dụng thứ mười lăm của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ trộn tăng, tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải theo các phương trình sau:

$$C = \alpha \times (L_{in} + R_{in})$$

$$L = L_{in} - C$$

$$R = R_{in} - C$$

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right)$$

trong đó, L_r biểu thị phần thực của tín hiệu audio âm thanh nổi kênh trái đầu vào, R_r biểu thị phần thực của tín hiệu audio âm thanh nổi kênh phải đầu vào, L_i biểu thị phần ảo của tín hiệu audio âm thanh nổi kênh trái đầu vào, R_i biểu thị phần ảo của tín hiệu audio âm thanh nổi kênh phải đầu vào, α biểu thị tham số trực giao, L_{in} biểu thị tín hiệu audio âm thanh nổi kênh trái đầu vào, R_{in} biểu thị tín hiệu audio âm thanh nổi kênh phải đầu vào, L biểu thị tín hiệu audio kênh trái, C biểu thị tín hiệu audio kênh trung tâm, và R biểu thị tín hiệu audio kênh phải. Do đó, có thể đạt được việc tách một cách hiệu quả cho kênh trung tâm của tín hiệu audio âm thanh nổi đầu vào sử dụng việc phân tách trực giao. Tín hiệu audio kênh trái và tín hiệu audio kênh phải tạo thành là trực giao với nhau.

Theo dạng áp dụng thứ mười bảy của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước xác định, bởi bộ trộn giảm, tín hiệu audio âm thanh nổi kênh trái đầu ra và tín hiệu audio âm thanh nổi kênh phải đầu ra dựa trên cơ sở của tín hiệu audio kênh trái được kết hợp, tín hiệu audio kênh trung tâm được kết hợp, và tín hiệu audio kênh phải được kết hợp. Do đó, hai kênh, tức là kênh trái và kênh phải, tín hiệu audio âm thanh nổi đầu ra sẽ được tạo ra một cách hiệu quả.

Theo dạng áp dụng thứ mười tám của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, số đo độ lớn chứa công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu. Do đó, số đo độ lớn có thể chỉ thị các trị số khác nhau tại các thang khác nhau.

Theo dạng áp dụng thứ mười chín của phương pháp xử lý tín hiệu theo khía cạnh thứ hai như vậy hoặc dạng áp dụng bất kỳ ở trên của khía cạnh thứ hai, phương pháp chứa bước định trọng số, bởi bộ phận kết hợp, tín hiệu audio kênh trái, tín hiệu audio kênh trung tâm, và tín hiệu audio kênh phải bởi thông số khuếch đại đầu vào được xác định từ trước, và định trọng số, bởi bộ phận kết hợp, tín hiệu audio kênh trái được định trọng số, tín hiệu audio kênh trung tâm được định trọng số, và tín hiệu audio kênh phải được định trọng số bởi thông số khuếch đại lời nói được xác định từ trước. Do đó, có thể đạt được việc điều khiển một cách hiệu quả độ lớn của thành phần giọng nói liên quan tới độ lớn của thành phần không phải là giọng nói.

Theo khía cạnh thứ ba, sáng chế đề xuất chương trình máy tính chứa mã chương trình để thực hiện phương pháp theo khía cạnh thứ hai như vậy hoặc dạng bất kỳ trong số các dạng áp dụng của khía cạnh thứ hai khi được thực thi trên máy tính. Do đó, phương pháp có thể được thực hiện một cách tự động.

Thiết bị xử lý tín hiệu có thể lập trình được, được bố trí để thực thi chương trình máy tính và/hoặc mã chương trình.

Sáng chế có thể được áp dụng trong phần cứng và/hoặc phần mềm.

Mô tả vắn tắt các hình vẽ

Các phương án thực hiện của sáng chế sẽ được mô tả liên quan tới các hình vẽ kèm theo, trong đó:

Fig.1 thể hiện giản đồ của thiết bị xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện;

Fig.2 thể hiện giản đồ của phương pháp xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện;

Fig.3 thể hiện giản đồ của thiết bị xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện;

Fig.4 thể hiện giản đồ của bộ trộn tăng của thiết bị xử lý tín hiệu theo một phương án thực hiện;

Fig.5 thể hiện giản đồ của bộ lọc của thiết bị xử lý tín hiệu theo một phương án thực hiện;

Fig.6 thể hiện giản đồ của bộ phát hiện hoạt động giọng nói của thiết bị xử lý tín hiệu theo một phương án thực hiện; và

Fig.7 thể hiện giản đồ của thiết bị xử lý tín hiệu để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện.

Các số chỉ dẫn giống nhau được sử dụng cho các đặc điểm giống nhau hoặc tương đương.

Mô tả chi tiết sáng chế

Fig.1 thể hiện giản đồ của thiết bị xử lý tín hiệu 100 để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện. Tín hiệu audio đa kênh chứa tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R. thiết bị xử lý tín hiệu 100 chứa bộ lọc 101 và bộ phận kết hợp 103.

Bộ lọc 101 được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên cơ sở của tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R, để thu hàm tăng ích ăng ten G dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm C và số đo thể hiện

độ lớn tổng thể của tín hiệu audio đa kênh, và để định trọng số tín hiệu audio kênh trái L bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trái được định trọng số L_E , để định trọng số tín hiệu audio kênh trung tâm C bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trung tâm được định trọng số C_E , và để định trọng số tín hiệu audio kênh phải R bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh phải được định trọng số R_E .

Bộ phận kết hợp 103 được đặt cấu hình để kết hợp tín hiệu audio kênh trái L với tín hiệu audio kênh trái được định trọng số L_E để thu tín hiệu audio kênh trái được kết hợp L_{EV} , để kết hợp tín hiệu audio kênh trung tâm C với tín hiệu audio kênh trung tâm được định trọng số C_E để thu tín hiệu audio kênh trung tâm được kết hợp C_{EV} , và để kết hợp tín hiệu audio kênh phải R với tín hiệu audio kênh phải được định trọng số R_E để thu tín hiệu audio kênh phải được kết hợp R_{EV} .

Các tín hiệu audio đa kênh có thể chứa, ví dụ các tín hiệu audio âm thanh nổi ba kênh, chứa chỉ tín hiệu audio kênh trái L, tín hiệu audio kênh phải và tín hiệu audio kênh trung tâm C, và cũng có thể được đề cập tới như là âm thanh nổi LCR hoặc các tín hiệu audio âm thanh nổi 3.0, các tín hiệu audio đa kênh 5.1, chứa tín hiệu audio kênh trái L, tín hiệu audio kênh phải R, tín hiệu audio kênh trung tâm C, tín hiệu audio kênh bao quanh trái L_S , tín hiệu audio kênh bao quanh phải R_S , và tín hiệu kênh trầm B, hoặc các tín hiệu đa kênh khác có tín hiệu audio kênh trung tâm và ít nhất hai tín hiệu audio kênh khác. Các tín hiệu audio khác với tín hiệu audio kênh trung tâm C, tức là tín hiệu audio kênh trái L, tín hiệu audio kênh phải R, tín hiệu audio kênh bao quanh trái L_S , tín hiệu audio kênh bao quanh phải R_S và tín hiệu kênh trầm B, cũng có thể được đề cập tới như là các tín hiệu audio không phải là kênh trung tâm. Trong trường hợp của tín hiệu audio đa kênh 5.1, số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh có thể được thu làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm, số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái và tín hiệu audio kênh phải, số đo độ lớn của chênh lệch giữa tín hiệu audio kênh bao quanh trái và tín hiệu audio kênh bao quanh phải, và số đo độ lớn của tín hiệu audio kênh ảnh hưởng tần số thấp. Trong trường hợp của tín hiệu audio đa kênh 5.1, bộ lọc thu được có thể được sử dụng để định trọng số tất cả các tín hiệu trong các tín hiệu audio được chứa.

Fig.2 thể hiện giản đồ của phương pháp xử lý tín hiệu 200 để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện. Tín hiệu audio đa kênh chứa tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R.

Phương pháp xử lý tín hiệu 200 chứa bước xác định 201 số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên cơ sở của tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R, thu 203 hàm tăng ích ăng ten G dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm C và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, định trọng số 205 tín hiệu audio kênh trái L bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trái được định trọng số L_E , định trọng số 207 tín hiệu audio kênh trung tâm C bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trung tâm được định trọng số C_E , định trọng số 209 tín hiệu audio kênh phải R bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh phải được định trọng số R_E , kết hợp 211 tín hiệu audio kênh trái L với tín hiệu audio kênh trái được định trọng số L_E để thu tín hiệu audio kênh trái được kết hợp L_{EV} , kết hợp 213 tín hiệu audio kênh trung tâm C với tín hiệu audio kênh trung tâm được định trọng số C_E để thu tín hiệu audio kênh trung tâm được kết hợp C_{EV} , và kết hợp 215 tín hiệu audio kênh phải R với tín hiệu audio kênh phải được định trọng số R_E để thu tín hiệu audio kênh phải được kết hợp R_{EV} .

Phương pháp xử lý tín hiệu 200 có thể được thực hiện bởi thiết bị xử lý tín hiệu 100, tức là bởi bộ lọc 101 và bộ phận kết hợp 103.

Trong phần dưới đây, các dạng áp dụng khác và các phương án thực hiện của thiết bị xử lý tín hiệu 100 và phương pháp xử lý tín hiệu 200 sẽ được mô tả.

Sáng chế đề cập đến lĩnh vực xử lý tín hiệu audio. Thiết bị xử lý tín hiệu 100 và phương pháp xử lý tín hiệu 200 có thể được áp dụng cho việc tăng cường giọng nói, tức là tăng cường cho hội thoại, nằm trong các tín hiệu audio, tức là các tín hiệu audio âm thanh nói. Cụ thể, thiết bị xử lý tín hiệu 100 và phương pháp xử lý tín hiệu 200 có thể, trong kết hợp với bộ trộn tăng 301 hoặc trong kết hợp với bộ trộn tăng 301 và bộ trộn giảm 303, được áp dụng để xử lý các tín hiệu audio âm thanh nói để cải thiện sự rõ ràng trong hội thoại.

Có các thiết bị khác biệt có hai loa, như các TV, các máy tính xách tay, các máy tính bảng, các điện thoại di động và các điện thoại thông minh. Khi các tín hiệu audio âm thanh nổi được chơi lại sử dụng thiết bị này, các thành phần giọng nói của các rãnh âm thanh từ các phim, ví dụ, có thể khó để hiểu cho những người nghe thông thường và bị khuyết tật thính giác. Nó là đặc biệt cho trường hợp trong các môi trường ồn ào hoặc khi thành phần giọng nói bị lấn át bởi các thành phần không phải là giọng nói hoặc các âm thanh như âm nhạc hoặc các hiệu ứng âm thanh.

Các phương án thực hiện của sáng chế nhằm mục đích, cụ thể là tăng cường thành phần giọng nói của các tín hiệu audio âm thanh nổi để cải thiện độ rõ ràng của hội thoại. Một giả sử ngầm là giọng nói, hoặc lời nói tương đương, bị dịch chuyển trung tâm trong tín hiệu audio đa kênh, điều này nói chung là đúng cho hầu hết các tín hiệu audio âm thanh nổi. Mục đích là để tăng cường độ to của các thành phần giọng nói mà không ảnh hưởng tới chất lượng giọng nói, trong khi các thành phần không phải là giọng nói được để lại không thay đổi. Điều này là đặc biệt khả thi trong suốt các khoảng gián đoạn thời gian có đồng thời giọng nói và các thành phần không phải là giọng nói. Các phương án thực hiện của sáng chế cho phép, ví dụ, để sử dụng chỉ tín hiệu audio âm thanh nổi và không cần hoặc áp dụng hiệu biết khác từ kênh audio giọng nói tách biệt hoặc tín hiệu audio đa kênh 5.1 gốc. Có thể đạt được các mục đích này bằng cách tách tín hiệu audio kênh trung tâm ảo và tăng cường tín hiệu audio kênh trung tâm này cũng như các tín hiệu audio khác sử dụng thiết bị xử lý tín hiệu được mô tả 100 hoặc phương pháp xử lý tín hiệu 200. Ngoài ra, giải pháp cho việc phát hiện hoạt động giọng nói có thể được sử dụng để đảm bảo rằng các thành phần không phải là giọng nói có thể không bị ảnh hưởng bởi việc xử lý. Các phương án thực hiện khác của sáng chế có thể được sử dụng để xử lý các tín hiệu audio đa kênh khác, như tín hiệu audio đa kênh 5.1.

Các phương án thực hiện của sáng chế là dựa trên giải pháp sau, trong đó, từ việc ghi tín hiệu audio âm thanh nổi, tín hiệu audio kênh trung tâm được tách sử dụng giải pháp trộn lên. Tín hiệu audio kênh trung tâm này còn có thể được xử lý sử dụng việc tăng cường giọng nói và việc phát hiện hoạt động giọng nói, để thu ước lượng của thành phần giọng nói gốc. Đặc điểm của giải pháp có thể là thành phần giọng nói có thể không chỉ được tách từ tín hiệu audio kênh trung tâm, mà còn từ các tín hiệu audio

kênh còn lại. Do quy trình trộn tăng không làm việc một cách hoàn hảo nên các tín hiệu audio kênh còn lại có thể vẫn chứa thành phần giọng nói. Khi các thành phần giọng nói cũng được tách và được gia tăng, tín hiệu audio đầu ra tạo thành sẽ có chất lượng giọng nói và độ rộng được cải thiện.

Trong phần dưới đây, cụ thể là các phương án thực hiện của sáng chế để tăng cường thành phần giọng nói của tín hiệu audio đa kênh LCR (chứa tín hiệu audio kênh trung tâm, tín hiệu audio kênh trái, và tín hiệu audio kênh phải), được thu từ tín hiệu audio âm thanh nổi hai kênh bởi việc trộn tăng từ 2 thành 3, sẽ được mô tả dựa trên các hình vẽ từ Fig.3 đến Fig.7.

Tuy nhiên, các phương án thực hiện của sáng chế không bị hạn chế vào các tín hiệu audio đa kênh này và cũng có thể chứa việc xử lý của ba tín hiệu audio kênh LCR, tức là nhận được từ các thiết bị khác, hoặc việc xử lý của các tín hiệu đa kênh khác chứa tín hiệu audio kênh trung tâm, tức là các tín hiệu đa kênh 5.1 hoặc 7.1. Các phương án thực hiện khác thậm chí còn có thể được đặt cấu hình để xử lý các tín hiệu đa kênh, không chứa tín hiệu audio kênh trung tâm, tức là tín hiệu đa kênh 4.0 chứa tín hiệu kênh audio trái và phải và tín hiệu kênh bao quanh trái và phải, bằng cách trộn tăng tín hiệu đa kênh để thu tín hiệu audio kênh trung tâm ảo trước khi áp dụng việc tăng cường giọng nói hoặc hội thoại với hoặc không với việc phát hiện hoạt động giọng nói.

Fig.3 thể hiện giản đồ của thiết bị xử lý tín hiệu 100 để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện. Thiết bị xử lý tín hiệu 100 chứa bộ lọc 101, bộ phận kết hợp 103, bộ trộn tăng 301, và bộ trộn giảm 303. Bộ lọc 101 và bộ phận kết hợp 103 chứa bộ xử lý kênh trái 305, bộ xử lý kênh trung tâm 307, và bộ xử lý kênh phải 309.

Bộ trộn tăng 301 được đặt cấu hình để xác định tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R dựa trên cơ sở của tín hiệu audio âm thanh nổi kênh trái đầu vào L_{in} và tín hiệu audio âm thanh nổi kênh phải đầu vào R_{in} . Theo cách khác, bộ trộn tăng 301 tạo ra việc trộn tăng từ 2 thành 3, như sẽ được giải thích chi tiết hơn dưới dạng ví dụ dựa trên Fig.4.

Bộ xử lý kênh trái 305 được đặt cấu hình để xử lý tín hiệu audio kênh trái L để tạo ra tín hiệu audio kênh trái được kết hợp L_{EV} . Bộ xử lý kênh trung tâm 307 được đặt

cấu hình để xử lý tín hiệu audio kênh trung tâm C để tạo ra tín hiệu audio kênh trung tâm được kết hợp C_{EV} . Bộ xử lý kênh phải 309 được đặt cấu hình để xử lý tín hiệu audio kênh phải R để tạo ra tín hiệu audio kênh phải được kết hợp R_{EV} . Bộ xử lý kênh trái 305, bộ xử lý kênh trung tâm 307, và bộ xử lý kênh phải 309 được đặt cấu hình để thực hiện việc tăng cường giọng nói, ENH, như sẽ được giải thích chi tiết hơn ở dưới, dưới dạng ví dụ dựa trên Fig.5. Bộ xử lý kênh trái 305, bộ xử lý kênh trung tâm 307, và bộ xử lý kênh phải 309 có thể còn được đặt cấu hình để xử lý bộ chỉ thị hoạt động giọng nói được tạo ra bởi việc phát hiện hoạt động giọng nói, VAD, như sẽ được giải thích chi tiết hơn ở dưới, dưới dạng ví dụ dựa trên Fig.6.

Bộ trộn giảm 303 được đặt cấu hình để xác định tín hiệu audio âm thanh nổi kênh trái đầu ra L_{out} và tín hiệu audio âm thanh nổi kênh phải đầu ra R_{out} dựa trên cơ sở của tín hiệu audio kênh trái được kết hợp L_{EV} , tín hiệu audio kênh trung tâm được kết hợp C_{EV} , và tín hiệu audio kênh phải được kết hợp R_{EV} . Theo cách khác, bộ trộn giảm 303 tạo ra việc trộn giảm từ 3 thành 2.

Do đó, các tín hiệu audio được tăng cường giọng nói được xử lý theo cách mà tín hiệu âm thanh nổi hai kênh được trộn giảm L_{out} và R_{out} có thể được đưa ra một cách trực tiếp tới thiết bị chơi lại hai kênh truyền thống, tức là bộ TV âm thanh nổi truyền thống.

Theo một phương án thực hiện của sáng chế, giải pháp chung được sử dụng bởi bộ trộn tăng 301 cho việc tách kênh trung tâm từ tín hiệu audio âm thanh nổi đầu vào chứa tín hiệu audio âm thanh nổi kênh trái đầu vào L_{in} và tín hiệu audio âm thanh nổi kênh phải đầu vào R_{in} . Nó tạo thành các tín hiệu audio kênh trái, trung tâm và kênh phải, được biểu diễn như là L, C, và R. Các phương án thực hiện khác của sáng chế có thể sử dụng các giải pháp cho việc trộn lên. Các phương án thực hiện khác của sáng chế là có thể hiểu được, trong đó, tức là tín hiệu audio đa kênh 5.1 là sẵn có và các kênh trái, trung tâm và kênh phải được tạo thành được sử dụng một cách trực tiếp.

Các tín hiệu audio kênh trái, trung tâm, và kênh phải L, C, và R được xử lý theo cách được cải tiến để ước lượng bộ lọc tăng cường giọng nói phụ thuộc thời gian và/hoặc tần số 101 mà sau đó có thể được áp dụng trên tất cả các kênh của tín hiệu audio đa kênh. Bộ lọc 101 này được đặt cấu hình để làm suy giảm các thành phần

không phải là giọng nói có thể có mặt đồng thời với thành phần giọng nói. Khác biệt liên quan tới các giải pháp khác là việc không chỉ tín hiệu audio kênh trung tâm, mà các tín hiệu audio khác, tức là tín hiệu audio kênh trái và tín hiệu audio kênh phải trong trường hợp LCR như được mô tả trên Fig.3, được xử lý với cùng một bộ lọc 101. Các phương án thực hiện của sáng chế sử dụng giải pháp được cải tiến để định nghĩa việc tăng cường giọng nói bộ lọc 101.

Ngoài ra, việc phát hiện hoạt động giọng nói có thể được thực hiện bằng cách sử dụng giải pháp được cải tiến, khai thác thông tin từ tất cả các kênh của tín hiệu audio đa kênh. Đầu ra của bộ phát hiện hoạt động giọng nói, tức là bộ chỉ thị hoạt động giọng nói, có thể là quyết định mềm mà có thể chỉ thị hoạt động giọng nói. Tổ hợp của việc tăng cường giọng nói và việc phát hiện hoạt động giọng nói tạo ra tín hiệu audio đa kênh chỉ hoặc ít nhất là hầu hết chỉ chứa thành phần giọng nói. Thành phần giọng nói tín hiệu audio đa kênh này có thể được gia tăng và được cộng vào tín hiệu audio đa kênh gốc bởi bộ phận kết hợp 103 để thu các tín hiệu audio kênh được kết hợp LEV , CEV , và REV . Việc trộn giảm âm thanh nổi có thể được thực hiện bởi bộ trộn giảm 303 để tạo ra các tín hiệu audio âm thanh nổi kênh đầu ra cuối cùng L_{out} và R_{out} .

Fig.4 thể hiện giản đồ của bộ trộn tăng 301 của thiết bị xử lý tín hiệu 100 theo một phương án thực hiện. Bộ trộn tăng 301 được đặt cấu hình để xác định tín hiệu audio kênh trái L , tín hiệu audio kênh trung tâm C , và tín hiệu audio kênh phải R dựa trên cơ sở của tín hiệu audio âm thanh nổi kênh trái đầu vào L_{in} và tín hiệu audio âm thanh nổi kênh phải đầu vào R_{in} . Bộ trộn tăng 301 tạo ra việc trộn tăng từ 2 đến 3. Bộ trộn tăng 301 được đặt cấu hình để thực hiện việc tách của tín hiệu audio kênh trung tâm C từ tín hiệu audio âm thanh nổi hai kênh đầu vào sử dụng giải pháp trộn lên.

Quy trình để thu tín hiệu audio kênh trung tâm ảo C từ, ví dụ, tín hiệu audio âm thanh nổi đầu vào hai kênh cũng được đề cập tới như là việc tách trung tâm. Điều này có thể được mong muốn khi chỉ tín hiệu audio âm thanh nổi truyền thống của việc ghi là có sẵn. Có các giải pháp khác nhau để đạt được việc tách trung tâm. Một họ các giải pháp trộn tăng là dựa trên việc giải mã ma trận. Các giải pháp này là các giải pháp độc lập với tín hiệu tuyến tính cho việc trộn lên. Chúng có thể được ghép với bộ giải mã ma trận và làm việc trong miền thời gian. Các giải pháp hình học, mặt khác lại phụ

thuộc vào tín hiệu. Các giải pháp này có thể dựa trên giả sử rằng tín hiệu audio kênh trái L và tín hiệu audio kênh phải R là không tương quan với nhau. Các giải pháp này làm việc trong miền tần số.

Trong phần dưới đây, giải pháp đặc biệt được mô tả như là ví dụ của việc tách trung tâm, có thể được sử dụng trong phương án thực hiện bất kỳ của sáng chế. Giải pháp này được thực hiện trong miền tần số. Nó có nghĩa là tín hiệu audio âm thanh nổi đầu vào được biến đổi vào miền tần số, tức là bằng cách áp dụng thuật toán biến đổi Fourier rời rạc (discrete Fourier transform - DFT) trong các cửa sổ thời gian ngắn. Lựa chọn thích hợp cho kích cỡ khối của biến đổi Fourier rời rạc (discrete Fourier transform - DFT) có thể là 1024 khi tần số lấy mẫu là 48000Hz được sử dụng.

Giải pháp xây dựng trên giả sử rằng các tín hiệu audio kênh trái và kênh phải L và R là trực giao với nhau. Ý tưởng là để thu tín hiệu audio kênh trung tâm C theo

$$C = \alpha \times (L_{in} + R_{in}) \quad (1)$$

trong đó, α là tham số được xác định. Các tín hiệu audio kênh trái và kênh phải L và R sau đó có thể được dẫn ra như sau

$$L = L_{in} - C \quad (2)$$

$$R = R_{in} - C \quad (3)$$

từ tín hiệu audio kênh trung tâm C tạo thành. Tham số α có thể được tối ưu hóa theo cách để đáp ứng ràng buộc

$$L \times R^* = 0 \quad (4)$$

mô tả tính trực giao của các tín hiệu audio. Lời giải toán học cho vấn đề này có thể được dẫn ra, tạo ra kết quả

$$\alpha = \frac{1}{2} \times \left(1 - \sqrt{\frac{(L_r - R_r)^2 + (L_i - R_i)^2}{(L_r + R_r)^2 + (L_i + R_i)^2}} \right) \quad (5)$$

trong đó, L_r , L_i , R_r và R_i biểu thị các phần thực và phần ảo của các thành phần phổ của các tín hiệu audio âm thanh nổi trái và phải đầu ra L_{in} và R_{in} , một cách tương ứng. Tham số α là phụ thuộc vào thời gian và phụ thuộc vào tần số và do đó có thể

được tính toán cho tất cả các ngăn tần số (frequency bin) của khung đã cho của các mẫu tín hiệu audio.

Các giải pháp hình học cụ thể khác cho việc tách trung tâm có thể được áp dụng. Các giải pháp cụ thể khác sử dụng, ví dụ, phân tích thành phần chính cho việc tách trung tâm.

Fig.5 thể hiện giản đồ của bộ lọc 101 của thiết bị xử lý tín hiệu 100 theo một phương án thực hiện. Bộ lọc 101 chứa bộ trừ 501, bộ phận xác định 503, bộ phận xác định 505, bộ phận xác định 507, bộ phận định trọng số 509, bộ phận định trọng số 511, và bộ phận định trọng số 513. Giản đồ minh họa giải pháp tăng cường giọng nói.

Bộ trừ 501 được đặt cấu hình để trừ tín hiệu audio kênh phải R từ tín hiệu audio kênh trái L để thu tín hiệu audio còn dư S.

Bộ phận xác định 503 được đặt cấu hình để xác định độ lớn hoặc công suất được bình phương của tín hiệu audio kênh trung tâm C để thu số đo độ lớn P_C của tín hiệu audio kênh trung tâm C. Bộ phận xác định 505 được đặt cấu hình để xác định độ lớn hoặc công suất được bình phương của tín hiệu audio còn dư S để thu số đo độ lớn P_S của tín hiệu audio còn dư S.

Bộ phận xác định 507 được đặt cấu hình để xác định tỉ lệ giữa số đo độ lớn P_C của tín hiệu audio kênh trung tâm C và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh để thu hàm tăng ích ăng ten G. Số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh được tạo thành bởi tổng của số đo độ lớn P_C của tín hiệu audio kênh trung tâm C và số đo độ lớn P_S của tín hiệu audio còn dư S. Hàm tăng ích ăng ten G có thể là độc lập với thời gian và/hoặc độc lập với tần số. Chỉ số thời gian mẫu được biểu diễn là m. Chỉ số ngăn tần số (frequency bin) được biểu diễn là k.

Bộ phận định trọng số 509 được đặt cấu hình để định trọng số tín hiệu audio kênh trái L bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trái được định trọng số L_E . Bộ phận định trọng số 511 được đặt cấu hình để định trọng số tín hiệu audio kênh trung tâm C bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh trung tâm được định trọng số C_E . Bộ phận định trọng số 513 được đặt cấu hình để định trọng số tín hiệu audio kênh phải R bởi hàm tăng ích ăng ten G để thu tín hiệu audio kênh phải được định trọng số R_E .

Các phương án thực hiện của sáng chế sử dụng thông tin từ các tín hiệu audio kênh trái, trung tâm và kênh phải L, C, và R để ước lượng hàm tăng ích ăng ten G theo giải pháp lọc Wiener cho việc tăng cường giọng nói. Giải pháp lọc Wiener có thể được áp dụng trên tất cả các kênh của tín hiệu audio đa kênh để loại bỏ các thành phần không phải là giọng nói. Trong trường hợp tín hiệu audio kênh trung tâm C chứa thành phần giọng nói, giải pháp lọc Wiener (hầu hết) chỉ giữ các thành phần giọng nói của tất cả các kênh của tín hiệu audio đa kênh.

Nói chung, giải pháp tăng cường giọng nói được sử dụng có thể giải quyết nhiều phụ. Do đó, tín hiệu đầu vào Y của kênh bất kỳ có thể được xem như là $Y = X + N$, trong đó, X chứa thành phần giọng nói sạch và N có thể được coi là nhiễu phụ. Nó được giả sử rằng X và N là không tương quan so với nhau. Để loại bỏ N từ tín hiệu audio quan sát được Y, mật độ phổ công suất nhiễu của nhiễu phụ N hoặc tỉ lệ tín hiệu so với nhiễu trước đó X/N có thể được ước lượng. Hàm tăng ích ăng ten phụ thuộc tần số G hoặc $G(m,k)$ có thể được thu như

$$G = \frac{\frac{X}{N}}{1 + \frac{X}{N}} = \frac{X}{X + N} \quad (6)$$

và ước lượng của tín hiệu audio chứa thành phần giọng nói sạch có thể được xác định như là $\hat{X} = G \times Y$, làm việc trên tất cả các ngăn tần số (frequency bin) của tín hiệu audio.

Giải pháp tăng cường giọng nói khai thác giả sử rằng tín hiệu audio kênh trung tâm C chứa hầu hết là giọng nói. Do thông thường không có giải pháp tách trung tâm nào tạo ra được việc tách trung tâm hoàn hảo nên tín hiệu audio kênh trung tâm C có thể chứa các thành phần không phải là giọng nói và các kênh khác của tín hiệu audio đa kênh có thể chứa các thành phần giọng nói. Do đó, mục đích là để loại bỏ các thành phần không phải là giọng nói trong tín hiệu audio kênh trung tâm C và để cách ly các thành phần giọng nói trong các kênh khác của tín hiệu audio đa kênh. Để đạt được mục đích này, giải pháp lọc Wiener có thể được áp dụng để ước lượng hàm tăng ích ăng ten G. Thay cho việc sử dụng các cách tiếp cận phức tạp để ước lượng mật độ phổ công suất nhiễu của nhiễu phụ N, giải pháp đơn giản nhưng hiệu quả để định nghĩa X

và N cho giải pháp lọc Wiener được sử dụng, như được định nghĩa bởi các phương trình (7), (8), và (9). Tín hiệu audio kênh trung tâm C được coi như chứa thành phần giọng nói, tương ứng với X, trong khi thành phần của các kênh khác của tín hiệu audio đa kênh được coi như là chứa nhiễu, tương ứng với N.

Theo một phương án thực hiện, tín hiệu audio còn dư S được thu từ các tín hiệu audio kênh trái và kênh phải bởi bộ trừ 501, tức là theo $S = L - R$. Theo cách này, các thành phần trung tâm được loại bỏ khỏi tín hiệu còn dư. Các công suất có thể được xác định từ phổ của tín hiệu audio kênh trung tâm C bởi bộ phận xác định 503 và phổ của tín hiệu audio còn dư S bởi bộ phận xác định 505 theo

$$P_C(m, k) = |C(m, k)|^2 \quad (7)$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2 \quad (8)$$

trong đó, m là chỉ số thời gian mẫu và k là chỉ số tần số (frequency bin). Giải pháp có thể khác là sử dụng độ lớn thay cho công suất, hoặc độ lớn logarit hoặc công suất logarit. Theo cách phương án thực hiện khác, các công suất có thể được làm trơn tru qua thời gian để làm giảm việc xử lý các thành phần lạ.

Hàm tăng ích ăng ten G sau đó được xác định bởi bộ phận xác định 507 theo giải pháp lọc Wiener theo

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)} \quad (9)$$

Hàm tăng ích ăng ten G tiếp theo được áp dụng cho các tín hiệu audio kênh trái, trung tâm và kênh phải L, C, và R bởi các bộ phận định trọng số 509-513, một cách tương ứng. Nó tạo thành tín hiệu audio kênh trái được định trọng số L_E , tín hiệu audio kênh trung tâm được định trọng số C_E , và tín hiệu audio kênh phải được định trọng số R_E .

Trong trường hợp tín hiệu audio kênh trung tâm gốc C chỉ chứa thành phần giọng nói, các tín hiệu audio được định trọng số được tăng cường cũng chỉ chứa các thành phần giọng nói.

Theo phương án thực hiện của sáng chế, định dạng tín hiệu audio đa kênh khác được sử dụng. Với tín hiệu audio đa kênh 5.1 ví dụ, lựa chọn để xác định tín hiệu audio còn dư S là

$$S = L - R + L_s - R_s, \quad (10)$$

trong đó, L biểu thị tín hiệu audio kênh trái, R biểu thị tín hiệu audio kênh phải, L_s biểu thị tín hiệu audio kênh bao quanh trái, và R_s biểu thị tín hiệu audio kênh bao quanh phải. Theo phương án thực hiện khác, công suất P_s có thể được xác định làm tổng của công suất của L-R và công suất của $L_s - R_s$.

Tín hiệu audio còn dư S và công suất của tín hiệu audio còn dư P_s có thể được xác định một cách tương ứng sử dụng các định dạng tín hiệu audio đa kênh khác, như định dạng tín hiệu audio đa kênh 7.1.

Để tiếp tục làm giảm độ phức tạp tính toán, các ngăn tần số (frequency bin) của các tín hiệu audio có thể được nhóm lại cùng nhau thành các băng tần số, tức là theo thang tần số Mel. Trong trường hợp này, hàm tăng ích ăng ten G có thể được xác định cho từng ngăn tần số (frequency bin).

Ngoài ra, việc xử lý chỉ trên các tần số có thể có khả năng chứa giọng nói của người, tức là nằm trong khoảng giới hạn tần số từ 100Hz đến 8000Hz, giúp lọc ra các thành phần không phải là giọng nói.

Các phương án thực hiện của việc tăng cường giọng nói loại bỏ các thành phần không phải là giọng nói không được mong muốn bị rò rỉ vào trong tín hiệu audio kênh trung tâm C trong suốt quá trình trộn lên. Ngoài ra, nó gia tăng các thành phần trực tiếp bị rò rỉ vào trong các kênh khác của tín hiệu audio đa kênh.

Fig.6 thể hiện giản đồ của bộ phát hiện hoạt động giọng nói 601 của thiết bị xử lý tín hiệu 100 theo một phương án thực hiện. Bộ phát hiện hoạt động giọng nói 601 được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói V dựa trên cơ sở của tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R, trong đó, bộ chỉ thị hoạt động giọng nói V chỉ thị độ lớn của thành phần giọng nói nằm trong tín hiệu audio đa kênh qua thời gian. Bộ phát hiện hoạt động giọng nói 601 chứa bộ trừ 603, bộ phận xác định 605, bộ phận xác định 607, bộ làm trễ 609, bộ làm trễ

611, bộ trừ 613, bộ trừ 615, bộ phận xác định 617, bộ phận xác định 619, và bộ phận xác định 621.

Bộ trừ 603 được đặt cấu hình để trừ tín hiệu audio kênh phải R từ tín hiệu audio kênh trái L để thu tín hiệu audio còn dư S. Bộ phận xác định 605 được đặt cấu hình để xác định độ lớn của tín hiệu audio kênh trung tâm C để thu $|C(m,k)|$, trong đó, m biểu thị chỉ số thời gian mẫu và k biểu thị chỉ số ngăn tần số (frequency bin). Bộ phận xác định 607 được đặt cấu hình để xác định độ lớn của tín hiệu audio còn dư S để thu $|S(m,k)|$, trong đó, m biểu thị chỉ số thời gian mẫu và k biểu thị chỉ số ngăn tần số (frequency bin). Bộ làm trễ 609 được đặt cấu hình để làm trễ $|C(m,k)|$ bởi chu kỳ thời gian lấy mẫu để thu $|C(m-1,k)|$. Bộ làm trễ 611 được đặt cấu hình để làm trễ $|S(m,k)|$ bởi chu kỳ thời gian lấy mẫu để thu $|S(m-1,k)|$. Bộ trừ 613 được đặt cấu hình để trừ $|C(m-1,k)|$ từ $|C(m,k)|$ để thu $|C(m,k)| - |C(m-1,k)|$. Bộ trừ 615 được đặt cấu hình để trừ $|S(m-1,k)|$ từ $|S(m,k)|$ để thu $|S(m,k)| - |S(m-1,k)|$.

Bộ phận xác định 617 được đặt cấu hình để xác định số đo của sự biến đổi phổ F_C của tín hiệu audio kênh trung tâm C, ví dụ thông lượng phổ, tức là dựa trên cơ sở của tổng được bình phương Σ^2 qua tất cả các ngăn tần số (frequency bin) qua $|C(m,k)| - |C(m-1,k)|$. Bộ phận xác định 619 được đặt cấu hình để xác định số đo của sự biến đổi phổ F_S của khác biệt giữa tín hiệu audio kênh trái L và tín hiệu audio kênh phải R, ví dụ thông lượng phổ, tức là dựa trên cơ sở của tổng được bình phương Σ^2 qua tất cả các ngăn tần số (frequency bin) qua $|S(m,k)| - |S(m-1,k)|$. Bộ phận xác định 621 được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói V dựa trên cơ sở của số đo của sự biến đổi phổ F_C và số đo của sự biến đổi phổ F_S , tức là dựa trên cơ sở của tỷ số $F_C / (F_C + F_S)$.

Việc phát hiện hoạt động giọng nói chứa quy trình phát hiện theo thời gian và việc phân mảnh của giọng nói. Mục đích của việc phát hiện hoạt động giọng nói là để phát hiện giọng nói trong im lặng hoặc trong các âm thanh khác. Giải pháp này được mong muốn cho hầu hết các loại công nghệ giọng nói.

Các giải pháp khác cho việc phát hiện hoạt động giọng nói có thể được áp dụng trong các phương án thực hiện của sáng chế. Các giải pháp đơn giản là, ví dụ dựa trên năng lượng. Việc đặt ngưỡng năng lượng có thể được sử dụng để phát hiện giọng nói.

Thông thường, giải pháp này chỉ có hiệu quả cho giọng nói trong tình trạng im lặng. Các giải pháp khác chứa các giải pháp dựa trên mô hình thống kê, là dựa trên ước lượng tỉ lệ tín hiệu so với nhiễu (signal-to-noise ratio - SNR) và là tương tự như các giải pháp tăng cường giọng nói theo thống kê. Các giải pháp dựa trên mô hình tham số thường xuyên gắn các đặc điểm audio mức thấp với bộ phân loại như mô hình trộn Gaussian. Các đặc điểm audio có thể là năng lượng điều biến 4Hz, tỉ lệ cắt không, trọng tâm phổ, hoặc thông lượng phổ.

Theo phương án thực hiện của sáng chế, việc phát hiện hoạt động giọng nói được áp dụng để đảm bảo rằng chỉ các thành phần giọng nói hoặc hội thoại được gia tăng và các thành phần không phải là giọng nói được để lại không thay đổi. Tổng quan của giải pháp tăng cường giọng nói được cho trên Fig.6.

Bộ chỉ thị hoạt động giọng nói V được dẫn ra từ tín hiệu audio kênh trung tâm C và tín hiệu audio còn dư $S = L - R$, do nó có thể được thực hiện nằm trong giải pháp tăng cường giọng nói. Từ các tín hiệu audio này, thông lượng phổ được tách ra. Thông lượng phổ là số đo cho thay đổi theo thời gian của phổ. Thông lượng phổ của DFT hoặc tín hiệu miền tần số X có thể được định nghĩa như

$$F_x(m) = \sum_k (|X(m, k)| - |X(m-1, k)|)^2 \quad (11)$$

Các định nghĩa tương tự khác của thông lượng phổ cũng có thể được áp dụng cho các phương án thực hiện khác của sáng chế. Thông lượng phổ chỉ thị các thay đổi trong phân bố năng lượng phổ và thể hiện đạo hàm theo thời gian qua thời gian. Thau cho định nghĩa trong phương trình (11), trong đó, khác biệt được xác định qua hai khung phụ tín hiệu audio liên tiếp, thông lượng phổ cũng có thể được xác định như là khác biệt qua hai khối liên tiếp chứa nhiều khung phụ tín hiệu audio. Với các tín hiệu audio có các thành phần giọng nói, các trị số cao hơn của thông lượng phổ được mong đợi so sánh với âm nhạc và các âm thanh khác.

Theo phương án thực hiện của sáng chế, thiết lập kênh cụ thể, trong đó, tức là một kênh của tín hiệu audio đa kênh chứa giọng nói sơ cấp, được khai thác để dẫn ra bộ chỉ thị hoạt động giọng nói liên tục độc lập với tần số V. Thông lượng phổ F_c của

tín hiệu audio kênh trung tâm C và thông lượng phổ F_s của tín hiệu audio còn dư S sau đó có thể được xác định theo phương trình (11).

Để thu bộ chỉ thị hoạt động giọng nói V là độc lập với bất kỳ quy trình chuẩn hóa nào, bộ chỉ thị hoạt động giọng nói V có thể, tức là được tính toán như

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right) \quad (12)$$

Định nghĩa này của bộ chỉ thị hoạt động giọng nói V đảm bảo rằng $V = 0$ trong trường hợp mà $F_c = F_s$. Cuối cùng, V bị hạn chế vào $V \in [0;1]$. Tham số a biểu thị thông số định cỡ được xác định từ trước điều khiển khoảng giới hạn động của V, trong đó, $a = 4$ có thể là trị số có thể chấp nhận được tạo thành

$$V = 4 \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right) \quad (13)$$

Ngoài ra, bộ chỉ thị hoạt động giọng nói V có thể được đặt tới $V = 0$ trong trường hợp mà F_c không vượt quá ngưỡng cụ thể t. Để thu đường cong bộ chỉ thị hoạt động giọng nói trơn tru qua thời gian, việc làm trơn tru theo thời gian có thể được áp dụng cho V.

Tương tự như giải pháp tăng cường giọng nói, các giải pháp việc phát hiện hoạt động giọng nói cũng có thể được thực hiện khi các ngăn tần số (frequency bin) được nhóm thành các băng tần số, tức là theo thang tần số Mel. Ngoài ra, việc hạn chế các tần số xem xét vào khoảng giới hạn tần số của giọng nói của người, tức là từ 100Hz tới 8000Hz, còn tiếp tục cải thiện hiệu quả hoạt động.

Kết quả của giải pháp phát hiện hoạt động giọng nói là quyết định liên tục độc lập với tần số được thu bằng cách sử dụng thuật toán đơn giản và hiệu quả. Nó có thể áp dụng chỉ một số tham số có thể điều chỉnh được và có thể không sử dụng dữ liệu bất kỳ khác, ví dụ để học mô hình. Giải pháp có thể phân biệt một cách mạnh mẽ giữa giọng nói và các âm thanh khác, như âm nhạc.

Fig.7 thể hiện giản đồ của thiết bị xử lý tín hiệu 100 để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh theo một phương án thực hiện. Giản đồ minh họa quy trình trộn. Thiết bị xử lý tín hiệu 100 tạo thành việc áp dụng khả thi của

thiết bị xử lý tín hiệu như được mô tả theo Fig.1. Thiết bị xử lý tín hiệu 100 chứa bộ lọc 101, bộ phận kết hợp 103, và bộ phát hiện hoạt động giọng nói 601.

Bộ lọc 101 tạo ra chức năng được mô tả liên quan tới bộ lọc 101 trên Fig.5. Bộ phát hiện hoạt động giọng nói 601 tạo ra chức năng được mô tả liên quan tới bộ phát hiện hoạt động giọng nói 601 trên Fig.6.

Theo một phương án thực hiện, bộ phận kết hợp 103 được đặt cấu hình để kết hợp tín hiệu audio kênh trái L với tín hiệu audio kênh trái được định trọng số L_E để thu tín hiệu audio kênh trái được kết hợp L_{EV} , để kết hợp tín hiệu audio kênh trung tâm C với tín hiệu audio kênh trung tâm được định trọng số C_E để thu tín hiệu audio kênh trung tâm được kết hợp C_{EV} , và để kết hợp tín hiệu audio kênh phải R với tín hiệu audio kênh phải được định trọng số R_E để thu tín hiệu audio kênh phải được kết hợp R_{EV} . Bộ phận kết hợp chứa bộ cộng 701, bộ cộng 703, bộ cộng 705, bộ phận định trọng số 707, bộ phận định trọng số 709, bộ phận định trọng số 711, và bộ phận định trọng số 713.

Theo một phương án thực hiện, bộ phận định trọng số 713 được đặt cấu hình để định trọng số bộ chỉ thị hoạt động giọng nói $V(m)$ bởi thông số khuếch đại lời nói được xác định từ trước G_S để thu bộ chỉ thị hoạt động giọng nói được định trọng số $V_G = G_S V(m)$, trong đó, m biểu thị chỉ số thời gian mẫu. Bộ phận kết hợp có thể chứa bộ phận định trọng số khác, không được thể hiện trên các hình vẽ, đang được đặt cấu hình để định trọng số tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R bởi thông số khuếch đại đầu vào được xác định từ trước G_{in} .

Bộ phận định trọng số 707 được đặt cấu hình để định trọng số tín hiệu audio kênh trái được định trọng số L_E với bộ chỉ thị hoạt động giọng nói được định trọng số $V_G = G_S V(m)$, và bộ cộng 701 được đặt cấu hình để cộng kết quả vào tín hiệu audio kênh trái L để thu tín hiệu audio kênh trái được kết hợp L_{EV} . Bộ phận định trọng số 709 được đặt cấu hình để định trọng số tín hiệu audio kênh trung tâm được định trọng số C_E với bộ chỉ thị hoạt động giọng nói được định trọng số $V_G = G_S V(m)$, và bộ cộng 703 được đặt cấu hình để cộng kết quả vào tín hiệu audio kênh trung tâm C để thu tín hiệu audio kênh trung tâm được kết hợp C_{EV} . Bộ phận định trọng số 711 được đặt cấu hình để định trọng số tín hiệu audio kênh phải được định trọng số R_E với bộ chỉ thị

hoạt động giọng nói được định trọng số $V_G = G_S V(m)$, và bộ cộng 705 được đặt cấu hình để cộng kết quả vào tín hiệu audio kênh phải R để thu tín hiệu audio kênh phải được kết hợp R_{EV} .

Theo một phương án thực hiện, bộ phận định trọng số 713 được đặt cấu hình để định trọng số tín hiệu audio kênh trái được định trọng số L_E , tín hiệu audio kênh trung tâm được định trọng số C_E , và tín hiệu audio kênh phải được định trọng số R_E bởi thông số khuếch đại lời nói được xác định từ trước G_S . Bộ phận kết hợp 103 có thể còn chứa bộ phận định trọng số, không được thể hiện trên hình vẽ, đang được đặt cấu hình để định trọng số tín hiệu audio kênh trái L, tín hiệu audio kênh trung tâm C, và tín hiệu audio kênh phải R bởi thông số khuếch đại đầu vào được xác định từ trước G_{in} .

Thông số khuếch đại lời nói được xác định từ trước G_S cũng có thể được áp dụng trong trường hợp mà bộ phát hiện hoạt động giọng nói 601 không được sử dụng. Để đơn giản, bộ phận định trọng số 713 được thể hiện như là bộ phận định trọng số 713 đơn trên hình vẽ. Theo áp dụng khả thi, bộ phận định trọng số 713 được sử dụng ba lần, cụ thể là giữa bộ phận định trọng số 709 và bộ cộng 703, giữa bộ phận định trọng số 707 và bộ cộng 701, và giữa bộ phận định trọng số 711 và bộ cộng 705. Trong trường hợp mà bộ phát hiện hoạt động giọng nói 601 không được sử dụng, $V = 1$ có thể được giả sử, và G_S có thể được sử dụng để thay đổi V .

Các kết quả của việc tăng cường giọng nói và việc phát hiện hoạt động giọng nói do đó có thể được kết hợp để thu ước lượng của tín hiệu audio giọng nói sạch. Việc tăng cường giọng nói và việc phát hiện hoạt động giọng nói có thể được thực hiện song song như đã được mô tả. Bộ chỉ thị hoạt động giọng nói V có thể được định trọng số hoặc có thể được nhân bởi bộ phận định trọng số 713 với thông số khuếch đại lời nói G_S , trong đó, $V_G = V G_S$ có thể được sử dụng để điều khiển việc gia tăng giọng nói. V_G có thể được kết hợp bởi các bộ phận định trọng số 707, 709, 711 theo cách nhân lên nhiều lần với các tín hiệu audio được định trọng số L_E , C_E , và R_E và các tín hiệu audio tạo thành có thể được cộng bởi các bộ cộng 701, 703, 705 vào các tín hiệu audio gốc L, C, và R để thu các tín hiệu audio được kết hợp L_{EV} , C_{EV} , và R_{EV} cuối cùng của thiết bị xử lý tín hiệu 100 theo các phương trình sau:

$$C_{EV}(m, k) = G_{in} \times C + G_S \times V(m) \times G(m, k) \times C(m, k) \quad (14)$$

$$L_{EV}(m, k) = G_{in} \times L + G_S \times V(m) \times G(m, k) \times L(m, k) \quad (15)$$

$$R_{EV}(m, k) = G_{in} \times R + G_S \times V(m) \times G(m, k) \times R(m, k) \quad (16)$$

trong đó, G_{in} là thông số khuếch đại đầu vào được áp dụng trên các tín hiệu audio gốc. Thông số này điều khiển hệ số khuếch đại của các thành phần không phải là giọng nói được tạo thành bởi tín hiệu audio đa kênh. Các kết hợp cụ thể của G_{in} và G_S , tức là $G_{in} = 1$ và $G_S = -1$, có thể được sử dụng để loại bỏ thành phần giọng nói từ tín hiệu audio đa kênh. Các thiết lập phù hợp để gia tăng thành phần giọng nói có thể là $G_{in} = 1$ trong khi G_S có thể ở trong khoảng giới hạn 1 và 4. Các tín hiệu audio được kết hợp L_{EV} , C_{EV} , và R_{EV} cuối cùng sau đó có thể được biến đổi ngược trở lại miền thời gian và có thể được sử dụng để tạo việc trộn giảm âm thanh nổi.

Kết quả là, có thể tạo ra giải pháp không đặt về mặt tính toán mà vẫn hiệu quả cho vấn đề tăng cường giọng nói hoặc hội thoại. Tất cả các thành phần có thể vận hành trong miền tần số DFT. Được so sánh với giải pháp đơn giản trong đó tín hiệu audio kênh trung tâm C, tức là trong tín hiệu audio bao quanh 5.1, được gia tăng và tất cả âm thanh nằm trong tín hiệu audio kênh trung tâm C được tăng cường, theo các phương án thực hiện của sáng chế chỉ các thành phần giọng nói trong tín hiệu audio kênh trung tâm C được gia tăng, tức là do việc phát hiện hoạt động giọng nói. Hơn nữa, các phương án thực hiện của sáng chế cũng xử lý một cách đồng thời giọng nói và các thành phần không phải là giọng nói, trong đó, chỉ các thành phần giọng nói được gia tăng tức là do giải pháp tăng cường giọng nói.

Thực tế là không chỉ tín hiệu audio kênh trung tâm C, mà các tín hiệu audio khác (tức là L và R) sẽ được xử lý sử dụng việc tăng cường giọng nói và việc phát hiện hoạt động giọng nói, đảm bảo rằng các tín hiệu audio cuối cùng chứa thành phần giọng nói rộng theo không gian với chất lượng cao. Đó không phải là trường hợp khi chỉ tín hiệu audio kênh trung tâm C được xử lý. Các phương án thực hiện của sáng chế là độc lập với codec, việc trộn, hoặc định dạng tín hiệu audio đa kênh cụ thể, như tín hiệu audio bao quanh 5.1, và có thể được mở rộng tới các cấu hình kênh khác.

Các phương án thực hiện của sáng chế, và cụ thể là của thiết bị xử lý tín hiệu, có thể bao gồm bộ xử lý đơn hoặc nhiều bộ xử lý được đặt cấu hình để áp dụng các chức năng khác nhau của thiết bị và các phương pháp được mô tả ở đây, tức là của bộ lọc

101, bộ phận kết hợp 103 và/hoặc các bộ phận khác hoặc các bước khác được mô tả ở đây dựa trên các hình vẽ từ Fig.1 đến Fig.7.

Phụ thuộc vào các yêu cầu áp dụng cụ thể của các phương pháp theo sáng chế, các phương pháp theo sáng chế có thể được áp dụng trong phần cứng hoặc trong phần mềm hoặc trong kết hợp bất kỳ của chúng.

Các áp dụng có thể được thực hiện bằng cách sử dụng vật ghi lưu trữ dạng số, cụ thể là đĩa mềm, các đĩa CD, DVD hoặc Blu-Ray, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ tác động nhanh có các tín hiệu điều khiển có thể đọc được bằng điện được lưu trên đó mà chúng phối hợp hoặc có khả năng phối hợp với hệ thống máy tính có thể lập trình được sao cho phương án thực hiện của ít nhất một phương pháp theo sáng chế sẽ được thực hiện.

Phương án thực hiện khác của sáng chế này là hoặc bao gồm, theo đó, sản phẩm chương trình máy tính với mã chương trình được lưu trên bộ mang đọc được bởi máy, mã chương trình là vận hành để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế nêu trên khi sản phẩm chương trình máy tính chạy trên máy tính.

Theo cách khác, các phương án thực hiện của các phương pháp theo sáng chế là hoặc bao gồm, theo đó, chương trình máy tính có mã chương trình để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế khi chương trình máy tính chạy trên máy tính, trên bộ xử lý hoặc dạng tương tự.

Do đó, phương án thực hiện khác của sáng chế là, hoặc bao gồm vật ghi lưu trữ dạng số đọc được bởi máy, chứa, được lưu trên đó là, chương trình máy tính vận hành để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế khi sản phẩm chương trình máy tính chạy trên máy tính, bộ xử lý hoặc dạng tương tự.

Do đó, phương án thực hiện khác của sáng chế là, hoặc bao gồm luồng dữ liệu hoặc chuỗi các tín hiệu hiển thị chương trình máy tính vận hành để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế khi sản phẩm chương trình máy tính chạy trên máy tính, trên bộ xử lý hoặc dạng tương tự.

Do đó, phương án thực hiện khác của sáng chế là, hoặc bao gồm máy tính, bộ xử lý hoặc thiết bị logic có thể lập trình được khác được làm thích ứng để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế.

Do đó, phương án thực hiện khác của sáng chế là, hoặc bao gồm máy tính, bộ xử lý hoặc thiết bị logic có khả năng lập trình được khác mà lưu trên đó là chương trình máy tính vận hành để thực hiện ít nhất một phương pháp trong các phương pháp theo sáng chế khi sản phẩm chương trình máy tính chạy trên máy tính, bộ xử lý hoặc thiết bị logic lập trình được khác, tức là mạng cổng lập trình được theo trường (FPGA - Field Programmable Gate Array) hoặc mạch tích hợp ứng dụng cụ thể (ASIC - Application Specific Integrated Circuit).

Mặc dù phần nêu trên được thể hiện một cách cụ thể và được mô tả liên quan tới các phương án thực hiện cụ thể của nó, nhưng người có hiểu biết trung bình trong lĩnh vực sẽ hiểu rằng có thể tạo ra các thay đổi khác ở dạng và các chi tiết khác mà không tách khỏi tinh thần và phạm vi bảo hộ của sáng chế. Do đó, cần hiểu rằng các thay đổi khác nhau có thể được tạo ra trong việc làm thích ứng các phương án thực hiện khác nhau mà không tách khỏi khái niệm rộng hơn được bộc lộ ở đây và được hiểu bởi các yêu cầu bảo hộ kèm theo.

YÊU CẦU BẢO HỘ

1. Thiết bị xử lý tín hiệu (100) để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh bao gồm tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), thiết bị xử lý tín hiệu (100) bao gồm:

bộ lọc (101) được đặt cấu hình để:

xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R),

thu hàm tăng ích ăng ten (G) dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, trong đó hàm tăng ích ăng ten là phụ thuộc vào tần số,

định trọng số tín hiệu audio kênh trái (L) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trái được định trọng số (L_E),

định trọng số tín hiệu audio kênh trung tâm (C) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trung tâm được định trọng số (C_E) và

định trọng số tín hiệu audio kênh phải (R) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh phải được định trọng số (R_E); và

bộ phận kết hợp (103) được đặt cấu hình để:

kết hợp tín hiệu audio kênh trái (L) với tín hiệu audio kênh trái được định trọng số (L_E) để thu tín hiệu audio kênh trái (L_{EV}),

kết hợp tín hiệu audio kênh trung tâm (C) với tín hiệu audio kênh trung tâm được định trọng số (C_E) để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và

kết hợp tín hiệu audio kênh phải (R) với tín hiệu audio kênh phải được định trọng số (R_E) để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).

2. Thiết bị xử lý tín hiệu (100) theo điểm 1, trong đó bộ lọc (101) còn được đặt cấu hình để xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng

của số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R).

3. Thiết bị xử lý tín hiệu (100) theo điểm 1, trong đó bộ lọc (101) được đặt cấu hình để xác định hàm tăng ích ăng ten (G) theo các phương trình sau:

$$G(m, k) = \frac{P_C(m, k)}{P_C(m, k) + P_S(m, k)}$$

$$P_C(m, k) = |C(m, k)|^2$$

$$P_S(m, k) = |L(m, k) - R(m, k)|^2$$

trong đó G biểu thị hàm tăng ích ăng ten, L biểu thị tín hiệu audio kênh trái, C biểu thị tín hiệu audio kênh trung tâm, R biểu thị tín hiệu audio kênh phải, P_C biểu thị công suất của tín hiệu audio kênh trung tâm (C) làm số đo thể hiện độ lớn của tín hiệu audio kênh trung tâm (C), P_S biểu thị công suất của khác biệt giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R), và tổng của P_C và P_S biểu thị số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngăn tần số (frequency bin).

4. Thiết bị xử lý tín hiệu (100) theo điểm 1, trong đó tín hiệu audio đa kênh còn bao gồm tín hiệu audio kênh bao quanh trái (LS) và tín hiệu audio kênh bao quanh phải (RS),

trong đó bộ lọc (101) còn được đặt cấu hình để:

xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số còn dựa trên tín hiệu audio kênh bao quanh trái (LS) và tín hiệu audio kênh bao quanh phải (RS), và

xác định số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm (C), của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R), và của số đo độ lớn của chênh lệch giữa tín hiệu audio kênh bao quanh trái (LS) và tín hiệu audio kênh bao quanh phải (RS).

5. Thiết bị xử lý tín hiệu (100) theo điểm 1, còn bao gồm:

bộ phát hiện hoạt động giọng nói (601) được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói (V) dựa trên tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), bộ chỉ thị hoạt động giọng nói (V) chỉ thị độ lớn của thành phần giọng nói nằm trong tín hiệu audio đa kênh qua thời gian,

trong đó bộ phận kết hợp (103) còn được đặt cấu hình để:

kết hợp tín hiệu audio kênh trái được định trọng số (L_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh trái được kết hợp (L_{EV}),

kết hợp tín hiệu audio kênh trung tâm được định trọng số (C_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và

kết hợp tín hiệu audio kênh phải được định trọng số (R_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).

6. Thiết bị xử lý tín hiệu (100) theo điểm 5, trong đó bộ phát hiện hoạt động giọng nói (601) còn được đặt cấu hình để:

xác định số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh dựa trên tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), và

thu bộ chỉ thị hoạt động giọng nói (V) dựa trên tỉ lệ giữa số đo của sự biến đổi phổ (F_c) của tín hiệu audio kênh trung tâm (C) và số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh.

7. Thiết bị xử lý tín hiệu (100) theo điểm 6, trong đó bộ phát hiện hoạt động giọng nói (601) còn được đặt cấu hình để xác định bộ chỉ thị hoạt động giọng nói (V) theo phương trình sau:

$$V = a \times \left(\frac{F_c}{F_c + F_s} - 0.5 \right)$$

trong đó V biểu thị bộ chỉ thị hoạt động giọng nói, F_c biểu thị số đo của sự biến đổi phổ của tín hiệu audio kênh trung tâm (C), F_s biểu thị số đo của sự biến đổi phổ của khác biệt giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R), và tổng của F_c và F_s biểu thị số đo thể hiện sự biến đổi phổ tổng thể của tín hiệu audio đa kênh, và a biểu thị thông số định cỡ được xác định từ trước.

8. Thiết bị xử lý tín hiệu (100) theo điểm 7, trong đó bộ phát hiện hoạt động giọng nói (601) còn được đặt cấu hình để xác định số đo của sự biến đổi phổ (F_c) của tín hiệu audio kênh trung tâm (C) làm thông lượng phổ và số đo của sự biến đổi phổ (F_s) của khác biệt giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R) làm thông lượng phổ theo các phương trình sau:

$$F_c(m) = \sum_k \left(|C(m, k)| - |C(m-1, k)| \right)^2$$

$$F_s(m) = \sum_k \left(|S(m, k)| - |S(m-1, k)| \right)^2$$

trong đó F_c biểu thị thông lượng phổ của tín hiệu audio kênh trung tâm (C), F_s biểu thị thông lượng phổ của khác biệt giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R), C biểu thị tín hiệu audio kênh trung tâm, S biểu thị khác biệt giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R), m biểu thị chỉ số thời gian mẫu, và k biểu thị chỉ số ngăn tần số (frequency bin).

9. Thiết bị xử lý tín hiệu (100) theo điểm 5, trong đó bộ phát hiện hoạt động giọng nói (601) còn được đặt cấu hình để lọc bộ chỉ thị hoạt động giọng nói (V) theo thời gian dựa trên hàm lọc thông thấp được xác định từ trước.

10. Thiết bị xử lý tín hiệu (100) theo điểm 5, trong đó bộ phận kết hợp (103) còn được đặt cấu hình để:

định trọng số tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R) bởi thông số khuếch đại đầu vào được xác định từ trước (G_{in}), và

định trọng số bộ chỉ thị hoạt động giọng nói (V) bởi thông số khuếch đại lời nói được xác định từ trước (G_s).

11. Thiết bị xử lý tín hiệu (100) theo điểm 5, trong đó bộ phận kết hợp (103) còn được đặt cấu hình để:

cộng tín hiệu audio kênh trái (L) vào tổ hợp của tín hiệu audio kênh trái được định trọng số (L_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh trái được kết hợp (L_{EV}),

cộng tín hiệu audio kênh trung tâm (C) vào tổ hợp của tín hiệu audio kênh trái được định trọng số (L_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và

cộng tín hiệu audio kênh phải (R) vào tổ hợp của tín hiệu audio kênh trái được định trọng số (L_E) với bộ chỉ thị hoạt động giọng nói (V) để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).

12. Thiết bị xử lý tín hiệu (100) theo điểm 1, còn bao gồm:

bộ trộn tăng (301) được đặt cấu hình để xác định tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R) dựa trên tín hiệu audio âm thanh nổi kênh trái đầu vào (L_{in}) và tín hiệu audio âm thanh nổi kênh phải đầu vào (R_{in}).

13. Thiết bị xử lý tín hiệu (100) theo điểm 12, còn bao gồm:

bộ trộn giảm (303) đang được đặt cấu hình để xác định tín hiệu audio âm thanh nổi kênh trái đầu ra (L_{out}) và tín hiệu audio âm thanh nổi kênh phải đầu ra (R_{out}) dựa trên tín hiệu audio kênh trái được kết hợp (L_{EV}), tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và tín hiệu audio kênh phải được kết hợp (R_{EV}).

14. Thiết bị xử lý tín hiệu (100) theo điểm 1, còn bao gồm:

bộ trộn giảm (303) được đặt cấu hình để xác định tín hiệu audio âm thanh nổi kênh trái đầu ra (L_{out}) và tín hiệu audio âm thanh nổi kênh phải đầu ra (R_{out}) dựa trên tín hiệu audio kênh trái được kết hợp (L_{EV}), tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và tín hiệu audio kênh phải được kết hợp (R_{EV}).

15. Thiết bị xử lý tín hiệu (100) theo điểm 1, trong đó số đo độ lớn bao gồm công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu.

16. Phương pháp xử lý tín hiệu (200) để tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh bao gồm tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), phương pháp xử lý tín hiệu (200) bao gồm các bước:

xác định (201) số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R),

thu (203) hàm tăng ích ăng ten (G) dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, trong đó hàm tăng ích ăng ten là phụ thuộc vào tần số,

định trọng số (205) tín hiệu audio kênh trái (L) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trái được định trọng số (L_E),

định trọng số (207) tín hiệu audio kênh trung tâm (C) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trung tâm được định trọng số (C_E),

định trọng số (209) tín hiệu audio kênh phải (R) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh phải được định trọng số (R_E),

kết hợp (211) tín hiệu audio kênh trái (L) với tín hiệu audio kênh trái được định trọng số (L_E) để thu tín hiệu audio kênh trái được kết hợp (L_{EV}),

kết hợp (213) tín hiệu audio kênh trung tâm (C) với tín hiệu audio kênh trung tâm (C_E) được định trọng số để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và

kết hợp (215) tín hiệu audio kênh phải (R) với tín hiệu audio kênh phải (R_E) được định trọng số để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).

17. Phương pháp xử lý tín hiệu (200) theo điểm 16, trong đó bước xác định (201) số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh làm tổng của số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo độ lớn của chênh lệch giữa tín hiệu audio kênh trái (L) và tín hiệu audio kênh phải (R).

18. Phương pháp xử lý tín hiệu (200) theo điểm 16, trong đó số đo độ lớn bao gồm công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu.

19. Vật ghi đọc được bằng máy tính chứa mã chương trình, mà khi được thực thi bởi bộ xử lý, sẽ làm cho hệ thống máy tính tăng cường thành phần giọng nói nằm trong tín hiệu audio đa kênh, tín hiệu audio đa kênh bao gồm tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R), bằng cách thực hiện các bước sau:

xác định (201) số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh qua tần số dựa trên tín hiệu audio kênh trái (L), tín hiệu audio kênh trung tâm (C), và tín hiệu audio kênh phải (R),

thu (203) hàm tăng ích ăng ten (G) dựa trên tỉ lệ giữa số đo độ lớn của tín hiệu audio kênh trung tâm (C) và số đo thể hiện độ lớn tổng thể của tín hiệu audio đa kênh, trong đó hàm tăng ích ăng ten là phụ thuộc vào tần số,

định trọng số (205) tín hiệu audio kênh trái (L) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trái được định trọng số (L_E),

định trọng số (207) tín hiệu audio kênh trung tâm (C) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh trung tâm được định trọng số (C_E),

định trọng số (209) tín hiệu audio kênh phải (R) bởi hàm tăng ích ăng ten (G) để thu tín hiệu audio kênh phải được định trọng số (R_E),

kết hợp (211) tín hiệu audio kênh trái (L) với tín hiệu audio kênh trái được định trọng số (L_E) để thu tín hiệu audio kênh trái được kết hợp (L_{EV}),

kết hợp (213) tín hiệu audio kênh trung tâm (C) với tín hiệu audio kênh trung tâm (C_E) được định trọng số để thu tín hiệu audio kênh trung tâm được kết hợp (C_{EV}), và

kết hợp (215) tín hiệu audio kênh phải (R) với tín hiệu audio kênh phải (R_E) được định trọng số để thu tín hiệu audio kênh phải được kết hợp (R_{EV}).

20. Vật ghi đọc được bằng máy tính theo điểm 19, trong đó số đo độ lớn bao gồm công suất, công suất logarit, độ lớn hoặc độ lớn logarit của tín hiệu.

1/7

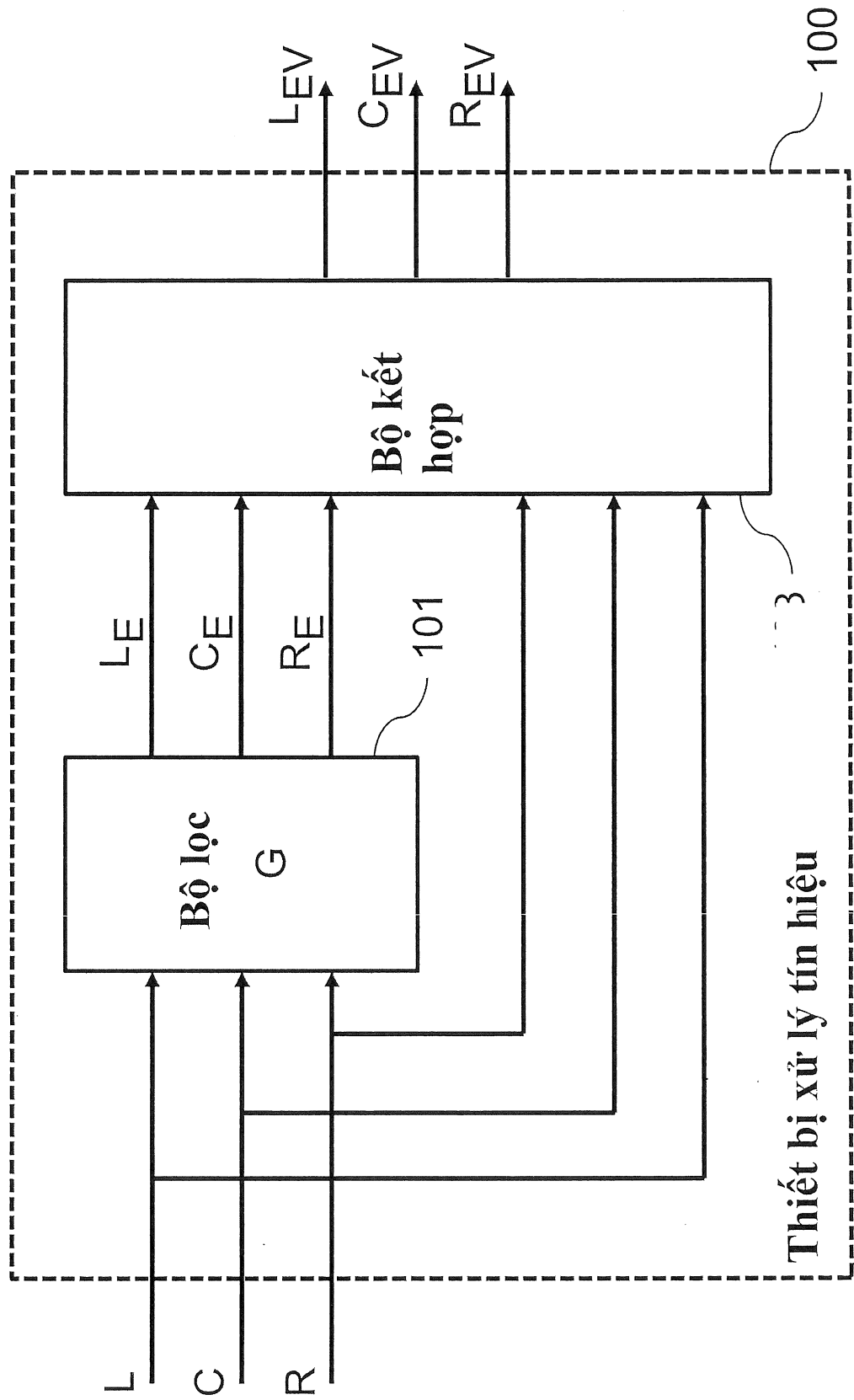


Fig. 1

2/7

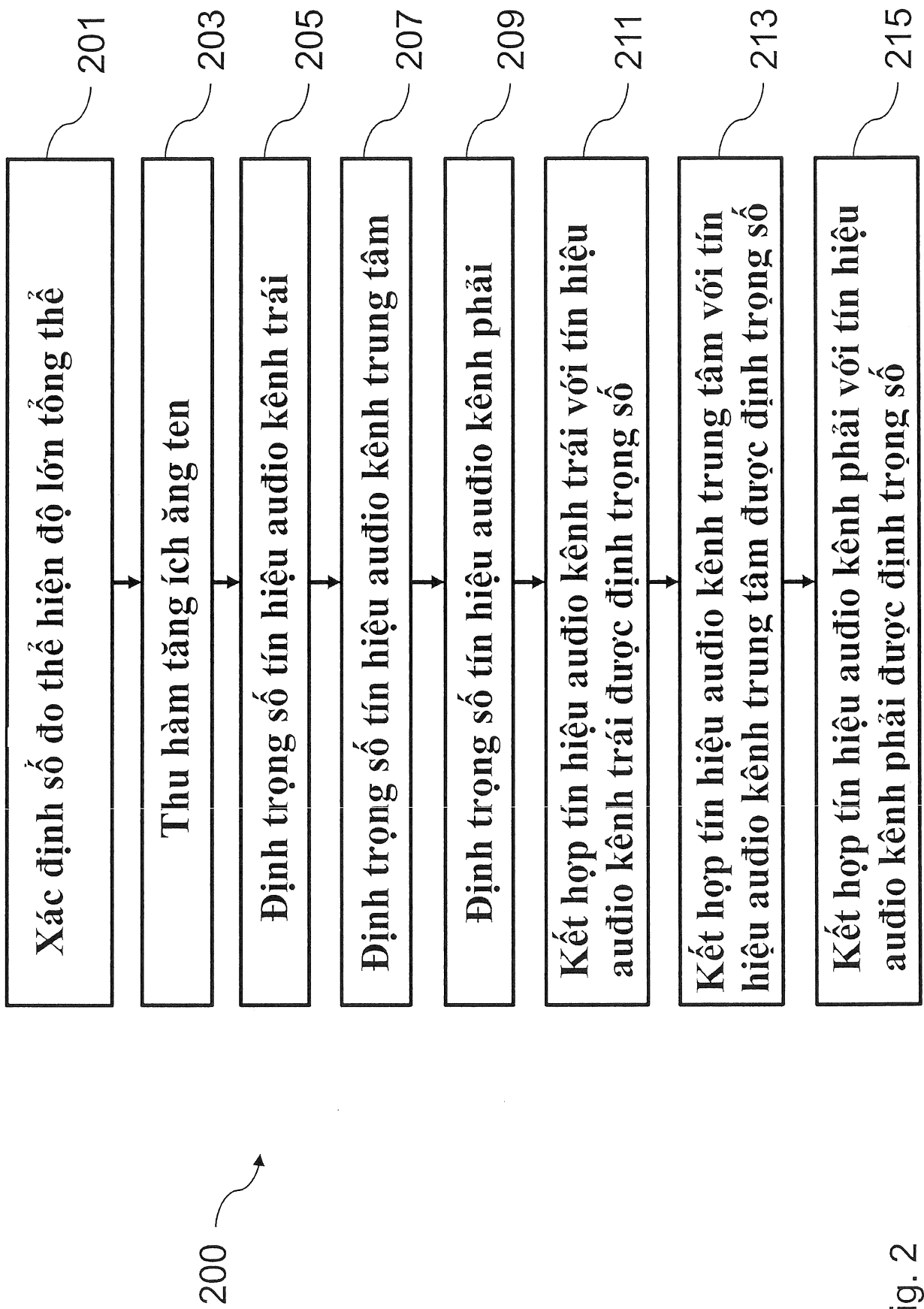
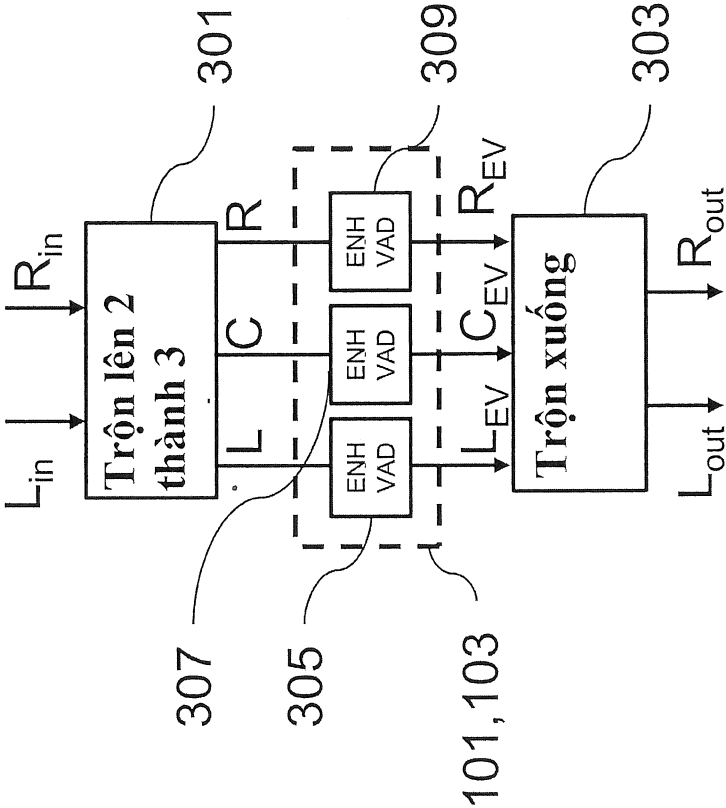


Fig. 2

3/7



100

Fig. 3

4/7

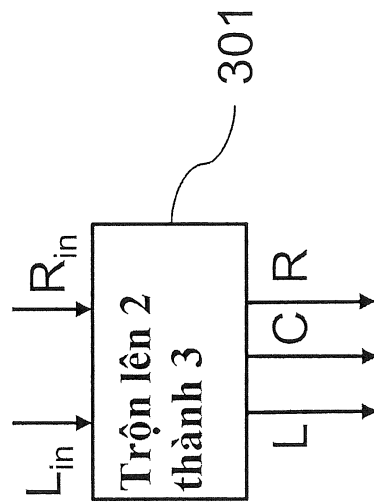


Fig. 4

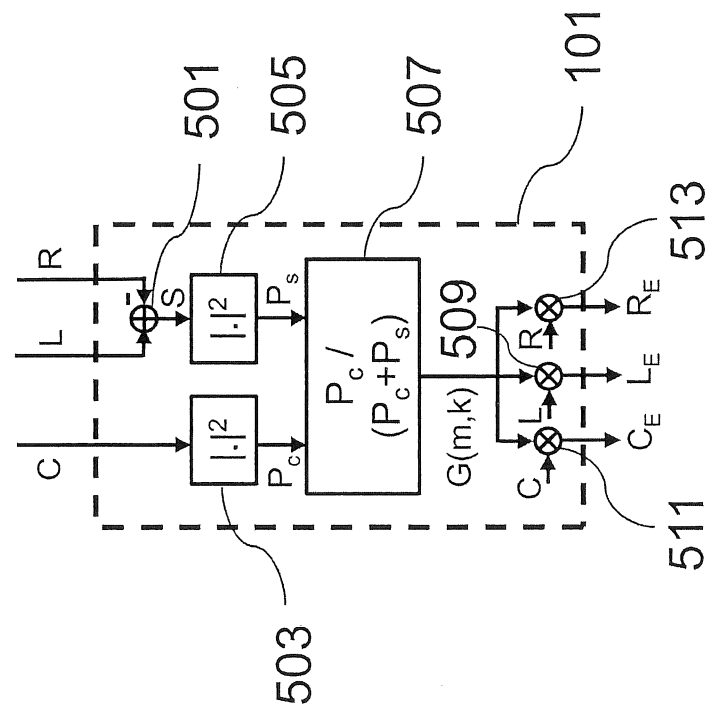


Fig. 5

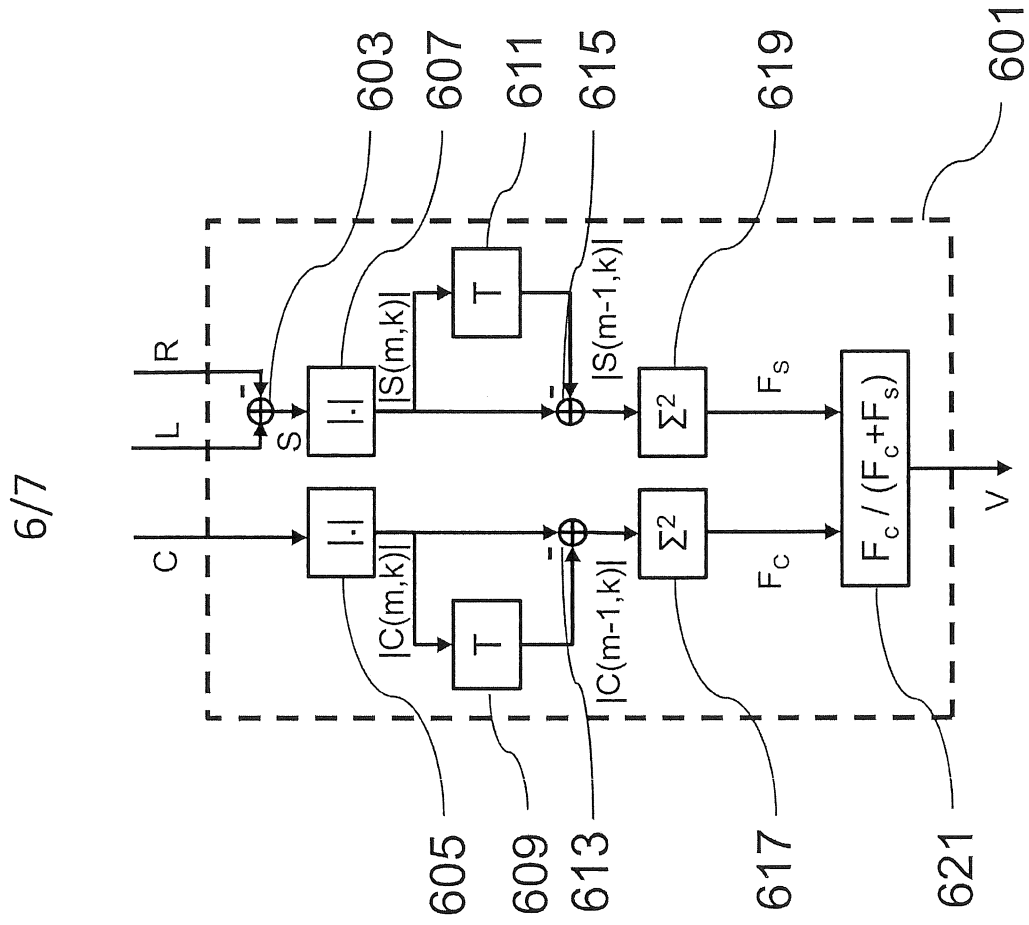


Fig. 6

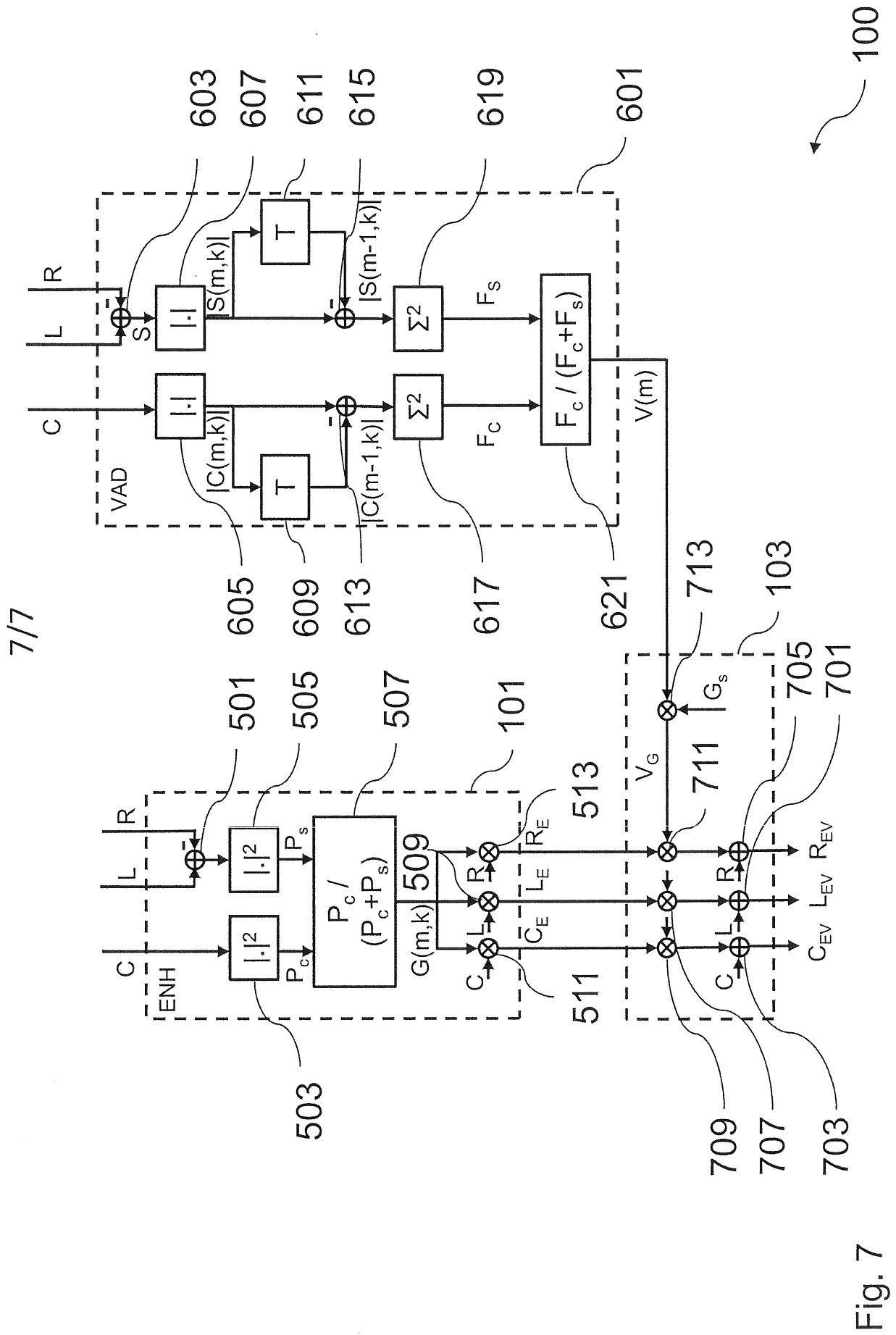


Fig. 7