



(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)



1-0047456

(51)^{2020.01} G10L 19/005; G10L 19/04

(13) B

(21) 1-2020-04405

(22) 13/11/2013

(62) 1-2015-02031

(86) PCT/KR2013/010310 13/11/2013

(87) WO 2014/077591 22/05/2014

(30) 61/725,694 13/11/2012 US

(45) 25/06/2025 447

(43) 25/02/2021 395A

(73) SAMSUNG ELECTRONICS CO., LTD. (KR)

129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-742, Republic of Korea

(72) CHOO, Ki-hyun (KR); POROV, Anton Victorovich (RU); OSIPOV, Konstantin Sergeevich (RU); LEE, Nam-suk (KR).

(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP MÃ HÓA TÍN HIỆU ÂM THANH

(21) 1-2020-04405

(57) Sáng chế đề cập đến phương pháp và thiết bị để xác định chế độ mã hóa để nâng cao chất lượng của tín hiệu âm thanh được khôi phục. Phương pháp xác định chế độ mã hóa bao gồm bước xác định một trong số các chế độ mã hóa bao gồm chế độ mã hóa thứ nhất và chế độ mã hóa thứ hai là chế độ mã hóa ban đầu tương ứng với các đặc tính của tín hiệu âm thanh, và nếu có lỗi ở quá trình xác định chế độ mã hóa ban đầu, thì tạo ra chế độ mã hóa được hiệu chỉnh bằng cách hiệu chỉnh chế độ mã hóa ban đầu thành chế độ mã hóa thứ ba.

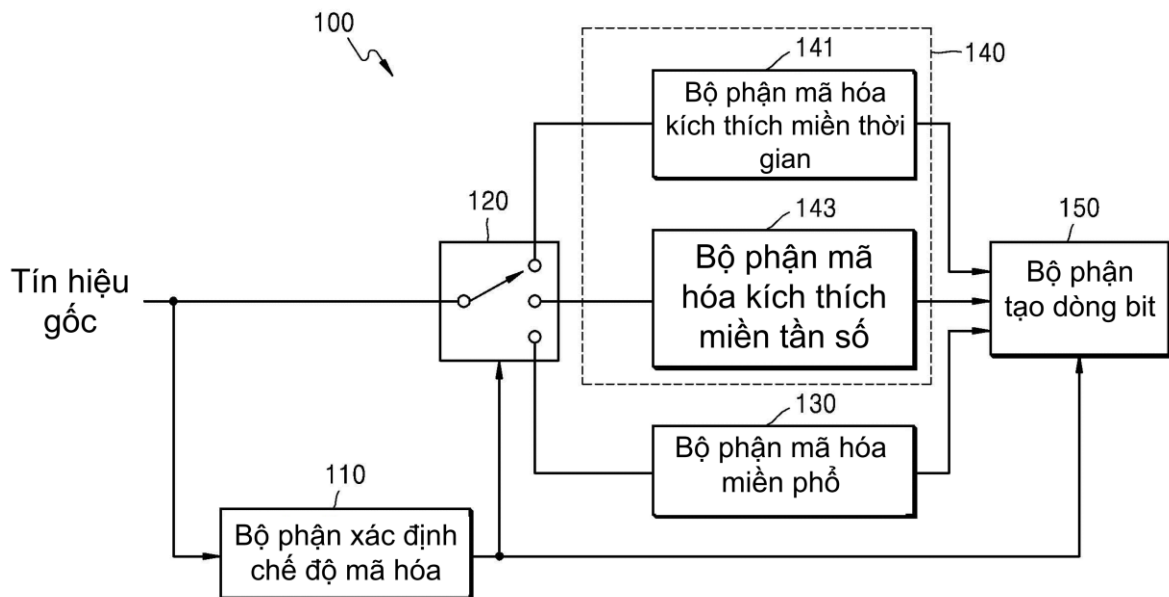


FIG. 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực mã hóa và giải mã âm thanh, và cụ thể hơn là, phương pháp và thiết bị để xác định chế độ mã hóa để nâng cao chất lượng của tín hiệu âm thanh được khôi phục, bằng cách xác định chế độ mã hóa phù hợp với các đặc tính của tín hiệu âm thanh và ngăn chặn chuyển chế độ mã hóa thường xuyên, và phương pháp và thiết bị để mã hóa tín hiệu âm thanh và phương pháp và thiết bị để giải mã tín hiệu âm thanh.

Tình trạng kỹ thuật của sáng chế

Hiện được biết đến một cách rộng rãi rằng có hiệu quả để mã hóa tín hiệu nhạc ở miền tần số và có hiệu quả để mã hóa tín hiệu tiếng nói ở miền thời gian. Do đó, đã đề xuất một số kỹ thuật dùng để xác định loại tín hiệu âm thanh, trong đó tín hiệu nhạc và tín hiệu tiếng nói được pha trộn, và xác định chế độ mã hóa tương ứng với loại đã xác định.

Tuy nhiên, do chuyển chế độ mã hóa tần số, nên không chỉ xuất hiện độ trễ, mà chất lượng âm thanh được giải mã cũng bị suy giảm. Ngoài ra, vì không có kỹ thuật để hiệu chỉnh chế độ mã hóa đã xác định một cách căn bản, nghĩa là, loại, nếu lỗi xuất hiện trong quá trình xác định chế độ mã hóa, chất lượng tín hiệu âm thanh được khôi phục bị suy giảm.

Bản chất kỹ thuật của sáng chế

Vấn đề kỹ thuật

Các khía cạnh của một hoặc nhiều phương án làm ví dụ đề xuất phương pháp và thiết bị dùng để xác định chế độ mã hóa để nâng cao chất lượng của tín hiệu âm thanh được khôi phục, bằng cách xác định chế độ mã hóa phù hợp với các đặc tính của tín hiệu âm thanh, phương pháp và thiết bị để mã hóa tín hiệu âm thanh và phương pháp và thiết bị để giải mã tín hiệu âm thanh.

Các khía cạnh của một hoặc nhiều phương án làm ví dụ đề xuất phương pháp và thiết bị dùng để xác định chế độ mã hóa phù hợp với các đặc tính của tín hiệu âm thanh và giảm các trễ do chuyển chế độ mã hóa thường xuyên, phương pháp và thiết bị để mã hóa tín hiệu âm thanh và phương pháp và thiết bị để giải mã tín hiệu âm thanh.

Giải pháp kỹ thuật

Theo một khía cạnh của một hoặc nhiều phương án làm ví dụ, sáng chế đề xuất phương pháp xác định chế độ mã hóa, phương pháp này bao gồm bước xác định một trong số các chế độ mã hóa bao gồm chế độ mã hóa thứ nhất và chế độ mã hóa thứ hai là chế độ mã hóa ban đầu tương ứng với các đặc tính của tín hiệu âm thanh, và nếu có lỗi ở bước xác định chế độ mã hóa ban đầu, thì tạo ra chế độ mã hóa được hiệu chỉnh bằng cách hiệu chỉnh chế độ mã hóa ban đầu thành chế độ mã hóa thứ ba.

Theo một khía cạnh của một hoặc nhiều phương án làm ví dụ, sáng chế đề xuất phương pháp mã hóa tín hiệu âm thanh, phương pháp này bao gồm bước xác định một trong số các chế độ mã hóa bao gồm chế độ mã hóa thứ nhất và chế độ mã hóa thứ hai là chế độ mã hóa ban đầu tương ứng với các đặc tính của tín hiệu âm thanh, nếu có lỗi ở bước xác định chế độ mã hóa ban đầu, thì tạo ra chế độ mã hóa được hiệu chỉnh bằng cách hiệu chỉnh chế độ mã hóa ban đầu thành chế độ mã hóa thứ ba, và thực hiện các quy trình mã hóa khác nhau đối với tín hiệu âm thanh dựa vào chế độ mã hóa ban đầu hoặc chế độ mã hóa được hiệu chỉnh.

Theo một khía cạnh của một hoặc nhiều phương án làm ví dụ, sáng chế đề xuất phương pháp giải mã tín hiệu âm thanh, phương pháp này gồm bước phân tách dòng bit bao gồm một trong số chế độ mã hóa ban đầu thu được bằng cách xác định một trong số các chế độ mã hóa bao gồm chế độ mã hóa thứ nhất và chế độ mã hóa thứ hai tương ứng với các đặc tính của tín hiệu âm thanh và chế độ mã hóa thứ ba được hiệu chỉnh từ chế độ mã hóa ban đầu nếu có lỗi ở bước xác định chế độ mã hóa ban đầu, và thực hiện các quy trình giải mã khác nhau đối với dòng bit dựa vào chế độ mã hóa ban đầu hoặc chế độ mã hóa thứ ba.

Hiệu quả đạt được của sáng chế

Theo các phương án làm ví dụ, bằng cách xác định chế độ mã hóa cuối cùng của khung hiện thời dựa vào sự hiệu chỉnh chế độ mã hóa ban đầu và các chế độ mã hóa của các khung tương ứng với chiều dài tồn lưu (hangover length), chế độ mã hóa thích ứng với các đặc tính của tín hiệu âm thanh có thể được chọn trong khi ngăn chặn việc chuyển chế độ mã hóa thường xuyên giữa các khung này.

Mô tả vắn tắt các hình vẽ

Fig.1 là sơ đồ khối minh họa cấu hình của thiết bị mã hóa âm thanh theo một phương án làm ví dụ;

Fig.2 là sơ đồ khối minh họa cấu hình của thiết bị mã hóa âm thanh theo một phương án làm ví dụ khác;

Fig.3 là sơ đồ khối minh họa cấu hình của bộ phận xác định chế độ mã hóa theo một phương án làm ví dụ;

Fig.4 là sơ đồ khối minh họa cấu hình của bộ phận xác định chế độ mã hóa ban đầu theo một phương án làm ví dụ;

Fig.5 là sơ đồ khối minh họa cấu hình của bộ phận trích xuất tham số đặc trưng theo một phương án làm ví dụ;

Fig.6 là sơ đồ minh họa phương pháp chuyển thích ứng giữa sự mã hóa miền dự đoán tuyến tính và miền phổ theo một phương án làm ví dụ;

Fig.7 là sơ đồ minh họa hoạt động của bộ phận hiệu chỉnh chế độ mã hóa theo một phương án làm ví dụ;

Fig.8 là sơ đồ khối minh họa cấu hình của thiết bị giải mã âm thanh theo một phương án làm ví dụ; và

Fig.9 là sơ đồ khối minh họa cấu hình của thiết bị giải mã âm thanh theo một phương án làm ví dụ khác.

Mô tả chi tiết sáng chế

Sáng chế hiện sẽ được thực hiện một cách chi tiết theo các phương án, các ví dụ về các phương án này được minh họa trên các hình vẽ kèm theo, trong đó các số chỉ dẫn giống nhau đề cập đến các bộ phận giống nhau trong toàn bộ sáng chế. Theo đó, các phương án của sáng chế có thể có các dạng khác nhau và không nên được hiểu là bị giới hạn ở các phần mô tả được nêu ở đây. Theo đó, các phương án chỉ được mô tả sau đây, dựa trên các hình vẽ, để giải thích các khía cạnh của sáng chế.

Các thuật ngữ chẳng hạn như “được nối” và “được liên kết” có thể được sử dụng để biểu thị trạng thái được nối hoặc liên kết trực tiếp, nhưng sẽ được hiểu là thành phần khác có thể được đặt ở giữa chúng.

Các thuật ngữ chẳng hạn như “thứ nhất” và “thứ hai” có thể sử dụng để mô tả các thành phần khác nhau, nhưng các thành phần này sẽ không bị giới hạn ở các thuật ngữ này. Các thuật ngữ này có thể được sử dụng chỉ để phân biệt một thành phần này với thành phần khác.

Các bộ phận được mô tả theo các phương án làm ví dụ được minh họa độc lập để biểu thị các chức năng đặc trưng khác nhau, và không có nghĩa là mỗi bộ phận được tạo ra từ một thành phần phần cứng hoặc phần mềm riêng biệt. Mỗi bộ phận này được minh họa để thuận tiện cho việc giải thích, và các bộ phận có thể tạo ra một bộ phận, và một bộ phận này có thể được chia thành nhiều bộ phận.

Fig.1 là sơ đồ khối minh họa cấu hình của thiết bị mã hóa âm thanh 100 theo một phương án làm ví dụ.

Thiết bị mã hóa âm thanh 100 được thể hiện trên Fig.1 có thể bao gồm bộ phận xác định chế độ mã hóa 110, bộ phận chuyển 120, bộ phận mã hóa miền phổ 130, bộ phận mã hóa miền dự đoán tuyến tính 140 và bộ phận tạo dòng bit 150. Bộ phận mã hóa miền dự đoán tuyến tính 140 có thể bao gồm bộ phận mã hóa kích thích miền thời gian 141 và bộ phận mã hóa kích thích miền tần số 143, trong đó bộ phận mã hóa miền dự đoán tuyến tính 140 này có thể được thể hiện dưới dạng ít nhất một trong số hai bộ phận mã hóa kích thích 141 và 143. Trừ khi cần phải được thể hiện dưới dạng phần cứng riêng biệt, các thành phần nêu trên có thể có thể được tích hợp vào ít nhất một môđun và có thể được thực hiện dưới dạng ít nhất một bộ xử lý (không được thể hiện trên hình vẽ). Ở đây, thuật ngữ tín hiệu âm thanh có thể đề cập đến tín hiệu nhạc, tín hiệu tiếng nói hoặc tín hiệu được pha trộn của các tín hiệu này.

Dựa vào Fig.1, bộ phận xác định chế độ mã hóa 110 có thể phân tích các đặc tính của tín hiệu âm thanh để xác định loại tín hiệu âm thanh, và xác định chế độ mã hóa tương ứng với kết quả phân loại. Bước xác định chế độ mã hóa có thể được thực hiện ở các đơn vị của các siêu khung, khung hoặc dải tần. Theo cách khác, bước xác định chế độ mã hóa có thể được thực hiện ở các đơn vị của các nhóm siêu khung, nhóm khung hoặc nhóm dải tần. Ở đây, các ví dụ về các chế độ mã hóa có thể bao gồm miền phổ và miền thời gian hoặc miền dự đoán tuyến tính, nhưng không bị giới hạn ở đó. Nếu hiệu suất và tốc độ xử lý của bộ xử lý là đủ và các độ trễ do chuyển chế độ mã hóa có thể được giải quyết, thì các chế độ mã hóa này có thể được chia nhỏ, và các sơ đồ mã hóa có thể còn được chia nhỏ tương ứng với chế

độ mã hóa. Theo một phương án làm ví dụ, bộ phận xác định chế độ mã hóa 110 có thể xác định chế độ mã hóa ban đầu của tín hiệu âm thanh là một trong số chế độ mã hóa miền phổ và chế độ mã hóa miền thời gian. Theo một phương án làm ví dụ khác, bộ phận xác định chế độ mã hóa 110 có thể xác định chế độ mã hóa ban đầu của tín hiệu âm thanh là một trong số chế độ mã hóa miền phổ, chế độ mã hóa kích thích miền thời gian và chế độ mã hóa kích thích miền tần số. Nếu chế độ mã hóa miền phổ được xác định là chế độ mã hóa ban đầu, thì bộ phận xác định chế độ mã hóa 110 có thể hiệu chỉnh chế độ mã hóa ban đầu thành một trong số chế độ mã hóa miền phổ và chế độ mã hóa kích thích miền tần số. Nếu chế độ mã hóa miền thời gian, nghĩa là, chế độ mã hóa kích thích miền thời gian được xác định là chế độ mã hóa ban đầu, thì bộ phận xác định chế độ mã hóa 110 có thể hiệu chỉnh chế độ mã hóa ban đầu thành một trong số chế độ mã hóa kích thích miền thời gian và chế độ mã hóa kích thích miền tần số. Nếu chế độ mã hóa kích thích miền thời gian được xác định là chế độ mã hóa ban đầu, thì bước xác định chế độ mã hóa cuối cùng có thể được thực hiện một cách chọn lọc. Nói cách khác, chế độ mã hóa ban đầu, nghĩa là, chế độ mã hóa kích thích miền thời gian có thể được duy trì. Bộ phận xác định chế độ mã hóa 110 có thể xác định các chế độ mã hóa của các khung tương ứng với chiều dài tồn lưu, và có thể xác định chế độ mã hóa cuối cùng đối với khung hiện thời. Theo một phương án làm ví dụ, nếu chế độ mã hóa ban đầu hoặc chế độ mã hóa được hiệu chỉnh của khung hiện thời giống với các chế độ mã hóa của các khung trước đó, ví dụ, 7 khung trước đó, thì chế độ mã hóa ban đầu tương ứng hoặc chế độ mã hóa được hiệu chỉnh có thể được xác định là chế độ mã hóa cuối cùng của khung hiện thời. Trong khi đó, nếu chế độ mã hóa ban đầu hoặc chế độ mã hóa được hiệu chỉnh của khung hiện thời không giống với các chế độ mã hóa của các khung trước đó, ví dụ, 7 khung trước đó, thì bộ phận xác định chế độ mã hóa 110 có thể xác định chế độ mã hóa của khung ngay trước khung hiện thời là chế độ mã hóa cuối cùng của khung hiện thời.

Như được mô tả ở trên, bằng cách xác định chế độ mã hóa cuối cùng của khung hiện thời dựa vào sự hiệu chỉnh chế độ mã hóa ban đầu và các chế độ mã hóa của các khung tương ứng với chiều dài tồn lưu, chế độ mã hóa thích ứng với các đặc tính của tín hiệu âm thanh có thể được chọn trong khi ngăn chặn chuyển chế độ mã hóa thường xuyên giữa các khung.

Nói chung, sự mã hóa miền thời gian, nghĩa là, sự mã hóa kích thích miền thời gian có thể có hiệu quả đối với tín hiệu tiếng nói, sự mã hóa miền phổ có thể có

hiệu quả đối với tín hiệu nhạc, và sự mã hóa kích thích miền tần số có thể có hiệu quả đối với tín hiệu thanh âm và/hoặc hòa âm.

Trương ứng với chế độ mã hóa được xác định bởi bộ phận xác định chế độ mã hóa 110, bộ phận chuyển 120 có thể tạo ra tín hiệu âm thanh cho bộ phận mã hóa miền phổ 130 hoặc bộ phận mã hóa miền dự đoán tuyến tính 140. Nếu bộ phận mã hóa miền dự đoán tuyến tính 140 được thể hiện dưới dạng bộ phận mã hóa kích thích miền thời gian 141, thì bộ phận chuyển 120 có thể bao gồm tổng cộng hai nhánh. Nếu bộ phận mã hóa miền dự đoán tuyến tính 140 được thể hiện dưới dạng bộ phận mã hóa kích thích miền thời gian 141 và bộ phận mã hóa kích thích miền tần số 143, thì bộ phận chuyển 120 có thể có tổng cộng 3 nhánh.

Bộ phận mã hóa miền phổ 130 có thể mã hóa tín hiệu âm thanh ở miền phổ. Miền phổ này có thể đề cập đến miền tần số hoặc miền biến đổi. Các ví dụ về các phương pháp mã hóa có thể áp dụng được cho bộ phận mã hóa miền phổ 130 có thể bao gồm sự mã hóa âm thanh nâng cao (advance audio coding, AAC), hoặc kết hợp biến đổi cosin rời rạc được thay đổi (modified discrete cosine transform, MDCT) và sự mã hóa xung giai thừa (factorial pulse coding, FPC), nhưng không bị giới hạn ở đó. Cụ thể, các kỹ thuật lượng tử hóa và kỹ thuật mã hóa entropy có thể được sử dụng thay cho FPC. Các kỹ thuật này có thể có hiệu quả để mã hóa tín hiệu nhạc trong bộ phận mã hóa miền phổ 130.

Bộ phận mã hóa miền dự đoán tuyến tính 140 có thể mã hóa tín hiệu âm thanh ở miền dự đoán tuyến tính. Miền dự đoán tuyến tính này có thể đề cập đến miền kích thích hoặc miền thời gian. Bộ phận mã hóa miền dự đoán tuyến tính 140 có thể được thể hiện dưới dạng bộ phận mã hóa kích thích miền thời gian 141 hoặc có thể được thể hiện để bao gồm bộ phận mã hóa kích thích miền thời gian 141 và bộ phận mã hóa kích thích miền tần số 143. Các ví dụ về các phương pháp mã hóa có thể áp dụng được cho bộ phận mã hóa kích thích miền thời gian 141 có thể bao gồm dự đoán tuyến tính kích thích mã (code excited linear prediction, CELP) hoặc CELP đại số (algebraic CELP, ACELP), nhưng không bị giới hạn ở đó. Các ví dụ về các phương pháp mã hóa có thể áp dụng được cho bộ phận mã hóa kích thích miền tần số 143 có thể bao gồm sự mã hóa tín hiệu tổng quát (general signal coding, GSC) hoặc kích thích được mã hóa biến đổi (transform coded excitation, TCX), nhưng không bị giới hạn ở đó. Có thể có hiệu quả để mã hóa tín hiệu tiếng nói trong bộ phận mã hóa kích thích miền thời gian 141, trong đó có thể có hiệu quả để mã

hóa tín hiệu thanh âm và/hoặc hòa âm trong bộ phận mã hóa kích thích miền tần số 143.

Bộ phận tạo dòng bit 150 có thể tạo ra dòng bit để bao gồm chế độ mã hóa được tạo ra bởi bộ phận xác định chế độ mã hóa 110, là kết quả mã hóa được tạo ra bởi bộ phận mã hóa miền phổ 130, và là kết quả mã hóa được tạo ra bởi bộ phận mã hóa miền dự đoán tuyến tính 140.

Fig.2 là sơ đồ khối minh họa cấu hình của thiết bị mã hóa âm thanh 200 theo một phương án làm ví dụ khác.

Thiết bị mã hóa âm thanh 200 được thể hiện trên Fig.2 có thể bao gồm môđun xử lý trước thông thường 205, bộ phận xác định chế độ mã hóa 210, bộ phận chuyển 220, bộ phận mã hóa miền phổ 230, bộ phận mã hóa miền dự đoán tuyến tính 240 và bộ phận tạo dòng bit 250. Ở đây, bộ phận mã hóa miền dự đoán tuyến tính 240 có thể bao gồm bộ phận mã hóa kích thích miền thời gian 241 và bộ phận mã hóa kích thích miền tần số 243 và bộ phận mã hóa miền dự đoán tuyến tính 240 có thể được thể hiện dưới dạng bộ phận mã hóa kích thích miền thời gian 241 hoặc bộ phận mã hóa kích thích miền tần số 243. So với thiết bị mã hóa âm thanh 100 được thể hiện trên Fig.1, thiết bị mã hóa âm thanh 200 có thể còn bao gồm môđun xử lý trước thông thường 205, và do đó các phần mô tả về các thành phần giống với các phần mô tả về các thành phần của thiết bị mã hóa âm thanh 100 sẽ được lược bỏ.

Dựa vào Fig.2, môđun xử lý trước thông thường 205 có thể thực hiện bước xử lý âm thanh nổi chung, xử lý xung quanh và/hoặc xử lý mở rộng băng thông. Bước xử lý âm thanh nổi chung, xử lý xung quanh và xử lý mở rộng băng thông có thể giống với xử lý âm thanh nổi chung, xử lý xung quanh và xử lý mở rộng băng thông được thể hiện bởi tiêu chuẩn cụ thể, ví dụ, tiêu chuẩn MPEG, nhưng không bị giới hạn ở đó. Đầu ra của môđun xử lý trước thông thường 205 có thể nằm ở kênh âm thanh đơn, kênh âm thanh nổi hoặc đa kênh. Theo số lượng kênh của tín hiệu được kết xuất bởi môđun xử lý trước thông thường 205, bộ phận chuyển 220 có thể bao gồm ít nhất một bộ chuyển. Ví dụ, nếu môđun xử lý trước thông thường 205 kết xuất tín hiệu của hai hoặc nhiều hơn hai kênh, nghĩa là, kênh âm thanh nổi hoặc đa kênh, các bộ chuyển tương ứng với các kênh tương ứng có thể được bố trí. Ví dụ, kênh thứ nhất của tín hiệu âm thanh nổi có thể là kênh tiếng nói, và kênh thứ hai của tín hiệu âm thanh nổi có thể là kênh nhạc. Trong trường hợp này, tín hiệu âm thanh

có thể được tạo ra đồng thời cho hai bộ chuyển. Thông tin bổ sung được tạo ra bởi môđun xử lý trước thông thường 205 có thể được tạo ra cho bộ phận tạo dòng bit 250 và có trong dòng bit. Thông tin bổ sung này có thể cần thiết để thực hiện xử lý âm thanh nổi chung, xử lý xung quanh và/hoặc xử lý mở rộng băng thông ở đầu giải mã và có thể bao gồm các tham số không gian, thông tin đường bao, thông tin năng lượng, v.v. Tuy nhiên, có thể có thông tin bổ sung khác nhau dựa vào các kỹ thuật xử lý được áp dụng cho các thông tin này.

Theo một phương án làm ví dụ, ở môđun xử lý trước thông thường 205, việc xử lý mở rộng băng thông có thể được thực hiện khác nhau dựa vào các miền mã hóa. Tín hiệu âm thanh ở dải lõi có thể được xử lý bằng cách sử dụng chế độ mã hóa kích thích miền thời gian hoặc chế độ mã hóa kích thích miền tần số, trong đó tín hiệu âm thanh ở dải mở rộng băng thông có thể được xử lý ở miền thời gian. Việc xử lý mở rộng băng thông ở miền thời gian có thể bao gồm các chế độ bao gồm chế độ được hữu thanh hoặc chế độ vô thanh. Theo cách khác, tín hiệu âm thanh ở dải lõi có thể được xử lý bằng cách sử dụng chế độ mã hóa miền phổ, trong đó tín hiệu âm thanh ở dải mở rộng băng thông có thể được xử lý ở miền tần số. Việc xử lý mở rộng băng thông ở miền tần số có thể bao gồm các chế độ bao gồm chế độ chuyển tiếp, chế độ thông thường hoặc chế độ điều hòa. Để thực hiện việc xử lý mở rộng băng thông ở các miền khác nhau, chế độ mã hóa được xác định bởi bộ phận xác định chế độ mã hóa 110 có thể được tạo ra cho môđun xử lý trước thông thường 205 làm thông tin báo hiệu. Theo một phương án làm ví dụ, phần cuối cùng của dải lõi và phần bắt đầu của dải mở rộng băng thông có thể chồng lên nhau ở mức độ nào đó. Vị trí và kích thước của các phần được chồng đó có thể được thiết lập trước.

Fig.3 là sơ đồ khối minh họa cấu hình của bộ phận xác định chế độ mã hóa 300 theo một phương án làm ví dụ.

Bộ phận xác định chế độ mã hóa 300 được thể hiện trên Fig.3 có thể bao gồm bộ phận xác định chế độ mã hóa ban đầu 310 và bộ phận hiệu chỉnh chế độ mã hóa 330.

Dựa vào Fig.3, bộ phận xác định chế độ mã hóa ban đầu 310 có thể xác định xem tín hiệu âm thanh là tín hiệu nhạc hay tín hiệu tiếng nói bằng cách sử dụng các tham số đặc trưng được trích xuất từ tín hiệu âm thanh. Nếu tín hiệu âm thanh này được xác định là tín hiệu tiếng nói, thì sự mã hóa miền dự đoán tuyến tính có thể

phù hợp. Trong khi đó, nếu tín hiệu âm thanh được xác định là tín hiệu nhạc, thì sự mã hóa miền phổ có thể phù hợp. Bộ phận xác định chế độ mã hóa ban đầu 310 có thể xác định loại tín hiệu âm thanh biểu thị xem sự mã hóa miền phổ, mã hóa kích thích miền thời gian hoặc mã hóa kích thích miền tần số có phù hợp đối với tín hiệu âm thanh hay không bằng cách sử dụng các tham số đặc trưng được trích xuất từ tín hiệu âm thanh. Chế độ mã hóa tương ứng có thể được xác định dựa vào loại tín hiệu âm thanh. Nếu bộ phận chuyển (120 trên Fig.1) có hai nhánh, thì chế độ mã hóa có thể được biểu thị bằng 1-bit. Nếu bộ phận chuyển (120 trên Fig.1) có ba nhánh, thì chế độ mã hóa có thể được biểu thị bằng 2-bit. Bộ phận xác định chế độ mã hóa ban đầu 310 có thể xác định xem tín hiệu âm thanh là tín hiệu nhạc hay tín hiệu tiếng nói bằng cách sử dụng kỹ thuật bất kỳ trong số các kỹ thuật khác nhau đã biết trong lĩnh vực kỹ thuật. Các ví dụ về các kỹ thuật này có thể bao gồm phân loại FD/LPD hoặc phân loại ACELP/TCX được bộc lộ ở một phần bộ mã hóa theo tiêu chuẩn USAC và phân loại ACELP/TCX được sử dụng theo tiêu chuẩn AMR, nhưng không bị giới hạn ở đó. Nói cách khác, chế độ mã hóa ban đầu có thể được xác định bằng cách sử dụng phương pháp bất kỳ trong số các phương pháp khác nhau khác với phương pháp theo các phương án được mô tả ở đây.

Bộ phận hiệu chỉnh chế độ mã hóa 330 có thể xác định chế độ mã hóa được hiệu chỉnh bằng cách hiệu chỉnh chế độ mã hóa ban đầu được xác định bởi bộ phận xác định chế độ mã hóa ban đầu 310 bằng cách sử dụng các tham số hiệu chỉnh. Theo một phương án làm ví dụ, nếu chế độ mã hóa miền phổ được xác định là chế độ mã hóa ban đầu, thì chế độ mã hóa ban đầu này có thể được hiệu chỉnh thành chế độ mã hóa kích thích miền tần số dựa vào các tham số hiệu chỉnh. Nếu chế độ mã hóa miền thời gian được xác định là chế độ mã hóa ban đầu, thì chế độ mã hóa ban đầu này có thể được hiệu chỉnh thành chế độ mã hóa kích thích miền tần số dựa vào các tham số hiệu chỉnh. Nói cách khác, xác định được xem có lỗi trong quá trình xác định chế độ mã hóa ban đầu hay không bằng cách sử dụng các tham số hiệu chỉnh. Nếu xác định được rằng không có lỗi trong quá trình xác định chế độ mã hóa ban đầu, thì chế độ mã hóa ban đầu này có thể được duy trì. Ngược lại, nếu xác định được rằng có lỗi trong quá trình xác định chế độ mã hóa ban đầu, thì chế độ mã hóa ban đầu này có thể được hiệu chỉnh. Sự hiệu chỉnh chế độ mã hóa ban đầu có thể thu được từ chế độ mã hóa miền phổ thành chế độ mã hóa kích thích miền tần số và từ chế độ mã hóa kích thích miền thời gian thành chế độ mã hóa kích thích miền tần số.

Trong khi đó, chế độ mã hóa ban đầu hoặc chế độ mã hóa được hiệu chỉnh có thể là chế độ mã hóa tạm thời đối với khung hiện thời, trong đó chế độ mã hóa tạm thời đối với khung hiện thời này có thể được so sánh với các chế độ mã hóa đối với các khung trước đó nằm trong chiều dài tồn lưu được thiết lập trước và chế độ mã hóa cuối cùng đối với khung hiện thời có thể được xác định.

Fig.4 là sơ đồ khối minh họa cấu hình của bộ phận xác định chế độ mã hóa ban đầu 400 theo một phương án làm ví dụ.

Bộ phận xác định chế độ mã hóa ban đầu 400 được thể hiện trên Fig.4 có thể bao gồm bộ phận trích xuất tham số đặc trưng 410 và bộ phận xác định 430.

Dựa vào Fig.4, bộ phận trích xuất tham số đặc trưng 410 có thể trích xuất các tham số đặc trưng cần thiết để xác định chế độ mã hóa từ tín hiệu âm thanh. Các ví dụ về các tham số đặc trưng được trích xuất bao gồm ít nhất một hoặc hai trong số các tham số cao độ, tham số giọng nói, tham số tương quan và lỗi dự đoán tuyến tính, nhưng không bị giới hạn ở đó. Các phần mô tả chi tiết về các tham số riêng biệt sẽ được đưa ra sau đây.

Trước tiên, tham số đặc trưng thứ nhất F_1 liên quan đến tham số cao độ, trong đó trạng thái độ cao có thể được xác định bằng cách sử dụng các giá trị cao độ N được phát hiện ở khung hiện thời và ít nhất một khung trước đó. Để ngăn chặn ảnh hưởng khỏi độ lệch ngẫu nhiên hoặc giá trị cao độ sai, thì có thể loại bỏ các giá trị cao độ M khác đáng kể với giá trị trung bình của các giá trị cao độ N . Ở đây, N và M có thể là các giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước. Ngoài ra, N có thể được thiết lập trước, và chênh lệch giữa giá trị cao độ cần được loại bỏ và giá trị trung bình của các giá trị cao độ N có thể được xác định thông qua thực nghiệm hoặc mô phỏng trước. Tham số đặc trưng thứ nhất F_1 có thể được biểu diễn như được thể hiện ở phương trình sau đây bằng cách sử dụng giá trị trung bình m_p và phương sai σ_p đối với các giá trị cao độ $(N-M)$.

【Phương trình 1】

$$F_1 = \frac{\sigma_p}{m_p}$$

Tham số đặc trưng thứ hai F_2 còn liên quan đến tham số cao độ và có thể biểu thị độ tin cậy của giá trị cao độ được phát hiện ở khung hiện thời. Tham số đặc trưng thứ hai F_2 này có thể được biểu diễn như được thể hiện ở phương trình 2 sau đây bằng cách sử dụng các phương sai σ_{SF_1} và σ_{SF_2} có giá trị cao độ lần lượt được phát hiện ở hai khung phụ SF_1 và SF_2 của khung hiện thời.

【Phương trình 2】

$$F_2 = \frac{cov(SF_1, SF_2)}{\sigma_{SF_1} \sigma_{SF_2}}$$

Ở đây, $cov(SF_1, SF_2)$ biểu thị hiệp phương sai giữa các khung phụ SF_1 và SF_2 . Nói cách khác, tham số đặc trưng thứ hai F_2 biểu thị tương quan giữa hai khung phụ dưới dạng khoảng cách độ cao. Theo một phương án làm ví dụ, khung hiện thời có thể bao gồm hai hoặc nhiều hơn hai khung phụ, và phương trình 2 có thể được thay đổi dựa vào số lượng khung phụ.

Tham số đặc trưng thứ ba F_3 có thể được biểu diễn như được thể hiện ở phương trình 3 sau đây dựa vào tham số giọng nói Voicing và tham số tương quan Corr.

【Phương trình 3】

$$F_3 = \sqrt{Q \frac{cVoicing - CorrC^2}{N}}$$

Ở đây, tham số giọng nói Voicing liên quan đến các đặc trưng thanh âm của âm thanh và có thể thu được bằng phương pháp bất kỳ trong số các phương pháp khác nhau được biết trong lĩnh vực kỹ thuật, trong đó tham số tương quan Corr có thể thu được bằng cách tổng hợp các tương quan giữa các khung đối với mỗi dải tần.

Tham số đặc trưng thứ tư F_4 liên quan đến lỗi dự đoán tuyến tính $ELPC$ và có thể được biểu diễn như được thể hiện ở phương trình 4 sau đây.

【Phương trình 4】

$$F_4 = \frac{\sqrt{(E_{LPCi} - M(E_{LPC}))^2}}{N}$$

Ở đây, $M(E_{LPC})$ biểu thị trung bình của các lỗi dự đoán tuyến tính N .

Bộ phận xác định 430 có thể xác định loại tín hiệu âm thanh bằng cách sử dụng ít nhất một tham số đặc trưng được tạo ra bởi bộ phận trích xuất tham số đặc trưng 410 và có thể xác định chế độ mã hóa ban đầu dựa vào loại đã xác định này. Bộ phận xác định 430 có thể sử dụng cơ chế quyết định mềm, trong đó ít nhất một hỗn hợp có thể được tạo ra trên mỗi tham số đặc trưng. Theo một phương án làm ví dụ, loại tín hiệu âm thanh có thể được xác định bằng cách sử dụng mô hình hỗn hợp Gauss (Gaussian mixture model, GMM) dựa vào các xác suất hỗn hợp. Xác suất $f(x)$ liên quan đến một hỗn hợp có thể được tính toán theo phương trình 5 sau đây.

【Phương trình 5】

$$f(x) = \frac{1}{\sqrt{(2\pi)^N \det(C^{-1})}} e^{-0.5(x-m)^T C^{-1}(x-m)}$$

$$x = (x_1, \dots, x_N)$$

$$m = (Cx_1C, \dots, Cx_NC)$$

Ở đây, x biểu thị vectơ đầu vào của tham số đặc trưng, m biểu thị hỗn hợp, và c biểu thị ma trận hiệp phương sai.

Bộ phận xác định 430 có thể tính toán xác suất nhạc P_m và xác suất tiếng nói P_s bằng cách sử dụng phương trình 6 sau đây.

【Phương trình 6】

$$P_m = \sum_{i \in M} p_i, P_s = \sum_{i \in S} p_i$$

Ở đây, xác suất nhạc P_m có thể được tính toán bằng cách thêm các xác suất P_i của các hỗn hợp M có liên quan đến các tham số đặc trưng vượt trội để xác định nhạc, trong đó xác suất tiếng nói P_s có thể được tính toán bằng cách thêm các xác suất P_i của các hỗn hợp S có liên quan đến các tham số đặc trưng vượt trội để xác định tiếng nói.

Trong khi đó, để cải thiện độ chính xác, xác suất nhạc P_m và xác suất tiếng nói P_s có thể được tính toán theo phương trình 7 sau đây.

【Phương trình 7】

$$P_m = \sum_{i \in M} Q_i P_i (1 - p_i^{err}) + \sum_{i \in S} Q_i P_i (p_i^{err})$$

$$P_s = \sum_{i \in S} Q_i P_i (1 - p_i^{err}) + \sum_{i \in M} Q_i P_i (p_i^{err})$$

Ở đây, p_i^{err} biểu thị xác suất sai số của mỗi hỗn hợp. Xác suất sai số này có thể thu được bằng cách phân loại dữ huấn luyện bao gồm các tín hiệu tiếng nói rõ ràng và các tín hiệu nhạc rõ ràng bằng cách sử dụng mỗi trong số các hỗn hợp và đếm số lượng phân loại sai.

Tiếp theo, xác suất P^M mà tất cả các khung chỉ bao gồm các tín hiệu nhạc và xác suất tiếng nói P^S mà tất cả các khung chỉ bao gồm các tín hiệu tiếng nói đối với các khung này cũng như chiều dài tồn lưu không đổi có thể được tính toán theo phương trình 8 sau đây. Chiều dài tồn lưu này có thể được thiết lập thành 8, nhưng không bị giới hạn ở đó. Tám khung có thể bao gồm khung hiện thời và 7 khung trước đó.

【Phương trình 8】

$$P^M = \frac{\bigcirc_{i=0}^{-7} P_m^{(i)}}{\bigcirc_{i=0}^{-7} P_m^{(i)} + \bigcirc_{i=0}^{-7} P_s^{(i)}}$$

$$P^S = \frac{\bigcirc_{i=0}^{-7} P_s^{(i)}}{\bigcirc_{i=0}^{-7} P_m^{(i)} + \bigcirc_{i=0}^{-7} P_s^{(i)}}$$

Tiếp theo, các tập trạng thái $\{D_i^M\}$ và $\{D_i^S\}$ có thể được tính toán bằng cách sử dụng xác suất nhạc Pm hoặc xác suất tiếng nói Ps thu được bằng cách sử dụng phương trình 5 hoặc phương trình 6. Các phần mô tả chi tiết của chúng được đưa ra sau đây dựa trên Fig.6. Ở đây, có thể thiết lập sao cho mỗi trạng thái có giá trị là 1 đối với nhạc và có giá trị là 0 đối với tiếng nói.

Dựa vào Fig.6, ở hoạt động 610 và hoạt động 620, tổng các trạng thái nhạc M và tổng các trạng thái giọng nói S có thể thu được từ các tập trạng thái $\{D_i^M\}$ và $\{D_i^S\}$ được tính toán bằng cách sử dụng xác suất nhạc Pm và xác suất tiếng nói Ps. Nói cách khác, tổng các trạng thái nhạc M và tổng các trạng thái tiếng nói S có thể được biểu diễn như được thể hiện ở phương trình 9 sau đây.

【Phương trình 9】

$$M = \bigcirc_i D_i^M$$

$$S = \bigcirc_i D_i^S$$

Ở hoạt động 630, tổng các trạng thái nhạc M được so sánh với giá trị ngưỡng được chỉ định Tm. Nếu tổng các trạng thái nhạc M lớn hơn giá trị ngưỡng Tm, thì chế độ mã hóa của khung hiện thời được chuyển sang chế độ nhạc, nghĩa là, chế độ mã hóa miền phổ. Nếu tổng các trạng thái nhạc M nhỏ hơn hoặc bằng giá trị ngưỡng Tm, thì chế độ mã hóa của khung hiện thời không bị thay đổi.

Ở hoạt động 640, tổng các trạng thái tiếng nói S được so sánh với giá trị ngưỡng được chỉ định Ts. Nếu tổng các trạng thái tiếng nói S lớn hơn giá trị ngưỡng Ts, thì chế độ mã hóa của khung hiện thời được chuyển sang chế độ tiếng nói, nghĩa là, chế độ mã hóa miền dự đoán tuyến tính. Nếu tổng các trạng thái tiếng nói S nhỏ hơn hoặc bằng giá trị ngưỡng Ts, thì chế độ mã hóa của khung hiện thời không bị thay đổi.

Giá trị ngưỡng Tm và giá trị ngưỡng Ts có thể được thiết lập thành các giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Fig.5 là sơ đồ khối minh họa cấu hình của bộ phận trích xuất tham số đặc trưng 500 theo một phương án làm ví dụ.

Bộ phận xác định chế độ mã hóa ban đầu 500 được thể hiện trên Fig.5 có thể bao gồm bộ phận biến đổi 510, bộ phận trích xuất tham số phổ 520, bộ phận trích xuất tham số thời gian 530 và bộ phận xác định 540.

Trên Fig.5, bộ phận biến đổi 510 có thể biến đổi tín hiệu âm thanh gốc từ miền thời gian sang miền tần số. Ở đây, bộ phận biến đổi 510 này có thể áp dụng kỹ thuật biến đổi bất kỳ trong số các kỹ thuật biến đổi khác nhau để thể hiện tín hiệu âm thanh từ miền thời gian sang miền phổ. Các ví dụ về các kỹ thuật này có thể bao gồm biến đổi Fourier nhanh (fast Fourier transform, FFT), biến đổi cosin rời rạc (discrete cosine transform, DCT) hoặc biến đổi cosin rời rạc được sửa đổi (MDCT), nhưng không bị giới hạn các kỹ thuật này.

Bộ phận trích xuất tham số phổ 520 có thể trích xuất ít nhất một tham số phổ từ tín hiệu âm thanh miền tần số được tạo ra bởi bộ phận biến đổi 510. Các tham số phổ có thể được phân hạng thành các tham số đặc trưng ngắn hạn và các tham số đặc trưng dài hạn. Các tham số đặc trưng ngắn hạn có thể thu được từ khung hiện thời, trong đó các tham số đặc trưng dài hạn có thể thu được từ các khung bao gồm khung hiện thời và ít nhất một khung trước đó.

Bộ phận trích xuất tham số thời gian 530 có thể trích xuất ít nhất một tham số thời gian từ tín hiệu âm thanh miền thời gian. Các tham số thời gian này có thể còn được phân hạng thành các tham số đặc trưng ngắn hạn và các tham số đặc trưng dài hạn. Các tham số đặc trưng ngắn hạn có thể thu được từ khung hiện thời, trong

đó các tham số đặc trưng dài hạn có thể thu được từ các khung bao gồm khung hiện thời và ít nhất một khung trước đó.

Bộ phận xác định (430 trên Fig.4) có thể xác định loại tín hiệu âm thanh bằng cách sử dụng các tham số phổ được tạo ra bởi bộ phận trích xuất tham số phổ 520 và các tham số thời gian được tạo ra bởi bộ phận trích xuất tham số thời gian 530 và có thể xác định chế độ mã hóa ban đầu dựa vào nhóm đã xác định này. Bộ phận xác định (430 trên Fig.4) có thể sử dụng cơ chế quyết định mềm.

Fig.7 là sơ đồ minh họa hoạt động của bộ phận hiệu chỉnh chế độ mã hóa 310 theo một phương án làm ví dụ.

Dựa vào Fig.7, ở hoạt động 700, chế độ mã hóa ban đầu được xác định bởi bộ phận xác định chế độ mã hóa ban đầu 310 được thu nhận và nó có thể được xác định xem chế độ mã hóa này là chế độ miền thời gian, nghĩa là, chế độ kích thích miền thời gian hay chế độ miền phổ.

Ở hoạt động 701, nếu xác định được ở hoạt động 700 rằng chế độ mã hóa ban đầu là chế độ miền phổ ($state_{TS} = 1$), thì chỉ số $state_{TTSS}$ biểu thị xem sự mã hóa kích thích miền tần số thích hợp hơn có thể được kiểm tra hay không. Chỉ số $state_{TTSS}$ này biểu thị xem sự mã hóa kích thích miền tần số (ví dụ, GSC) thích hợp hơn có thể thu được bằng cách sử dụng các âm điệu có các dải tần số khác nhau hay không. Các phần mô tả chi tiết về hoạt động này sẽ được đưa ra sau đây.

Âm điệu của tín hiệu dải tần thấp có thể thu được là tỷ số giữa tổng các hệ số phổ có các giá trị nhỏ bao gồm giá trị nhỏ nhất và hệ số phổ có giá trị lớn nhất so với một dải tần đã cho. Nếu các dải tần đã cho này là 0~1 kHz, 1~2 kHz và 2~4 kHz, thì các âm điệu t_{01} , t_{12} và t_{24} có các dải tần tương ứng và âm điệu t_L của tín hiệu dải tần thấp, nghĩa là, dải lõi có thể được biểu diễn như được thể hiện ở phương trình 10 sau đây.

【Phương trình 10】

$$t_{01} = 0.2 \log_{10} \left(\frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [0, \dots, 1kHz]$$

$$t_{12} = 0.2 \log_{10} \left(\frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [1, \dots, 2kHz]$$

$$t_{24} = 0.2 \log_{10} \left(\frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [2, \dots, 4kHz]$$

$$t_L = \max(t_{01}, t_{12}, t_{24})$$

Trong khi đó, lỗi dự đoán tuyến tính err có thể thu được bằng cách sử dụng bộ lọc mã hóa dự đoán tuyến tính (linear prediction coding, LPC) và có thể được sử dụng để loại bỏ các thành phần âm mạnh. Nói cách khác, chế độ mã hóa miền phổ có thể có hiệu quả hơn đối với các thành phần âm mạnh so với chế độ mã hóa kích thích miền tần số.

Trạng thái phía trước $cond_{front}$ để chuyển sang chế độ mã hóa kích thích miền tần số bằng cách sử dụng các âm điệu và lỗi dự đoán tuyến tính thu được như được mô tả ở trên có thể được thể hiện như được thể hiện ở phương trình 11 sau đây.

【Phương trình 11】

$$cond_{front} = t_{12} > t_{12front} \text{ và } t_{24} > t_{24front} \text{ và } t_L > t_{Lfront} \text{ và } err > err_{front}$$

Ở đây, $t_{12front}$, $t_{24front}$, t_{Lfront} và err_{front} là các giá trị ngưỡng và có thể có các giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Trong khi đó, trạng thái phía sau $cond_{back}$ để kết thúc chế độ mã hóa kích thích miền tần số bằng cách sử dụng các âm điệu và lỗi dự đoán tuyến tính thu được như được mô tả ở trên có thể được biểu diễn như được thể hiện ở phương trình 12 sau đây.

【Phương trình 12】

$$cond_{back} = t_{12} < t_{12back} \text{ và } t_{24} < t_{24back} \text{ và } t_I < t_{Iback}$$

Ở đây, t_{12back} , t_{24back} , t_{Iback} là các giá trị ngưỡng và có thể có các giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Nói cách khác, có thể xác định được xem chỉ số $state_{TTSS}$ biểu thị xem sự mã hóa kích thích miền tần số (ví dụ, GSC) có thích hợp hơn so với sự mã hóa miền phổ là 1 hay không bằng cách xác định xem trạng thái phía trước được thể hiện ở phương trình 11 có được thỏa mãn hay không hoặc trạng thái phía sau được thể hiện ở phương trình 12 không được thỏa mãn hay không. Ở đây, bước xác định trạng thái phía sau được thể hiện ở phương trình 12 có thể là tùy chọn.

Ở hoạt động 702, nếu chỉ số $state_{TTSS}$ là 1, thì chế độ mã hóa kích thích miền tần số có thể được xác định là chế độ mã hóa cuối cùng. Trong trường hợp này, chế độ mã hóa miền phổ, là chế độ mã hóa ban đầu, được hiệu chỉnh thành chế độ mã hóa kích thích miền tần số, là chế độ mã hóa cuối cùng.

Ở hoạt động 705, nếu xác định được ở hoạt động 701 rằng chỉ số $state_{TTSS}$ là 0, thì chỉ số $statess$ để xác định xem tín hiệu âm thanh bao gồm đặc tính tiếng nói mạnh hay không có thể được kiểm tra. Nếu có lỗi trong quá trình xác định chế độ mã hóa miền phổ, thì chế độ mã hóa kích thích miền tần số có thể có hiệu quả hơn so với chế độ mã hóa miền phổ. Chỉ số $statess$ để xác định xem tín hiệu âm thanh bao gồm đặc tính tiếng nói mạnh có thể thu được hay không bằng cách sử dụng chênh lệch vc giữa tham số giọng nói và tham số tương quan.

Trạng thái phía trước $cond_{front}$ để chuyển sang chế độ tiếng nói mạnh bằng cách sử dụng chênh lệch vc giữa tham số giọng nói và tham số tương quan có thể được biểu diễn như được thể hiện ở phương trình 13 sau đây.

[Phương trình 13]

$$cond_{front} = vc > vc_{front}$$

Ở đây, vc_{front} là giá trị ngưỡng và có thể có giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Trong khi đó, trạng thái phía sau $cond_{back}$ để kết thúc chế độ tiếng nói mạnh bằng cách sử dụng chênh lệch vc giữa tham số giọng nói và tham số tương quan có thể được biểu diễn như được thể hiện ở phương trình 14 sau đây.

【Phương trình 14】

$$cond_{back} = vc < vc_{back}$$

Ở đây, vc_{back} là giá trị ngưỡng và có thể có giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Nói cách khác, ở hoạt động 705, có thể xác định được xem chỉ số $statess$ biểu thị xem sự mã hóa kích thích miền tần số (ví dụ GSC) có thích hợp hơn so với sự mã hóa miền phổ là 1 hay không bằng cách xác định xem trạng thái phía trước được thể hiện ở phương trình 13 có được thỏa mãn hay không hoặc trạng thái phía sau được thể hiện ở phương trình 14 không được thỏa mãn hay không. Ở đây, bước xác định trạng thái phía sau được thể hiện ở phương trình 14 có thể là tùy chọn.

Ở hoạt động 706, nếu xác định được ở hoạt động 705 rằng chỉ số $statess$ là 0, nghĩa là, tín hiệu âm thanh không bao gồm đặc tính tiếng nói mạnh, thì chế độ mã hóa miền phổ có thể được xác định là chế độ mã hóa cuối cùng. Trong trường hợp này, chế độ mã hóa miền phổ, là chế độ mã hóa ban đầu, được duy trì là chế độ mã hóa cuối cùng.

Ở hoạt động 707, nếu xác định được ở hoạt động 705 rằng chỉ số $statess$ là 1, nghĩa là, tín hiệu âm thanh bao gồm đặc tính tiếng nói mạnh, thì chế độ mã hóa kích thích miền tần số có thể được xác định là chế độ mã hóa cuối cùng. Trong trường hợp này, chế độ mã hóa miền phổ, là chế độ mã hóa ban đầu, được hiệu chỉnh thành chế độ mã hóa kích thích miền tần số, là chế độ mã hóa cuối cùng.

Bằng cách thực hiện các hoạt động 700, 701 và 705, lỗi trong quá trình xác định chế độ mã hóa miền phổ là chế độ mã hóa ban đầu có thể được hiệu chỉnh. Chi tiết là, chế độ mã hóa miền phổ, là chế độ mã hóa ban đầu, có thể được duy trì hoặc chuyển sang chế độ mã hóa kích thích miền tần số là chế độ mã hóa cuối cùng.

Trong khi đó, nếu xác định được ở hoạt động 700 rằng chế độ mã hóa ban đầu là chế độ mã hóa miền dự đoán tuyến tính ($state_{TS} = 0$), thì chỉ số $statesM$

dùng để xác định xem tín hiệu âm thanh bao gồm đặc tính nhạc mạnh có thể được kiểm tra hay không. Nếu có lỗi trong quá trình xác định chế độ mã hóa miền dự đoán tuyến tính, nghĩa là, chế độ mã hóa kích thích miền thời gian, thì chế độ mã hóa kích thích miền tần số có thể hiệu quả hơn so với chế độ mã hóa kích thích miền thời gian. Chỉ số $states_{SM}$ để xác định xem tín hiệu âm thanh bao gồm đặc tính nhạc mạnh có thể thu được hay không bằng cách sử dụng giá trị 1-vc thu được bằng cách lấy 1 trừ chênh lệch vc giữa tham số giọng nói và tham số tương quan.

Trạng thái phía trước $cond_{front}$ để chuyển sang chế độ nhạc mạnh bằng cách sử dụng giá trị 1-vc thu được bằng cách lấy 1 trừ chênh lệch vc giữa tham số giọng nói và tham số tương quan, có thể được biểu diễn như được thể hiện ở phương trình 15 sau đây.

【Phương trình 15】

$$cond_{front} = 1 - vc > vcm_{front}$$

Ở đây, vcm_{front} là giá trị ngưỡng và có thể có giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Trong khi đó, trạng thái phía sau $cond_{back}$ để kết thúc chế độ nhạc mạnh bằng cách sử dụng giá trị 1-vc thu được bằng cách lấy 1 trừ chênh lệch vc giữa tham số giọng nói và tham số tương quan, có thể được biểu diễn như được thể hiện ở phương trình 16 sau đây.

【Phương trình 16】

$$cond_{back} = 1 - vc < vcm_{back}$$

Ở đây, vcm_{back} là giá trị ngưỡng và có thể có giá trị thu được thông qua thực nghiệm hoặc mô phỏng trước.

Nói cách khác, ở hoạt động 709, có thể xác định được xem chỉ số $states_{SM}$ biểu thị xem sự mã hóa kích thích miền tần số (ví dụ, GSC) có thích hợp hơn so với sự mã hóa kích thích miền thời gian là 1 hay không bằng cách xác định xem trạng

thái phía trước được thể hiện ở phương trình 15 có được thỏa mãn hay không hoặc trạng thái phía sau được thể hiện ở phương trình 16 không được thỏa mãn hay không. Ở đây, bước xác định trạng thái phía sau được thể hiện ở phương trình 16 có thể là tùy chọn.

Ở hoạt động 710, nếu xác định được ở hoạt động 709 rằng chỉ số $states_M$ là 0 nghĩa là, tín hiệu âm thanh không bao gồm đặc tính nhạc mạnh, thì chế độ mã hóa kích thích miền thời gian có thể được xác định là chế độ mã hóa cuối cùng. Trong trường hợp này, chế độ mã hóa miền dự đoán tuyến tính, là chế độ mã hóa ban đầu, được chuyển sang chế độ mã hóa kích thích miền thời gian là chế độ mã hóa cuối cùng. Theo một phương án làm ví dụ, có thể xem xét rằng chế độ mã hóa ban đầu được duy trì mà không có các thay đổi, nếu chế độ mã hóa miền dự đoán tuyến tính tương ứng với chế độ mã hóa kích thích miền thời gian.

Ở hoạt động 707, nếu xác định được ở hoạt động 709 rằng chỉ số $states_M$ là 1 nghĩa là, tín hiệu âm thanh bao gồm đặc tính nhạc mạnh, thì chế độ mã hóa kích thích miền tần số có thể được xác định là chế độ mã hóa cuối cùng. Trong trường hợp này, chế độ mã hóa miền dự đoán tuyến tính, là chế độ mã hóa ban đầu, được hiệu chỉnh thành chế độ mã hóa kích thích miền tần số, là chế độ mã hóa cuối cùng.

Bằng cách thực hiện các hoạt động 700 và 709, lỗi trong quá trình xác định chế độ mã hóa ban đầu có thể được hiệu chỉnh. Chi tiết là, chế độ mã hóa miền dự đoán tuyến tính (ví dụ, chế độ mã hóa kích thích miền thời gian), là chế độ mã hóa ban đầu, có thể được duy trì hoặc chuyển sang chế độ mã hóa kích thích miền tần số là chế độ mã hóa cuối cùng.

Theo một phương án làm ví dụ, hoạt động 709 để xác định xem tín hiệu âm thanh bao gồm đặc tính nhạc mạnh để hiệu chỉnh lỗi trong quá trình xác định chế độ mã hóa miền dự đoán tuyến tính có thể là tùy chọn hay không.

Theo một phương án làm ví dụ khác, trình tự thực hiện hoạt động 705 để xác định xem tín hiệu âm thanh có bao gồm đặc tính tiếng nói mạnh hay không và hoạt động 701 để xác định xem chế độ mã hóa kích thích miền tần số thích hợp có thể bị đảo ngược hay không. Nói cách khác, sau hoạt động 700, hoạt động 705 có thể được thực hiện trước tiên, và sau đó hoạt động 701 có thể được thực hiện. Trong trường hợp này, các tham số được sử dụng đối với các bước xác định đó có thể được thay đổi theo yêu cầu của tình huống.

Fig.8 là sơ đồ khối minh họa cấu hình của thiết bị giải mã âm thanh 800 theo một phương án làm ví dụ.

Thiết bị giải mã âm thanh 800 được thể hiện trên Fig.8 có thể bao gồm bộ phận phân tách dòng bit 810, bộ phận giải mã miền phổ 820, bộ phận giải mã miền dự đoán tuyến tính 830 và bộ phận chuyển 840. Bộ phận giải mã miền dự đoán tuyến tính 830 có thể bao gồm bộ phận giải mã kích thích miền thời gian 831 và bộ phận giải mã kích thích miền tần số 833, trong đó bộ phận giải mã miền dự đoán tuyến tính 830 có thể được thể hiện dưới dạng ít nhất một trong số bộ phận giải mã kích thích miền thời gian 831 và bộ phận giải mã kích thích miền tần số 833. Trừ khi cần phải được thể hiện dưới dạng phần cứng riêng biệt, các thành phần nêu trên có thể có thể được tích hợp vào ít nhất một môđun và có thể được thực hiện dưới dạng ít nhất một bộ xử lý (không được thể hiện trên hình vẽ).

Dựa vào Fig.8, bộ phận phân tách dòng bit 810 có thể phân tách dòng bit đã nhận và tách thông tin về chế độ mã hóa và dữ liệu được mã hóa. Chế độ mã hóa này có thể tương ứng với chế độ mã hóa ban đầu thu được bằng cách xác định một trong số các chế độ mã hóa bao gồm chế độ mã hóa thứ nhất và chế độ mã hóa thứ hai tương ứng với các đặc tính của tín hiệu âm thanh hoặc chế độ mã hóa thứ ba được hiệu chỉnh từ chế độ mã hóa ban đầu nếu có lỗi trong quá trình xác định của chế độ mã hóa ban đầu.

Bộ phận giải mã miền phổ 820 có thể giải mã dữ liệu được mã hóa ở miền phổ từ dữ liệu mã hóa được tách.

Bộ phận giải mã miền dự đoán tuyến tính 830 có thể giải mã dữ liệu được mã hóa ở miền dự đoán tuyến tính từ dữ liệu mã hóa được tách. Nếu bộ phận giải mã miền dự đoán tuyến tính 830 bao gồm bộ phận giải mã kích thích miền thời gian 831 và bộ phận giải mã kích thích miền tần số 833, thì bộ phận giải mã miền dự đoán tuyến tính 830 có thể thực hiện giải mã kích thích miền thời gian hoặc giải mã kích thích miền tần số đối với dữ liệu mã hóa được tách này.

Bộ phận chuyển 840 có thể chuyển tín hiệu được khôi phục bởi bộ phận giải mã miền phổ 820 hoặc tín hiệu được khôi phục bởi bộ phận giải mã miền dự đoán tuyến tính 830 và có thể tạo ra tín hiệu được chuyển dưới dạng tín hiệu được khôi phục cuối cùng.

Fig.9 là sơ đồ khối minh họa cấu hình của thiết bị giải mã âm thanh 900 theo một phương án làm ví dụ khác.

Thiết bị giải mã âm thanh 900 có thể bao gồm bộ phận phân tách dòng bit 910, bộ phận giải mã miền phổ 920, bộ phận giải mã miền dự đoán tuyến tính 930, bộ phận chuyển 940 và môđun xử lý sau thông thường 950. Bộ phận giải mã miền dự đoán tuyến tính 930 có thể bao gồm bộ phận giải mã kích thích miền thời gian 931 và bộ phận giải mã kích thích miền tần số 933, trong đó bộ phận giải mã miền dự đoán tuyến tính 930 này có thể được thể hiện dưới dạng ít nhất một trong số bộ phận giải mã kích thích miền thời gian 931 và bộ phận giải mã kích thích miền tần số 933. Trừ khi cần phải được thể hiện dưới dạng phần cứng riêng biệt, các thành phần nêu trên có thể có thể được tích hợp vào ít nhất một môđun và có thể được thực hiện dưới dạng ít nhất một bộ xử lý (không được thể hiện trên hình vẽ). So với thiết bị giải mã âm thanh 800 được thể hiện trên Fig.8, thiết bị giải mã âm thanh 900 có thể còn bao gồm môđun xử lý sau thông thường 950, và do đó các phần mô tả về các thành phần giống với các phần mô tả về các thành phần của thiết bị giải mã âm thanh 800 sẽ được lược bỏ.

Dựa vào Fig.9, môđun xử lý sau thông thường 950 có thể thực hiện xử lý âm thanh nổi chung, xử lý xung quanh và/hoặc xử lý mở rộng băng thông, tương ứng với môđun xử lý trước thông thường (205 trên Fig.2).

Các phương pháp theo các phương án làm ví dụ có thể được ghi dưới dạng các chương trình thực thi được bằng máy tính và có thể được thực hiện trong các máy tính kỹ thuật số sử dụng cho mục đích chung mà thực thi các chương trình bằng cách sử dụng vật ghi đọc được bằng máy tính không khả biến. Ngoài ra, các cấu trúc dữ liệu, lệnh chương trình hoặc tệp dữ liệu, có thể được sử dụng theo các phương án, có thể được ghi trên vật ghi đọc được bằng máy tính không khả biến theo các cách khác nhau. Vật ghi đọc được bằng máy tính không khả biến là thiết bị lưu trữ dữ liệu bất kỳ mà có thể lưu trữ dữ liệu có thể sau đó được đọc bởi hệ thống máy tính. Các ví dụ về vật ghi đọc được bằng máy tính không khả biến bao gồm các vật ghi lưu trữ từ, chẳng hạn như đĩa cứng, đĩa mềm và băng từ, các vật ghi quang, chẳng hạn như CD-ROM và DVD, các vật ghi quang-từ, chẳng hạn như đĩa quang, và các thiết bị phần cứng, chẳng hạn như ROM, RAM và bộ nhớ chớp, được tạo cấu hình cụ thể để lưu trữ và thực thi các lệnh chương trình. Ngoài ra, vật ghi đọc được bằng máy tính không khả biến có thể là vật ghi truyền dẫn để truyền tín hiệu chỉ

định các lệnh chương trình, cấu trúc dữ liệu hoặc dạng tương tự. Các ví dụ về các lệnh chương trình này có thể bao gồm không chỉ các mã ngôn ngữ cơ học được tạo bởi trình biên dịch mà cả các mã ngôn ngữ bậc cao thực thi được bằng máy tính sử dụng trình thông dịch hoặc dạng tương tự.

Trong khi các phương án làm ví dụ đã được thể hiện và mô tả ở trên một cách cụ thể, người có hiểu biết trung bình trong lĩnh vực kỹ thuật sẽ hiểu rằng các thay đổi khác nhau về hình thức và các chi tiết có thể được thực hiện ở sáng chế này mà không lệch khỏi mục đích và phạm vi của sáng chế như được xác định bởi các điểm yêu cầu bảo hộ kèm theo. Các phương án làm ví dụ này chỉ nên được xem xét theo nghĩa mô tả và không nhằm mục đích làm giới hạn sáng chế. Do đó, phạm vi của sáng chế được xác định không phải bởi phần mô tả chi tiết theo các phương án làm ví dụ mà bởi các điểm yêu cầu bảo hộ kèm theo, và tất cả các khác biệt nằm trong phạm vi này sẽ được hiểu là nằm trong sáng chế này.

YÊU CẦU BẢO HỘ

1. Phương pháp để mã hóa tín hiệu âm thanh, phương pháp này bao gồm các bước:

nhận tín hiệu âm thanh;

thu, được thực hiện bởi ít nhất một bộ xử lý, các tham số thứ nhất của khung hiện thời của tín hiệu âm thanh;

chọn, được thực hiện bởi ít nhất một bộ xử lý, loại khung hiện thời trong tín hiệu âm thanh trong số các loại khung bao gồm loại nhạc và loại tiếng nói, dựa vào các tham số thứ nhất của khung hiện thời bằng cách sử dụng mô hình hỗn hợp Gauss (Gaussian mixture model, GMM);

thu các tham số thứ hai bao gồm âm điệu thứ nhất, âm điệu thứ hai và âm điệu thứ ba;

tạo ra các điều kiện, trong đó mỗi trong số các điều kiện này được tạo ra dựa vào sự kết hợp của các tham số thứ hai đã thu;

xác định, được thực hiện bởi ít nhất một bộ xử lý, xem lỗi có xảy ra ở loại khung hiện thời đã chọn hay không dựa vào việc xem ít nhất một trong số các điều kiện có được đáp ứng hay không;

khi lỗi xảy ra ở loại khung hiện thời đã chọn, thì hiệu chỉnh, được thực hiện bởi ít nhất một bộ xử lý, loại khung hiện thời đã chọn;

mã hóa, được thực hiện bởi ít nhất một bộ xử lý, khung hiện thời, dựa vào loại đã hiệu chỉnh hoặc loại khung hiện thời đã chọn; và

tạo ra dòng bit dựa vào khung hiện thời đã mã hóa,

trong đó âm điệu thứ nhất thu được từ dải băng con nằm trong khoảng từ 0 đến 1 kHz, âm điệu thứ hai thu được từ dải băng con nằm trong khoảng từ 1 đến 2 kHz và âm điệu thứ ba thu được từ dải băng con nằm trong khoảng từ 2 đến 4 kHz, và

trong đó bước hiệu chỉnh bao gồm các bước:

khi lỗi xảy ra ở loại khung hiện thời đã chọn và loại khung hiện thời đã chọn này là loại tiếng nói, thì hiệu chỉnh loại khung hiện thời đã chọn từ loại tiếng nói sang loại nhạc; và

khi lỗi xảy ra ở loại khung hiện thời đã chọn và loại khung hiện thời đã chọn này là loại nhạc, thì hiệu chỉnh loại khung hiện thời đã chọn từ loại nhạc sang loại tiếng nói.

2. Phương pháp theo điểm 1, trong đó bước hiệu chỉnh được thực hiện dựa vào ít nhất hai trạng thái độc lập.

3. Phương pháp theo điểm 1, trong đó các tham số thứ hai còn bao gồm chênh lệch

giữa tham số giọng nói và tham số tương quan.

4. Phương pháp theo điểm 1, trong đó bước xác định xem lỗi xảy ra ở loại khung hiện thời đã chọn có xảy ra hay không bao gồm các bước:

xác định xem khung hiện thời có các đặc tính tiếng nói hay không khi khung hiện thời được phân loại làm loại nhạc; và

xác định xem khung hiện thời có các đặc tính nhạc hay không khi khung hiện thời được phân loại làm loại tiếng nói.

5. Phương pháp theo điểm 1, trong đó bước hiệu chỉnh bao gồm các bước:

hiệu chỉnh sự phân loại của khung hiện thời, khi khung hiện thời được phân loại làm loại nhạc và có các đặc tính tiếng nói; và

hiệu chỉnh sự phân loại của khung hiện thời, khi khung hiện thời được phân loại làm loại tiếng nói và có các đặc tính nhạc.

6. Phương pháp theo điểm 1, trong đó bước xác định còn được thực hiện dựa vào tham số tồn lưu (hangover parameter) được sử dụng để ngăn chặn chuyển mạch tần số giữa các chế độ mã hóa.

FIG.1

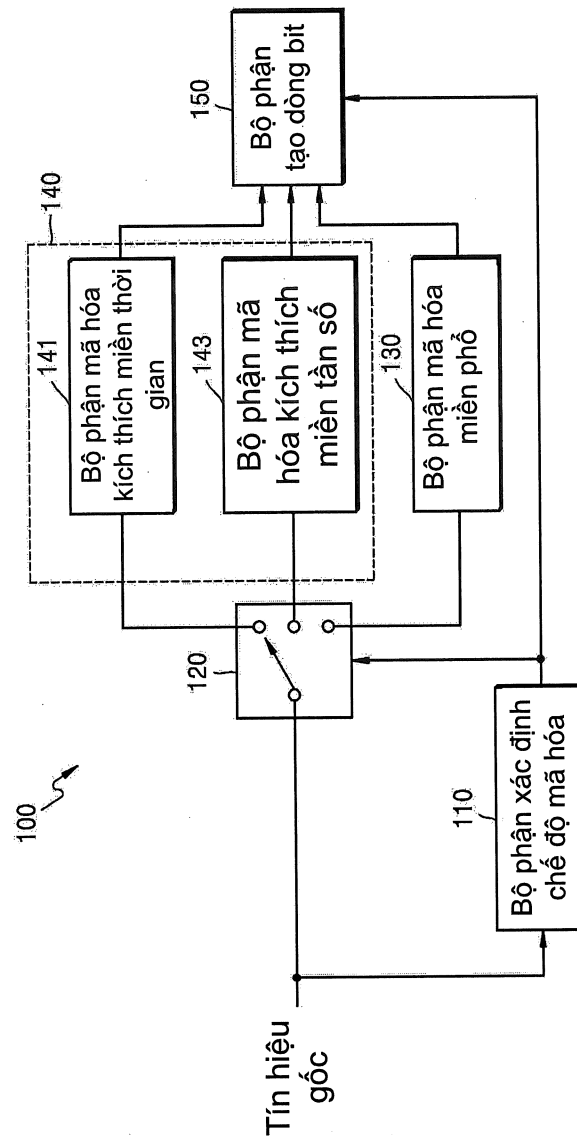


FIG.2

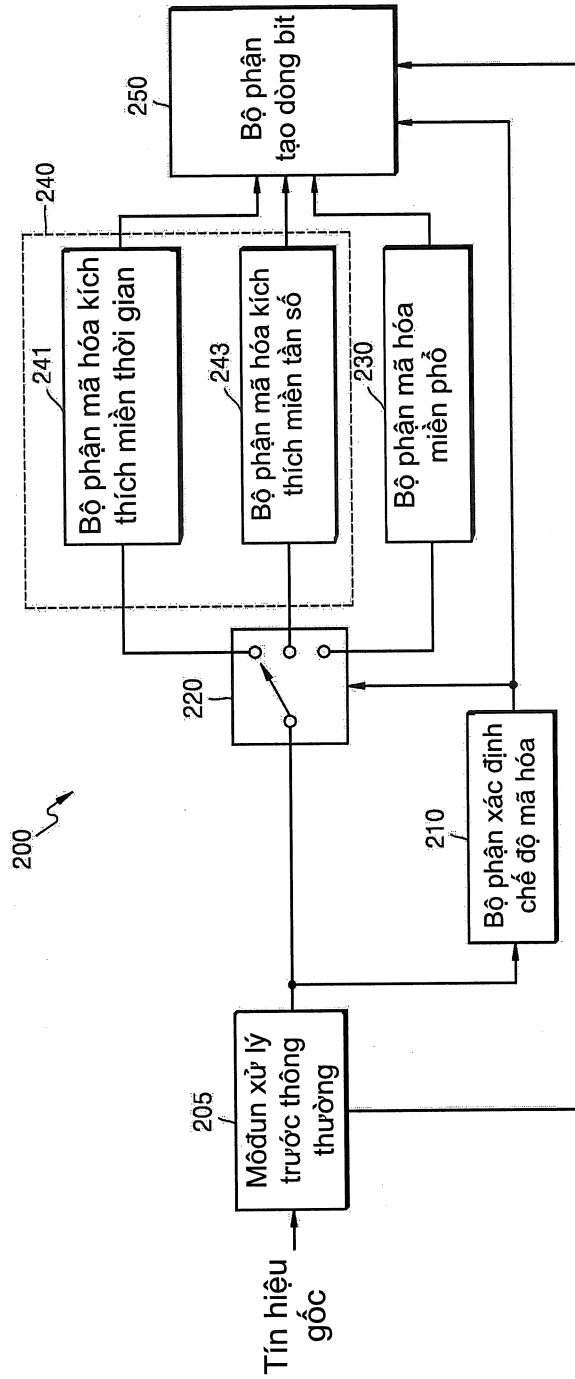


FIG.3

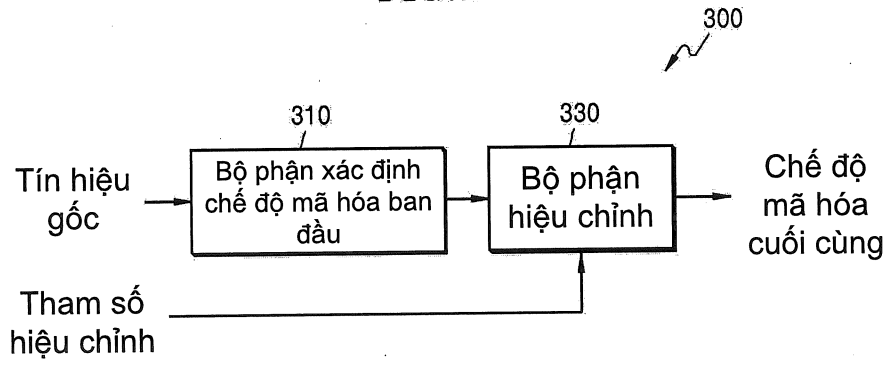


FIG.4

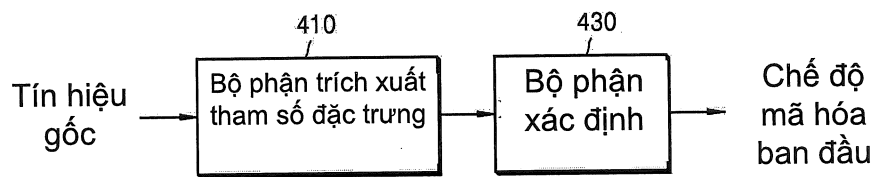


FIG.5

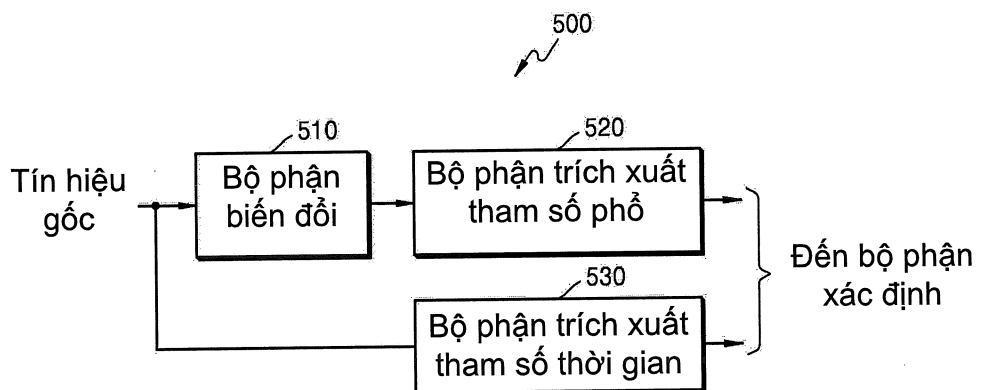


FIG. 6

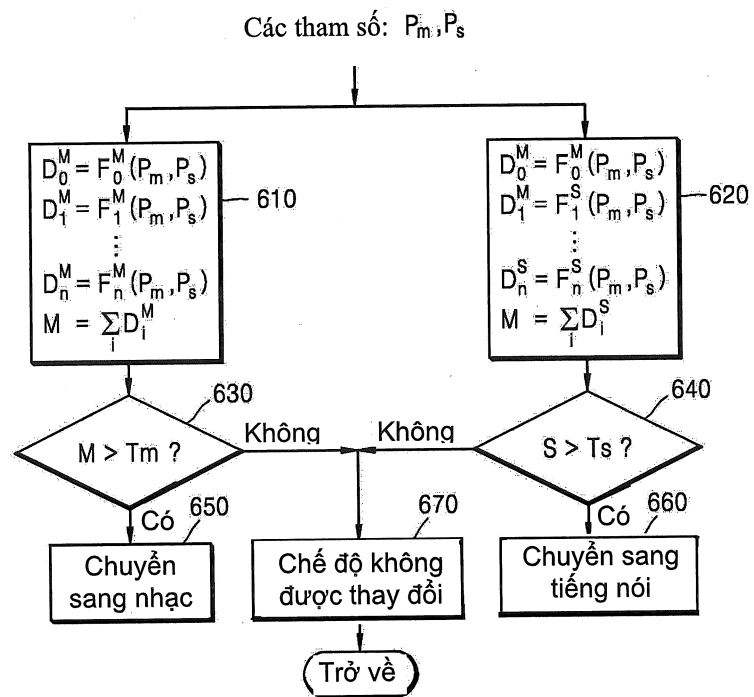


FIG. 7

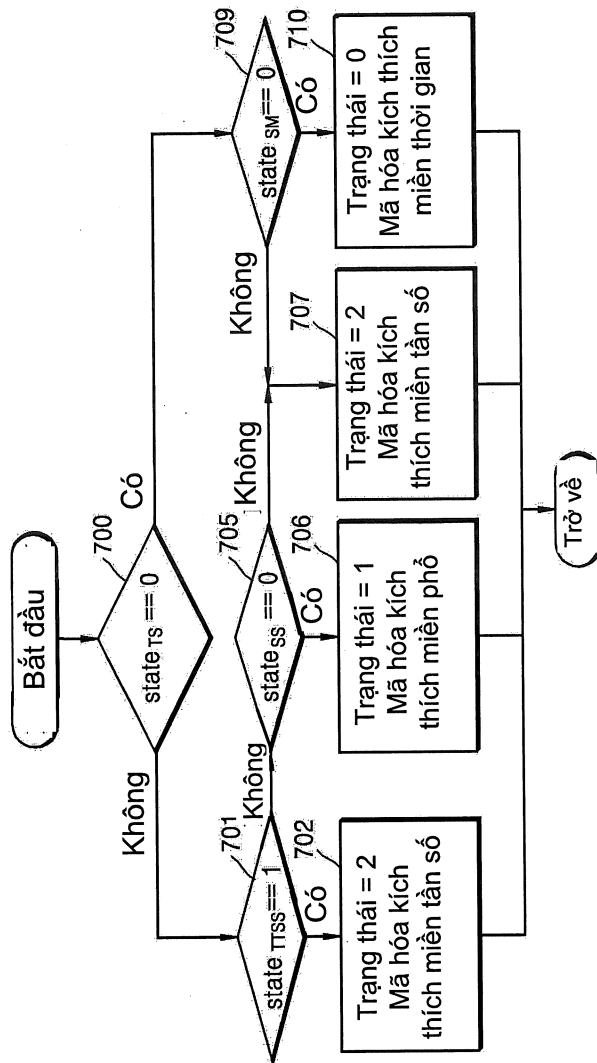


FIG. 8

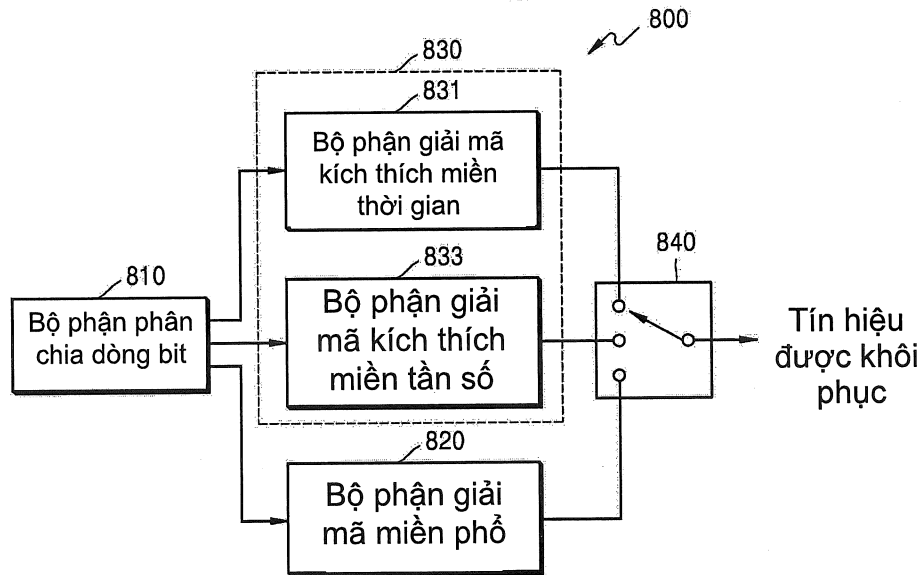


FIG. 9

