



(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)



1-0051763

(51)^{2020.01} G06Q 50/10; H04W 4/14; H04W 12/06 (13) B

(21) 1-2019-07197

(22) 19/12/2019

(30) TW 108146063 16/12/2019 TW

(45) 25/09/2025 450

(43) 25/06/2021 399A

(73) ATEN International Co., Ltd. (TW)

3F, No.125, Sec.2, Datung Rd., Xizhi Dist., New Taipei City 221, Taiwan

(72) Huang, Hsien-Chao (TW); Ting, Yi-Tzu (TW); Chen, Yu-Yun (TW); Yang, Chung-Wen (TW).

(74) Công ty TNHH Sở hữu trí tuệ Việt Á (VIET A IP CO.,LTD)

(54) THIẾT BỊ HỖ TRỢ HÁT, PHƯƠNG PHÁP HỖ TRỢ HÁT VÀ PHƯƠNG TIỆN LƯU TRỮ DỮ LIỆU ĐƯỢC BẰNG MÁY TÍNH PHI TẠM THỜI BAO GỒM CÁC LỆNH ĐỀ THỰC HIỆN PHƯƠNG PHÁP TRÊN

(21) 1-2019-07197

(57) Sáng chế đề xuất hệ thống hỗ trợ hát, phương pháp hỗ trợ hát và phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời bao gồm các lệnh để thực hiện phương pháp trên. Khi đoạn hát đã được trình bày không xuất hiện trong giai đoạn cần trình bày, quy trình duy trì việc hát được thực hiện. Khi đoạn hát đã được trình bày sai cao độ, quy trình điều chỉnh cao độ được thực hiện.

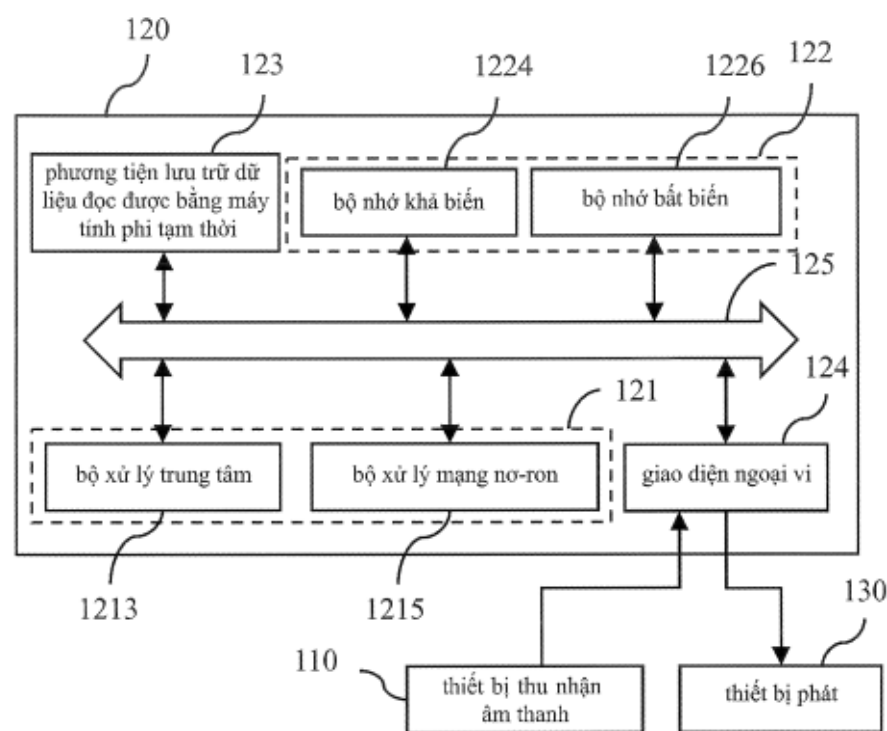


FIG. 2

Lĩnh vực kỹ thuật được đề cập

Sáng chế liên quan đến công nghệ hỗ trợ thông minh, cụ thể là hệ thống hỗ trợ hát, phương pháp hỗ trợ hát và phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời bao gồm các lệnh để thực hiện phương pháp trên.

Tình trạng kỹ thuật của sáng chế

Các thiết bị hát đã biết bởi các nhà sáng chế được cung cấp cho người dùng để điều chỉnh tông giọng (tone) của bản nhạc đệm khi phát. Tuy nhiên, các thiết bị hát đã biết không thể giải quyết được các vấn đề khi người dùng hát sai cao độ (sai nốt nhạc). Ngoài ra, trong khi hát, khi người dùng quên lời hát hoặc giai điệu, các thiết bị hát đã biết chỉ có thể phát đoạn nhạc gốc để giúp người dùng hát bài hát (cụ thể là chế độ hướng dẫn hát), có thể có ích cho việc luyện tập nhưng có thể không có tác dụng khi biểu diễn.

Mục đích của sáng chế

Trước những vấn đề trên, một hoặc một số phương án của sáng chế này đề xuất hệ thống hỗ trợ hát, phương pháp hỗ trợ hát và phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời bao gồm các lệnh để thực hiện phương pháp hỗ trợ hát.

Bản chất kỹ thuật của sáng chế

Trong một phương án, hệ thống hỗ trợ hát bao gồm thiết bị thu nhận âm thanh, thiết bị xử lý và thiết bị phát, và phương pháp hỗ trợ hát liên quan đến quy trình duy trì việc hát. Thiết bị thu nhận âm thanh được điều chỉnh để thu nhận đoạn hát được trình bày. Thiết bị xử lý được điều chỉnh để dò tìm xem đoạn hát đã được trình bày có xuất hiện trong giai đoạn cần trình bày hay không, nếu không, thiết bị xử lý thực hiện quy trình duy trì việc hát. Quy trình duy trì việc hát được thực hiện bằng cách đi kèm với mô hình âm thanh bao gồm bộ mã hóa và bộ giải mã. Mô hình âm thanh có thể được huấn luyện bởi dữ liệu huấn luyện (training data) liên quan đến ca sĩ gốc để thu được mô hình âm thanh ca sĩ gốc, và mô hình âm thanh có thể được huấn luyện bởi dữ liệu huấn luyện liên quan đến người dùng để thu được mô hình âm thanh người dùng. Quy trình duy trì việc hát bao gồm biến đổi phân đoạn âm thanh ca sĩ gốc thành đặc trưng âm thanh ca sĩ gốc; nhập đặc trưng âm thanh ca sĩ gốc vào bộ mã hóa của mô hình âm thanh ca sĩ gốc; thu nhận đầu ra của bộ mã hóa của mô hình âm thanh ca sĩ gốc bằng bộ giải mã của mô hình âm thanh người dùng; thu được đặc trưng âm thanh người dùng từ đầu ra của bộ giải mã của mô hình âm thanh người dùng; và biến đổi đặc trưng âm thanh người dùng thành đoạn hát tổng hợp bằng bộ phân tích và tổng hợp giọng nói. Thiết bị phát được điều chỉnh để trích xuất đoạn hát tổng hợp.

Trong một phương án của sáng chế, hệ thống hỗ trợ hát bao gồm thiết bị thu nhận âm thanh, thiết bị xử lý và thiết bị phát, và phương pháp hỗ trợ hát có liên quan đến quy trình điều chỉnh cao độ. Thiết bị thu nhận âm thanh được điều chỉnh để thu nhận đoạn hát đã được trình bày. Thiết bị xử lý được điều chỉnh để xác định xem cao độ của đoạn hát đã được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, nếu không, thiết bị xử lý sẽ thực hiện quy trình điều chỉnh cao độ. Quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa thứ hai và bộ giải mã thứ hai. Mô hình cao độ có thể được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ có thể được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng. Quy trình điều chỉnh cao độ bao gồm biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng; biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc; nhập phổ âm thanh người dùng vào bộ mã hóa của mô hình cao độ người dùng để thu được đặc trưng âm thanh người dùng; nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc; nhập đặc trưng âm thanh người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã về mô hình cao độ người dùng; thu được phổ âm thanh người dùng đã điều chỉnh từ đầu ra của bộ giải mã của mô

hình cao độ người dùng; và biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ. Thiết bị phát được điều chỉnh để trích xuất đoạn hát tổng hợp.

Theo đó, dựa trên một hoặc một số phương án của sáng chế, bằng cách dò tìm xem người dùng có hát trong giai đoạn cần hát hay không nhằm xác định xem quy trình duy trì việc hát có cần được thực hiện hay không, quy trình duy trì việc hát có thể được cung cấp một cách tự động bằng cách sử dụng âm thanh của chính người dùng khi người dùng quên lời bài hát hoặc ngừng hát vì một số lý do. Bằng cách áp dụng quy trình duy trì việc hát, đoạn hát duy trì được tạo ra có thể duy trì cao độ của ca sĩ gốc; hơn nữa, do mô hình âm thanh có cấu trúc tự động mã hóa, nên đoạn hát duy trì được tạo ra có thể mang âm sắc của người dùng. Ngoài ra, bằng cách dò tìm cao độ của đoạn hát được trình bày bởi người dùng, quy trình điều chỉnh cao độ có thể được thực hiện khi người dùng hát sai cao độ. Bằng việc thực hiện quy trình điều chỉnh cao độ, cao độ của người dùng có thể được điều chỉnh một cách tự động; hơn nữa, vì mô hình cao độ có cấu trúc tự động mã hóa, đoạn hát được trình bày có thể mang âm sắc của người dùng.

Mô tả vắn tắt các hình vẽ

Sáng chế sẽ được hiểu đầy đủ hơn từ mô tả chi tiết được nêu dưới đây với các hình vẽ chỉ nhằm mục đích minh họa, và do đó việc này không làm giới hạn sáng chế, trong đó:

FIG. 1 minh họa sơ đồ khối của hệ thống hỗ trợ hát theo một phương án của sáng chế;

FIG. 2 minh họa sơ đồ khối của thiết bị xử lý theo một phương án của sáng chế;

FIG. 3 minh họa sơ đồ khối của sản phẩm chương trình máy tính để thực hiện phương pháp hỗ trợ hát theo một phương án của sáng chế;

FIG. 4 minh họa lưu đồ của phương pháp hỗ trợ hát theo một phương án của sáng chế;

FIG. 5 minh họa sơ đồ khối của mô hình âm thanh theo một phương án của sáng chế;

FIG. 6 minh họa lưu đồ của quy trình duy trì việc hát theo một phương án của sáng chế;

FIG. 7 minh họa sơ đồ khối của mô-đun duy trì việc hát theo một phương án của sáng chế;

FIG. 8 minh họa sơ đồ khối của mô hình cao độ theo một phương án của sáng chế;

FIG. 9 minh họa lưu đồ của quy trình điều chỉnh cao độ theo một phương án của sáng chế; và

FIG. 10 minh họa sơ đồ khối của mô-đun điều chỉnh cao độ theo một phương án của sáng chế.

Mô tả chi tiết sáng chế

Vui lòng tham khảo FIG. 1, minh họa sơ đồ khối của hệ thống hỗ trợ hát 100 theo một phương án của sáng chế. Trong phương án này, hệ thống hỗ trợ hát 100 bao gồm thiết bị thu nhận âm thanh 110, thiết bị xử lý 120 và thiết bị phát 130 được kết nối liên tiếp. Thiết bị thu nhận âm thanh 110 và thiết bị xử lý 120 có thể được kết nối với nhau thông qua kết nối dây (ví dụ: qua dây dẫn, qua cáp tuân theo giao thức truyền thông cụ thể (như chuẩn kết nối tuần tự đa dụng (USB)) hoặc không dây (ví dụ: qua Bluetooth hoặc mạng không dây). Tương tự, thiết bị phát 130 và thiết bị xử lý 120 có thể được kết nối với nhau thông qua kết nối dây (ví dụ: qua dây dẫn hoặc cáp) hoặc không dây (ví dụ: qua Bluetooth hoặc mạng không dây).

Thiết bị thu nhận âm thanh 110 được điều chỉnh để truy xuất đoạn hát đã được trình bày của người dùng và thiết bị thu nhận âm thanh có thể chứa micro đơn lẻ hoặc nhiều micro (ví dụ: dải micro). Micro có thể là micro động, micro tụ, micro hệ thống vi cơ điện tử (MEMS), v.v.

Thiết bị xử lý 120 thực hiện một cách chọn lọc quy trình hỗ trợ hát (cụ thể, phương pháp hỗ trợ hát) cho đoạn hát đã được trình bày được truy xuất bởi thiết bị thu nhận âm thanh 110, và quy trình này sẽ được mô tả trong các đoạn sau. Thiết bị xử lý 120 có thể là một hoặc nhiều hệ thống máy tính có khả năng tính toán; ví dụ, thiết bị xử lý 120 có thể là máy tính cá nhân, máy tính xách tay, điện thoại thông minh, máy tính bảng, cụm sơ-vơ (server), v.v. Vui lòng tham khảo FIG. 2. FIG. 2 minh họa sơ đồ khối của thiết bị xử lý 120 theo một phương án của sáng chế. Thiết bị xử lý 120 có bộ xử lý 121, bộ nhớ 122, phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời 123, giao diện ngoại vi 124 được điều chỉnh để kết nối với thiết bị thu nhận âm thanh 110 và thiết bị phát 130, và bus 125 được điều chỉnh để cho phép các linh kiện của thiết bị

xử lý 120 giao tiếp với nhau. Bus 125 có thể bao gồm, nhưng không bị giới hạn bởi, bus hệ thống, bus bộ nhớ, bus ngoại vi hoặc một hay nhiều sự kết hợp của chúng. Bộ xử lý 121 có thể bao gồm, nhưng không bị giới hạn bởi, bộ xử lý trung tâm (CPU) 1213 và bộ xử lý mạng nơ-ron (NPU) 1215. Bộ nhớ 122 có thể bao gồm, nhưng không bị giới hạn bởi, bộ nhớ khả biến 1224 (ví dụ: bộ nhớ truy cập ngẫu nhiên (RAM)) và bộ nhớ bất biến (bộ nhớ điện tĩnh) 1226 (ví dụ: bộ nhớ chỉ đọc (ROM)). Ví dụ, phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời 123 có thể là ổ cứng, ổ cứng thể rắn (SSD), v.v., để lưu trữ sản phẩm chương trình máy tính 300 (như thể hiện trong FIG. 3) bao gồm nhiều lệnh. Do đó, khi bộ xử lý 121 của hệ thống máy tính thực hiện các lệnh, hệ thống máy tính sẽ thực hiện phương pháp hỗ trợ hát.

Vui lòng tham khảo FIG. 1. Thiết bị phát 130 được điều chỉnh để phát đoạn hát đã được trình bày trong đó quy trình hỗ trợ hát đã được thực hiện hoặc không được thực hiện bởi thiết bị xử lý 120. Thiết bị phát có thể bao gồm một hoặc nhiều loa. Loa có thể là loa động lực (dynamic speaker), loa lõi điện cứng cân bằng (balanced armature speaker), v.v.

Trong một số phương án, bất kỳ hai trong số các thiết bị thu nhận âm thanh 110, thiết bị xử lý 120 và thiết bị phát 130 có thể được tích hợp thành một chi tiết. Ví dụ: thiết bị nhận âm thanh 110 và thiết bị phát 130 có thể là cùng một thiết bị trong micro tai nghe. Trong ví dụ khác, thiết bị thu nhận âm thanh 110 và thiết bị xử lý 120 có thể là cùng thiết bị trong điện thoại thông minh và được kết nối với thiết bị phát bên ngoài 130. Thay vào đó, trong một ví dụ, thiết bị phát 130 và thiết bị xử lý 120 có thể là cùng thiết bị trong máy tính cá nhân và được kết nối với thiết bị nhận âm thanh bên ngoài 110. Trong ví dụ khác, thiết bị thu nhận âm thanh 110, thiết bị xử lý 120 và thiết bị phát 130 có thể là cùng một thiết bị trong máy tính xách tay.

Trong một số phương án của sáng chế, thiết bị thu nhận âm thanh 110, thiết bị xử lý 120 và thiết bị phát 130 có thể là các thiết bị riêng biệt. Ví dụ: thiết bị xử lý 120 là máy tính cá nhân và máy tính cá nhân được kết nối với thiết bị thu nhận âm thanh ngoài 110 và thiết bị phát ngoài 130.

Trong một số phương án, thiết bị xử lý 120 bao gồm hai hoặc nhiều hệ thống máy tính, ví dụ: máy tính cá nhân và máy chủ, trong đó máy chủ cung cấp quy trình hỗ trợ hát. Máy tính cá nhân có thể có thiết bị thu nhận âm thanh ngoài hoặc tích hợp 110 và thiết bị phát ngoài hoặc tích hợp 130, và đoạn hát đã được trình bày có thể được truyền đến máy chủ qua mạng để xử lý, và đoạn hát được trình bày đã qua xử lý được gửi trở lại máy tính cá nhân bằng máy chủ thông qua mạng.

Vui lòng tham khảo FIG. 3 và FIG. 4. FIG. 3 minh họa sơ đồ khối của sản phẩm chương trình máy tính 300 để thực hiện phương pháp hỗ trợ hát theo một phương án của sáng chế. FIG. 4 minh họa lưu đồ của phương pháp hỗ trợ hát theo một phương án của sáng chế. Trong phương án này, sản phẩm chương trình máy tính 300 bao gồm mô-đun dò tìm giọng hát con người 310, mô-đun duy trì việc hát 320, mô-đun điều chỉnh cao độ 330, mô-đun tách nhạc 340 và mô-đun trộn âm thanh 350.

Đầu tiên, mô-đun tách nhạc 340 thực hiện việc tách rời tệp bài hát mà người dùng sẽ trình bày, cụ thể là mô-đun tách nhạc truy xuất nhạc đệm và đoạn hát của ca sĩ gốc (sau đây gọi là đoạn hát ca sĩ gốc) từ bài hát. Mô-đun tách nhạc 340 có thể được thực hiện bởi các thuật toán như phân tích thành phần chính mạnh mẽ (RPCA), lặp lại kỹ thuật trích xuất mẫu (REPET), mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi quy sâu (DRNN), v.v. Trong các ví dụ mà ở đó RPCA hoặc REPET được sử dụng, nhạc đệm trong bài hát được coi là xuất hiện lặp đi lặp lại (vì âm thanh của nhạc cụ đều đặn hơn so với âm thanh của con người), và hai thuật toán được áp dụng để tìm ra các phần lặp đi lặp lại từ bài hát. Do đó, các phần lặp đi lặp lại (tức là nhạc đệm) có thể được tách rời khỏi bài hát, và phần hát của con người có thể được tìm ra, bằng cách đó, phần hát của con người và nhạc đệm có thể được tách rời. Thuật toán CNN và thuật toán DRNN sử dụng mô hình mạng nơ-ron học sâu (deep-learned neural network model) để tách rời nhạc đệm và phần hát của con người với nhau; nói cách khác, đầu vào cho trước dành cho mô hình mạng nơ-ron là đoạn hát có nhạc đệm, và sau đó, tham số của mô hình mạng nơ-ron thu được nhờ việc học sâu (deep-learning) có thể tự động được trích xuất thành đoạn nhạc đệm thuần túy và đoạn hát thuần túy của con người. Cụ thể, trong phương án này, mô hình mạng nơ-ron có thể được huấn luyện để cho phép mạng nơ-ron học được cách tách rời đoạn hát của con người khỏi bài hát một cách chính xác.

Trong bước S401, mô-đun dò tìm giọng hát con người 310 dò tìm âm thanh (đoạn hát được trình bày) mà thiết bị thu nhận âm thanh 110 thu nhận được để dò tìm xem đoạn hát được trình bày có xuất hiện trong giai đoạn cần trình bày không. Giai đoạn cần trình bày có thể thu nhận được dựa theo giai đoạn nơi đoạn hát ca sĩ gốc xuất hiện. Mô-đun dò tìm giọng hát con người 310 có thể được thực hiện bởi các thuật toán như cây phân loại và hồi quy (CART), dò tìm hoạt động giọng nói (VAD), VadNet, v.v. Nếu đoạn hát được trình bày không xuất hiện trong giai đoạn cần trình bày, bước S402 được thực hiện; nếu đoạn hát được trình bày xuất hiện trong giai đoạn cần trình bày, bước S403 được thực hiện.

Trong bước S402, mô-đun duy trì việc hát 320 thực hiện quy trình duy trì việc hát sẽ được mô tả trong các đoạn sau, từ đó quy trình duy trì việc hát có thể tạo ra đoạn hát cần được trình bày (sau đây gọi là đoạn hát tổng hợp), bằng cách đó hỗ trợ cho tình huống quên lời bài hát.

Trong bước S403, mô-đun điều chỉnh cao độ 330 xác định xem cao độ của đoạn hát đã được trình bày (sau đây gọi cao độ trình bày) có nhất quán với cao độ của đoạn hát ca sĩ gốc (sau đây gọi là cao độ ca sĩ gốc) hay không. Nếu các cao độ nhất quán với nhau, bước S404 được thực hiện; nếu các cao độ không nhất quán với nhau, bước S405 được thực hiện.

Việc xác định liệu các cao độ có nhất quán với nhau hay không sử dụng mô-đun điều chỉnh cao độ 330 có thể được thực hiện bởi các thuật toán như thuật toán tương quan chéo chuẩn hóa (normalized cross correlation), thuật toán dò tìm cao độ đa âm (polyphonic pitch detection), v.v. . Trong các cách tiếp cận này, các dạng sóng xuất hiện lặp đi lặp lại được suy ra từ đoạn âm thanh, và khoảng thời gian giữa các lần xuất hiện của các dạng sóng được tính toán để thu được thời gian cần thiết cho từng dạng sóng (cụ thể là chu kỳ của dạng sóng), bằng cách đó cao độ (tần số âm thanh) của đoạn âm thanh có thể được tính bằng cách sử dụng các chu kỳ của dạng sóng. Theo đó, cao độ của ca sĩ gốc và cao độ của người dùng có thể được suy ra và so sánh với nhau (ví dụ, bằng phép trừ), để xác định xem chúng có nhất quán với nhau hay không. Nếu các cao độ được xác định là nhất quán với nhau, thì cao độ của ca sĩ gốc tương đương với cao độ của người dùng; nếu các cao độ được xác định là không nhất quán với nhau, thì cao độ của ca sĩ gốc không tương đương với cao độ của người dùng. Các phương án không bị giới hạn ở đây. Trong một số phương án, có

thể áp dụng các phần mềm mã nguồn mở như “World” hoặc “Straight” để truy xuất cao độ của con người để so sánh. Nói cách khác, bất kỳ phần mềm nguồn mở nào trong số này có thể biến đổi cao độ của con người thành tham số âm thanh cho việc so sánh tiếp theo.

Trong bước S404, mô-đun điều chỉnh cao độ 330 thực hiện quy trình điều chỉnh cao độ mà sẽ được mô tả trong các đoạn sau để điều chỉnh cao độ đã trình bày trở lại cao độ gốc. Do đó, tình trạng hát sai nốt có thể được cải thiện.

Trong bước S405, mô-đun trộn âm thanh 350 trộn phần nhạc đệm được tách rời khỏi mô-đun tách nhạc 340 với đoạn hát đã được trình bày mà không cần điều chỉnh, với đoạn hát tổng hợp được tạo trong bước S402, hoặc với đoạn hát đã được trình bày trong đó cao độ được điều chỉnh, và đoạn âm thanh đã trộn này được trích xuất bởi thiết bị phát 130.

Mô hình âm thanh được sử dụng trong quy trình duy trì việc hát được giới thiệu trước phần giới thiệu quy trình duy trì việc hát. Vui lòng tham khảo FIG. 5, minh họa sơ đồ khối của mô hình âm thanh 500 theo một phương án của sáng chế. Trong phương án này, mô hình âm thanh 500 là mô hình mạng nơ-ron với chế độ học tập có giám sát, và mô hình âm thanh 500 bao gồm bộ mã hóa 510 và bộ giải mã 520. Nói cách khác, trong phương án này, mô hình âm thanh 500 có cấu trúc giải mã tự động. Bộ mã hóa 510 được điều chỉnh để biến đổi đặc trưng âm thanh đầu vào thành vector đặc trưng, trong khi đó bộ giải mã 520 được điều chỉnh để biến đổi vector đặc trưng thành đặc trưng âm thanh. Sau khi có nhiều dữ liệu huấn luyện (ví dụ như âm thanh kèm văn bản) của người nào đó được đưa vào mô hình âm thanh 500, các tham số trọng số (weighting parameters) có thể được đồng quy, và các tham số trọng số được đi kèm với mô hình âm thanh 500 để trở thành một mô hình âm thanh 500 đã được huấn luyện tương ứng với một người nhất định, trong đó dữ liệu huấn luyện là nhiều đoạn âm thanh của một người nhất định. Ví dụ: nếu nhiều dữ liệu âm thanh liên quan đến người dùng (tức là dữ liệu huấn luyện người dùng) được cung cấp làm dữ liệu huấn luyện, mô hình âm thanh 500 có thể được huấn luyện thành mô hình âm thanh người dùng 500; ngược lại, nếu nhiều dữ liệu âm thanh của ca sĩ (tức là dữ liệu huấn luyện ca sĩ gốc) được cung cấp làm dữ liệu huấn luyện, mô hình âm thanh 500 có thể được huấn luyện thành mô hình âm thanh ca sĩ gốc 500. Nói cách khác, mô hình âm thanh 500 với các tham số trọng số khác nhau có thể được

huấn luyện bằng việc sử dụng dữ liệu huấn luyện liên quan đến nhiều người khác nhau. Ngoài ra, trong một số phương án, mô hình âm thanh có thể được thiết lập bởi cá nhân nhất định, thông qua việc nhập nhiều từ hoặc cụm từ bằng cách nói hoặc bằng cách nhập vào nhiều bài hát bằng cách hát theo thời gian thực.

Trong một số phương án, như thể hiện trong FIG. 5, bộ mã hóa 510 bao gồm lớp chập (convolution layer) 511, đơn vị tuyến tính cổng (gated linear unit - GLU) 512 và khối dư (residual block) sáu lớp 513; ngược lại, bộ giải mã 520 bao gồm lớp giải chập (deconvolution layer) 521 và đơn vị tuyến tính cổng 522. Tuy nhiên, cần hiểu rằng các thành phần của bộ mã hóa 510 và bộ giải mã 520 không giới hạn ở các yếu tố và sự kết hợp nêu trên.

Vui lòng tham khảo FIG. 6, minh họa lưu đồ của quy trình duy trì duy trì việc hát theo một phương án của sáng chế. Trong bước S601, phân đoạn âm thanh ca sĩ gốc được biến đổi thành đặc trưng âm thanh (sau đây gọi là đặc trưng âm thanh ca sĩ gốc). Trong phương án này, đặc trưng âm thanh ca sĩ gốc có thể là các hệ số âm tần Mel (Mel frequency cepstral coefficients – MFCC, sau đây gọi là MFCC), nhưng các phương án của sáng chế không bị giới hạn tại đây; đặc trưng âm thanh ca sĩ gốc có thể là, ví dụ, đường bao âm phổ, tần số cơ bản và tính phi chu kỳ. Trong phương án này, quá trình tiền xử lý có thể được áp dụng cho phân đoạn âm thanh ca sĩ gốc, ví dụ, đoạn hát ca sĩ gốc và nhạc đệm được tách rời, do đó chỉ có đoạn hát ca sĩ gốc được xử lý để thực hiện việc biến đổi đặc trưng âm thanh.

Trong bước S602, đặc trưng âm thanh ca sĩ gốc được nhập vào bộ mã hóa 510 của mô hình âm thanh ca sĩ gốc 500; vì mô hình âm thanh ca sĩ gốc 500 có các tham số trọng số tương ứng với ca sĩ gốc, bộ mã hóa 510 có thể tạo ra vector đặc trưng tương ứng với đặc trưng âm thanh ca sĩ gốc. Trong bước S603, bộ giải mã 520 của mô hình âm thanh người dùng 500 được sử dụng để nhận vector đặc trưng được trích xuất bởi bộ mã hóa 510 của mô hình âm thanh ca sĩ gốc 500. Trong bước S604, vì mô hình âm thanh người dùng 500 có các tham số trọng số tương ứng với người dùng, đặc trưng âm thanh người dùng (sau đây gọi là đặc trưng âm thanh thứ nhất của người dùng) có thể được thu từ đầu ra của bộ giải mã 520 của mô hình âm thanh người dùng 500. Trong bước S605, bộ phân tích và tổng hợp giọng nói (như thể hiện trong FIG. 7) được sử dụng để biến đổi đặc

trung âm thanh thứ nhất của người dùng thành đoạn hát tổng hợp. Theo đó, đoạn hát duy trì được tạo ra có thể duy trì cao độ của ca sĩ gốc, và đoạn hát duy trì có thể có âm sắc của người dùng.

Vui lòng tham khảo FIG. 7, minh họa sơ đồ khối của mô-đun duy trì hát 320 theo một phương án của sáng chế. Trong một số phương án của sáng chế, mô-đun duy trì việc hát 320 bao gồm bộ phân tích đặc trưng 321 để thu được bất kỳ thông tin âm thanh nào của các MFCC, tần số cơ bản và tính phi chu kỳ từ phân đoạn âm thanh ca sĩ gốc. Bộ phân tích đặc trưng 321 bao gồm một số thuật toán để ước tính thông tin mẫu âm thanh nói trên. Ví dụ, tần số cơ bản có thể thu được bằng các thuật toán DIO, YIN hoặc SWIPE. Trong các cách tiếp cận này, các dạng sóng mà xuất hiện lặp đi lặp lại được suy ra từ đoạn âm thanh, và khoảng thời gian giữa các lần xuất hiện của các dạng sóng được tính toán để thu được thời gian cần thiết cho từng dạng sóng (cụ thể là chu kỳ của dạng sóng), bằng cách đó cao độ (tần số cơ bản) của đoạn âm thanh có thể được tính bằng cách sử dụng các chu kỳ của dạng sóng. Ví dụ, tính phi chu kỳ có thể thu được bằng thuật toán Platinum (suy luận nền tảng bằng cách loại bỏ yếu tố ẩn dưới). Tính phi chu kỳ là phần tần số cực cao của đoạn âm thanh. Các phần tần số cực cao là các phần của đoạn âm thanh với sự thay đổi nhanh chóng. Thuật toán Platinum là để suy ra giá trị thay đổi tối đa của đoạn âm thanh từ đoạn âm thanh, tức là, tính phi chu kỳ. Cách để thu được các MFCC được mô tả như sau. Phân đoạn âm thanh ca sĩ gốc được chia thành các khung và các khung này có dạng cửa sổ, và sau đó thuật toán Biến đổi Fourier (Fourier Transform - FT) được áp dụng cho từng khung. Tiếp theo, các khung đã biến đổi được xếp chồng lên nhau để thu được ảnh phổ. Sau đó, các ngân hàng bộ lọc Mel (Mel-scale filter banks) được sử dụng để biến đổi ảnh phổ thành âm phổ tần số Mel. Và, sau khi xử lý bằng thuật toán và hàm FT nghịch đảo được áp dụng cho âm phổ tần số Mel, có thể thu được các MFCC.

Trong một số phương án, bộ phân tích đặc trưng 321 có thể thu được đường bao âm phổ bằng cách sử dụng, ví dụ, thuật toán CheapTrick.

Như được minh họa trong FIG. 7, các MFCC được đưa vào mô hình âm thanh trộn 500' (bao gồm bộ mã hóa 510 của mô hình âm thanh ca sĩ gốc 500 và bộ giải mã 520 của mô hình âm thanh người dùng 500), và thông qua việc thực hiện các bước từ S602 đến S604, có thể thu được đặc trưng âm thanh thứ nhất của người dùng (trong phương án này là các MFCC). Bộ phân tích và tổng

hợp giọng nói 322 có thể tạo ra đoạn hát tổng hợp theo các MFCC, nhưng các phương án của sáng chế không bị giới hạn tại đây; bộ phân tích và tổng hợp giọng nói 322 có thể tạo ra đoạn hát tổng hợp không chỉ theo các MFCC mà còn theo tần số cơ bản, tính phi chu kỳ (sự phát âm bật hơi trong âm thanh của con người) và đường bao âm phổ, sao cho đoạn hát tổng hợp có thể tự nhiên hơn. Cụ thể, trong phương án này, bộ phân tích và tổng hợp giọng nói lấy các MFCC hoặc các đặc trưng âm thanh người dùng khác (ví dụ: tần số cơ bản, đường bao âm phổ và tính phi chu kỳ) làm đầu vào và bộ phân tích và tổng hợp giọng nói liên tục đưa ra giá trị của dạng sóng của đoạn âm thanh ở mỗi thời điểm (nghĩa là trục X là trục thời gian và trục Y là giá trị âm thanh tại mỗi thời điểm). Đường bao âm phổ có liên quan đến âm sắc, trong khi tần số cơ bản có liên quan đến cao độ. Bộ phân tích và tổng hợp giọng nói 322 và ít nhất một phần của bộ phân tích đặc trưng 321 có thể được thực hiện bằng phần mềm mã nguồn mở “World” hay “Straight”, nhưng các phương án không bị giới hạn tại đây.

Mô hình cao độ được sử dụng trong quy trình điều chỉnh cao độ được giới thiệu trước phần giới thiệu của quy trình điều chỉnh cao độ. Vui lòng tham khảo FIG. 8, minh họa sơ đồ khối của mô hình cao độ 800 theo một phương án của sáng chế. Trong phương án này, mô hình cao độ 800 là mô hình mạng nơ-ron với chế độ học tập có giám sát, và mô hình cao độ 800 bao gồm bộ mã hóa 810, bộ giải mã 820 và PostNet 830. Nói cách khác, trong phương án này, mô hình cao độ 800 có cấu trúc tự động mã hóa. Bộ mã hóa 810 được điều chỉnh để biến đổi đặc trưng âm thanh đầu vào thành vector đặc trưng, trong khi bộ giải mã 820 được điều chỉnh để biến đổi vector đặc trưng thành đặc trưng âm thanh. PostNet 830 được điều chỉnh để thực hiện việc điều chỉnh tối ưu hóa cho đặc trưng âm thanh. Ví dụ, PostNet có thể giảm các vấn đề liên quan đến tiếng ồn của thông tin âm thanh được trích xuất, tiếng nổ siêu thanh (sonic boom) và sự gián đoạn âm thanh, bằng cách đó cải thiện chất lượng của thông tin âm thanh được trích xuất. Sau khi nhiều dữ liệu huấn luyện của người nào đó được đưa vào mô hình cao độ 800, các tham số trọng số (weighting parameters) có thể được đồng quy, và các tham số trọng số được đi kèm với mô hình cao độ để trở thành mô hình cao độ 800 được huấn luyện tương ứng với một người nhất định, trong đó dữ liệu huấn luyện là nhiều đoạn âm thanh của một người nhất định. Ví dụ: nếu nhiều dữ liệu âm thanh liên quan đến người dùng (tức là dữ liệu huấn luyện người dùng) được cung cấp làm dữ liệu huấn luyện, mô hình cao độ 800 có thể được huấn luyện thành mô hình cao độ người dùng 800; ngược lại, nếu nhiều dữ liệu âm thanh của ca sĩ (tức là dữ liệu huấn luyện ca sĩ gốc) được cung cấp làm dữ liệu huấn luyện, mô hình cao độ 800 có thể được huấn luyện thành mô hình cao độ ca sĩ gốc

800. Nói cách khác, mô hình cao độ 800 với các tham số trọng số khác nhau có thể được huấn luyện bằng cách sử dụng dữ liệu huấn luyện liên quan đến người khác. Trong phương án này, liên quan đến cùng một người, dữ liệu huấn luyện để huấn luyện mô hình cao độ 800 có thể khác với dữ liệu huấn luyện để huấn luyện mô hình âm thanh.

Trong một số phương án, như trong FIG. 8, bộ mã hóa 810 bao gồm ba lớp chập 811, bộ giải mã 820 bao gồm một lớp chập 821 và đơn vị truy toán cổng (GRU) 822 (sau đây gọi là đơn vị GRU) và PostNet 830 bao gồm một lớp giải chập 831. Tuy nhiên, cần hiểu rằng các thành phần của bộ mã hóa 810, bộ giải mã 820 và PostNet 830 không giới hạn ở các yếu tố và sự kết hợp nêu trên.

Vui lòng tham khảo FIG. 9 và FIG. 10. FIG. 9 minh họa lưu đồ của quy trình điều chỉnh cao độ theo một phương án của sáng chế. FIG. 10 minh họa sơ đồ khối của mô hình điều chỉnh cao độ 330 theo một phương án của sáng chế. Trong bước S901, đơn vị chuyển đổi 331 được sử dụng để biến đổi đoạn hát đã được trình bày bởi người dùng thành phổ âm thanh (sau đây gọi là phổ âm thanh người dùng). Bộ biến đổi 331 có thể được thực hiện thuật toán biến đổi Fourier hoặc các thuật toán khác có khả năng thực hiện việc biến đổi miền thời gian sang miền tần số. Trong bước S902, bộ biến đổi 331 được sử dụng một lần nữa để biến đổi đoạn hát của ca sĩ gốc thành phổ âm thanh (sau đây gọi là phổ âm thanh ca sĩ gốc). Mặc dù hai bộ biến đổi 331 được minh họa trong FIG. 9, cần hiểu rằng bước S901 và bước S902 có thể được thực hiện bởi một bộ biến đổi 331.

Trong bước S903, phổ âm thanh người dùng được đưa vào bộ mã hóa 810 của mô hình cao độ người dùng 800 (sau đây gọi là bộ mã hóa mô hình cao độ người dùng 810a); vì mô hình cao độ người dùng 800 có các tham số trọng số tương ứng với người dùng, đặc trưng âm thanh người dùng (sau đây gọi là đặc trưng âm thanh thứ hai của người dùng) có thể thu được. Trong phương án này, đặc trưng âm thanh thứ hai của người dùng có thể là một hoặc các sự kết hợp của tần số cơ bản, đường bao âm phổ, tính phi chu kỳ và hệ số âm tần tổng quát Mel (Mel generalized cepstral coefficients).

Trong bước S904, phổ âm thanh ca sĩ gốc được đưa vào bộ mã hóa 810 của mô hình cao độ ca sĩ gốc 800 (sau đây gọi là bộ mã hóa mô hình cao độ ca sĩ gốc 810b); vì mô hình cao độ ca sĩ gốc 800 có các tham số trọng số tương ứng với ca sĩ gốc, tần số cơ bản của ca sĩ gốc (sau đây gọi là tần số cơ bản của ca sĩ gốc) có thể thu được.

Trong một số phương án, thứ tự thực hiện bước S901 và bước S902 có thể được hoán đổi và thứ tự thực hiện bước S903 và bước S904 có thể được hoán đổi. Hơn nữa, thứ tự thực hiện các bước S901 đến S904 có thể được hoán đổi miễn là bước S901 được thực hiện trước bước S903 và miễn là bước S902 được thực hiện trước bước S904.

Trong bước S905, đặc trưng âm thanh thứ hai của người dùng thu được từ bước S903 và tần số cơ bản của ca sĩ gốc thu được trong bước S904 được đưa vào bộ giải mã 820 của mô hình âm thanh người dùng 800 (sau đây gọi là bộ giải mã mô hình cao độ người dùng 820a) để duy trì tông giọng của người dùng và cao độ của ca sĩ gốc (tức là cao độ chính xác).

Trong bước S906, vì mô hình cao độ người dùng 800 có các tham số trọng số tương ứng với người dùng, phổ âm thanh người dùng đã điều chỉnh có thể thu được từ đầu ra của bộ giải mã mô hình cao độ người dùng 820a.

Trong bước S907, bộ phân tích và tổng hợp giọng nói được sử dụng để biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ, trong đó bộ phân tích và tổng hợp giọng nói lấy các MFCC hoặc các đặc trưng âm thanh của người dùng khác (ví dụ: tần số cơ bản, đường bao phổ và tính phi chu kỳ) làm đầu vào và bộ phân tích và tổng hợp giọng nói liên tiếp trích xuất giá trị của dạng sóng của âm thanh tại mỗi thời điểm. Bộ phân tích và tổng hợp giọng nói có thể được thực hiện bởi phần mềm mã nguồn mở “World” hoặc “Straight”, nhưng các phương án không bị giới hạn ở đây. Theo đó, đoạn hát được trình bày bởi người dùng có thể được điều chỉnh để có cao độ chính xác trong khi âm sắc của người dùng vẫn được duy trì.

Trong một số phương án, mô-đun điều chỉnh cao độ 330 có thể được thực hiện bởi thuật toán cộng chồng chéo đồng bộ (PSOLA). Trong thuật toán PSOLA, các dạng sóng xuất hiện lặp đi lặp lại được suy ra từ đoạn âm thanh và các dạng sóng được thêm hoặc bớt để đạt được sự gia tăng tần số hoặc giảm tần số, sao cho có thể điều chỉnh cao độ của đoạn âm thanh.

Trong một số phương án, thiết bị xử lý 120 nhận dạng được nét nhận diện của ca sĩ gốc theo đoạn hát của ca sĩ gốc đã được tách rời và nạp mô hình cao độ ca sĩ gốc 800 tương ứng với ca sĩ gốc theo nét nhận diện của ca sĩ gốc.

Trong một số phương án, thiết bị xử lý 120 nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát của ca sĩ gốc đã được tách rời và nạp mô hình âm thanh ca sĩ gốc 500 tương ứng với ca sĩ gốc theo nét nhận diện của ca sĩ gốc.

Trong một số phương án, thiết bị xử lý 120 thu được nét nhận diện của ca sĩ gốc từ siêu dữ liệu (metadata) của tệp bài hát hoặc thông qua nguồn đầu vào của người dùng, từ đó thiết bị xử lý có thể nạp mô hình âm thanh ca sĩ gốc 500 và mô hình cao độ ca sĩ gốc 800 tương ứng với ca sĩ gốc.

Trong một số phương án, sản phẩm chương trình máy tính 300 có thể không có mô-đun tách nhạc 340 và mô-đun trộn âm thanh 350, và mô-đun xử lý 120 được điều chỉnh để phát nhạc đệm mà không có đoạn hát ca sĩ gốc. Trong điều kiện này, đoạn hát ca sĩ gốc cho mô-đun điều chỉnh cao độ 330 có thể thu được từ tệp bài hát được lưu trữ trong phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời 123 hoặc từ tệp đoạn hát đã được xử lý khác, và giai đoạn cần trình bày có thể thu được bằng dấu thời gian (time stamping).

Theo đó, dựa trên một hoặc một số phương án của sáng chế, bằng cách dò tìm xem người dùng có hát trong giai đoạn cần hát hay không nhằm xác định quy trình duy trì việc hát có cần được thực hiện hay không, việc duy trì việc hát được cung cấp một cách tự động bằng cách sử dụng âm thanh của chính người dùng khi người dùng quên lời bài hát hoặc ngừng hát vì một số lý do. Bằng cách áp dụng quy trình duy trì việc hát, đoạn hát duy trì được tạo ra có thể duy trì cao độ của ca sĩ gốc; hơn nữa, vì mô hình âm thanh có cấu trúc tự động mã hóa, đoạn hát duy trì được tạo ra có thể có âm sắc của người dùng. Hơn nữa, bằng cách dò tìm cao độ của đoạn hát được trình bày bởi người dùng, quy trình điều chỉnh cao độ có thể được thực hiện khi người dùng hát sai cao độ. Bằng cách thực hiện quy trình điều chỉnh cao độ, cao độ của người dùng có thể được điều chỉnh tự động; hơn nữa, vì mô hình cao độ có cấu trúc tự động mã hóa, đoạn hát đã được trình bày có thể mang âm sắc của người dùng.

YÊU CẦU BẢO HỘ

Các điểm yêu cầu bảo hộ bao gồm:

1. Phương pháp hỗ trợ hát, bao gồm:

thu nhận đoạn hát đã được trình bày bằng thiết bị thu nhận âm thanh;

dò tìm xem đoạn hát đã được trình bày có xuất hiện trong giai đoạn cần trình bày hay không bởi thiết bị xử lý, nếu không, thực hiện quy trình duy trì việc hát, trong đó quy trình duy trì việc hát được thực hiện bằng cách đi kèm với mô hình âm thanh bao gồm bộ mã hóa và bộ giải mã, mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình âm thanh ca sĩ gốc, và mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình âm thanh người dùng, quy trình duy trì việc hát bao gồm:

biến đổi phân đoạn âm thanh ca sĩ gốc thành đặc trưng âm thanh ca sĩ gốc;

nhập đặc trưng âm thanh ca sĩ gốc vào bộ mã hóa của mô hình âm thanh ca sĩ gốc;

thu nhận đầu ra của bộ mã hóa của mô hình âm thanh ca sĩ gốc bằng bộ giải mã của mô hình âm thanh người dùng;

thu được đặc trưng âm thanh thứ nhất của người dùng từ đầu ra của bộ giải mã của mô hình âm thanh người dùng; và

biến đổi đặc trưng âm thanh thứ nhất của người dùng thành đoạn hát tổng hợp bằng bộ phân tích và tổng hợp giọng nói; và

trích xuất đoạn hát tổng hợp bằng thiết bị phát.

2. Phương pháp hỗ trợ hát theo điểm 1, trong đó nếu đoạn hát đã được trình bày xuất hiện trong giai đoạn cần trình bày, xác định xem cao độ được trình bày của đoạn hát đã được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, bằng thiết bị xử lý, nếu không, thực hiện quy trình điều chỉnh cao độ cho đoạn hát đã được trình bày, bằng thiết bị xử lý.

3. Phương pháp hỗ trợ hát theo điểm 2, trong đó quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa thứ hai và bộ giải mã thứ hai, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng;

biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

nhập phổ âm thanh người dùng vào bộ mã hóa thứ hai của mô hình cao độ người dùng để thu được đặc trưng âm thanh thứ hai của người dùng;

nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa thứ hai của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

nhập đặc trưng âm thanh thứ hai của người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã thứ hai của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã điều chỉnh từ đầu ra của bộ giải mã thứ hai của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ.

4. Phương pháp hỗ trợ hát theo điểm 3, bao gồm thêm:

tách rời tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc bằng thiết bị xử lý;

nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc bằng thiết bị xử lý; và

nhập mô hình cao độ ca sĩ gốc tương ứng với ca sĩ gốc bằng thiết bị xử lý, theo nét nhận diện của ca sĩ gốc.

5. Phương pháp hỗ trợ hát theo điểm 1, bao gồm thêm:

tách rời tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc bằng thiết bị xử lý.

6. Phương pháp hỗ trợ hát theo điểm 5, bao gồm thêm:

nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc bằng thiết bị xử lý; và

nạp mô hình âm thanh ca sĩ gốc tương ứng với ca sĩ gốc bằng thiết bị xử lý, theo nét nhận diện của ca sĩ gốc.

7. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi ngắn hạn, bao gồm nhiều lệnh, trong đó khi có ít nhất một bộ xử lý của hệ thống máy tính thực hiện các lệnh, hệ thống máy tính sẽ thực hiện các bước:

dò tìm xem đoạn hát đã được trình bày có xuất hiện trong giai đoạn cần trình bày hay không, nếu không, thực hiện quy trình duy trì việc hát, trong đó quy trình duy trì việc hát được thực hiện bằng cách đi kèm với mô hình âm thanh bao gồm bộ mã hóa và bộ giải mã, mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình âm thanh ca sĩ gốc và mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình âm thanh người dùng, quy trình duy trì việc hát bao gồm:

biến đổi phân đoạn âm thanh ca sĩ gốc thành đặc trưng âm thanh ca sĩ gốc;

nạp đặc trưng âm thanh ca sĩ gốc vào bộ mã hóa của mô hình âm thanh ca sĩ gốc;

thu nhận đầu ra của bộ mã hóa của mô hình âm thanh ca sĩ gốc bằng bộ giải mã của mô hình âm thanh người dùng;

thu được đặc trưng âm thanh thứ nhất của người dùng từ đầu ra của bộ giải mã của mô hình âm thanh người dùng; và

biến đổi đặc trưng âm thanh thứ nhất của người dùng thành đoạn hát tổng hợp bằng bộ phân tích và tổng hợp giọng nói.

8. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời theo điểm 7, trong đó nếu đoạn hát đã được trình bày xuất hiện trong giai đoạn cần trình bày, hệ thống máy tính sẽ thực hiện các lệnh để xác định xem cao độ của đoạn hát đã được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, nếu không, hệ thống máy tính sẽ thực hiện các lệnh để tiến hành quy trình điều chỉnh cao độ cho đoạn hát đã được trình bày.

9. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời theo điểm 8, trong đó quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa thứ hai và bộ giải mã thứ hai, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng;

biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

nhập phổ âm thanh người dùng vào bộ mã hóa thứ hai của mô hình cao độ người dùng để thu được đặc trưng âm thanh thứ hai của người dùng;

nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa thứ hai của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

nhập đặc trưng âm thanh thứ hai của người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã thứ hai của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã điều chỉnh từ đầu ra của bộ giải mã thứ hai của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ.

10. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời theo điểm 9, trong đó các bước được thực hiện bởi hệ thống máy tính bao gồm thêm:

tách rời tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc;

nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc; và

nạp mô hình cao độ ca sĩ gốc tương ứng với ca sĩ gốc, theo nét nhận diện của ca sĩ gốc.

11. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời theo điểm 7, trong đó các bước được thực hiện bởi hệ thống máy tính bao gồm thêm:

tách rời tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc.

12. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời theo điểm 11, trong đó các bước được thực hiện bởi hệ thống máy tính bao gồm thêm:

nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc; và

nạp mô hình âm thanh ca sĩ gốc tương ứng với ca sĩ gốc, theo nét nhận diện của ca sĩ gốc.

13. Hệ thống hỗ trợ hát, bao gồm:

thiết bị thu nhận âm thanh, được điều chỉnh để thu nhận đoạn hát đã được trình bày của người dùng;

thiết bị xử lý, được điều chỉnh để dò tìm xem đoạn hát đã được trình bày có xuất hiện trong giai đoạn cần trình bày hay không, nếu không, thiết bị xử lý thực hiện quy trình duy trì việc hát, trong đó quy trình duy trì việc hát được thực hiện bằng cách đi kèm với mô hình âm thanh bao gồm bộ mã hóa và bộ giải mã, mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình âm thanh ca sĩ gốc và mô hình âm thanh được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình âm thanh người dùng, quy trình duy trì việc hát bao gồm:

biến đổi phân đoạn âm thanh ca sĩ gốc thành đặc trưng âm thanh ca sĩ gốc;

nhập đặc trưng âm thanh ca sĩ gốc vào bộ mã hóa của mô hình âm thanh ca sĩ gốc;

thu nhận đầu ra của bộ mã hóa của mô hình âm thanh ca sĩ gốc bằng bộ giải mã của mô hình âm thanh người dùng;

thu được đặc trưng âm thanh thứ nhất của người dùng từ đầu ra của bộ giải mã của mô hình âm thanh người dùng; và

biến đổi đặc trưng âm thanh thứ nhất của người dùng thành đoạn hát tổng hợp bằng bộ phân tích và tổng hợp giọng nói; và

thiết bị phát được điều chỉnh để trích xuất đoạn hát tổng hợp.

14. Hệ thống hỗ trợ hát theo điểm 13, trong đó nếu đoạn hát đã được trình bày xuất hiện trong giai đoạn cần trình bày, thiết bị xử lý xác định xem cao độ của đoạn hát đã được trình bày có nhất quán

với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, nếu không, thiết bị xử lý thực hiện quy trình điều chỉnh cao độ cho đoạn hát đã được trình bày.

15. Hệ thống hỗ trợ hát theo điểm 14, trong đó quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa thứ hai và bộ giải mã thứ hai, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng;

biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

nhập phổ âm thanh người dùng vào bộ mã hóa thứ hai của mô hình cao độ người dùng để thu được đặc trưng âm thanh thứ hai của người dùng;

nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa thứ hai của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

nhập đặc trưng âm thanh thứ hai của người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã thứ hai của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã điều chỉnh từ đầu ra của bộ giải mã thứ hai của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ.

16. Hệ thống hỗ trợ hát theo điểm 15, trong đó thiết bị xử lý tách rời thêm tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc, nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc, và nạp mô hình cao độ ca sĩ gốc tương ứng với ca sĩ gốc theo nét nhận diện của ca sĩ gốc.

17. Hệ thống hỗ trợ hát theo điểm 13, trong đó thiết bị xử lý tách rời thêm tệp bài hát thành nhạc đệm và đoạn hát ca sĩ gốc.

18. Hệ thống hỗ trợ hát theo điểm 17, trong đó thiết bị xử lý nhận dạng nét nhận diện của ca sĩ gốc theo đoạn hát ca sĩ gốc và thiết bị xử lý nạp mô hình âm thanh ca sĩ gốc tương ứng với ca sĩ gốc theo nét nhận diện của ca sĩ gốc.

19. Phương pháp hỗ trợ hát, bao gồm:

thu nhận đoạn hát đã được trình bày bằng thiết bị thu nhận âm thanh;

xác định xem cao độ đã được trình bày của đoạn hát đã được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, nếu không, thực hiện quy trình điều chỉnh cao độ cho đoạn hát đã được trình bày, trong đó quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa và bộ giải mã, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng;

biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

nhập phổ âm thanh người dùng vào bộ mã hóa của mô hình cao độ người dùng để thu được đặc trưng âm thanh của người dùng;

nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

nhập đặc trưng âm thanh người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã được điều chỉnh từ đầu ra của bộ giải mã của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ; và

trích xuất đoạn hát tổng hợp bằng thiết bị phát.

20. Phương tiện lưu trữ dữ liệu đọc được bằng máy tính phi tạm thời, bao gồm nhiều lệnh, trong đó khi có ít nhất một bộ xử lý của hệ thống máy tính thực hiện các lệnh, hệ thống máy tính sẽ thực hiện các bước:

xác định xem cao độ được trình bày của bài hát đã được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát ca sĩ gốc hay không, nếu không, thực hiện quy trình điều chỉnh cao độ cho bài hát đã được trình bày, trong đó quy trình điều chỉnh cao độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa và bộ giải mã, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi bài hát được trình bày thành phổ âm thanh người dùng;

biến đổi bài hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

đưa phổ âm thanh người dùng vào một bộ mã hóa của mô hình cao độ người dùng để thu được đặc trưng âm thanh người dùng;

đưa phổ âm thanh ca sĩ gốc vào bộ mã hóa của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

đưa đặc trưng âm thanh người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã điều chỉnh từ đầu ra của bộ giải mã của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ.

21. Hệ thống hỗ trợ hát, bao gồm:

thiết bị thu nhận âm thanh, được điều chỉnh để thu nhận đoạn hát đã được trình bày của người dùng;

thiết bị xử lý, được điều chỉnh để xác định xem cao độ của đoạn hát được trình bày có nhất quán với cao độ ca sĩ gốc của đoạn hát của ca sĩ gốc hay không, nếu không, thiết bị xử lý sẽ thực hiện quy trình điều chỉnh cao độ cho đoạn hát đã được trình bày, trong đó quy trình điều chỉnh cao

độ được thực hiện bằng cách đi kèm với mô hình cao độ bao gồm bộ mã hóa và bộ giải mã, mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện ca sĩ gốc để thu được mô hình cao độ ca sĩ gốc, và mô hình cao độ được huấn luyện bởi dữ liệu huấn luyện của người dùng để thu được mô hình cao độ người dùng, quy trình điều chỉnh cao độ bao gồm:

biến đổi đoạn hát đã được trình bày thành phổ âm thanh người dùng;

biến đổi đoạn hát ca sĩ gốc thành phổ âm thanh ca sĩ gốc;

nhập phổ âm thanh người dùng vào bộ mã hóa của mô hình cao độ người dùng để thu được đặc trưng âm thanh của người dùng;

nhập phổ âm thanh ca sĩ gốc vào bộ mã hóa của mô hình cao độ ca sĩ gốc để thu được tần số cơ bản ca sĩ gốc;

nhập đặc trưng âm thanh người dùng và tần số cơ bản ca sĩ gốc vào bộ giải mã của mô hình cao độ người dùng;

thu được phổ âm thanh người dùng đã được điều chỉnh từ đầu ra của bộ giải mã của mô hình cao độ người dùng; và

biến đổi phổ âm thanh người dùng đã được điều chỉnh thành đoạn hát đã được trình bày đã điều chỉnh cao độ; và

thiết bị phát được điều chỉnh để trích xuất đoạn hát đã được trình bày đã điều chỉnh cao độ.

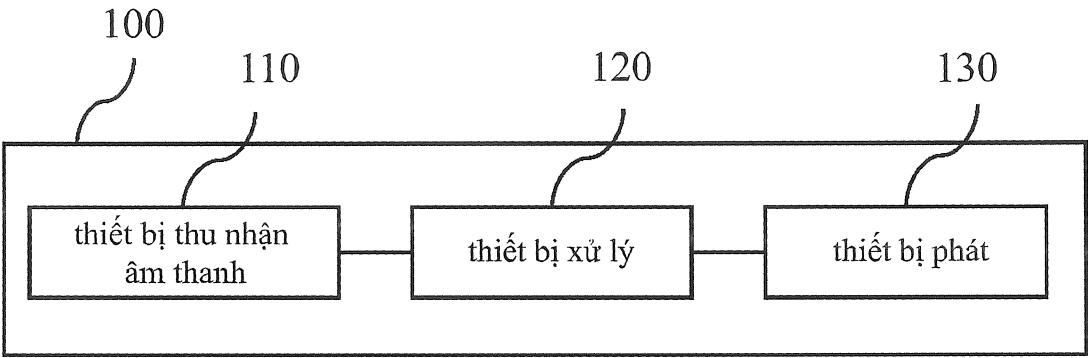


FIG. 1

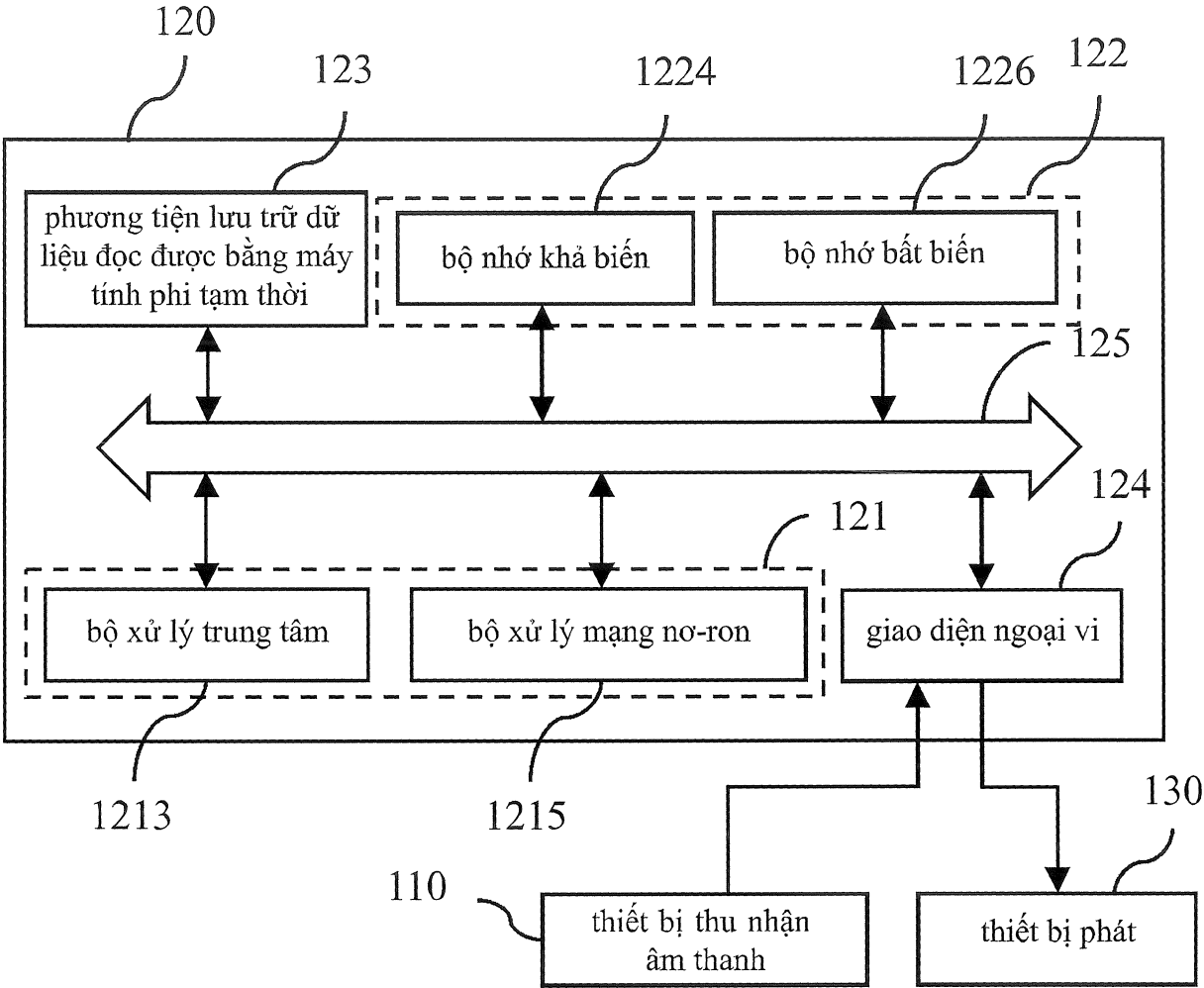


FIG. 2

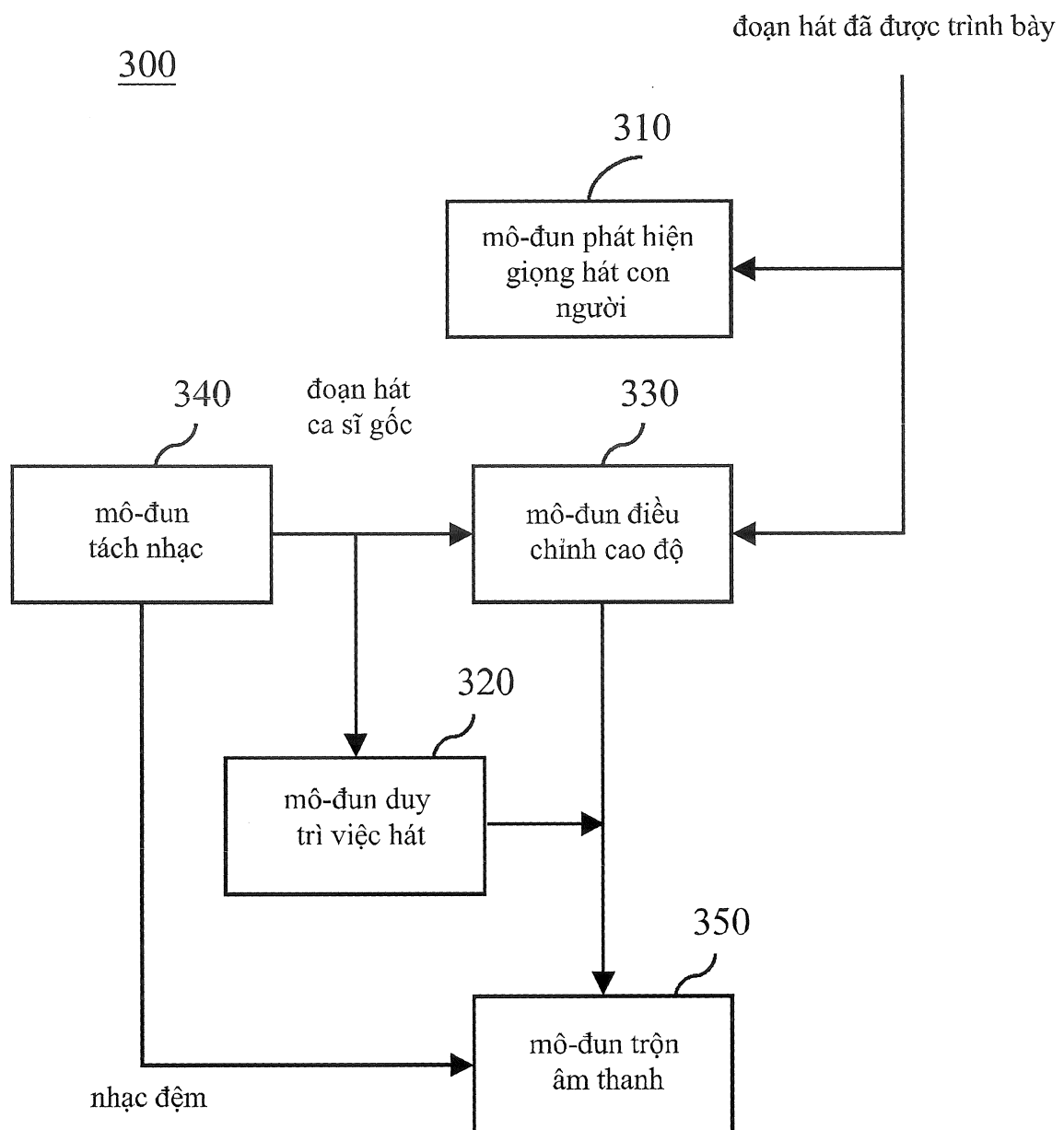


FIG. 3

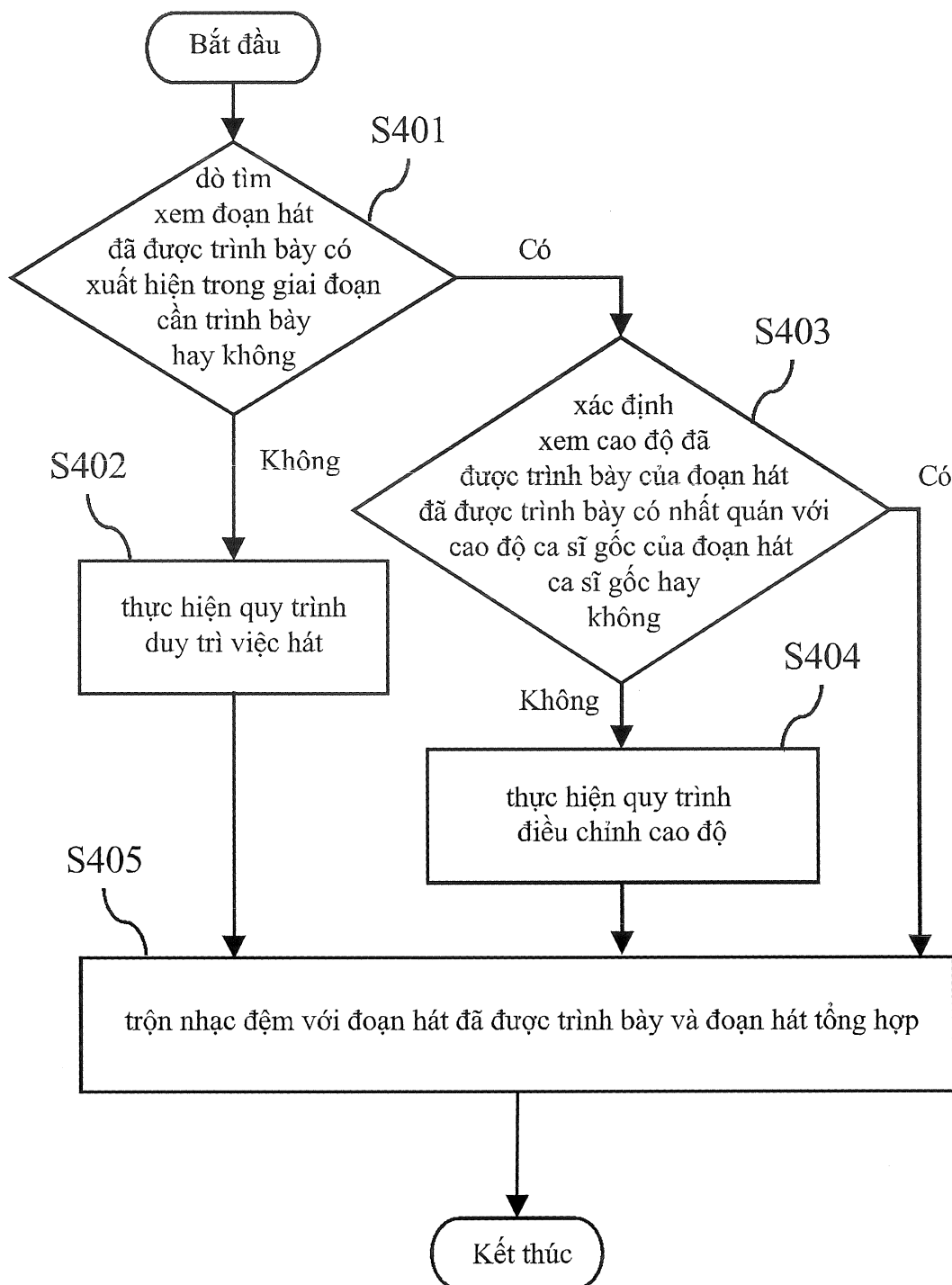


FIG. 4

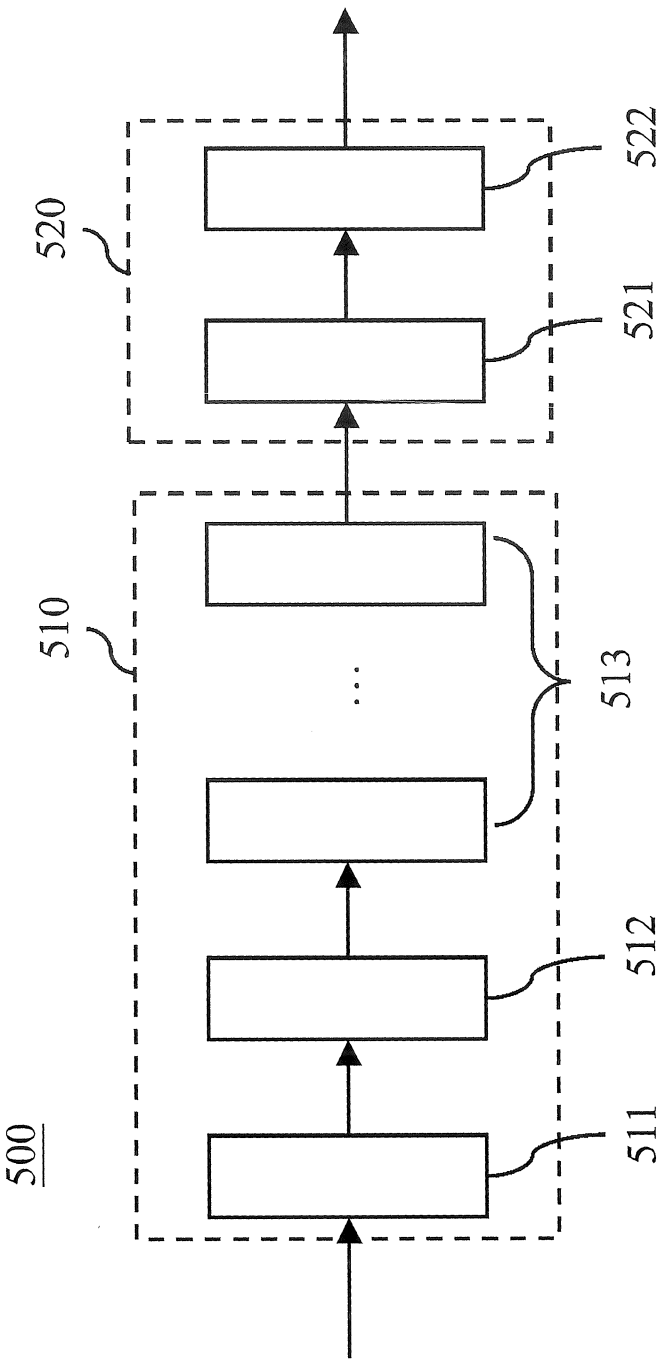


FIG. 5

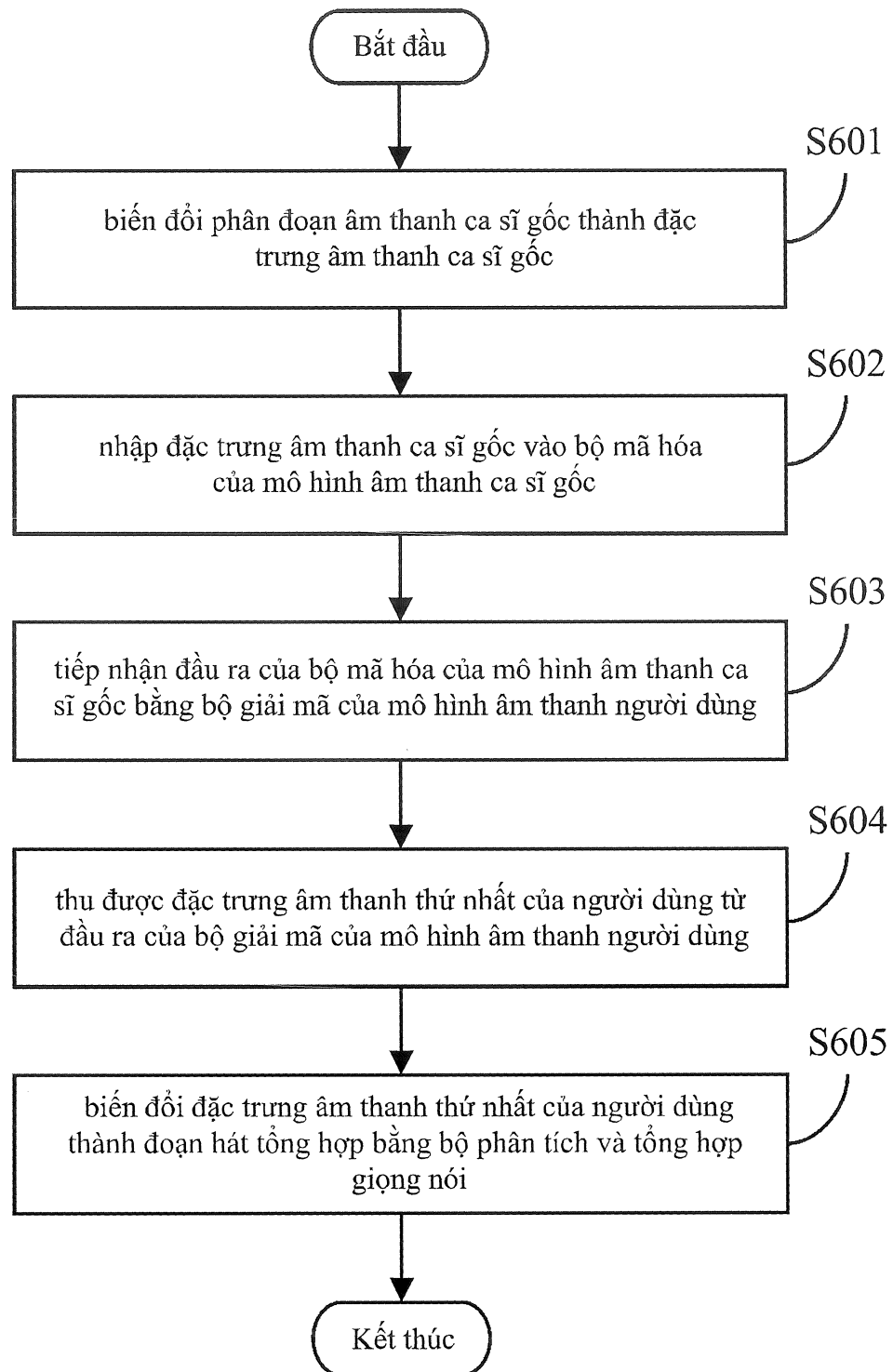


FIG. 6

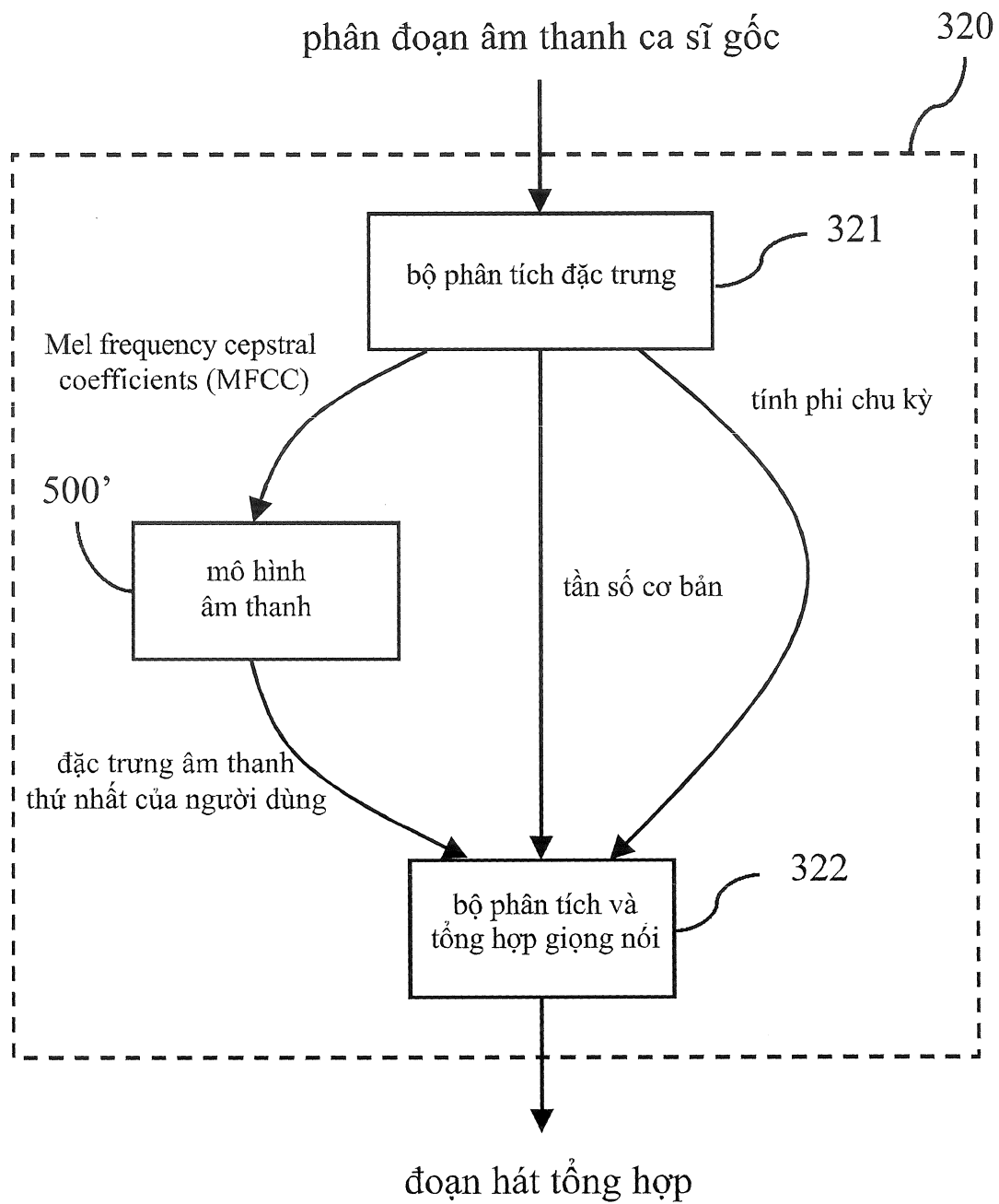


FIG. 7

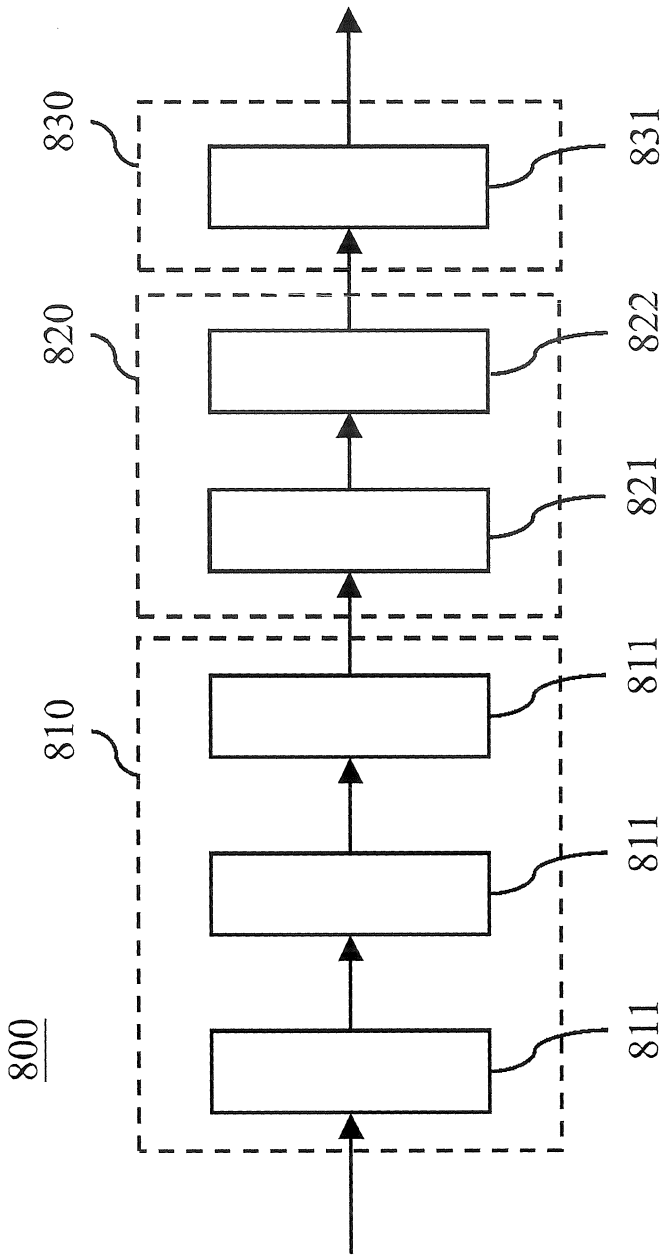


FIG. 8

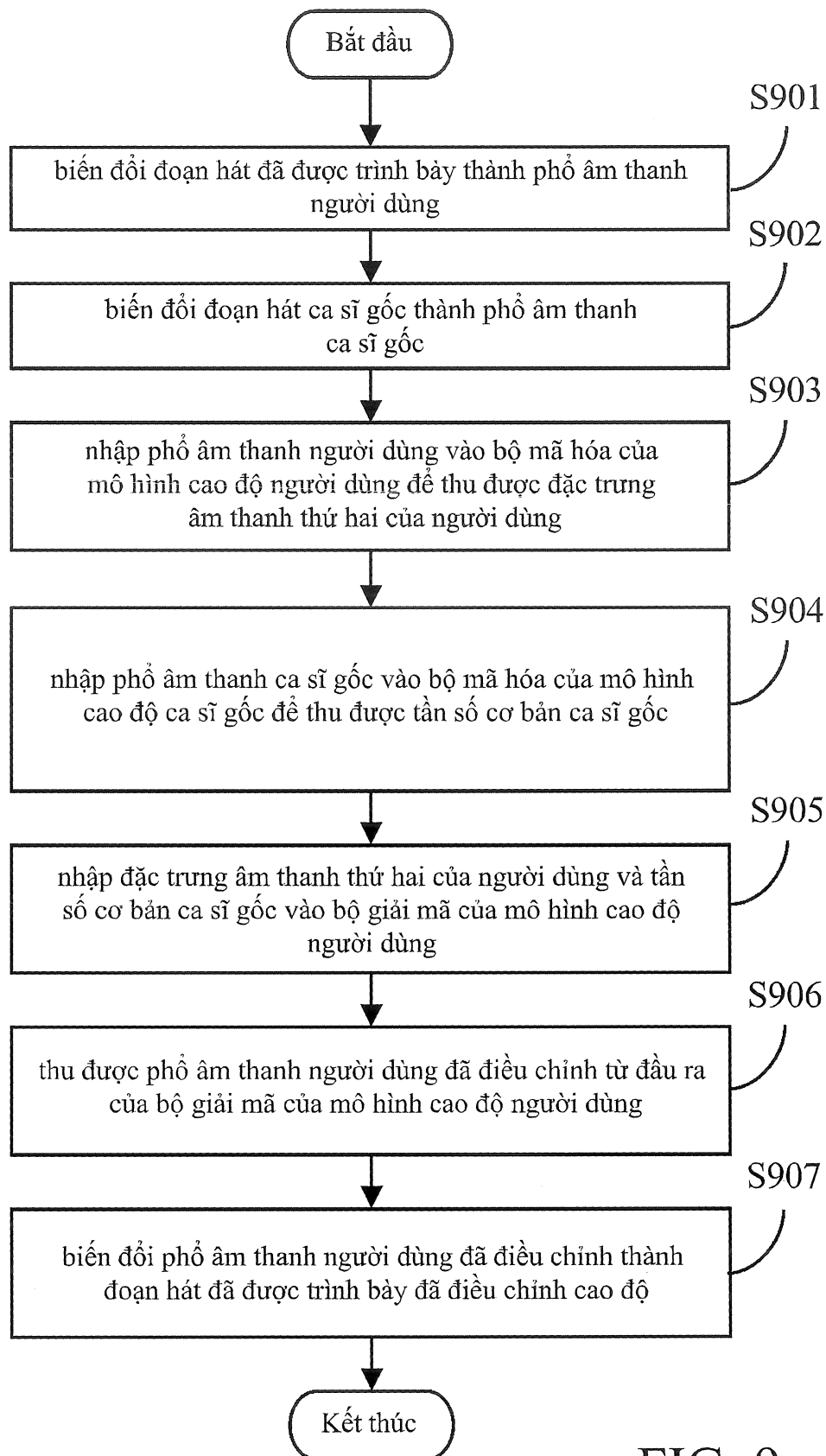


FIG. 9

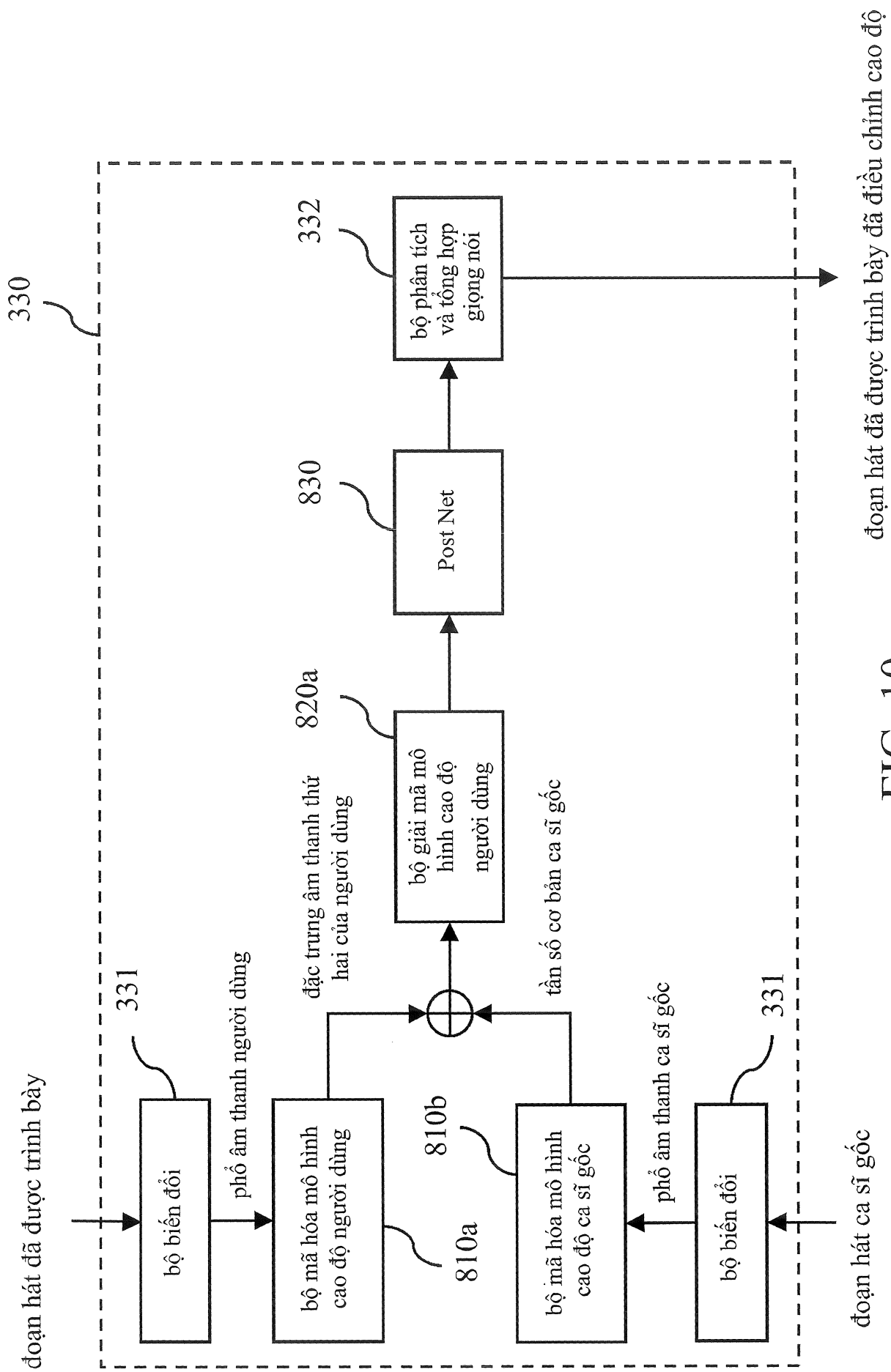


FIG. 10