



(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) 

1-0019796

(51)⁷ **G10L 19/04**

(13) **B**

(21) 1-2016-00789

(22) 26.09.2013

(86) PCT/CN2013/084252 26.09.2013

(87) WO2015/018121 12.02.2015

(30) 201310339218.5 06.08.2013 CN

(45) 25.09.2018 366

(43) 25.05.2016 338

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

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

(72) WANG, Zhe (CN)

(74) Văn phòng luật sư Phạm và Liên danh (PHAM & ASSOCIATES)

(54) **PHƯƠNG PHÁP VÀ THIẾT BỊ PHÂN LOẠI TÍN HIỆU ÂM THANH**

(57) Sáng chế đề xuất phương pháp phân loại tín hiệu âm thanh, trong đó phương pháp gồm: xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số (101); cập nhật, theo việc liệu khung âm thanh là âm nhạc bộ gõ hoặc hoạt động của khung âm thanh lịch sử, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số (102); và phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số (103). Ngoài ra, sáng chế cũng đề cập đến thiết bị phân loại tín hiệu âm thanh.

	Khung thứ N trước	...	Khung thứ hai trước	Khung trước	Khung hiện tại	
--	----------------------	-----	------------------------	----------------	-------------------	--

Lĩnh vực kỹ thuật được đề cập

Sáng chế liên quan đến lĩnh vực các công nghệ xử lý tín hiệu số, và cụ thể là, đến phương pháp và thiết bị phân loại tín hiệu âm thanh.

Tình trạng kỹ thuật của sáng chế

Để giảm các tài nguyên bị chiếm bởi tín hiệu video trong quá trình lưu trữ hoặc truyền, tín hiệu âm thanh được nêu ở đầu truyền và sau đó được truyền đến đầu nhận, và đầu nhận phục hồi tín hiệu âm thanh thông qua giải nén.

Trong ứng dụng xử lý âm thanh, việc phân loại tín hiệu âm thanh là công nghệ quan trọng được áp dụng rộng rãi. Chẳng hạn, trong ứng dụng mã hóa/giải mã âm thanh (codec – encoding/decoding), codec tương đối phổ biến hiện tại là loại lai mã hóa và giải mã. Nói chung, codec này gồm bộ mã hóa (như CELP) dựa trên mô hình tạo thoại và bộ mã hóa dựa trên phép chuyển đổi (như bộ mã hóa dựa trên MDCT). Ở tốc độ bit thấp hoặc trung gian, bộ mã hóa dựa trên mô hình tạo thoại có thể thu được chất lượng mã hóa thoại tương đối tốt, nhưng có chất lượng mã hóa âm nhạc tương đối kém, trong khi bộ mã hóa dựa trên biến đổi có thể thu được chất lượng mã hóa âm nhạc tương đối tốt, nhưng có chất lượng mã hóa thoại tương đối kém. Do vậy, codec lai mã hóa tín hiệu thoại bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại, và mã hóa tín hiệu âm nhạc bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi, nói chung nhờ đó thu thập hiệu quả mã hóa tối ưu. Ở đây, công nghệ lõi là việc phân loại tín hiệu âm thanh, hoặc lựa chọn kiểu mã hóa trong phạm vi sáng chế.

Codec lai cần thu được thông tin kiểu tín hiệu chính xác trước khi codec lai có thể thu được lựa chọn kiểu mã hóa tối ưu. Bộ phân loại tín hiệu âm thanh ở đây cũng có thể được xem gần như bộ phân loại thoại/âm

nhạc. Tốc độ nhận dạng thoại và tốc độ nhận dạng âm nhạc là các chỉ báo quan trọng để đo hiệu năng của bộ phân loại thoại/âm nhạc. Cụ thể đối với tín hiệu âm nhạc, do sự đa dạng/phức tạp của các đặc tính tín hiệu của nó, việc nhận dạng tín hiệu âm nhạc thường khó khăn hơn nhận dạng tín hiệu thoại. Ngoài ra, độ trễ nhận dạng cũng là một trong các chỉ báo quan trọng. Do sự không rõ ràng các đặc tính thoại/âm nhạc trong thời gian ngắn, cần mất khá nhiều thời gian trước khi thoại/âm nhạc có thể được nhận dạng tương đối chính xác. Nói chung, ở đoạn trung gian của cùng loại tín hiệu, độ trễ dài hơn chỉ báo nhận dạng chính xác hơn. Tuy nhiên, ở đoạn chuyển tiếp hai loại tín hiệu, độ trễ nhận dạng dài hơn chỉ báo độ chính xác nhận dạng thấp hơn, vốn đặc biệt nghiêm trọng trong trường hợp mà tín hiệu lai (như thoại có nhạc nền) được đưa vào. Do vậy, việc có cả tốc độ nhận dạng cao và độ trễ nhận dạng thấp là thuộc tính cần của bộ nhận dạng thoại/âm nhạc hiệu năng cao. Ngoài ra, độ ổn định phân loại cũng là thuộc tính quan trọng ảnh hưởng chất lượng mã hóa của bộ mã hóa lai. Nói chung, khi bộ mã hóa lai chuyển đổi giữa các loại bộ mã hóa khác nhau, việc suy giảm chất lượng có thể xuất hiện. Nếu việc chuyển đổi loại tần số xuất hiện trong bộ phân loại trong cùng loại tín hiệu, chất lượng mã hóa bị ảnh hưởng tương đối đáng kể; do vậy, yêu cầu tín hiệu phân loại đầu ra của bộ phân loại cần chính xác và trơn tru. Ngoài ra, trong một số ứng dụng, như thuật toán phân loại trong hệ thống truyền thông, cũng yêu cầu độ phức tạp tính toán và các chi phí phụ thêm lưu trữ của thuật toán phân loại càng thấp càng tốt, để thỏa mãn các yêu cầu thị trường.

Chuẩn ITU-T G.720.1 gồm bộ phân loại thoại/âm nhạc. Bộ phân loại này sử dụng tham số chính: biến thiên dao động phổ tần số `var_flux` làm cơ sở chính để phân loại tín hiệu, và sử dụng hai tham số lớn nhất phổ tần số khác nhau `p1` và `p2` làm cơ sở phụ. Việc phân loại tín hiệu đầu vào theo `var_flux` được hoàn thành trong bộ đệm FIFO (first in first out – vào

trước ra trước) var_flux theo các thống kê cục bộ của var_flux. Quá trình cụ thể được tóm tắt như sau: Trước hết, mật độ dao động phổ tần số được trích rút khỏi mỗi khung âm thanh đầu vào và được đệm trong bộ đệm thứ nhất, và mật độ ở đây được tính toán trong bốn khung mới nhất gồm khung đầu vào hiện tại, hoặc có thể được tính toán bằng cách sử dụng phương pháp khác. Sau đó, biến thiên mật độ của N khung mới nhất gồm khung đầu vào hiện tại được tính toán, để thu được var_flux của khung đầu vào hiện tại, và var_flux được đệm trong bộ đệm thứ hai. Sau đó, số lượng K khung của var_flux lớn hơn ngưỡng thứ nhất trong số M khung mới nhất gồm khung đầu vào hiện tại trong bộ đệm thứ hai được đếm. Nếu tỷ lệ K trên M lớn hơn ngưỡng thứ hai, thì xác định được rằng khung đầu vào hiện tại là khung thoại; ngược lại khung đầu vào hiện tại là khung âm nhạc. Các tham số phụ p1 và p2 chủ yếu được sử dụng để chỉnh sửa phân loại, và cũng được tính toán cho mỗi khung âm thanh đầu vào hiện tại. Khi p1 và/hoặc p2 lớn hơn ngưỡng thứ ba và/hoặc ngưỡng thứ tư, xác định được ngay rằng khung âm thanh đầu vào hiện tại là khung âm nhạc.

Các nhược điểm của bộ phân loại thoại/âm nhạc này là như sau: một mặt, tốc độ nhận dạng tuyệt đối cho âm nhạc vẫn cần được cải thiện, và mặt khác, do các ứng dụng đích của bộ phân loại không dành riêng cho trường hợp ứng dụng của tín hiệu lai, cũng vẫn còn có chỗ cải thiện ở hiệu năng nhận dạng cho tín hiệu lai.

Nhiều bộ phân loại thoại/âm nhạc hiện tại được thiết kế dựa trên nguyên lý nhận dạng chế độ. Loại bộ phân loại này thường lấy nhiều (một tá đến vài tá) tham số đặc trưng từ khung âm thanh đầu vào, và cấp các tham số này vào bộ phân loại dựa trên mô hình lai Gaussian, hoặc mạng nơron, hoặc phương pháp phân loại kinh điển khác để thực hiện phân loại.

Loại bộ phân loại này có nền tảng lý thuyết tương đối chắc chắn,

nhưng nói chung có độ phức tạp lưu trữ hoặc tính toán tương đối cao, và do vậy, chi phí triển khai tương đối cao.

Bản chất kỹ thuật của sáng chế

Mục đích của các phương án thực hiện sáng chế là đề xuất phương pháp và thiết bị phân loại tín hiệu âm thanh, để giảm độ phức tạp phân loại tín hiệu trong khi đảm bảo tốc độ nhận dạng phân loại tín hiệu âm thanh lại.

Theo khía cạnh thứ nhất, phương pháp phân loại tín hiệu âm thanh được đề xuất, trong đó phương pháp gồm:

xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh;

cập nhật, theo việc liệu khung âm thanh là âm nhạc bộ gõ, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số; và

phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số.

Theo cách thức triển khai khả thi thứ nhất, bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số gồm:

nếu khung âm thanh hiện tại là khung hoạt động, lưu trữ dao động phổ tần số của khung âm thanh hiện tại trong bộ nhớ dao động phổ tần số.

Theo cách thức triển khai khả thi thứ hai, bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số

trong bộ nhớ dao động phổ tần số gồm:

nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, lưu trữ dao động phổ tần số của khung âm thanh hiện tại trong bộ nhớ dao động phổ tần số.

Theo cách thức triển khai khả thi thứ ba, bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số gồm:

nếu khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, lưu trữ dao động phổ tần số của khung âm thanh hiện tại trong bộ nhớ dao động phổ tần số.

Dựa vào khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ nhất, theo cách thức triển khai khả thi thứ tư, việc cập nhật, theo việc liệu khung âm thanh hiện tại là âm nhạc bộ gõ, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số gồm:

nếu khung âm thanh hiện tại thuộc âm nhạc bộ gõ, chỉnh sửa các giá trị của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số.

Dựa vào khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ tư của khía cạnh thứ nhất, theo cách thức triển khai khả thi thứ năm, việc phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số gồm:

thu thập giá trị trung bình của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số; và

khi giá trị trung bình thu được của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Dựa vào khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ nhất hoặc cách thức triển khai khả thi thứ tư của khía cạnh thứ nhất, theo cách thức triển khai khả thi thứ sáu, phương pháp phân loại tín hiệu âm thanh còn gồm:

thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kề, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; và độ dốc năng lượng dự báo tuyến tính ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng; và

xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự báo tuyến tính trong các bộ nhớ,

trong đó việc phân loại khung âm thanh theo các thống kê của một phần hoặc tất cả dữ liệu của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số gồm:

thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ

tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thể của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Theo khía cạnh thứ hai, thiết bị phân loại tín hiệu âm thanh được đề xuất, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và gồm:

khởi xác định lưu trữ, được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập và lưu trữ dao động phổ tần số của khung âm thanh hiện tại, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh;

bộ nhớ, được cấu hình để lưu trữ dao động phổ tần số khi khởi xác định lưu trữ xuất ra kết quả rằng dao động phổ tần số cần được lưu trữ;

khởi cập nhật, được cấu hình để cập nhật, theo việc liệu khung âm thanh là âm nhạc bộ gõ, các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khởi phân loại, được cấu hình để phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong

bộ nhớ.

Theo cách thức triển khai khả thi thứ nhất, khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Theo cách thức triển khai khả thi thứ hai, khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Theo cách thức triển khai khả thi thứ ba, khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Dựa vào khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ hai, theo cách thức triển khai khả thi thứ tư, khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại thuộc âm nhạc bộ gõ, chỉnh sửa các giá trị của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số.

Dựa vào khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ tư của khía cạnh thứ hai, theo cách thức triển khai khả thi thứ năm, khối phân loại gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của một

phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khối xác định, được cấu hình để so sánh giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số với điều kiện phân loại âm nhạc; và khi giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Dựa vào khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ nhất của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ hai của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ ba của khía cạnh thứ hai hoặc cách thức triển khai khả thi thứ tư của khía cạnh thứ hai, theo cách thức triển khai khả thi thứ sáu, thiết bị phân loại tín hiệu âm thanh còn gồm:

khối thu thập tham số, được cấu hình để thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, tham số thanh âm, và độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kế, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; tham số thanh âm ký hiệu mức độ tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch; và độ dốc năng lượng dự báo tuyến tính ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng; trong đó

khối xác định lưu trữ còn được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số hay không, mức độ tương quan phổ tần số, và độ

đốc năng lượng dự dự báo tuyến tính trong các bộ nhớ;

khối bộ nhớ còn được cấu hình để: khi khối xác định lưu trữ xuất ra kết quả mà giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ đốc năng lượng dự dự báo tuyến tính cần được lưu trữ, lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ đốc năng lượng dự dự báo tuyến tính; và

khối phân loại được cấu hình cụ thể để thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ đốc năng lượng dự dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng.

Dựa vào cách thức triển khai khả thi thứ sáu của khía cạnh thứ hai, theo cách thức triển khai khả thi thứ bảy, khối phân loại gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ đốc năng lượng dự dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khối xác định, được cấu hình để: khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn

ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Theo các phương án thực hiện sáng chế, tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các dao động phổ tần số; do vậy, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp. Ngoài ra, các dao động phổ tần số được điều chỉnh khi xem xét các yếu tố như độ hoạt động tiếng nói và âm nhạc bộ gõ; do vậy, sáng chế có tốc độ nhận dạng cao hơn đối với tín hiệu âm nhạc, và thích hợp cho việc phân loại tín hiệu âm thanh lai.

Mô tả vắn tắt các hình vẽ

Để mô tả các giải pháp kỹ thuật theo các phương án thực hiện sáng chế hoặc theo giải pháp kỹ thuật đã biết rõ ràng hơn, phần dưới đây giới thiệu vắn tắt các hình vẽ đi kèm cần để mô tả các phương án thực hiện hoặc giải pháp kỹ thuật đã biết. Rõ ràng, các hình vẽ đi kèm trong phần mô tả dưới đây chỉ thể hiện một số phương án thực hiện sáng chế, và người có kiến thức trung bình trong lĩnh vực vẫn có thể duy ra các hình vẽ khác từ các hình vẽ đi kèm này mà không cần nỗ lực sáng tạo.

Fig.1 là sơ đồ của phân chia tín hiệu âm thanh thành các khung;

Fig.2 là lưu đồ của phương pháp phân loại tín hiệu âm thanh theo phương án thực hiện sáng chế;

Fig.3 là lưu đồ thu thập dao động phổ tần số theo phương án thực hiện sáng chế;

Fig.4 là lưu đồ của phương pháp phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.5 là lưu đồ của phương pháp phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.6 là lưu đồ của phương pháp phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.7 đến Fig.10 là các lưu đồ phân loại cụ thể của việc phân loại tín hiệu âm thanh theo sáng chế;

Fig.11 là lưu đồ của phương pháp phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.12 là lưu đồ phân loại cụ thể của việc phân loại tín hiệu âm thanh theo sáng chế;

Fig.13 là sơ đồ cấu trúc của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện sáng chế;

Fig.14 là sơ đồ cấu trúc của khối phân loại theo phương án thực hiện sáng chế;

Fig.15 là sơ đồ cấu trúc của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.16 là sơ đồ cấu trúc của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế;

Fig.17 là sơ đồ cấu trúc của khối phân loại theo phương án thực hiện sáng chế;

Fig.18 là sơ đồ cấu trúc của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế; và

Fig.19 là sơ đồ cấu trúc của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác của sáng chế.

Mô tả chi tiết các phương án thực hiện sáng chế

Phần dưới đây mô tả rõ ràng và đầy đủ các giải pháp kỹ thuật theo các phương án thực hiện sáng chế dựa vào các hình vẽ đi kèm theo các phương án thực hiện sáng chế. Rõ ràng là, các phương án thực hiện được mô tả chỉ là một số chứ không phải tất cả các phương án thực hiện sáng chế. Tất cả các phương án thực hiện khác thu thập bởi người có kiến thức trung bình trong lĩnh vực dựa trên các phương án thực hiện sáng chế mà không cần nỗ lực sáng tạo sẽ nằm trong phạm vi bảo hộ của sáng chế.

Trong lĩnh vực xử lý tín hiệu số, các codec âm thanh và các codec video được áp dụng rộng rãi trong các thiết bị điện tử khác nhau, chẳng hạn, điện thoại di động, thiết bị không dây, PDA (personal digital assistant – thiết bị hỗ trợ cá nhân số), máy tích xách tay, bộ tiếp nhận/điều hướng GPS (Global positioning system – hệ thống định vị toàn cầu), camera, máy phát âm thanh/video, video camera, máy ghi video, và thiết bị giám sát. Nói chung, loại thiết bị điện tử này gồm bộ mã hóa âm thanh hoặc bộ giải mã âm thanh, trong đó bộ mã hóa hoặc giải mã âm thanh có thể được cài đặt trực tiếp bởi mạch số hoặc vi mạch, chẳng hạn, DSP (digital signal processor – bộ xử lý tín hiệu số), hoặc được triển khai bởi mã phần mềm điều khiển bộ xử lý để thực thi quá trình trong mã phần mềm. Trong bộ mã hóa âm thanh, trước hết tín hiệu âm thanh được phân loại, các loại tín hiệu âm thanh khác nhau được mã hóa theo các chế độ mã hóa khác nhau, và sau đó dòng bit được thu thập sau khi mã hóa được truyền đến phía bộ giải mã.

Nói chung, tín hiệu âm thanh được xử lý theo kiểu phân khung, và mỗi khung tín hiệu biểu thị tín hiệu âm thanh của khoảng thời gian xác định. Như được thể hiện trên Fig.1, khung âm thanh hiện được đưa vào và cần được phân loại có thể được gọi là khung âm thanh hiện tại, và khung âm thanh bất kỳ trước khi khung âm thanh hiện tại có thể được gọi là khung âm thanh lịch sử. Theo chuỗi thời gian từ khung âm thanh hiện tại đến các khung âm thanh lịch sử, các khung âm thanh lịch sử có thể lần lượt trở thành khung âm thanh trước đó, khung âm thanh thứ hai trước đó, khung âm thanh thứ ba trước đó, và khung âm thanh thứ N trước đó, trong đó N lớn hơn hoặc bằng 4.

Theo phương án thực hiện này, tín hiệu âm thanh đầu vào là tín hiệu âm thanh băng rộng được lấy mẫu ở 16kHz, và tín hiệu âm thanh đầu vào được phân chia thành các khung bằng cách sử dụng 20ms làm khung, tức là, mỗi khung có 320 điểm lấy mẫu miền thời gian. Trước khi tham số

đặc trưng được trích rút, trước hết khung tín hiệu âm thanh đầu vào được lấy mẫu xuống ở tốc độ lấy mẫu 12,8kHz, tức là, có 256 điểm lấy mẫu trong mỗi khung. Mỗi khung tín hiệu âm thanh đầu vào dưới đây đề cập đến khung tín hiệu âm thanh được thu thập sau khi lấy mẫu xuống.

Như được thể hiện trên Fig.2, phương án thực hiện phương pháp phân loại tín hiệu âm thanh gồm:

S101: Thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào, và xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số hay không, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh.

Việc phân loại tín hiệu âm thanh thường được thực hiện trên cơ sở từng khung, và tham số được trích rút từ mỗi khung tín hiệu âm thanh để thực hiện phân loại, để xác định xem liệu khung tín hiệu âm thanh thuộc khung thoại hoặc khung âm nhạc, và thực hiện mã hóa ở chế độ mã hóa tương ứng. Theo phương án thực hiện, dao động phổ tần số của khung âm thanh hiện tại có thể được thu thập sau khi việc xử lý phân khung được thực hiện trên tín hiệu âm thanh, và sau đó được xác định theo độ hoạt động tiếng nói của khung âm thanh hiện tại liệu có lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số. Theo phương án thực hiện khác, sau khi việc xử lý phân khung được thực hiện trên tín hiệu âm thanh, có thể được xác định theo độ hoạt động tiếng nói của khung âm thanh hiện tại liệu có lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số, và khi dao động phổ tần số cần được lưu trữ, dao động phổ tần số được thu thập và lưu trữ.

Mật độ dao động phổ tần số ký hiệu dao động năng lượng thời gian ngắn hoặc thời gian dài của phổ tần số của tín hiệu, và là giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các

tần số tương ứng của khung âm thanh hiện tại và khung lịch sử trên phổ dải giữa và thấp, trong đó khung lịch sử đề cập đến khung bất kỳ trước khung âm thanh hiện tại. Theo phương án thực hiện, dao động phổ tần số là giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các tần số tương ứng của khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại trên phổ dải giữa và thấp. Theo phương án thực hiện khác, dao động phổ tần số là giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các giá trị lớn nhất phổ tần số tương ứng của khung âm thanh hiện tại và khung lịch sử trên phổ dải giữa và thấp.

Như được thể hiện trên Fig.3, phương án thực hiện thu thập dao động phổ tần số gồm các bước dưới đây:

S1011: Thu thập phổ tần số của khung âm thanh hiện tại.

Theo phương án thực hiện, phổ tần số của khung âm thanh có thể được thu thập trực tiếp; theo phương án thực hiện khác, các phổ tần số, tức là, các phổ năng lượng, của hai khung phụ bất kỳ của khung âm thanh hiện tại được thu thập, và phổ tần số của khung âm thanh hiện tại được thu thập bằng cách sử dụng giá trị trung bình của các phổ tần số của hai khung phụ.

S1012: Thu thập phổ tần số của khung lịch sử của khung âm thanh hiện tại.

Khung lịch sử đề cập đến khung âm thanh bất kỳ trước khung âm thanh hiện tại, và có thể là khung âm thanh thứ ba trước khung âm thanh hiện tại theo phương án thực hiện.

S1013: Tính toán giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các tần số tương ứng của khung âm thanh hiện tại và khung lịch sử trên phổ dải giữa và thấp, để sử dụng giá trị trung bình làm dao động phổ tần số của khung âm thanh hiện tại.

Theo phương án thực hiện, giá trị trung bình của các giá trị tuyệt đối

của các hiệu số giữa năng lượng logarit của tất cả các bin tần số của khung âm thanh hiện tại trên phổ dải giữa và thấp và năng lượng logarit của các bin tần số tương ứng của khung lịch sử trên phổ dải giữa và thấp có thể được tính toán.

Theo phương án thực hiện khác, giá trị trung bình của các giá trị tuyệt đối của các hiệu số giữa năng lượng logarit của các giá trị lớn nhất phổ tần số của khung âm thanh hiện tại trên phổ dải giữa và thấp và năng lượng logarit của các giá trị lớn nhất phổ tần số tương ứng của khung lịch sử trên phổ dải giữa và thấp có thể được tính toán.

Phổ dải giữa và thấp là, chẳng hạn, khoảng phổ tần số từ 0 đến $f_s/4$ hoặc từ 0 đến $f_s/3$.

Ví dụ được sử dụng trong đó tín hiệu âm thanh đầu vào là tín hiệu âm thanh băng rộng được lấy mẫu ở 16kHz và tín hiệu âm thanh đầu vào sử dụng 20ms làm khung, FFT sớm 256 điểm và FFT muộn 256 điểm được thực hiện trên khung âm thanh hiện tại cách 20ms, hai cửa sổ FFT bị chồng lấp 50%, và các phổ tần số (các phổ năng lượng) của hai khung phụ của khung âm thanh hiện tại được thu thập, và lần lượt được đánh dấu như là $C^0(i)$ và $C^1(i)$, $i = 0, 1, \dots, 127$, trong đó $C^x(i)$ ký hiệu phổ tần số của khung phụ thứ x^{th} . Dữ liệu của khung phụ thứ hai của khung trước cần được sử dụng cho FFT của khung phụ thứ nhất của khung âm thanh hiện tại, trong đó

$$C^x(i) = \text{rel}^2(i) + \text{img}^2(i),$$

trong đó $\text{rel}(i)$ và $\text{img}(i)$ lần lượt ký hiệu phần thực và phần ảo của hệ số FFT của bin tần số thứ i^{th} . Phổ tần số $C(i)$ của khung âm thanh hiện tại được thu thập bằng cách tính trung bình các phổ tần số của hai khung phụ, trong đó

$$C(i) = \frac{1}{2}(C^0(i) + C^1(i)).$$

Mật độ dao động phổ tần số của khung âm thanh hiện tại là giá trị

trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các tần số tương ứng của khung âm thanh hiện tại và khung 60ms trước khung âm thanh hiện tại trên phổ dải giữa và thấp theo phương án thực hiện, và khoảng này không thể bằng 60ms theo phương án thực hiện khác, trong đó

$$flux = \frac{1}{42} \sum_{i=0}^{42} [10 \log(C(i)) - 10 \log(C_{-3}(i))]$$

trong đó $C_{-3}(i)$ ký hiệu phổ tần số của khung lịch sử thứ ba trước khi khung âm thanh hiện tại, tức là, khung lịch sử 60ms trước khung âm thanh hiện tại khi chiều dài khung bằng 20ms theo phương án thực hiện này. Mỗi biểu mẫu tương ứng $X_n()$ theo sáng chế ký hiệu tham số X của khung lịch sử thứ n của khung âm thanh hiện tại, và chỉ số dưới 0 có thể bị bỏ qua đối với khung âm thanh hiện tại. $\log(.)$ ký hiệu hàm logarit với 10 làm cơ số.

Theo phương án thực hiện khác, mật độ dao động phổ tần số của khung âm thanh hiện tại cũng có thể được thu thập bằng cách sử dụng phương pháp dưới đây, tức là, mật độ dao động phổ tần số là giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các giá trị lớn nhất phổ tần số tương ứng của khung âm thanh hiện tại và khung 60ms trước khung âm thanh hiện tại trên phổ dải giữa và thấp, trong đó

$$flux = \frac{1}{K} \sum_{i=0}^K [10 \log(P(i)) - 10 \log(P_{-3}(i))]$$

trong đó $P(i)$ ký hiệu năng lượng của giá trị lớn nhất cục bộ thứ i^{th} của phổ tần số của khung âm thanh hiện tại, bin tần số ở đó giá trị lớn nhất cục bộ được đặt là bin tần số, trên phổ tần số, có năng lượng lớn hơn năng lượng của bin tần số cao hơn liền kề và năng lượng của bin tần số thấp hơn liền kề, và K ký hiệu số lượng các giá trị lớn nhất cục bộ trên phổ dải giữa và thấp.

Bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số có thể được triển khai theo nhiều cách:

Theo phương án thực hiện, nếu tham số hoạt động tiếng nói của khung âm thanh ký hiệu rằng khung âm thanh là khung hoạt động, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ.

Theo phương án thực hiện khác, được xác định, theo độ hoạt động tiếng nói của khung âm thanh và liệu khung âm thanh có phải là tấn công năng lượng hay không, liệu có lưu trữ dao động phổ tần số trong bộ nhớ hay không. Nếu tham số hoạt động tiếng nói của khung âm thanh ký hiệu rằng khung âm thanh là khung hoạt động, và tham số ký hiệu xem liệu khung âm thanh có phải là tấn công năng lượng ký hiệu mà khung âm thanh không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ. Theo phương án thực hiện khác, nếu khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ. Chẳng hạn, nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại, khung âm thanh trước và khung âm thanh thứ hai trước không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ.

Cờ hoạt động tiếng nói vad_flag ký hiệu liệu tín hiệu đầu vào hiện tại là tín hiệu nền trước hoạt động (thoại, âm nhạc, hoặc tương tự) hoặc tín hiệu nền sau im lặng (như nhiều nền sau hoặc câm) của tín hiệu nền trước, và được thu thập bởi VAD (voice activity detector – bộ dò hoạt động

tiếng nói). $\text{vad_flag} = 1$ ký hiệu rằng khung tín hiệu đầu vào là khung hoạt động, tức là, khung tín hiệu nền trước; ngược lại, $\text{vad_flag} = 0$ ký hiệu khung tín hiệu nền sau. Do VAD không thuộc nội dung sáng tạo của sáng chế, nên thuật toán cụ thể của VAD không được mô tả chi tiết ở đây.

Cờ tấn công tiếng nói attack_flag ký hiệu liệu khung âm thanh hiện tại thuộc tấn công năng lượng ở âm nhạc. Khi một số khung lịch sử trước khi khung âm thanh hiện tại chủ yếu là các khung âm nhạc, nếu năng lượng khung của khung âm thanh hiện tại tăng tương đối lớn so với năng lượng của khung lịch sử thứ nhất trước khung âm thanh hiện tại, và tăng tương đối lớn so với năng lượng trung bình của các khung âm thanh nằm trong chu kỳ thời gian trước khung âm thanh hiện tại, và đường bao miền thời gian của khung âm thanh hiện tại cũng tăng tương đối lớn so với đường bao trung bình của các khung âm thanh nằm trong chu kỳ thời gian trước khung âm thanh hiện tại, xem xét được rằng khung âm thanh hiện tại thuộc tấn công năng lượng ở âm nhạc.

Theo độ hoạt động tiếng nói của khung âm thanh hiện tại, dao động phổ tần số của khung âm thanh hiện tại được lưu trữ chỉ khi khung âm thanh hiện tại là khung hoạt động, có thể giảm tỷ lệ đánh giá sai của khung không hoạt động, và cải thiện tốc độ nhận dạng phân loại âm thanh.

Khi các điều kiện dưới đây được thỏa mãn, attack_flag được chọn bằng 1, tức là, ký hiệu rằng khung âm thanh hiện tại là tấn công năng lượng trong một đoạn nhạc:

$$\left\{ \begin{array}{l} \text{etot} - \text{etot}_{-1} > 6 \\ \text{etot} - \text{lp_speec} > 5 \\ \text{mode_mov} > 0,9 \\ \log_max_spl - \text{mov_log_max_spl} > 5 \end{array} \right. ,$$

trong đó etot ký hiệu năng lượng khung logarit của khung âm thanh hiện tại; etot_{-1} ký hiệu năng lượng khung logarit của khung âm thanh

trước đó; lp_speech ký hiệu trung bình di chuyển thời gian dài của năng lượng khung logarit etot; log_max_spl và $mov_log_max_spl$ lần lượt ký hiệu biên độ điểm lấy mẫu logarit lớn nhất miền thời gian của khung âm thanh hiện tại và trung bình di chuyển thời gian dài của biên độ điểm lấy mẫu logarit lớn nhất miền thời gian; và $mode_mov$ ký hiệu trung bình di chuyển thời gian dài của các kết quả phân loại cuối cùng lịch sử khi phân loại tín hiệu.

Ý nghĩa của công thức nêu trên là: khi một số khung lịch sử trước khi khung âm thanh hiện tại chủ yếu là các khung âm nhạc, nếu năng lượng khung của khung âm thanh hiện tại tăng tương đối lớn so với năng lượng của khung lịch sử thứ nhất trước khung âm thanh hiện tại, và tăng tương đối lớn so với năng lượng trung bình của các khung âm thanh nằm trong chu kỳ thời gian trước khung âm thanh hiện tại, và đường bao miền thời gian của khung âm thanh hiện tại cũng tăng tương đối lớn so với đường bao trung bình của các khung âm thanh nằm trong chu kỳ thời gian trước khung âm thanh hiện tại, xem xét được rằng khung âm thanh hiện tại thuộc tần công năng lượng ở âm nhạc.

Năng lượng khung logarit etot được ký hiệu bằng năng lượng tổng dải phụ logarit của khung âm thanh đầu vào:

$$etot = 10 \log \left(\sum_{j=0}^{19} \left[\frac{1}{hb(j) - lb(j) + 1} \cdot \sum_{i=lb(j)}^{hb(j)} C(i) \right] \right),$$

trong đó $hb(j)$ và $lb(j)$ lần lượt ký hiệu biên tần số cao và biên tần số thấp của dải phụ thứ j^{th} trong phổ tần số của khung âm thanh đầu vào; và $C(i)$ ký hiệu phổ tần số của khung âm thanh đầu vào.

Trung bình di chuyển thời gian dài $mov_log_max_spl$ của biên độ điểm lấy mẫu logarit lớn nhất miền thời gian của khung âm thanh hiện tại chỉ được cập nhật trong khung tiếng nói hoạt động:

$$mov_log_max_spl = \begin{cases} 0,95 \cdot mov_log_max_spl_{-1} + 0,05 \cdot log_max_spl & log_max_spl > mov_log_max_spl_{-1} \\ 0,995 \cdot mov_log_max_spl_{-1} + 0,005 \cdot log_max_spl & log_max_spl \leq mov_log_max_spl_{-1} \end{cases}$$

Theo phương án thực hiện, mật độ dao động phổ tần số của khung âm thanh hiện tại được đệm trong bộ đệm lịch sử mật độ FIFO. Theo phương án thực hiện này, chiều dài của bộ đệm lịch sử mật độ bằng 60 (60 khung). Độ hoạt động tiếng nói của khung âm thanh hiện tại và liệu khung âm thanh có phải là tấn công năng lượng hay không được xác định, và khi khung âm thanh hiện tại là khung tín hiệu nền trước và khung âm thanh hiện tại nào và hai khung trước khung âm thanh hiện tại không thuộc tấn công năng lượng của âm nhạc, mật độ dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ.

Trước khi mật độ của khung âm thanh hiện tại được đệm, nó kiểm tra liệu các điều kiện dưới đây được thỏa mãn:

$$\begin{cases} vad_flag \neq 0 \\ attack_flag \neq 1 \\ attack_flag_{-1} \neq 1 \\ attack_flag_{-2} \neq 1, \end{cases}$$

nếu các điều kiện được thỏa mãn, mật độ được đệm; ngược lại mật độ không được đệm.

vad_flag ký hiệu liệu tín hiệu đầu vào hiện tại là tín hiệu nền trước hoạt động hoặc tín hiệu nền sau im lặng của tín hiệu nền trước, và $vad_flag = 0$ ký hiệu khung tín hiệu nền sau; và $attack_flag$ ký hiệu liệu khung âm thanh hiện tại thuộc tấn công năng lượng ở âm nhạc, và $attack_flag = 1$ ký hiệu rằng khung âm thanh hiện tại là tấn công năng lượng trong một đoạn nhạc.

Ý nghĩa của công thức nêu trên là: khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại nào, khung âm thanh trước đó, và khung âm thanh thứ hai trước đó không thuộc tấn công năng lượng.

S102: Cập nhật, theo việc liệu khung âm thanh là âm nhạc bộ gõ hoặc hoạt động của khung âm thanh lịch sử, các dao động phổ tần số được lưu

trữ trong bộ nhớ dao động phổ tần số.

Theo phương án thực hiện, nếu tham số ký hiệu xem liệu khung âm thanh thuộc âm nhạc bộ gõ ký hiệu rằng khung âm thanh hiện tại thuộc âm nhạc bộ gõ, các giá trị của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số được điều chỉnh, và các giá trị dao động phổ tần số hợp lệ trong bộ nhớ dao động phổ tần số được điều chỉnh thành giá trị nhỏ hơn hoặc bằng ngưỡng âm nhạc, trong đó khi dao động phổ tần số của khung âm thanh nhỏ hơn ngưỡng âm nhạc, âm thanh được phân loại như là khung âm nhạc. Theo phương án thực hiện, các giá trị dao động phổ tần số hợp lệ được đặt lại bằng 5. Tức là, khi âm thanh bộ gõ flag percus_flag được chọn bằng 1, tất cả dữ liệu bộ đệm hợp lệ trong bộ đệm lịch sử mật độ được đặt lại bằng 5. Ở đây, dữ liệu bộ đệm hợp lệ tương đương giá trị dao động phổ tần số hợp lệ. Nói chung, giá trị dao động phổ tần số của khung âm nhạc tương đối nhỏ, trong khi giá trị dao động phổ tần số của khung thoại tương đối lớn. Khi khung âm thanh thuộc âm nhạc bộ gõ, các giá trị dao động phổ tần số hợp lệ được điều chỉnh thành giá trị nhỏ hơn hoặc bằng ngưỡng âm nhạc, có thể cải thiện xác suất mà khung âm thanh được phân loại như là khung âm nhạc, nhờ đó cải thiện độ chính xác phân loại tín hiệu âm thanh.

Theo phương án thực hiện khác, các dao động phổ tần số trong bộ nhớ được cập nhật theo độ hoạt động của khung lịch sử của khung âm thanh hiện tại. Một cách cụ thể, theo phương án thực hiện, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và khung âm thanh trước đó là khung không hoạt động, dữ liệu của các dao động phổ tần số khác được lưu trữ trong bộ nhớ dao động phổ tần số ngoại trừ dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa thành dữ liệu không hiệu dụng. Khi khung âm thanh trước đó là khung không hoạt động trong khi khung âm thanh hiện tại là khung hoạt động, độ hoạt động tiếng nói của khung âm thanh

hiện tại khác với độ hoạt động của khung lịch sử, dao động phổ tần số của khung lịch sử bị mất hiệu lực, có thể giảm tác động của khung lịch sử khi phân loại âm thanh, nhờ đó cải thiện độ chính xác phân loại tín hiệu âm thanh.

Theo phương án thực hiện khác, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa thành giá trị thứ nhất. Giá trị thứ nhất có thể là ngưỡng thoại, trong đó khi dao động phổ tần số của khung âm thanh lớn hơn ngưỡng thoại, âm thanh được phân loại như là khung thoại. Theo phương án thực hiện khác, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và kết quả phân loại của khung lịch sử là khung âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa thành giá trị thứ hai, trong đó giá trị thứ hai lớn hơn giá trị thứ nhất.

Nếu mật độ của khung âm thanh hiện tại được đếm, và khung âm thanh trước đó là khung không hoạt động ($\text{vad_flag} = 0$), ngoại trừ mật độ khung âm thanh hiện tại mới được đếm trong bộ đếm lịch sử mật độ, dữ liệu còn lại trong bộ đếm lịch sử mật độ đều được đặt lại bằng -1 (tương đương việc dữ liệu bị mất hiệu lực).

Nếu mật độ được đếm trong bộ đếm lịch sử mật độ, và ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động ($\text{vad_flag} = 1$), mật độ khung âm thanh hiện tại vừa được đếm trong bộ đếm lịch sử mật độ được chỉnh sửa thành 16; tức là, kiểm tra được liệu các điều kiện dưới đây có được thỏa mãn hay không:

$$\begin{cases} vad_flag_{-1} = 1 \\ vad_flag_{-2} = 1 \\ vad_flag_{-3} = 1, \end{cases}$$

nếu các điều kiện không được thỏa mãn, mật độ khung âm thanh hiện tại vừa được đếm trong bộ đếm lịch sử mật độ được chỉnh sửa thành 16; và

nếu ba khung liên tiếp trước khung âm thanh hiện tại đều là các khung hoạt động ($vad_flag = 1$), được kiểm tra liệu các điều kiện dưới đây được thỏa mãn:

$$\begin{cases} mode_mov > 0,9 \\ flux > 20 \end{cases};$$

nếu các điều kiện được thỏa mãn, mật độ khung âm thanh hiện tại vừa được đếm trong bộ đếm lịch sử mật độ được chỉnh sửa thành 20; ngược lại không thực hiện hoạt động,

trong đó $mode_mov$ ký hiệu trung bình di chuyển thời gian dài của các kết quả phân loại cuối cùng lịch sử khi phân loại tín hiệu; $mode_mov > 0,9$ ký hiệu rằng tín hiệu là trong tín hiệu âm nhạc, và mật độ bị giới hạn theo kết quả phân loại lịch sử của tín hiệu âm thanh, để giảm xác suất đặc tính thoại xuất hiện trong mật độ và cải thiện độ ổn định xác định phân loại.

Khi ba khung lịch sử liên tiếp trước khung âm thanh hiện tại đều là các khung không hoạt động, và khung âm thanh hiện tại là khung hoạt động, hoặc khi ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, và khung âm thanh hiện tại là khung hoạt động, việc phân loại ở pha khởi tạo. Theo phương án thực hiện, để khiến kết quả phân loại nghiêng về thoại (âm nhạc), dao động phổ tần số của khung âm thanh hiện tại có thể được chỉnh sửa thành ngưỡng thoại (âm nhạc) hoặc giá trị gần với ngưỡng thoại (âm nhạc). Theo phương án thực hiện khác, nếu tín hiệu trước tín hiệu hiện tại là tín hiệu thoại (âm nhạc),

dao động phổ tần số của khung âm thanh hiện tại có thể được chỉnh sửa thành ngưỡng thoại (âm nhạc) hoặc giá trị gần với ngưỡng thoại (âm nhạc), để cải thiện độ ổn định xác định phân loại. Theo phương án thực hiện khác, để khiến kết quả phân loại nghiêng về phía âm nhạc, dao động phổ tần số có thể bị giới hạn, tức là, dao động phổ tần số của khung âm thanh hiện tại có thể được chỉnh sửa, sao cho dao động phổ tần số không lớn hơn ngưỡng, để giảm xác suất xác định rằng dao động phổ tần số là đặc tính thoại.

Cờ âm thanh bộ gõ `percus_flag` ký hiệu liệu âm thanh bộ gõ tồn tại trong khung âm thanh. Việc `percus_flag` được chọn bằng 1 ký hiệu rằng âm thanh bộ gõ được dò thấy, và việc `percus_flag` được chọn bằng 0 ký hiệu việc âm thanh bộ gõ không được dò thấy.

Khi phần nhô năng lượng tương đối rõ rệt xuất hiện trong tín hiệu hiện tại (tức là, một số khung tín hiệu mới nhất gồm khung âm thanh hiện tại và một số khung lịch sử của khung âm thanh hiện tại) trong cả thời gian ngắn lẫn thời gian dài, và tín hiệu hiện tại không có đặc tính âm thanh có thanh âm rõ ràng, nếu một số khung lịch sử trước khi khung âm thanh hiện tại chủ yếu là các khung âm nhạc, được xem xét rằng tín hiệu hiện tại là đoạn âm nhạc bộ gõ; ngược lại, ngoài ra, nếu không khung phụ nào của tín hiệu hiện tại có đặc tính âm thanh có thanh âm rõ ràng và độ tăng tương đối rõ ràng cũng xuất hiện trong đường bao miền thời gian của tín hiệu hiện tại so với độ trung bình thời gian dài của đường bao miền thời gian, cũng được xem xét rằng tín hiệu hiện tại là đoạn âm nhạc bộ gõ.

Cờ âm thanh bộ gõ `percus_flag` được thu thập bằng cách thực hiện bước dưới đây:

Năng lượng khung logarit etot của khung âm thanh đầu vào trước hết được thu thập, trong đó năng lượng khung logarit etot được ký hiệu bằng năng lượng tổng dải phụ logarit của khung âm thanh đầu vào:

$$etot = 10 \log \left(\sum_{j=0}^{19} \left[\frac{1}{hb(j) - lb(j) + 1} \cdot \sum_{i=lb(j)}^{hb(j)} C(i) \right] \right),$$

trong đó $hb(j)$ và $lb(j)$ lần lượt ký hiệu biên tần số cao và biên tần số thấp của dải phụ thứ j^{th} trong phổ tần số của khung đầu vào, và $C(i)$ ký hiệu phổ tần số của khung âm thanh đầu vào.

Khi các điều kiện dưới đây được thỏa mãn, `percus_flag` được chọn bằng 1; ngược lại `percus_flag` được chọn bằng 0:

$$\left\{ \begin{array}{l} etot_{-2} - etot_{-3} > 6 \\ etot_{-2} - etot_{-1} > 0 \\ etot_{-2} - etot > 3 \\ etot_{-1} - etot > 0 \\ etot_{-2} - lp_speech > 3 \\ 0,5 \cdot voicing_{-1}(1) + 0,25 \cdot voicing(0) + 0,25 \cdot voicing(1) < 0,75 \\ mode_mov > 0,9 \end{array} \right. ,$$

Hoặc

$$\left\{ \begin{array}{l} etot_{-2} - etot_{-3} > 6 \\ etot_{-2} - etot_{-1} > 0 \\ etot_{-2} - etot > 3 \\ etot_{-1} - etot > 0 \\ etot_{-2} - lp_speech > 3 \\ 0,5 \cdot voicing_{-1}(1) + 0,25 \cdot voicing(0) + 0,25 \cdot voicing(1) < 0,75' \\ voicing_{-1}(0) < 0,8 \\ voicing_{-1}(1) < 0,8 \\ voicing(0) < 0,8 \\ log_max_spl_{-2} - mov_log_max_spl_{-2} > 10 \end{array} \right.$$

trong đó $etot$ ký hiệu năng lượng khung logarit của khung âm thanh hiện tại; lp_speech ký hiệu trung bình di chuyển thời gian dài của năng lượng khung logarit $etot$; $voicing(0)$, $voicing_{-1}(0)$, và $voicing_{-1}(1)$ lần lượt ký hiệu các mức độ tương quan pitch vòng mở chuẩn hóa của khung phụ thứ nhất của khung âm thanh đầu vào hiện tại và các khung phụ thứ nhất và thứ hai của khung lịch sử thứ nhất, và tham số thanh âm có thanh âm được thu thập bằng cách phân tích và dự báo tuyến tính, biểu thị mức độ

tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch, và có giá trị giữa 0 và 1; mode_mov ký hiệu trung bình di chuyển thời gian dài của các kết quả phân loại cuối cùng lịch sử khi phân loại tín hiệu; log_max_spl₂ và mov_log_max_spl₂ lần lượt ký hiệu biên độ điểm lấy mẫu logarit lớn nhất miền thời gian của khung lịch sử thứ hai và trung bình di chuyển thời gian dài của biên độ điểm lấy mẫu logarit lớn nhất miền thời gian. lp_speech được cập nhật trong mỗi khung tiếng nói hoạt động (tức là, khung có vad_flag = 1), và phương pháp cập nhật lp_speech là:

$$lp_speech = 0,99 \cdot lp_speech_{-1} + 0,01 \cdot etot.$$

Ý nghĩa của hai công thức nêu trên là: khi phần nhô năng lượng tương đối rõ rệt xuất hiện trong tín hiệu hiện tại (tức là, một số khung tín hiệu mới nhất gồm khung âm thanh hiện tại và một số khung lịch sử của khung âm thanh hiện tại) trong cả thời gian ngắn lẫn thời gian dài, và tín hiệu hiện tại không có đặc tính âm thanh có thanh âm rõ ràng, nếu một số khung lịch sử trước khi khung âm thanh hiện tại chủ yếu là các khung âm nhạc, xem xét được rằng tín hiệu hiện tại là đoạn âm nhạc bộ gõ; ngược lại, ngoài ra, nếu không khung phụ nào của tín hiệu hiện tại có đặc tính âm thanh có tiếng nói rõ ràng và việc tăng tương đối rõ ràng cũng xuất hiện trong đường bao miền thời gian của tín hiệu hiện tại so với độ trung bình thời gian dài của nó, cũng xem xét được rằng tín hiệu hiện tại là đoạn âm nhạc bộ gõ.

Tham số thanh âm, tức là, mức độ tương quan pitch vòng mở chuẩn hóa, ký hiệu mức độ tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch, có thể được thu thập thông qua tìm kiếm pitch vòng hở ACELP, và có giá trị giữa 0 và 1. Điều này thuộc giải pháp kỹ thuật đã biết và do vậy không được mô tả chi tiết sáng chế. Theo phương án thực hiện này, thanh âm được tính toán cho mỗi một trong hai khung phụ của khung âm thanh hiện tại, và các thanh âm được tính trung

bình để thu thập tham số thanh âm của khung âm thanh hiện tại. Tham số thanh âm của khung âm thanh hiện tại cũng được đệm trong bộ đệm lịch sử thanh âm, và theo phương án thực hiện này, chiều dài của bộ đệm lịch sử thanh âm bằng 10.

`mode_mov` được cập nhật trong mỗi khung tiếng nói hoạt động và khi nhiều hơn 30 khung tiếng nói hoạt động liên tiếp đã xuất hiện trước khung này, và phương pháp cập nhật là:

$$\text{mode_mov} = 0,95 \cdot \text{move_mov}_{-1} + 0,05 \cdot \text{mode},$$

trong đó `mode` là kết quả phân loại của khung âm thanh đầu vào hiện tại, và có giá trị nhị phân, trong đó “0” ký hiệu phân loại tiếng nói, và “1” ký hiệu phân loại âm nhạc.

S103: Phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo các thống kê của một phần hoặc tất cả dữ liệu của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số. Khi các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại tiếng nói, khung âm thanh hiện tại được phân loại như là khung thoại; khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, khung âm thanh hiện tại được phân loại như là khung âm nhạc.

Các thống kê ở đây là giá trị được thu thập bằng cách thực hiện hoạt động thống kê trên dao động phổ tần số hợp lệ (tức là, dữ liệu hiệu dụng) được lưu trữ trong bộ nhớ dao động phổ tần số. Chẳng hạn, hoạt động thống kê có thể là hoạt động thu thập giá trị trung bình hoặc biến thiên. Các thống kê theo các phương án thực hiện dưới đây có ý nghĩa tương tự.

Theo phương án thực hiện, bước S103 gồm:

thu thập giá trị trung bình của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số; và

khi giá trị trung bình thu được của dữ liệu hiệu dụng của các dao động

phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Chẳng hạn, khi giá trị trung bình thu được của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng phân loại âm nhạc, khung âm thanh hiện tại được phân loại như là khung âm nhạc; ngược lại khung âm thanh hiện tại được phân loại như là khung thoại.

Nói chung, giá trị dao động phổ tần số của khung âm nhạc tương đối nhỏ, trong khi giá trị dao động phổ tần số của khung thoại tương đối lớn. Do vậy, khung âm thanh hiện tại có thể được phân loại theo các dao động phổ tần số. Rõ ràng là, phân loại tín hiệu cũng có thể được thực hiện trên khung âm thanh hiện tại bằng cách sử dụng phương pháp phân loại khác. Chẳng hạn, số lượng các chi tiết dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số được đếm; bộ nhớ dao động phổ tần số được phân chia, theo số lượng các chi tiết dữ liệu hiệu dụng, thành ít nhất hai khoảng có các chiều dài khác nhau từ đầu gần đến đầu xa, và giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số tương ứng với mỗi khoảng được thu thập, trong đó điểm bắt đầu của các khoảng là vị trí lưu trữ của dao động phổ tần số của khung hiện tại, đầu gần là đầu ở đó dao động phổ tần số của khung hiện tại được lưu trữ, và đầu xa là đầu ở đó dao động phổ tần số của khung lịch sử được lưu trữ; khung âm thanh được phân loại theo các thống kê của các dao động phổ tần số trong khoảng tương đối ngắn, và nếu các thống kê của các tham số này trong khoảng này đủ để phân biệt loại của khung âm thanh, quá trình phân loại kết thúc; ngược lại quá trình phân loại được tiếp tục trong khoảng ngắn nhất của các khoảng tương đối dài còn lại, và phần còn lại có thể được kết luận bằng phép tương tự. Trong quá trình phân loại của mỗi khoảng, khung âm thanh hiện tại được phân loại theo ngưỡng phân loại tương ứng với mỗi khoảng, khung âm thanh hiện tại

được phân loại như là khung thoại hoặc khung âm nhạc, và khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại thoại, khung âm thanh hiện tại được phân loại như là khung thoại; khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, khung âm thanh hiện tại được phân loại như là khung âm nhạc.

Sau khi phân loại tín hiệu, các tín hiệu khác nhau có thể được mã hóa theo các kiểu mã hóa khác nhau. Chẳng hạn, tín hiệu thoại được mã hóa bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại (như CELP), và tín hiệu âm nhạc được mã hóa bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi (như bộ mã hóa dựa trên MDCT).

Theo phương án thực hiện nêu trên, do tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các dao động phổ tần số, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp. Ngoài ra, các dao động phổ tần số được điều chỉnh khi xem xét các yếu tố như độ hoạt động tiếng nói và âm nhạc bộ gõ; do vậy, sáng chế có tốc độ nhận dạng cao hơn đối với tín hiệu âm nhạc, và thích hợp cho việc phân loại tín hiệu âm thanh lại.

Như được thể hiện trên Fig.4, theo phương án thực hiện khác, sau bước S102, phương pháp còn gồm:

S104: Thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, và lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính trong các bộ nhớ, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên tiếp, của cấu trúc điều hòa tín hiệu; và độ dốc năng lượng dự dự báo tuyến tính ký hiệu độ dốc năng lượng dự dự báo

tuyến tính ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh đầu vào thay đổi khi cấp độ đoán tuyến tính tăng.

Một cách tùy chọn, trước khi các tham số này được lưu trữ, phương pháp còn gồm: xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự báo tuyến tính trong các bộ nhớ hay không; và nếu khung âm thanh hiện tại là khung hoạt động, lưu trữ các tham số này; ngược lại bỏ qua lưu trữ các tham số này.

Giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại. Theo phương án thực hiện, giá trị lớn nhất dải cao tần phổ tần số ph được tính toán bằng cách sử dụng công thức dưới đây:

$$ph = \sum_{i=64}^{126} p2v_map(i),$$

trong đó $p2v_map(i)$ ký hiệu giá trị lớn nhất của bin tần số thứ i^{th} của phổ tần số, và giá trị lớn nhất $p2v_map(i)$ được thu thập bằng cách sử dụng công thức dưới đây:

$$p2v_map(i) = \begin{cases} 20\log(peak(i)) - 10\log(vl(i)) - 10\log(vr(i)) & peak(i) \neq 0 \\ 0 & peak(i) = 0 \end{cases},$$

trong đó $peak(i) = C(i)$ nếu bin tần số thứ i^{th} là giá trị lớn nhất cục bộ của phổ tần số; ngược lại $peak(i) = 0$; và $vl(i)$ và $vr(i)$ lần lượt ký hiệu các giá trị thấp nhất phổ tần số cục bộ $v(n)$ liên kề với bin tần số thứ i^{th} ở phía tần số cao và phía tần số thấp của bin tần số thứ i^{th} , trong đó

$$peak(i) = \begin{cases} C(i) & C(i) > C(i-1), C(i) > C(i+1) \\ 0 & \text{else} \end{cases}, \text{ và}$$

$$v = \forall C(i) \quad C(i) < C(i-1), C(i) < C(i+1).$$

Giá trị lớn nhất dải cao tần phổ tần số ph của khung âm thanh hiện tại cũng được đệm ở bộ đệm lịch sử ph, và theo phương án thực hiện này, chiều dài của bộ đệm lịch sử ph bằng 60.

Mức độ tương quan phổ tần số cor_map_sum ký hiệu độ ổn định, giữa các khung liên kế, của cấu trúc điều hòa tín hiệu, và được thu thập bằng cách thực hiện các bước dưới đây:

Trước hết, phổ tần số được loại bỏ sàn $C'(i)$ của khung âm thanh đầu vào $C(i)$ được thu thập, trong đó

$$C'(i) = C(i) - floor(i),$$

trong đó $floor(i)$ ký hiệu sàn phổ của phổ tần số của khung âm thanh đầu vào, trong đó $i = 0, 1, \dots, 127$; và

$$floor(i) = \begin{cases} C(i) & C(i) \in v \\ vl(i) + (i - idx[vl(i)]) \cdot \frac{vr(i) - vl(i)}{idx[vr(i)] - idx[vl(i)]} & \text{else} \end{cases},$$

trong đó $idx[x]$ ký hiệu vị trí của x trên phổ tần số, trong đó $idx[x] = 0, 1, \dots, 127$.

Sau đó, giữa cứ hai giá trị thấp nhất phổ tần số liên kế, tương quan $cor(n)$ giữa phổ tần số được loại bỏ sàn của khung âm thanh đầu vào và phổ tần số được loại bỏ sàn của khung trước đó được thu thập, trong đó

$$cor(n) = \frac{\left(\sum_{i=lb(n)}^{hb(n)} C'(i) \cdot C'_{-1}(i) \right)^2}{\left(\sum_{i=lb(n)}^{hb(n)} C'(i) \cdot C'(i) \right) \cdot \left(\sum_{i=lb(n)}^{hb(n)} C'_{-1}(i) \cdot C'_{-1}(i) \right)},$$

trong đó $lb(n)$ và $hb(n)$ lần lượt ký hiệu các vị trí điểm cuối của khoảng phổ tần số giá trị thấp nhất thứ n^{th} (tức là, vị trí được đặt giữa hai giá trị thấp nhất liên kế), tức là, các vị trí giới hạn hai giá trị thấp nhất phổ tần số của khoảng giá trị thấp nhất.

Cuối cùng, mức độ tương quan phổ tần số cor_map_sum của khung âm thanh đầu vào được tính toán bằng cách sử dụng công thức dưới đây:

$$cor_map_sum = \sum_{i=0}^{127} cor(inv[lb(n) \leq i, hb(n) \geq i]),$$

trong đó $inv[f]$ ký hiệu hàm đảo của hàm f .

Độ dốc năng lượng dự dự báo tuyến tính $epsP_tilt$ ký hiệu mức độ mà năng lượng dự dự báo tuyến tính của tín hiệu âm thanh đầu vào thay đổi khi cấp dự đoán tuyến tính tăng, và có thể được tính toán và được thu thập bằng cách sử dụng công thức dưới đây:

$$epsP_tilt = \frac{\sum_{i=1}^n epsP(i) \cdot epsP(i+1)}{\sum_{i=1}^n epsP(i) \cdot epsP(i)},$$

trong đó $epsP(i)$ ký hiệu năng lượng dự dự báo của dự báo tuyến tính cấp i^{th} ; và n là số nguyên dương, ký hiệu cấp dự đoán tuyến tính, và nhỏ hơn hoặc bằng cấp dự đoán tuyến tính lớn nhất. Chẳng hạn, theo phương án thực hiện, $n = 15$.

Do vậy, bước S103 có thể được thay thế bằng bước dưới đây:

S105: Thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu hiệu dụng được lưu trữ trong các bộ nhớ, trong đó hoạt động tính toán có thể gồm hoạt động thu thập giá trị trung bình, hoạt động thu thập biến thiên, hoặc hoạt động tương tự.

Theo phương án thực hiện, bước này gồm:

thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến

tính được lưu trữ một cách riêng rẽ; và

khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Nói chung, giá trị dao động phổ tần số của khung âm nhạc tương đối nhỏ, trong khi giá trị dao động phổ tần số của khung thoại tương đối lớn; giá trị lớn nhất dải cao tần phổ tần số của khung âm nhạc tương đối lớn, và giá trị lớn nhất dải cao tần phổ tần số của khung thoại tương đối nhỏ; giá trị mức độ tương quan phổ tần số của khung âm nhạc tương đối lớn, và giá trị mức độ tương quan phổ tần số của khung thoại tương đối nhỏ; thay đổi ở độ dốc năng lượng dự dự báo tuyến tính của khung âm nhạc tương đối nhỏ, và thay đổi ở độ dốc năng lượng dự dự báo tuyến tính của khung thoại tương đối lớn. Do vậy, khung âm thanh hiện tại có thể được phân loại theo các thống kê của các tham số nêu trên. Rõ ràng là, phân loại tín hiệu cũng có thể được thực hiện trên khung âm thanh hiện tại bằng cách sử dụng phương pháp phân loại khác. Chẳng hạn, số lượng các chi tiết dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số được đếm; bộ nhớ được phân chia, theo số lượng các chi tiết dữ liệu hiệu dụng, thành ít nhất hai khoảng có các chiều dài khác nhau từ đầu gần đến đầu xa, giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số tương ứng với mỗi khoảng, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan

phổ tần số, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được thu thập, trong đó điểm bắt đầu của các khoảng là vị trí lưu trữ của dao động phổ tần số của khung hiện tại, đầu gần là đầu ở đó dao động phổ tần số của khung hiện tại được lưu trữ, và đầu xa là đầu ở đó dao động phổ tần số của khung lịch sử được lưu trữ; khung âm thanh được phân loại theo các thống kê của dữ liệu hiệu dụng của các tham số nêu trên trong khoảng tương đối ngắn, và nếu các thống kê của các tham số này trong khoảng này đủ để phân biệt loại của khung âm thanh, quá trình phân loại kết thúc; ngược lại quá trình phân loại được tiếp tục trong khoảng ngắn nhất của các khoảng tương đối dài còn lại, và phần còn lại có thể được kết luận bằng phép tương tự. Trong quá trình phân loại của mỗi khoảng, khung âm thanh hiện tại được phân loại theo ngưỡng phân loại tương ứng với mỗi khoảng, và khi một trong các điều kiện dưới đây được thỏa mãn, khung âm thanh hiện tại được phân loại như là khung âm nhạc; ngược lại khung âm thanh hiện tại được phân loại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Sau khi phân loại tín hiệu, các tín hiệu khác nhau có thể được mã hóa theo các kiểu mã hóa khác nhau. Chẳng hạn, tín hiệu thoại được mã hóa bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại (như CELP), và tín hiệu âm nhạc được mã hóa bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi (như bộ mã hóa dựa trên MDCT).

Theo phương án thực hiện nêu trên, tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các dao động phổ tần số, giá trị lớn nhất dải

cao tần phổ tần số, các mức độ tương quan phổ tần số, và các độ dốc năng lượng dự báo tuyến tính; do vậy, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp. Ngoài ra, các dao động phổ tần số được điều chỉnh khi xem xét các yếu tố như độ hoạt động tiếng nói và âm nhạc bộ gõ, và các dao động phổ tần số được điều chỉnh theo môi trường tín hiệu trong đó khung âm thanh hiện tại được đặt; do vậy, sáng chế cải thiện tốc độ nhận dạng phân loại, và thích hợp cho việc phân loại tín hiệu âm thanh lai.

Như được thể hiện trên Fig.5, theo phương án thực hiện khác, phương pháp phân loại tín hiệu âm thanh gồm:

S501: Thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào.

Việc phân loại tín hiệu âm thanh thường được thực hiện trên cơ sở từng khung, và tham số được trích rút từ mỗi khung tín hiệu âm thanh để thực hiện phân loại, để xác định xem liệu khung tín hiệu âm thanh thuộc khung thoại hoặc khung âm nhạc, và thực hiện mã hóa ở chế độ mã hóa tương ứng.

S502: Thu thập độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại, trong đó độ dốc năng lượng dự báo tuyến tính ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng.

Theo phương án thực hiện, độ dốc năng lượng dự báo tuyến tính $epsP_tilt$ có thể được tính toán và thu thập bằng cách sử dụng công thức dưới đây:

$$epsP_tilt = \frac{\sum_{i=1}^n epsP(i) \cdot epsP(i+1)}{\sum_{i=1}^n epsP(i) \cdot epsP(i)}$$

trong đó $epsP(i)$ ký hiệu năng lượng dự báo của dự báo tuyến tính cấp i^{th} ; và n là số nguyên dương, ký hiệu cấp dự đoán tuyến tính, và nhỏ hơn hoặc bằng cấp dự đoán tuyến tính lớn nhất. Chẳng hạn, theo phương

án thực hiện, $n = 15$.

S503: Lưu trữ độ dốc năng lượng dự dự báo tuyến tính trong bộ nhớ.

Độ dốc năng lượng dự dự báo tuyến tính có thể được lưu trữ trong bộ nhớ. Theo phương án thực hiện, bộ nhớ có thể là bộ đệm FIFO, và chiều dài của bộ đệm bằng 60 khối lưu trữ (tức là, 60 độ dốc năng lượng dự dự báo tuyến tính có thể được lưu trữ).

Một cách tùy chọn, trước khi lưu trữ độ dốc năng lượng dự dự báo tuyến tính, phương pháp còn gồm: xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ độ dốc năng lượng dự dự báo tuyến tính trong bộ nhớ hay không; và nếu khung âm thanh hiện tại là khung hoạt động, lưu trữ độ dốc năng lượng dự dự báo tuyến tính; ngược lại bỏ qua lưu trữ độ dốc năng lượng dự dự báo tuyến tính.

S504: Phân loại khung âm thanh theo các thống kê của một phần dữ liệu của các độ nghiêng năng lượng dự dự báo trong bộ nhớ.

Theo phương án thực hiện, các thống kê một phần dữ liệu của các độ nghiêng năng lượng dự dự báo là biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự dự báo, và do vậy bước S504 gồm:

so sánh biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự dự báo với ngưỡng phân loại âm nhạc, và khi biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự dự báo nhỏ hơn ngưỡng phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Nói chung, thay đổi ở giá trị độ dốc năng lượng dự dự báo tuyến tính của khung âm nhạc tương đối nhỏ, và thay đổi ở giá trị độ dốc năng lượng dự dự báo tuyến tính của khung thoại tương đối lớn. Do vậy, khung âm thanh hiện tại có thể được phân loại theo các thống kê của các độ dốc năng lượng dự dự báo tuyến tính. Rõ ràng là, phân loại tín hiệu cũng có thể được thực hiện trên khung âm thanh hiện tại dựa vào tham số khác bằng cách sử dụng phương pháp phân loại khác.

Theo phương án thực hiện khác, trước bước S504, phương pháp còn gồm: thu thập dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, và mức độ tương quan phổ tần số của khung âm thanh hiện tại, và lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, và mức độ tương quan phổ tần số trong các bộ nhớ tương ứng. Do vậy, bước S504 cụ thể là:

thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu hiệu dụng được lưu trữ trong các bộ nhớ.

Ngoài ra, việc thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng gồm:

thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung

âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Nói chung, giá trị dao động phổ tần số của khung âm nhạc tương đối nhỏ, trong khi giá trị dao động phổ tần số của khung thoại tương đối lớn; giá trị lớn nhất dải cao tần phổ tần số của khung âm nhạc tương đối lớn, và giá trị lớn nhất dải cao tần phổ tần số của khung thoại tương đối nhỏ; giá trị mức độ tương quan phổ tần số của khung âm nhạc tương đối lớn, và giá trị mức độ tương quan phổ tần số của khung thoại tương đối nhỏ; thay đổi ở giá trị độ dốc năng lượng dự báo tuyến tính của khung âm nhạc tương đối nhỏ, và thay đổi ở giá trị độ dốc năng lượng dự báo tuyến tính của khung thoại tương đối lớn. Do vậy, khung âm thanh hiện tại có thể được phân loại theo các thống kê của các tham số nêu trên.

Theo phương án thực hiện khác, trước bước S504, phương pháp còn gồm: thu thập lượng âm thanh phổ tần số của khung âm thanh hiện tại và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, và lưu trữ lượng âm thanh phổ tần số và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp trong các bộ nhớ tương ứng. Do vậy, bước S504 cụ thể là:

thu thập các thống kê của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và các thống kê của các lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ; và

phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của các độ dốc năng lượng dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh

phổ tần số trên dải tần số thấp, trong đó các thống kê đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu được lưu trữ trong các bộ nhớ.

Ngoài ra, việc thu thập các thống kê của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ và các thống kê của các lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ gồm: thu thập biến thiên của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ; và thu thập giá trị trung bình của các lượng âm thanh phổ tần số được lưu trữ. Việc phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của các độ dốc năng lượng dự dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp gồm:

khi khung âm thanh hiện tại là khung hoạt động, và một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại:

biến thiên của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ năm; hoặc

giá trị trung bình của các lượng âm thanh phổ tần số lớn hơn ngưỡng thứ sáu; hoặc

tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp nhỏ hơn ngưỡng thứ bảy.

Việc thu thập lượng âm thanh phổ tần số của khung âm thanh hiện tại và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp gồm:

đếm số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, sử dụng số lượng này làm lượng âm thanh phổ tần số; và

tính toán tỷ lệ của số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 4kHz và có các giá trị lớn nhất bin tần số

lớn hơn giá trị định trước với số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, để sử dụng tỷ lệ này làm tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp. Theo phương án thực hiện, giá trị định trước bằng 50.

Lượng âm thanh phổ tần số Ntonal ký hiệu số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước. Theo phương án thực hiện, số lượng có thể được thu thập theo cách thức dưới đây: đếm số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất $p2v_map(i)$ lớn hơn 50, tức là, Ntonal, trong đó $p2v_map(i)$ ký hiệu giá trị lớn nhất của bin tần số thứ i^{th} của phổ tần số, và for a cách thức tính toán of $p2v_map(i)$, tham khảo phần mô tả phương án thực hiện nêu trên.

Tỷ lệ $ratio_Ntonal_lf$ của lượng âm thanh phổ tần số trên dải tần số thấp ký hiệu tỷ lệ của lượng âm dải tần số thấp trên lượng âm thanh phổ tần số. Theo phương án thực hiện, tỷ lệ có thể được thu thập theo cách thức dưới đây: đếm lượng Ntonal_lf của khung âm thanh hiện tại trên dải tần số từ 0kHz đến 4kHz và có $p2v_map(i)$ lớn hơn 50. $ratio_Ntonal_lf$ là tỷ lệ của Ntonal_lf trên Ntonal, tức là, $Ntonal_lf/Ntonal$. $p2v_map(i)$ ký hiệu giá trị lớn nhất của bin tần số thứ i^{th} của phổ tần số, và đối với cách thức tính toán $p2v_map(i)$, tham khảo phần mô tả phương án thực hiện nêu trên. Theo phương án thực hiện khác, trung bình của các giá trị Ntonal được lưu trữ và trung bình của các giá trị Ntonal_lf được lưu trữ được thu thập riêng rẽ, và tỷ lệ của trung bình các giá trị Ntonal_lf trên trung bình của các giá trị Ntonal được tính toán được sử dụng làm tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp.

Theo phương án thực hiện này, tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các độ dốc năng lượng dự báo tuyến tính.

Ngoài ra, cả độ vững phân loại lẫn tốc độ nhận dạng phân loại được xem xét; do vậy, không có tham số phân loại, nhưng kết quả tương đối chính xác, độ phức tạp thấp, và các chi phí bổ sung bộ nhớ thấp.

Như được thể hiện trên Fig.6, theo phương án thực hiện khác, phương pháp phân loại tín hiệu âm thanh gồm:

S601: Thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào.

S602: Thu thập dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại.

Mật độ dao động phổ tần số ký hiệu dao động năng lượng thời gian ngắn hoặc thời gian dài của phổ tần số của tín hiệu, và là giá trị trung bình của các giá trị tuyệt đối của các hiệu số năng lượng logarit giữa các tần số tương ứng của khung âm thanh hiện tại và khung lịch sử trên phổ dải giữa và thấp, trong đó khung lịch sử đề cập đến khung bất kỳ trước khi khung âm thanh hiện tại. Giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại. Mức độ tương quan phổ tần số cor_map_sum ký hiệu độ ổn định, giữa các khung liên kề, của cấu trúc điều hòa tín hiệu. Độ dốc năng lượng dự báo tuyến tính $epsP_tilt$ ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh đầu vào thay đổi khi cấp dự đoán tuyến tính tăng. Đối với phương pháp cụ thể tính toán các tham số này, tham khảo phương án thực hiện nêu trên.

Ngoài ra, tham số thanh âm có thể được thu thập; và tham số thanh âm ký hiệu mức độ tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch. Tham số thanh âm thanh âm được thu thập bằng cách phân tích và dự báo tuyến tính, biểu thị mức độ tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch, và có giá trị giữa 0 và 1. Điều này thuộc giải pháp kỹ thuật đã biết,

và do vậy không được mô tả chi tiết sáng chế. Theo phương án thực hiện này, thanh âm được tính toán cho mỗi trong hai khung phụ của khung âm thanh hiện tại, và các thanh âm được tính trung bình để thu thập tham số thanh âm của khung âm thanh hiện tại. Tham số thanh âm của khung âm thanh hiện tại cũng được đệm trong bộ đệm lịch sử thanh âm, và theo phương án thực hiện này, chiều dài của bộ đệm lịch sử thanh âm bằng 10.

S603: Lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính trong các bộ nhớ tương ứng.

Một cách tùy chọn, trước khi các tham số này được lưu trữ, phương pháp còn gồm:

Theo phương án thực hiện, được xác định theo độ hoạt động tiếng nói của khung âm thanh hiện tại liệu có lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số. Nếu khung âm thanh hiện tại là khung hoạt động, dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số.

Theo phương án thực hiện khác, được xác định, theo độ hoạt động tiếng nói của khung âm thanh và liệu khung âm thanh có phải là tấn công năng lượng hay không, liệu có lưu trữ dao động phổ tần số trong bộ nhớ. Nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số. Theo phương án thực hiện khác, nếu khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ. Chẳng hạn, nếu khung âm thanh hiện tại là khung hoạt động, và khung trước đó của khung âm thanh hiện tại hoặc khung lịch sử thứ hai của khung âm thanh

hiện tại không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ.

Đối với các định nghĩa và cách thức thu thập của cờ hoạt động tiếng nói vad_flag và tiếng nói attack flag attack_flag, tham khảo phần mô tả phương án thực hiện nêu trên.

Một cách tùy chọn, trước khi các tham số này được lưu trữ, phương pháp còn gồm:

xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự báo tuyến tính trong các bộ nhớ; và nếu khung âm thanh hiện tại là khung hoạt động, lưu trữ các tham số này; ngược lại bỏ qua lưu trữ các tham số này.

S604: Thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu hiệu dụng được lưu trữ trong các bộ nhớ, trong đó hoạt động tính toán có thể gồm hoạt động thu thập giá trị trung bình, hoạt động thu thập biến thiên, hoặc hoạt động tương tự.

Một cách tùy chọn, trước bước S604, phương pháp có thể còn gồm:

cập nhật, theo việc liệu khung âm thanh hiện tại có phải là âm nhạc bộ gõ hay không, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số. Theo phương án thực hiện, nếu khung âm thanh hiện tại là âm nhạc bộ gõ, các giá trị dao động phổ tần số hợp lệ trong bộ nhớ dao

động phổ tần số được điều chỉnh thành giá trị nhỏ hơn hoặc bằng ngưỡng âm nhạc, trong đó khi dao động phổ tần số của khung âm thanh nhỏ hơn ngưỡng âm nhạc, âm thanh được phân loại như là khung âm nhạc. Theo phương án thực hiện, nếu khung âm thanh hiện tại là âm nhạc bộ gõ, các giá trị dao động phổ tần số hợp lệ trong bộ nhớ dao động phổ tần số được đặt lại bằng 5.

Một cách tùy chọn, trước khi bước S604, phương pháp có thể còn gồm:

cập nhật các dao động phổ tần số trong bộ nhớ theo độ hoạt động của khung lịch sử của khung âm thanh hiện tại. Theo phương án thực hiện, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và khung âm thanh trước đó là khung không hoạt động, dữ liệu của các dao động phổ tần số khác được lưu trữ trong bộ nhớ dao động phổ tần số ngoại trừ dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa trong dữ liệu hiệu dụng. Theo phương án thực hiện khác, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa thành giá trị thứ nhất. Giá trị thứ nhất có thể là ngưỡng thoại, trong đó khi dao động phổ tần số của khung âm thanh lớn hơn ngưỡng thoại, âm thanh được phân loại như là khung thoại. Theo phương án thực hiện khác, nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và kết quả phân loại của khung lịch sử là khung âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, dao động phổ tần số của khung âm thanh hiện tại được chỉnh sửa thành giá trị thứ hai, trong đó giá trị thứ hai lớn hơn giá trị thứ nhất.

Chẳng hạn, nếu khung trước đó của khung âm thanh hiện tại là khung

không hoạt động ($\text{vad_flag} = 0$), ngoại trừ mật độ khung âm thanh hiện tại mới được đếm trong bộ đệm lịch sử mật độ, dữ liệu còn lại trong bộ đệm lịch sử mật độ đều được đặt lại bằng -1 (tương đương việc dữ liệu bị mất hiệu lực). Nếu ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động ($\text{vad_flag} = 1$), mật độ khung âm thanh hiện tại vừa được đếm trong bộ đệm lịch sử mật độ được chỉnh sửa thành 16. Nếu ba khung liên tiếp trước khung âm thanh hiện tại đều là các khung hoạt động ($\text{vad_flag} = 1$), kết quả tron tru thời gian dài của kết quả phân loại tín hiệu lịch sử là tín hiệu âm nhạc và mật độ khung âm thanh hiện tại lớn hơn 20, dao động phổ tần số của khung âm thanh hiện tại được đếm được chỉnh sửa thành 20. Để tính toán khung hoạt động và kết quả tron tru thời gian dài of the kết quả phân loại tín hiệu lịch sử, tham khảo phương án thực hiện nêu trên.

Theo phương án thực hiện, bước S604 gồm:

thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Nói chung, giá trị dao động phổ tần số của khung âm nhạc tương đối nhỏ, trong khi giá trị dao động phổ tần số của khung thoại tương đối lớn; giá trị lớn nhất dải cao tần phổ tần số của khung âm nhạc tương đối lớn, và giá trị lớn nhất dải cao tần phổ tần số của khung thoại tương đối nhỏ; giá trị mức độ tương quan phổ tần số của khung âm nhạc tương đối lớn, và giá trị mức độ tương quan phổ tần số của khung thoại tương đối nhỏ; giá trị độ dốc năng lượng dự báo tuyến tính của khung âm nhạc tương đối nhỏ, và giá trị độ dốc năng lượng dự báo tuyến tính của khung thoại tương đối lớn. Do vậy, khung âm thanh hiện tại có thể được phân loại theo các thống kê của các tham số nêu trên. Rõ ràng là, việc phân loại tín hiệu cũng có thể được thực hiện trên khung âm thanh hiện tại bằng cách sử dụng phương pháp phân loại khác. Chẳng hạn, số lượng các chi tiết dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số được đếm; bộ nhớ được phân chia, theo số lượng các chi tiết dữ liệu hiệu dụng, thành ít nhất hai khoảng có các chiều dài khác nhau từ đầu gần đến đầu xa, giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số tương ứng với mỗi khoảng, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được thu thập, trong đó điểm bắt đầu của các khoảng là vị trí lưu trữ của dao động phổ tần số của khung hiện tại, đầu gần là đầu ở đó dao động phổ tần số của khung hiện tại được lưu trữ, và đầu xa là đầu ở đó dao động phổ tần số của khung lịch sử được lưu trữ; khung âm thanh được phân loại theo các thống kê của dữ liệu hiệu dụng của các tham số nêu trên trong khoảng tương đối ngắn, và nếu các thống kê tham số trong khoảng này đủ để phân biệt loại của khung âm thanh, quá trình phân loại kết thúc; ngược lại quá trình phân loại được tiếp tục trong khoảng ngắn nhất của các khoảng tương đối dài còn lại, và phần

còn lại có thể được kết luận bằng phép tương tự. Trong quá trình phân loại của mỗi khoảng, khung âm thanh hiện tại được phân loại theo ngưỡng phân loại tương ứng với mỗi khoảng, và khi một trong các điều kiện dưới đây được thỏa mãn, khung âm thanh hiện tại được phân loại như là khung âm nhạc; ngược lại khung âm thanh hiện tại được phân loại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Sau khi phân loại tín hiệu, các tín hiệu khác nhau có thể được mã hóa theo các kiểu mã hóa khác nhau. Chẳng hạn, tín hiệu thoại được mã hóa bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại (như CELP), và tín hiệu âm nhạc được mã hóa bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi (như bộ mã hóa dựa trên MDCT).

Theo phương án thực hiện này, việc phân loại được thực hiện theo các thống kê lâu dài của các dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, các mức độ tương quan phổ tần số, và các độ dốc năng lượng dự báo tuyến tính. Ngoài ra, cả độ vững phân loại lẫn tốc độ nhận dạng phân loại được xem xét; do vậy, không có tham số phân loại, nhưng kết quả tương đối chính xác, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp.

Theo phương án thực hiện, sau khi mật độ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số ph , mức độ tương quan phổ tần số cor_map_sum , và độ dốc năng lượng dự báo tuyến tính $epsP_tilt$ được lưu trữ trong các bộ nhớ tương ứng, việc phân loại có thể được thực hiện theo số lượng các chi tiết dữ liệu hiệu dụng của các dao động phổ tần số

được lưu trữ bằng cách sử dụng các quá trình xác định khác nhau. Nếu cờ hoạt động tiếng nói được chọn bằng 1, tức là, khung âm thanh hiện tại là khung tiếng nói hoạt động, lượng N chi tiết dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ được kiểm tra.

Nếu giá trị của lượng N chi tiết dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ thay đổi, quá trình xác định cũng thay đổi.

(1) Như được thể hiện trên Fig.7, nếu $N = 60$, giá trị trung bình của tất cả dữ liệu trong bộ đệm lịch sử mật độ được thu thập và được đánh dấu là flux60, giá trị trung bình của 30 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là flux30, và giá trị trung bình của 10 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là flux10. Giá trị trung bình của tất cả dữ liệu trong bộ đệm lịch sử ph được thu thập và được đánh dấu là ph60, giá trị trung bình của 30 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là ph30, và giá trị trung bình của 10 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là ph10. Giá trị trung bình của tất cả dữ liệu trong bộ đệm lịch sử cor_map_sum được thu thập và được đánh dấu là cor_map_sum60, giá trị trung bình của 30 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là cor_map_sum30, và giá trị trung bình của 10 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là cor_map_sum10. Ngoài ra, biến thiên của tất cả dữ liệu trong bộ đệm lịch sử epsP_tilt được thu thập và được đánh dấu là epsP_tilt60, biến thiên của 30 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là epsP_tilt30, và biến thiên của 10 chi tiết dữ liệu ở đầu gần được thu thập và được đánh dấu là epsP_tilt10. Lượng voicing_cnt của chi tiết dữ liệu có giá trị lớn hơn 0,9 trong bộ đệm lịch sử thanh âm được thu thập. Đầu gần là đầu ở đó các tham số nêu trên tương ứng với khung âm thanh hiện tại được lưu trữ.

Trước hết, kiểm tra được liệu flux10, ph10, epsP_tilt10,

cor_map_sum10, và voicing_cnt có thỏa mãn các điều kiện dưới đây hay không: $\text{flux10} < 10$ hoặc $\text{epsPtilt10} < 0,0001$ hoặc $\text{ph10} > 1050$ hoặc $\text{cor_map_sum10} > 95$, và $\text{voicing_cnt} < 6$. Nếu các điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại như là loại âm nhạc (tức là, Mode = 1). Ngược lại, kiểm tra liệu flux10 có lớn hơn 15 hay không và liệu voicing_cnt có lớn hơn 2 hay không, hoặc liệu flux10 có lớn hơn 16 hay không. Nếu các điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại như là loại thoại (tức là, Mode = 0). Ngược lại, được kiểm tra liệu flux30, flux10, ph30, epsP_tilt30, cor_map_sum30, và voicing_cnt thỏa mãn các điều kiện dưới đây: $\text{flux30} < 13$ và $\text{flux10} < 15$, hoặc $\text{epsPtilt30} < 0,001$ hoặc $\text{ph30} > 800$ hoặc $\text{cor_map_sum30} > 75$. Nếu các điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại như là loại âm nhạc. Ngược lại, kiểm tra được liệu flux60, flux30, ph60, epsP_tilt60, và cor_map_sum60 có thỏa mãn các điều kiện dưới đây hay không: $\text{flux60} < 14,5$ hoặc $\text{cor_map_sum30} > 75$ hoặc $\text{ph60} > 770$ hoặc $\text{epsP_tilt10} < 0,002$, và $\text{flux30} < 14$. Nếu các điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại như là loại âm nhạc; ngược lại khung âm thanh hiện tại được phân loại như là loại thoại.

(2) Như được thể hiện trên Fig.8, nếu $N < 60$ và $N \geq 30$, giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử mật độ, giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử ph, và giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử cor_map_sum được thu thập riêng rẽ và được đánh dấu là fluxN, phN, và cor_map_sumN. Ngoài ra, biến thiên của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử epsP_tilt được thu thập và được đánh dấu là epsP_tiltN. Kiểm tra liệu fluxN, phN, epsP_tiltN, và cor_map_sumN thỏa mãn điều kiện dưới đây: $\text{fluxN} < 13 + (N - 30)/20$ hoặc $\text{cor_map_sumN} > 75 + (N - 30)/6$ hoặc $\text{phN} > 800$ hoặc $\text{epsP_tiltN} < 0,001$. Nếu điều kiện được thỏa mãn, khung âm thanh hiện tại được phân

loại thành loại âm nhạc; ngược lại khung âm thanh hiện tại được phân loại thành loại thoại.

(3) Như được thể hiện trên Fig.9, nếu $N < 30$ và $N \geq 10$, giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử mật độ, giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử ph, và giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử cor_map_sum được thu thập riêng rẽ và được đánh dấu là $fluxN$, phN , và cor_map_sumN . Ngoài ra, biến thiên của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử $epsP_tilt$ được thu thập và được đánh dấu là $epsP_tiltN$.

Trước hết, kiểm tra được liệu trung bình di chuyển thời gian dài $mode_mov$ của kết quả phân loại lịch sử có lớn hơn 0,8 hay không. Nếu có, được kiểm tra liệu $fluxN$, phN , $epsP_tiltN$, và cor_map_sumN thỏa mãn điều kiện dưới đây: $fluxN < 16 + (N - 10)/20$ hoặc $phN > 1000 - 12,5 \times (N - 10)$ hoặc $epsP_tiltN < 0,0005 + 0,000045 \times (N - 10)$ hoặc $cor_map_sumN > 90 - (N - 10)$. Ngược lại, lượng voicing_cnt của chi tiết dữ liệu có giá trị lớn hơn 0,9 trong bộ đệm lịch sử thanh âm được thu thập, và được kiểm tra liệu các điều kiện dưới đây được thỏa mãn: $fluxN < 12 + (N - 10)/20$ hoặc $phN > 1050 - 12,5 \times (N - 10)$ hoặc $epsP_tiltN < 0,0001 + 0,000045 \times (N - 10)$ hoặc $cor_map_sumN > 95 - (N - 10)$, và $voicing_cnt < 6$. Nếu nhóm bất kỳ trong hai nhóm điều kiện nêu trên được thỏa mãn, khung âm thanh hiện tại được phân loại là loại âm nhạc; ngược lại khung âm thanh hiện tại được phân loại là loại thoại.

(4) Như được thể hiện trên Fig.10, nếu $N < 10$ và $N > 5$, giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử ph và giá trị trung bình của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử cor_map_sum được thu thập và được đánh dấu là phN và cor_map_sumN , và biến thiên của N chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử $epsP_tilt$ được thu thập và được đánh dấu là $epsP_tiltN$. Ngoài ra, lượng

voicing_cnt6 của chi tiết dữ liệu có giá trị lớn hơn 0,9 trong sáu chi tiết dữ liệu ở đầu gần trong bộ đệm lịch sử thanh âm được thu thập.

Kiểm tra liệu các điều kiện dưới đây có được thỏa mãn hay không: $\text{epsP_tiltN} < 0,00008$ hoặc $\text{phN} > 1100$ hoặc $\text{cor_map_sumN} > 100$, và $\text{voicing_cnt} < 4$. Nếu các điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại là loại âm nhạc; ngược lại khung âm thanh hiện tại được phân loại là loại thoại.

(5) Nếu $N \leq 5$, kết quả phân loại của khung âm thanh trước đó được sử dụng làm loại phân loại của khung âm thanh hiện tại.

Phương án thực hiện nêu trên là quá trình phân loại cụ thể trong đó việc phân loại được thực hiện theo các thống kê lâu dài của các dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, các mức độ tương quan phổ tần số, và các độ dốc năng lượng dự dự báo tuyến tính, và chuyên gia trong lĩnh vực có thể hiểu rằng, việc phân loại có thể được thực hiện bằng cách sử dụng quá trình khác. Quá trình phân loại theo phương án thực hiện này có thể được áp dụng cho các bước tương ứng theo phương án thực hiện nêu trên, để dùng làm, chẳng hạn, phương pháp phân loại cụ thể của bước 103 trên Fig.2, bước 105 trên Fig.4, hoặc bước 604 trên Fig.6.

Như được thể hiện trên Fig.11, theo phương án thực hiện khác, phương pháp phân loại tín hiệu âm thanh gồm:

S1101: Thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào.

S1102: Thu thập độ dốc năng lượng dự dự báo tuyến tính và lượng âm thanh phổ tần số của khung âm thanh hiện tại và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp.

Độ dốc năng lượng dự dự báo tuyến tính epsP_tilt ký hiệu mức độ mà năng lượng dự dự báo tuyến tính của tín hiệu âm thanh đầu vào thay đổi khi cấp dự đoán tuyến tính tăng; lượng âm thanh phổ tần số N_{tonal} ký hiệu số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định

trước; tỷ lệ ratio_Ntonal_lf của lượng âm thanh phổ tần số trên dải tần số thấp ký hiệu tỷ lệ của lượng âm dải tần số thấp trên lượng âm thanh phổ tần số. Để tính toán cụ thể, tham khảo phần mô tả phương án thực hiện nêu trên.

S1103: Lưu trữ độ dốc năng lượng dự báo tuyến tính epsP_tilt , lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp trong các bộ nhớ tương ứng.

Độ dốc năng lượng dự báo tuyến tính epsP_tilt và lượng âm thanh phổ tần số của khung âm thanh hiện tại được đệm trong các bộ đệm lịch sử tương ứng, và theo phương án thực hiện này, các chiều dài của hai bộ đệm cũng đều bằng 60.

Một cách tùy chọn, trước khi các tham số này được lưu trữ, phương pháp còn gồm: xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ độ dốc năng lượng dự báo tuyến tính, lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp trong các bộ nhớ; và lưu trữ độ dốc năng lượng dự báo tuyến tính trong bộ nhớ khi xác định được rằng độ dốc năng lượng dự báo tuyến tính cần được lưu trữ. Nếu khung âm thanh hiện tại là khung hoạt động, các tham số này được lưu trữ; ngược lại các tham số này không được lưu trữ.

S1104: Thu thập các thống kê của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và các thống kê của các lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ, trong đó các thống kê đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu được lưu trữ trong các bộ nhớ, trong đó hoạt động tính toán có thể gồm hoạt động thu thập giá trị trung bình, hoạt động thu thập biến thiên, hoặc hoạt động tương tự.

Theo phương án thực hiện, việc thu thập các thống kê của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và các thống kê của các

lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ gồm: thu thập biến thiên của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ; và thu thập giá trị trung bình của các lượng âm thanh phổ tần số được lưu trữ.

S1105: Phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của các độ dốc năng lượng dự dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp.

Theo phương án thực hiện, bước này gồm:

khi khung âm thanh hiện tại là khung hoạt động, và một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại:

biến thiên của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ năm; hoặc

giá trị trung bình của các lượng âm thanh phổ tần số lớn hơn ngưỡng thứ sáu; hoặc

tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp nhỏ hơn ngưỡng thứ bảy.

Nói chung, giá trị độ dốc năng lượng dự dự báo tuyến tính của khung âm nhạc tương đối nhỏ, và giá trị độ dốc năng lượng dự dự báo tuyến tính của khung thoại tương đối lớn; lượng âm thanh phổ tần số của khung âm nhạc tương đối lớn, và lượng âm thanh phổ tần số của khung thoại tương đối nhỏ; tỷ lệ của lượng âm thanh phổ tần số của khung âm nhạc trên dải tần số thấp tương đối thấp, và tỷ lệ của lượng âm thanh phổ tần số của khung thoại trên dải tần số thấp tương đối cao (năng lượng của khung thoại chủ yếu được tập trung trên dải tần số thấp). Do vậy, khung âm thanh hiện tại có thể được phân loại theo các thống kê của các tham số nêu trên. Rõ ràng là, phân loại tín hiệu cũng có thể được thực hiện trên

khung âm thanh hiện tại bằng cách sử dụng phương pháp phân loại khác.

Sau khi phân loại tín hiệu, các tín hiệu khác nhau có thể được mã hóa theo các kiểu mã hóa khác nhau. Chẳng hạn, tín hiệu thoại được mã hóa bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại (như CELP), và tín hiệu âm nhạc được mã hóa bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi (như bộ mã hóa dựa trên MDCT).

Theo phương án thực hiện nêu trên, tín hiệu âm thanh được phân loại theo các thông kê lâu dài của các độ dốc năng lượng dự báo tuyến tính và các lượng âm thanh phổ tần số và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp; do vậy, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp.

Theo phương án thực hiện, sau khi độ dốc năng lượng dự báo tuyến tính epsP_tilt , lượng âm thanh phổ tần số Ntonal , và tỷ lệ ratio_Ntonal_lf của lượng âm thanh phổ tần số trên dải tần số thấp được lưu trữ trong các bộ đệm tương ứng, biến thiên của tất cả dữ liệu trong bộ đệm lịch sử epsP_tilt được thu thập và được đánh dấu là epsP_tilt60 . Giá trị trung bình của tất cả dữ liệu trong bộ đệm lịch sử Ntonal được thu thập và được đánh dấu là Ntonal 60 . Giá trị trung bình của tất cả dữ liệu trong bộ đệm lịch sử Ntonal_lf được thu thập, và tỷ lệ của giá trị trung bình trên Ntonal60 được tính toán và được đánh dấu là ratio_Ntonal_lf60 . Như được thể hiện trên Fig.12, khung âm thanh hiện tại được phân loại theo quy tắc dưới đây:

Nếu cờ hoạt động tiếng nói bằng 1 (tức là, $\text{vad_flag} = 1$), tức là, khung âm thanh hiện tại là khung tiếng nói hoạt động, kiểm tra được liệu điều kiện dưới đây có được thỏa mãn hay không: $\text{epsP_tilt60} < 0,002$ hoặc $\text{Ntonal60} > 18$ hoặc $\text{ratio_Ntonal_lf60} < 0,42$, Nếu điều kiện được thỏa mãn, khung âm thanh hiện tại được phân loại thành loại âm nhạc (tức là, $\text{Mode} = 1$); ngược lại khung âm thanh hiện tại được phân loại thành loại thoại (tức là, $\text{Mode} = 0$).

Phương án thực hiện nêu trên là quá trình phân loại cụ thể trong đó việc phân loại được thực hiện theo các thống kê của các độ dốc năng lượng dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, và chuyên gia trong lĩnh vực có thể hiểu rằng, việc phân loại có thể được thực hiện bằng cách sử dụng quá trình khác. Quá trình phân loại theo phương án thực hiện này có thể được áp dụng cho các bước tương ứng theo phương án thực hiện nêu trên, để dùng làm, chẳng hạn, phương pháp phân loại cụ thể của bước 504 trên Fig.5 hoặc bước 1105 trên Fig.11.

Sáng chế đề xuất phương pháp lựa chọn kiểu mã hóa âm thanh có độ phức tạp thấp và chi phí bổ sung bộ nhớ thấp. Ngoài ra, cả độ vững phân loại và tốc độ nhận dạng phân loại được xem xét.

Cùng với phương pháp nêu trên theo phương án thực hiện, sáng chế còn đề xuất thiết bị phân loại tín hiệu âm thanh, và thiết bị có thể được đặt trong thiết bị đầu cuối hoặc thiết bị mạng. Thiết bị phân loại tín hiệu âm thanh có thể thực hiện các bước của phương pháp nêu trên theo phương án thực hiện.

Như được thể hiện trên Fig.13, sáng chế đề xuất thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và gồm:

khối xác định lưu trữ 1301, được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập và lưu trữ dao động phổ tần số của khung âm thanh hiện tại, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh;

bộ nhớ 1302, được cấu hình để lưu trữ dao động phổ tần số khi khối xác định lưu trữ xuất ra kết quả mà dao động phổ tần số cần được lưu trữ;

khối cập nhật 1304, được cấu hình để cập nhật, theo việc liệu khung thoại là âm nhạc bộ gõ hoặc hoạt động của khung âm thanh lịch sử, các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khối phân loại 1303, được cấu hình để phân loại khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ; và khi các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại tiếng nói, phân loại khung âm thanh hiện tại như là khung thoại; hoặc khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc.

Theo phương án thực hiện, khối xác định lưu trữ 1301 được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Theo phương án thực hiện khác, khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Theo phương án thực hiện khác, khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

Theo phương án thực hiện, khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại thuộc âm nhạc bộ gõ, chỉnh sửa các giá trị của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số.

Theo phương án thực hiện khác, khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh trước đó là khung không hoạt động, chỉnh sửa dữ liệu của các dao động

phổ tần số khác được lưu trữ trong bộ nhớ ngoại trừ dao động phổ tần số của khung âm thanh hiện tại thành dữ liệu không hiệu dụng; hoặc nếu khung âm thanh hiện tại là khung hoạt động, và ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ nhất; hoặc nếu khung âm thanh hiện tại là khung hoạt động, và kết quả phân loại lịch sử là tín hiệu âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ hai, trong đó giá trị thứ hai lớn hơn giá trị thứ nhất.

Như được thể hiện trên Fig.14, theo phương án thực hiện, khối phân loại 1303 gồm:

khối tính toán 1401, được cấu hình để thu thập giá trị trung bình của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khối xác định 1402, được cấu hình để so sánh giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số với điều kiện phân loại âm nhạc; và khi giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Chẳng hạn, khi giá trị trung bình thu được của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng phân loại âm nhạc, khung âm thanh hiện tại được phân loại như là khung âm nhạc; ngược lại khung âm thanh hiện tại được phân loại như là khung thoại.

Theo phương án thực hiện nêu trên, do tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các dao động phổ tần số, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp. Ngoài ra, các dao động phổ tần số được điều chỉnh khi xem xét

các yếu tố như độ hoạt động tiếng nói và âm nhạc bộ gõ; do vậy, sáng chế có tốc độ nhận dạng cao hơn đối với tín hiệu âm nhạc, và thích hợp cho việc phân loại tín hiệu âm thanh lai.

Theo phương án thực hiện khác, thiết bị phân loại tín hiệu âm thanh còn gồm:

khối thu thập tham số, được cấu hình để thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kế, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; và độ dốc năng lượng dự dự báo tuyến tính ký hiệu mức độ mà năng lượng dự dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng; trong đó

khối xác định lưu trữ còn được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính;

khối lưu trữ còn được cấu hình để: khi khối xác định lưu trữ xuất ra kết quả mà giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính cần được lưu trữ, lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính; và

khối phân loại được cấu hình cụ thể để thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ, và phân loại khung âm

thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng; và khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại tiếng nói, phân loại khung âm thanh hiện tại như là khung thoại; hoặc khi các thống kê dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc.

Theo phương án thực hiện, khối phân loại cụ thể gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khối xác định, được cấu hình để: khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Theo phương án thực hiện nêu trên, tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, các mức độ tương quan phổ tần số, và các độ dốc năng lượng dự báo tuyến tính; do vậy, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp. Ngoài ra, các dao động phổ tần số được điều chỉnh khi xem xét các yếu tố như độ hoạt động tiếng nói và âm nhạc bộ gõ, và các dao động phổ tần số được điều

chỉnh theo môi trường tín hiệu trong đó khung âm thanh hiện tại được đặt; do vậy, sáng chế cải thiện tốc độ nhận dạng phân loại, và thích hợp cho việc phân loại tín hiệu âm thanh lai.

Như được thể hiện trên Fig.15, sáng chế đề xuất phương án thực hiện khác của thiết bị phân loại tín hiệu âm thanh, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và gồm:

khối phân khung 1501, được cấu hình để thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào;

khối thu thập tham số 1502, được cấu hình để thu thập độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại, trong đó độ dốc năng lượng dự báo tuyến tính ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng;

khối lưu trữ 1503, được cấu hình để lưu trữ độ dốc năng lượng dự báo tuyến tính; và

khối phân loại 1504, được cấu hình để phân loại khung âm thanh theo các thống kê của một phần dữ liệu của các độ nghiêng năng lượng dự báo trong bộ nhớ.

Như được thể hiện trên Fig.16, thiết bị phân loại tín hiệu âm thanh còn gồm:

khối xác định lưu trữ 1505, được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ độ dốc năng lượng dự báo tuyến tính trong bộ nhớ, trong đó

khối lưu trữ 1503 được cấu hình cụ thể để: khi khối xác định lưu trữ xác định rằng độ dốc năng lượng dự báo tuyến tính cần được lưu trữ, lưu trữ độ dốc năng lượng dự báo tuyến tính trong bộ nhớ.

Theo phương án thực hiện, các thống kê một phần dữ liệu của các độ nghiêng năng lượng dự báo là biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự báo; và

khối phân loại được cấu hình cụ thể để so sánh biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự báo với ngưỡng phân loại âm nhạc, và khi biến thiên một phần dữ liệu của các độ nghiêng năng lượng dự báo nhỏ hơn ngưỡng phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại.

Theo phương án thực hiện khác, khối thu thập tham số còn được cấu hình để: thu thập dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, và mức độ tương quan phổ tần số của khung âm thanh hiện tại, và lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, và mức độ tương quan phổ tần số trong các bộ nhớ tương ứng; và

khối phân loại được cấu hình cụ thể để thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu hiệu dụng được lưu trữ trong các bộ nhớ.

Như được thể hiện trên Fig.17, một cách cụ thể, theo phương án thực hiện, khối phân loại 1504 gồm:

khối tính toán 1701, được cấu hình để thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ một cách riêng rẽ;

và

khối xác định 1702, được cấu hình để: khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Theo phương án thực hiện khác, khối thu thập tham số còn được cấu hình để thu thập lượng âm thanh phổ tần số của khung âm thanh hiện tại và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, và lưu trữ lượng âm thanh phổ tần số và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp trong các bộ nhớ; và

khối phân loại được cấu hình cụ thể để thu thập các thống kê của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và các thống kê của các lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ; và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của các độ dốc năng lượng dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu được lưu trữ trong các bộ nhớ.

Một cách cụ thể, khối phân loại gồm:

khối tính toán, được cấu hình để thu thập biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và giá trị trung bình của các lượng âm thanh phổ tần số được lưu trữ; và

khối xác định, được cấu hình để: khi khung âm thanh hiện tại là khung

hoạt động, và một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: biến thiên của các độ dốc năng lượng dư dự báo tuyến tính nhỏ hơn ngưỡng thứ năm; hoặc giá trị trung bình của các lượng âm thanh phổ tần số lớn hơn ngưỡng thứ sáu; hoặc tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp nhỏ hơn ngưỡng thứ bảy.

Một cách cụ thể, khối thu thập tham số thu thập độ dốc năng lượng dư dự báo tuyến tính của khung âm thanh hiện tại theo công thức dưới đây:

$$epsP_tilt = \frac{\sum_{i=1}^n epsP(i) \cdot epsP(i+1)}{\sum_{i=1}^n epsP(i) \cdot epsP(i)},$$

trong đó $epsP(i)$ ký hiệu năng lượng dư dự báo của dự báo tuyến tính cấp i^{th} của khung âm thanh hiện tại; và n là số nguyên dương, ký hiệu cấp dự đoán tuyến tính, và nhỏ hơn hoặc bằng cấp dự đoán tuyến tính lớn nhất.

Một cách cụ thể, khối thu thập tham số được cấu hình để đếm số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, sử dụng số lượng này làm lượng âm thanh phổ tần số; và khối thu thập tham số được cấu hình để tính toán tỷ lệ của số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 4kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước trên số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, sử dụng tỷ lệ này làm tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp.

Theo phương án thực hiện này, tín hiệu âm thanh được phân loại theo các thống kê lâu dài của các độ dốc năng lượng dư dự báo tuyến tính. Ngoài ra, cả độ vững phân loại và tốc độ nhận dạng phân loại được xem

xét; do vậy, không có tham số phân loại, nhưng kết quả tương đối chính xác, độ phức tạp thấp, và các chi phí bổ sung bộ nhớ thấp.

Sáng chế đề xuất thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và gồm:

khối phân khung, được cấu hình để thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào;

khối thu thập tham số, được cấu hình để thu thập dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh; giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kế, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; và độ dốc năng lượng dự dự báo tuyến tính ký hiệu mức độ mà năng lượng dự dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng;

khối lưu trữ, được cấu hình để lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính; và

khối phân loại, được cấu hình để thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ

liệu thu được sau khi hoạt động tính toán được thực hiện trên dữ liệu hiệu dụng được lưu trữ trong các bộ nhớ, trong đó hoạt động tính toán có thể gồm hoạt động thu thập giá trị trung bình, hoạt động thu thập biến thiên, hoặc hoạt động tương tự.

Theo phương án thực hiện, thiết bị phân loại tín hiệu âm thanh có thể còn gồm:

khối xác định lưu trữ, được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại; và

khối lưu trữ được cấu hình cụ thể để: khi khối xác định lưu trữ xuất ra kết quả mà dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính cần được lưu trữ, lưu trữ dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính.

Một cách cụ thể, theo phương án thực hiện, khối xác định lưu trữ xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số hay không. Nếu khung âm thanh hiện tại là khung hoạt động, khối xác định lưu trữ xuất ra kết quả rằng tham số cần được lưu trữ; ngược lại khối xác định lưu trữ xuất ra kết quả rằng tham số không cần được lưu trữ. Theo phương án thực hiện khác, khối xác định lưu trữ xác định, theo độ hoạt động tiếng nói của khung âm thanh và liệu khung âm thanh có phải là tấn công năng lượng hay không, liệu có lưu trữ dao động phổ tần số trong bộ nhớ hay không. Nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số. Theo phương án thực hiện khác, nếu khung âm thanh hiện tại

là khung hoạt động, và các khung liên tiếp không gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại thuộc tần công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ. Chẳng hạn, nếu khung âm thanh hiện tại là khung hoạt động, và khung trước đó của khung âm thanh hiện tại hoặc khung lịch sử thứ hai của khung âm thanh hiện tại không thuộc tần công năng lượng, dao động phổ tần số của khung âm thanh được lưu trữ trong bộ nhớ dao động phổ tần số; ngược lại dao động phổ tần số không được lưu trữ.

Theo phương án thực hiện, khối phân loại gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khối xác định, được cấu hình để: khi một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

Đối với cách thức cụ thể tính toán dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, đề cập đến phương pháp nêu trên theo phương án thực hiện.

Ngoài ra, thiết bị phân loại tín hiệu âm thanh có thể còn gồm:

khối cập nhật, được cấu hình để cập nhật, theo việc liệu khung thoại là âm nhạc bộ gõ hoặc hoạt động của khung âm thanh lịch sử, các dao động phổ tần số được lưu trữ trong bộ nhớ. Theo phương án thực hiện, khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại thuộc âm nhạc bộ gõ, chỉnh sửa các giá trị của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số. Theo phương án thực hiện khác, khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh trước đó là khung không hoạt động, chỉnh sửa dữ liệu của các dao động phổ tần số khác được lưu trữ trong bộ nhớ ngoại trừ dao động phổ tần số của khung âm thanh hiện tại thành dữ liệu không hiệu dụng; hoặc nếu khung âm thanh hiện tại là khung hoạt động, và ba khung liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ nhất; hoặc nếu khung âm thanh hiện tại là khung hoạt động, và kết quả phân loại lịch sử là tín hiệu âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ hai, trong đó giá trị thứ hai lớn hơn giá trị thứ nhất.

Theo phương án thực hiện này, việc phân loại được thực hiện theo các thống kê lâu dài của các dao động phổ tần số, giá trị lớn nhất dải cao tần phổ tần số, các mức độ tương quan phổ tần số, và các độ dốc năng lượng dư dự báo tuyến tính. Ngoài ra, cả độ vững phân loại lẫn tốc độ nhận dạng phân loại được xem xét; do vậy, không có tham số phân loại, nhưng kết quả tương đối chính xác, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp.

Sáng chế đề xuất của thiết bị phân loại tín hiệu âm thanh theo phương án thực hiện khác, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và gồm:

khối phân khung, được cấu hình để thực hiện xử lý phân khung trên tín hiệu âm thanh đầu vào;

khối thu thập tham số, được cấu hình để thu thập độ dốc năng lượng dự báo tuyến tính và lượng âm thanh phổ tần số của khung âm thanh hiện tại và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, trong đó độ dốc năng lượng dự báo tuyến tính epsP_tilt ký hiệu mức độ mà năng lượng dự báo tuyến tính của tín hiệu âm thanh đầu vào thay đổi khi cấp dự đoán tuyến tính tăng; lượng âm thanh phổ tần số Ntonal ký hiệu số lượng bins tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước; và tỷ lệ ratio_Ntonal_lf của lượng âm thanh phổ tần số trên dải tần số thấp ký hiệu tỷ lệ của lượng âm thanh dải tần số thấp trên lượng âm thanh phổ tần số, trong đó để tính toán cụ thể, tham khảo phần mô tả phương án thực hiện nêu trên;

khối lưu trữ, được cấu hình để lưu trữ độ dốc năng lượng dự báo tuyến tính, lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp; và

khối phân loại, được cấu hình để thu thập các thống kê của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và các thống kê của các lượng âm thanh phổ tần số được lưu trữ một cách riêng rẽ; và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của các độ dốc năng lượng dự báo tuyến tính, các thống kê của các lượng âm thanh phổ tần số, và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp, trong đó các thống kê dữ liệu hiệu dụng đề cập đến giá trị dữ liệu thu được sau hoạt động tính toán được thực hiện trên dữ liệu được lưu trữ trong các bộ nhớ.

Một cách cụ thể, khối phân loại gồm:

khối tính toán, được cấu hình để thu thập biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính được lưu trữ và giá

trị trung bình của các lượng âm thanh phổ tần số được lưu trữ; và

khối xác định, được cấu hình để: khi khung âm thanh hiện tại là khung hoạt động, và một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: biến thiên của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ năm; hoặc giá trị trung bình của các lượng âm thanh phổ tần số lớn hơn ngưỡng thứ sáu; hoặc tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp nhỏ hơn ngưỡng thứ bảy.

Một cách cụ thể, khối thu thập tham số thu thập độ dốc năng lượng dự báo tuyến tính của khung âm thanh hiện tại theo công thức dưới đây:

$$epsP_tilt = \frac{\sum_{i=1}^n epsP(i) \cdot epsP(i+1)}{\sum_{i=1}^n epsP(i) \cdot epsP(i)}$$

trong đó $epsP(i)$ ký hiệu năng lượng dự báo của dự báo tuyến tính cấp i^{th} của khung âm thanh hiện tại; và n là số nguyên dương, ký hiệu cấp dự đoán tuyến tính, và nhỏ hơn hoặc bằng cấp dự đoán tuyến tính lớn nhất.

Một cách cụ thể, khối thu thập tham số được cấu hình để đếm số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, để sử dụng số lượng này làm lượng âm thanh phổ tần số; và khối thu thập tham số được cấu hình để tính toán tỷ lệ của số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 4kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước trên số lượng bin tần số của khung âm thanh hiện tại nằm trên dải tần số từ 0kHz đến 8kHz và có các giá trị lớn nhất bin tần số lớn hơn giá trị định trước, để sử dụng tỷ lệ này làm tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp.

Theo phương án thực hiện nêu trên, tín hiệu âm thanh được phân loại

theo các thống kê lâu dài của các độ dốc năng lượng dự dự báo tuyến tính và các lượng âm thanh phổ tần số và tỷ lệ của lượng âm thanh phổ tần số trên dải tần số thấp; do vậy, hầu như không có tham số, tốc độ nhận dạng tương đối cao, và độ phức tạp tương đối thấp.

Thiết bị phân loại tín hiệu âm thanh nêu trên có thể được kết nối với các bộ mã hóa khác nhau, và mã hóa các tín hiệu khác nhau bằng cách sử dụng các bộ mã hóa khác nhau. Chẳng hạn, thiết bị phân loại tín hiệu âm thanh được kết nối với hai bộ mã hóa, mã hóa tín hiệu thoại bằng cách sử dụng bộ mã hóa dựa trên mô hình tạo thoại (như CELP), và mã hóa tín hiệu âm nhạc bằng cách sử dụng bộ mã hóa dựa trên phép biến đổi (như bộ mã hóa dựa trên MDCT). Đối với định nghĩa và phương pháp thu thập của mỗi tham số cụ thể trong thiết bị nêu trên theo phương án thực hiện, tham khảo phần mô tả liên quan của phương pháp theo phương án thực hiện.

Cùng với phương pháp nêu trên theo phương án thực hiện, sáng chế còn đề xuất thiết bị phân loại tín hiệu âm thanh, và thiết bị có thể được đặt trong thiết bị đầu cuối hoặc thiết bị mạng. Thiết bị phân loại tín hiệu âm thanh có thể được triển khai bởi mạch phần cứng, hoặc được triển khai bởi phần mềm phối hợp với phần cứng. Chẳng hạn, như được thể hiện trên Fig.18, bộ xử lý gọi thiết bị phân loại tín hiệu âm thanh để triển khai phân loại trên tín hiệu âm thanh. Thiết bị phân loại tín hiệu âm thanh có thể thực hiện các phương pháp và quá trình khác trong phương pháp nêu trên theo phương án thực hiện. Đối với các mô đun và chức năng cụ thể của thiết bị phân loại tín hiệu âm thanh, tham khảo phần mô tả liên quan của thiết bị nêu trên theo phương án thực hiện.

Ví dụ về thiết bị 1900 trên Fig.19 là bộ mã hóa. Thiết bị 1900 gồm bộ xử lý 1910 và bộ nhớ 1920.

Bộ nhớ 1920 có thể gồm bộ nhớ ngẫu nhiên, bộ nhớ nhanh, bộ nhớ chỉ đọc, bộ nhớ chỉ đọc lập trình được, bộ nhớ bất biến, thanh ghi, hoặc

tương tự. Bộ xử lý 1910 có thể là bộ xử lý trung tâm (Central Processing Unit, CPU).

Bộ nhớ 1920 được cấu hình để lưu trữ lệnh thực thi được. Bộ xử lý 1910 có thể thực thi lệnh thực thi được được lưu trữ trong bộ nhớ 1920, và được cấu hình để:

Đối với các chức năng và hoạt động khác của thiết bị 1900, tham khảo các quá trình của phương pháp theo các phương án thực hiện trên Fig.3 đến Fig.12, không được mô tả lại ở đây để tránh lặp lại.

Người có kiến thức trung bình trong lĩnh vực có thể hiểu rằng tất cả hoặc một số quá trình của các phương pháp theo các phương án thực hiện có thể được triển khai bởi chương trình máy tính ra lệnh phần cứng liên quan. Chương trình có thể được lưu trữ trong vật lưu trữ máy tính đọc được. Khi chương trình chạy, các quá trình của các phương pháp trong các phương án thực hiện được thực hiện. Vật lưu trữ nêu trên có thể gồm: đĩa từ, đĩa quang, bộ nhớ chỉ đọc, hoặc bộ nhớ truy nhập ngẫu nhiên.

Theo một số phương án thực hiện sáng chế, nên hiểu rằng hệ thống, thiết bị, và phương pháp được bộc lộ có thể được triển khai theo các cách khác. Chẳng hạn, thiết bị được mô tả theo phương án thực hiện chỉ lấy làm ví dụ. Chẳng hạn, việc phân chia khối chỉ là phân chia chức năng logic và có thể là phân chia khác khi triển khai thực. Chẳng hạn, các khối hoặc các thành phần có thể được kết hợp hoặc tích hợp vào hệ thống khác, hoặc một số dấu hiệu có thể bị bỏ qua hoặc không được thực hiện. Ngoài ra, các ghép nối lẫn nhau được đề cập hoặc hiển thị hoặc các ghép nối trực tiếp hoặc các kết nối truyền thông có thể được triển khai bằng cách sử dụng một số giao diện. Các ghép nối gián tiếp hoặc các kết nối truyền thông giữa các thiết bị hoặc các khối có thể được triển khai dưới dạng điện tử, cơ khí, hoặc các dạng khác.

Các khối được mô tả như là các phần riêng rẽ có thể hoặc không thể tách rời về mặt vật lý, và các phần được hiển thị dưới dạng các khối có

thể hoặc không thể là các khối vật lý, có thể được đặt ở một vị trí, hoặc có thể được phân tán trên các khối mạng. Một số hoặc tất cả các khối có thể được lựa chọn theo nhu cầu thực để đạt được các mục đích của các giải pháp theo các phương án thực hiện.

Ngoài ra, các khối chức năng theo các phương án thực hiện sáng chế có thể được tích hợp vào một khối xử lý, hoặc mỗi một khối có thể tồn tại độc lập về mặt vật lý, hoặc hai hoặc nhiều khối được tích hợp vào một khối

Phần nêu trên chỉ là các phương án thực hiện làm ví dụ của sáng chế. Chuyên gia trong lĩnh vực có thể thực hiện các cải biến và biến thể đối với sáng chế mà không xa rời phạm vi sáng chế.

YÊU CẦU BẢO HỘ

1. Phương pháp phân loại tín hiệu âm thanh bao gồm các bước:

xác định (101), theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh;

cập nhật (102), theo việc liệu khung âm thanh là âm nhạc bộ gõ, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số; và

phân loại (103) khung âm thanh hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số.

2. Phương pháp theo điểm 1, trong đó bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số hay không bao gồm bước:

nếu khung âm thanh hiện tại là khung hoạt động, lưu trữ dao động phổ tần số của khung âm thanh hiện tại trong bộ nhớ dao động phổ tần số.

3. Phương pháp theo điểm 1, trong đó bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số hay không bao gồm bước:

nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tấn công năng lượng, lưu trữ dao động phổ tần số của khung âm thanh hiện tại trong bộ nhớ dao động phổ tần số.

4. Phương pháp theo điểm 1, trong đó bước xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập dao động phổ tần số của khung âm thanh hiện tại và lưu trữ dao động phổ tần số trong bộ nhớ dao động phổ tần số hay không bao gồm bước:

nếu khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp bao gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại không thuộc tấn công năng lượng, lưu trữ dao động phổ tần số của khung âm thanh hiện tại này trong bộ nhớ dao động phổ tần số.

5. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 4, trong đó bước cập nhật, theo độ hoạt động của khung âm thanh lịch sử, các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số bao gồm các bước:

nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và khung âm thanh trước đó là khung không hoạt động, chỉnh sửa dữ liệu của các dao động phổ tần số khác được lưu trữ trong bộ nhớ dao động phổ tần số ngoại trừ dao động phổ tần số của khung âm thanh hiện tại thành dữ liệu không hiệu dụng; hoặc

nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và ba khung lịch sử liên tiếp trước khung âm thanh hiện tại đều không phải là khung hoạt động, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ nhất; hoặc

nếu xác định được rằng dao động phổ tần số của khung âm thanh hiện tại được lưu trữ trong bộ nhớ dao động phổ tần số, và kết quả phân loại lịch sử là tín hiệu âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ hai.

6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó bước phân loại khung âm thanh hiện tại như là khung thoải hoặc khung âm nhạc theo các thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số bao gồm các bước:

thu thập giá trị trung bình của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số; và

khi giá trị trung bình thu được của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc.

7. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5 còn bao gồm các bước:

thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kề, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; và độ dốc năng lượng dự dự báo tuyến tính ký hiệu mức độ mà năng lượng dự dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng; và

xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dự dự báo tuyến tính trong bộ nhớ,

trong đó bước phân loại khung âm thanh theo các thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ dao động phổ tần số bao gồm các bước:

thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thể của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

một trong các điều kiện dưới đây được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

8. Thiết bị phân loại tín hiệu âm thanh, trong đó thiết bị được cấu hình để phân loại tín hiệu âm thanh đầu vào, và bao gồm:

khối xác định lưu trữ (1301), được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có thu thập và lưu trữ dao động phổ tần số của khung âm thanh hiện tại hay không, trong đó dao động phổ tần số ký hiệu dao động năng lượng của phổ tần số của tín hiệu âm thanh;

bộ nhớ (1302), được cấu hình để lưu trữ dao động phổ tần số khi khối xác định lưu trữ xuất ra kết quả mà dao động phổ tần số cần được lưu trữ;

khối cập nhật (1304), được cấu hình để cập nhật, theo việc liệu khung âm thanh này là âm nhạc bộ gõ, các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khối phân loại (1303), được cấu hình để phân loại khung âm thanh

hiện tại như là khung thoại hoặc khung âm nhạc theo thống kê của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ.

9. Thiết bị theo điểm 8, trong đó khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

10. Thiết bị theo điểm 8, trong đó khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và khung âm thanh hiện tại không thuộc tần công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

11. Thiết bị theo điểm 8, trong đó khối xác định lưu trữ được cấu hình cụ thể để: khi xác định được rằng khung âm thanh hiện tại là khung hoạt động, và các khung liên tiếp không bao gồm khung âm thanh hiện tại và khung lịch sử của khung âm thanh hiện tại thuộc tần công năng lượng, xuất ra kết quả mà dao động phổ tần số của khung âm thanh hiện tại cần được lưu trữ.

12. Thiết bị theo điểm bất kỳ trong số các điểm từ 8 đến 11, trong đó khối cập nhật được cấu hình cụ thể để: nếu khung âm thanh hiện tại là khung hoạt động, và khung âm thanh trước đó là khung không hoạt động, chỉnh sửa dữ liệu của các dao động phổ tần số khác được lưu trữ trong bộ nhớ ngoại trừ dao động phổ tần số của khung âm thanh hiện tại thành dữ liệu không hiệu dụng; hoặc

nếu khung âm thanh hiện tại là khung hoạt động, và ba khung liên tiếp

trước khung âm thanh hiện tại đều không phải là khung hoạt động, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ nhất; hoặc

nếu khung âm thanh hiện tại là khung hoạt động, và kết quả phân loại lịch sử là tín hiệu âm nhạc và dao động phổ tần số của khung âm thanh hiện tại lớn hơn giá trị thứ hai, chỉnh sửa dao động phổ tần số của khung âm thanh hiện tại thành giá trị thứ hai.

13. Thiết bị theo điểm bất kỳ trong số các điểm từ 8 đến 12, trong đó khối phân loại bao gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của một phần hoặc tất cả dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ trong bộ nhớ; và

khối xác định, được cấu hình để so sánh giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số với điều kiện phân loại âm nhạc; và khi giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số thỏa mãn điều kiện phân loại âm nhạc, phân loại khung âm thanh hiện tại như là khung âm nhạc.

14. Thiết bị theo điểm bất kỳ trong số các điểm từ 8 đến 12, còn bao gồm:

khối thu thập tham số, được cấu hình để thu thập giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, tham số thanh âm, và độ dốc năng lượng dự dự báo tuyến tính của khung âm thanh hiện tại, trong đó giá trị lớn nhất dải cao tần phổ tần số ký hiệu giá trị lớn nhất hoặc độ rõ năng lượng, trên dải cao tần, của phổ tần số của khung âm thanh hiện tại; độ bền vững mức độ tương quan phổ tần số ký hiệu, giữa các khung liên kế, của cấu trúc điều hòa tín hiệu của khung âm thanh hiện tại; tham số thanh âm ký hiệu mức độ tương quan miền thời gian giữa khung âm thanh hiện tại và tín hiệu trước chu kỳ pitch; và độ dốc năng

lượng dư dự báo tuyến tính ký hiệu mức độ mà năng lượng dư dự báo tuyến tính của tín hiệu âm thanh thay đổi khi cấp dự báo tuyến tính tăng; trong đó

khối xác định lưu trữ còn được cấu hình để xác định, theo độ hoạt động tiếng nói của khung âm thanh hiện tại, liệu có lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dư dự báo tuyến tính trong các bộ nhớ hay không;

khối bộ nhớ còn được cấu hình để: khi khối xác định lưu trữ xuất ra kết quả mà giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dư dự báo tuyến tính cần được lưu trữ, lưu trữ giá trị lớn nhất dải cao tần phổ tần số, mức độ tương quan phổ tần số, và độ dốc năng lượng dư dự báo tuyến tính; và

khối phân loại được cấu hình cụ thể để thu thập các thống kê của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, các thống kê của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và các thống kê của dữ liệu hiệu dụng của các độ dốc năng lượng dư dự báo tuyến tính được lưu trữ, và phân loại khung âm thanh như là khung thoại hoặc khung âm nhạc theo các thống kê của dữ liệu hiệu dụng.

15. Thiết bị theo điểm 14, trong đó khối phân loại bao gồm:

khối tính toán, được cấu hình để thu thập giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số được lưu trữ, giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số được lưu trữ, và biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dư dự báo tuyến tính được lưu trữ một cách riêng rẽ; và

khối xác định, được cấu hình để: khi một trong các điều kiện dưới đây

được thỏa mãn, phân loại khung âm thanh hiện tại như là khung âm nhạc; ngược lại phân loại khung âm thanh hiện tại như là khung thoại: giá trị trung bình của dữ liệu hiệu dụng của các dao động phổ tần số nhỏ hơn ngưỡng thứ nhất; hoặc giá trị trung bình của dữ liệu hiệu dụng của giá trị lớn nhất dải cao tần phổ tần số lớn hơn ngưỡng thứ hai; hoặc giá trị trung bình của dữ liệu hiệu dụng của các mức độ tương quan phổ tần số lớn hơn ngưỡng thứ ba; hoặc biến thiên của dữ liệu hiệu dụng của các độ dốc năng lượng dự báo tuyến tính nhỏ hơn ngưỡng thứ tư.

	Khung thứ N trước	...	Khung thứ hai trước	Khung trước	Khung hiện tại	
--	----------------------	-----	------------------------	----------------	-------------------	--

Fig.1

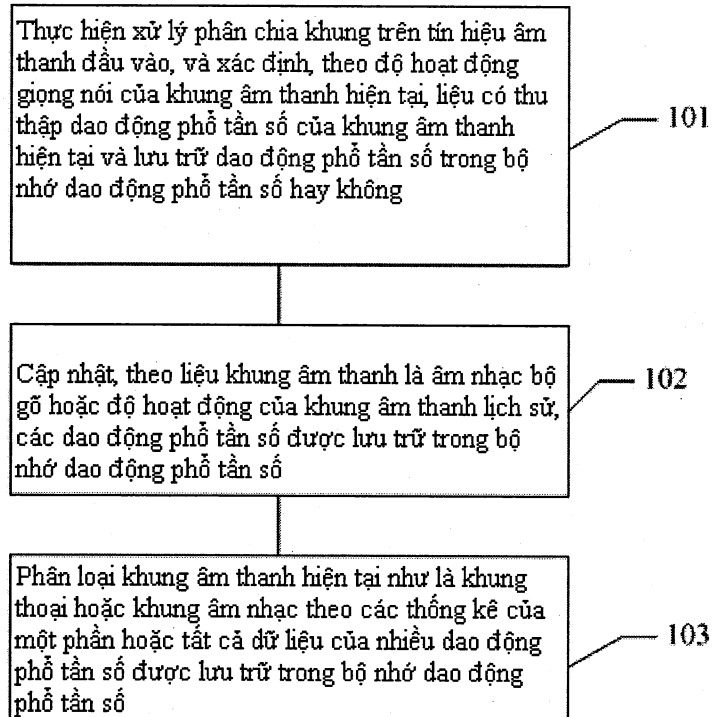


Fig.2

2/11

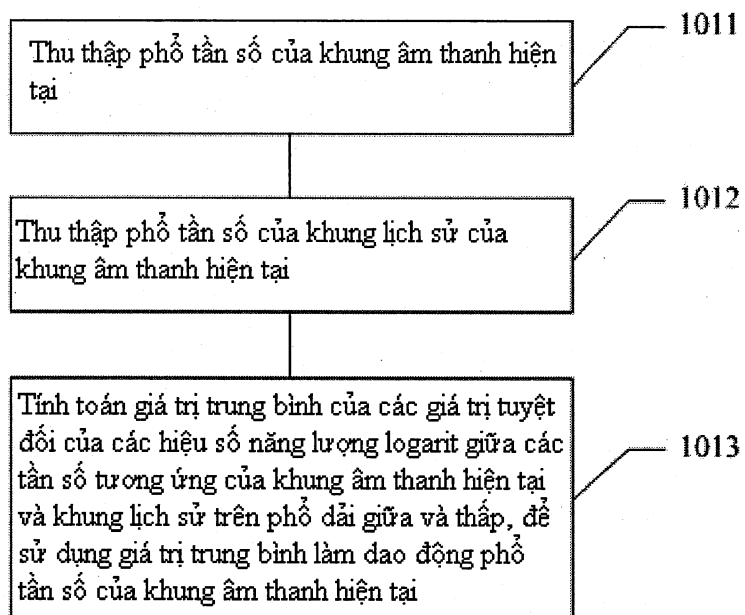


Fig.3

3/11

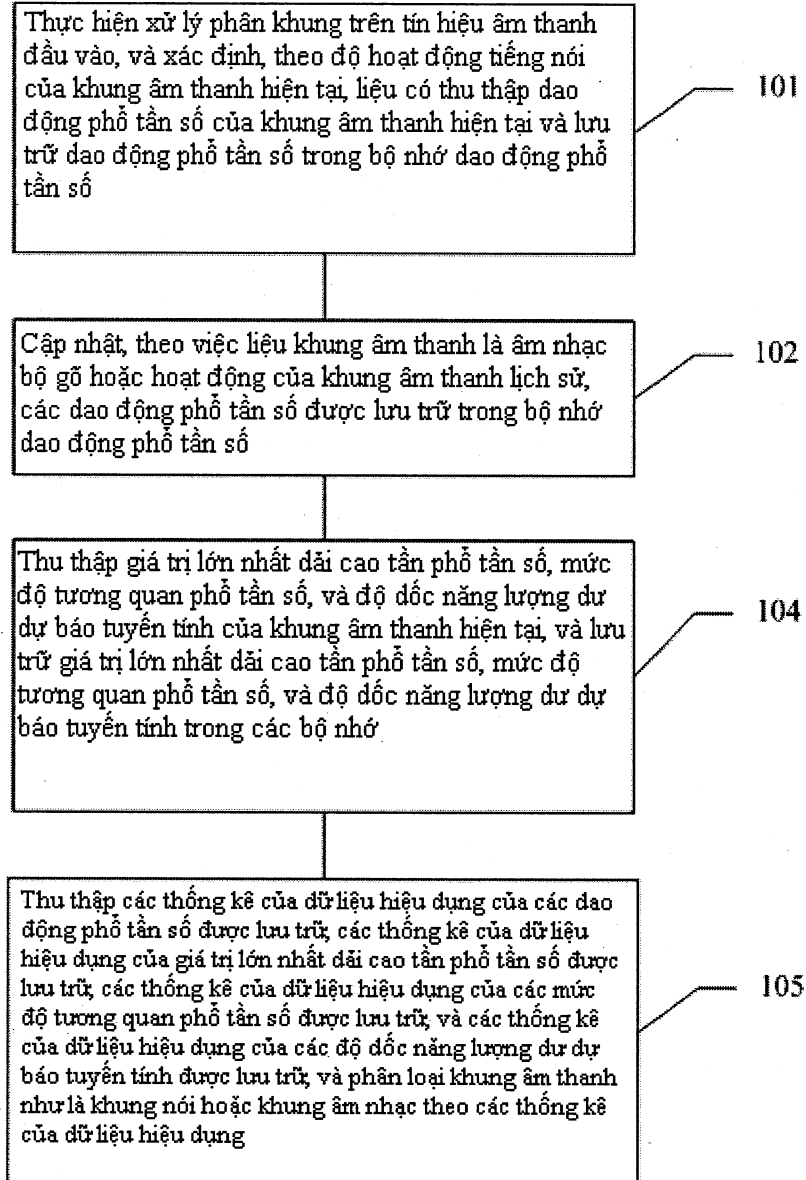


Fig.4

4/11

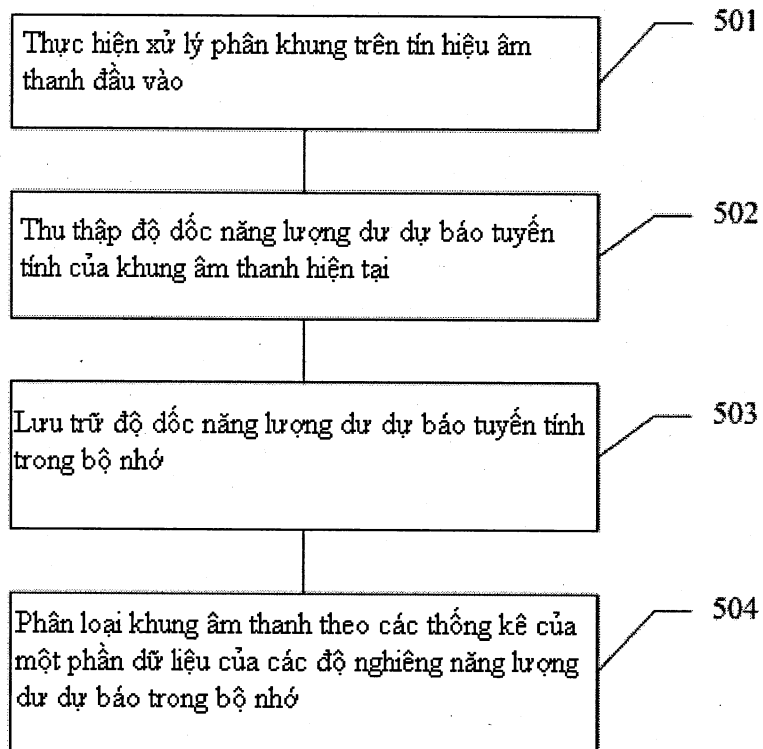


Fig.5

5/11

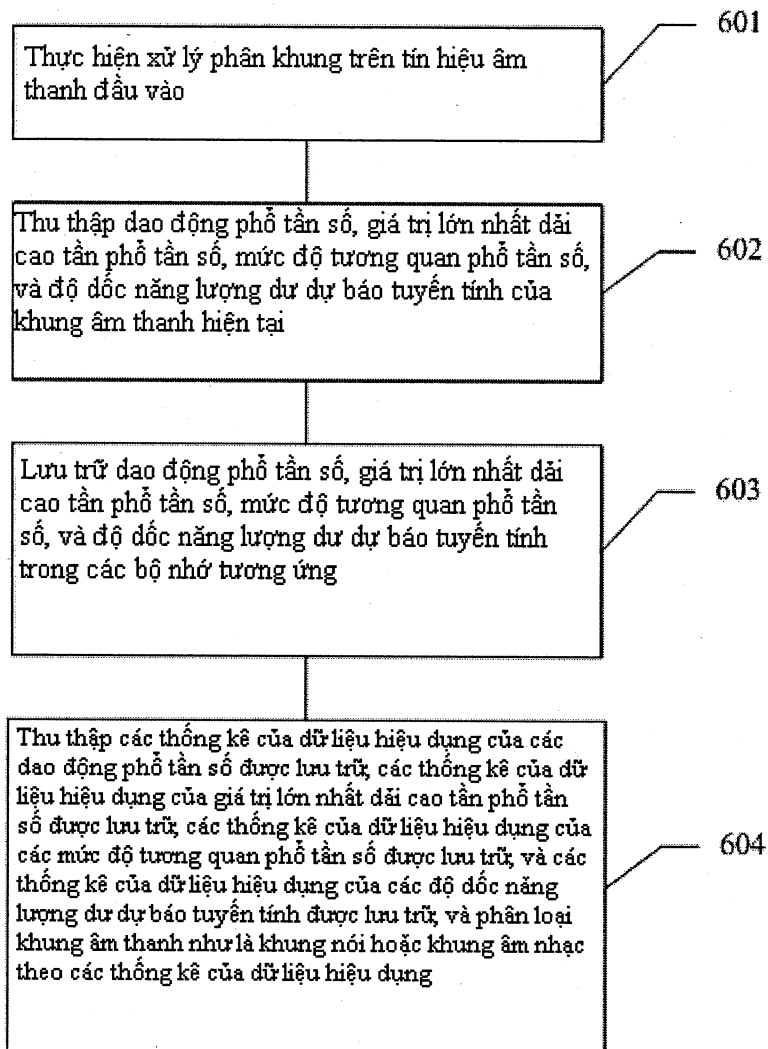


Fig.6

6/11

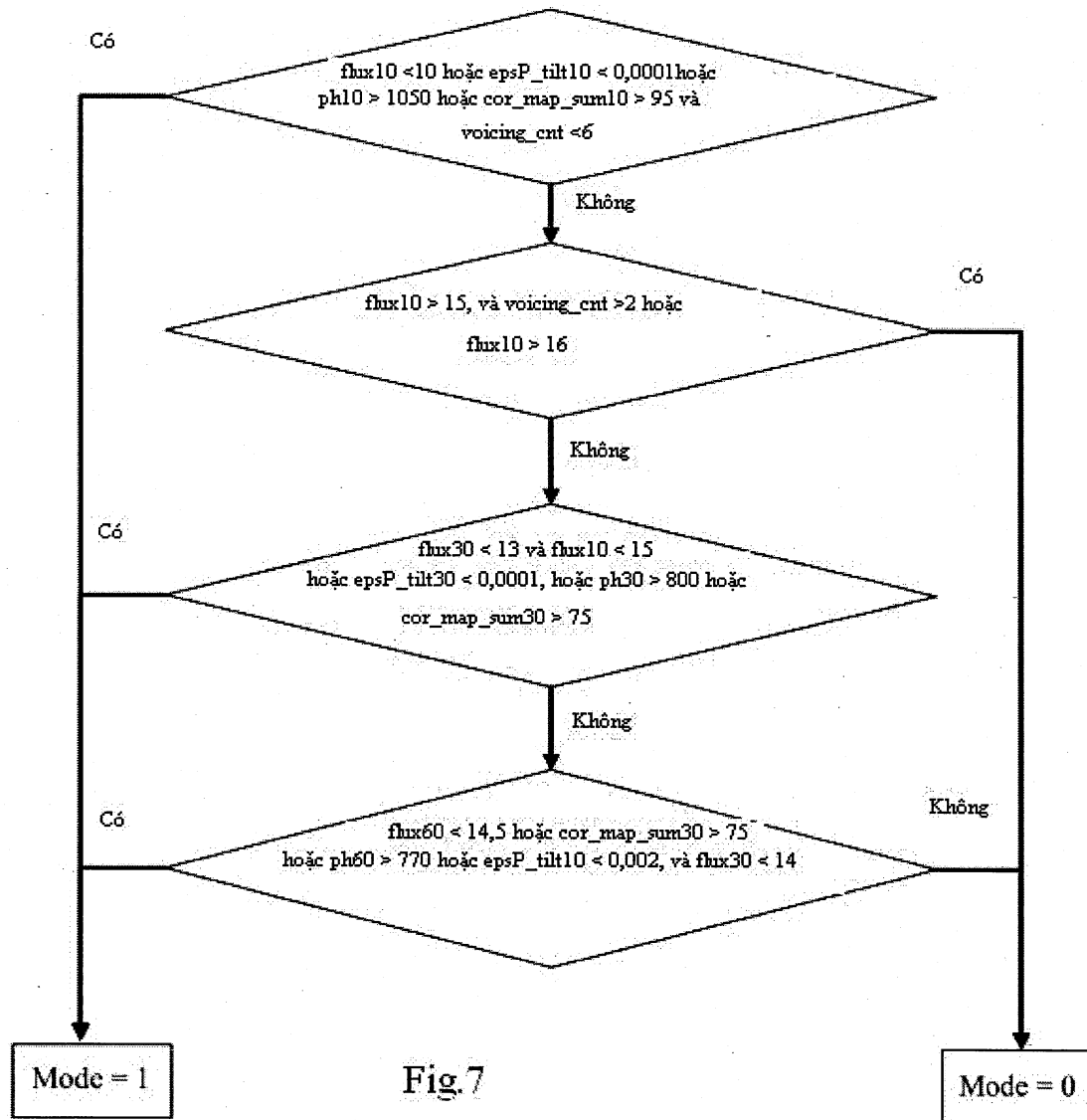


Fig.7

7/11

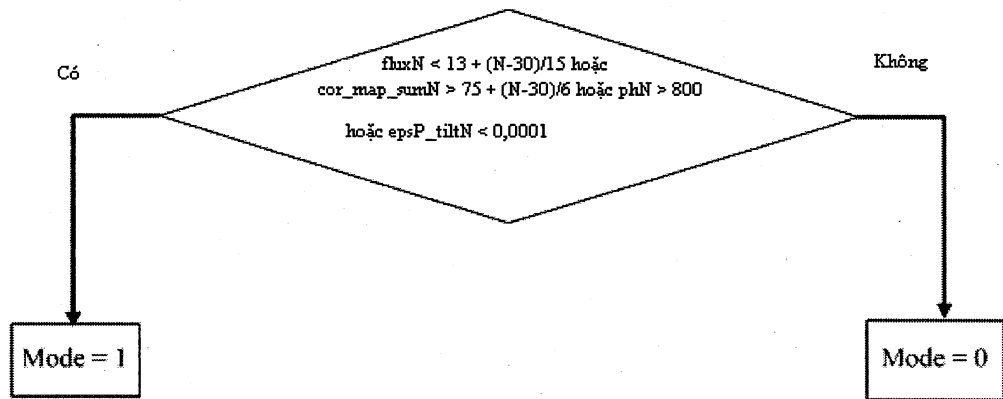


Fig. 8

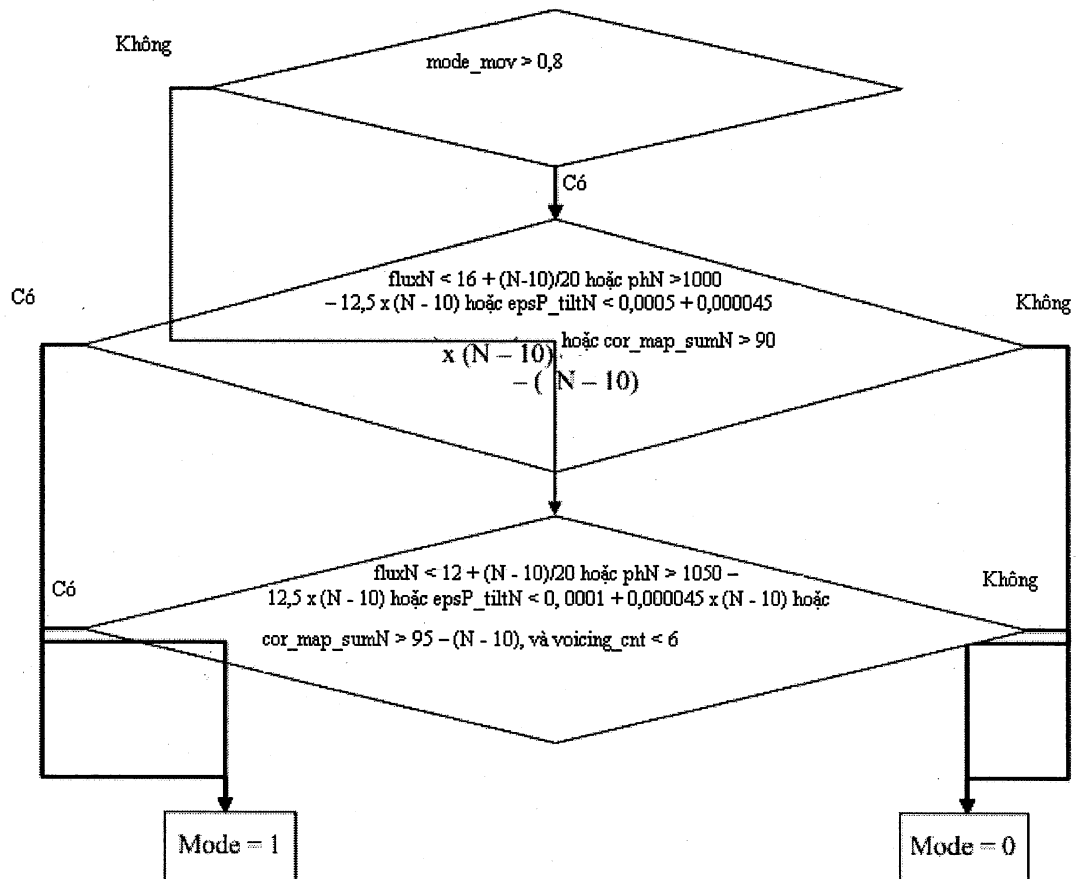
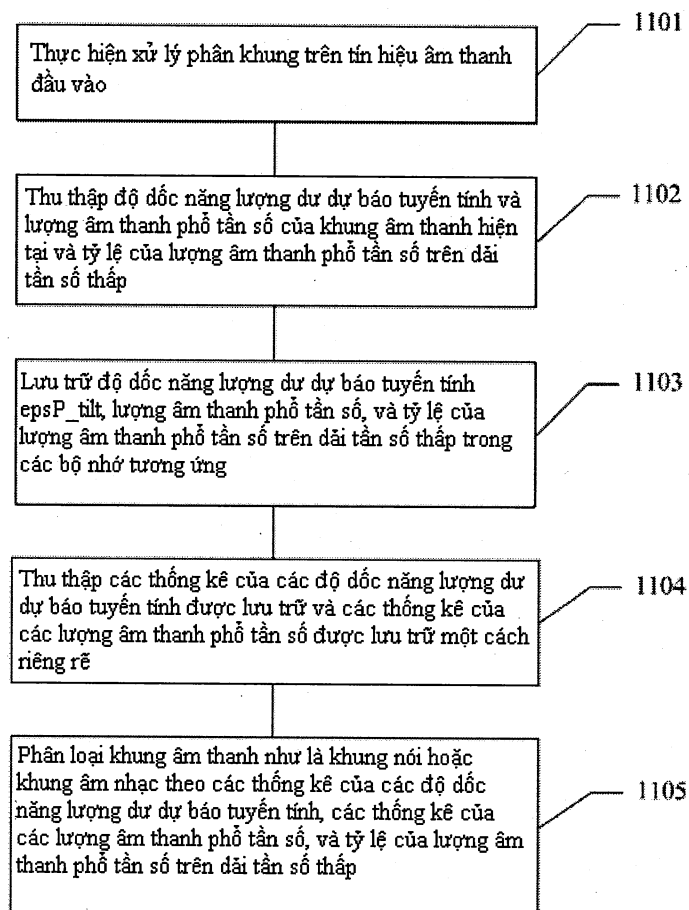
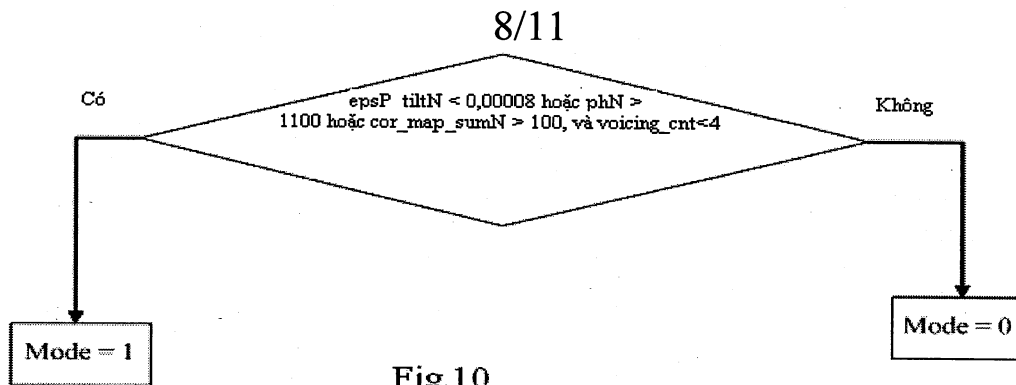
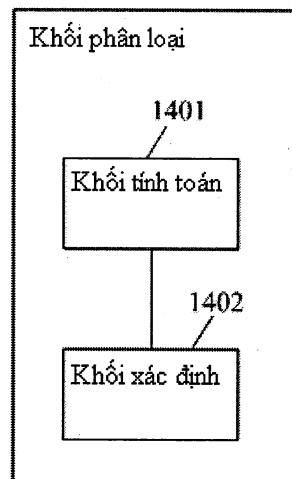
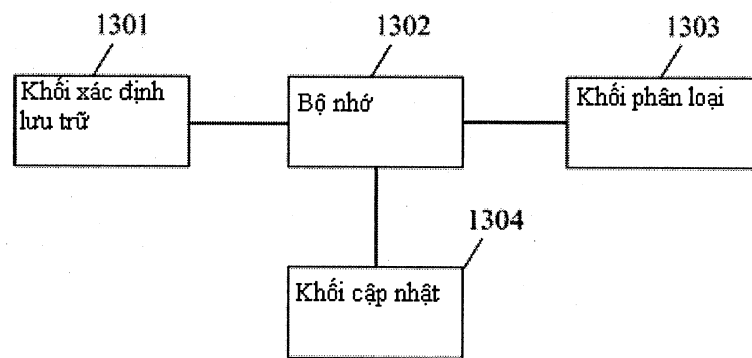
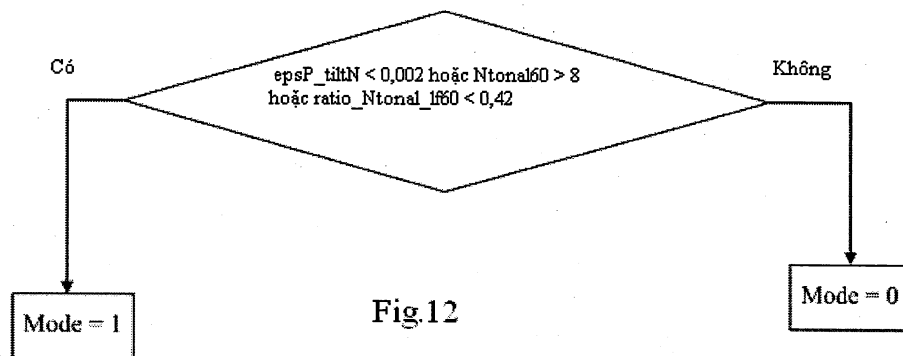


Fig. 9



9/11



10/11

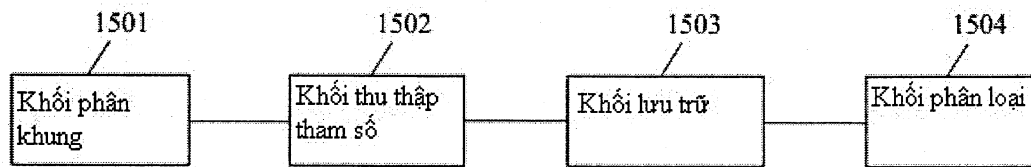


Fig.15

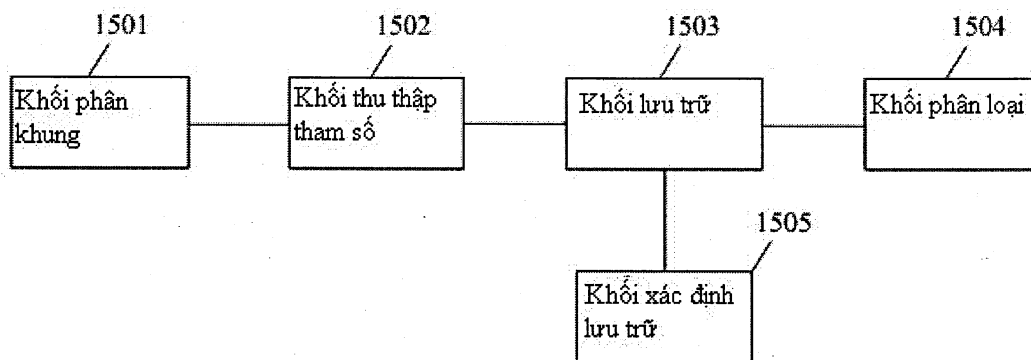


Fig.16

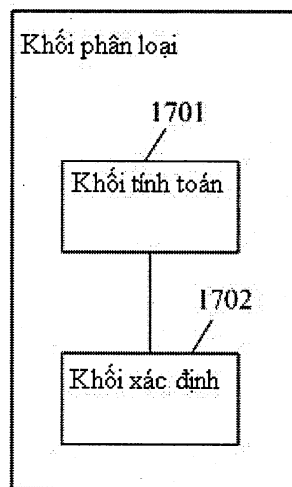


Fig.17

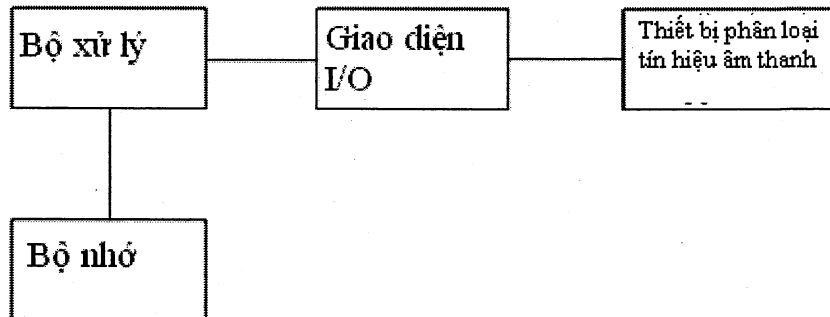


Fig.18

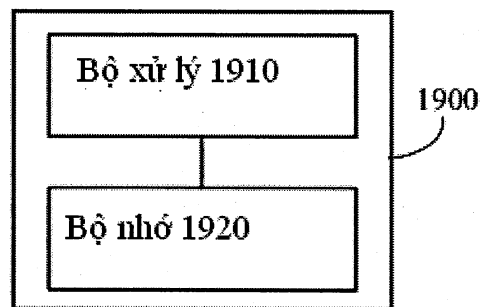


Fig.19